


# The Genetic and Metabolic Basis of Kidney Stone Disease



Catherine Elizabeth Lovegrove

Balliol College

*University of Oxford*

A thesis presented for the degree of

*Doctor of Philosophy in Surgical Sciences*

Hilary 2025

# Abstract

Kidney stone disease (KSD) is a common, heritable, frequently recurrent condition affecting approximately 10–20% of people. Patients may be asymptomatic or experience severe pain, haematuria, urinary tract infections, and, in extreme cases, multi-organ failure. KSD and its treatments reduce health-related quality of life and its rising prevalence will cost the National Health Service over £300 million annually in England by 2030. Despite their relevance, the mechanisms underlying stone formation remain poorly understood, limiting effective preventive strategies. In this thesis, I investigate the genetic architecture of KSD to uncover molecular mechanisms and therapeutic targets for the disease.

I conduct the largest genetic association studies of KSD to date, integrating combined-sex, and sex-specific analyses, European- and trans-ancestry data, and common and rare genetic variants. These studies identify novel genetic associations with KSD and highlight genes linked to mineral metabolism and renal structural disorders.

Using Mendelian randomisation (MR) and colocalisation analyses, I reveal a shared genetic architecture between adiposity and KSD, implicating a causal *GIPR* missense variant which I propose alters renal tubular osteopontin secretion, affecting risk of stone formation. Drug-target MR analyses suggest that modulating gastric inhibitory polypeptide receptor pathways may reduce KSD risk.

To investigate the mechanisms linking mineral metabolism and renal cysts to KSD, I integrate multiomic and chromatin interaction data with MR and

colocalisation analyses. I identify three biological pathways contributing to approximately 12–17% of KSD cases: impaired DGK$\delta$-mediated calcium-sensing receptor signalling; increased NaPi-IIa-mediated renal phosphate excretion; and defective 24-$\alpha$-hydroxylase-mediated 1,25-dihydroxyvitamin D inactivation. I further establish a causal, bidirectional relationship between non-polycystic renal cysts and KSD, implicating *TBX2*-mediated effects.

These findings support a "lithogenic triad" of urine supersaturation, stasis, and tubular epithelial injury in KSD pathogenesis. By integrating multiple analytical approaches with distinct and orthogonal sources of bias , I provide insights into causal mechanisms and therapeutic targets for KSD, laying the foundation for future translational research.

# Acknowledgements

I am deeply grateful to my supervisors, Dr. Sarah Howles, Professor Dominic Furniss, and Professor Michael Holmes. You have provided invaluable support, guidance, and freedom to pursue my research questions, with the occasional nudge back on track when needed. Beyond your supervision, your mentorship has shaped my academic growth and my development as a clinical academic and an individual. It has been a privilege to learn from you—thank you.

Thank you, Dr. Michelle Goldsworthy, Professor Rajesh Thakker, Professor Fadil Hannan, Dr. Michael Ng, and Professor John Sayer for equipping me with the resources and collaborations essential for my DPhil studies.

I am very grateful to the Medical Research Council for funding my studies and facilitating my training.

To fellow clinical academics- Becky, Rob, Alex, Max, Rachel, Akira, Ayesha, and Yos- thank you for your counsel, perspective, and camaraderie. The future of clinical academia is bright with you in it.

Finally, I would like to thank my friends, many of whom have joined me cycling, running, or on a mountain, and others who have supported me in different ways. Your encouragement, perspective, and laughter when I needed them most have been invaluable throughout this journey. And to my loving family, your constant, unwavering support has been an indescribable strength—thank you. Without you, there would be no thesis.

# Declaration

The work in this thesis is my own original work, except as detailed in this declaration or specified in the text.

**Chapter 2 Section 2.6**: Dr. Sarah Howles prepared snRNA- and snATAC-seq data.

**Chapter 2 Section 2.7**: Dr. Michelle Goldsworthy performed Capture-C experiments.

**Chapter 6, Appendices D, E, G**: Dr. Sarah Howles identified inclusion and exclusion codes for analyses using *All of Us* data.

During the course of my DPhil, I have published several articles of which material is included in this thesis with permission from co-authors:

- **Lovegrove CE**, Howles SA, Furniss D, et al. Causal inference in health and disease: a review of the principles and applications of Mendelian randomization. Journal of Bone and Mineral Research. 2024 Aug 21; 10.1093/jbmr/zjae136.
- **Lovegrove CE**, Bešević J, Wiberg A, et al. Central Adiposity Increases Risk of Kidney Stone Disease through Effects on Serum Calcium Concentrations. Journal of the American Society of Nephrology. 2023 Dec; 10.1681/ASN.0000000000000238.
- Geraghty R, **Lovegrove CE** , Howles S, et al. Role of Genetic Testing in Kidney Stone Disease: A Narrative Review. Curr Urol Rep. 2024 Dec; 10.1007/s11934-024-01225-5.

I would like to acknoelwdge the participants and investigators of the UK Biobank, FinnGen, and *All of Us* studies.

# List of Abbreviations

| | |
|---|---|
| 3C | Chromosome conformation capture |
| ABC | ATP-binding cassette |
| ADPKD | Autosomal dominant polycystic kidney disease |
| ALPL | Alkaline phosphatase |
| APRT | Adenine phosphoribosyltransferas2 |
| AQP1 | Aquaporin-1 |
| $b^{0,+}$AT | b(0,+)-type amino acid transporter 1 |
| BCAS3 | Breast carcinoma-amplified sequence 3 microtubule-associated cell migration factor |
| BCRP | Breast cancer resistance protein |
| BMI | Body mass index |
| bp | Base-pairs |
| CAD | Coronary artery disease |
| CADD | Combined Annotation Dependent Depletion |
| CaSR | Calcium sensing receptor |
| CHD | Coronary heart disease |
| CI | Confidence interval |
| CKD | Chronic kidney disease |

| | |
|---|---|
| ClC-5 | Chloride channel protein 5 |
| ClC-Kb | Chloride channel protein ClC-Kb |
| CT | Computed tomography |
| DGK$\delta$ | Diacylglycerol kinase $\delta$ |
| DNA | Deoxyribonucleic acid |
| DPP4i | Dipeptidyl peptidase IV inhibitors |
| dRTA | Distal renal tubular acidosis |
| eQTL | Expression quantitative trait loci |
| ExWAS | Exome-wide association study |
| FGF | Fibroblast growth factor |
| FGF23 | Fibroblast growth factor 23 |
| FHH | Familial hypocalciuric hypercalcemia |
| GCTA | Genome-Wide Complex Trait Analysis |
| GIANT | Genetic Investigation of Anthropometric Traits |
| GIP | Gastric inhibitory polypeptide |
| GIPR | Gastric inhibitory polypeptide receptor |
| GLP1 | Glucagon-like peptide-1 |
| GLP1-RA | Glucagon-like peptide-1 receptor agonist |
| GLUT9 | Glucose transporter type 9 |
| GR/HPR | Glyoxylate reductase/hydroxypyruvate reductas |
| GRK | G protein-coupled receptor kinase 6 |

| | |
|---|---|
| GWAS | Genome-wide association study/studies |
| HEK293 cells | Human embryonic kidney 293 cells |
| HGPRT | Hypoxanthine-guanine phosphoribosyltransferase |
| HK2 cells | Human kidney 2 cells |
| HOG | 4-hydroxy-2-oxoglutarate aldolase |
| HOGA1 | 4-hydroxy-2-oxoglutarate aldolase |
| HPO | Human phenotype ontology |
| ICD | International Classification of Diseases |
| IV | Instrumental variable |
| Kbp | Kilobase pairs |
| KSD | Kidney stone disease |
| LD | Linkage disequilibrium |
| LDL | Low-density lipoprotein |
| LDSC | LD score regression |
| LOCO | Leave-one-chromosome-out |
| LOVO | Leave-one-variant-out |
| MAC | Minor allele count |
| MAF | Minor allele frequency |
| Mbp | Megabase pairs |
| MR | Mendelian randomisation |
| MRI | Magnetic resonance imaging |

| | |
|---|---|
| NaPi-IIa | Sodium-dependent phosphate transport protein 2A |
| NaPi-IIb | Sodium-dependent phosphate transport protein 2B |
| NaPi-IIc | Sodium-dependent phosphate transport protein 2C |
| NHE3 | Sodium-hydrogen exchanger 3 |
| NHERF1 | $Na^+/H^+$ exchanger regulatory factor |
| NKCC2 | Sodium-potassium-chloride cotransporter 2 |
| NPHLOP | Hypophosphataemia with nephrolithiasis or osteoporosis |
| OPCS | Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures |
| OPN | Osteopontin |
| OR | Odds ratio |
| PCA | Principal component analysis |
| PCR | Polymerase chain reaction |
| pcRNA | Positionally conserved long non-coding RNA |
| PCSK9 | Proprotein convertase subtilisin/kexin type 9 |
| PP | Posterior probability |
| PRLEID1 | PRELI domain-containing protein 1 |
| PRPP | Phosphoribosylpyrophosphate |
| PRPPS | Phosphoribosyl pyrophosphate synthase |
| PRS | Polygenic risk score/scores |
| PTH | Parathyroid hormone |
| PTV | Protein-truncating variants |

| | |
|---|---|
| QC | Quality control |
| rBAT | b(0,+)-type amino acid transporter-related heavy chain |
| RCT | Randomised control trial |
| REVEL | Rare Exome Variant Ensemble Learner |
| ROMK | Renal outer medullary potassium channel |
| RPTEC | Renal proximal tubule epithelial cells |
| SAT-1 | Sulphate anion transporter 1 |
| SD | Standard deviation |
| SGLT2i | Sodium-glucose co-transporter-2 inhibitors |
| SIFT | Sorting intolerant from tolerant |
| SKAT-O | Sequence Kernel Association Test- |
| SLC26A6 | Solute carrier family 26 member 6 |
| snATAC-seq | Single nucleus assay for transposase-accessible chromatin sequencing |
| SNP | Single nucleotide polymorphism |
| snRNA-seq | Single nucleus ribonucleic acid sequencing |
| TAD | Topologically associating domain |
| TFAP2B | Transcription factor AP-2-$\beta$ |
| TRIAP1 | TP53-regulated inhibitor of apoptosis 1 |
| TRPV5 | Transient receptor potential cation channel subfamily V |
| TSS | Transcription start site |
| UK | United Kingdom |

| | |
|---|---|
| URAT1 | Urate anion exchanger 1 |
| USA | United States of America |
| USP40 | Ubiquitin carboxyl-terminal hydrolase 40 |
| WDR72 | WD Repeat Domain 72 |
| WES | Whole exome sequencing |
| WHO | World Health Organisation |
| WHR | Waist-to-hip ratio |
| $\lambda$GC | Genomic control $\lambda$ |

# Contents

**Abstract** **ii**

**Acknowledgements** **iv**

**Declaration** **v**

**Abbreviations** **vi**

**List of Figures** **xix**

**List of Tables** **xxiv**

**1 Introduction** **1**

1.1 Kidney stone disease . . . . . 1

1.1.1 Definition . . . . . 1

1.1.2 Epidemiology . . . . . 1

1.1.3 Classification of kidney stones . . . . . 2

1.1.4 Principles of kidney stone formation . . . . . 5

1.1.5 Principles of calcium and phosphate metabolism . . . . . 8

1.1.6 Aetiology . . . . . 12

1.1.7 Diagnosis . . . . . 44

1.1.8 Treatment . . . . . 46

1.1.9 Summary of kidney stone disease . . . . . 48

1.2 Genetic epidemiology . . . . . 49

1.2.1 Definition . . . . . 49

1.2.2 Genetic architecture and genome-wide association studies . . 50

1.2.3 Exome-wide association studies . . . . . 55

1.2.4 Mendelian randomisation . . . . . 57

1.2.5 Colocalisation . . . . . 67
1.3 Single-cell multi-omics . . . . . 71
1.4 Chromosome conformation capture . . . . . 73
1.5 Thesis aims . . . . . 75

**2 Methods** **76**

2.1 Cohorts . . . . . 76
2.1.1 UK Biobank . . . . . 76
2.1.2 FinnGen . . . . . 78
2.1.3 Biobank Japan . . . . . 80
2.1.4 *All of Us* . . . . . 81
2.2 Genome-wide association studies . . . . . 82
2.2.1 Genome-wide association studies in the UK Biobank . . . . 82
2.2.2 Genome-wide association studies in the FinnGen study . . . 84
2.2.3 Genome-wide association studies in Biobank Japan . . . . . 85
2.2.4 Meta-analysis of genome-wide association studies . . . . . 86
2.2.5 Defining index association signals . . . . . 86
2.2.6 Estimating heritability . . . . . 87
2.2.7 Gene and gene set analyses . . . . . 87
2.3 Exome-wide association studies . . . . . 88
2.4 Mendelian randomisation . . . . . 91
2.5 Colocalisation . . . . . 93
2.6 Single-cell analyses . . . . . 94
2.6.1 snRNA-seq preparation . . . . . 94
2.6.2 snATAC-seq preparation . . . . . 95
2.6.3 Predicting *cis*-regulatory interactions . . . . . 96
2.6.4 Calculating gene activity scores . . . . . 97
2.7 Chromosome conformational capture . . . . . 98

**3 The role of genetic variants in kidney stone disease** **99**

3.1 Introduction . . . 99
3.1.1 Aim and Objectives . . . 99
3.2 Methods . . . 100
3.2.1 Cohorts . . . 100
3.2.2 Case ascertainment . . . 101
3.2.3 Genome-wide association studies . . . 101
3.2.4 Exome-wide association study . . . 102
3.3 Results . . . 103
3.3.1 Genome-wide association study of kidney stone disease in the UK Biobank . . . 103
3.3.2 European ancestry GWAS meta-analysis . . . 108
3.3.3 Trans-ancestry GWAS meta-analysis . . . 117
3.3.4 Heritability estimates . . . 128
3.3.5 Gene and gene set based analyses . . . 129
3.3.6 Exome-wide association study of kidney stone disease in the UK Biobank . . . 134
3.4 Discussion and conclusions . . . 140
3.4.1 Genome-wide association studies . . . 140
3.4.2 Downstream bioinformatic analyses . . . 142
3.4.3 Exome-wide association study of kidney stone disease in the UK Biobank . . . 144
3.4.4 Conclusion . . . 146

**4 The role of adiposity in kidney stone disease 148**

4.1 Introduction . . . 148
4.1.1 Definitions . . . 148
4.1.2 The relevance of adiposity to kidney stone disease . . . 150
4.1.3 Aim and Objectives . . . 151
4.2 Methods . . . 152
4.2.1 Data sources . . . 152

4.2.2 Mendelian randomisation . . . . . 152
4.2.3 Colocalisation . . . . . 153
4.2.4 Drug-target Mendelian randomisation . . . . . 153
4.3 Results . . . . . 153
4.3.1 Genetic variants increase the risk of kidney stone disease via alterations in body mass index or waist-to-hip ratio . . . . . 153
4.3.2 A candidate causal variant in *GIPR* has functional effects . . 163
4.3.3 Modulating adiposity at *GIPR* may prevent kidney stone disease 164
4.4 Discussion and conclusions . . . . . 165
4.4.1 rs1800437- a candidate causal variant in *GIPR* . . . . . 166
4.4.2 Implications for clinical practice . . . . . 168
4.4.3 Drug repurposing for kidney stone disease . . . . . 170
4.4.4 Conclusion . . . . . 172

**5 The role of mineral metabolism in kidney stone disease 173**
5.1 Introduction . . . . . 173
5.1.1 Disorders of calcium and phosphate homeostasis associated with kidney stone disease . . . . . 173
5.1.2 Causal effects of calcium and phosphate in kidney stone disease 173
5.1.3 The unmet need in kidney stone disease . . . . . 174
5.1.4 Aim and Objectives . . . . . 175
5.2 Study design . . . . . 176
5.3 Methods . . . . . 177
5.3.1 Data sources . . . . . 177
5.3.2 Mendelian randomisation . . . . . 177
5.3.3 Colocalisation . . . . . 178
5.3.4 Single-cell Multiomics . . . . . 178
5.3.5 Chromatin conformational capture . . . . . 179
5.3.6 Drug-target Mendelian randomisation . . . . . 179
5.4 Results . . . . . 180

5.4.1 Genetic variants increase the risk of kidney stone disease via alterations in serum calcium or phosphate concentrations . . 180
5.4.2 The functional relevance of rs838717 . . . . . 192
5.4.3 The functional relevance of rs10051765 . . . . . 201
5.4.4 The functional relevance of rs6127099 . . . . . 211
5.4.5 The clinical relevance of rs838717, rs10051765, and rs6127099 in treating kidney stone disease . . . . . 218
5.5 Discussion and conclusions . . . . . 222
5.5.1 Implications of DGK$\delta$-mediated impaired calcium-sensing receptor signalling for kidney stone disease . . . . . 224
5.5.2 Implications of NaPi-IIa-mediated increased renal phosphate excretion for kidney stone disease . . . . . 225
5.5.3 Implications of 24-$\alpha$-hydroxylase-mediated impaired inactivation of 1,25-dihydroxyvitamin D for kidney stone disease . . . . . 227
5.5.4 Limitations . . . . . 228
5.5.5 Conclusion . . . . . 228

**6 The role of renal cysts in kidney stone disease 230**
6.1 Introduction . . . . . 230
6.1.1 Gene and gene set associations between renal cysts and kidney stone disease . . . . . 230
6.1.2 Aim and Objectives . . . . . 231
6.2 Methods . . . . . 232
6.2.1 Cohorts . . . . . 232
6.2.2 Renal cyst case ascertainment . . . . . 232
6.2.3 Observational analyses in the UK Biobank and *All of Us* study 232
6.2.4 A genome-wide association study of renal cysts in the UK Biobank . . . . . 233
6.2.5 Colocalisation and Mendelian randomisation analyses . . . . 233

6.2.6 Single-cell multiomics . . . . . 234
6.3 Results . . . . . 234
6.3.1 Renal cysts are associated with kidney stone disease . . . . . 234
6.3.2 Genome-wide association study of renal cysts in the UK Biobank 235
6.3.3 Renal cysts and kidney stone disease colocalise at the *BCAS3* locus . . . . . 237
6.3.4 Evidence for a bidirectional causal relationship between renal cysts and kidney stone disease . . . . . 241
6.3.5 The *BCAS3* locus mediates the causal effect of renal cysts on kidney stones . . . . . 243
6.3.6 The functional relevance of rs9895661 . . . . . 245
6.4 Discussion and conclusions . . . . . 252
6.4.1 The bidirectional, causal relationship between renal cysts and kidney stone disease . . . . . 253
6.4.2 The role of *TBX2-AS1* in renal cyst formation . . . . . 254
6.4.3 The clinical relevance of a shared genetic architecture between renal cysts and kidney stone disease . . . . . 255
6.4.4 Limitations . . . . . 256
6.4.5 Conclusions . . . . . 257

**7 Discussion** **258**

7.1 Summary . . . . . 258
7.2 The relevance of results from this thesis . . . . . 260
7.3 Key limitations . . . . . 261
7.3.1 Ancestral representation . . . . . 261
7.3.2 Potential bias in MR and colocalisation studies . . . . . 262
7.3.3 Tissue specificity . . . . . 262
7.4 Future direction . . . . . 262
7.5 Conclusions . . . . . 263

**References** **265**

**Appendices**

**A Inclusion criteria for kidney stone disease association analyses in the UK Biobank** **349**

**B Inclusion criteria for renal cyst association analyses in the UK Biobank** **353**

**C Exclusion criteria for renal cyst association analyses in the UK Biobank** **355**

**D Inclusion criteria for renal cyst association analyses in *All of Us*** **356**

**E Exclusion criteria for renal cyst association analyses in *All of Us*** **357**

**F Imaging codes for renal cyst observational association analyses in the UK Biobank** **358**

**G Imaging codes for renal cyst observational association analyses in *All of Us*** **359**

**Scientific outputs** **361**

# List of Figures


1.1 Crystals constituting kidney stones. . . . . 3

1.2 Principles of kidney stone formation. . . . . 6

1.3 Randall's plaques and Randall's plugs- foci for kidney stone formation. 7

1.4 Abnormalities of purine metabolism causing hyperuricosuria. . . . . 20

1.5 Enzyme deficiencies in type 1, 2, and 3 primary hyperoxaluria. . . . 32

1.6 Pathophysiology of Bartter syndrome. . . . . 40

1.7 Challenges in observational studies. . . . . 57

1.8 Analogy of a randomised control trial and Mendelian randomisation. 58

1.9 Assumptions in Mendelian randomisation. . . . . 59

1.10 Mediation Mendelian randomisation. . . . . 66

1.11 Genetic colocalisation. . . . . 67

3.1 Results of a genome-wide association study of kidney stone disease in the UK Biobank. . . . . 103

3.2 Quantile-quantile plot of observed vs. expected p-values for a genome-wide association study of kidney stone disease in the UK Biobank. . . . . 104

3.3 Results of genome-wide association studies of kidney stone disease in males and females in the UK Biobank. . . . . 106

3.4 Quantile-quantile plot of observed vs. expected p-values for male- and female-sex genome-wide association studies of kidney stone disease in the UK Biobank. . . . . 106

3.5 Manhattan plot for a kidney stone disease genome-wide association study meta-analysis in the UK Biobank and FinnGen R12 Study. . 109

3.6 Quantile-quantile plot of observed vs. expected p-values for a kidney stone disease genome-wide association study meta-analysis in the UK Biobank and FinnGen R12 Study. . . . . 109
3.7 Manhattan plot for a kidney stone disease genome-wide association study meta-analysis in the UK Biobank, FinnGen R12 Study, and Biobank Japan. . . . . 118
3.8 Quantile-quantile plot of observed vs. expected p-values for a kidney stone disease genome-wide association study meta-analysis in the UK Biobank, FinnGen R12 Study, and Biobank Japan. . . . . 119
3.9 Manhattan plot for an exome-wide association study of kidney stone disease in the UK Biobank. . . . . 134
3.10 Quantile-quantile plot of observed vs. expected p-values for exome-wide association study of kidney stone disease in the UK Biobank. . . . . 135
3.11 Comparisons of rare variant effect sizes with common effects. . . . . 136
3.12 Exome-wide gene-based association testing for kidney stone disease in 11,225 cases and 381,369 controls in the UK Biobank, based on rare variants. . . . . 139

4.1 Forest plot demonstrating regional Mendelian randomisation estimates for independent body mass index-associated variants (±500 kilobase pairs) on odds of kidney stone disease. . . . . 157
4.2 Forest plot demonstrating regional Mendelian randomisation estimates for independent waist-to-hip ratio-associated variants (±500 kilobase pairs) on odds of kidney stone disease. . . . . 157
4.3 Locus-zoom plots of 1 megabase pair regions with multi-trait colocalisation between body mass index, waist-to-hip ratio, and kidney stone disease. . . . . 160

4.4 Forest plot demonstrating drug-target Mendelian randomisation estimates for the effects of modulating body mass index and waist-to-hip ratio via *GIPR*. . . . . 165
4.5 Antibody staining from The Human Protein Atlas shows *GIPR* expression in normal human renal tissue. . . . . 167
4.6 A hypothesis for the role of *GIPR* in obesity and risk of kidney stone disease. . . . . 168

5.1 Identifying genetic variants predisposing to increased risk of kidney stone disease via alterations in serum calcium and phosphate concentrations. . . . . 176
5.2 Forest plot demonstrating regional Mendelian randomisation estimates for independent serum calcium and phosphate index-associated variants (±500 kilobase pairs) on odds of kidney stone disease. . . . . 188
5.3 Locus-zoom plots of 1 megabase pair regions with multi-trait colocalisation and causal effects between mineral metabolism traits and kidney stone disease. . . . . 190
5.4 Pearson correlation coefficients of gene activity, inferred from rs838717 coaccessibility, and RNA expression of *INPP5D*, *NGEF*, *DGKD*, *UGT1A6*, and *USP40*. . . . . 195
5.5 *Cis*-coaccessibility of rs838717 and protein-coding gene promoters correlates with RNA expression in renal cell lines. . . . . 196
5.6 Cature-C: chromatin interactions of rs838717 and gene promoters across three renal cell lines. . . . . 199
5.7 Pearson correlation coefficients of gene activity, inferred from rs10051765 coaccessibility, and RNA expression of *SLC34A1*, *GRK6*, *FAM193B*, *B4GALT7*, and *PRELID1*. . . . . 203
5.8 *Cis*-coaccessibility of rs10051765 and protein-coding gene promoters correlates with RNA expression in renal cell lines. . . . . 204

5.9 Cature-C: chromatin interactions of rs10051765 and gene promoters across three renal cell lines. . . . . 207
5.10 Implications of increased NaPi-IIa-mediated phosphate leak. . . . . 209
5.11 Pearson correlation coefficients of gene activity, inferred from rs6127099 coaccessibility, and RNA expression of *CYP24A1* and *DOK5*. . . . . 213
5.12 *Cis*-coaccessibility of rs6127099 and protein-coding gene promoters correlates with RNA expression in renal cell lines. . . . . 214
5.13 Cature-C: chromatin interactions of rs6127099 and gene promoters across three renal cell lines. . . . . 216
5.14 Forest plot demonstrating drug-target Mendelian randomisation estimates for the effects of modulating serum calcium concentrations via *DGKD*, *CASR*, and *CYP24A1*, and serum phosphate concentrations via *SLC34A1*. . . . . 222

6.1 Manhattan plot for a genome-wide association study of renal cysts in the UK Biobank. . . . . 236
6.2 Quantile-quantile plot of observed vs. expected p-values for a genome-wide association study of renal cysts in the UK Biobank. . 236
6.3 Locus-zoom plots of 1 megabase pair regions with colocalisation between renal cysts and kidney stone disease. . . . . 239
6.4 Bidirectional Mendelian randomisation analyses between renal cysts and kidney stone disease. . . . . 242
6.5 Leave-one-out analyses for the effect of increasing genetic liability to renal cysts on the odds of kidney stone disease. . . . . 242
6.6 Fixed-effect meta-analysis of bidirectional Mendelian randomisation estimates between liability to renal cysts and liability to kidney stone disease. . . . . 243
6.7 Regional bidirectional Mendelian randomisation between renal cysts and kidney stone disease mediated by rs9895661±500 kilobase pairs. 244

6.8 Pearson correlation coefficients of gene activity, inferred from rs9895661 coaccessibility, and RNA expression of *TBX2* and *TBX4*. 246

6.9 CaptureC- chromatin interactions of rs9895661 and gene promoters across three renal cell lines. . . . . 251

7.1 A lithogenic triad of factors that increase the risk of kidney stone disease. . . . . 260

# List of Tables


1.1 Laboratory results differentiating primary, secondary, and tertiary hyperparathyroidism. . . . . 15

1.2 Monogenic disorders causing kidney stone disease. . . . . 25

1.3 Published genome-wide association studies of kidney stone disease. . 43

2.1 Cohorts leveraged in this thesis. . . . . 76

2.2 Definitions of whole-exome sequencing variants based on SnpEff annotations. . . . . 90

2.3 snRNA cluster annotation based on cell-type specific markers. . . . 95

3.1 Single-nucleotide polymorphisms significantly associated with kidney stone disease in the UK Biobank. . . . . 105

3.2 Single nucleotide polymorphisms significantly associated with kidney stone disease from male sex genome-wide association study in the UK Biobank. . . . . 107

3.3 Single nucleotide polymorphisms significantly associated with kidney stone disease from female sex genome-wide association study in the UK Biobank. . . . . 107

3.4 Variants significantly associated with kidney stone disease in a meta-analysis of UK Biobank and FinnGen R12 study. . . . . 110

3.5 Variants significantly associated with kidney stone disease in a meta-analysis of UK Biobank, FinnGen R12 study, and Biobank Japan. . . . . 120

3.6 Heritability estimates for genome-wide association studies in UK Biobank, European ancestry genome-wide association study meta-analysis, and trans-ancestry meta-analysis. . . . . 128

3.7 MAGMA gene association test for European ancestry kidney stone disease genome-wide association study meta-analysis. . . . . 130
3.8 MAGMA gene association test for trans-ancestry kidney stone disease genome-wide association study meta-analysis. . . . . 131
3.9 Gene set analysis in European ancestry kidney stone disease genome-wide association study meta-analysis. . . . . 132
3.10 Gene set analysis in trans-ancestry kidney stone disease genome-wide association study meta-analysis. . . . . 133
3.11 Variants significantly associated with kidney stone disease from an exome-wide association study of kidney stone disease in the UK Biobank. . . . . 135
3.12 Summary of significant results from exome-wide, gene-based collapsing analyses of rare variants and kidney stone disease in the UK Biobank. . . . . 138

4.1 Number of 1 megabase regions involved in evaluating the causal effect of adiposity on odds of kidney stone disease. . . . . 154
4.2 Genomic regions of 1 megabase where body mass index increases odds of kidney stone disease. . . . . 155
4.3 Genomic regions of 1 megabase where waist-to-hip ratio increases odds of kidney stone disease. . . . . 156
4.4 Genomic regions of 1 megabase with evidence of multi-trait colocalisation between body-mass index, waist-to-hip ratio, and kidney stone disease. . . . . 158
4.5 95% credible sets for genomic regions with evidence for causal effects on kidney stone risk and colocalisation with adiposity traits. . . . . 159
4.6 Associations of candidate causal variants identified from HyPrColoc with adiposity traits and kidney stone disease. . . . . 161
4.7 Kidney stone population attributable risk and population attributable fraction for rs1800437-G in the UK Biobank. . . . . 163

4.8 Drug target Mendelian randomisation; effects of modulating adiposity traits on odds of kidney stone disease. . . . . 164

5.1 Estimate of bias for Mendelian randomisation with sample overlap in the UK Biobank. . . . . 180

5.2 Number of 1 megabase regions involved in evaluating the causal effect of serum calcium and phosphate concentrations on odds of kidney stone disease. . . . . 181

5.3 Genomic regions of 1 megabase where increasing serum calcium concentration affects odds of kidney stone disease with evidence of colocalisation. . . . . 182

5.4 Genomic regions of 1 megabase where decreasing serum phosphate concentration increases odds of kidney stone disease with evidence of colocalisation. . . . . 186

5.5 Genomic regions of 1 megabase with evidence of multi-trait colocalisation between serum calcium, phosphate, and parathyroid hormone concentrations and kidney stone disease. . . . . 189

5.6 Associations of candidate causal variants identified from HyPrColoc with mineral metabolism traits and kidney stone disease. . . . . 191

5.7 Significant Pearson correlation coefficients between gene activity and expression on chromosome 2 across renal cell types. . . . . 194

5.8 Significant interactions of rs838717 with gene promoters in renal cell lines identified in Capture-C analysis. . . . . 198

5.9 Significant Pearson correlation coefficients between gene activity and expression on chromosome 5 across renal cell types. . . . . 202

5.10 Significant interactions of rs10051765 with gene promoters in renal cell lines identified in Capture-C analysis. . . . . 206

5.11 Significant Pearson correlation coefficients between gene activity and expression on chromosome 20 across renal cell types. . . . . 212

5.12 Significant interactions of rs6127099 with gene promoters in renal cell lines identified in Capture-C analysis. . . . . 215
5.13 Kidney stone population attributable risk and population attributable fraction for candidate causal variants in the UK Biobank. . . . . 219
5.14 Drug target Mendelian randomisation; effects of modulating mineral metabolism traits on odds of kidney stone disease. . . . . 221

6.1 Associations of kidney stone disease in participants with and without renal cysts in *All of Us* and the UK Biobank. . . . . 234
6.2 Single nucleotide polymorphisms significantly associated with renal cysts from a genome-wide association study in the UK Biobank. . . 235
6.3 Results from regional colocalisation of renal cysts and kidney stone disease. . . . . 237
6.4 95% credible sets of colocalisation analyses between renal cysts and kidney stone disease at rs9895661± 500 kilobase pairs. . . . . 238
6.5 Associations of candidate causal variants identified from colocalisation analyses between renal cysts and kidney stone disease. . . . . 240
6.6 Bidirectional Mendelian randomisation analyses between renal cysts and kidney stone disease. . . . . 241
6.7 Regional bidirectional Mendelian randomisation between renal cysts and kidney stone disease mediated by rs9895661±500 kilobase pairs. 244
6.8 Pearson correlation coefficients between gene activity and expression on chromosome 17 across renal cell types. . . . . 246
6.9 Significant interactions of rs9895661 with gene promoters in renal cell lines identified in Capture-C analysis. . . . . 248

# 1 | Introduction

## 1.1 Kidney stone disease

### 1.1.1 Definition

Kidney stone disease (KSD) results from the accumulation of minerals in the renal calyces and pelvis, where stones may attach to the renal papillae or migrate downstream into the urinary tract [1]. It is a common condition, affecting ≈10% of individuals and has been documented since ancient civilisations, with signs and symptoms described by Hippocrates [2]:

*"An acute pain is felt in the kidney, the loins, the flank and the testis of the affected side; the patient passes urine frequently; gradually the urine is suppressed. With the urine, sand is passed; as the sand passes along the urethra, it causes severe pain which is relieved when it is expelled; then the same sufferings begin again."* [3, 4]

Understanding the causes of KSD is a linchpin to managing nephrolithiasis, which can have profound effects on physical function and quality of life [2]. Given its clinical significance, KSD is the central focus of this thesis. In the following sections, I will outline its epidemiology, classifications, principles of stone formation, aetiology, diagnosis, and treatment.

### 1.1.2 Epidemiology

KSD is common and affects ≈1 in 11 people by 70 years of age. Up to 50% of people with KSD will experience a second stone episode within 5 years of their first presentation [5, 6]. KSD accounts for >40,000 interventions, >88,000 acute hospital admissions, and ≈130 deaths each year [7]. By 2030 it is estimated that the National

Health Service in England will spend ≈£300 million per year treating kidney stones, a cost comparable to the cost of the first 5 years of bladder and prostate cancer treatment combined in the United Kingdom (UK). Geographical location is linked to varying prevalence of KSD. To date, this variation has primarily been attributed to environmental risk factors, such as seasonal variability and climate, while the role of genetic predisposition across different genetic ancestries remains relatively unexplored. Hotter and drier climates, rising temperatures, and increased sunlight are each independently associated with a higher prevalence of kidney stones [8]. Our understanding of the epidemiology of KSD is largely based on observational studies implicating age, male gender, obesity, occupation, and diet as some of the risk factors associated with the condition [5, 9, 10]. Historically, KSD has affected more males than females. In the United States of America (USA), in 2017-18, prevalence of disease was reported as 10.9% in men and 9.5% in women. However, this gap is narrowing; between 2007 and 2018 prevalence of KSD in females increased from 6.5% to 9.4% yet this trend was not seen in males [5].

Socioeconomic status and ancestry affect the prevalence of KSD [9]. In the USA, non-Hispanic, white individuals bear the greatest burden of KSD, accounting for ≈73% of cases, whereas non-Hispanic black and Hispanic populations have a lower prevalence of KSD among both men and women [9]. The predominance of KSD in non-Hispanic white populations remains unexplained; however, researchers hypothesise that it is due to genetic and non-genetic risk factors related to varying sensitivity to parathyroid hormone (PTH), and differences in oxalate, citrate, and calcium metabolism [11–14]. Individuals of lower socioeconomic status are disproportionately affected by KSD. They present with larger and more complex stones, report worse health-related quality of life, and are more vulnerable to recurrent disease, in part due to diagnostic and treatment delays [15].

### 1.1.3 Classification of kidney stones

Calcium oxalate, calcium phosphate, magnesium ammonium phosphate (struvite), and uric acid stones are the four most common constituents of kidney stones. While most kidney stones are of mixed composition, they are generally classified based on their predominant chemical component [1]. Calcium-based stones account for

≈80% of all renal calculi [1]. Magnesium ammonium phosphate stones (also known as "infection" stones) make up another ≈10%, uric acid stones comprise 9%, and cystine stones account for ≈1%. Rare stone types may also occur due to iatrogenic causes, such as therapy with indinavir, triamterene, or xanthine.

Kidney stones range in size, from a grain of sand to multiple centimetres, and comprise crystals surrounded by an organic matrix [16, 17]. Calcium oxalate and calcium phosphate crystals induce the production of macromolecules that modulate crystallisation [18]. Calcific stones' matrices contain lipids, that induce crystal nucleation, and urinary proteins that modulate crystallisation, including osteopontin, inter-$\alpha$-inhibitor, and urinary prothrombin fragment 1 [18, 19].

Calcium oxalate stones occur in either monohydrate or dihydrate forms, each with distinct crystal structures [1]. Calcium oxalate monohydrate stones have a spindle- or dumbbell-shaped morphology with well-defined concentric layers, while calcium oxalate dihydrate stones display a tetragonal, bipyramidal shape (Figure 1.1). Pure calcium oxalate monohydrate stones are more common and typically have smooth surfaces, in contrast to the rough, jagged appearance of calcium oxalate dihydrate stones or mixed stones, which have more irregular, sharp features.

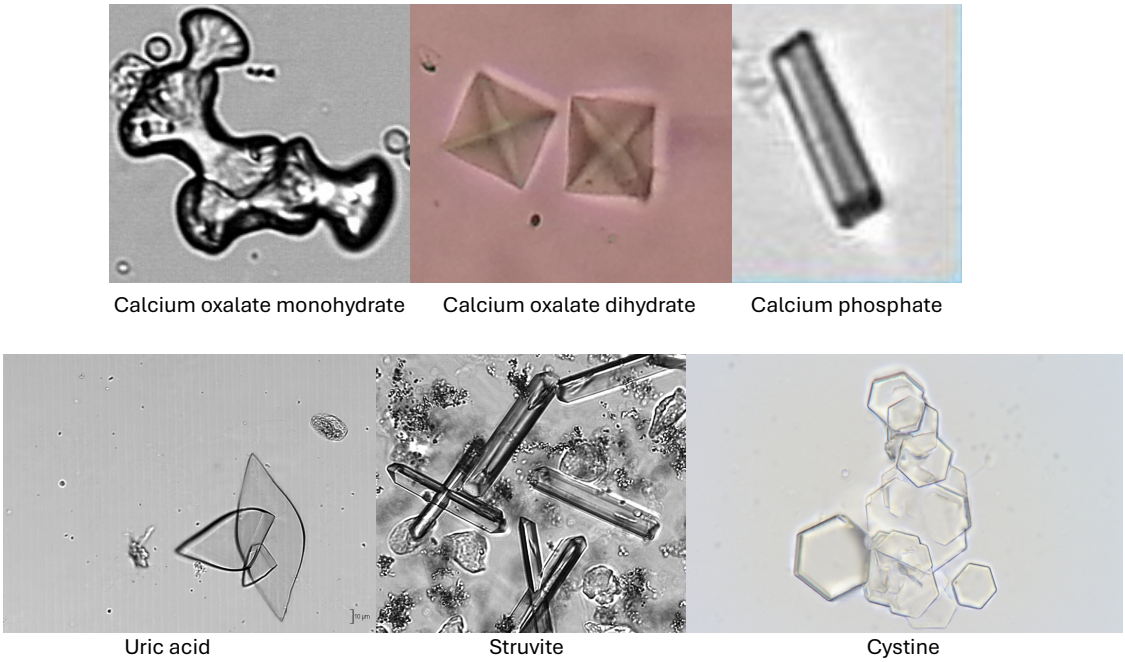



***Figure 1.1. Crystals constituting kidney stones.***

Urine microscopy of different crystal forms showcasing their unique morphologies. Images are credited to artists and shared under Wikimedia Creative Commons Attribution-Share Alike Licenses: [20–25].

Calcium phosphate stones can be basic calcium phosphate (apatite stones), calcium hydrogen phosphate dihydrate (brushite stones), or tricalcium phosphate (whitlockite). Despite being the most common crystal in kidney stones and filling spaces between other types of crystals, pure calcium phosphate stones are rare. Whitlockite is exceedingly rare, however brushite is frequently found in kidney stones as rosettes of radially arranged, thin, blade-like crystals (Figure 1.1) [26]. Major risk factors for calcium phosphate stones include hypercalciuria, hypocitraturia, and elevated urine pH which may reflect underlying metabolic abnormalities such as distal renal tubular acidosis (dRTA) or hyperparathyroidism [26].

Uric acid stones are associated with metabolic syndrome, obesity, and insulin resistance (Figure 1.1) [5]. Uric acid crystal morphology differs according to the crystal composition; uric acid dihydrate (pictured in Figure 1.1) are the most common form and may resemble rhombi, prisms, ovals with pointed ends, wedges, rosettes, or irregular plates [27]. Unlike calcium stones, the risk of uric acid stone formation is higher in acidic urine (pH$<$5.5) due to the decreased solubility of uric acid below this threshold. Dehydration exacerbates this risk by increasing urinary uric acid excretion, predisposing individuals to hyperuricosuria and, consequently, uric acid stone formation [1]. Medications that inhibit renal uric acid reabsorption, such as probenecid, further promote hyperuricosuria [28]. It was widely believed that hyperuricaemia also contributed to uric acid stone formation, under the assumption that elevated serum uric acid, from excess purine breakdown, for example from high animal protein intake, chemotherapy, or from overproduction, would lead to increased urinary uric acid excretion. However, there is no correlation between serum and urine uric acid levels [29], thus routine serum uric acid testing in patients with kidney stones is not recommended by the National Institute for Health and Care Excellence (NICE) in the UK [30] although features in European and American clinical guidelines [6, 31].

Struvite stones typically form due to infection with urease-producing bacteria, such as *Klebsiella pneumoniae* and *Proteus spp.* These bacteria break down urea into ammonia, raising urine pH and promoting orthorhombic, "coffin-lid" shaped, magnesium ammonium phosphate hexahydrate crystal formation (Figure 1.1). These

crystals become encrusted with apatite and mixed with cellular debris, accelerating the growth of large, rapidly forming stones. Struvite and calcium phosphate crystals can quickly coalesce into "staghorn" stones, which occupy the renal pelvis and calyces [32].

Cystine stones result from a rare autosomal recessive defect in the cysteine transporter of the kidney [33]. Cysteine is freely filtered into renal tubular fluid at the glomerulus. Cysteine dimerises into cystine, which, at a normal urinary pH (5.5–7.5), is relatively insoluble, in contrast to the other dibasic amino acids (ornithine, lysine, and arginine) [34]. Thus, cystine crystallises into hexagonal, amber-coloured, slightly opaque crystals with homogeneous interiors (Figure 1.1) [35]. Furthermore, cystine crystals can promote calcium oxalate crystal precipitation. Thus, while most individuals with cystine stones form pure cystine stones, up to 40% may have mixed stones containing calcium oxalate, phosphate, or struvite [36].

### 1.1.4 Principles of kidney stone formation

Urine becomes saturated when the product of solute concentrations exceeds the solubility product ($K_{sp}$)(Figure 1.2). Below $K_{sp}$, crystals do not form. However, above $K_{sp}$, crystals can form, though inhibitors in urine may prevent this. Once solute concentrations reach a critical threshold (the *formation product*, $K_f$), urine becomes *supersaturated*, making the crystallisation inhibitors ineffective and allowing crystals to form [37](Figure 1.2). Urine between $K_{sp}$ and $K_f$ is in a *metastable* state. Several factors promote urine supersaturation, including the post-prandial state, dehydration, alkaline urine, hypercalciuria, hyperuricosuria, and hyperoxaluria. The degree of supersaturation for each solute determines the type of kidney stone that forms [38].

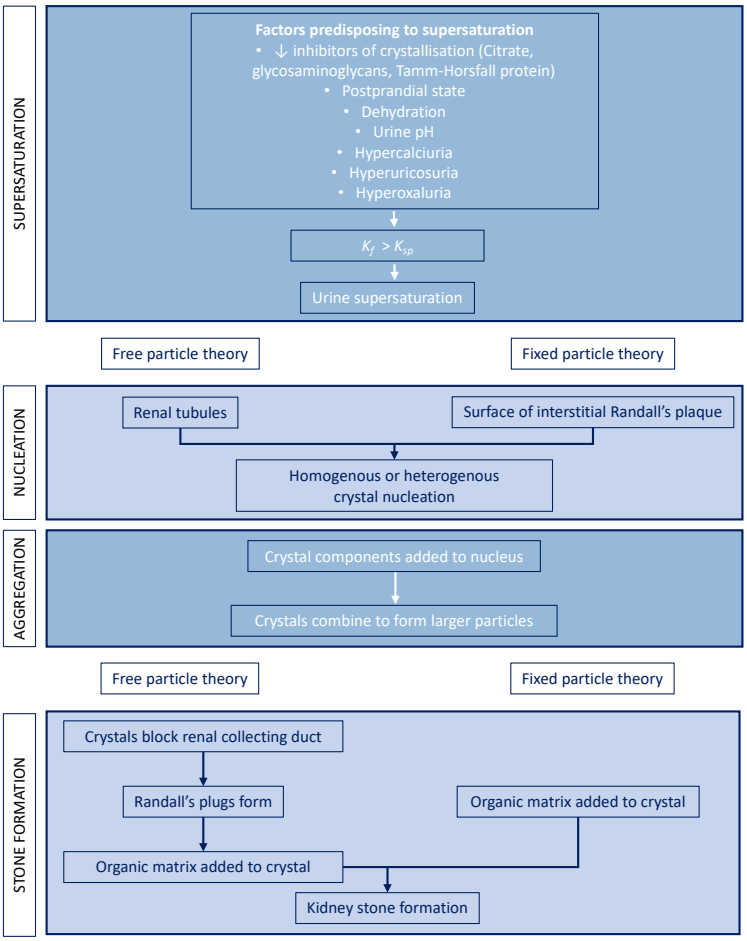



***Figure 1.2. Principles of kidney stone formation.***

**Supersaturation:** Urine supersaturation occurs when the product of solute concentrations ($K_f$) exceeds the solubility product ($K_{sp}$). **Nucelation:** Under the free particle theory, crystal nucleation occurs in the renal tubules. Under the fixed particle theory, crystal nucleation occurs on interstitial Randall's plaques. **Aggregation:** Growth of crystals. **Stone formation:** Under the free particle theory, crystals grow and become lodged in the collecting duct and form Randall's plugs which become coated in organic matrix to form kidney stones. Under the fixed particle theory, crystals adhered to Randall's plaques in the renal interstitium become covered in organic matrix to form kidney stones.

*Crystal nucleation*, the first stage of crystallisation, refers to the process where solute molecules begin clustering. Homogeneous nucleation involves a high concentration of one mineral solute, whereas heterogeneous nucleation involves a combination of proteins, organic polymers, mineral particulates, and chemically active cell surfaces [39]. Two major mechanisms explain kidney stone formation: the *free particle theory* and the *fixed particle theory* (Figure 1.3 and Figure 1.2). The free particle theory suggests that crystals nucleate within renal tubules, forming Randall's plugs, also referred to as "Bellini duct plugs" (Figure 1.3). In contrast, the fixed particle theory proposes that Randall's plaques themselves act as sites of crystal nucleation

and aggregation (Figure 1.3) [40]. There are two theories regarding the origins of Randall's plaques and both are based on morphological examinations of kidneys from stone formers [41]. Evan et al. suggest that Randall's plaques originate in the basement membrane of the thin limb of the loop of Henle [42–44]. In contrast, Stoller et al. propose that Randall's plaques result from vascular injury to the papillary vasculature, which triggers an atherosclerotic-like reaction [45]. This reaction may lead to calcification of the damaged vessel wall, which can erode into the renal papilla and subsequently into the papillary ducts of Bellini, creating a nidus for stone formation within the surrounding supersaturated urine [45]. In idiopathic stone formers, stones are often found attached to these plaques [46, 47].

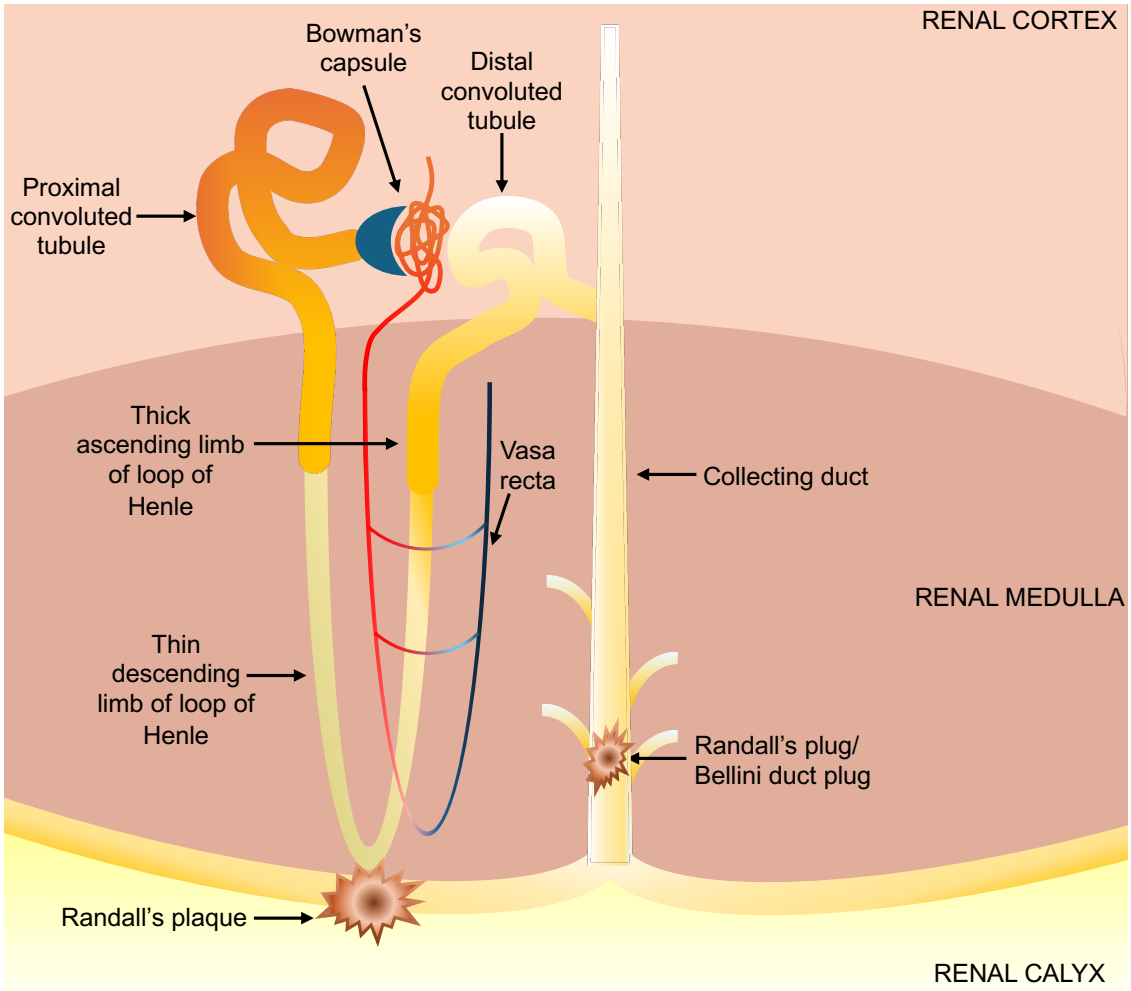



***Figure 1.3. Randall's plaques and Randall's plugs- foci for kidney stone formation.***

An illustration of the differences between Randall's plaques and Randall's plugs, also known as Bellini duct plugs, in one nephron. **Randall's plaques:** A breach in the urothelium at the renal papilla causes interstitial calcium phosphate to be exposed to urine, creating a nidus for stone formation. **Randall's plugs:** Crystal nuclei form in the lumen of nephrons at sites of cell injury resulting in crystal attachment and growth.

After nucleation, crystals grow through aggregation (Figure 1.2). The rate of crystal growth depends on the relative supersaturation of the solutes and the balance between inhibitors and promoters of stone formation in urine [48, 49]. Citrate, for example, inhibits crystallisation by forming complexes with calcium, preventing it from binding with oxalate or phosphate. Other inhibitors include magnesium, pyrophosphate, adenosine diphosphate, adenosine triphosphate, glycosaminoglycans, non-polymerised Tamm-Horsfall protein (uromodulin), and osteopontin when suspended in solution [48, 50]. In contrast, promoters include matrix substance A, the polymerised form of Tamm-Horsfall protein (uromucoid), and osteopontin when immobilised on surfaces such as collagen [50].

Under the free particle mechanism, crystals nucleate, grow, and aggregate within the renal tubules (Figure 1.2). Once they reach a critical size, they are either retained in the kidney, because they are too large to pass, or attach to the tubular epithelium [1, 39]. When urine becomes highly supersaturated, these crystal deposits can also block the kidney's collecting ducts and form Randall's plugs (Figure 1.3). These plugs, once exposed to urine in the renal pelvis, promote stone formation, particularly in areas of urine stasis within occluded tubules or renal calyces [51]. In contrast, the fixed particle mechanism suggests that stones form by adhering to calcium phosphate Randall's plaques in the renal interstitium (Figure 1.2 and Figure 1.3). When the surface of a plaque is breached, for instance, due to matrix metalloproteinases or shear force, the plaque is exposed to urine, promoting further crystal growth and providing a substrate for organic matrix to bind to [41, 51].

### 1.1.5 Principles of calcium and phosphate metabolism

Given that most kidney stones are composed of calcium and/or phosphate, and that disturbances in their metabolism play a pivotal role in KSD development, discussing the principles of calcium and phosphate metabolism is pertinent when considering the pathophysiology of KSD. Maintaining serum concentrations within normal limits is critical in maintaining normal physiological activity. Calcium and phosphate homeostasis operates through a complex network of feedback loops, tightly regulated by PTH, 1,25-dihydroxyvitamin D (calcitriol), calcitonin, and

fibroblast growth factor-23 (FGF23). These regulators act on key tissues, including the intestine, bone, parathyroid, and kidneys [52, 53].

#### 1.1.5.1 Calcium metabolism

Disruptions of calcium homeostasis cause serum and urine calcium levels to fluctuate and impact bone mineralisation, neuromuscular function, coagulation, intracellular signal transduction, and kidney stone formation. Approximately 40% of circulating serum calcium binds to albumin, while smaller amounts bind to citrate and lactate. The remaining calcium exists in a free, non-ionised form, which is biologically active. Variations in albumin levels alter total serum calcium measurements, even when biologically active calcium levels remain normal. Thus, to accurately assess an individual's calcium status, total calcium levels must be adjusted for albumin concentration.

Extracellular calcium balance relies on a tightly regulated interplay between the parathyroid glands, kidneys, intestines, and bones. The parathyroid glands sense extracellular calcium concentrations and adjust PTH secretion accordingly. The kidneys and small intestine mediate shifts in calcium between extracellular fluid and the external environment, and the skeletal system acts as both a calcium reservoir and buffer, managing acute changes in calcium concentrations. The calcitropic hormones PTH and 1,25-dihydroxyvitamin D act on target organs to mediate these interactions [53].

The transmembrane calcium sensing receptor (CaSR) is a G-protein coupled receptor that is expressed in the parathyroid glands, kidneys, and bones and plays a central role in calcium homeostasis [54]. CaSR expression is regulated by endocytic complexes composed of clathrin, $\beta$-arrestin, and the adaptor protein AP2$\sigma$ [55], which mediate receptor recycling. CaSR signalling also enhances the anterograde trafficking of newly synthesised receptors to the plasma membrane, a process termed "agonist-driven insertional signalling" [56]. This prevents CaSR desensitisation from prolonged exposure to extracellular calcium [56].

Decreases in serum calcium concentration, detected by the CaSR in the parathyroid

glands, cause reduced intracellular signalling, triggering PTH secretion. PTH activates the PTH1 receptor and stimulates the following processes:

1. Osteoclast-mediated calcium resorption from bone.
2. Renal calcium reabsorption from urine in the thick ascending limb of the loop of Henle.
3. Phosphate excretion in the proximal renal tubule.
4. Expression of 1-$\alpha$-hydroxylase in the kidney, which converts inactive 25-hydroxyvitamin D into the active form, 1,25-dihydroxyvitamin D.

Once activated, 1,25-dihydroxyvitamin D binds to the vitamin D receptor, facilitating intestinal calcium and phosphate absorption to support bone mineralisation, and modulating PTH and FGF23 levels [53]. Elevated serum calcium levels and vitamin D receptor activation, suppress PTH secretion by negative feedback mechanisms.

**Calcium metabolism in the kidney**

Renal calcium handling occurs through different mechanisms along the nephron. Most renal calcium reabsorption is by passive, hormone-independent, paracellular transport in the proximal tubules (≈60%) and thick ascending limb of the loop of Henle (≈20%) [57]. In the proximal tubule, calcium reabsorption is driven by active, transcellular sodium reabsorption through the sodium-hydrogen exchanger (NHE3) which generates an osmotic gradient and promotes water reabsorption [58]. Passive calcium reabsorption from the proximal tubule lumen then occurs via a paracellular channel formed by claudins-2 and -12 , driven by a combination of two mechanisms: 1) water reabsorption concentrates calcium in the lumen and generates a concentration gradient that promotes calcium reabsorption; 2) the convective movement of water drives calcium reabsorption.

In the thick ascending limb of the loop of Henle, calcium reabsorption is regulated by CaSR independently of PTH [59]. The sodium-potassium-chloride cotransporter (NKCC2) and renal outer medullary potassium (ROMK) channel establish a

positive transepithelial potential that promotes paracellular calcium and magnesium transport. Claudins-16 and -19 form calcium-permeable channels, which are regulated by claudin-14 in response to extracellular calcium levels. Increased CaSR signalling enhances claudin-14 expression, reducing calcium permeability and promoting calcium excretion to prevent hypercalcaemia [60, 61].

In contrast to the proximal nephron, where calcium transport is largely passive, fine-tuning of calcium reabsorption occurs in the distal convoluted tubule by active, transcellular mechanisms. Here, PTH and 1,25-dihydroxyvitamin D regulate calcium transport, accounting for ≈15% of renal calcium reabsorption [57]. Calcium passes from the tubule lumen, through the transient receptor potential cation channel subfamily V (TRPV5), binds calbindin D-28k for intracellular transport, and exits the cell through the $Ca^{2+}$-ATPase and the $Na^{+}/Ca^{2+}$ exchange on the basolateral membrane [57].

### Phosphate metabolism

Like calcium, phosphate metabolism is essential for a range of biological activities, including muscle contraction, cell signalling, and membrane stabilisation. Phosphate uptake occurs along the gastrointestinal tract via passive and active mechanisms, namely tight junctions and the sodium-dependent phosphate transport protein 2B (NaPi-IIb) respectively [62]. NaPi-IIb, upregulated by 1,25-dihydroxyvitamin D, enhances phosphate absorption, and nearly all unbound inorganic phosphate in the intestinal lumen is absorbed. Renal phosphate excretion is minimal under normal dietary conditions but increases in response to excessive phosphate intake [63].

Renal phosphate handling primarily occurs in the proximal tubule, where phosphate is reabsorbed through transcellular mechanisms. Phosphate excretion is almost entirely dependent on glomerular filtration and is tightly regulated by phosphaturic hormones, namely FGF23 and PTH. A phosphate load triggers FGF23 release from osteocytes and PTH secretion from the parathyroid glands. Both hormones promote phosphaturia by promoting sodium-dependent phosphate transport protein 2A (NaPi-IIa) internalisation at the apical membrane of renal tubular epithelial

cells, thereby reducing phosphate reabsorption and increasing its urinary excretion [62].

FGF23 binds to FGF receptors and the co-receptor, $\alpha$-Klotho, to enhance phosphate excretion and suppress 1,25-dihydroxyvitamin D synthesis. While 1,25-dihydroxyvitamin D upregulates NaPi-IIb to promote phosphate absorption, it also stimulates FGF23 production. In turn, FGF23 inhibits 1-$\alpha$-hydroxylase, reducing active vitamin D levels, and induces its degradation into the inactive form, 24,25-dihydroxyvitamin D, via 24-$\alpha$-hydroxylase. PTH similarly increases FGF23 secretion, while FGF23 negatively regulates PTH [62].

These tightly coordinated mechanisms maintain calcium and phosphate homeostasis by balancing absorption, reabsorption, excretion, and hormonal control. Disruptions in this equilibrium can promote kidney stone formation.

### 1.1.6 Aetiology

KSD is a complex trait, resulting from the interplay of both environmental and genetic risk factors. In the subsections that follow, I will explore the modifiable and non-modifiable risk factors for KSD, including dietary influences on calcium metabolism, metabolic syndrome, renal disorders, and our understanding of the genetic causes of stone formation.

#### 1.1.6.1 Calcium metabolism and kidney stone disease

**Idiopathic hypercalciuria**

Idiopathic hypercalciuria is the most common metabolic abnormality in patients with kidney stones, affecting ≈50% of individuals [64, 65]. It is characterised by normocalcaemia and elevated urinary calcium excretion (>0.1 mmol/kg or >4 mg/kg per 24 hours) in the absence of an underlying disease [66, 67]. Idiopathic hypercalciuria significantly promotes stone formation by increasing urinary supersaturation of calcium oxalate [68].

Although first-degree relatives of affected individuals have a higher likelihood of hypercalciuria, suggesting a genetic component [69], idiopathic hypercalciuria is a polygenic trait. Thus, genetic and environmental factors contribute to increased intestinal calcium absorption, bone resorption, and reduced renal calcium reabsorption [67].

Idiopathic hypercalciuria is influenced by intestinal dietary calcium absorption, calcium release from bones, and the extent to which filtered calcium is reabsorbed in the renal tubule [64, 65]. Compared to healthy individuals, patients with idiopathic hypercalciuria excrete more calcium than they absorb, even on a normal diet, and display reduced efficiency in calcium uptake when dietary intake is low [67, 70]. Furthermore, the imbalance between calcium excretion and absorption can lead to higher bone resorption, often resulting in reduced bone mineral density [71]. Traditionally, the condition has been classified into absorptive and renal subtypes. However, calcium-balance studies indicate that many patients remain in negative calcium balance even on a low-calcium diet, leading to the recommendation for a normal-calcium intake [30, 64]. The role of calcitriol in this process remains debated; some studies report elevated levels that could drive increased intestinal calcium absorption, while others suggest a more complex pathophysiology [72, 73].

### Hyperparathyroidism

Primary hyperparathyroidism is a common endocrine disorder characterised by hypercalcaemia and elevated or inappropriately normal levels of PTH. In 80% of cases, a solitary parathyroid adenoma is the underlying cause, while four-gland hyperplasia accounts for 10-15%, multiple adenomas for 5%, and parathyroid carcinoma for less than 1% of cases [74]. The incidence is higher in women than in men, more prevalent among African Americans compared to other ancestral groups, and increases with age [75]. In primary hyperparathyroidism, an increase in parathyroid cell mass disrupts the normal feedback between serum calcium levels and PTH secretion, requiring higher calcium concentrations to suppress PTH release [76]. This may be partly due to reduced CaSR expression in parathyroid adenomas, which diminishes the glands' sensitivity to extracellular calcium. Studies have shown an inverse

relationship between CaSR expression and PTH secretion, highlighting its role in the dysregulation of calcium-mediated feedback in benign parathyroid tumours [77, 78].

Primary hyperparathyroidism is frequently asymptomatic, although subclinical target organ involvement, such as osteoporosis, vertebral fractures, hypercalciuria, and KSD, may be present [74]. Most diagnoses are incidental findings during routine blood tests, which reveal elevated albumin-adjusted serum calcium levels, increased PTH concentrations, and low phosphate levels due to the phosphaturic effects of PTH [79]. The primary renal manifestations include hypercalciuria and kidney stones; asymptomatic renal stones are identified in ≈7% patients, particularly younger and male patients, compared to ≈1.6% in the general population [80]. For patients experiencing end-organ sequelae, parathyroidectomy remains the only curative treatment. This surgical intervention increases bone mineral density and reduces the incidence of nephrolithiasis to levels comparable to those in patients with idiopathic KSD [80].

Primary hyperparathyroidism can be distinguished from secondary and tertiary forms through laboratory investigations (Table 1.1) [74]. Secondary hyperparathyroidism arises due to elevated PTH concentrations as a physiological response to hypocalcaemia or hyperphosphataemia, often caused by vitamin D deficiency, malabsorption, renal disease, or hypercalciuria [81]. Hypocalcaemia drives PTH secretion and parathyroid hyperplasia, with elevated PTH promoting normal or raised serum calcium levels, phosphate resorption in the intestine, osteoclast-driven bone resorption, and increased activity of renally-expressed 1-$\alpha$-hydroxylase to promote 1,25-dihydroxyvitamin production [81]. 1,25-dihydroxyvitamin enhances calcium and phosphate absorption in the gut and their reabsorption in the renal tubules. Tertiary hyperparathyroidism represents the progression of prolonged and severe secondary hyperparathyroidism, as often observed in end-stage renal failure. This condition evolves into hypercalcaemia due to autonomous functioning of one or more hyperplastic parathyroid glands leading to excessively high PTH levels alongside elevated calcium and phosphate concentrations.

***Table 1.1. Laboratory results differentiating primary, secondary, and tertiary hyperparathyroidism.***

| Serum concentration | Primary hyperparathyroidism | Secondary hyperparathyroidism | Tertiary hyperparathyroidism |
|---|---|---|---|
| **Calcium** | High | Low / normal | High |
| **Phosphate** | Low | High | High |
| **Parathyroid hormone** | High | High | High |

#### 1.1.6.2 Diet and kidney stone disease

Observational studies associate low fluid intake with an increased risk of kidney stone formation [82]. Inadequate hydration reduces diuresis, leading to more concentrated urine and promoting supersaturation of stone-forming substances. Analyses from the UK Biobank suggest that higher consumption of total fluids, tea, coffee, and alcohol lowers the risk of stone formation, likely due to diuretic effects [83].

The relationship between dietary protein intake and KSD risk remains unclear. Idiopathic hypercalciuria is linked to higher urea excretion, suggesting an increased sensitivity to the hypercalciuric effects of dietary protein in this group [71]. A recent analysis of 439,072 individuals identified a higher risk of kidney stones with greater meat consumption [83]. Additionally, a trial of 120 men found that a diet low in animal protein and salt halved the risk of recurrent stone formation compared to a low-calcium diet [84]. However, population-based cross-sectional studies report no significant association between dietary protein intake and KSD prevalence [85], though dietary modifications post-diagnosis may confound these findings. Small randomised controlled trials (RCTs) have also shown no significant benefit of reducing animal protein intake on KSD risk [86, 87]. The conflicting evidence likely reflects the inherent limitations of observational studies, such as confounding variables, and the challenges of accurately assessing nutrient intake.

Oxalate homeostasis is the result of the delicate balance between endogenous production by the liver, exogenous dietary supply, and renal and faecal excretion. Hyperoxaluria can arise from dietary or enteric factors. It not only promotes frequent stone formation, but can also cause oxalate nephropathy, an acute tubulointerstitial renal disease caused by intrarenal calcium oxalate crystal deposition and which can lead to chronic kidney disease (CKD) [73, 88]. Diets high in oxalate- such as those rich in spinach, rhubarb, cocoa, beetroot, and peanuts- can cause peaks in renal excretion to increase urinary oxalate concentrations. Similarly, diets low in calcium can also cause increased enteric oxalate absorption and hence higher urinary oxalate concentrations. In cases of dietary hyperoxaluria, treatment focuses on dietary modifications to reduce oxalate and normalise calcium intake.

Conversely, enteric hyperoxaluria is associated with small bowel malabsorption, as seen in Crohn's disease or following gastrointestinal resection. Increased intestinal concentrations of bile salts and fatty acids enhance the permeability to oxalate and amino acids, leading to hyperoxaluria. Additionally, saponification from calcium binding to unabsorbed fatty acids causes an excess absorption of intestinal oxalate which is then filtered and excreted in urine [88]. Treatment options for enteric hyperoxaluria include reducing dietary oxalate and fat intake, administering oral calcium supplements to precipitate oxalate in the intestinal lumen, using cholestyramine to bind fatty acids, bile acids, and oxalate, supplementing with oral citrate, and maintaining high fluid intake [89, 90].

#### 1.1.6.3 Metabolic syndrome and kidney stone disease

Several systemic comorbidities are associated with an increased risk of KSD. Among these, metabolic syndrome, a disorder of energy utilisation and storage, poses significant challenges to clinical practice and public health. Observational studies reveal links between KSD and the components of metabolic syndrome, which include central obesity, hypertension, dyslipidaemia, and phenotypes related to type 2 diabetes [91–98]. Metabolic syndrome arises from a complex interplay of environmental factors, genetic predisposition, insulin resistance, dysfunctional adipose tissue, ectopic lipid accumulation, and systemic inflammation. The

mechanisms linking obesity and metabolic syndrome to an increased risk of KSD are not fully understood. Notably, the global incidence of KSD has risen in recent decades, paralleling increases in the prevalence of obesity and type 2 diabetes, possibly alluding to a shared pathophysiology [5].

### Obesity and insulin resistance

Obesity and insulin resistance are associated with alterations in urinary profiles. There is substantial evidence linking obesity to higher plasma uric acid concentrations [99, 100]. It has long been assumed that obesity and insulin resistance drive increased uric acid excretion, thereby raising the prevalence of uric acid nephrolithiasis [101]. However, studies suggest that although uric acid nephrolithiasis may be more prevalent in obese or insulin-resistant individuals, the overall proportion of uric acid stones within the stone population does not differ significantly when comparing obese and non-obese individuals, unless body mass index (BMI) reaches levels $\geq 40 kg/m^2$ [101]. Furthermore, the correlation between plasma and urinary uric acid levels is weak [29]. Individuals with a higher BMI demonstrate similar fractional excretion of uric acid in the fasted state and may exhibit increased renal reabsorption and reduced renal clearance of uric acid in the non-fasted state when compared to individuals with a normal BMI [102, 103].

Although obesity may not directly increase urinary uric acid load, higher production of organic acids could reduce urine pH and promote crystallisation of calcium- and uric acid-based stones. Insulin resistance may further contribute to lower urinary ammonium and pH, alongside hypercalciuria and hypocitraturia [100, 104–107]. Diabetic fatty rat models have lower prevalence of the renal brush-border sodium-hydrogen exchanger, NHE3, lower urinary ammonium, and lower urinary pH [108]. Similarly, *in vitro* studies of opossum kidney cells suggest that renal lipotoxicity, resulting from steatosis, reduces ammonium secretion in proximal tubular cells through decreased NHE3 activity and impaired ammonia regulation [108].

### Dyslipidaemia and inflammation

Increased adiposity and dyslipidaemia are linked to altered serum concentrations of calcium, phosphate, vitamin D, sodium, oxalate, and uric acid - all of which influence KSD risk [102, 109, 110]. Type 2 diabetes further exacerbates this risk, with higher plasma levels of the oxalate precursors, glyoxal and glyoxylate, observed in affected individuals (Figure 1.5). Elevated C-reactive protein and interleukin-6 levels in obesity-induced systemic inflammation may also alter gut oxalate absorption, particularly in individuals with both diabetes and obesity, promoting hyperoxaluria and increasing nephrolithiasis risk [111, 112]. Furthermore, pro-inflammatory cytokines, including interleukins-1$\beta$, -6, -8, and tumour necrosis factor-$\alpha$, can directly damage renal tubules by promoting apoptosis and acute kidney injury, further increasing the risk of KSD [113].

### Hypertension

Hypertension, a feature of metabolic syndrome, shows a strong association with KSD in observational studies, which report a two- to five-fold increased risk of stone episodes in hypertensive compared with normotensive patients [114, 115]. Several mechanisms are postulated to link hypertension to increased KSD risk, including vascular injury, increased urinary calcium excretion, and renal dysfunction. Elevated blood pressure disrupts renal papillary circulation, induces renal vasoconstriction, reduces renal blood flow, and increases plasma volume and natriuresis [45, 114, 116, 117]. Analyses of 24-hour urinalysis profiles in hypertensive patients reveal lower urinary citrate concentrations, a factor that increases risk of stone formation [118]. Furthermore, hypertensive individuals often have a higher body weight and consume more dietary sodium than normotensive controls, compounding their risk of KSD, and antihypertensive therapies, such as thiazide diuretics, also alter urine composition [114]. Although multivariable analyses attempt to adjust for these factors, residual confounding remains a major concern in observational studies and distinguishing the independent effects of hypertension from those of obesity and other metabolic disturbances remains challenging.

#### Insights from Mendelian randomisation

The evidence which I have outlined for the role of obesity and metabolic syndrome as risk factors for KSD is largely founded on retrospective observational studies. Mendelian randomisation (MR) is one method of studying the effects of modifiable exposures on an outcome using genetic variants to avoid some of the biases that are inherent to observational analyses (discussed in more depth in Section 1.2.4). Prior to commencing my DPhil studies, I performed MR analyses, incorporating data from the UK Biobank and FinnGen study, and revealed that KSD risk increases by $\approx 21\%$ per 1 standard deviation (SD) higher BMI, a marker of general adiposity, independent of waist-to-hip ratio (WHR), a measure of central adiposity [119]. Similarly, the risk increases by ≈24% per 1 SD higher WHR, independent of BMI. Mediation MR analyses further demonstrate that 12–15% of the effect of higher WHR, but not BMI, on risk of KSD is mediated via increases in serum calcium concentrations. Moreover, there is no evidence for causal effects between diabetes, hypertension, dyslipidaemia, or inflammatory processes, and KSD. These findings indicate that both general and central adiposity contribute to the risk of KSD, likely through independent mechanisms. This may be related to the distinct roles of central and peripheral adipose depots in energy storage, glucose homeostasis, endocrine function, cardiovascular disease, cancer risk, and all-cause mortality [120–122].

Further research is needed to unravel causal mechanisms linking adiposity and KSD and to determine whether weight loss and metabolic interventions affect risk of kidney stone formation.

### 1.1.6.4 Uric acid metabolism and kidney stone disease

Hyperuricosuria, defined as urinary uric acid excretion exceeding 800 mg per day in males and exceeding 750 mg per day in females, increases the risk of uric acid stones and promotes calcium oxalate stone formation. Uric acid acts as a nucleus for crystallisation, reduces crystallisation inhibitors, and lowers the metastable limit of calcium oxalate [123].

Uric acid forms through purine breakdown, with purine load depending on dietary

intake (mainly from meat), intrinsic production (elevated by conditions featuring high cell turnover, for example myeloproliferative disorders, or during chemotherapy), and renal excretory function. Guanine and adenosine compounds are processed by purine nucleoside phosphorylase to produce hypoxanthine and guanine (Figure 1.4). Hypoxanthine is then converted to xanthine by xanthine oxidase, while guanine is converted by guanase. Xanthine oxidase catalyses the final step, converting xanthine to uric acid [124].

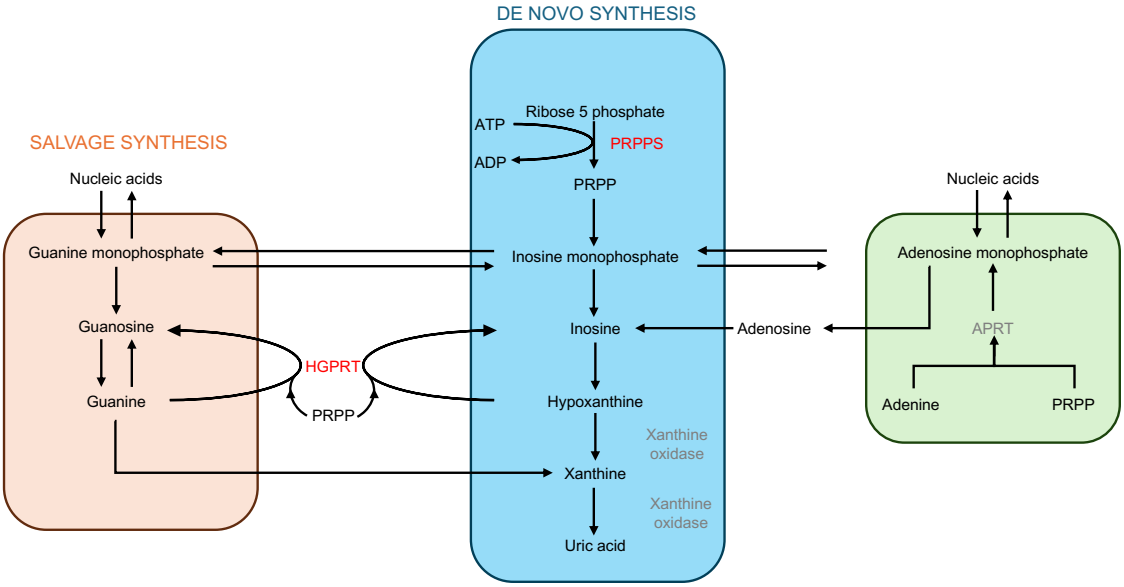



***Figure 1.4. Abnormalities of purine metabolism causing hyperuricosuria.***
Mutations in HGPRT and PRPPS, that cause hereditary hyperuricosuria, are highlighted in red.
ADP=Adenosine diphosphate; ATP=Adenosine triphosphate; HGPRT=Hypoxanthine-guanine phosphoribosyltransferase; PRPP=Phosphoribosylpyrophosphate; PRPPS=Phosphoribosyl pyrophosphate synthase.

#### 1.1.6.5 Renal diseases and kidney stone disease

**Chronic kidney disease**

Observational analyses report associations of renal diseases, such as CKD, autosomal dominant polycystic kidney disease (ADPKD) and medullary sponge kidney, with KSD [93, 125, 126]. Over nearly nine years of follow-up, researchers found that a cohort of ≈5,000 stone formers had a higher likelihood of being diagnosed with CKD, sustained higher serum creatinine levels, and lower estimated glomerular filtration rates compared to ≈13,000 controls [93]. They proposed that medications, infections, stone treatments, obstruction, and chronic crystalluria

contributed to these associations. More recently, the same research group used MR to demonstrate that, although stone formers statistically have higher rates of CKD and cardiometabolic disorders, KSD is not an independent risk factor for CKD and CKD lowers the risk of calcium kidney stones. They further showed that KSD and CKD are not genetically correlated, suggesting distinct underlying genetic architectures and causal mechanisms [127].

**Renal cystic disease**

Patients with ADPKD are at a heightened risk of KSD compared to healthy individuals of a similar age, with studies reporting a prevalence of KSD ranging from ≈3% to 59% in this population [125, 128]. Additionally, individuals with ADPKD have a higher prevalence of kidney stones than unaffected family members [129]. These findings may be attributable to metabolic and mechanical factors and micro- and macroscopic abnormalities in renal tract development [130]. Cysts, abnormal nephron morphology, and pre-caliceal duct or tubular ectasias are present in ≈15% of ADPKD patients with KSD, as are abnormalities in renal acidification pathways that promote acidic urine and hypocitraturia [125, 128].

ADPKD is caused by mutations in the *PKD1* or *PKD2* genes, which encode polycystin-1 and polycystin-2, respectively. Polycystin-1 is involved in cell-cycle regulation and intracellular calcium transport and is concentrated in the primary cilia of renal epithelial cells. Polycystin-2, a member of the voltage-activated calcium channel family, is also concentrated in the primary cilia of renal epithelial cells. Defective polycystins can lead to abnormalities in primary cilia formation and function, including elongation following polycystin inactivation. These defects impact cilia in the embryonic node, which is crucial for establishing the left-right axis during embryogenesis, and in renal tubular epithelial cells, leading to impaired calcium and citrate handling and increasing the risk of stone formation [131, 132]. Moreover, defects in polycystin-2 disrupt calcium signalling in the endoplasmic reticulum [133–135]. Studies using porcine kidney cell lines demonstrate that lower polycystin-2 or inositol 1,4,5-trisphosphate expression, a second messenger that mediates intracellular calcium release, can trigger cyst formation and that cyst

growth can continue independently of cilia function [135]. Given the critical role of intracellular calcium signalling in monogenic disorders associated with KSD, these findings may represent a potential shared pathophysiological mechanism.

Patients with ADPKD frequently form uric acid stones due to hypocitraturia and low urinary pH, which result from defects in ammonia excretion [136, 137]. They also present with hyperuricosuria, hyperoxaluria, and hypercalciuria, further heightening their lithogenic risk. Cystic expansion and structural distortion of renal tubules in ADPKD can cause urinary stasis, again, promoting stone formation. Medullary cysts and fibrosis create areas of altered urine flow and impaired drainage, facilitating crystal formation.

A bi-directional relationship between cyst formation and KSD has been suggested [126]. KSD in ADPKD patients is associated with significant pain and is an independent risk factor for progression to end-stage renal disease in observational studies [138–140]. Abnormal renal development, duct and tubular ectasias can promote nephrocalcinosis and nephrolithiasis through an altered urine phenotype that promotes stone formation. In rodent models of polycystic kidney disease, renal tubular microcrystals trigger cystogenesis, potentially as a renoprotective response. Lodged microcrystals activate mTOR and STAT3 signalling pathways, causing the renal tubule to dilate via hypertrophy to enable crystals to be flushed out [126, 130]. While this process is reversible in healthy kidneys, through inactivation of the mTOR and STAT3 pathways, individuals with a genetic predisposition to renal disease, such as ADPKD, exhibit persistent pathway activation. This results in sustained tubular dilation and proliferation, ultimately leading to cystogenesis [130].

Similarly, medullary sponge kidney, a rare renal disorder characterised by ectasia and cystic dilatation of the intrapapillary portions of the medullary collecting ducts, is strongly associated with KSD. Medullary sponge kidney is found in ≈3% to 20% of recurrent stone formers [141–143]. Patients have microcysts in the papilla and medulla, multiple calcifications, dRTA, hypocitraturia, and hypercalciuria which all contribute to their increased lithogenic risk.

While the heightened risk of KSD in ADPKD and medullary sponge kidney is

established, the relationship between non-polycystic, "simple" renal cysts and KSD remains unclear. Solitary and multiple renal simple cysts are the most common cystic lesions in the adult kidney and are typically unilateral and located in the renal cortex. These cysts arise in non-diseased kidneys and are incidentally identified in ≈5% of individuals undergoing abdominal ultrasound, with prevalence increasing to ≈20% in 40-year-olds and ≈33% in those over 60 years of age [144, 145]. Generally, simple renal cysts are considered benign and are rarely observed at birth, suggesting an acquired, rather than congenital, origin. They are thought to result from obstructive, degenerative, stromal-epithelial disruptive, or neoplastic mechanisms in the distal convoluted tubule or collecting ducts [146, 147]. However, studies report that up to 86% of individuals with simple renal cysts have hypercalciuria, hypocitraturia, or a family history of KSD [148]. Additionally, patients with renal cysts identified through ultrasound health check-ups show over a two-fold greater frequency of KSD compared to controls [149]. If patients with renal cysts have an increased risk of KSD, nephrologists and urologists may need to consider whether targeted lifestyle advice or medical prophylaxis would be beneficial to mitigate the burden of KSD-related complications, including pain, infection, and progressive renal function decline.

#### Distal renal tubular acidosis

dRTA leads to alkaline urine due to dysfunction in the $\alpha$-intercalated cells of the distal renal tubule. These cells are responsible for acid secretion, and their impairment results in a higher proportion of phosphate in divalent and trivalent forms in the renal tubule, increasing the risk of calcium phosphate stone formation [150]. Over one third of patients with calcium phosphate stones suffer from dRTA and cannot regulate urine pH properly [73, 151, 152]. Furthermore, acidaemia exacerbates stone formation by increasing bone resorption to buffer excess $H^+$ ions, resulting in elevated urinary calcium and phosphate concentrations alongside reduced urinary citrate levels. Citrate, a natural inhibitor of stone formation, as discussed in Section 1.1.4, complexes with calcium to prevent its binding with oxalate or phosphate. Thus, lower urinary citrate concentrations promote

a lithogenic urinary profile [150]. Therapeutic use of potassium alkali salts can increase urinary citrate levels, reduce the risk of stone formation, and help to preserve bone density [153]. If hypercalciuria persists despite alkali treatment, thiazide diuretics may be used. dRTA can coexist with idiopathic hypercalciuria and cause calcium phosphate stones. In these cases, potassium alkali salts are less effective as the primary abnormality lies in the hypercalciuria rather than a metabolic acidosis [73, 154].

#### 1.1.6.6 Heritability of kidney stone disease

Twin studies estimate the population-level heritability of KSD as ≈50% [155]. Additionally, 35-65% of individuals with KSD have a family history of kidney stone formation, compared to only 5-20% in individuals without KSD [156].

Monogenic causes of KSD account for ≈3–7% of cases [157–159], with this figure rising to as high as 30% in those who form stones before the age of 25 years [160]. Monogenic disorders can be classified based on their associations with calcium, oxalate, cystine, uric acid, xanthine, adenine, and renal tubular metabolism (Table 1.2).

***Table 1.2. Monogenic disorders causing kidney stone disease.***

| Gene | Protein | Inheritance | Monogenic disorder |
|---|---|---|---|
| **Calcium metabolism** | | | |
| *CASR* | Extracellular calcium-sensing receptor | Autosomal dominant | Familial hypocalciuric hypercalcaemia type I |
| *GNA11* | Guanine nucleotide- binding protein subunit $\alpha - 11$ | Autosomal dominant | Familial hypocalciuric hypercalcaemia type II |
| *AP2S1* | Adaptor-related protein complex 2 subunit $\sigma$ | Autosomal dominant | Familial hypocalciuric hypercalcaemia type III |
| *VDR* | Vitamin D3 receptor | Autosomal dominant | Idiopathic hypercalciuria |
| *ADCY10* | Adenylate cyclase type 10 | Autosomal dominant | Idiopathic hypercalciuria |
| *SLC34A3* | Sodium-dependent phosphate transport protein 2C | Autosomal recessive | Hereditary hypophosphataemic rickets with hypercalciuria |
| *CYP24A1* | 1,25-dihydroxyvitamin D(3) 24-hydroxylase | Autosomal recessive | Infantile hypercalcaemia type I |
| *SLC34A1* | Sodium-dependent phosphate transport protein 2A | Autosomal recessive | Infantile hypercalcaemia type II |
| *CLDN16* | Claudin-16 | Autosomal recessive | Familial hypomagnesaemia with hypercalciuria and nephrocalcinosis |
| *CLDN19* | Claudin-19 | Autosomal recessive | Familial hypomagnesaemia with hypercalciuria and nephrocalcinosis |
| **Oxalate metabolism** | | | |
| *AGXT* | Alanine–glyoxylate aminotransferase | Autosomal recessive | Primary hyperoxaluria type I |
| *GRHPR* | Glyoxylate reductase /hydroxypyruvate reductase | Autosomal recessive | Primary hyperoxaluria type II |
| *HOGA1* | 4-hydroxy-2-oxoglutarate aldolase | Autosomal recessive | Primary hyperoxaluria type III |
| *SLC26A1* | Sulphate anion transporter 1 | Autosomal recessive | Hyperoxaluria |
| *SLC26A6* | Solute carrier family 26 member 6 | Autosomal dominant | Hyperoxaluria |
| **Cysteine metabolism** | | | |
| *SLC3A1* | $b^{0,+}$-type amino acid transporter-related heavy chain | Autosomal recessive | Cystinuria type A |

| Gene | Protein | Inheritance | Monogenic disorder |
|---|---|---|---|
| *SLC7A9* | $b^{0,+}$-type amino acid transporter 1 | Autosomal recessive | Cystinuria type B |
| **Uric acid metabolism** | | | |
| *SLC2A9* | Solute carrier family 2, facilitated glucose transporter member 9 | Autosomal recessive | Hereditary hyperuricosuria |
| *SLC22A12* | Solute carrier family 22 member 12 | Autosomal recessive | Hereditary hyperuricosuria |
| *HPRT1* | Hypoxanthine-guanine phosphoribosyltransferase | X-linked recessive | Hyperuricaemia and hyperuricosuria |
| *PRPS1* | Ribose-phosphate pyrophosphokinase 1 | X-linked recessive | Hyperuricaemia and hyperuricosuria |
| **Xanthine metabolism** | | | |
| *XDH* | Xanthine dehydrogenase/oxidase | Autosomal recessive | Type I hereditary xanthinuria |
| *MOCOS* | Molybdenum cofactor sulfurase | Autosomal recessive | Type II hereditary xanthinuria |
| **Adenine metabolism** | | | |
| *APRT* | Adenine phosphoribosyltransferase | Autosomal recessive | Adenine phosphoribosyltransferase deficiency |
| **Renal tubular metabolism** | | | |
| *SLC4A1* | Band 3 anion transport protein | Autosomal dominant | Distal renal tubular acidosis |
| *ATP6V0A4* | ATPase H+ transporting V0 subunit a4 | Autosomal recessive | Distal renal tubular acidosis |
| *ATP6V1B1* | V-ATPase subunit B 1 | Autosomal recessive | Distal renal tubular acidosis |
| *CA2* | Carbonic anhydrase 2 | Autosomal recessive | Distal renal tubular acidosis |
| *FOXI1* | Forkhead box protein I1 | Autosomal recessive | Distal renal tubular acidosis |
| *WDR72* | WD repeat-containing protein 72 | Autosomal recessive | Distal renal tubular acidosis |
| *SLC12A1* | Solute carrier family 12 member 1 | Autosomal recessive | Bartter syndrome type I |
| *KCNJ1* | ATP-sensitive inward rectifier potassium channel 1 | Autosomal recessive | Bartter syndrome type II |

| Gene | Protein | Inheritance | Monogenic disorder |
|---|---|---|---|
| *CLCNKB* | Chloride channel protein ClC-Kb | Autosomal recessive | Bartter syndrome type III |
| *BSND* | Barttin | Autosomal recessive | Bartter syndrome type IV |
| *CASR* | Extracellular calcium-sensing receptor | Autosomal recessive | Bartter syndrome type V |
| *CLCN5* | H(+)/Cl(-) exchange transporter 5 | X-linked recessive | Bartter syndrome type VI |
| *CLCN5* | H(+)/Cl(-) exchange transporter 5 | X-linked recessive | Dent disease type I |
| *OCRL* | Inositol polyphosphate 5-phosphatase | X-linked recessive | Dent disease type II |

#### 1.1.6.7 Monogenic disorders of calcium metabolism associated with kidney stone disease

**Familial hypocalciuric hypercalcemia**

Familial hypocalciuric hypercalcemia (FHH) is an autosomal dominant disorder of mineral homeostasis that results in reduced signal transduction in the CaSR signalling pathway. Patients typically present with elevated serum calcium levels, elevated or inappropriately normal PTH levels that are not suppressed despite hypercalcemia, and disproportionate hypocalciuria [161]. As in primary hyperparathyroidism, patients with FHH are often asymptomatic, with diagnosis frequently occurring through routine biochemical assessments. In some cases, FHH has been identified from clinical manifestations of rickets [161].

FHH is distinguishable from primary hyperparathyroidism by calculating the fractional excretion of calcium over 24 hours, assuming vitamin D levels are replete [74]. A fractional excretion of calcium >1% is consistent with a diagnosis of FHH. Differentiating FHH from primary hyperparathyroidism is important, as surgical intervention is neither indicated nor curative in FHH. FHH exists in three forms- FHH1, FHH2, and FHH3- corresponding to inactivating mutations in the Calcium Sensing Receptor gene (*CASR*), encoding the CaSR, the G Protein Subunit Alpha 11 gene (*GNA11*) encoding the G$\alpha$11 protein, a subunit of a guanine nucleotide-binding protein, and the Adaptor Related Protein Complex 2 Subunit $\sigma$ 1 gene (*AP2S1*) encoding the protein adaptor protein-2 complex subunit $\sigma$, respectively [162–164].

**Autosomal dominant hypocalcaemia**

Autosomal dominant hypocalcaemia arises from gain-of-function mutations in components of the CaSR signalling pathway [163, 165]. These mutations increase the sensitivity of the CaSR to extracellular calcium concentrations, leading to reduced serum calcium levels and increased urinary calcium excretion from diminished renal tubular calcium reabsorption. Additionally, patients may present with hypoparathyroidism [166]. Autosomal dominant hypocalcaemia is classified

based on the underlying genetic mutation: type 1 is caused by abnormalities in the CaSR itself, encoded by *CASR*, while type 2 is associated with variants in the G protein subunit $\alpha$11, encoded by *GNA11*, which is implicated in CaSR signal transduction.

### Autosomal dominant idiopathic hypercalciuria

Autosomal dominant idiopathic hypercalciuria is a syndrome characterised by unexplained hypercalciuria despite normal serum calcium levels, distinguishing it from primary hyperparathyroidism [73]. Affected individuals display higher-than-average fractional intestinal calcium absorption and renal calcium excretion. Normally, this would suppress PTH release. Indeed, patients with idiopathic hypercalciuria often have normal or low PTH concentrations [72]. Vitamin D receptor (*VDR*) and adenylate cyclase 10 (*ADCY10*) genes are implicated in autosomal dominant idiopathic hypercalciuria [167, 168]. Variants in *VDR*, encoding the vitamin D receptor, are associated with kidney stones and increased urinary calcium excretion in genetic linkage studies [168]. *ADCY10* encodes adenylate cyclase 10 which is regulated by bicarbonate concentrations and produces cyclic adenosine monophosphate and phosphate from adenosine triphosphate [167]. The *ADCY10* region was mapped in studies of three kindreds with absorptive hypercalciuria and subsequently identified as a candidate gene by Reed et al. [167]. Patients with variants in *ADCY10* reportedly have lower vertebral bone density and over double the increased risk of autosomal dominant idiopathic hypercalciuria [169].

### Familial hypomagnesaemia with hypercalciuria and nephrocalcinosis

Familial hypomagnesaemia with hypercalciuria and nephrocalcinosis is a rare, autosomal recessive disorder caused by biallelic loss-of-function mutations in the claudin-16 and claudin-19 genes (*CLDN16* and *CLDN19*), which encode claudin-16 and claudin-19, respectively. These claudins are expressed in the thick ascending limb of the loop of Henle where they form tight junctions in conjunction with occludin and junctional adhesion molecules [170, 171]. Claudin-16 and claudin-19 regulate cation selectivity, facilitating renal tubular calcium and magnesium reabsorption,

driven by an electrochemical gradient. Loss of function in these proteins results in renal calcium and magnesium wasting, leading to nephrocalcinosis and progressive renal dysfunction.

Patients should avoid vitamin D supplementation as it can exacerbate hypercalciuria. Magnesium supplementation is beneficial to offset urinary losses [172]. While claudin-16 is expressed only in the kidney, claudin-19 is also expressed in retinal pigment epithelium. Thus, claudin-19 variants are associated with ocular abnormalities including macular colobomata, myopia, and horizontal nystagmus [173].

#### 1.1.6.8 Monogenic disorders of phosphate metabolism associated with kidney stone disease

**Hereditary hypophosphataemic rickets with hypercalciuria**

Hereditary hypophosphataemic rickets with hypercalciuria results from biallelic loss-of-function mutations in the solute carrier family 34 member 3 (*SLC34A3*) gene, which encodes the sodium-dependent phosphate transport protein 2C (NaPi-IIc). NaPi-IIc is expressed in the proximal renal tubule and plays a critical role in sodium-dependent phosphate transport. Affected patients typically present with short stature, renal phosphate wasting, hypercalciuria, normocalcaemia, increased serum 1,25-dihydroxyvitamin D, and lower PTH levels [174–178].

Renal phosphate wasting can lead to hypercalciuria by reducing FGF23-driven 1$\alpha$-hydroxylase inhibition, thereby increasing levels of 1$\alpha$,25-dihydroxyvitamin $D_3$. This promotes increased intestinal calcium absorption and contributes to hypercalciuria. Patients can be managed with oral phosphate supplements to raise serum phosphate concentrations and alleviate co-existing rickets [174–178].

Heterozygous mutations in *SLC34A3* cause a milder, hypomorphic phenotype, where patients have an elevated risk of KSD, potentially accompanied by higher urinary calcium excretion and lower bone mineral density [176, 179]. Association studies show that rare variants in *SLC34A3* fall into an intermediate category of pathogenicity, and do not cause fully penetrant Mendelian disease but confer a

higher risk of KSD compared to the aggregated effects of common risk alleles [180].

Patients with heterozygous mutations in the solute carrier family 34 member 1 (*SLC34A1*) gene, encoding NaPi-IIa, also expressed in the proximal renal tubule, may present with hypophosphataemia, osteoporosis, and kidney stones [181]. Variants in *SLC34A1* are associated with hypophosphataemic hypercalciuric stone formation without obvious skeletal abnormalities in population-based studies [160, 179, 182]. Additionally, biallelic mutations in *SLC34A1* cause infantile hypercalcaemia type 2, characterised by vomiting, dehydration, and nephrocalcinosis [183].

**Infantile hypercalcaemia**

Infantile hypercalcaemia falls into two classes: type 1 is attributed to biallelic loss-of-function mutations in the cytochrome P450 family 24 subfamily A member 1 gene (*CYP24A1*) [184], while type 2 is attributed to biallelic loss-of-function mutations in *SLC34A1* [183]. Affected patients typically exhibit elevated serum and urine calcium concentrations and nephrocalcinosis. There is considerable variability in phenotype and age of presentation; infants may present with dehydration, polyuria, lethargy, and failure to thrive, while some individuals are asymptomatic until adulthood when vitamin D supplementation in pregnancy can precipitate symptoms [185]. *CYP24A1* encodes 24-$\alpha$-hydroxylase, which hydroxylates active 1,25-dihydroxyvitamin D to inactive 1,24,25-trihydroxyvitamin D and converts 25-hydroxyvitamin D to 24,25-dihydroxyvitamin D. Patients with impaired 24-$\alpha$-hydroxylase function are highly sensitive to vitamin D. They can benefit from avoiding vitamin D supplements and from treatment with fluconazole or ketoconazole to inhibit 1-$\alpha$-hydroxylase and 1,25-dihydroxyvitamin D synthesis [186, 187].

#### 1.1.6.9 Monogenic disorders of oxalate metabolism associated with kidney stone disease

Primary hyperoxaluria is an autosomal recessive inborn error of metabolism. Patients often present in childhood with urinary oxalate excretion levels exceeding 40–45 mg per 24 hours. Accumulated insoluble calcium oxalate crystals can lead to nephrocalcinosis, kidney stones, tubular obstruction, and, particularly in cases of concurrent tubulointerstitial oxalate nephropathy, CKD [88].

Primary hyperoxaluria is categorised according to the underlying monogenic abnormality. Type 1 primary hyperoxaluria constitutes ≈80% of cases and affects 1-3 individuals per million [188]. This results from biallelic mutations in *AGXT*, encoding alanine glyoxylate aminotransferase, a vitamin B6-dependent hepatic peroxisomal enzyme (Figure 1.5). Decreased enzyme function diminishes glyoxylate metabolism to glycine, leaving residual glyoxylate which is converted to oxalate or glycolate which then depletes the available citrate in urine [88, 188].

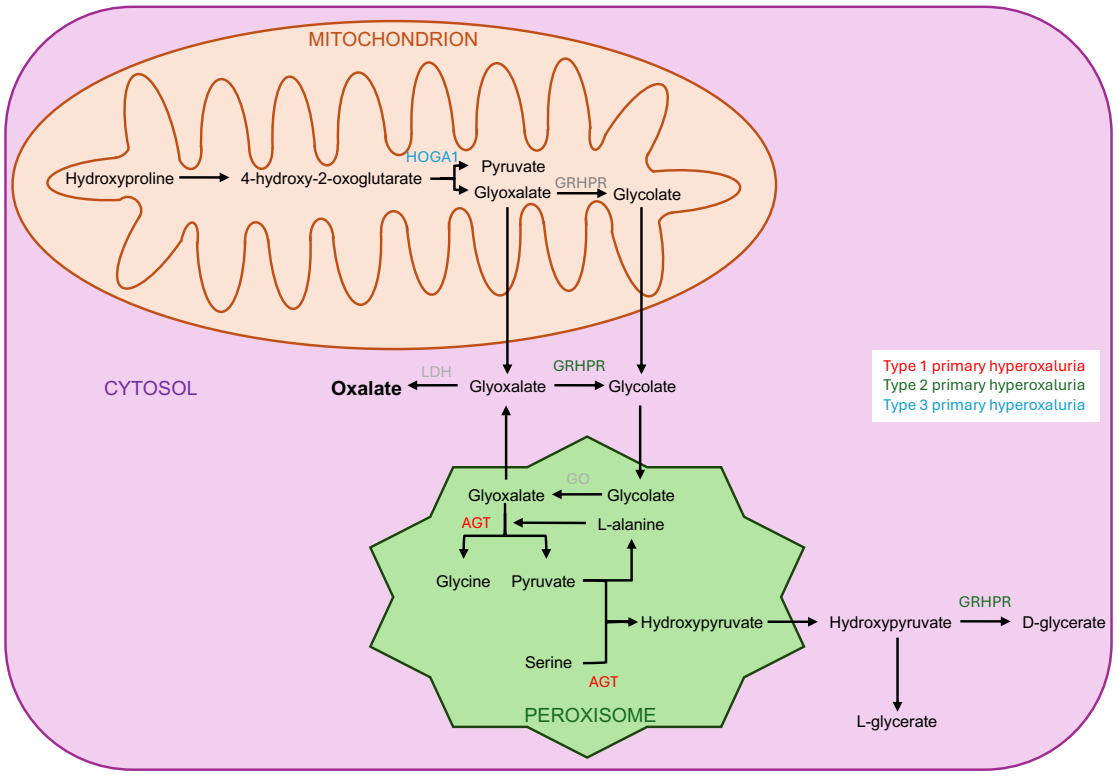



***Figure 1.5. Enzyme deficiencies in type 1, 2, and 3 primary hyperoxaluria.***

Type 1, 2, and 3 primary hyperoxaluria are associated with peroxisomal AGT (red), cytosolic GRHPR (green), and mitochondrial HOGA (blue) deficiencies, respectively.

GO=Glycolate oxidase; GRHPR=Glyoxylate reductase-hydroxypyruvate reductase; HOGA=4-hydroxy-2-oxoglutarate aldolase; LDH=Lactate dehydrogenase.

Type 2 primary hyperoxaluria accounts for ≈10% of cases and arises from biallelic mutations in glyoxylate and hydroxypyruvate reductase, *GRHPR*, that encodes glyoxylate reductase/hydroxypyruvate reductase (GR/HPR) (Figure 1.5). Dysfunctional GR/HPR causes glyoxylate to be converted to oxalate, rather than glycolate, and hydroxypyruvate to L-glycerate, rather than D-glycerate, leading to increased urinary L-glyceric acid and oxalate excretion [189]. Patients with type 2 primary hyperoxaluria generally present with a milder phenotype than those with type 1. Affected individuals typically have normal urinary citrate concentrations, because GR/HPR dysfunction causes less oxalate to be produced overall, hence there is less depletion of urinary citrate levels, which partially mitigates disease severity.

Type 3 primary hyperoxaluria represents ≈10% of cases and results from biallelic mutations in 4-hydroxy-2-oxoglutarate aldose 1 (*HOGA*) encoding 4-hydroxy-2-oxoglutarate aldolase (HOGA1), an enzyme implicated in hydroxyproline catabolism (Figure 1.5). Affected individuals have higher urinary oxalate, 4-hydroxy-L-glutamate, 4-hydroxy-2-oxoglutarate (HOG), and 2,4-dihydroxyglutarate concentrations [190].

Oxalate handling in the kidneys is also influenced by glomerular filtration, which depends on plasma oxalate levels, and by tubular secretion and reabsorption, which involves specific transporters. The sulphate anion transporter 1 (SAT-1), encoded by *SLC26A1*, mediates oxalate-sulphate exchange at the basolateral membrane. Loss-of-function mutations in *SLC26A1* are associated with hyperoxaluria and kidney stones in animal models and humans [191]. Additionally, chloride-oxalate exchange at the apical tubular membrane is facilitated by solute carrier family 26 member 6 (SLC26A6), encoded by *SLC26A6*, and is implicated in cases of kidney stones in humans and mice [192, 193].

Patients with primary hyperoxaluria are advised to drink at least 3 litres of fluid daily and may benefit from urinary alkalinisation to prevent calcium oxalate crystal formation. In the absence of CKD, dietary restriction of oxalate is not recommended [194]. An emerging therapeutic approach involves ribonucleic acid interference

technologies such as lumasiran and nedosiran [195–198]. Lumasiran reduces peroxisomal glycolate oxidase expression, thereby inhibiting the conversion of glycolate to glyoxylate. Conversely, nedosiran reduces hepatic lactate dehydrogenase expression, preventing over-production of oxalate. Both therapies lower urinary oxalate excretion. A double-blind, phase 3 trial of lumasiran in 39 patients with primary hyperoxaluria reported that, after six months of therapy, 84% of patients taking lumasiran had 24-hour urinary oxalate excretion no higher than 1.5 times the upper limit of normal, compared with 0% in the placebo group [195]. Similarly, a double-blind, placebo-controlled study of nedosiran in 35 patients with primary hyperoxaluria found that half of participants receiving nedosiran achieved normal or near-normal 24-hour urinary oxalate excretion on $\geq$2 consecutive visits starting at day 90, compared with none of the patients randomised to placebo [195, 196].

#### 1.1.6.10 Monogenic disorders of cysteine metabolism associated with kidney stone disease

Cystinuria is an autosomal recessive disorder characterised by defects in cysteine transport in the renal proximal tubule, leading to excessive cystine excretion and cystine stone formation. It is a relatively common disorder, affecting $\approx$1 in 7,000 people globally. It accounts for about 1% of kidney stones in adults and 7% in paediatric populations [34]. While the disease can manifest at any age, 80% of patients experience their first stone by age 20 years.

Cystinuria arises when cysteine, which dimerises to form cystine, and dibasic amino acids (ornithine, lysine, and arginine) are not resorbed after free filtration at the glomerulus. The inability to resorb cysteine from tubular fluid may be attributed to mutations in genes encoding amino acid transport proteins, specifically solute carrier family 3 member 1 (*SLC3A1*) and solute carrier family 7 member 9 (*SLC7A9*). *SLC3A1* encodes rBAT, the neutral and basic amino acid transport protein, a heavy subunit of heteromeric amino acid transporters, while *SLC7A9* encodes the $b^{0,+}$ type amino acid transporter 1 ($b^{0,+}$AT), a light subunit. The heterodimerisation of these subunits facilitates cysteine, ornithine, lysine, and arginine resorption from the renal tubule [199]. Importantly, $b^{0,+}$AT, which encompasses the catalytic element

of the transporter, relies on rBAT for trafficking to the plasma membrane of renal proximal tubule cells [34].

Type A cystinuria arises from mutations in *SLC3A1*, type B cystinuria from mutations in *SLC7A9*, and type AB cystinuria from mutations in both genes [200].

For children over two years old, the age at which renal tubular transport has matured, with a positive family history, biochemical analysis may be performed to identify elevated urinary cystine concentrations, along with genetic testing. Treatment focusses on increasing urinary cystine solubility, by optimising urinary volume and pH to above 7.5 through hydration, and oral alkalinising medications. Additionally, urinary cystine levels can be reduced by limiting the intake of animal protein and sodium, and by using cystine binding agents, for example thiols, penicillamine, and $\alpha$-mercaptopropionylglycine [34].

#### 1.1.6.11 Monogenic disorders of uric acid metabolism associated with kidney stone disease

Hereditary hyperuricosuria is a rare disorder caused by abnormal renal tubular urate reabsorption, PRPPS (phosphoribosyl pyrophosphate synthase) overactivity, or HGPRT (hypoxanthine-guanine phosphoribosyltransferase) deficiency (Figure 1.4) [124]. PRPPS, encoded by *PRPS1* on the X-chromosome, catalyses ribose-5-phosphate phosphorylation to phosphoribosyl pyrophosphate in purine synthesis [201]. Overactive PRPPS leads to hyperuricaemia and hyperuricosuria [202, 203]. HGPRT, encoded by *HPRT1*, salvages hypoxanthine by converting it to inosine monophosphate (Figure 1.4). Loss-of-function mutations in *HPRT1* increase hypoxanthine levels which are converted to uric acid via xanthine oxidase [204]. The extent of HGPRT deficiency determines the clinical presentation: complete deficiency causes Lesch-Nyhan syndrome, characterised by hyperuricemia, hyperuricosuria, uric acid stones, renal failure, and neurological symptoms; partial deficiency leads to Kelley-Seegmiller syndrome, which lacks the neurological features [204]. Finally, some heterozygous females have uric acid nephrolithiasis, gout arthropathy, and

hearing loss [201].

Hereditary renal uric acid wasting and uric acid nephrolithiasis can also result from autosomal mutations in the solute carrier genes *SLC22A12* and *SLC2A9*. *SLC22A12* encodes URAT1, a proximal tubular urate transporter, while *SLC2A9* encodes GLUT9, a glucose transporter [205].

Treatment aims to increase urinary uric acid solubility through high fluid intake and urinary alkalinisation. Xanthine oxidase inhibitors, such as allopurinol or febuxostat, can reduce uric acid overproduction from PRPPS hyperactivity or HGPRT deficiency. Lowering uric acid production reduces the renal uric acid load, which can alleviate KSD associated with *SLC22A12* or *SLC2A9* mutations. However, xanthine oxidase inhibitors can increase levels of upstream oxypurines (xanthine and hypoxanthine), risking cellular damage and xanthinuria. Therefore, careful dose management is essential [206].

#### 1.1.6.12 Monogenic disorders of adenine metabolism associated with kidney stone disease

Adenine phosphoribosyltransferase (APRT) catalyses synthesis of 5'-adenosine monophosphate and inorganic pyrophosphate from adenine and phosphoribosyl pyrophosphate. When this pathway is disrupted, a greater proportion of adenine undergoes oxidation by xanthine dehydrogenase to form 2,8-dihydroxyadenine, which accumulates in urine [207, 208]. At urinary pH values below 8.5, 2,8-dihydroxyadenine is insoluble and crystallises. These crystals can form radiolucent stones and cause crystalline nephropathy, leading to end-stage renal failure due to tubular obstruction or crystal deposition [201, 207].

APRT deficiency results from homozygous or compound heterozygous mutations in *APRT*, inherited in an autosomal recessive pattern. More than 40 pathogenic mutations have been identified in ≈400 individuals across over 25 countries [209, 210]. However, with homozygous APRT loss-of-function mutations estimated to occur in 1 in 50,000 to 1 in 100,000 individuals, the prevalence of this disorder is likely under-reported [208, 211, 212]. APRT deficiency is classified into type I

(complete loss of enzymatic function) and type II (partial loss of enzymatic function) based on *in vitro* studies, although no evidence supports a correlation between *in vitro* and *in vivo* phenotypes [213].

Clinicians can diagnose APRT deficiency through several methods including stone composition analysis, renal biopsy, identifying 2,8-dihydroxyadenine crystals in urine, or identifying low APRT activity in erythrocytes [208]. Current recommendations advise adult patients to maintain fluid intake that achieves a urine output of 2.5 litres daily. Treatment with xanthine oxidase inhibitors, such as allopurinol or febuxostat, can reduce 2,8-dihydroxyadenine stone formation and prevent further deterioration of renal function [208, 214].

#### 1.1.6.13 Monogenic disorders of xanthine metabolism associated with kidney stone disease

Xanthinuria is an autosomal recessive disorder of defective purine metabolism characterised by large amounts of xanthine excretion in urine and a tendency to form xanthine stones [215]. This occurs in two distinct forms: type I and type II hereditary xanthinuria. Both forms promote low serum and urinary uric acid levels and elevated concentrations of xanthine and hypoxanthine, which predispose patients to kidney stones and renal impairment [215–217]. Additionally, xanthine crystals can accumulate in muscles, causing pain and cramping.

Mutations in xanthine dehydrogenase (*XDH*) cause type I hereditary xanthinuria. *XDH* encodes xanthine dehydrogenase and xanthine oxidase. In contrast, mutations in molybdenum cofactor sulfurase (*MOCOS*), encoding molybdenum cofactor sulfurase, result in type II hereditary xanthinuria. This enzyme sulphurates the molybdenum cofactor, essential for both xanthine dehydrogenase and aldehyde oxidase function. Consequently, patients with type II hereditary xanthinuria lack both enzymes [215, 216, 218, 219].

Several methods can differentiate type I from type II hereditary xanthinuria. Abnormalities in the molybdenum cofactor synthesis pathway, observed in type II xanthinuria, elevate serum sulphite concentrations in addition to the typical

biochemical abnormalities of type I xanthinuria [220]. Elevated sulphite levels lead to psychomotor retardation, failure to thrive, seizures, and hypotonia [219, 220]. Furthermore, the allopurinol loading test can distinguish between the two forms. Unlike type II, patients with type I hereditary xanthinuria metabolise allopurinol to oxypurinol via aldehyde oxidase. The additional deficiency of aldehyde oxidase in type II hereditary xanthinuria blocks this conversion.

In contrast to hereditary xanthinuria, xanthinuria as a complex disease can also result from congenital absence of a molybdenum cofactor. Mutations in molybdenum cofactor synthesis 1 (*MOCS1*), molybdenum cofactor synthesis 2 (*MOCS2*), or gephyrin (*GPHN*) disrupt the biosynthesis of this cofactor [220]. As with type II hereditary xanthinuria, patients with congenital molybdenum cofactor deficiency exhibit elevated sulphite concentrations and may present with neonatal seizures, severe neurological symptoms, and dysmorphic features [221]. Finally, xanthine stones can also form as a consequence of treating severe hyperuricaemia with allopurinol, such as in cases of myeloproliferative disorders or Lesch-Nyhan syndrome.

#### 1.1.6.14 Monogenic disorders of renal tubule metabolism associated with kidney stone disease

**Distal renal tubular acidosis**

Autosomal dominant disorders of dRTA may arise from mutations in the solute carrier family 4 member 1 (*SLC4A1*) gene, which encodes a chloride–bicarbonate exchanger expressed on the basolateral membrane of $\alpha$-intercalated cells [153]. Autosomal recessive disorders of dRTA may be the result of mutations in ATPase $H^+$ transporting V1 subunit B1 (*ATP6V1B1*), ATPase $H^+$ transporting V0 subunit A4 (*ATP6V0A4*) [222, 223], carbonic anhydrase 2 (*CA2*), forkhead box I1 (*FOXI1*) [224], or WD repeat domain 72 (*WDR72*) [225] genes [226]. *ATP6V1B1* and *ATP6V0A4* encode subunits B1 and A4, respectively, of the $H^+$-ATPase pump in $\alpha$-intercalated cells, which are essential for acid secretion and maintaining urine pH balance [222, 223]. Patients harbouring variants in *ATP6V1B1* can also have sensorineural hearing loss as the $H^+$-ATPase pump plays a role in maintaining the pH balance of the ear endolymph. Alkaline conditions in the inner ear may damage

hair cells and impair hearing [223].

Mouse studies show that *FOXI1* regulates $H^+$-ATPase pump subunits B1 and A4 and anion exchanger 1 expression, all three of which are critical to $\alpha$-intercalated cells maintaining acid–base balance [224]. Given the role of *FOXI1* in regulating $H^+$-ATPase pump subunit B1 expression, patients with variants in this gene can also present with early-onset sensorineural deafness [224].

*WDR72* encodes WD Repeat Domain 72 (WDR72) which is implicated in amelogenesis imperfecta, a disease of abnormal dental enamel formation [227]. Homozygous loss of function WDR72 variants are also associated with dRTA and kidney stones [228]. Homozygous variants in *CA2*, encoding carbonic anhydrase 2 which catalyses the conversion of carbon dioxide into $H^+$ and bicarbonate ions in renal $\alpha$-intercalated cells, are implicated in dRTA and KSD [229, 230].

### Bartter syndrome

Bartter syndrome is caused by abnormal sodium chloride reabsorption in the thick ascending limb of the loop of Henle. Ordinarily, calcium ions are reabsorbed from the renal tubule via the $Ca^{2+}$-ATPase and $Na^+/Ca^{2+}$ exchangers on the basolateral membrane, while chloride ions pass through the voltage-gated chloride channel (ClC-Kb) on the apical membrane (Figure 1.6) [226]. There are six types of Bartter syndrome, classified by the underlying genetic mutations. They share a common phenotype of hypokalaemic alkalosis, hypotension, hyper-reninaemic hyperaldosteronism, increased urinary prostaglandin excretion, and hypercalciuria with nephrocalcinosis. Patients are treated with non-steroidal anti-inflammatory drugs to reduce renal prostaglandin production, and spironolactone to correct the hypokalaemic metabolic alkalosis.

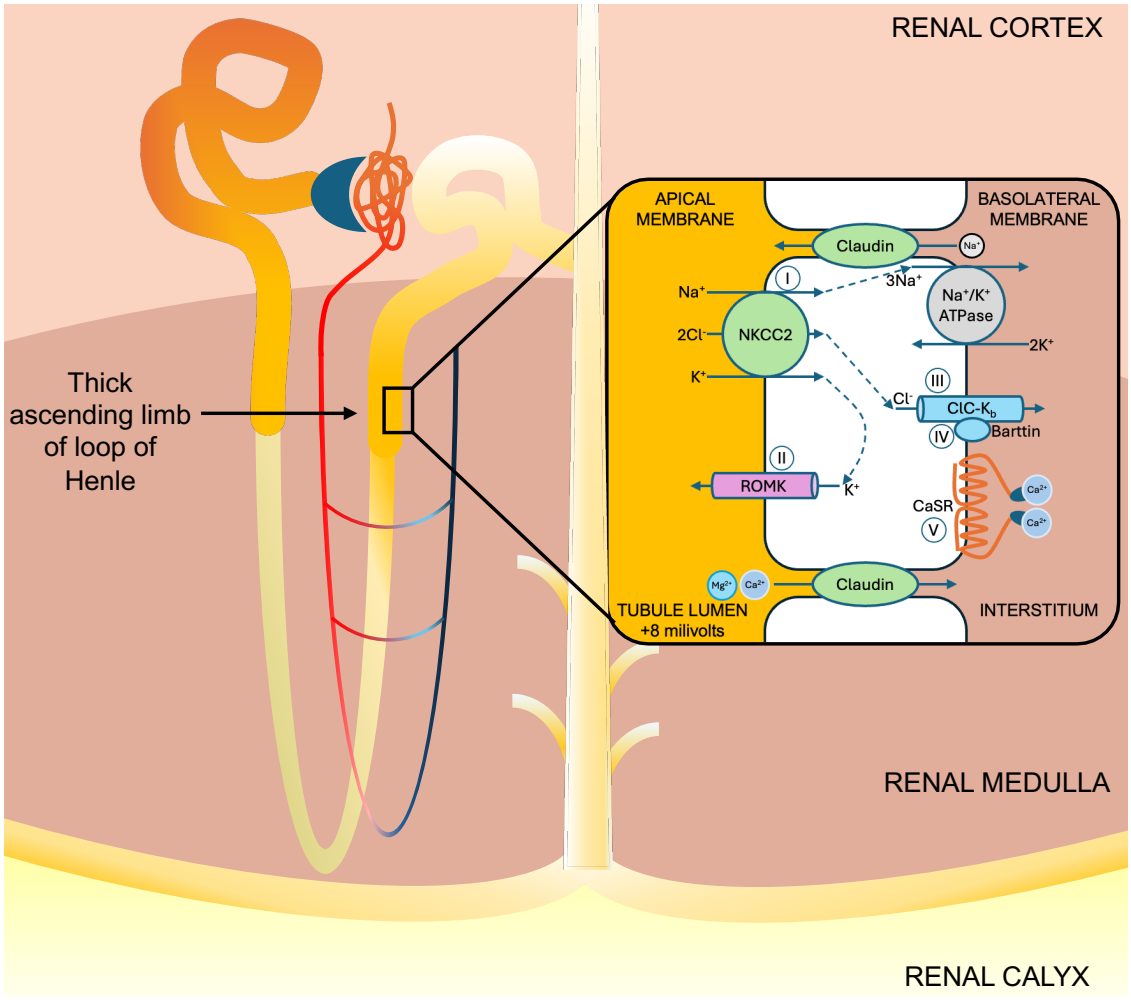



***Figure 1.6. Pathophysiology of Bartter syndrome.***

An epithelial cell of the thick ascending limb of the Loop of Henle. The sodium-potassium-chloride co-transporter 2 (NKCC2) and renal outer medullary potassium channel (ROMK) generate a lumen-positive potential difference driving transcellular and paracellular calcium ($Ca^{2+}$) resorption. This is regulated by the calcium-sensing receptor (CaSR) which inhibits NKCC2 activity. Chloride ions ($Cl^{-}$) pass through the voltage-gated chloride channel CLC-Kb, which requires Barttin for proper function. Sodium ions ($Na^{+}$) are transported across the basolateral membrane via $Na^{+}/K^{+}$-ATPase while $Ca^{2+}$ is removed through $Ca^{2+}$-ATPase and $Na^{+}/Ca^{2+}$ exchangers (not shown). Claudins regulate tight junction permeability to cations including magnesium ($Mg^{2+}$), $Na^{+}$, and $Ca^{2+}$. Transport proteins labelled I–V correspond to the dysfunctional proteins responsible for Bartter syndrome types I–V, respectively. The chloride/proton exchanger ClC-5, associated with Bartter syndrome type VI, is not shown. Most ClC-5 is located in subapical endosomes, with only a small portion present at the apical membrane.

Bartter syndrome type I is attributed to mutations in solute carrier family 12 member 1 (*SLC12A1*) which encodes the kidney-specific NKCC2 (Figure 1.6). Type II is due to mutations in the potassium inwardly rectifying channel subfamily J member 1 (*KCNJ1*) gene, encoding ROMK (Figure 1.6). Type III Bartter syndrome is caused by variants in chloride voltage-gated channel Kb (*CLCNKB*), which encodes the voltage-gated chloride channel ClC-Kb (Figure 1.6). Type IV Bartter syndrome

results from mutations in the Barttin CLCNK type accessory subunit $\beta$ gene (*BSND*), which encodes an essential beta subunit for CLC chloride channels (Barttin) (Figure 1.6). Type IV Bartter syndrome is also linked to deafness as Barttin is expressed in the potassium-secreting marginal cells of the scala media of the inner ear [231]. Type V is secondary to mutations in *CASR*, encoding the CaSR (Figure 1.6) [231–235]. Type VI Bartter syndrome is described in one boy who displayed signs of vomiting, rickets, hypotrophy, metabolic alkalosis, hypochloraemia, hyperkalaemia, hypophosphataemia, hypomagnesaemia, alkaline urine, and hypercalciuria [236]. Genetic sequencing revealed a single hemizygous nucleotide transition (c.1073G>A) in *CLCN5* exon 7, which causes a p.G261E amino acid substitution in chloride channel protein 5 (ClC-5), also known as $H^+/Cl^-$ exchange transporter 5 [236].

Types I-IV exhibit autosomal recessive inheritance, type V has autosomal dominant inheritance, and type VI has been described with X-linked recessive inheritance. Mutations in *CLCN5* are more commonly associated with Dent disease.

### Dent disease

Dent disease is an X-linked recessive disorder of the proximal renal tubule characterised by low molecular weight proteinuria, hypercalciuria, KSD, nephrocalcinosis, and renal impairment [212]. There are two subtypes, classified by their associated genetic mutations.

Type 1 Dent disease is caused by loss-of-function mutations in *CLCN5*. ClC-5 acts in concert with the V-ATPase in the proximal tubule, thick ascending limb of the loop of Henle, and the $\alpha$-intercalated cells of the collecting duct [237]. Variants in *CLCN5* result in abnormal endosomal acidification, disrupted endosomal trafficking, and reduced tubular solute reabsorption. Type 2 Dent disease is secondary to mutations in OCRL inositol polyphosphate-5-phosphatase (*OCRL*), which encodes an inositol polyphosphate 5-phosphatase. Variants in *OCRL* are also implicated in the oculocerebrorenal syndrome of Lowe, which is associated with congenital cataracts, glaucoma, mental retardation, decreased seizure threshold, muscular hypotonia, and behavioural problems [234, 238]. Inositol polyphosphate

5-phosphatase plays a role in vesicle trafficking, phagocytosis, cell adhesion and migration, cellular polarity, ciliogenesis, cytokinesis, and intracellular signalling [239]. The precise mechanisms by which *OCRL* mutations cause Dent disease are incompletely understood but are postulated to involve a loss of phosphatidylinositol 4,5-bisphosphate 5-phosphatase activity or a reduction in functional protein due to misfolding or decreased expression [234].

**Concluding remarks on monogenic associations with kidney stone disease**

As explained in Section 1.1.6.6, while monogenic conditions associated with KSD provide important insights into its pathophysiology, they account for only a minority of cases. The predominant genetic contribution to KSD at the population level arises from polygenic risk, where multiple common variants each exert small effects that collectively increase susceptibility. Having outlined the monogenic causes, I will now discuss the insights gained on the polygenic basis of KSD from large-scale genetic association studies and candidate gene studies.

#### 1.1.6.15 The landscape of genetic association studies in kidney stone disease

**Common polygenic variants associated with kidney stone disease**

At the start of my DPhil studies, in 2022, four publications had identified genetic variants associated with KSD through genome-wide association studies (GWAS, outlined in detail in Section 1.2.2) in large, population-based cohorts [158, 240–242]. Our research group published additional analyses in 2023 [119] as did Hao et al. [243] (Table 1.3). GWAS have identified genetic variants in proximity to *CASR* [119, 242, 243], *CYP24A1* [119, 158, 242, 243], and *CLDN14* [119, 158, 240, 242, 243] associated with KSD. These findings suggest key roles for the CaSR signalling pathway (*CASR*), vitamin D metabolism (*CYP24A1*), and tight junction assembly and cell-cell adhesion in the thick ascending limb of the loop of Henle in KSD pathophysiology *CLDN14*.

***Table 1.3. Published genome-wide association studies of kidney stone disease.***

| Author | Title | Ancestry | Cases | Controls | Independent associations |
|---|---|---|---|---|---|
| Thorleifsson et al. 2009 [240] | Sequence variants in the *CLDN14* gene associate with kidney stones and bone mineral density. | European | 3,773 | 42,510 | 1 |
| Urabe et al. 2012 [241] | A genome-wide association study of nephrolithiasis in the Japanese population identifies novel susceptible loci at 5q35.3, 7p14.3, and 13q14.1. | East Asian | 5,892 | 17,809 | 3 |
| Oddsson et al. 2015 [242] | Common and rare variants associated with kidney stones and biochemical traits. | European | 5,419 | 279,870 | 4 |
| Howles et al. 2019 [158] | Genetic variants of calcium and vitamin D metabolism in kidney stone disease. | European | 6,536 | 388,508 | 10 |
| | | Trans-ancestry | 12,123 (UK:6,536 Japan:5,587) | 417,378 (UK:388,508 Japan:28,870) | 20 |
| Lovegrove et al. 2023 [119] | Central Adiposity Increases Risk of Kidney Stone Disease through Effects on Serum Calcium Concentrations. | European meta-analysis | 17,101 (UK:8,504 FinnGen:8,597) | 721,947 (UK:388,819 FinnGen:333,128) | 47 |
| | | European males | 5,633 | 176,738 | 13 |
| | | European females | 2,871 | 212,081 | 3 |
| Hao et al. 2023 [243] | Integrative genome-wide analyses identify novel loci associated with kidney stones and provide insights into its genetic architecture | European meta-analysis | 17,969 (UK:9,372 FinnGen:8,597) | 702,230 (UK:369,102 FinnGen:333,128) | 44 |

### Rare polygenic variants associated with kidney stone disease

Studies focusing on rare pathogenic variants associated with KSD have largely been conducted in sub-specialist renal or recurrent clinics, where individuals are at higher risk of monogenic forms of KSD compared to those in general population cohorts [157, 180].

At the time of commencing my DPhil, there was only one published exome-wide association study (ExWAS) of KSD [180]. Sadeghi-Alavijeh et al. (2022) considered data from the 100,000 genomes cohort (374 KSD cases and 24,930 controls) and identified enrichment of rare, monogenic, predicted damaging variants in *SLC34A3* in 5% participants with KSD compared with 1.6% of controls without KSD [244]. The genebass.org database tested for associations of genomic regions with KSD in a subset of the UK Biobank population with incomplete case ascertainment (5,879 cases; 388,962 controls; heritability 7.34%), but failed to identify any genes with p-value for association $<2.5\times10^{-7}$ (the study's defined threshold gene-based

tests) and identified only one variant (21-36461453-G-A; c.243C>T; p=$2.61\times10^{-10}$) with a p-value for association $<8.0\times10^{-9}$ (the study's defined p-value threshold for single-variant tests) [245].

**Limitations of existing genetic association studies in kidney stone disease**

Despite the valid efforts to disentangle the role of genetic variants in KSD, salient gaps in our understanding remain unfilled. Studies to date have largely been conducted in combined-sex populations of European ancestry. Although historically there has been a disparity in the prevalence of KSD between males and females [9], only the Lovegrove et al. 2023 study [119] has investigated whether genetic factors may underlie this observation using sex-stratified GWAS. Sex-stratified analyses may uncover sex-specific genetic factors contributing to the observed disparity in KSD prevalence. Furthermore, despite the estimated heritability of KSD being $\approx 45\%$, the majority of this remains unexplained; the trans-ancestry KSD GWAS by Howles et al. describes a heritability of roughly 5% [158]. Even after the addition of rare *SC34A3* variants identified by the ExWAS by Sadeghi-Alavijeh et al. to the findings of Howles et al., the explained heritability is 14.2% [180]. Thus, most of the genetic basis of KSD remains to be elucidated. Since these studies were published, the UK Biobank has released data from primary care records, providing an opportunity for improved case ascertainment to increase the power of GWAS. Efforts towards enhancing our understanding of the genetic basis of KSD may be realised through better-powered genetic association studies, leveraging improved case ascertainment, and through trans-ancestry studies to investigate the generalisability of findings across populations. Addressing these gaps can improve our understanding of KSD pathophysiology and inform personalised prevention and treatment strategies.

### 1.1.7 Diagnosis

Kidney stones can cause a wide range of symptoms or, in some cases, none at all. The classic presentation of renal colic involves pain radiating from the *"loin to*

*groin*" [73], which can intensify as the stone reaches the vesico-ureteric junction, one of the narrowest points of the ureter. However, patients may also present with isolated loin pain, haematuria, or symptoms of life-threatening urinary tract obstruction, such as fever, vomiting, dysuria, and even multi-organ failure. The severity of symptoms does not necessarily reflect the size or seriousness of the stone [246]. Prolonged obstructive uropathy can lead to chronic renal failure, potentially requiring renal replacement therapy [247]. A definitive diagnosis requires radiological imaging, most commonly non-contrast computed tomography (CT), which has a sensitivity of 97% and specificity of 95% [248, 249], or direct visualisation of the stone after it passes through the urinary tract. A key challenge in managing KSD is identifying which patients will experience isolated stone episodes and which are at risk of recurrence. Stratifying patients in this way would allow urologists to adopt a personalised, patient-centred approach to treatment. In the UK, current guidelines recommend measuring serum calcium at initial presentation, with further PTH testing if calcium levels are abnormal. This is crucial, as ≈3–5% of KSD patients have hyperparathyroidism, which can be treated to prevent recurrence [30, 250–252].

American Urology Association and European Association of Urology guidelines advise metabolic testing in "high-risk" stone formers, including individuals under 18 years, those with disorders of mineral metabolism, renal or gastrointestinal disease, anatomical abnormalities, recurrent stones, a family history of KSD, or a genetic diagnosis linked to stone formation [6, 31]. This more comprehensive evaluation includes serum calcium, phosphate, urate, and magnesium testing, along with 24-hour urine collection to assess urine volume, pH, oxalate, citrate, and calcium concentrations [73]. In cases with a strong family history, clinicians may also screen for cystinuria.

Currently, no clinical guidelines or consensus exist on which patients should undergo genetic screening to facilitate genetic counselling, clarify clinical uncertainty, or guide treatment decisions. In the UK, genetic testing for KSD relies on a panel of 39 genes, reviewed by experts and offered at clinicians' discretion [253]. While gene panels provide a cost-effective balance between validity and affordability, they limit the scope of findings compared to whole exome or whole genome sequencing. A

recent review recommends genetic testing for patients aged ≤25 years, and those >25 years with metabolic abnormalities and at least two stone episodes, bilateral disease, or a strong family history of KSD [254]. Whether these recommendations will influence future guidelines remains to be seen. As sequencing costs decline and our understanding of the genetic basis of KSD advances, genetic testing may shift towards broader sequencing approaches rather than targeted panel testing [254].

### 1.1.8 Treatment

Despite the prevalence of KSD and the burden which it confers on individuals and healthcare systems, our understanding of the mechanisms underlying stone formation is limited and there is no high-quality evidence to support medical or dietary interventions to prevent KSD. For patients with no overt underlying cause for their KSD, current stone prevention strategies are few and relatively ineffective.

#### 1.1.8.1 Lifestyle management of kidney stone disease

Lifestyle advice is the cornerstone of management for patients with idiopathic KSD [6, 30]. Key recommendations include increasing daily water intake to 2.5–3.0 litres to reduce urinary saturation of stone-forming compounds [255]. Additionally, patients are advised to reduce salt intake to less than 6 g per day as this has been shown to decrease urinary calcium excretion. A study underlying this recommendation observed significantly lower urinary calcium and oxalate excretion in patients following a low-sodium diet compared to those in the control arm of the study [256].

Other recommended dietary measures include avoiding carbonated drinks, particularly those containing phosphoric acid as their main acidifier. This advice is corroborated by an RCT involving 1,009 men with KSD, where the group avoiding carbonated drinks had a 6.4% greater freedom from recurrence compared to the control group. Secondary outcomes demonstrated that avoiding phosphoric acid-containing drinks was advantageous, whereas avoidance of drinks acidified with citric acid showed no impact on recurrence rates [257].

For patients with calcium-based stones and hypocitraturia, adding fresh lemon juice to water as an alternative to potassium citrate therapy has been found to increase urinary citrate concentration, which may help to prevent stone formation and recurrence. A small study demonstrated the benefit of this approach [258], and current guidelines now extend this advice to all idiopathic stone formers [30]. Elevated urinary citrate concentrations inhibit calcium binding to other stone constituents, thus reducing stone formation risk.

Maintaining a normal daily calcium intake (700–1,200 mg), rather than adopting a low-calcium diet is another key recommendation. This advice is based on a study of 60 men with recurrent calcium oxalate stones and hypercalciuria, which demonstrated lower rates of stone recurrence in those consuming a normal calcium diet compared to those on a low-calcium diet [84]. Similarly, this guidance has been generalised to apply to all idiopathic stone formers [6, 30].

#### 1.1.8.2 Medical management of kidney stone disease

Evidence supporting the efficacy of medical therapies to prevent kidney stone formation is limited. In the UK, NICE recommends potassium citrate for patients with recurrent KSD whose stone composition is more than 50% calcium oxalate [30]. However, this recommendation is based on low-quality evidence derived from small, mostly non-randomised studies conducted in adults and children with calcium-based stones, with or without hypocitraturia [30, 259–264].

Clinicians are also advised to consider prescribing thiazide diuretics if stones are more than 50% calcium oxalate and persist after sodium restriction [30]. This recommendation also relies on evidence from small-scale, low-quality studies in adults with calcium-based stones [265–277].

More recently, the NOSTONE double-blind, RCT evaluated the efficacy of thiazide diuretics at several doses compared to placebo in 416 patients with recurrent calcium-containing kidney stones. There was no substantial difference in the incidence of recurrence, defined as symptomatic or radiological evidence of kidney stones, between patients receiving hydrochlorothiazide (at any dose) and those

receiving a placebo [278]. The equivocal efficacy of these therapies, combined with poor tolerability due to side effects such as orthostatic hypotension, gastrointestinal symptoms, and erectile dysfunction, may contribute to the low treatment adherence observed in patients with KSD [279].

The current literature also fails to provide sufficient evidence to recommend allopurinol or the combined use of allopurinol and thiazides for reducing stone recurrence in adults [30, 268, 269, 271, 280, 281]. Similarly, phosphate supplementation has limited evidence for effectiveness in preventing KSD [6, 282]. While magnesium-based therapies have shown some potential benefits, adverse effects preclude their recommendation for the prevention of idiopathic KSD by NICE [30, 265, 270].

#### 1.1.8.3 Procedural management of kidney stone disease

Given the limited effectiveness of conservative and medical management for KSD, patients frequently undergo procedural interventions [283]. Extracorporeal shockwave lithotripsy aims to fragment stones, allowing them to be excreted in the urine. Surgical approaches may involve percutaneous nephrolithotomy, a minimally-invasive, anterograde method providing direct, percutaneous access to the kidney through a small incision in the flank, or ureterorenoscopy, a retrograde, or sometimes antegrade, method that involves passing rigid and flexible endoscopes through the ureter. In larger or more complex cases of urolithiasis, a combined approach may be used. Overall, KSD and its treatment are associated with physical effects such as CKD [284] and reduced quality of life [285], which have further economic and biopsychosocial repercussions. Thus, there is a need to identify new treatment options for patients with KSD.

### 1.1.9 Summary of kidney stone disease

In summary, KSD is a common and frequently recurrent condition driven by both genetic and environmental factors. While monogenic disorders account for a small

proportion of cases, most arise from a complex polygenic basis, which I will discuss in the next section, that remains poorly understood. This knowledge gap limits our ability to predict, prevent, and treat KSD effectively. In this thesis, I aim to address this by identifying biological risk factors, uncovering genetic mechanisms underlying disease susceptibility, and translating these insights into clinical practice. To achieve this, I will apply a range of statistical, epidemiological, and genomic approaches, which I will now introduce.

## 1.2 Genetic epidemiology

### 1.2.1 Definition

Genetic factors play a crucial role in shaping complex traits where genetic variation arises from polygenic effects. Over the past two decades, the field of *genetic epidemiology* has rapidly evolved, largely in response to the limitations of traditional observational epidemiology, particularly when addressing issues like selection bias, confounding, and misclassification [286]. Large-scale association studies are necessary to achieve the statistical power to uncover the polygenic basis of these traits, where many genetic variants exert small individual effects [287]. A key feature of genetic epidemiology is the systematic application of epidemiological methods to evaluate the role of genetic variation in health and disease. Major projects, such as the International Human Genome Project [288] and the International HapMap Project [289, 290], have driven significant progress in this area.

A deeper understanding of the human genome, combined with advances in mathematical and computational techniques, has led to a "renaissance" of observational epidemiology to identify relationships between genetic variants and phenotypes, and to clarify relationships between risk factors and traits. These aims have informed a plethora of methods including genetic association studies, MR, and colocalisation analyses. Research leveraging these approaches can be applied to enhance diagnostic, prognostication, prevention, and therapeutic strategies [291]. Thus, genomics transforms conventional epidemiological research; large, longitudinal

studies that integrate demographics, medical and prescription records, biospecimens, and genetic data have emerged, enabling researchers to explore and characterise genes associated with common diseases. Prominent examples include the UK Biobank (N=≈500,000) [292], the Estonian Genome Project (N=≈1,000,000) [293], and the FinnGen study (N=≈400,000) [294] [286]. This expanded capacity for health research, driven by these comprehensive datasets and technological advances, has shaped modern-day personalised medicine.

### 1.2.2 Genetic architecture and genome-wide association studies

"Genetic architecture" refers to the landscape of genetic contributions to a given phenotype, encompassing the number, effect sizes, frequencies, and interactions of genetic variants influencing a trait [295]. Understanding this architecture is essential for uncovering the biological mechanisms that predispose individuals to disease and can inform novel therapeutic strategies for treatment and prevention. GWAS play a key role in this process by identifying common genomic loci associated with phenotypic traits, such as albumin-adjusted serum calcium concentrations or the presence of KSD [296]. These studies provide insight into the familial risk of polygenic diseases and illuminate their underlying biological pathways.

In GWAS, the outcome (dependent variable) is regressed on the genotype at each locus - often common (minor allele frequency >1%) single nucleotide polymorphisms (SNPs, the independent variable)-using linear regression for continuous traits or logistic regression for dichotomous traits. Statistical tests are conducted to assess the associations between genetic variants and the phenotype, under the assumption of an additive allele effect; that is, the estimate reflects the effect of each additional allele [297]. To account for multiple testing and minimise the likelihood of false-positive associations, p-values are adjusted using methods like Bonferroni correction, with a commonly accepted threshold for genome-wide significance set at $p < 5 \times 10^{-8}$ (i.e. 0.05 divided by roughly 1 million tests) [296].

#### 1.2.2.1 Applications of genome-wide association studies

The utility of GWAS extends beyond identifying genetic associations, and they can provide insights into genetic architecture through bioinformatic analyses that detect independent association signals, estimate heritability, and identify genes and pathways implicated in disease pathophysiology. These insights can inform clinical applications, including identifying opportunities for drug repurposing or highlighting novel therapeutic targets, polygenic risk scores (PRS) to quantify genetic susceptibility to disease [298, 299], MR for causal inference [300, 301], and colocalisation analyses to identify shared causal mechanisms between traits [296, 302].

**Genomics-driven drug discovery**

Genomics-driven drug discovery harnesses genotype-phenotype associations identified in GWAS, alongside genetic epidemiological techniques such as MR and colocalisation analyses. These approaches, discussed in detail below, help identify novel therapeutic targets and uncover opportunities for repurposing existing medications by providing genetic evidence for disease-relevant pathways.

Developing new drugs is a costly and time-intensive process. Advancing a drug from Phase I trials to regulatory approval takes over a decade, costs more than $1.3 billion, and has a failure rate exceeding 90% [303, 304]. Most failures result from adverse effects or lack of efficacy, often due to biological differences between humans and the model organisms used in preclinical research [305–307]. Drug responses and side effect profiles observed in non-human models frequently fail to translate to humans, while experimental medicine using cell lines cannot fully replicate systemic effects across human physiological systems.

Repurposing existing medications that have already undergone safety testing provides a cost-effective and efficient alternative to traditional drug development. Drug targets supported by strong genetic evidence are more likely to progress to Phase III trials and gain regulatory approval [308]. By integrating genetic insights with candidate drug targets, genomics-driven drug discovery can improve therapeutic development, anticipate potential side effects, and mitigate inefficiencies

associated with non-human models and in vitro studies [309].

### Polygenic risk scores

PRS offer a means of stratifying individuals based on genetic predisposition, enabling early intervention and personalised treatment strategies. As sequencing costs decline and large biobank and consortia datasets expand, studies of larger sample sizes emerge, increasing the predictive power of PRS, enabling the discovery of more significant associations than was previously possible, and enhancing the feasibility of integrating genetic risk profiling into routine healthcare [310]. By identifying high-risk subgroups, patients can receive more individualised healthcare from targeted interventions and avoid unnecessary side effects from treatments that are less likely to be effective for them.

### Genotype-phenotype relationships

GWAS has become a powerful tool for uncovering genotype-phenotype relationships that refine our understanding or disease mechanisms and enable us to generate hypotheses that drive biomedical research. For example, GWAS has identified key obesity-related genes such as *LEP*, *PCSK1*, and *POMC* [311], which have informed the development of genotype-tailored therapies. These findings exemplify the role of GWAS in linking genetic discoveries to clinical translation. The associations of *LEP*, *GLP-1*, and *GIPR* and obesity [312, 313] have influenced treatment strategies, for example, recombinant human leptin for leptin-deficient patients with *LEP* mutations [314], and glucagon-like peptide-1 analogues, such as liraglutide and semaglutide, have been successfully implemented in obesity management [315–318].

Thus, leveraging GWAS to identify genetic drivers of disease can uncover novel therapeutic targets, refine risk prediction models, and develop personalised interventions. As genetic datasets grow and analytical techniques advance, the potential to translate GWAS findings into clinical practice continues to expand, offering new opportunities to transform how we treat patients with complex diseases.

#### 1.2.2.2 Limitations of genome-wide association studies

While GWAS is a valuable genetic epidemiological tool in a researcher's armamentarium, they are not effective at fully elucidating the genetic basis of a phenotype and interpreting results poses significant challenges [319]. Identifying specific genes and mechanisms underlying associations with rare and common, non-coding variants is challenging [320]. Many identified variants have modest effects on traits, are often associated with multiple traits, and frequently reside in non-coding regions, possibly because regulation regions are highly relevance in polygenic inheritance or because most protein-coding variants are not accessible through imputation [321]. Additionally, these variants are commonly inherited together (a phenomenon referred to as *"linkage disequilibrium"*, LD), which makes it difficult to distinguish causal variants from nearby non-causal ones [322]. Furthermore, differences in allele frequencies across populations complicate the generalisation of GWAS findings, making it challenging to interpret and apply results across diverse genetic ancestries.

Despite its wide-ranging applications, GWAS is not without limitations. Some of the most notable challenges include:

**Population stratification**

Population stratification (or structure) is an important source of systemic bias in GWAS. It refers to differences in allele frequencies between subpopulations that arise due to underlying genetic ancestry, rather than a true association between genetic variants and a trait [297]. In GWAS, uncorrected population stratification can lead to false-positive (type 1) or false-negative (type 2) associations. Genomic control $\lambda$ ($\lambda_{GC}$) is a commonly used method to detect and correct for population stratification by assessing genome-wide inflation in GWAS results [323]. Specifically, $\lambda_{GC}$ is calculated as the ratio of the observed median test statistic to the expected median of a chi-square distribution with 1 degree of freedom under the null hypothesis, which is $\approx$0.455 [324]. A $\lambda_{GC} \approx 1$ indicates minimal population stratification, while

$\lambda_{GC} > 1$ suggests the presence of stratification or another confounding factor [325]. Values $< 1.05$ are typically considered acceptable in scientific literature [323].

To correct for population stratification, various methods exist that infer genetic ancestry including principal component analysis (PCA) and structured association. PCA is a linear dimensionality reduction technique that identifies uncorrelated principle components, ranked based on the variance they explain in the data. These components can then be included as covariates in a GWAS to adjust for stratification. Importantly, PCA applies greater correction to variants that show large differences in allele frequency across ancestries [323]. Conversely, structured association assigns individuals to subpopulation clusters and associations between genetic variants and a trait are stratified by cluster [323].

However, both PCA and structured association methods fail to address the complexities introduced by family structure (i.e., known familial relationships within the sample) and cryptic relatedness (i.e., distant, unknown familial relationships). For studies that use family-based recruitment, family-based association tests that focus on within-family information can mitigate the impact of family structure [323]. Cryptic relatedness, however, can occur even in seemingly unrelated samples and poses additional challenges.

To address these issues, mixed model approaches offer robust methods to correct for population stratification, family structure, and cryptic relatedness [320, 326]. Mixed models allow fixed and random effects to be included in analyses. Fixed effects typically include the SNPs of interest and relevant covariates, such as age and sex, while random effects account for covariance from heritable and non-heritable random variation [323]. For studies in populations with minimal population stratification - such as those with homogeneous ancestry - mixed models can be applied alone. In cases where stratification is a concern, mixed models can be used alongside principal component covariates for more comprehensive adjustment [323].

### Linkage disequilibrium

In a large population, in the absence of selection, pairs of genetic loci that are not close together on a chromosome are not associated at a population level [327].

This *linkage equilibrium* arises from genetic recombination, which promotes the independent assortment of alleles across generations [322, 328, 329]. However, because recombination events are relatively rare in a population, tightly linked loci are often correlated and exist in *linkage disequilibrium*, leading to their alleles co-occurring more frequently than would be expected under the principle of random assortment [330]. Since correlated SNPs are inherently not independent, LD reflects the pattern of correlation between these SNPs [297].

LD is crucial in the context of GWAS. SNP arrays only genotype between 500,000 and 1,000,000 SNPs and thus capture a fraction ($\approx 1\%$) of genotypic variation. Given this limitation, direct testing of functional polymorphisms that directly cause a trait is unlikely. Instead, signals from ungenotyped, causal variants can only be identified from genotyped markers that are in strong LD with them [331]. To enhance analyses, researchers can impute additional SNPs using LD structures specific to ancestral populations as demonstrated by data from the HapMap Consortium and 1000 Genomes Project [288–290]. While LD increases the likelihood of identifying genetic regions associated with phenotypes, it complicates efforts in fine-mapping and identifying specific susceptibility genes as multiple highly-correlated SNPs can be located in intergenic regions and/or have similar strengths of association with a phenotype [330, 331]. For instance, the SNP with the lowest p-value at a locus identified by GWAS is often not the causal SNP but rather one that "tags" the causal SNP [331]. LD limits the generalisability of GWAS results because LD patterns vary across ancestral populations [330]. A SNP deemed significant in one GWAS may be correlated with the causal SNP in the population where the GWAS was conducted, but this correlation may not hold true in a different population.

### 1.2.3 Exome-wide association studies

As discussed, GWAS can identify hundreds of genetic loci associated with common variants. However, to assess the role of rare, coding variants that impact protein-coding genes and contribute to disease, large-scale patient cohorts with whole exome sequencing (WES) are essential for conducting exome-wide association

studies (ExWAS). WES enables a substantially larger repertoire of rare and ultra-rare protein-coding variants to be identified [332–334]. Beyond single-variant association testing, ExWAS facilitates "collapsing analyses" which aggregate multiple low-frequency variants within a gene to detect gene-based associations [296]. This approach can identify rare variants with potentially high effect sizes and uncover disease mechanisms that GWAS, primarily detecting common variants of small effect sizes, may miss. For example, collapsing analyses of loss-of-function variants can estimate the direction of effect for associated genes [334, 335]. By leveraging WES data, researchers can identify protein-altering variants with clearer functional consequences than non-coding variants. These discoveries help address the "missing heritability" in complex traits [336–338], provide valuable mechanistic insights, and inform therapeutic target discovery and validation [339–342]. Furthermore, such findings advance precision medicine efforts [343, 344].

The availability of WES data from resources such as the UK Biobank [345], along with the capability to impute data from genotyped samples that have not been sequenced, provides new opportunities to address the limitations of genomic studies in KSD. WES studies have successfully identified rare and low-frequency coding variants associated with numerous traits and diseases, including type 2 diabetes [346], haematological traits [347–349], blood lipid levels [350], blood pressure [351], immune-mediated diseases [352, 353], height [354], and BMI [355]. In nephrological diseases, exome sequencing has revealed novel risk genes significantly associated with kidney-related traits. Examples include ADPKD [356], glomerular disease [357], and CKD [358, 359]. These studies highlight the overlap between risk genes and kidney-related biomarkers, underscoring the importance of investigating rare coding variants to advance understanding of kidney disease pathophysiology.

### 1.2.4 Mendelian randomisation

[1] Assessing causality between an exposure and an outcome is a mainstay of epidemiological research. Conventional observational epidemiological studies - such as cohort, case-control, and cross-sectional designs - provide valuable insights into associations between exposures and outcomes. However, they are often vulnerable to biases introduced by unmeasured or imprecisely measured confounders, measurement errors, and reverse causation (Figure 1.7) [360–362]. Furthermore, efforts to quantify complex variables like socio-economic status are frequently insufficient, leading to residual confounding. These limitations cannot be fully overcome through statistical adjustment and increase the risk of drawing incorrect conclusions about exposure-outcome relationships.

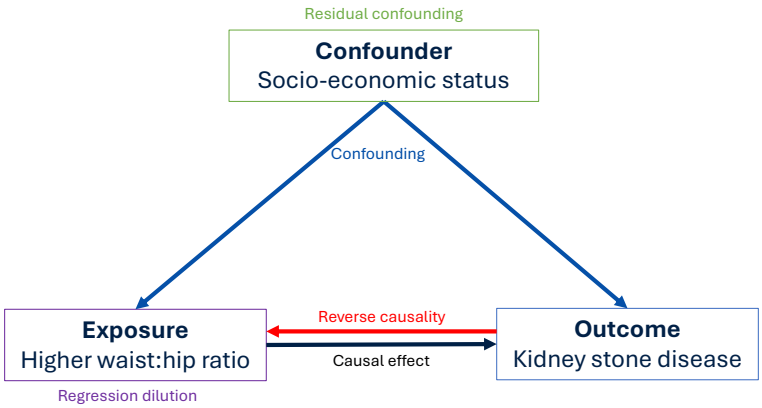



***Figure 1.7. Challenges in observational studies.***

**Confounding:** A variable affects the exposure and outcome. Measured confounders can be accounted for in multivariable analyses; unmeasured confounders can bias results. **Residual confounding:** Confounding that persists despite adjusting for measured confounders. **Reverse causality:** An effect of the outcome on the exposure. **Regression dilution bias:** Random measurement errors that bias the estimate between an exposure and outcome. Adapted from Lovegrove et al. [300]

Well-designed RCTs are considered the "gold standard" for addressing biases and establishing causal relationships between an exposure and an outcome (Figure 1.8) [363]. However, not all research questions are amenable to RCTs. For instance, RCTs may be impractical when the outcome of interest is rare or takes a long time to develop. Additionally, in situations where the exposure is known to be harmful -

[1] Content from this section includes work that I wrote from ***Lovegrove CE**, Howles SA, Furniss D, et al. Causal inference in health and disease: a review of the principles and applications of Mendelian randomization. Journal of Bone and Mineral Research. 2024 Aug 21; 10.1093/jbmr/zjae136* which I have permission from co-authors to include.

such as obesity - randomising individuals to interventions aimed at increasing their WHR would be unethical [364, 365].

MR is an innovative epidemiological technique that aims to investigate causal relationships between exposures and outcomes, addressing some of the practical challenges posed by RCTs [366]. While MR is often likened to an RCT as an approach to estimate causal effects [367], this analogy has limitations and fails to fully capture the complexities involved in causal inference. Nonetheless, it serves as a useful framework for conceptualising MR (Figure 1.8).

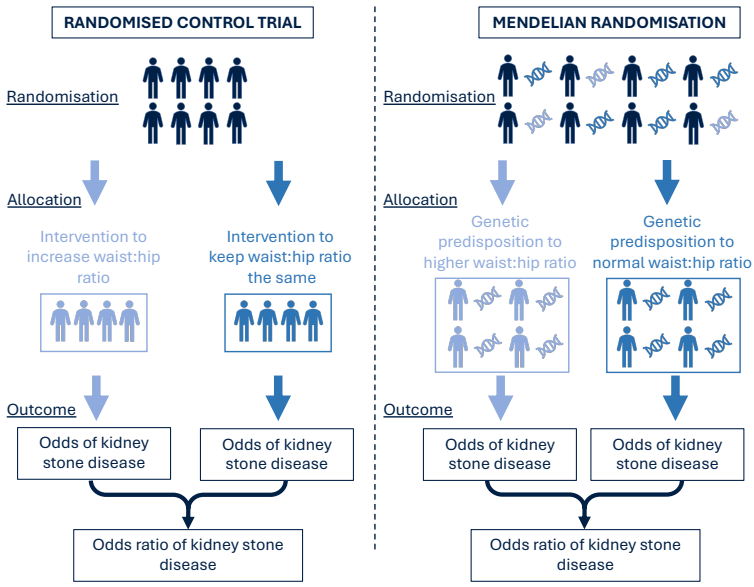



***Figure 1.8. Analogy of a randomised control trial and Mendelian randomisation.***

Mendelian randomisation leverages genetic variants as "instrumental variables" to simulate the random assignment of an exposure, analogous to randomisation in a randomised control trial. This reduces the influence of confounding factors and strengthens causal inference; hence, Mendelian randomisation can be a powerful tool for investigating relationships between exposures and outcomes. Adapted from Lovegrove et al. [300]

#### 1.2.4.1 Principles and assumptions of Mendelian randomisation

Gregor Mendel's laws of genetic inheritance - specifically the laws of segregation and independent assortment - form the foundation of MR [368]. The law of "independent segregation" states that at each autosomal site one allele is inherited from the mother and one from the father, with the specific allele inherited from each parent being determined randomly [328, 329]. Meanwhile, the law of "independent assortment" asserts that the inheritance of alleles at a locus is independent of alleles at other loci, except for genetically-linked regions that are in LD and inherited together [322,

328, 329]. MR uses these genetic alleles that are inherited and fixed at conception arising from the random assortment of chromosomes during meiosis. These alleles are independent of environmental confounding variables, enabling them to function as unbiased proxies, or instrumental variables (IVs), for exposure traits (Figure 1.8) [369].

To reliably draw causal inference from MR, three fundamental conditions must be met (Figure 1.9):

1. *Relevance*: The IV must be reliably associated with the exposure.
2. *Exchangeability*: The exposure IV is not caused by a factor that also causes the outcome, that is, there are no unmeasured confounding variables between the exposure IV and the outcome.
3. *Exclusion restriction*: The exposure IV only affects the outcome via the exposure trait and not by any other trait with a downstream effect on the outcome of interest.

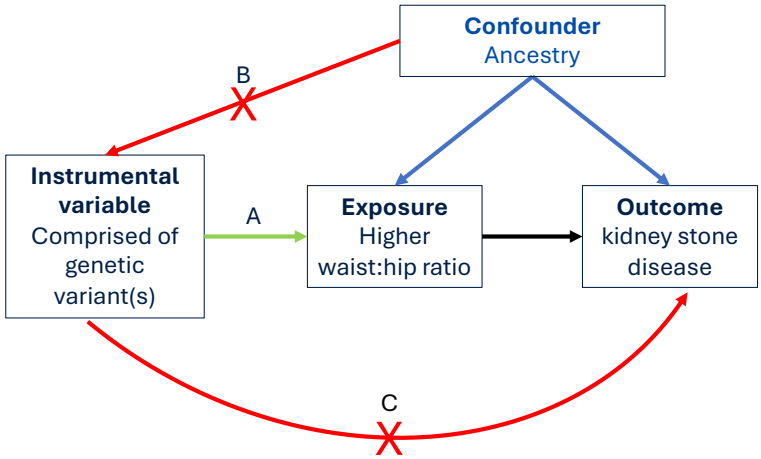



***Figure 1.9. Assumptions in Mendelian randomisation.***
**(A) Relevance:** Genetic variant(s) serving as instrumental variables (IVs) must be strongly associated with the exposure variable. **(B) Independence:** IVs must be independent of confounding factors to ensure that genetic associations with confounders are random, similar to randomisation process in a randomised control trial. **(C) Exclusion restriction:** IVs must influence the outcome exclusively through the exposure variable and not by other, horizontally pleiotropic, pathways. Adapted from Lovegrove et al. [300]

Although it is impossible to definitively prove that the assumptions of *exchangeability* and *exclusion* restriction hold, researchers can demonstrate that these assumptions are not violated by applying multivariable approaches and sensitivity analyses, such

as MR-Egger [300]. Beyond the core tenets of MR, several additional assumptions play a crucial role. While a comprehensive discussion of all the principles underlying MR falls outside the scope of this thesis, this section will highlight those particularly relevant to investigating causal risk factors for KSD.

The principle of *gene-environment equivalence*, commonly referred to as the consistency assumption in causal inference [370, 371], posits that an exposure influences an outcome in the same way, regardless of whether the exposure arises from genetic variation or environmental factors. For example, in MR analysis, researchers assume that a 5 mmHg increase in systolic blood pressure confers the same risk of stroke whether the rise results from a genetic predisposition or from lifestyle factors.

Temporality is another key consideration in MR studies. Because individuals inherit genetic variants at conception, MR analyses reflect lifetime exposure rather than short-term changes. For example, MR studies investigating the relationship between BMI and coronary artery disease capture the effects of long-term obesity, whereas RCTs on weight-loss therapies typically assess only short-term outcomes. Additionally, MR analyses generally assume that genetic variants maintain consistent associations with exposures over time [372]. However, the impact of an exposure on an outcome may depend on when in life the exposure occurs. For instance, evidence suggests that early-life adiposity can influence breast cancer risk independently of adiposity later in life [373].

To ensure reliable causal inference in MR studies, researchers must derive IVs from the same population in which they assess the outcome [374, 375]. Differences in LD patterns across ancestral groups can lead to biased results if an MR analysis includes populations with different genetic backgrounds. A genetic variant may not have the same association with an exposure trait in both the exposure and outcome datasets, violating the exchangeability assumption and introducing confounding due to genetic ancestry [376]. For example, causal inference would be flawed if researchers used an exposure IV from a population that consumes alcohol but an outcome IV from a population with minimal alcohol consumption, such as women in certain East Asian cultures [377]. Unfortunately, these challenges are compounded by the

Eurocentric bias that plagues GWAS and the scarcity of well-powered studies in non-European populations. These disparities not only reinforce inequalities in health research but also limit the generalisability of findings across diverse populations [299].

#### 1.2.4.2 Mendelian randomisation study designs

MR studies may be conducted using individual-level data or summary statistics from GWAS and may involve one or two populations. In both scenarios, increasing the sample size and the variance of the exposure explained by the IVs enhances statistical power, thereby improving the precision of the findings [375]. Since common genetic variants usually explain only a small fraction of the variation in a trait, large GWAS datasets are often employed to conduct well-powered MR analyses.

Genetic variants can function as discrete instruments or be combined into an allele score, which is calculated as the sum of risk alleles [378]. Combining carefully selected genetic variants into an allele score can enhance the power of analyses, particularly when individual variants exert small effects on the exposure. Using them as discrete instruments risks violating the *relevance* assumption due to weak instrument bias [378]. Prior to performing MR analyses, variants within the exposure IV that are in LD should be "clumped" to retain only one variant per locus based on the population's linkage pattern, or an LD matrix should be included in conditional analyses when generating MR estimates [379]. These steps ensure that the IV comprises independent genetic variants, thereby avoiding "double counting" any single variant-outcome estimate. Additionally, to minimise bias, allele scores must be weighted to reflect the estimated effect of each genetic variant on the exposure, accounting for varying variant-effect associations among the SNPs that contribute to the score.

In one-sample MR, researchers derive both exposure and outcome IVs from the same population. This approach carries a risk of overfitting, which introduces bias away from the null and may lead to incorrect causal inferences [369, 378, 380]. Overfitting arises because the exposure IV predicts the outcome within the same dataset, increasing the likelihood of chance correlations with confounding variables

[381]. Both one-sample and two-sample MR studies face the additional challenge of weak instrument bias, which can further distort causal estimates when overfitting is present. However, researchers can mitigate these risks by using stronger instruments, derived from large consortia, and well-powered genetic databases.

In two-sample MR, researchers derive IVs for the exposure and outcome from separate populations. This approach enhances statistical power, particularly when datasets are small or biospecimens are limited, by allowing genetic associations with the exposure and outcome to be assessed independently using larger datasets [369]. However, two-sample MR assumes that both populations share similar genetic ancestry, age, and sex distributions to avoid bias, as discussed in section 1.2.4.1 [376, 382]. Additionally, researchers must ensure robust case ascertainment in the population used to derive the outcome IV to maintain sufficient power for reliable causal estimates.

As meta-analyses of GWAS become more common in MR analyses, sample overlap increasingly affects two-sample MR [382]. In continuous outcomes, the degree of overlap determines the extent of bias; greater overlap makes the study design resemble one-sample MR. For binary outcomes, bias depends on whether case participants contribute to the exposure IV [381]. If only control participants overlap, estimates remain unbiased. However, when both cases and controls overlap, bias mirrors that seen in continuous outcomes [381]. Recent evidence suggests that sample overlap may not be as problematic as once thought. Researchers can assess power, bias, and type 1 error rates using statistical tools such as https://sb452.shinyapps.io/overlap [381].

#### Mendelian randomisation estimates

There are various methods of performing MR, each designed to overcome or quantify the effects of different biases. Estimates resulting from MR analyses for continuous outcomes are interpreted as the unit change in outcome per unit increase in exposure. For binary outcomes, the estimate is interpreted as the log odds ratio for risk of outcome per unit increase in exposure.

### Wald ratio

The simplest method of estimating the causal effect of a single IV on an outcome is known as the "Wald ratio". The Wald ratio is calculated by regressing the outcome (Y) on the IV (Z), regressing the exposure (X) on the IV (Z), and then deriving the ratio of these two regression estimates [383].

$$\text{Wald Ratio} = \frac{\hat{\beta}_{Y|Z}}{\hat{\beta}_{X|Z}}$$

Alternatively, this is represented as

$$\text{Wald Ratio} = \frac{\hat{\beta}_{\text{outcome}}}{\hat{\beta}_{\text{exposure}}}$$

where $\hat{\beta}$ is the $\beta$-coefficient for the regression of the exposure or outcome on the IV. While the Wald ratio suggests that the variant representing the exposure IV has a direct causal effect on the outcome, it is important to consider that the variant may be in LD with a disease-causing variant. This linkage can lead to erroneous conclusions if the assumption of direct causality is made without further investigation. This concern is also relevant when using IVs that consist of multiple genetic variants.

### Inverse variance weighted

An inverse variance weighted estimate (IVW) is a method used to perform a meta-analysis of the Wald ratios for each genetic variant within an IV. This approach generates an overall estimate of the causal effect of an exposure IV that includes multiple genetic variants. The IVW can be considered as a regression analysis, where the effects of genetic variants on the outcome are regressed against their effects on the exposure. The regression is weighted by the inverse of the variance of the variant-outcome effects, ensuring that more precise estimates contribute more to the overall result. The intercept of this regression is constrained

to zero; hence, a variant with no effect on the exposure is assumed to have no effect on the outcome. A fixed-effect IVW is calculated on the assumption that no variants in the IV violate the three core assumptions of MR (Figure 1.9). If this is satisfied, the IVW estimate has the highest statistical power [384]. However, if horizontal pleiotropy is present - where genetic variants influence the outcome through pathways other than the exposure - the IVW estimate may be biased [384]. A random-effects model can account for balanced pleiotropy, where the deviation of a variant's effect on the outcome is independent of its effect on the exposure. This approach accounts for the variability across genetic variants, allowing the IVW method to produce an unbiased estimate of the causal effect, even in the presence of pleiotropic variants that do not systematically affect the outcome [385].

### MR-Egger

The assumption of *exclusion restriction* is violated in the presence of directional horizontal pleiotropy, where variants affect the outcome via pathways not mediated by the exposure. In such circumstances, the IVW estimate is biased. This issue can be addressed by using the *MR-Egger* method, which allows for a non-zero intercept when regressing variant-outcome effects on variant-exposure effects, thus accounting for directional (or "unbalanced") horizontal pleiotropy [386, 387]. The intercept in MR-Egger reflects the average pleiotropic effect across all variants, while the slope provides an estimate of the causal effect after adjusting for pleiotropy.

Even when all variants in an IV show horizontal pleiotropy, the MR-Egger estimate can remain unbiased as long as the pleiotropic effects are independent of the variant-exposure effects. This is known as the *InSIDE* assumption (Instrument Strength Independent of Direct Effect). For example, if a study investigates the relationship between urine calcium concentration and the risk of KSD, and variants in the exposure IV are associated with serum calcium concentration, it is plausible that these variants affect risk of KSD through both urine and serum calcium levels. If the genetic effects on serum calcium concentration are correlated with their effects on urine calcium, the *InSIDE* assumption would be violated, leading to a biased MR-Egger estimate [388]. Moreover, as the variability of the variant-exposure

associations increases, the precision of the MR-Egger estimate decreases, making it less reliable when the IV contains only a few genetic variants [386].

#### Weighted median

An alternative approach to managing unbalanced pleiotropy is to use the median effect of all variants within an IV [389]. While a simple median estimate yields a reliable causal effect if at least 50% of the variants are valid, the weighted median method offers a consistent estimate when at least half of the overall weight is derived from valid variants that adhere to the MR assumptions (Figure 1.9). Variants are assigned weights based on the inverse variance of their relationship with the outcome, allowing variants with stronger associations to have a greater impact on the weighted median estimate.

#### Simple and weighted mode

Mode-based estimation involves grouping variants into clusters that share similar causal effects using a kernel density function [390]. The causal estimate is taken from the largest cluster. If this cluster consists of valid variants, the mode-based estimate is unbiased. In a weighted mode approach, each variant's contribution to its cluster is weighted based on the inverse variance of its association with the outcome.

### 1.2.4.3 Multivariable Mendelian randomisation

Multivariable analysis can be useful to determine the effect of an exposure on an outcome while adjusting for relevant covariates. Multivariable MR is especially valuable when the *"exclusion restriction"* assumption is violated, as it allows genetic variants to influence the outcome through other risk factors [391]. Moreover, multivariable MR can include multiple, correlated exposures in analyses, helping researchers to determine whether all included exposures have direct effects on the outcome, and to uncover mediating relationships. Multivariable MR relies on the

same assumptions as univariable MR (section 1.2.4.1, Figure 1.9) [392]. It also requires that the genetic variants influence the outcome only through the risk factors included in the analysis [391].

#### 1.2.4.4 Mediation Mendelian randomisation

Mediation analyses explore how an exposure affects an outcome by estimating the proportion of the effect that is mediated by an additional trait (Figure 1.10) [393].

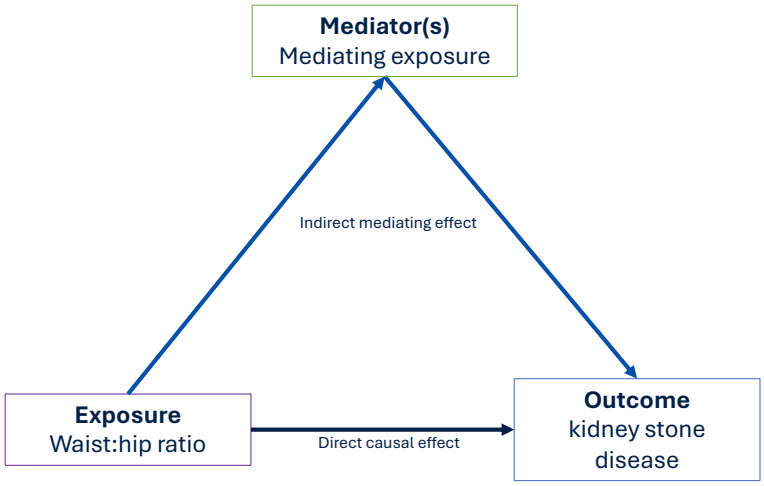



***Figure 1.10. Mediation Mendelian randomisation.***
Mediation analysis estimates the effect of an exposure on an outcome via mediating factors (indirect mediating effect) and via independent mechanisms (direct causal effect). The proportion mediated effect can be derived. Adapted from Lovegrove et al. [300]

Mediation MR analyses estimate the *total* effect of an exposure on an outcome through all pathways, the *direct* effect of the exposure independent of any mediators, and the *indirect* effect through specified mediators [393]. Analyses can be performed using either 2-step or multivariable MR methods. In the 2-step MR approach, the indirect effect is calculated by multiplying the exposure's effect on the mediator with the mediator's effect on the outcome (product of coefficients method) [394]. In multivariable MR, the direct effect of the exposure on the outcome, adjusting for the mediator, is estimated and the direct effect is subtracted from the total effect to find the indirect effect (difference method) [394]. The *proportion mediated effect* is derived to estimate the proportion of the exposure's total effect on the outcome that is explained by the mediating pathway (Figure 1.10).

#### 1.2.4.5 Drug-target Mendelian randomisation

MR can be applied to predict drug effects and toxicities, providing evidence for the potential utility of a drug before committing to costly clinical trials [395]. When a drug target is a protein, MR analyses can estimate the effects of modulating that protein by selecting genetic variants that serve as an IV for the target protein. The impact of altering the protein (i.e. the exposure) on an outcome can subsequently be estimated. Limiting the IV to variants near transcription start and stop sites (*cis*-protein quantitative loci) enhances confidence that the estimated effects are due to the gene that encodes the protein [396]. As previously mentioned, the MR estimate is unlikely to match those of clinical interventions; MR studies reflect life-long exposure whereas clinical interventions are assessed over a defined period.

### 1.2.5 Colocalisation

GWAS can effectively identify genetic loci linked to specific traits. However, challenges such as LD and incomplete genomic coverage complicate efforts to uncover the causal mechanisms underlying these associations [302]. It is common for GWAS of different traits or diseases to reveal overlapping genetic signals, but these shared signals do not necessarily imply that the traits are driven by a common causal genetic variant [397]. A causal variant, in this context, refers to a genetic polymorphism where variation in the genotype leads to changes in molecular mechanisms, ultimately influencing the trait of interest (Figure 1.11) [397].




***Figure 1.11. Genetic colocalisation.***

**A:** Genetic colocalisation of two traits with a shared causal genetic variant. **B:** Absence of colocalisation. Association signal is likely due to tagging variants being in linkage disequilibrium.

For example, analyses have revealed that genetic variants in the *PCSK9* region, which encodes proprotein convertase subtilisin/kexin type 9 (PCSK9), regulate cholesterol metabolism. MR analyses demonstrate that elevated low-density lipoprotein (LDL) cholesterol, driven by PCSK9, increases the risk of coronary heart disease (CHD) [398]. Colocalisation analysis in this region shows a high posterior probability (>99.9%) that the same genetic signal influences both LDL cholesterol levels and CHD risk, suggesting that LDL cholesterol is directly causal for CHD at this locus (Figure 1.11A) [399]. When colocalisation analyses also include plasma levels of PCSK9, weaker evidence of colocalisation emerges (posterior probability 82.4%). This supports the efficacy of PCSK9 inhibitors in significantly lowering LDL cholesterol levels and reducing cardiovascular events in both familial and nonfamilial hypercholesterolaemia patients, as well as healthy individuals while also suggesting that there may be multiple variants affecting PCSK9 expression [400].

In contrast to causal variants, tagging variants are in LD with causal variants but do not directly influence the trait. Such variants may appear associated with multiple traits in GWAS, but they could be directly causal for only one of them while being in LD with the true causal variant for other traits [401]. For instance, signals in the *GLP1R* region have been identified for both type 2 diabetes (T2D) [402] and BMI (BMI) [313] in GWAS. However, colocalisation analysis reveals separate causal variants for each trait, indicating that different mechanisms drive glycaemic control and body weight regulation in this region (Figure 1.11B) [315]. This lack of genetic colocalisation complicates MR analyses, as it violates the assumption of exchangeability; a genetic instrument might affect the outcome through an alternative pathway, bypassing the exposure of interest [397]. Understanding this genetic architecture is essential for developing targeted therapies for complex diseases.

#### 1.2.5.1 Approaches to colocalisation

Colocalisation analysis employs a formal statistical model to evaluate whether traits share a causal variant and, where appropriate, prioritise a shared putative causal variant. This approach is akin to fine-mapping across multiple traits [403]. Two primary approaches exist in colocalisation analysis:

### Proportional colocalisation

Proportional colocalisation tests the null hypothesis that genetic associations with two traits are proportional [397]. The method compares regression coefficients for each trait against multiple SNPs. If the coefficients are proportional, colocalisation is assumed, implying shared genetic associations [404, 405]. If they are not proportional, the null hypothesis is rejected, and colocalisation is unlikely.

To conduct this analysis, researchers select a set of SNPs, but this process can introduce biases. For example, "Winner's curse" may lead to the overestimation of effect sizes at these SNPs, and the most strongly associated SNP may not be the causal variant [406]. These biases can be mitigated by averaging across different SNP models to account for uncertainty, though there is still a risk of type 1 errors and incorrectly rejecting the null hypothesis, even when a shared causal SNP exists [407].

### Enumeration colocalisation

Enumeration colocalisation uses a Bayesian framework to evaluate evidence for different hypotheses about genetic associations. Researchers calculate posterior probabilities for five global hypotheses which are based on prior probabilities set to reflect the likelihood that any SNP in a region is associated with one or both traits [408]:

- H0: The genomic region is not associated with either trait
- H1: The genomic region is associated with trait 1 but not trait 2
- H2: The genomic region is associated with trait 2 but not trait 1
- H3: The genomic region is associated with both traits but has two distinct causal variants
- H4: The genomic region is associated with both traits and shares a causal variant

Unlike proportional colocalisation, the null hypothesis in enumeration colocalisation assumes that there is no colocalisation. Therefore, in the absence of strong evidence, posterior probabilities will be similar to the prior probabilities, reducing the risk of false-positive results [397].

#### 1.2.5.2 Multiple-trait colocalisation

By integrating data from intermediate traits, such as protein expression or metabolite levels, multiple-trait colocalisation studies can provide stronger evidence for a shared causal gene or pathway underlying several traits [407]. Multiple-trait colocalisation evaluates genetic signals shared by several traits in a single analysis and can identify clusters of traits with shared genetic predictors [397]. Some methods assume a single causal variant, while others allow for more complex patterns of colocalisation across traits. One example is HyPrColoc, which employs a divisive clustering algorithm to identify a small number of causal configurations across many traits [302]. This method avoids the limitations of iterative pairwise colocalisation analyses which can increase the risk of type 2 errors and obscure shared causal variants.

#### 1.2.5.3 Assumptions in colocalisation

Early colocalisation methods often assumed that only one causal variant existed per trait in a given genomic region [407]. Researchers would use GWAS effect estimates to calculate the likelihood of each causal variant pairing and assign them to one of the five hypotheses described above in section 1.2.5.1 [408]. This approach has the advantage of not requiring an LD reference matrix, making it possible to combine data from different populations [409]. However, the assumption of a single causal variant is overly simplistic. When this assumption is violated, colocalisation analysis takes a conservative approach, assessing whether the strongest signals for each trait colocalise [407]. Recent methods have moved away from this constraint by integrating LD reference matrices and removing prior assumptions about the likelihood of shared versus distinct causal variants [410]. Another strategy involves

conditioning on the most significant variant for each trait and testing colocalisation between conditioned signals [409].

More advanced approaches, such as the Sum of Single Effects (SuSiE) regression framework, allow researchers to account for multiple causal variants by summing individual regressions for each causal variant. SuSiE can quantify the strength of evidence for each variant being responsible for a signal and improve the detection of shared loci, reducing ambiguity in overlapping signals [410, 411].

## 1.3 Single-cell multi-omics

Human diseases often arise from the complex interplay of multiple organs and the diverse cell types within each tissue. While GWAS, MR, and genetic colocalisation analyses can identify genetic variants with likely causal effects on traits or disease risk, they fall short of pinpointing the specific cell types or cell states where these causal relationships occur. This limitation is particularly challenging for many disease-associated variants identified through GWAS because, as previously mentioned, they are often located in non-coding regions, hence, interpreting their functional relevance can be challenging. A detailed understanding of the cellular landscape of disease is therefore essential. It not only aids the interpretation of findings from genetic epidemiological studies but also provides a more comprehensive overview of how variants influence biological pathways, such as by altering chromatin states, gene regulation, or protein expression.

Recent advances have facilitated a shift from "bulk" tissue analyses to single-cell approaches, offering greater granularity and depth when investigating the causal mechanisms of disease pathophysiology. Single-cell technologies allow researchers to identify the cell subpopulations of biological interest in disease and to establish direct links between causal genetic or molecular changes and the cell types or states that they affect. Single-cell or single-nucleus ribonucleic acid sequencing (snRNA-seq) technologies can delineate cell sub-populations within organs and tissues based on transcription patterns. Several single cell RNA-seq methods exist and cell capture can be performed using fluorescence-activated cell sorting, microfluidics, or

microdroplets [412].

A more precise understanding of the cell types and molecular mechanisms driving disease provides a robust foundation for prioritising therapeutic targets, ultimately advancing efforts to develop targeted interventions. Loeb *et al* have recently identified that most heritable differences in adult renal function are because of altered gene expression in renal tubule epithelial cells and podocytes [413]. Hence, single-cell analysis provides one strategy to gain mechanistic insights from GWAS.

Multiomic analyses bridge the gap between statistical associations identified by GWAS and the functional mechanisms underlying them. These approaches can validate findings from MR or colocalisation analyses by uncovering their cellular and molecular basis. By profiling genomic, transcriptomic, and proteomic features in individual cells, researchers gain deeper insights into the biological mechanisms and genotype-phenotype relationships underlying disease [412]. Single nucleus assay for transposase-accessible chromatin using sequencing (snATAC-seq) is an extension of bulk ATAC-seq using Tn5 transposase to measure chromatin accessibility in cells [414]. Chromatin accessibility is the degree to which nuclear deoxyribonucleic acid (DNA) macromolecules can come into physical contact and is determined by the organisation of the nucleosome and chromatin-binding factors that can impede access to DNA [415]. It reflects the physical interactions by which regulatory elements, for example promoters and enhancers, control gene expression and chromatin accessibility can change in response to intrinsic and extrinsic factors such as developmental cues or injury, respectively [415]. In the kidney, chromatin accessibility influences processes of repair and regeneration [416].

Multimodal analyses integrating snATAC-seq with snRNA-seq can provide insights into how chromatin accessibility regulates gene transcription and uncover regulatory mechanisms such as enhancer-promoter interactions that mediate the effects of genetic variants [417].

## 1.4 Chromosome conformation capture

Single-cell multiomic approaches can help to identify the cell types involved in causal pathways that link genetic variants to specific traits. However, understanding chromatin architecture is crucial when investigating the functional significance of DNA regions [418, 419]. Long-range interactions of regulatory elements, such as enhancers, with chromatin are influenced by transcription factor binding and regulate the timing and magnitude of gene transcription in specific cell types. These regulatory elements may reside several megabases away from their target genes, sometimes even within the intronic region of an unrelated gene [420]. For these regulatory sequences to affect gene expression, *cis*-acting elements must come into close physical proximity with their targets through chromatin looping and interactions. Analysing chromatin accessibility and 3-dimensional architecture can provide insight into distant regulatory regions that influence transcription through long-range interactions.

Historically, methods such as fluorescent in situ hybridisation and electron microscopy were used to explore three-dimensional chromatin architecture. While these techniques provide valuable insights, they are labour-intensive and impractical for analysing multiple cell types and loci. The advent of chromosome conformation capture (3C) methods revolutionised the study of chromatin architecture, offering transformative insights into the genetic mechanisms underlying disease pathophysiology [421]. These methods measure the spatial proximity between DNA elements, enabling researchers to identify regulatory interactions and to compare them across various cell types.

The original 3C technique focused on analysing the frequency of chromatin contacts between two specific genomic loci, a "one-versus-one" approach [422]. Over time, 3C methods have evolved, integrating technologies like microarrays and next-generation sequencing, to provide more detailed interaction profiles using "one-versus-all", "many-versus-many", and "all-versus-all" approaches [420, 423]. These advances have dramatically increased the scale and resolution of chromatin interaction studies.

Capture-C methods are a subset of 3C that enable high-throughput analysis of *cis*

interactions from pre-specified oligonucleotide probes or "viewpoints". The approach comprises three steps [423]:

1. *Library generation*: Cells from a single-cell population are lysed and their chromatin cross-linked using formaldehyde to preserve the 3-dimensional structure of chromatin interactions. DNA is digested with a restriction enzyme, and fragments are ligated to reflect the spatial organisation of chromatin at the time of fixation.

2. *Library indexing and amplification*: The 3C library is indexed using next-generation sequencing kits and amplified via polymerase chain reaction.

3. *Enrichment and sequencing*: Targeted enrichment is performed using oligonucleotide probes that overlap restriction enzyme digestion sites. These probes capture specific DNA fragments, which are then amplified and sequenced.

The resulting contact data, representing interaction frequencies between loci, provide critical insights into regulatory mechanisms and help to identify the genes influenced by causal variants. This detailed understanding of chromatin interactions is essential to elucidate the cell-specific genetic mechanisms driving disease pathophysiology. For example, 3C methods have been used to profile chromatin contact a domain near the *BIN1* promoter region present in kidney cortex but absent in the medulla and papilla [424]. This finding aligns with single-cell analyses that reveal that *BIN1* is expressed in proximal tubular cells [425].

## 1.5 Thesis aims

As discussed, KSD is a complex condition with a genetic architecture that remains incompletely understood. The overarching goal of my thesis is to employ a range of genetic epidemiological and bioinformatic approaches, each with distinct sources of bias, to construct a robust body of evidence that deepens our understanding of the genetic and molecular basis of KSD. This evidence will clarify the contributions of genetic variants to the phenotype and identify novel therapeutic opportunities. To achieve this, my thesis is structured around the following aims:

1. To determine the genetic architecture of KSD by performing GWAS and meta-analyses in European and trans-ancestral populations and ExWAS in a European population (Chapter 3).

2. To establish the effect of adiposity on the risk of KSD and identify potential disease-mediating genes to guide future therapeutic approaches (Chapter 4).

3. To investigate the impact of alterations in mineral homeostasis on risk of KSD and uncover genetic and molecular pathways that may influence KSD and offer therapeutic potential (Chapter 5).

4. Based on findings from aim 1, to evaluate the relationship between renal cysts and risk of KSD to identify shared genetic and molecular mechanisms that may underpin a common pathophysiology (Chapter 6).

# 2 | Methods

## 2.1 Cohorts

In this section, I describe the population cohorts used in this thesis (Table 2.1) and explain how each dataset has curated genetic data for researchers.

***Table 2.1. Cohorts leveraged in this thesis.***

| Cohort | Study Design | Ancestry | Participants | GWAS software |
|---|---|---|---|---|
| UK Biobank | Longitudinal, healthy volunteer cohort study | Primarily European (white British) | 500,000 | BOLT-LMM v2.4, REGENIE v3.1.1 |
| FinnGenR12 | National health register | European (Finnish) | 500,348 | REGENIE v2.2.4 or v3.3 |
| Biobank Japan | Clinician-recruited, disease-oriented biobank | East Asian (Japanese) | ≈270,000 | SAIGE v0.29.4.2 |
| *All of Us* | Longitudinal, healthy volunteer cohort study | European:≈ 66.4%, African:≈ 19.5%, Asian:≈ 7.6%, American:≈ 6.3% | 413,000 | Not applicable |

GWAS=Genome-wide association study

### 2.1.1 UK Biobank

The UK Biobank is a large-scale biomedical database developed from a prospective cohort study to support research on major disease risk factors in middle and old age (Table 2.1) [426, 427]. Between 2006 and 2010, researchers recruited 500,000 individuals aged 40–69 years from 22 centres across the UK [292, 345]. With consent, participants completed detailed questionnaires, participated in interviews, and underwent physical measurements. They also provided blood and urine samples for genomic profiling. Researchers linked these data to primary and secondary

healthcare records and have followed up with participants at multiple intervals [292, 345].

The UK Biobank's size and scope make it a powerful resource, allowing researchers to examine a wide range of exposures and their effects on disease outcomes. Its prospective design enables researchers to assess exposures before disease diagnosis, reducing confounding from disease progression, treatment, or post-diagnosis lifestyle changes [426]. This approach strengthens confidence in exposure-outcome relationships while minimising bias.

Despite its strengths, the UK Biobank has limitations. As a voluntary cohort study, it suffers from healthy volunteer bias, which makes participants healthier than non-participants and less representative of the general population [428–431]. UK Biobank participants are more likely to be older, female, and from less socio-economically deprived areas than the general population. They are also less likely to be obese, smoke, drink alcohol daily, or report multiple health conditions [430]. Consequently, researchers may struggle to generalise findings to socio-economically deprived or less healthy individuals. The lower disease prevalence in this cohort may also lead to underestimating risk factors. Additionally, the selective nature of the study population may skew relationships between genetic, environmental, and behavioural factors and disease outcomes.

Another limitation is the under-representation of non-white ethnic groups [345]. Approximately 94% of participants identify as "white", while only ≈4% (≈22,000/502,000) report non-European ethnic backgrounds [345].

A third limitation is the restricted availability of hospital inpatient data. The UK Biobank captures hospital records through Hospital Episode Statistics for England, the Scottish Morbidity Record, and the Patient Episode Database for Wales [432]. These records have been available only since ≈1996, limiting the accuracy of disease prevalence estimates based on International Classification of Diseases (ICD-9/10) and Classification of Interventions and Procedures (OPCS-3/4) codes.

The National Information Governance Board for Health and Social Care and the NHS North West Centre for Research Ethics Committee (Ref: 11/NW/0382)

have approved the UK Biobank. This thesis is based on study application ID 83942 (*"Investigating the genetic and non-genetic causes of surgical disease"*). The respective sections detail the methods used for analyses with UK Biobank data.

#### 2.1.1.1 Genetic data curated by the UK Biobank

UK Biobank researchers phased and imputed genotype data using the UK BiLEVE and UK Biobank Axiom Arrays. They called genotypes from array intensity data with a custom genotype-calling pipeline [345]. For autosomal phasing, they used SHAPEIT3 with the 1000 Genomes Phase 3 dataset as a reference panel. They performed imputation with the Haplotype Reference Consortium reference panel and a merged UK10K/1000 Genomes Phase 3 panel. This imputation process generated a dataset containing 93,095,623 autosomal SNPs, short indels, and large structural variants across 487,442 individuals, along with an additional 3,963,705 markers on the X chromosome [345].

### 2.1.2 FinnGen

Finland's unique population history, shaped by a strong bottleneck, has led to an increased frequency of certain rare, deleterious alleles compared to other European populations [294]. Launched in 2017, the FinnGen study is a public-private partnership that integrates genetic data with digital health records from Finnish health registries to advance the understanding of disease genetics (Table 2.1) [294]. The combination of Finland's distinct genetic background and comprehensive longitudinal healthcare data provides valuable insights into the genetic basis of health and disease. Finland's nationwide electronic health registers cover the lifespan of each resident, offering near-complete documentation of hospital admissions, outpatient prescriptions, medical procedures, and mortality. Since FinnGen participants are recruited through research studies, hospitals, or blood donation centres, the cohort is enriched for individuals with prior healthcare interactions. Consequently, FinnGen is less affected by healthy volunteer bias than large-scale volunteer-based cohorts such as the UK Biobank.

FinnGen Release 12 includes data on 500,348 participants (282,064 females and 218,284 males) and 2,502 phenotypes [433]. The Coordinating Ethics Committee of the Helsinki and Uusimaa Hospital District has reviewed and approved the FinnGen study.

#### 2.1.2.1 Genetic data curated by the FinnGen R12 study

Researchers genotyped samples using both Illumina and Affymetrix arrays. They generated genotype calls with the GenCall and zCall algorithms for Illumina and the AxiomGT1 algorithm for Affymetrix data. Before imputation, they pre-phased the chip-genotyped data from 8,554 individuals using Eagle 2.3.5, adjusting the number of conditioning haplotypes to 20,000 while keeping the default settings. They performed genotype imputation with the population-specific SISu v4.0 reference panel and produced a variant call set using the GATK HaplotypeCaller algorithm, following GATK best practices for variant calling. They applied the following quality control (QC) parameters to the genetic data:

- Minor allele count (MAC) $> 3$
- Hardy–Weinberg equilibrium p-value $>1\times10^{-6}$
- Missingness $< 2\%$

In FinnGen R12, researchers conducted GWAS for 2,502 endpoints and 21,311,644 variants across 500,348 samples using REGENIE v2.2.4 [434]. For endpoints that failed to converge under v2.2.4, they used REGENIE v3.3, which relaxes the convergence criterion of the Frith likelihood ratio test from $1\times10^{-4}$ to $2.5\times10^{-4}$.

Before performing PCA, they iteratively pruned the SISu version 4.2 imputation panel, starting with 9,641,808 variants and retaining only bi-allelic, autosomal variants with a minimum info score of 0.9 and a minor allele frequency (MAF) greater than 0.01. They evaluated SNPs within a 500 kilobase pair (Kbp) window in 50 Kbp steps, removing one of any pair with an $r^2$ greater than 0.9 until the SNP count dropped below 200,000. If the higher $r^2$ was closer, they randomly selected

200,000 SNPs; otherwise, they returned the last pruned SNP. This process yielded 180,042 SNPs.

Next, they used PCA to identify outlier samples by integrating 1K Genome Project data and a Bayesian algorithm. This step excluded 15,898 samples, including those with non-European and southern European ancestries. To further refine population stratification and ensure genetic homogeneity, they conducted an additional PCA on the remaining samples and identified 879 more outliers of Western European origin.

Using kinship analysis, they identified all pairs of FinnGen samples up to the second degree of relatedness. They then calculated PCA on 296,829 "inliers" (unrelated individuals with Finnish ancestry) and projected 203,908 "outliers" (related to the inliers) onto the same principal component space to generate covariates for 500,737 samples. Finally, they excluded 355 samples that lacked the minimum phenotype data and 34 samples with mismatched recorded and inferred sex based on PLINK's F-statistic for the inbreeding coefficient on the X chromosome.

As a result, FinnGen R12 includes 500,348 samples (282,064 females and 218,284 males) in its core analyses.

### 2.1.3 Biobank Japan

Biobank Japan is one of the largest biobanks in East Asia, maintaining a registry of patients diagnosed with any of 47 pre-specified common diseases (Table 2.1) [435]. Between 2003 and 2008, researchers enrolled participants from 12 medical centres across Japan. Like the UK Biobank, they collected clinical data through interviews and medical records, with follow-up continuing until 2013. They also collected blood samples annually until 2013 to support genomic analyses [435]. Currently, Biobank Japan includes clinical data and samples from approximately 270,000 participants, with genotype data available for roughly 180,000 individuals [436].

#### 2.1.3.1 Genetic data curated by Biobank Japan

Researchers conducted sample QC by excluding individuals with a call rate <0.98, close relationships identified through identity-by-descent analysis, sex mismatches, and non-East Asian outliers detected via PCA. They used African, European, and East Asian reference panels from the HapMap Project [290].

For genotyping, they analysed ≈1 million SNPs using either the Illumina HumanOmniExpressExome BeadChip or a combination of the Illumina HumanOmniExpress and HumanExome BeadChips [437, 438]. They applied QC filters to include only variants that met the following criteria in analyses:

- MAF >1%
- Hardy–Weinberg equilibrium p-value $>1\times10^{-6}$
- Call rate >0.99

Researchers performed phasing for imputation using MACH [439] and imputed genotype dosages with Minimac, relying on the 1000 Genomes Project Phase 1 (version 3) East Asian reference haplotype panel [440]. After imputation, they identified 8,712,794 autosomal variants and 207,198 X chromosome variants for association analyses.

### 2.1.4 *All of Us*

The *All of Us* Research Programme, launched in the United States in May 2018, provides researchers with access to a large, diverse, and comprehensive health dataset from over one million participants (Table 2.1) [441]. This government-funded longitudinal cohort study integrates whole-genome sequencing, genotype, and phenotype data with electronic health records, biospecimens, laboratory assessments, survey responses, and physical measurements. By recruiting participants from diverse ancestries, it enhances the power and generalisability of epidemiological and genetic analyses for studying human diseases.

The programme prioritises enrolling adults (≥18 years) from demographics that

are often under-represented in biomedical research, acknowledging the disparities these groups face in accessing high-quality healthcare [441]. To achieve this goal, researchers focus on diversity in race, ethnicity, age, sex, gender, sexual orientation, disability status, healthcare access, income, education, and geographic location. The study aims to include at least 45% of participants from racial and ethnic minorities and over 75% from under-represented populations.

As of early 2024, the data release includes information from over 413,000 individuals, including clinical-grade whole-genome sequences for 245,388 participants and genome-wide genotyping for 312,925 participants. Among this cohort, 77% belong to communities typically under-represented in biomedical research, and 46% identify as racial or ethnic minorities [442]. Researchers have access to electronic health records standardised using the Observational Medical Outcomes Partnership Common Data Model [443] for 287,000 participants across more than 50 healthcare provider organisations [442]. A quarter of these participants have over 10 years of electronic health record data.

## 2.2 Genome-wide association studies

In this section, I describe the genome-wide association analyses that I conducted in the UK Biobank (Section 2.2.1) and explain how the research teams behind the FinnGen study (Section 2.2.2) and Biobank Japan (Section 2.2.3) generated the GWAS summary statistics for KSD, which I utilise in this thesis.

### 2.2.1 Genome-wide association studies in the UK Biobank

I performed GWAS using either BOLT-LMM v2.4 or REGENIE [434, 444]. I used BOLT-LMM for the GWAS conducted early in my DPhil studies and REGENIE later, when analysing the UK Biobank data launched their Research Analysis Platform (https://ukbiobank.dnanexus.com). For sample-level QC, I excluded individuals who had one or more of the following:

- Not white British ancestry based on principal component analysis and self-reported ethnicity (Field 22006)
- Call rate <98%
- Discordant genetic- (Field 22001) and self-reported sex (Field 21)
- Sex chromosome aneuploidy (Field 22019)
- Heterozygosity >3 standard deviations from the mean in UK Biobank principal component analysis-adjusted heterozygosity values (Field 20004)

I derived phenotypes for GWAS from International Classification of Diseases (ICD) revisions 9 and 10, the Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures (OPCS) revisions 3 and 4, primary care read codes version 2 and version 3, and self-report codes. I detail the inclusion and exclusion codes in the relevant chapters for each GWAS.

### BOLT-LMM

I conducted a GWAS of KSD using UK Biobank data, applying a linear mixed non-infinitesimal model in BOLT-LMM v2.4 [444]. I included genetic sex and genotyping platform as covariates. This model accounts for population substructure and cryptic relatedness [444]. I performed genotype-level QC using R (version 4.2.0) and PLINK v2.0 [445, 446]. For SNP-level QC, I excluded variants with a Hardy–Weinberg equilibrium p-value $<10^{-4}$, call rate <98%, and MAF <1%. After QC, 547,011 autosomal genotyped variants and 8,397,548 imputed variants remained for analysis.

In the association analyses, I utilised an hg19 reference genetic map and restricted the analysis to variants in HapMap3 [447]. I computed LD scores using data from the 1000 Genomes European sample [440, 448, 449]. SNPs of info score $< 0.9$ were excluded from resulting summary statistics.

**REGENIE**

I conducted a GWAS of renal cysts using UK Biobank data through a two-step procedure with REGENIE v3.1.1 [434], including sex, age at recruitment, and the first 10 principal components as covariates. In the first step, REGENIE fits a whole-genome regression model for trait prediction based on genetic data using the leave-one-chromosome-out (LOCO) scheme [434]. The second step performs variant association analyses. I applied the following QC parameters to genotyped and imputed genetic data to identify a set of high-quality variants for Step 1:

- MAC >10
- MAF >0.01
- Hardy–Weinberg equilibrium p-value $>1\times10^{-6}$
- Variant missingness <0.1
- LD pruning (window size 1, 000, step size 50, and $r^2 < 0.4$)
- Sample missingness <0.1

I used the resulting set of genetic variants for REGENIE Step 2, incorporating the LOCO predictions from Step 1. In Step 2, I applied Firth approximation when the p-value from the standard regression test was <0.01. I then used effect size estimates and likelihood ratio test p-values to compute standard errors.

### 2.2.2 Genome-wide association studies in the FinnGen study

I used publicly available data from a GWAS of KSD from FinnGen R12, comprising 499,184 participants (12,999 cases and 486,185 controls) in work underpinning this thesis [433]. The methods, adapted from the FinnGen R12 documentation [433], are summarised below:

ICD-8 and -9 code N592, and -10 code N20 (*"Calculus of kidney and ureter"*) identified 13,720 participants with KSD from hospital discharge notices and death registries. Individuals with ambiguous gender, genotype missingness>5%, heterozygosity ±4

standard deviations, or non-Finnish ancestry were excluded by FinnGen researchers during sample QC.

Sex, age, 10 principal components, FinnGen chip version, and genotyping batch were included as covariates in the GWAS. To generate the whole-genome regression model in Step 1, which informs association analyses in Step 2, LOCO relatedness was calculated using only genetic variants meeting the following QC parameters:

- MAF >0.01
- LD pruning (window size=150,000, and $r^2 < 0.2$)
- Missingness <3%
- Bi-allelic variants
- Info score >0.95

As a result, 188,153 well-imputed, common variants were used in the relatedness calculation. Association testing used variants with a minor allele count >5. Firth approximation was applied to improve accuracy for variants with a p-value <0.01 from standard regression.

### 2.2.3 Genome-wide association studies in Biobank Japan

Clinicians participating in Biobank Japan identified 5,587 patients with KSD. For GWAS, a control group of 28,870 individuals was drawn from four population-based cohorts: the Japan Public Health Center-based Prospective Study [450], the Japan Multi-Institutional Collaborative Cohort Study [451], the Iwate Tohoku Medical Megabank Organization, and the Tohoku Medical Megabank Organization [452]. GWAS were performed using a logistic regression model implemented in SAIGE v0.29.4.2 [453], adjusting for age, sex, and the top 10 principal components as covariates.

### 2.2.4 Meta-analysis of genome-wide association studies

To increase statistical power and provide more robust insights into the genetic basis of KSD, I performed a fixed-effect meta-analysis of SNPs with MAF >0.01 and info >0.6 using METAL and weighted effect sizes using the inverse of corresponding standard errors (Table 2.1) [454]. The GWAS of KSD in the UK Biobank that I conducted as described in Section 2.2.1 captured 11,186 cases and 390,488 controls. I downloaded publicly available summary statistics from FinnGen R12 [433] and converted the genomic build from GRCh38 to GRCh37 using CrossMap 0.6.4 [455]. FinnGen GWAS data comprised 12,999 cases and 486,185 controls, and 1,154,394 SNPs (Section 2.2.2). Data from the KSD GWAS conducted in Biobank Japan [456] included 5,587 cases, 28,870 controls, and 688,938 SNPs with imputation score $r^2$ >0.5, and MAF <0.01 (Section 2.2.3). I excluded variants with a high degree of heterogeneity across studies ($I^2$ >75.00).

### 2.2.5 Defining index association signals

To identify independent association signals in KSD GWAS in the UK Biobank and GWAS meta-analyses, I performed step-wise approximate conditional and joint analysis using Genome-Wide Complex Trait Analysis (GCTA) software version 1.94.1 with a European LD reference panel [457, 458]. For loci with a single association signal (defined as a chromosomal region where adjacent pairs of KSD-associated SNPs are <1 megabase pair (Mbp) apart [457, 459]), I defined the index SNP as the lead SNP from the unconditional analysis with the lowest p-value. I performed conditional analyses iteratively for loci with multiple association signals and defined the index SNP as the one with the lowest p-value in the conditional analysis. I define "novel loci" as those significantly associated with KSD ($p < 5 \times 10^{-8}$) and not reported in, or within 1 Mbp of, a variant/independent locus significantly associated with KSD in previously published GWAS studies.

### 2.2.6 Estimating heritability

To estimate heritability, I used LD score regression (LDSC) v1.0.1 with an hg19 reference genetic map and a European ancestry LD reference [448, 449]. I defined a population prevalence of 10% to estimate the SNP-based heritability of KSD using a liability transformation.

### 2.2.7 Gene and gene set analyses

I identified genes and gene sets associated with KSD using Multi-marker Analysis of GenoMic Annotation (MAGMA) v1.10 [460, 461]. MAGMA applies a multiple regression model to integrate variant-level p-values, LD, and gene positions to detect multi-SNP effects [460]. I mapped variants from the UK Biobank KSD GWAS, European ancestry KSD GWAS meta-analysis, and trans-ancestry KSD GWAS meta-analysis to 18,633 protein-coding genes based on genomic location. A SNP was assigned to a gene if it fell within its transcription start and stop sites. Gene associations were determined using a SNP-wise mean model, with LD patterns estimated from 1,000 Genomes Project genotypes. Bonferroni correction was applied for multiple testing ($p<2.68\times10^{-6}$).

To identify biological pathways or cell types relevant to KSD, I conducted self-contained and competitive analyses [460]. Self-contained analysis tested whether genes in a set were jointly associated with KSD, while competitive analysis assessed whether these genes were more strongly associated with KSD than other genes. Gene sets based on Human Phenotype Ontology (HPO) and GO terms were obtained from the Molecular Signature Database v2022.1 [461–463]. Bonferroni correction was again applied for significance thresholds ($P<0.05$/number of gene sets tested and $P<0.05$/number of genes in gene set, respectively).

## 2.3 Exome-wide association studies

### Exome sequencing and variant calling

Exome sequencing and variant calling by the UK Biobank have been previously described [292, 321]. Researchers transferred genomic DNA samples from peripheral blood to the Regeneron Genetics Centre, where they stored them at $-80^{\circ}c$. UK Biobank scientists enzymatically fragmented DNA to a mean size of 200 base-pairs (bp) and added asymmetric 10 bp barcodes to enable multiplexed exome capture and sequencing. They used a modified version of the Integrated DNA Technologies xGen probe library for overnight exome capture, then amplified and quantified the captured DNA fragments using polymerase chain reaction (PCR) before sequencing them in a multiplex manner with 75bp paired-end reads and two 10 bp index reads on the Illumina NovaSeq 6000 platform.

The UK Biobank platform mapped WES reads to the hg38 reference genome using BWA MEM [321]. It generated per-sample gVCF files with WeCall and aggregated them with GLnexus into a joint-genotyped, multi-sample project-level VCF (pVCF). The pipeline excluded SNPs with a read depth <7 and indels with a read depth <10. Additionally, a variant-level allele balance filter retained only variants with $\geq$ 1 homozygous variant carrier or $\geq$ 1 heterozygous carrier with an allele balance of $\geq$ 0.15 for SNVs and $\geq$ 0.2 for indels. Samples with discrepancies between genetic and reported sex, excess heterozygosity or contamination, inconsistencies between exome and array genotype calls, or $\geq$80% targeted bases achieving 20$\times$ coverage were omitted by the UK Biobank team. Genetically identified sample duplicates and WES variants discordant with genotyping chips were also excluded. The final pVCF, comprising 454,787 samples, was generated for downstream analysis.

### Single-variant association analyses

I performed exome-wide association analyses using REGENIE v3.1.1 [434] with the same two-step approach as described in Section 2.2.1. To fit a relatedness model that could be integrated into Step 2, I converted genotype array data from build

Hg37 to Hg38 using Picard liftover VCF [464]. Using PLINK2.0 [445, 446], I applied the following QC parameters to genotyped variants and participant samples prior to running REGENIE Step 1:

- MAF >0.01
- MAC>20
- LD pruning (window size 1, 000, step size 50, and $r^2 < 0.4$)
- Genotyped missingness <10%
- Sample missingness missing <10%

In REGENIE Step 1, I generated a whole-genome regression model for trait prediction based on genetic data using the LOCO scheme and included sex, age at recruitment, and the first 10 principal components as covariates.

In REGENIE Step 2 (association analyses), I included only exome sequence variants and participant samples that satisfied the following QC criteria:

- MAC >10
- Hardy–Weinberg equilibrium p-value $>1\times10^{-6}$
- Variant missingness <0.1
- Sample missingness <10%

Again, I included age at recruitment, sex, and the first 10 principal components as covariates in REGENIE Step 2, incorporating the LOCO predictions from Step 1. I applied Firth approximation when the p-value from the standard regression test was <0.01. I based standard errors on effect size estimates and likelihood ratio test p-values.

## Variant annotation

Using SnpEff, I annotated variants from WES with the most severe consequence for each variant across all protein transcripts [332, 465]. I defined gene coordinates based

on Ensembl [466]. Based on SnpEff annotations, I defined variants as described in 2.2 and used these masks, along with variant Combined Annotation Dependent Depletion (CADD) scores, to perform collapsing analyses.

***Table 2.2. Definitions of whole-exome sequencing variants based on SnpEff annotations.***

| Definition | Annotations included |
|---|---|
| Protein-truncating variants | *Exon_loss_variant*<br>*Frameshift_variant*<br>*Start_lost*<br>*Stop_gained*<br>*Stop_lost*<br>*Splice_acceptor_variant*<br>*Splice_donor_variant*<br>*Gene_fusion*<br>*Bidirectional_gene_fusion*<br>*Rare_amino_acid_variant*<br>*Transcript_ablation* |
| Missense variants | *Missense_variant_splice_region_variant*<br>*Missense_variant* |
| Non-synonymous variants | *Conservative_inframe_deletion*<br>*Conservative_inframe_insertion*<br>*Disruptive_inframe_insertion*<br>*Disruptive_inframe_deletion*<br>*Protein_altering_variant* |

## Collapsing analyses

In collapsing analyses, I considered rare (MAF <1%) protein-truncating variants (PTV), missense, and non-synonymous variants. I performed additional analyses considering variants with CADD scores ≥25. I included age at recruitment, sex, and the first 10 principal components as covariates and considered three burden masks per gene based on the frequency of the alternative allele of the variants screened in that group: MAF ≤0.01, MAF ≤0.001, and MAF ≤0.0001. I corrected Sequence Kernel Association Test-O (SKATO) p-values for multiple testing using Bonferroni adjustment according to the number of variants tested in the mask. This statistical method computes an optimal average of the SKAT

and burden tests to analyse rare variant effects in sequencing association studies [434].

Leave-one-variant-out (LOVO) analysis was conducted for gene masks significantly associated with KSD. This method involves iteratively testing the full set of variants within the mask, each time omitting a single variant. By systematically excluding individual variants, the LOVO analysis can identify whether the association is driven by a specific variant or the combined effect of multiple variants in the mask. In each gene-based LOVO analysis, omission of a variant was considered significant if the test p-value became non-significant ($p>(0.05/\text{number of variants in the mask})$).

## 2.4 Mendelian randomisation

MR assumes that IVs are associated with the exposure variable (relevance); there are no unmeasured confounding relationships (exchangeability); and variants are associated with the outcome only through changes in the exposure variable (exclusion restriction) [301].

For MR analyses, I considered three KSD outcome datasets. The primary analyses are based on the European-ancestry KSD GWAS meta-analysis, as described in Chapter 3 Section 3.3.2. Sensitivity analyses comprise results from the combined-sex KSD GWAS data from the UK Biobank, detailed in Chapter 3 Section 3.3.1, and the publicly available FinnGen study release 12 summary statistics [433].

### Pan-genome Mendelian randomisation

To generate IVs for phenotypes of interest, I selected SNPs with independent, GWAS-significant associations ($p < 5\times10^{-8}$) from relevant GWASs of individuals of European ancestry. To minimise the risk of type-1 error from IVs being in LD, I performed further pruning using the `clump_data()` function, which retains the SNP with the lowest p-value, with clump_r2 set at 0.01 using a European population reference panel. I harmonised exposure IVs with outcome IVs from KSD

GWAS summary statistics, using allele frequencies to infer positive strand alleles for palindromic IVs. I omitted IVs with ambiguous positive strand alleles that could not be harmonised. Using `https://sb452.shinyapps.io/overlap/`, I performed a power calculation to determine the minimum and maximum effects detectable with 80% statistical power [381, 467].

When interpreting results from MR analyses, I used the IVW method and Wald ratio estimates. I employed the MR-Egger intercept and estimate to explore pleiotropic relationships; where the MR-Egger intercept was significantly different from zero ($p < 0.05$) I interpreted the MR-Egger estimate as the estimate of best fit. Results are presented as effect estimates and corresponding 95% confidence intervals (CI) per standard deviation (SD) increase in exposure trait on the odds ratio of KSD. I corrected P-values for multiple testing using the Benjamini-Hochberg false discovery rate method, controlled at 5% and consider corrected p-values <0.05 as evidence against the null hypothesis [468]. To calculate the variance in exposure traits explained by the IVs, I used the following calculation where the genetic association with the risk factor (â) is in SD units, $\beta$ is the log-odds effect of the SNP, N is the sample size for the IV–outcome association, and K is the number of IVs [469]:

$$R^2 = [2 \times \text{MAF} \times (1\text{-MAF}) \times \beta^2]$$

To assess heterogeneity in causal estimates, I implemented Cochrane's Q test using `mr_heterogeneity()` in `TwoSampleMR` [385, 470].

## Regional Mendelian randomisation

To identify genomic regions with causal effects on KSD risk and prioritise those that may harbour causal variants, I estimated the regional (1 Mbp) effects of genetically predicted exposure variables on odds of KSD using the `TwoSampleMR` package in R (version 4.3.1) [385, 470] and GWAS summary statistics for the exposure trait. I selected independent genetic variants ($r^2$ <0.1) within ±500 Kbp of the lead independent variants that showed significant associations with the exposure trait ($p < 5 \times 10^{-8}$) as IVs [385, 470]. I pruned and harmonised IVs, interpreted estimates, corrected p-values, and assessed heterogeneity as described in pan-genome MR.

### Drug-target Mendelian randomisation

To estimate the potential utility of modulating drug targets to prevent KSD, I performed drug-target MR analyses. Exposure IVs comprise genetic proxies within ±300 Kbp of putative therapeutic genetic targets that are significantly associated with the relevant exposure traits ($p<5\times10^{-8}$). I used KSD GWAS data as the outcome IV for MR analyses, followed the principles described in Section 2.4, and conducted sensitivity analyses using the `clump_data` function with LD $r^2$ thresholds of 0.1 and 0.01. I identified possible off-target effects of putative therapeutics by collating phenotypes associated with variants linked to these genes ($p<5\times10^{-8}$) via the Open Targets Genetics portal (https://genetics.opentargets.org/) [471, 472].

## 2.5 Colocalisation

To evaluate the probability of a single shared causal variant between two or more traits, I performed colocalisation analyses using `coloc` and `HyPrColoc`, respectively, in R (version 4.3.1) [302, 407, 473]. *Coloc()* integrates evidence over all variants at a locus to enable evaluation of the following hypotheses for trait 1 and trait 2 [407, 473]:

- H0: Genomic region not associated with either trait
- H1: Genomic region associated with trait 1 but not trait 2
- H2: Genomic region associated with trait 2 but not trait 1
- H3: Genomic regions associated with trait 1 and trait 2 with two separate putative causal variants
- H4: Genomic region associated with trait 1 and trait 2 with one putative causal variant.

Where I used `coloc.abf()` from the `coloc` package, I set prior probabilities at $p_1 = 1 \times 10^{-4}$, $p_2 = 10 \times 10^{-4}$, and $p_{12} = 1 \times 10^{-5}$. I considered posterior probability

(PP) H4 ≥0.8 as strong evidence of colocalisation [407]. When using `hyprcoloc()` from the `HyPrColoc` package to facilitate multi-trait colocalisation analyses, I set the prior probabilities to the defaults (prior.1=$1\times10^{-5}$ and prior.12=$10\times10^{-5}$). Again, I interpreted PP ≥0.8 as strong evidence of colocalisation. Where strong evidence of colocalisation was present (PP H4 ≥0.8), I identified a 95% credible set to highlight candidate causal variants.

## 2.6 Single-cell analyses

To generate paired, cell-type specific transcriptional and chromatin accessibility profiles of healthy, adult human kidney cell lines and investigate the cell-specific effects of variants identified through GWAS, MR, and colocalisation analyses, I leveraged publicly available snRNA-seq and snATAC-seq data [417, 474]. Tissue originated from five patients undergoing partial or radical nephrectomy at Brigham and Women's Hospital (Boston, MA). "Control" cells were retrieved from parts of the kidney that were unaffected by a tumour. Sample processing, library preparation, and sequencing is detailed by Muto et al. [417]. Data aggregation and QC were performed by Dr Sarah Howles in R (version 4.3.1) with Seurat v5.1.0 [475], Signac v1.13.0 [476], DoubletFinder v2.0.4 [477], and Harmony [478].

### 2.6.1 snRNA-seq preparation

Raw snRNA-seq data were processed using CellRanger (v*3.1.0*), with the GRCh38 human genome as the reference [475]. Data from five control samples were aggregated into a single snRNA-seq object. Low-quality nuclei were excluded based on the following criteria:

- <500 or >3,000 features
- RNA counts >10,000
- Mitochondrial gene content >0.8%

- Ribosomal protein large or small subunit content >0.75%

Heterotypic doublets were identified and removed. The proportion of mitochondrial reads was regressed out to eliminate bias from cells with high mitochondrial contamination. The data were normalised to account for technical noise and to standardise expression data across cells. Batch effects were corrected using Harmony. Clustering was performed using a shared nearest neighbour (SNN) graph and Louvain algorithm. Dimensionality reduction was carried out with uniform manifold approximation and projection (UMAP) and clusters were annotated based on the expression of cell-type-specific markers (Table 2.3).

***Table 2.3. snRNA cluster annotation based on cell-type specific markers.***

| Cluster | Cell-type markers |
|---|---|
| Distal tubule cells | *AQP2, AQP3, CALB1, PTGS2, SCNN1G, SLC12A3, TRPM6, TRPV5* |
| Endothelial cells | *EMCN, FLT1, PECAM1* |
| Fibroblast cells | *ACTA2, COL1A1, LUM, TNC, CALD1, ITGA8, PDGFRB* |
| Glomerular cells | *AQP1, CD2AP, CFH, CLDN5, CLIC5, CRB2, ITGB5, NPHS1, PODXL, PTPRO, SYNPO, WT1, NPHS2, PTH2R* |
| Intercalated A cells | *ATP6V0D2, FOXI1, SLC26A7, SLC4A1* |
| Intercalated B cells | *ATP6V0D2, FOXI1, SLC26A4* |
| Parietal epithelial cells | *CFH* |
| Proximal tubule cells | *ALDOB, CUBN, DPEP1, FAM151A, GLYAT, HAVCR1, HNF1A, HNF4A, HNF4G, SLC22A8, SLC5A1, SLC5A2, VCAM1* |
| Thick ascending limb of loop of Henle cells | *CLDN16, SLC12A1, UMOD* |

The final snRNA-seq library comprised 19,812 cells with annotations for 19,918 genes. These data were subsequently integrated with snATAC-seq data.

### 2.6.2 snATAC-seq preparation

To detect chromatin accessibility, which plays a crucial role in understanding gene regulation and cellular identity, data from five control samples were aggregated into

a single snATAC-seq object. Low-quality cells were excluded based on the following criteria:

- Peak region fragments <2,000 or >25,000
- Percentage of reads in peaks <30%
- Blacklist ratio >0.002
- Nucleosome signal <2
- Transcription start site (TSS) enrichment score<2.5
- Mitochondrial gene ratio >0.2

Following QC, normalisation was performed using term-frequency inverse-document-frequency (TFIDF). Dimensionality reduction was applied using single value decomposition (SVD) with latent semantic indexing (LSI), followed by UMAP for further reduction. A SNN graph was constructed and clustering was performed using the original Louvain algorithm. Batch effects were corrected using Harmony.

snATAC-seq data were integrated with paired snRNA-seq data to provide opportunity to gain a comprehensive view of gene expression and chromatin accessibility across renal cell types. A log-normalised gene activity matrix was created by calculating counts per cell for gene bodies and promoter regions up to 2,000 bases upstream of protein-coding genes annotated in Ensembl [466]. Transfer anchors were identified to facilitate label transfer between the snATAC-seq and snRNA-seq objects using canonical correlation analysis (CCA). A cell-type confidence threshold of 97% was applied to filter heterotypic doublets. Genomic regions containing snATAC-seq peaks were annotated using MACS2 [479]. The final snATAC-seq library comprised 19,812 cells with annotations for 232,963 genes.

### 2.6.3 Predicting *cis*-regulatory interactions

*Cis*-coaccessibility reflects the correlation of chromatin accessibility between pairs of genomic sites, in 500 Kbp windows, and serves as a proxy for potential regulatory

interactions [480]. I predicted cell-type-specific cis-regulatory chromatin interactions of peaks using Cicero implemented in R [480]. I subsetted the snATAC-seq object according to cell type and converted it to a cell dataset object using the `as.cell_data_set()` function. I processed each cell-specific object using the `detect_genes()` and `estimate_size_factors()` functions before applying LSI dimensional reduction and converting to a Cicero object using the UMAP coordinates derived from accessible chromatin peaks. I set 50 cells to aggregate per bin except in the case of the glomerular cell type which I set to 30 cells to aggregate per bin due to a small sample size. I estimated cell-type-specific connections using the `run_cicero()` function and identified modules of cis co-accessible sites using the `generate_ccans()` function. I annotated co-accessible sites to a gene when they overlapped with the gene's coordinates as defined by Ensembl v113 [466].

### 2.6.4 Calculating gene activity scores

To identify regulatory regions where chromatin accessibility predicts transcriptional output across different cell types, I calculated a gene activity score for each cell type using Cicero [480]. I proxied gene activity according to cis-coaccessibility at the coordinates corresponding to a gene's promoter, as defined by Ensembl v113 [466]. I calculated the gene activity matrix using the `build_gene_activity_matrix()` function and normalised it using `normalize_gene_activities()` to enable comparability across genes and cell types. To identify active regulatory regions where chromatin accessibility predicts transcriptional output across cells, I calculated a Pearson $r^2$ co-efficient between mean normalised RNA expression and gene activity across cell types for each gene. A significant positive correlation indicates that chromatin accessibility at a promoter region, as inferred by the gene activity score, is associated with transcriptional activity, suggesting a regulatory role for the region of interest.

## 2.7 Chromosome conformational capture

Capture-C, a subset of 3C methods, was performed by Dr Michelle Goldsworthy for three biological replicates in three cell lines (renal proximal tubule epithelial (RPTEC) cells; normal adult human kidney immortalised proximal tubule epithelial (HK2) cells; and immortalised human embryonic kidney cells (HEK293) cells) as described in the published protocol [423]. Single-cell suspensions of $5\times10^6$ cells in 5 mL growth medium were crosslinked using formaldehyde and lysed. 3C libraries were constructed by proximity ligation following DpnII digestion. After de-crosslinking and DNA extraction, the libraries were sonicated to 200 bp fragments, indexed, and amplified by PCR. To enrich the libraries, biotinylated oligonucleotides were designed and used in two rounds of hybridisation, biotin-streptavidin bead pulldown, bead washes and PCR amplification. The final libraries were assessed on a fragment analyser and sequenced by 150bp paired-end sequencing using the Illumina NovaSeq 6000 S4 Flow cell platform. I analysed data using the CapCruncher v0.3.12 pipeline (https://github.com/sims-lab/CapCruncher) [423] and the Peaky protocol v0.0.0.9002 [481]. For each cell line, I identified interactions for each biological replicate, merged, and binned them into five bins without a maximum distance parameter to create a model predicting read counts under the null. I extracted adjusted read counts for putative interactions from these models. I assessed contacts between bait and prey fragments using recommended settings and applied a default value of $\omega = -3.8$ when estimating the best fit for omega. I inferred that interactions were significant when marginal posterior probability of contact (rjmcmc_pos) was $> 0.1$ for fragments $\leq$250 Kbp from the bait, or rjmcmc_pos $> 0.01$ for fragments >250 Kbp from the bait [482, 483].

# 3 | The role of genetic variants in kidney stone disease

## 3.1 Introduction

In the introduction, I highlighted evidence supporting KSD as a complex disorder with a strong genetic component. As discussed in Chapter 1, Section 1.2.2, GWAS play a crucial role in identifying genetic variants associated with complex diseases. The recent release of UK Biobank primary care records offers an opportunity to improve case ascertainment, thereby increasing the statistical power of KSD GWAS and enhancing our understanding of its genetic architecture. Unravelling the genetic basis of KSD is essential for elucidating disease mechanisms, identifying high-risk individuals, and developing genotype-driven therapeutic strategies.

### 3.1.1 Aim and Objectives

The principal aim for this chapter is to determine the genetic architecture of KSD through genome-wide and exome-wide approaches. Specifically, my research objectives are to:

1. **Genome-wide association studies**
    - Conduct a combined-sex GWAS of KSD in the UK Biobank using improved case ascertainment compared to previous studies.
    - Conduct sex-stratified GWAS to investigate potential sex-specific genetic associations in KSD.
    - Conduct meta-analyses of combined-sex GWAS results across European and trans-ancestral populations to improve power and generalisability.

2. **Downstream bioinformatic analyses**

   - Leverage GWAS results to perform downstream bioinformatic analyses, including:
     - Identifying independent genetic association signals.
     - Estimating heritability.
     - Gene and gene set analyses to uncover biological mechanisms underlying KSD.

3. **Exome-wide association studies**

   - Conduct an ExWAS of KSD in the UK Biobank, leveraging a larger sample size and improved case ascertainment compared to previous studies.
   - Apply collapsing models to test for associations of rare and predicted damaging coding variants with KSD at the gene level.

## 3.2 Methods

### 3.2.1 Cohorts

Data from the UK Biobank were leveraged for GWAS and ExWAS. The cohort and the ethical approval are described in detail in Chapter 2 Section 2.1.1. Publicly-available summary statistics from FinnGen Release 12 (Chapter 2 Section 2.1.2) and data gleaned from collaboration with Biobank Japan (Chapter 2 Section 2.1.3) were leveraged for European ancestry and trans-ancestry GWAS meta-analyses.

### 3.2.2 Case ascertainment

KSD cases were identified from diagnostic and operative codes derived from primary and secondary healthcare sources in the UK Biobank as of August 2022 (Appendix A). Individuals were selected as KSD cases if they had any one of the codes listed in Appendix A.

### 3.2.3 Genome-wide association studies

Combined-sex and sex-stratified GWAS of KSD in the UK Biobank were performed using BOLT-LMM v2.4 including genetic sex and genotyping platform as covariates [444]. Full details of genotype, sample, and SNP-level QC are detailed in Chapter 2 Section 2.2.1.

#### 3.2.3.1 Meta-analyses

Unfortunately, the FinnGen R12 study does not report sex-specific KSD GWAS summary statistics, precluding sex-stratified meta-analyses. Methods used to generate combined-sex GWAS summary statistics in the FinnGen R12 study and Biobank Japan are detailed in Chapter 2 Section 2.2.2 and Chapter 2 Section 2.2.3, respectively. Meta-analyses were performed as detailed in Chapter 2 Section 2.2.4.

#### 3.2.3.2 Downstream bioinformatic analyses

**Identifying independent genetic signals**

As described in Chapter 2 Section 2.2.5, GCTA software version 1.94.1 was used to identify independent genetic associations [457, 458]. Loci significantly associated with KSD ($p<5\times10^{-8}$) and not reported in, or within 1 Mbp of, a variant/independent locus significantly associated with KSD in previously published GWAS studies were defined as "novel".

**Estimating heritability**

Estimates of genomic inflation and the heritability of kidney stone disease in each cohort and subsequent meta-analysis were obtained using LDSC v1.0.1 as described in Chapter 2 Section 2.2.6 [454, 457].

**Gene and gene set analyses**

To identify genes, gene sets, biological pathways and cell types relevant to KSD, I undertook gene and gene set based analyses were implemented in MAGMA v1.10 [460]. A full explanation is provided in Chapter 2 Section 2.2.7.

### 3.2.4 Exome-wide association study

I performed an ExWAS of KSD in the UK Biobank using a two-step approach in REGENIE v3.1.1 [434], as described in Chapter 2, Section 2.3. Single-variant associations were considered significant at $p<5\times10^{-8}$. Independent genetic signals were identified using GCTA, following the same approach as for GWAS (Chapter 2 Section 2.2.5).

#### 3.2.4.1 Collapsing analyses

To test for associations between rare and damaging coding variants and KSD at the gene level, I grouped variants into masks based on PTV, missense, and non-synonymous variants, and those with a CADD score $\geq$25 (Table 2.2). I performed collapsing and LOVO analyses using the SKAT-O p-value to assess significance (Chapter 2 Section 2.3) [332].

## 3.3 Results

### 3.3.1 Genome-wide association study of kidney stone disease in the UK Biobank

#### 3.3.1.1 Combined-sex genome-wide association study

A combined-sex GWAS of 11,186 cases and 390,488 controls in the UK Biobank, using a greater number of inclusion codes than have been used in previous studies (Appendix A) to optimise ascertainment, identified 18 independent, genome-wide ($p<5\times10^{-8}$) significant associations at 17 loci across 13 chromosomes (Figure 3.1, Figure 3.2, Table 3.1). Of these, two loci were novel (candidate genes in proximity to novel association signals: *FLOT1, ABCC6*, Figure 3.1, Table 3.1).

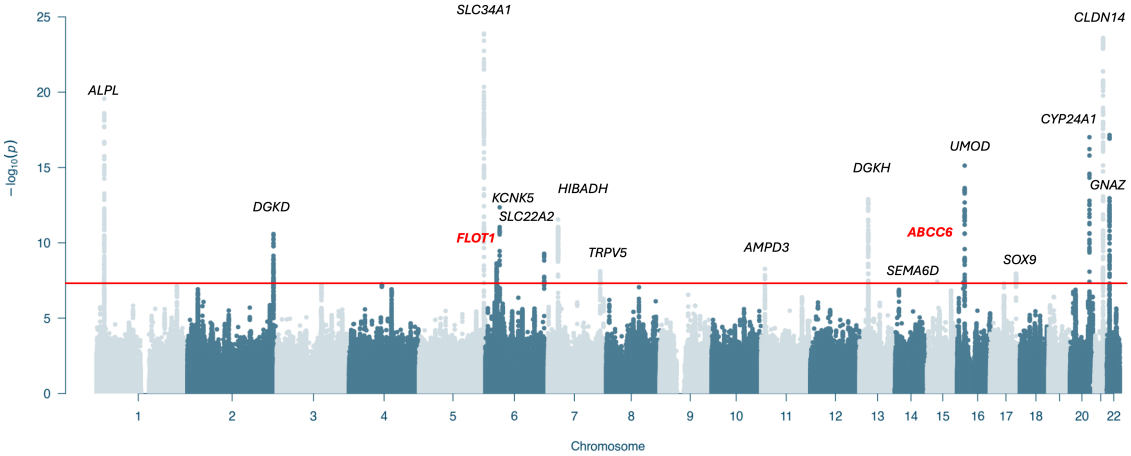



***Figure 3.1. Results of a genome-wide association study of kidney stone disease in the UK Biobank.***

Genome-wide p-values (-log10) plotted against their respective positions on each chromosome. The horizontal red line indicates the genome-wide significance threshold of $5.00\times10^{-8}$. Loci are labelled with the primary candidate gene at each locus, as shown in Table 3.1. Novel loci are shown in red.

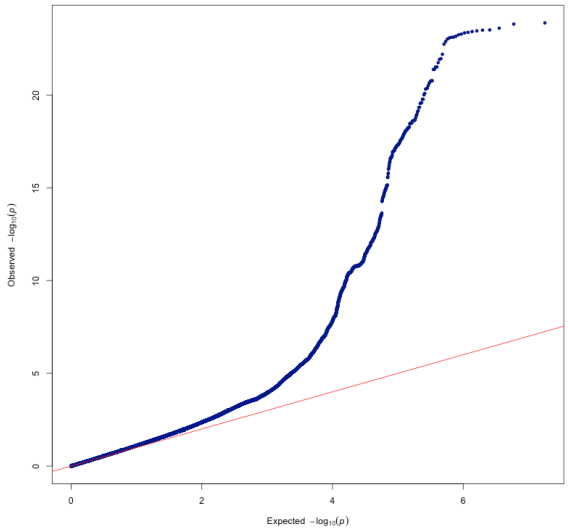



***Figure 3.2. Quantile-quantile plot of observed vs. expected p-values for a genome-wide association study of kidney stone disease in the UK Biobank.***

A genome-wide association study of kidney stone disease was performed for 11,186 individuals with kidney stone disease and 390,488 controls from the UK Biobank. The $\lambda$ GC demonstrated some inflation (1.15), but the LDSC intercept of 1.01, with an attenuation ratio of 0.09 indicated that the inflation was largely due to polygenicity and the large sample size.

***Table 3.1. Single-nucleotide polymorphisms significantly associated with kidney stone disease in the UK Biobank.***

| SNP | CHR | POS | Candidate Gene | EA | NEA | EAF | OR (95% CI) | P |
|---|---|---|---|---|---|---|---|---|
| rs77362499 | 1 | 21836204 | *ALPL* | G | C | 0.11 | 1.23 (1.17-1.28) | $2.70 \times 10^{-20}$ |
| rs1256332 | 1 | 21893344 | *ALPL* | A | C | 0.16 | 1.17 (1.13-1.21) | $3.81 \times 10^{-17}$ |
| rs838717 | 2 | 234296444 | *DGKD* | G | A | 0.43 | 1.10 (1.07-1.13) | $2.60 \times 10^{-11}$ |
| rs10051765 | 5 | 176799992 | *SLC34A1* | C | T | 0.33 | 1.16 (1.13-1.19) | $1.30 \times 10^{-24}$ |
| rs7774646 | 6 | 30760454 | *FLOT1** | T | G | 0.17 | 1.11 (1.07-1.15) | $2.30 \times 10^{-9}$ |
| rs1155347 | 6 | 39146230 | *KCNK5* | C | T | 0.22 | 1.13 (1.09-1.16) | $4.40 \times 10^{-13}$ |
| rs74348938 | 6 | 160614649 | *SLC22A2* | A | G | 0.03 | 1.28 (1.18-1.38) | $5.30 \times 10^{-10}$ |
| rs5883088 | 7 | 27610433 | *HIBADH* | G | GC | 0.33 | 1.11 (1.08-1.14) | $2.80 \times 10^{-12}$ |
| rs4252512 | 7 | 142605221 | *TRPV5* | C | T | 0.02 | 1.34 (1.22-1.49) | $7.70 \times 10^{-9}$ |
| rs7124537 | 11 | 10505718 | *AMPD3* | C | A | 0.09 | 1.15 (1.10-1.20) | $5.40 \times 10^{-9}$ |
| rs1182959 | 13 | 42674601 | *DGKH* | A | G | 0.18 | 1.14 (1.10-1.18) | $1.30 \times 10^{-13}$ |
| rs76594840 | 15 | 47900177 | *SEMA6D* | T | C | 0.92 | 1.15 (1.09-1.21) | $3.80 \times 10^{-8}$ |
| rs112414002 | 16 | 16284360 | *ABCC6** | C | T | 0.97 | 1.27 (1.16-1.38) | $4.40 \times 10^{-8}$ |
| rs77924615 | 16 | 20392332 | *UMOD* | A | G | 0.20 | 1.15 (1.11-1.19) | $7.50 \times 10^{-16}$ |
| rs9902482 | 17 | 70356993 | *SOX9* | C | T | 0.50 | 1.08 (1.05-1.11) | $1.10 \times 10^{-8}$ |
| rs17216707 | 20 | 52732362 | *CYP24A1* | T | C | 0.81 | 1.16 (1.12-1.20) | $9.70 \times 10^{-18}$ |
| rs219772 | 21 | 37835347 | *CLDN14* | A | T | 0.74 | 1.17 (1.13-1.21) | $2.50 \times 10^{-24}$ |
| rs13054904 | 22 | 23410918 | *GNAZ* | A | T | 0.26 | 1.14 (1.11-1.18) | $7.00 \times 10^{-18}$ |

Variants significantly associated with kidney stone disease from combined-sex genome-wide association study in the UK Biobank.
CHR=Chromosome; EA=Effect allele; EAF=Effect allele frequency; NEA=Non-effect allele; OR=Odds ratio; P=P-value; POS=Position based on NCBI Genome Build 37 (hg19); 95% CI=95% confidence interval. *denotes novel association.

#### 3.3.1.2 Sex-stratified genome-wide association studies

A GWAS of 7,427 male cases and 177,072 male controls in the UK Biobank identified 14 independent, genome-wide significant associations at 13 loci across 11 chromosomes (Figure 3.3, Figure 3.4, Table 3.2). Of these, one locus is not associated with KSD in previously published literature or in the combined-sex GWAS described in Section 3.3.1 (*CTCFL*, Table 3.1, Figure 3.1). A GWAS of

3,714 female cases and 213,416 female controls in the UK Biobank identified six independent, genome-wide significant associations at six loci across six chromosomes (Figure 3.3, Figure 3.4, Table 3.3). All loci identified in this analysis were associated with KSD in combined-sex GWAS (Figure 3.1, Table 3.1).

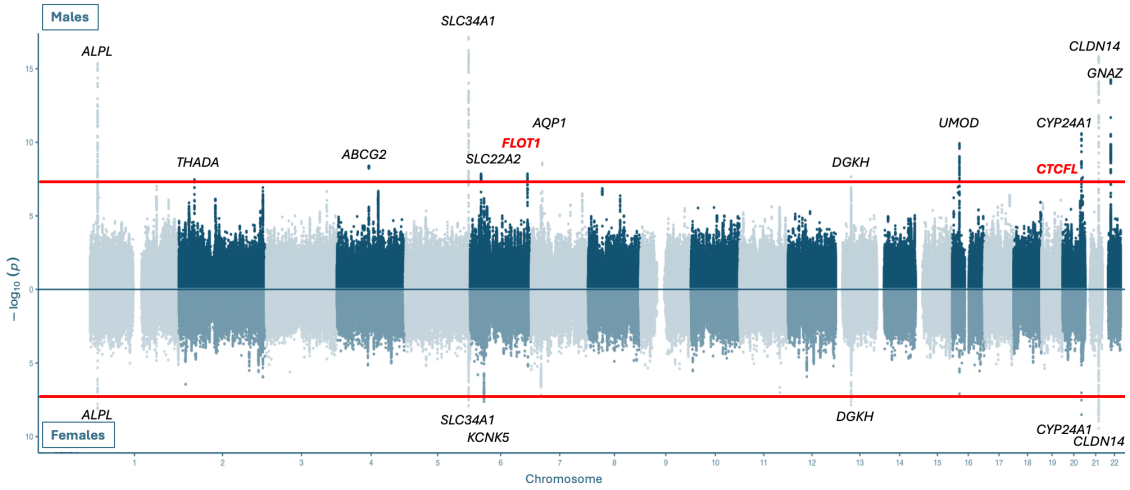



***Figure 3.3. Results of genome-wide association studies of kidney stone disease in males and females in the UK Biobank.***

Genome-wide p-values (-log10) plotted against their respective positions on each chromosome. The horizontal red line indicates the genome-wide significance threshold of $5.00\times10^{-8}$. Loci are labelled with the primary candidate gene at each locus, as shown in Table 3.2 and Table 3.3. Novel loci are shown in red. The upper plot shows results for male participants and the lower plot shows results for female participants.

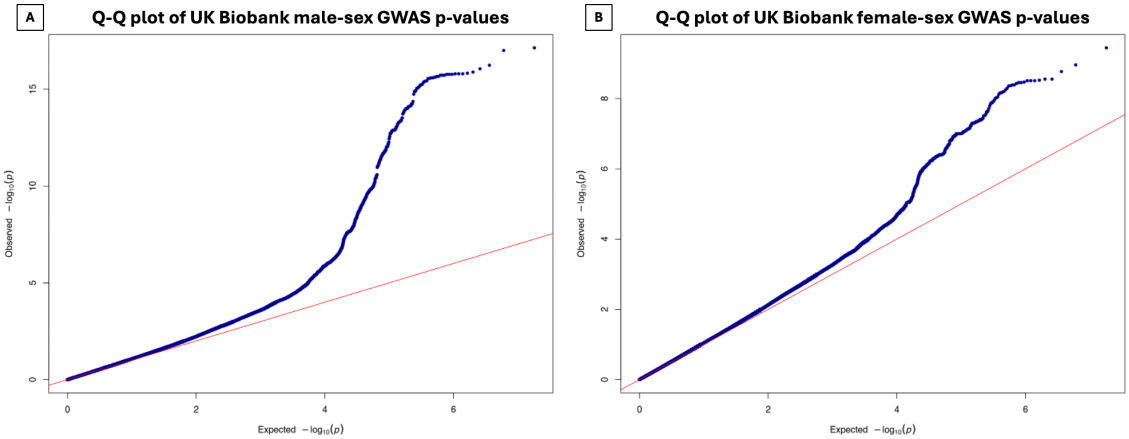



***Figure 3.4. Quantile-quantile plot of observed vs. expected p-values for male- and female-sex genome-wide association studies of kidney stone disease in the UK Biobank.***

**A**: A genome-wide association study (GWAS) of kidney stone disease (KSD) was performed for 7,427 males with KSD and 177,072 male controls. $\lambda$ GC demonstrated some inflation (1.10), but the LDSC intercept of 1.01, with an attenuation ratio of 0.06 indicated that inflation was largely due to polygenicity and sample size. **B**: A GWAS of KSD was performed for 3,741 females with KSD and 213,416 female controls. $\lambda$ GC demonstrated some inflation (1.05), but the LDSC intercept of 1.01, with an attenuation ratio of 0.12 indicated that inflation was largely due to polygenicity and sample size.

***Table 3.2. Single nucleotide polymorphisms significantly associated with kidney stone disease from male sex genome-wide association study in the UK Biobank.***

| SNP | CHR | POS | Candidate Gene | EA | NEA | EAF | OR (95% CI) | P |
|---|---|---|---|---|---|---|---|---|
| rs115239632 | 1 | 21826530 | *ALPL* | T | C | 0.04 | 1.43 (1.31-1.56) | $4.40 \times 10^{-16}$ |
| rs1256332 | 1 | 21893344 | *ALPL* | A | C | 0.16 | 1.21 (1.15-1.26) | $1.74 \times 10^{-16}$ |
| rs56380368 | 2 | 43149681 | *THADA* | G | C | 0.15 | 1.14 (1.09-1.20) | $3.40 \times 10^{-8}$ |
| rs2231142 | 4 | 89052323 | *ABCG2* | T | G | 0.11 | 1.17 (1.11-1.23) | $4.10 \times 10^{-9}$ |
| rs10051765 | 5 | 176799992 | *SLC34A1* | C | T | 0.33 | 1.17 (1.13-1.21) | $7.30 \times 10^{-18}$ |
| rs74348938 | 6 | 160614649 | *SLC22A2* | A | G | 0.03 | 1.32 (1.2-1.46) | $1.40 \times 10^{-8}$ |
| rs13197227 | 6 | 30755144 | *FLOT1** | C | A | 0.26 | 1.11 (1.07-1.16) | $1.40 \times 10^{-8}$ |
| rs2299905 | 7 | 30957616 | *AQP1* | T | A | 0.29 | 1.12 (1.08-1.16) | $2.50 \times 10^{-9}$ |
| rs9594689 | 13 | 42737268 | *DGKH* | G | T | 0.37 | 1.10 (1.07-1.14) | $2.20 \times 10^{-8}$ |
| rs13329952 | 16 | 20366507 | *UMOD* | C | T | 0.19 | 1.15 (1.10-1.20) | $1.20 \times 10^{-10}$ |
| rs17216707 | 20 | 52732362 | *CYP24A1* | T | C | 0.81 | 1.16 (1.11-1.21) | $2.60 \times 10^{-11}$ |
| rs113882401 | 20 | 56086652 | *CTCFL** | T | C | 0.05 | 1.26 (1.16-1.37) | $2.50 \times 10^{-8}$ |
| rs219771 | 21 | 37835501 | *CLDN14* | C | T | 0.74 | 1.17 (1.13-1.21) | $1.50 \times 10^{-16}$ |
| rs13054904 | 22 | 23410918 | *GNAZ* | A | T | 0.26 | 1.16 (1.12-1.20) | $5.70 \times 10^{-15}$ |

Variants significantly associated with kidney stone disease from male sex genome-wide association study in the UK Biobank.
CHR=Chromosome; EA=Effect allele; EAF=Effect allele frequency; NEA=Non-effect allele; OR=Odds ratio; P=P-value; POS=Position based on NCBI Genome Build 37 (hg19); 95% CI=95% confidence interval. *denotes novel association.

***Table 3.3. Single nucleotide polymorphisms significantly associated with kidney stone disease from female sex genome-wide association study in the UK Biobank.***

| SNP | CHR | POS | Candidate Gene | EA | NEA | EAF | OR (95% CI) | P |
|---|---|---|---|---|---|---|---|---|
| rs16825455 | 1 | 21837755 | *ALPL* | G | A | 0.89 | 1.24 (1.15-1.34) | $7.60 \times 10^{-9}$ |
| rs4131290 | 5 | 176780682 | *SLC34A1* | C | A | 0.71 | 1.17 (1.11-1.24) | $1.40 \times 10^{08}$ |
| rs1155347 | 6 | 39146230 | *KCNK5* | C | T | 0.78 | 1.17 (1.11-1.24) | $2.40 \times 10^{-8}$ |
| rs1290425764 | 13 | 42736966 | *DGKH* | C | CT | 0.75 | 1.16 (1.10-1.22) | $1.20 \times 10^{-8}$ |
| rs6127099 | 20 | 52731402 | *CYP24A1* | A | T | 0.72 | 1.16 (1.11-1.21) | $3.60 \times 10^{-10}$ |
| rs12626330 | 21 | 37835982 | *CLDN14* | G | C | 0.51 | 1.17 (1.11-1.23) | $3.10 \times 10^{-9}$ |

Variants significantly associated with kidney stone disease from female sex genome-wide association study in the UK Biobank.
CHR=Chromosome; EA=Effect allele; EAF=Effect allele frequency; NEA=Non-effect allele; OR=Odds ratio; P=P-value; POS=Position based on NCBI Genome Build 37 (hg19); 95% CI=95% confidence interval. *denotes novel association.

All signals identified in sex-specific GWAS were directionally concordant in analyses of male and female participants, however only *ALPL, SLC34A1, DGKH, UMOD, CYP24A1,* and *CLDN14* reached replication threshold ($p<5.00\times10^{-5}$) in both sexes.

### 3.3.2 European ancestry GWAS meta-analysis

To improve the power of European ancestry KSD GWAS findings, I meta-analysed combined-sex KSD GWAS data from the UK Biobank and FinnGen R12. Meta-analysis of 24,167 cases (11,186 UK Biobank, 12,999 FinnGen) and 876,673 controls (390,488 UK Biobank, 486,185 FinnGen) revealed 79 independent variants significantly associated with KSD at 71 loci across 20 chromosomes (Figure 3.5, Figure 3.6, Table 3.4). Of these, 33 loci are not associated with KSD in published GWAS (*CASZ1, CLDN19, HORMAD1, RBKS, CYP1B1, COL7A1, WNT5AA, HEG1, ADRA2C, ISL1, PDE4D, FLOT1, HLA-DQA1, TFAP2B, TCF21, PRKAG2, TMEM252, PARD3, AMPD3, SIK2, PRICKLE1, PRKD1, AP4E1, PDE8A, MAP2K4, CDK12, ARHGAP27, ARL17B, PTGER1, ZNF28, MACROD2, NRIP1, H1-0*, and *CHADL*) and 54 were not identified in combined-sex KSD GWAS of the UK Biobank cohort alone (Section 3.3.1).

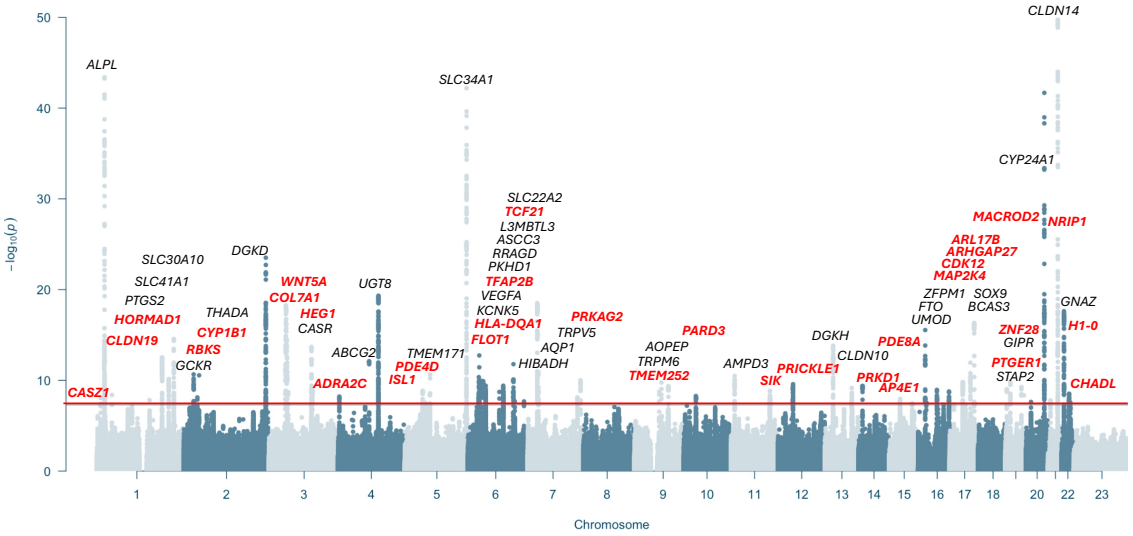



***Figure 3.5. Manhattan plot for a kidney stone disease genome-wide association study meta-analysis in the UK Biobank and FinnGen R12 Study.***

Genome-wide p-values (-log10) plotted against their respective positions on each chromosome. The horizontal red line indicates the genome-wide significance threshold of $5.00 \times 10^{-8}$. Loci are labelled with the primary candidate gene at each locus, as shown in Table 3.4. Novel loci are shown in red.

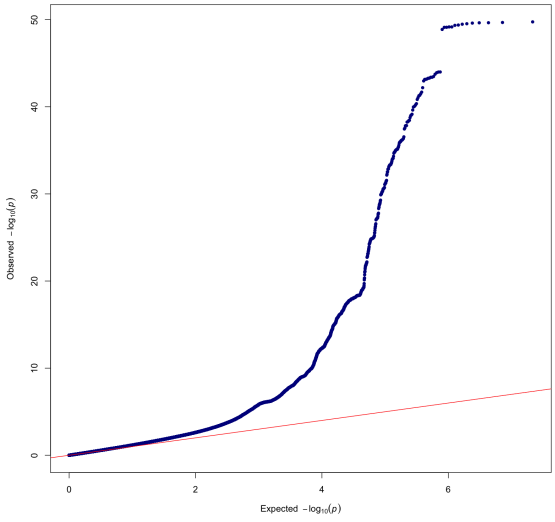



***Figure 3.6. Quantile-quantile plot of observed vs. expected p-values for a kidney stone disease genome-wide association study meta-analysis in the UK Biobank and FinnGen R12 Study.***

A meta-analysis of genome-wide association studies of kidney stone disease was performed for 24,167 individuals with kidney stone disease and 876,673 controls from the UK Biobank and FinnGen R12 study. $\lambda$ GC demonstrated some inflation (1.26), but the LDSC intercept of 1.05, with an attenuation ratio of 0.13 indicated that inflation was largely due to polygenicity and large sample size.

***Table 3.4. Variants significantly associated with kidney stone disease in a meta-analysis of UK Biobank and FinnGen R12 study.***

| Variant | | | | | | | Meta-analysis | | UK Biobank | | FinnGen | | Het | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **CHR** | **POS** | **Candidate gene** | **EA** | **NEA** | **EAF** | **OR (95% CI)** | **P** | **OR (95% CI)** | **P** | **OR (95% CI)** | **P** | $I^2$ | **P** |
| rs880315 | 1 | 10796866 | *CASZ1** | C | T | 0.38 | 1.06 (1.04-1.08) | $8.66\times10^{-9}$ | 1.07 (1.04-1.10) | $2.30\times10^{-6}$ | 1.05 (1.02-1.07) | $4.82\times10^{-4}$ | 29.4 | 0.23 |
| rs115239632 | 1 | 21826530 | *ALPL* | T | C | 0.05 | 1.33 (1.27-1.38) | $4.06\times10^{-44}$ | 1.33 (1.24-1.43) | $7.10\times10^{-16}$ | 1.32 (1.26-1.39) | $6.72\times10^{-30}$ | 0.0 | 0.9 |
| rs1256332 | 1 | 21893344 | *ALPL* | A | C | 0.14 | 1.15 (1.12-1.18) | $1.00\times10^{-24}$ | 1.16 (1.12-1.21) | $1.80\times10^{-16}$ | 1.10 (1.06-1.15) | $6.67\times10^{-7}$ | 73.6 | 0.05 |
| rs116804195 | 1 | 43201576 | *CLDN19** | T | C | 0.02 | 1.23 (1.15-1.31) | $4.07\times10^{-9}$ | 1.22 (1.09-1.35) | $3.60\times10^{-4}$ | 1.24 (1.13-1.35) | $2.88\times10^{-6}$ | 0.0 | 0.82 |
| rs77663766 | 1 | 150686011 | *HORMAD1** | A | C | 0.07 | 1.11 (1.07-1.14) | $1.53\times10^{-8}$ | 1.08 (1.02-1.15) | 0.01 | 1.12 (1.07-1.17) | $4.33\times10^{-7}$ | 0.0 | 0.4 |
| rs112851199 | 1 | 186621231 | *PTGS2* | C | T | 0.02 | 1.25 (1.18-1.33) | $3.02\times10^{-13}$ | 1.23 (1.12-1.34) | $6.60\times10^{-6}$ | 1.27 (1.17-1.38) | $8.23\times10^{-9}$ | 0.0 | 0.57 |
| rs2383517 | 1 | 186654633 | *PTGS2* | A | T | 0.89 | 1.10 (1.07-1.13) | $2.78\times10^{-10}$ | 1.06 (1.02-1.10) | $4.20\times10^{-3}$ | 1.27 (1.17-1.38) | $8.42\times10^{-9}$ | 0.0 | 0.53 |
| rs2383517 | 1 | 186654633 | *PTGS2* | A | T | 0.89 | 1.10 (1.07-1.13) | $2.78\times10^{-10}$ | 1.06 (1.02-1.10) | $4.20\times10^{-3}$ | 1.04 (0.99-1.09) | 0.09 | 0.0 | 0.53 |
| rs823101 | 1 | 205667006 | *SLC41A1* | C | T | 0.46 | 1.06 (1.04-1.08) | $7.28\times10^{-11}$ | 1.06 (1.03-1.09) | $3.40\times10^{-5}$ | 1.07 (1.04-1.09) | $4.58\times10^{-7}$ | 0.0 | 0.69 |
| rs884127 | 1 | 220082150 | *SLC30A10* | G | A | 0.59 | 1.08 (1.06-1.10) | $2.80\times10^{-15}$ | 1.08 (1.05-1.11) | $7.70\times10^{-8}$ | 1.08 (1.05-1.10) | $6.99\times10^{-9}$ | 0.0 | 0.89 |
| rs780093 | 2 | 27742603 | *GCKR* | T | C | 0.37 | 1.07 (1.05-1.09) | $2.19\times10^{-11}$ | 1.07 (1.05-1.10) | $2.40\times10^{-7}$ | 1.06 (1.03-1.09) | $1.54\times10^{-5}$ | 0.0 | 0.44 |
| rs10171517 | 2 | 28090059 | *RBKS** | C | T | 0.76 | 1.07 (1.05-1.09) | $1.03\times10^{-9}$ | 1.08 (1.04-1.11) | $4.00\times10^{-6}$ | 1.06 (1.03-1.09) | $4.77\times10^{-5}$ | 0.0 | 0.46 |

| Variant | | | | | | | Meta-analysis | | UK Biobank | | FinnGen | | Het | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **CHR** | **POS** | **Candidate gene** | **EA** | **NEA** | **EAF** | **OR (95% CI)** | **P** | **OR (95% CI)** | **P** | **OR (95% CI)** | **P** | $I^2$ | **P** |
| rs232573 | 2 | 38388481 | *CYP1B1** | A | G | 0.48 | 1.05 (1.04-1.07) | $7.63\times10^{-9}$ | 1.05 (1.02-1.08) | $4.80\times10^{-4}$ | 1.06 (1.03-1.09) | $3.46\times10^{-6}$ | 0.0 | 0.53 |
| rs1430083 | 2 | 43448479 | *THADA* | A | T | 0.76 | 1.08 (1.06-1.11) | $2.79\times10^{-11}$ | 1.10 (1.06-1.15) | $8.30\times10^{-7}$ | 1.07 (1.04-1.10) | $2.96\times10^{-6}$ | 44.7 | 0.18 |
| rs838717 | 2 | 234296444 | *DGKD* | G | A | 0.42 | 1.10 (1.08-1.12) | $3.08\times10^{-24}$ | 1.10 (1.07-1.13) | $2.60\times10^{-11}$ | 1.10 (1.08-1.13) | $1.71\times10^{-14}$ | 0.0 | 0.7 |
| rs74780677 | 3 | 48601774 | *COL7A1** | A | G | 0.94 | 1.23 (1.18-1.29) | $5.58\times10^{-19}$ | 1.14 (1.02-1.27) | 0.02 | 1.26 (1.19-1.32) | $2.04\times10^{-18}$ | 62.7 | 0.1 |
| rs34594364 | 3 | 55461778 | *WNT5AA** | AT | A | 0.79 | 1.10 (1.07-1.14) | $2.06\times10^{-9}$ | - | - | 1.10 (1.07-1.14) | $2.06\times10^{-9}$ | 0.0 | 1.00 |
| rs13059018 | 3 | 55473083 | *WNT5AA** | G | C | 0.91 | 1.11 (1.07-1.15) | $3.38\times10^{-9}$ | 1.08 (1.02-1.14) | 0.01 | 1.13 (1.08-1.17) | $5.03\times10^{-8}$ | 27.1 | 0.24 |
| rs34172859 | 3 | 121942713 | *CASR* | G | GA | 0.28 | 1.08 (1.06-1.10) | $2.03\times10^{-14}$ | 1.09 (1.06-1.12) | $6.20\times10^{-8}$ | 1.08 (1.05-1.11) | $5.72\times10^{-8}$ | 0.0 | 0.62 |
| rs61171818 | 3 | 124734646 | *HEG1** | A | C | 0.54 | 1.06 (1.04-1.08) | $3.68\times10^{-10}$ | 1.06 (1.03-1.09) | $2.00\times10^{-5}$ | 1.06 (1.03-1.09) | $6.38\times10^{-6}$ | 0.0 | 0.98 |
| rs369200 | 4 | 3743004 | *ADRA2C** | T | C | 0.45 | 1.06 (1.04-1.08) | $6.14\times10^{-9}$ | 1.04 (1.01-1.07) | $4.70\times10^{-3}$ | 1.07 (1.04-1.10) | $1.04\times10^{-7}$ | 59.7 | 0.12 |
| rs2231142 | 4 | 89052323 | *ABCG2* | T | G | 0.1 | 1.12 (1.09-1.15) | $7.79\times10^{-13}$ | 1.12 (1.08-1.17) | $6.20\times10^{-8}$ | 1.12 (1.07-1.17) | $2.62\times10^{-6}$ | 0.0 | 0.86 |
| rs10011293 | 4 | 115577381 | *UGT8* | T | C | 0.23 | 1.11 (1.08-1.13) | $4.73\times10^{-20}$ | 1.10 (1.06-1.14) | $1.30\times10^{-7}$ | 1.11 (1.08-1.14) | $5.16\times10^{-14}$ | 0.0 | 0.54 |
| rs66512073 | 5 | 51163093 | *ISL1** | T | C | 0.37 | 1.06 (1.04-1.08) | $1.44\times10^{-9}$ | 1.04 (1.02-1.07) | $2.00\times10^{-3}$ | 1.07 (1.05-1.10) | $6.49\times10^{-8}$ | 52.5 | 0.15 |
| rs11959132 | 5 | 58605916 | *PDE4D** | G | C | 0.75 | 1.06 (1.04-1.08) | $2.98\times10^{-8}$ | 1.06 (1.02-1.09) | $5.80\times10^{-4}$ | 1.07 (1.04-1.10) | $1.27\times10^{-5}$ | 0.0 | 0.68 |

| Variant | | | | | | | Meta-analysis | | UK Biobank | | FinnGen | | Het | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **CHR** | **POS** | **Candidate gene** | **EA** | **NEA** | **EAF** | **OR (95% CI)** | **P** | **OR (95% CI)** | **P** | **OR (95% CI)** | **P** | $I^2$ | **P** |
| rs638333 | 5 | 72419267 | *TMEM171* | C | T | 0.27 | 1.08 (1.05-1.10) | $2.92\times10^{-12}$ | 1.06 (1.03-1.09) | $1.50\times10^{-4}$ | 1.09 (1.06-1.12) | $1.77\times10^{-9}$ | 47 | 0.17 |
| rs10051765 | 5 | 176799992 | *SLC34A1* | C | T | 0.37 | 1.14 (1.12-1.16) | $6.45\times10^{-43}$ | 1.16 (1.13-1.19) | $1.30\times10^{-24}$ | 1.13 (1.10-1.15) | $1.94\times10^{-20}$ | 53.6 | 0.14 |
| rs9461633 | 6 | 30761168 | *FLOT1** | G | A | 0.18 | 1.11 (1.07-1.15) | $2.89\times10^{-9}$ | 1.11 (1.07-1.15) | $2.90\times10^{-9}$ | - | - | 0.0 | 1.00 |
| rs9271375 | 6 | 32587067 | *HLA-DQA1** | G | A | 0.56 | 1.07 (1.05-1.09) | $1.82\times10^{-13}$ | 1.08 (1.05-1.11) | $6.40\times10^{-8}$ | 1.07 (1.04-1.09) | $5.26\times10^{-7}$ | 0.0 | 0.64 |
| rs12190363 | 6 | 39180487 | *KCNK5* | A | G | 0.27 | 1.07 (1.05-1.09) | $1.20\times10^{-10}$ | 1.09 (1.05-1.12) | $1.40\times10^{-7}$ | 1.06 (1.03-1.08) | $7.48\times10^{-5}$ | 48.9 | 0.16 |
| rs68137036 | 6 | 43820215 | *VEGFA* | G | A | 0.3 | 1.07 (1.05-1.09) | $1.24\times10^{-10}$ | 1.06 (1.03-1.09) | $4.60\times10^{-5}$ | 1.07 (1.04-1.10) | $6.04\times10^{-7}$ | 0.0 | 0.73 |
| rs2206271 | 6 | 50786008 | *TFAP2B** | A | T | 0.32 | 1.06 (1.04-1.08) | $3.95\times10^{-10}$ | 1.05 (1.02-1.08) | $9.50\times10^{-4}$ | 1.07 (1.05-1.10) | $5.61\times10^{-8}$ | 21.9 | 0.26 |
| rs2465043 | 6 | 51180765 | *PKHD1* | G | A | 0.63 | 1.06 (1.04-1.08) | $2.13\times10^{-8}$ | 1.04 (1.02-1.07) | $2.20\times10^{-3}$ | 1.07 (1.04-1.09) | $1.62\times10^{-6}$ | 0.2 | 0.32 |
| rs148684631 | 6 | 90121976 | *RRAGD* | G | GGA GA | 0.68 | 1.06 (1.04-1.08) | $1.90\times10^{-9}$ | 1.08 (1.05-1.11) | $4.40\times10^{-7}$ | 1.05 (1.02-1.08) | $3.75\times10^{-4}$ | 51.9 | 0.15 |
| rs6928418 | 6 | 101175347 | *ASCC3* | T | C | 0.55 | 1.06 (1.04-1.08) | $3.92\times10^{-10}$ | 1.06 (1.04-1.09) | $4.40\times10^{-6}$ | 1.06 (1.03-1.08) | $1.94\times10^{-5}$ | 0.0 | 0.67 |
| rs7740107 | 6 | 130374461 | *L3MBTL3* | A | T | 0.74 | 1.08 (1.06-1.10) | $1.65\times10^{-12}$ | 1.07 (1.04-1.10) | $1.20\times10^{-5}$ | 1.08 (1.05-1.12) | $2.36\times10^{-8}$ | 0.0 | 0.51 |
| rs12190287 | 6 | 134214525 | *TCF21** | C | G | 0.59 | 1.05 (1.03-1.07) | $2.12\times10^{-8}$ | 1.07 (1.04-1.10) | $3.40\times10^{-6}$ | 1.04 (1.02-1.07) | $8.07\times10^{-4}$ | 27 | 0.24 |
| rs74495751 | 6 | 160624947 | *SLC22A2* | C | G | 0.03 | 1.24 (1.15-1.33) | $2.11\times10^{-8}$ | 1.24 (1.15-1.33) | $2.10\times10^{-8}$ | - | - | 0.0 | 1.00 |

| Variant | | | | | | | Meta-analysis | | UK Biobank | | FinnGen | | Het | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **CHR** | **POS** | **Candidate gene** | **EA** | **NEA** | **EAF** | **OR (95% CI)** | **P** | **OR (95% CI)** | **P** | **OR (95% CI)** | **P** | $I^2$ | **P** |
| rs1404278 | 7 | 27634726 | *HIBADH* | T | C | 0.31 | 1.09 (1.07-1.12) | $3.12\times10^{-19}$ | 1.11 (1.07-1.14) | $1.00\times10^{-11}$ | 1.08 (1.06-1.11) | $3.19\times10^{-9}$ | 0.0 | 0.32 |
| rs1004317 | 7 | 30956858 | *AQP1* | G | A | 0.39 | 1.07 (1.05-1.09) | $1.16\times10^{-11}$ | 1.06 (1.04-1.09) | $8.00\times10^{-6}$ | 1.07 (1.04-1.10) | $3.15\times10^{-7}$ | 0.0 | 0.81 |
| rs4252512 | 7 | 142605221 | *TRPV5* | C | T | 0.02 | 1.34 (1.22-1.49) | $7.68\times10^{-9}$ | 1.34 (1.22-1.49) | $7.70\times10^{-9}$ | - | - | 0.0 | 1.00 |
| rs6464165 | 7 | 151413124 | *PRKAG2** | T | C | 0.74 | 1.07 (1.05-1.09) | $9.80\times10^{-11}$ | 1.08 (1.05-1.11) | $5.10\times10^{-7}$ | 1.06 (1.03-1.10) | $3.77\times10^{-5}$ | 0.0 | 0.55 |
| rs4745107 | 9 | 71177110 | *TMEM252** | A | G | 0.73 | 1.06 (1.04-1.09) | $2.57\times10^{-9}$ | 1.05 (1.02-1.08) | $8.30\times10^{-4}$ | 1.08 (1.05-1.11) | $2.89\times10^{-7}$ | 27.3 | 0.24 |
| rs148349564 | 9 | 77466700 | *TRPM6* | C | T | 0.07 | 1.12 (1.08-1.16) | $1.74\times10^{-10}$ | 1.14 (1.08-1.20) | $2.80\times10^{-7}$ | 1.11 (1.05-1.16) | $1.07\times10^{-4}$ | 0.0 | 0.42 |
| rs143845824 | 9 | 97474225 | *AOPEP* | C | CA | 0.91 | 1.11 (1.08-1.15) | $4.36\times10^{-10}$ | 1.10 (1.04-1.16) | $1.00\times10^{-3}$ | 1.12 (1.08-1.17) | $9.08\times10^{-8}$ | 0.0 | 0.52 |
| rs2477001 | 10 | 34962926 | *PARD3** | C | A | 0.28 | 1.06 (1.04-1.08) | $5.43\times10^{-9}$ | 1.06 (1.03-1.09) | $2.30\times10^{-4}$ | 1.07 (1.04-1.10) | $5.16\times10^{-6}$ | 0.0 | 0.56 |
| rs61264398 | 11 | 10474497 | *AMPD3* | G | T | 0.56 | 1.06 (1.04-1.08) | $3.30\times10^{-11}$ | 1.07 (1.04-1.10) | $5.60\times10^{-7}$ | 1.06 (1.03-1.09) | $1.10\times10^{-5}$ | 0.0 | 0.53 |
| rs10789841 | 11 | 111469582 | *SIK2** | T | C | 0.75 | 1.07 (1.04-1.09) | $1.48\times10^{-9}$ | 1.08 (1.05-1.11) | $4.80\times10^{-7}$ | 1.05 (1.02-1.09) | $4.17\times10^{-4}$ | 19.2 | 0.27 |
| rs7133836 | 12 | 42861098 | *PRICKLE1** | C | T | 0.43 | 1.06 (1.04-1.08) | $2.66\times10^{-10}$ | 1.05 (1.02-1.08) | $4.10\times10^{-4}$ | 1.07 (1.04-1.10) | $9.40\times10^{-8}$ | 8.2 | 0.30 |
| rs73187207 | 13 | 42619919 | *DGKH* | T | G | 0.21 | 1.07 (1.05-1.10) | $8.72\times10^{-10}$ | 1.08 (1.04-1.11) | $6.50\times10^{-6}$ | 1.07 (1.04-1.11) | $4.11\times10^{-6}$ | 0.0 | 0.92 |
| rs12585865 | 13 | 42751707 | *DGKH* | C | T | 0.4 | 1.08 (1.06-1.10) | $1.51\times10^{-14}$ | 1.10 (1.07-1.13) | $2.70\times10^{-11}$ | 1.06 (1.03-1.08) | $1.65\times10^{-5}$ | 74.3 | 0.05 |

| Variant | | | | | | | Meta-analysis | | UK Biobank | | FinnGen | | Het | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **CHR** | **POS** | **Candidate gene** | **EA** | **NEA** | **EAF** | **OR (95% CI)** | **P** | **OR (95% CI)** | **P** | **OR (95% CI)** | **P** | $I^2$ | **P** |
| rs57719175 | 13 | 96175396 | *CLDN10* | G | A | 0.6 | 1.06 (1.04-1.08) | $7.05\times10^{-10}$ | 1.05 (1.02-1.08) | $7.70\times10^{-4}$ | 1.07 (1.04-1.10) | $1.13\times10^{-7}$ | 31.1 | 0.23 |
| rs11851035 | 14 | 30678761 | *PRKD1** | G | A | 0.85 | 1.09 (1.06-1.12) | $4.11\times10^{-10}$ | 1.12 (1.07-1.17) | $1.50\times10^{-7}$ | 1.07 (1.03-1.10) | $1.60\times10^{-4}$ | 64.2 | 0.09 |
| rs7181593 | 15 | 51146114 | *AP4E1** | C | T | 0.29 | 1.06 (1.04-1.08) | $1.14\times10^{-8}$ | 1.07 (1.04-1.10) | $1.30\times10^{-5}$ | 1.05 (1.02-1.08) | $1.82\times10^{-4}$ | 0.0 | 0.49 |
| rs11629962 | 15 | 85530028 | *PDE8A** | T | C | 0.26 | 1.06 (1.04-1.08) | $3.47\times10^{-8}$ | 1.07 (1.04-1.11) | $7.60\times10^{-6}$ | 1.05 (1.02-1.08) | $6.84\times10^{-4}$ | 13.0 | 0.28 |
| rs12921916 | 16 | 20407196 | *UMOD* | C | T | 0.28 | 1.09 (1.07-1.11) | $2.88\times10^{-16}$ | 1.10 (1.07-1.13) | $2.60\times10^{-10}$ | 1.08 (1.05-1.11) | $1.30\times10^{-7}$ | 0.0 | 0.34 |
| rs889299 | 16 | 23381914 | *SCNN1B* | G | A | 0.78 | 1.07 (1.04-1.09) | $2.84\times10^{-8}$ | 1.06 (1.02-1.09) | $6.00\times10^{-4}$ | 1.07 (1.04-1.11) | $9.67\times10^{-6}$ | 0.0 | 0.46 |
| rs11642015 | 16 | 53802494 | *FTO* | T | C | 0.41 | 1.06 (1.04-1.08) | $1.13\times10^{-9}$ | 1.05 (1.02-1.08) | $8.70\times10^{-4}$ | 1.07 (1.04-1.10) | $1.84\times10^{-7}$ | 17.2 | 0.27 |
| rs4782364 | 16 | 88516135 | *ZFPM1* | G | C | 0.41 | 1.06 (1.04-1.08) | $1.70\times10^{-9}$ | 1.04 (1.02-1.07) | $1.50\times10^{-3}$ | 1.07 (1.04-1.10) | $1.19\times10^{-7}$ | 43.6 | 0.18 |
| rs2240554 | 17 | 12264477 | *MAP2K4** | T | G | 0.54 | 1.05 (1.03-1.07) | $4.66\times10^{-8}$ | 1.05 (1.02-1.08) | $4.60\times10^{-4}$ | 1.05 (1.03-1.08) | $2.65\times10^{-5}$ | 0.0 | 0.77 |
| rs11868029 | 17 | 37650304 | *CDK12** | A | G | 0.22 | 1.07 (1.05-1.10) | $1.54\times10^{-10}$ | 1.08 (1.05-1.12) | $3.40\times10^{-7}$ | 1.07 (1.03-1.10) | $8.54\times10^{-5}$ | 0.0 | 0.49 |
| rs12150247 | 17 | 43480759 | *ARHGAP27** | C | G | 0.91 | 1.10 (1.06-1.13) | $2.88\times10^{-8}$ | 1.04 (0.99-1.09) | $1.40\times10^{-1}$ | 1.04 (1-1.09) | 0.08 | 0.0 | 0.93 |
| rs111443054 | 17 | 43489161 | *ARHGAP27** | TCAC ACA | T | 0.15 | 1.09 (1.06-1.12) | $1.46\times10^{-8}$ | 1.06 (1.03-1.10) | $5.40\times10^{-4}$ | 1.11 (1.05-1.17) | $2.00\times10^{-4}$ | 37.4 | 0.21 |
| rs2532340 | 17 | 44345653 | *ARL17B** | G | A | 0.18 | 1.08 (1.05-1.10) | $4.92\times10^{-8}$ | 1.07 (1.04-1.11) | $3.40\times10^{-5}$ | 1.08 (1.04-1.13) | $3.74\times10^{-4}$ | 0.0 | 0.67 |

| Variant | | | | | | | Meta-analysis | | UK Biobank | | FinnGen | | Het | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **CHR** | **POS** | **Candidate gene** | **EA** | **NEA** | **EAF** | **OR (95% CI)** | **P** | **OR (95% CI)** | **P** | **OR (95% CI)** | **P** | $I^2$ | **P** |
| rs117518238 | 17 | 59462062 | *BCAS3* | C | T | 0.92 | 1.19 (1.14-1.25) | $1.12\times10^{-12}$ | - | - | 1.19 (1.14-1.25) | $1.12\times10^{-12}$ | 0.0 | 1.00 |
| rs1058004 | 17 | 59486732 | *BCAS3* | T | C | 0.73 | 1.07 (1.05-1.09) | $5.32\times10^{-11}$ | 1.06 (1.02-1.09) | $4.50\times10^{-4}$ | 1.09 (1.06-1.12) | $9.57\times10^{-9}$ | 44.6 | 0.18 |
| rs9902482 | 17 | 70356993 | *SOX9* | C | T | 0.48 | 1.08 (1.06-1.10) | $4.76\times10^{-17}$ | 1.08 (1.05-1.11) | $1.10\times10^{-8}$ | 1.08 (1.05-1.11) | $8.16\times10^{-10}$ | 0.0 | 0.99 |
| rs11669940 | 19 | 4336379 | *STAP2* | C | G | 0.68 | 1.06 (1.04-1.08) | $9.56\times10^{-10}$ | 1.07 (1.04-1.10) | $2.60\times10^{-6}$ | 1.06 (1.03-1.08) | $6.41\times10^{-5}$ | 0.0 | 0.43 |
| rs3760702 | 19 | 14588237 | *PTGER1** | A | G | 0.29 | 1.07 (1.05-1.09) | $1.57\times10^{-10}$ | 1.08 (1.05-1.11) | $4.10\times10^{-7}$ | 1.06 (1.03-1.09) | $6.78\times10^{-5}$ | 0.0 | 0.45 |
| rs11672660 | 19 | 46180184 | *GIPR* | C | T | 0.77 | 1.07 (1.05-1.09) | $5.51\times10^{-10}$ | 1.07 (1.03-1.10) | $1.40\times10^{-4}$ | 1.07 (1.04-1.11) | $9.42\times10^{-7}$ | 0.0 | 0.77 |
| rs7259073 | 19 | 53357172 | *ZNF28** | T | C | 0.91 | 1.11 (1.07-1.15) | $4.58\times10^{-9}$ | 1.10 (1.04-1.16) | $7.40\times10^{-4}$ | 1.11 (1.07-1.16) | $1.48\times10^{-6}$ | 0.0 | 0.66 |
| rs1932940 | 20 | 14604464 | *MACROD2** | T | G | 0.63 | 1.06 (1.04-1.08) | $2.40\times10^{-8}$ | 1.07 (1.04-1.10) | $1.40\times10^{-6}$ | 1.04 (1.02-1.07) | $1.67\times10^{-3}$ | 49.7 | 0.16 |
| rs6127099 | 20 | 52731402 | *CYP24A1* | A | T | 0.73 | 1.16 (1.13-1.18) | $2.06\times10^{-42}$ | 1.14 (1.10-1.17) | $1.60\times10^{-16}$ | 1.17 (1.14-1.21) | $4.22\times10^{-28}$ | 60.3 | 0.11 |
| rs2585442 | 20 | 52737123 | *CYP24A1* | G | C | 0.29 | 1.08 (1.06-1.10) | $1.54\times10^{-13}$ | 1.11 (1.08-1.14) | $7.70\times10^{-11}$ | 1.12 (1.09-1.15) | $8.81\times10^{-18}$ | 0.0 | 0.56 |
| rs2823254 | 21 | 16783083 | *NRIP1** | A | T | 0.73 | 1.06 (1.04-1.08) | $1.65\times10^{-8}$ | 1.05 (1.02-1.09) | $3.00\times10^{-4}$ | 1.07 (1.04-1.10) | $1.20\times10^{-5}$ | 0.0 | 0.56 |
| rs128494 | 21 | 37834258 | *CLDN14* | C | T | 0.69 | 1.07 (1.05-1.10) | $5.66\times10^{-12}$ | 1.02 (0.99-1.06) | 0.16 | 1.02 (0.99-1.05) | 0.17 | 0.0 | 0.84 |
| rs219772 | 21 | 37835347 | *CLDN14* | A | T | 0.76 | 1.18 (1.15-1.21) | $1.79\times10^{-50}$ | 1.17 (1.13-1.21) | $2.50\times10^{-24}$ | 1.19 (1.15-1.23) | $5.55\times10^{-28}$ | 0.0 | 0.42 |

| Variant | | | | | | | Meta-analysis | | UK Biobank | | FinnGen | | Het | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **CHR** | **POS** | **Candidate gene** | **EA** | **NEA** | **EAF** | **OR (95% CI)** | **P** | **OR (95% CI)** | **P** | **OR (95% CI)** | **P** | $I^2$ | **P** |
| rs5996460 | 22 | 23393090 | *GNAZ* | A | G | 0.25 | 1.10 (1.08-1.12) | $2.42\times10^{-18}$ | 1.12 (1.09-1.15) | $1.10\times10^{-13}$ | 1.08 (1.05-1.11) | $9.85\times10^{-7}$ | 64.2 | 0.09 |
| rs6000889 | 22 | 38184851 | *H1-0** | C | T | 0.37 | 1.06 (1.04-1.08) | $2.99\times10^{-9}$ | 1.04 (1.01-1.07) | 0.01 | 1.08 (1.05-1.10) | $1.92\times10^{-8}$ | 70.7 | 0.06 |
| rs139495 | 22 | 41635425 | *CHADL** | C | T | 0.71 | 1.06 (1.04-1.08) | $2.85\times10^{-8}$ | 1.07 (1.04-1.11) | $2.30\times10^{-6}$ | 1.05 (1.02-1.07) | $1.44\times10^{-3}$ | 41.2 | 0.19 |

CHR=Chromosome; EA=Effect allele; EAF=Effect allele frequency; NEA=Non-effect allele; OR=Odds ratio; P=P-value; POS=Position based on NCBI Genome Build 37 (hg19); 95% CI=95% confidence interval. *denotes novel association.

All significant, independent signals in the European-ancestry GWAS meta-analysis were directionally concordant in UK Biobank and FinnGen GWAS (Table 3.4). Three coding-region variants associated with KSD were identified: 1) rs116804195 is a missense variant in *CLDN19* (C>T, p.R200Q). The presence of the T allele, the risk allele for KSD (OR=1.23, 95% CI=1.15-1.31, p=$4.07 \times 10^{-9}$), results in arginine replacing glutamine at position 200 of the claudin-19 protein. Despite this alteration, it is predicted to be tolerated/benign by both Sorting Intolerant From Tolerant (SIFT) and PolyPhen algorithms, with a scaled CADD score of 23 and a Rare Exome Variant Ensemble Learner (REVEL) score of 0.32; 2) rs2231142 (G>T, p.Q141K) is a missense variant located in exon 5 of *ABCG2*. The T allele, the KSD risk allele (OR=1.12, 95% CI=1.09-1.15, p=$7.79 \times 10^{-13}$), leads to a glutamine-to-lysine substitution. Like the previous variant, it is predicted to be tolerated/benign by SIFT and PolyPhen, with a scaled CADD score of 22 and a REVEL score of 0.11; 3) rs638333 is a missense variant in *TMEM171* (T>C, p.F23L). The C allele, conferring increased KSD risk (OR=1.08, 95% CI=1.05-1.10, p=$2.92 \times 10^{-12}$), causes a phenylalanine-to-leucine substitution at position 23 of the transmembrane protein 171. Unlike the previous variants, it is predicted to be deleterious/probably damaging by SIFT and PolyPhen, though has a scaled CADD score of 24 and a REVEL score of 0.12.

### 3.3.3 Trans-ancestry GWAS meta-analysis

To improve the power and generalisability of the KSD GWAS described in Sections 3.3.1 and 3.3.2, combined-sex KSD GWAS data from the UK Biobank, FinnGen R12, and Biobank Japan were meta-analysed. Meta-analysis of 29,772 cases (11,186 UK Biobank, 12,999 FinnGen, 5,587 Biobank Japan) and 905,543 controls (390,488 UK Biobank, 486,185 FinnGen, 28,870 Biobank Japan) revealed 95 independent variants significantly associated with KSD at 79 loci across 21 chromosomes (Figure 3.7, Figure 3.8, Table 3.5). Of these, 38 are not associated with KSD in published GWAS (*CASZ1, CLDN19, PKN2, SDC1, CYP1B1, COL7A1, WNT5A, ADRA2C, SHROOM3, ISL1, PDE4D, HLA-DQA1, TFAP2B, TCF21, PRKAG2, ZNF703, TMEM252, PARD3, SIK2, KLF12, PRKD1, SPPL2A, TLN2,*

*SCNN1B, SLC12A3, ZFHX3, MAP2K4, CDK12, ARHGAP27, ARL17B, FAM20A, PTGER1, ZNF28, CYP24A1, NRIP1, CLDN14, H1-0*, and *CHADL*) and 16 were not identified in the European ancestry meta-analysis described in Section 3.3.2.

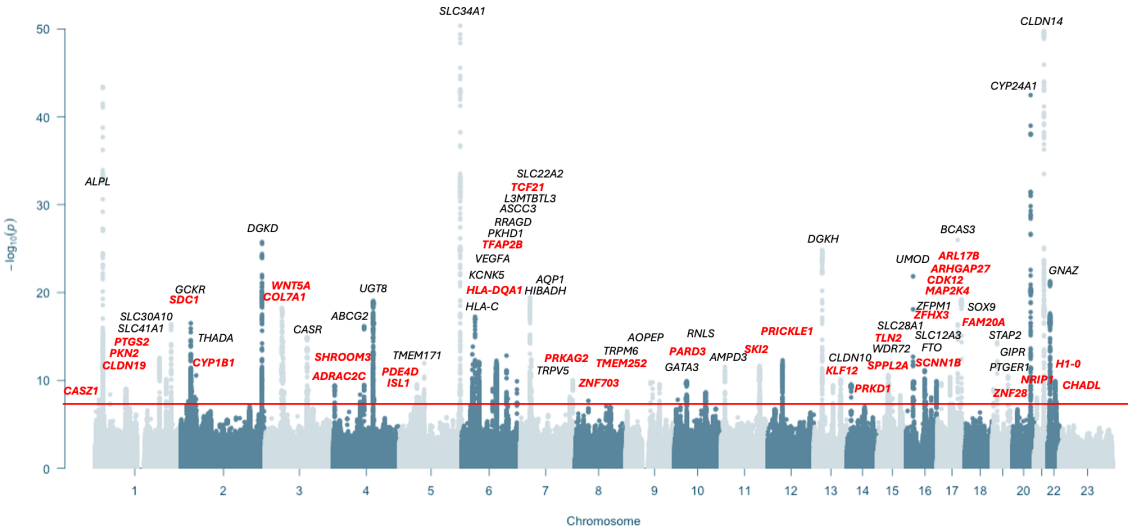



***Figure 3.7. Manhattan plot for a kidney stone disease genome-wide association study meta-analysis in the UK Biobank, FinnGen R12 Study, and Biobank Japan.***

Genome-wide p-values (-log10) plotted against their respective positions on each chromosome. The horizontal red line indicates the genome-wide significance threshold of $5.00 \times 10^{-8}$. Loci are labelled with the primary candidate gene at each locus, as shown in Table 3.5. Novel loci are shown in red.

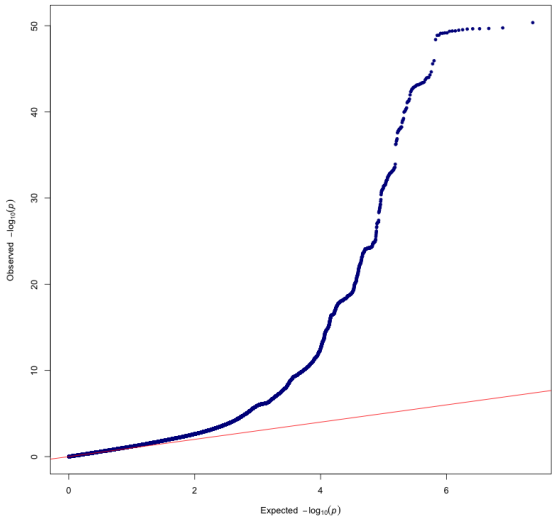



***Figure 3.8. Quantile-quantile plot of observed vs. expected p-values for a kidney stone disease genome-wide association study meta-analysis in the UK Biobank, FinnGen R12 Study, and Biobank Japan.***

A meta-analysis of genome-wide association studies of kidney stone disease was performed for 29, 754 individuals with kidney stone disease and 905, 543 controls from the UK Biobank and FinnGen R12 study. $\lambda$ GC demonstrated some inflation (1.29), but the LDSC intercept of 1.07, with an attenuation ratio of 0.17 indicated that inflation was largely due to polygenicity and large sample size.

***Table 3.5. Variants significantly associated with kidney stone disease in a meta-analysis of UK Biobank, FinnGen R12 study, and Biobank Japan.***

| Variant | | | | | | | Meta-analysis | | UK Biobank | | FinnGen | | Biobank Japan | | Het | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **CHR** | **POS** | **Candidate gene** | **EA** | **NEA** | **EAF** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | $I^2$ | **P** |
| rs880315 | 1 | 10796866 | *CASZ1** | C | T | 0.43 | 1.05 (1.03-1.07) | $1.48\times10^{-8}$ | 1.07 (1.04-1.10) | $2.30\times10^{-6}$ | 1.05 (1.02-1.07) | $4.82\times10^{-4}$ | 1.02 (0.98-1.07) | 0.38 | 39.3 | 0.19 |
| rs72659148 | 1 | 21817397 | *ALPL* | A | G | 0.14 | 1.08 (1.05-1.12) | $4.52\times10^{-8}$ | 1.07 (1.02-1.13) | $8.20\times10^{-3}$ | 1.02 (0.97-1.07) | $5.11\times10^{-1}$ | 1.11 (1.06-1.17) | $3.59\times10^{-5}$ | 67.2 | 0.05 |
| rs115239632 | 1 | 21826530 | *ALPL* | T | C | 0.05 | 1.33 (1.27-1.38) | $4.06\times10^{-44}$ | 1.33 (1.24-1.43) | $7.10\times10^{-16}$ | 1.32 (1.26-1.39) | $6.72\times10^{-30}$ | - | - | 0.0 | 0.90 |
| rs6656758 | 1 | 21897961 | *ALPL* | T | C | 0.21 | 1.10 (1.08-1.13) | $3.66\times10^{-19}$ | 1.11 (1.08-1.15) | $6.00\times10^{-11}$ | 1.07 (1.03-1.10) | $9.08\times10^{-5}$ | 1.09 (1.03-1.16) | $4.80\times10^{-3}$ | 45.0 | 0.16 |
| rs116804195 | 1 | 43201576 | *CLDN19** | T | C | 0.02 | 1.23 (1.15-1.31) | $4.07\times10^{-9}$ | 1.22 (1.09-1.35) | $3.60\times10^{-4}$ | 1.24 (1.13-1.35) | $2.88\times10^{-6}$ | - | - | 0.0 | 0.82 |
| rs6681274 | 1 | 89125905 | *PKN2** | G | A | 0.42 | 1.05 (1.04-1.07) | $9.18\times10^{-10}$ | 1.05 (1.03-1.08) | $1.20\times10^{-4}$ | 1.04 (1.01-1.07) | $2.73\times10^{-3}$ | 1.10 (1.05-1.15) | $1.51\times10^{-5}$ | 56.6 | 0.10 |
| rs112851199 | 1 | 186621231 | *PTGS2* | C | T | 0.02 | 1.25 (1.18-1.33) | $3.02\times10^{-13}$ | 1.23 (1.12-1.34) | $6.60\times10^{-6}$ | 1.27 (1.17-1.38) | $8.23\times10^{-9}$ | - | - | 0.0 | 0.57 |
| rs2383517 | 1 | 186654633 | *PTGS2* | A | T | 0.89 | 1.10 (1.07-1.13) | $2.78\times10^{-10}$ | 1.06 (1.02-1.10) | $4.20\times10^{-3}$ | 1.04 (0.99-1.09) | 0.09 | 1.08 (0.98-1.20) | 0.12 | 0.0 | 0.53 |
| rs823101 | 1 | 205667006 | *SLC41A1* | C | T | 0.46 | 1.06 (1.04-1.08) | $7.28\times10^{-11}$ | 1.06 (1.03-1.09) | $3.40\times10^{-5}$ | 1.07 (1.04-1.09) | $4.58\times10^{-7}$ | - | - | 0.0 | 0.69 |
| rs884127 | 1 | 220082150 | *SLC30A10* | G | A | 0.58 | 1.08 (1.06-1.09) | $4.26\times10^{-17}$ | 1.08 (1.05-1.11) | $7.70\times10^{-8}$ | 1.08 (1.05-1.10) | $6.99\times10^{-9}$ | 1.07 (1.02-1.11) | $3.68\times10^{-3}$ | 0.0 | 0.89 |
| rs11686716 | 2 | 20349985 | *SDC1** | G | C | 0.34 | 1.05 (1.03-1.07) | $2.47\times10^{-8}$ | 1.03 (1-1.06) | 0.02 | 1.06 (1.03-1.09) | $7.34\times10^{-6}$ | 1.13 (1.06-1.22) | $5.02\times10^{-4}$ | 69.3 | 0.04 |

| Variant | | | | | | | Meta-analysis | | UK Biobank | | FinnGen | | Biobank Japan | | Het | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **CHR** | **POS** | **Candidate gene** | **EA** | **NEA** | **EAF** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | $I^2$ | **P** |
| rs780093 | 2 | 27742603 | *GCKR* | T | C | 0.4 | 1.08 (1.06-1.10) | $3.15\times10^{-17}$ | 1.07 (1.05-1.10) | $2.40\times10^{-7}$ | 1.06 (1.03-1.09) | $1.54\times10^{-5}$ | 1.14 (1.09-1.19) | $1.10\times10^{-8}$ | 73.2 | 0.02 |
| rs232573 | 2 | 38388481 | *CYP1B1** | A | G | 0.46 | 1.05 (1.03-1.07) | $7.72\times10^{-9}$ | 1.05 (1.02-1.08) | $4.80\times10^{-4}$ | 1.06 (1.03-1.09) | $3.46\times10^{-6}$ | 1.03 (0.96-1.10) | 0.45 | 0.0 | 0.62 |
| rs56380368 | 2 | 43149681 | *THADA* | G | C | 0.12 | 1.08 (1.05-1.11) | $8.21\times10^{-9}$ | 1.08 (1.04-1.13) | $4.00\times10^{-5}$ | 1.07 (1.02-1.11) | $1.68\times10^{-3}$ | 1.12 (1.03-1.23) | 0.01 | 0.0 | 0.58 |
| rs1430083 | 2 | 43448479 | *THADA* | A | T | 0.76 | 1.08 (1.06-1.11) | $2.79\times10^{-11}$ | 1.10 (1.06-1.15) | $8.30\times10^{-7}$ | 1.07 (1.04-1.10) | $2.96\times10^{-6}$ | - | - | 44.7 | 0.18 |
| rs838717 | 2 | 234296444 | *DGKD* | G | A | 0.42 | 1.10 (1.08-1.11) | $1.82\times10^{-26}$ | 1.10 (1.07-1.13) | $2.60\times10^{-11}$ | 1.10 (1.08-1.13) | $1.71\times10^{-14}$ | 1.08 (1.03-1.12) | $1.07\times10^{-3}$ | 0.0 | 0.61 |
| rs74780677 | 3 | 48601774 | *COL7A1** | A | G | 0.94 | 1.23 (1.18-1.29) | $5.58\times10^{-19}$ | 1.14 (1.02-1.27) | 0.02 | 1.26 (1.19-1.32) | $2.04\times10^{-18}$ | - | - | 62.7 | 0.10 |
| rs13059018 | 3 | 55473083 | *WNT5A** | G | C | 0.91 | 1.11 (1.07-1.15) | $3.38\times10^{-9}$ | 1.08 (1.02-1.14) | 0.01 | 1.13 (1.08-1.17) | $5.03\times10^{-8}$ | - | - | 27.1 | 0.24 |
| rs11707541 | 3 | 121644529 | *CASR* | C | A | 0.47 | 1.07 (1.05-1.10) | $7.48\times10^{-9}$ | 1.07 (1.04-1.09) | $3.10\times10^{-6}$ | - | - | 1.08 (1.03-1.14) | $1.45\times10^{-3}$ | 0.0 | 0.58 |
| rs2202126 | 3 | 121942812 | *CASR* | A | G | 0.33 | 1.08 (1.06-1.10) | $1.19\times10^{-15}$ | 1.09 (1.05-1.12) | $6.70\times10^{-8}$ | 1.08 (1.05-1.11) | $5.72\times10^{-8}$ | 1.06 (1.01-1.11) | 0.01 | 0.0 | 0.64 |
| rs74830731 | 3 | 124753028 | *CASR* | G | A | 0.55 | 1.06 (1.04-1.08) | $8.77\times10^{-10}$ | 1.06 (1.03-1.09) | $3.60\times10^{-5}$ | 1.06 (1.03-1.09) | $7.93\times10^{-6}$ | - | - | 0.0 | 0.93 |
| rs369200 | 4 | 3743004 | *ADRA2C** | T | C | 0.44 | 1.06 (1.04-1.08) | $3.86\times10^{-10}$ | 1.04 (1.01-1.07) | $4.70\times10^{-3}$ | 1.07 (1.04-1.10) | $1.04\times10^{-7}$ | 1.07 (1.01-1.13) | 0.018 | 24.3 | 0.27 |
| rs11723275 | 4 | 77528821 | *SHROOM3** | C | A | 0.52 | 1.05 (1.04-1.07) | $4.24\times10^{-9}$ | 1.04 (1.01-1.07) | $2.70\times10^{-3}$ | 1.06 (1.03-1.08) | $8.55\times10^{-6}$ | 1.08 (1.02-1.14) | 0.01 | 0.0 | 0.48 |
| rs2231142 | 4 | 89052323 | *ABCG2* | T | G | 0.16 | 1.12 (1.09-1.15) | $7.12\times10^{-17}$ | 1.12 (1.08-1.17) | $6.20\times10^{-8}$ | 1.12 (1.07-1.17) | $2.62\times10^{-6}$ | 1.11 (1.06-1.16) | $1.47\times10^{-5}$ | 0.0 | 0.93 |

| Variant | | | | | | | Meta-analysis | | UK Biobank | | FinnGen | | Biobank Japan | | Het | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **CHR** | **POS** | **Candidate gene** | **EA** | **NEA** | **EAF** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | $I^2$ | **P** |
| rs10011293 | 4 | 115577381 | *UGT8* | T | C | 0.22 | 1.10 (1.08-1.13) | $9.38\times10^{-20}$ | 1.10 (1.06-1.14) | $1.30\times10^{-7}$ | 1.11 (1.08-1.14) | $5.16\times10^{-14}$ | 1.03 (0.91-1.15) | 0.67 | 0.0 | 0.38 |
| rs4865689 | 5 | 51164136 | *ISL1** | A | C | 0.39 | 1.06 (1.04-1.08) | $2.87\times10^{-10}$ | 1.05 (1.02-1.08) | $1.20\times10^{-3}$ | 1.07 (1.04-1.10) | $1.58\times10^{-7}$ | 1.04 (1-1.09) | 0.06 | 0.0 | 0.38 |
| rs11959132 | 5 | 58605916 | *PDE4D** | G | C | 0.76 | 1.06 (1.04-1.08) | $9.06\times10^{-9}$ | 1.06 (1.02-1.09) | $5.80\times10^{-4}$ | 1.07 (1.04-1.10) | $1.27\times10^{-5}$ | 1.07 (0.98-1.17) | 0.13 | 0.0 | 0.90 |
| rs638333 | 5 | 72419267 | *TMEM171* | C | T | 0.26 | 1.07 (1.05-1.10) | $1.12\times10^{-12}$ | 1.06 (1.03-1.09) | $1.50\times10^{-4}$ | 1.09 (1.06-1.12) | $1.77\times10^{-9}$ | 1.06 (0.98-1.14) | 0.15 | 4.6 | 0.35 |
| rs10051765 | 5 | 176799992 | *SLC34A1* | C | T | 0.37 | 1.14 (1.12-1.16) | $4.38\times10^{-51}$ | 1.16 (1.13-1.19) | $1.30\times10^{-24}$ | 1.13 (1.1-1.15) | $1.94\times10^{-20}$ | 1.15 (1.1-1.20) | $6.17\times10^{-10}$ | 13.1 | 0.32 |
| rs2249742 | 6 | 31240721 | *HLA-C* | T | C | 0.47 | 1.06 (1.04-1.08) | $2.49\times10^{-11}$ | 1.07 (1.04-1.10) | $1.30\times10^{-6}$ | 1.06 (1.03-1.09) | $7.12\times10^{-6}$ | 1.08 (1.03-1.13) | $5.75\times10^{-3}$ | 0.0 | 0.71 |
| rs9271375 | 6 | 32587067 | *HLA-DQA1** | G | A | 0.56 | 1.07 (1.05-1.09) | $1.47\times10^{-14}$ | 1.08 (1.05-1.11) | $6.40\times10^{-8}$ | 1.07 (1.04-1.09) | $5.26\times10^{-7}$ | 1.07 (1.01-1.14) | 0.03 | 0.0 | 0.90 |
| rs55856036 | 6 | 39148170 | *KCNK5* | C | A | 0.22 | 1.09 (1.07-1.11) | $5.73\times10^{-18}$ | 1.12 (1.08-1.16) | $9.00\times10^{-12}$ | 1.07 (1.04-1.10) | $1.44\times10^{-5}$ | 1.12 (1.06-1.19) | $7.26\times10^{-5}$ | 66.3 | 0.05 |
| rs729761 | 6 | 43804571 | *VEGFA* | T | G | 0.28 | 1.07 (1.05-1.09) | $1.87\times10^{-11}$ | 1.06 (1.03-1.09) | $1.40\times10^{-4}$ | 1.07 (1.04-1.10) | $2.63\times10^{-7}$ | 1.08 (1.01-1.16) | 0.03 | 0.0 | 0.76 |
| rs2076308 | 6 | 50791640 | *TFAP2B** | C | G | 0.22 | 1.08 (1.06-1.10) | $8.83\times10^{-13}$ | 1.07 (1.03-1.10) | $3.00\times10^{-4}$ | 1.08 (1.05-1.11) | $1.25\times10^{-6}$ | 1.09 (1.05-1.14) | $1.20\times10^{-4}$ | 0.0 | 0.67 |
| rs2465043 | 6 | 51180765 | *PKHD1* | G | A | 0.6 | 1.06 (1.04-1.08) | $3.87\times10^{-10}$ | 1.04 (1.02-1.07) | $2.20\times10^{-3}$ | 1.07 (1.04-1.09) | $1.62\times10^{-6}$ | 1.08 (1.03-1.13) | $1.13\times10^{-3}$ | 0.0 | 0.45 |
| rs72912510 | 6 | 90118764 | *RRAGD* | G | A | 0.8 | 1.07 (1.05-1.09) | $9.32\times10^{-10}$ | 1.08 (1.04-1.11) | $1.60\times10^{-5}$ | 1.06 (1.03-1.09) | $1.34\times10^{-4}$ | 1.11 (1.01-1.21) | 0.02 | 0.0 | 0.62 |
| rs6928418 | 6 | 101175347 | *ASCC3* | T | C | 0.52 | 1.06 (1.05-1.08) | $5.88\times10^{-13}$ | 1.06 (1.04-1.09) | $4.40\times10^{-6}$ | 1.06 (1.03-1.08) | $1.94\times10^{-5}$ | 1.09 (1.04-1.14) | $2.41\times10^{-3}$ | 0.0 | 0.49 |

| Variant | | | | | | | Meta-analysis | | UK Biobank | | FinnGen | | Biobank Japan | | Het | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **CHR** | **POS** | **Candidate gene** | **EA** | **NEA** | **EAF** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | $I^2$ | **P** |
| rs7740107 | 6 | 130374461 | *L3MBTL3* | A | T | 0.74 | 1.08 (1.06-1.10) | $1.52\times10^{-13}$ | 1.07 (1.04-1.10) | $1.20\times10^{-5}$ | 1.08 (1.05-1.12) | $2.36\times10^{-8}$ | 1.25 (1.08-1.46) | $3.82\times10^{-3}$ | 51.7 | 0.13 |
| rs12190287 | 6 | 134214525 | *TCF21** | C | G | 0.59 | 1.05 (1.03-1.07) | $1.69\times10^{-8}$ | 1.07 (1.04-1.10) | $3.40\times10^{-6}$ | 1.04 (1.02-1.07) | $8.07\times10^{-4}$ | 1.03 (0.99-1.08) | 0.18 | 12.4 | 0.32 |
| rs74495751 | 6 | 160624947 | *SLC22A2* | C | G | 0.03 | 1.24 (1.16-1.32) | $1.29\times10^{-10}$ | 1.24 (1.15-1.33) | $2.10\times10^{-8}$ | - | - | 1.23 (1.08-1.39) | $1.52\times10^{-3}$ | 0.0 | 0.92 |
| rs12539707 | 7 | 27626165 | *HIBADH* | T | C | 0.29 | 1.09 (1.07-1.11) | $3.63\times10^{-20}$ | 1.10 (1.07-1.14) | $1.50\times10^{-11}$ | 1.08 (1.06-1.11) | $3.17\times10^{-9}$ | 1.10 (1.01-1.19) | 0.03 | 0.0 | 0.63 |
| rs73301968 | 7 | 30917675 | *AQP1* | T | A | 0.07 | 1.14 (1.1-1.18) | $3.95\times10^{-13}$ | 1.19 (1.11-1.28) | $2.90\times10^{-6}$ | 1.10 (1.04-1.16) | $3.68\times10^{-4}$ | 1.17 (1.09-1.24) | $2.79\times10^{-6}$ | 41.2 | 0.18 |
| rs1004317 | 7 | 30956858 | *AQP1* | G | A | 0.38 | 1.05 (1.04-1.07) | $1.40\times10^{-9}$ | 1.06 (1.04-1.09) | $8.00\times10^{-6}$ | 1.07 (1.04-1.10) | $3.15\times10^{-7}$ | 1.03 (0.98-1.08) | 0.27 | 7.00 | 0.34 |
| rs4252512 | 7 | 142605221 | *TRPV5* | C | T | 0.02 | 1.34 (1.22-1.49) | $7.68\times10^{-9}$ | 1.34 (1.22-1.49) | $7.70\times10^{-9}$ | - | - | - | - | 0.0 | 1.00 |
| rs6464165 | 7 | 151413124 | *PRKAG2** | T | C | 0.74 | 1.07 (1.05-1.09) | $9.80\times10^{-11}$ | 1.08 (1.05-1.11) | $5.10\times10^{-7}$ | 1.06 (1.03-1.10) | $3.77\times10^{-5}$ | - | - | 0.0 | 0.55 |
| rs4739515 | 8 | 37391203 | *ZNF703** | G | C | 0.77 | 1.08 (1.05-1.12) | $2.17\times10^{-8}$ | 1.09 (1.02-1.17) | $8.90\times10^{-3}$ | 1.06 (1.01-1.11) | 0.02 | 1.11 (1.06-1.15) | $7.54\times10^{-6}$ | 0.0 | 0.38 |
| rs4745107 | 9 | 71177110 | *TMEM252** | A | G | 0.71 | 1.06 (1.04-1.08) | $1.63\times10^{-10}$ | 1.05 (1.02-1.08) | $8.30\times10^{-4}$ | 1.08 (1.05-1.11) | $2.89\times10^{-7}$ | 1.06 (1.01-1.11) | 0.03 | 0.0 | 0.48 |
| rs148349564 | 9 | 77466700 | *TRPM6* | C | T | 0.07 | 1.12 (1.08-1.16) | $2.62\times10^{-10}$ | 1.14 (1.08-1.20) | $2.80\times10^{-7}$ | 1.11 (1.05-1.16) | $1.07\times10^{-4}$ | - | - | 0.0 | 0.42 |
| rs188856714 | 9 | 97582244 | *AOPEP* | T | G | 0.91 | 1.10 (1.07-1.14) | $2.87\times10^{-10}$ | 1.09 (1.04-1.16) | $1.60\times10^{-3}$ | 1.12 (1.08-1.17) | $7.40\times10^{-8}$ | 1.06 (0.99-1.14) | 0.10 | 0.0 | 0.40 |
| rs17486892 | 10 | 9280394 | *GATA3* | T | C | 0.78 | 1.06 (1.04-1.08) | $1.62\times10^{-8}$ | 1.06 (1.03-1.10) | $4.30\times10^{-4}$ | 1.06 (1.03-1.10) | $3.89\times10^{-5}$ | 1.05 (0.99-1.11) | 0.10 | 0.0 | 0.90 |

| Variant | | | | | | | Meta-analysis | | UK Biobank | | FinnGen | | Biobank Japan | | Het | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **CHR** | **POS** | **Candidate gene** | **EA** | **NEA** | **EAF** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | $I^2$ | **P** |
| rs7908758 | 10 | 34994352 | *PARD3** | G | A | 0.33 | 1.06 (1.04-1.08) | $1.24\times10^{-10}$ | 1.06 (1.03-1.09) | $2.20\times10^{-4}$ | 1.07 (1.04-1.10) | $5.69\times10^{-6}$ | 1.06 (1.02-1.11) | 0.01 | 0.0 | 0.85 |
| rs55894538 | 10 | 90102403 | *RNLS* | T | G | 0.31 | 1.06 (1.04-1.08) | $2.37\times10^{-9}$ | 1.05 (1.02-1.08) | $1.80\times10^{-3}$ | 1.06 (1.03-1.09) | $2.00\times10^{-5}$ | 1.08 (1.03-1.14) | $2.72\times10^{-3}$ | 0.0 | 0.60 |
| rs61264398 | 11 | 10474497 | *AMPD3* | G | T | 0.53 | 1.06 (1.04-1.08) | $2.96\times10^{-12}$ | 1.07 (1.04-1.10) | $5.60\times10^{-7}$ | 1.06 (1.03-1.09) | $1.10\times10^{-5}$ | 1.05 (1.01-1.10) | 0.03 | 0.0 | 0.75 |
| rs1815945 | 11 | 111235756 | *SIK2** | G | T | 0.37 | 1.07 (1.05-1.08) | $2.25\times10^{-12}$ | 1.06 (1.03-1.09) | $3.40\times10^{-5}$ | 1.05 (1.02-1.08) | $3.84\times10^{-4}$ | 1.12 (1.07-1.17) | $5.30\times10^{-7}$ | 69 | 0.04 |
| rs7133836 | 12 | 42861098 | *PRICKLE1** | C | T | 0.43 | 1.06 (1.05-1.08) | $5.26\times10^{-13}$ | 1.05 (1.02-1.08) | $4.10\times10^{-4}$ | 1.07 (1.04-1.10) | $9.40\times10^{-8}$ | 1.08 (1.04-1.13) | $3.72\times10^{-3}$ | 0.0 | 0.41 |
| rs73187207 | 13 | 42619919 | *DGKH* | T | G | 0.2 | 1.08 (1.06-1.10) | $4.27\times10^{-12}$ | 1.08 (1.04-1.11) | $6.50\times10^{-6}$ | 1.07 (1.04-1.11) | $4.11\times10^{-6}$ | 1.12 (0.97-1.30) | 0.13 | 0.0 | 0.85 |
| rs36122479 | 13 | 42758435 | *DGKH* | A | T | 0.21 | 1.13 (1.11-1.16) | $1.51\times10^{-25}$ | 1.13 (1.08-1.18) | $8.00\times10^{-8}$ | 1.10 (1.06-1.14) | $3.83\times10^{-7}$ | 1.19 (1.14-1.24) | $2.53\times10^{-15}$ | 75.0 | 0.02 |
| rs10871066 | 13 | 73990436 | *KLF12** | G | A | 0.73 | 1.06 (1.04-1.08) | $3.66\times10^{-10}$ | 1.06 (1.03-1.09) | $3.20\times10^{-5}$ | 1.05 (1.02-1.08) | $2.61\times10^{-3}$ | 1.09 (1.04-1.14) | $1.08\times10^{-4}$ | 2.8 | 0.36 |
| rs57719175 | 13 | 96175396 | *CLDN10* | G | A | 0.61 | 1.06 (1.04-1.08) | $8.07\times10^{-11}$ | 1.05 (1.02-1.08) | $7.70\times10^{-4}$ | 1.07 (1.04-1.10) | $1.13\times10^{-7}$ | 1.05 (1-1.10) | 0.04 | 0.0 | 0.46 |
| rs12885720 | 14 | 30670202 | *PRKD1** | G | A | 0.86 | 1.09 (1.06-1.12) | $3.14\times10^{-10}$ | 1.12 (1.07-1.17) | $1.70\times10^{-7}$ | 1.07 (1.03-1.10) | $1.78\times10^{-4}$ | 1.12 (0.91-1.39) | 0.29 | 30.7 | 0.24 |
| rs7181593 | 15 | 51146114 | *SPPL2A** | C | T | 0.31 | 1.06 (1.04-1.08) | $2.41\times10^{-11}$ | 1.07 (1.04-1.10) | $1.30\times10^{-5}$ | 1.05 (1.02-1.08) | $1.82\times10^{-4}$ | 1.09 (1.04-1.14) | $2.27\times10^{-3}$ | 0.0 | 0.48 |
| rs2459379 | 15 | 54002712 | *WDR72* | G | A | 0.6 | 1.05 (1.03-1.07) | $7.60\times10^{-9}$ | 1.06 (1.03-1.09) | $7.80\times10^{-6}$ | 1.03 (1-1.06) | 0.03 | 1.10 (1.05-1.15) | $8.31\times10^{-5}$ | 71.0 | 0.032 |
| rs4775542 | 15 | 63078377 | *TLN2** | C | T | 0.25 | 1.06 (1.04-1.08) | $1.05\times10^{-8}$ | 1.06 (1.03-1.09) | $2.10\times10^{-4}$ | 1.04 (1.01-1.08) | $6.78\times10^{-3}$ | 1.10 (1.05-1.15) | $1.43\times10^{-3}$ | 34.6 | 0.22 |

| Variant | | | | | | | Meta-analysis | | UK Biobank | | FinnGen | | Biobank Japan | | Het | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **CHR** | **POS** | **Candidate gene** | **EA** | **NEA** | **EAF** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | $I^2$ | **P** |
| rs11629962 | 15 | 85530028 | *SLC28A1* | T | C | 0.25 | 1.06 (1.04-1.08) | $5.17\times10^{-9}$ | 1.07 (1.04-1.11) | $7.60\times10^{-6}$ | 1.05 (1.02-1.08) | $6.84\times10^{-4}$ | 1.18 (1.04-1.35) | 0.01 | 47.6 | 0.15 |
| rs118190679 | 16 | 20369797 | *UMOD* | C | T | 0.94 | 1.11 (1.07-1.16) | $4.12\times10^{-8}$ | 1.09 (1.03-1.16) | $1.60\times10^{-3}$ | 1.01 (0.96-1.07) | $6.75\times10^{-1}$ | - | - | 74.6 | 0.05 |
| rs12921916 | 16 | 20407196 | *UMOD* | C | T | 0.27 | 1.10 (1.08-1.12) | $1.45\times10^{-22}$ | 1.10 (1.07-1.13) | $2.60\times10^{-10}$ | 1.08 (1.05-1.11) | $1.30\times10^{-7}$ | 1.17 (1.11-1.24) | $1.37\times10^{-9}$ | 74.4 | 0.02 |
| rs889299 | 16 | 23381914 | *SCNN1B** | G | A | 0.76 | 1.07 (1.05-1.09) | $1.45\times10^{-10}$ | 1.06 (1.02-1.09) | $6.00\times10^{-4}$ | 1.07 (1.04-1.11) | $9.67\times10^{-6}$ | 1.09 (1.03-1.14) | $9.34\times10^{-4}$ | 0.0 | 0.60 |
| rs11642015 | 16 | 53802494 | *FTO* | T | C | 0.39 | 1.06 (1.04-1.08) | $7.60\times10^{-12}$ | 1.05 (1.02-1.08) | $8.70\times10^{-4}$ | 1.07 (1.04-1.10) | $1.84\times10^{-7}$ | 1.09 (1.04-1.15) | $8.70\times10^{-4}$ | 21.1 | 0.28 |
| rs2289114 | 16 | 56938522 | *SLC12A3** | T | C | 0.22 | 1.06 (1.04-1.08) | $4.03\times10^{-9}$ | 1.06 (1.02-1.10) | $1.20\times10^{-3}$ | 1.06 (1.03-1.09) | $3.56\times10^{-5}$ | 1.07 (1.02-1.12) | 0.01 | 0.0 | 0.95 |
| rs1858800 | 16 | 73024276 | *ZFHX3** | T | C | 0.27 | 1.06 (1.04-1.08) | $8.94\times10^{-9}$ | 1.05 (1.02-1.08) | $7.10\times10^{-4}$ | 1.07 (1.04-1.10) | $1.41\times10^{-5}$ | 1.05 (1-1.11) | 0.07 | 0.0 | 0.70 |
| rs4782364 | 16 | 88516135 | *ZFPM1* | G | C | 0.43 | 1.06 (1.04-1.08) | $1.25\times10^{-10}$ | 1.04 (1.02-1.07) | $1.50\times10^{-3}$ | 1.07 (1.04-1.10) | $1.19\times10^{-7}$ | 1.05 (1.01-1.10) | 0.02 | 0.0 | 0.40 |
| rs2721830 | 17 | 12278913 | *MAP2K4** | A | G | 0.51 | 1.06 (1.04-1.07) | $2.12\times10^{-10}$ | 1.06 (1.03-1.08) | $1.10\times10^{-4}$ | 1.05 (1.02-1.08) | $1.67\times10^{-4}$ | 1.08 (1.04-1.13) | $3.57\times10^{-4}$ | 0.0 | 0.49 |
| rs11868029 | 17 | 37650304 | *CDK12** | A | G | 0.21 | 1.07 (1.05-1.09) | $4.36\times10^{-11}$ | 1.08 (1.05-1.12) | $3.40\times10^{-7}$ | 1.07 (1.03-1.10) | $8.54\times10^{-5}$ | 1.05 (0.99-1.11) | 0.09 | 0.0 | 0.61 |
| rs12150247 | 17 | 43480759 | *ARHGAP27** | C | G | 0.91 | 1.10 (1.06-1.13) | $2.85\times10^{-8}$ | 1.04 (0.99-1.09) | $1.40\times10^{-1}$ | 1.04 (1-1.09) | 0.08 | - | - | 0.0 | 0.93 |
| rs111443054 | 17 | 43489161 | *ARHGAP27** | TCACCA | T | 0.15 | 1.09 (1.06-1.12) | $1.44\times10^{-8}$ | 1.06 (1.03-1.10) | $5.40\times10^{-4}$ | 1.11 (1.05-1.17) | $2.00\times10^{-4}$ | - | - | 37.4 | 0.21 |

| Variant | | | | | | | Meta-analysis | | UK Biobank | | FinnGen | | Biobank Japan | | Het | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **CHR** | **POS** | **Candidate gene** | **EA** | **NEA** | **EAF** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | $I^2$ | **P** |
| rs2532340 | 17 | 44345653 | *ARL17B** | G | A | 0.18 | 1.08 (1.05-1.10) | $4.95\times10^{-8}$ | 1.07 (1.04-1.11) | $3.40\times10^{-5}$ | 1.08 (1.04-1.13) | $3.74\times10^{-4}$ | - | - | 0.0 | 0.67 |
| rs9905274 | 17 | 59450441 | *BCAS3* | C | T | 0.75 | 1.12 (1.1-1.15) | $1.07\times10^{-26}$ | 1.08 (1.04-1.12) | $2.10\times10^{-5}$ | 1.15 (1.11-1.19) | $1.70\times10^{-16}$ | 1.14 (1.09-1.19) | $4.25\times10^{-9}$ | 69.6 | 0.04 |
| rs12325724 | 17 | 66709890 | *FAM20A** | A | C | 0.2 | 1.06 (1.04-1.09) | $1.31\times10^{-8}$ | 1.05 (1.02-1.09) | $1.70\times10^{-3}$ | 1.06 (1.03-1.10) | $2.58\times10^{-4}$ | 1.10 (1.04-1.16) | $1.16\times10^{-3}$ | 0.0 | 0.47 |
| rs4793434 | 17 | 70352537 | *SOX9* | G | C | 0.45 | 1.08 (1.06-1.10) | $5.89\times10^{-20}$ | 1.08 (1.05-1.11) | $2.10\times10^{-8}$ | 1.08 (1.05-1.11) | $6.77\times10^{-10}$ | 1.09 (1.04-1.14) | $2.04\times10^{-4}$ | 0.0 | 0.92 |
| rs11669940 | 19 | 4336379 | *STAP2* | C | G | 0.65 | 1.06 (1.04-1.08) | $1.08\times10^{-9}$ | 1.07 (1.04-1.10) | $2.60\times10^{-6}$ | 1.06 (1.03-1.08) | $6.41\times10^{-5}$ | 1.03 (0.98-1.07) | 0.20 | 19.5 | 0.29 |
| rs3760702 | 19 | 14588237 | *PTGER1* | A | G | 0.29 | 1.08 (1.06-1.10) | $4.39\times10^{-15}$ | 1.08 (1.05-1.11) | $4.10\times10^{-7}$ | 1.06 (1.03-1.09) | $6.78\times10^{-5}$ | 1.14 (1.08-1.19) | $3.78\times10^{-7}$ | 65.5 | 0.06 |
| rs11672660 | 19 | 46180184 | *GIPR* | C | T | 0.77 | 1.07 (1.05-1.09) | $3.43\times10^{-11}$ | 1.07 (1.03-1.10) | $1.40\times10^{-4}$ | 1.07 (1.04-1.11) | $9.42\times10^{-7}$ | 1.06 (1.01-1.12) | 0.02 | 0.0 | 0.92 |
| rs7259073 | 19 | 53357172 | *ZNF28** | T | C | 0.91 | 1.11 (1.07-1.15) | $4.58\times10^{-9}$ | 1.10 (1.04-1.16) | $7.40\times10^{-4}$ | 1.11 (1.07-1.16) | $1.48\times10^{-6}$ | - | - | 0.0 | 0.66 |
| rs6123359 | 20 | 52714706 | *CYP24A1* | G | A | 0.23 | 1.10 (1.07-1.12) | $1.84\times10^{-14}$ | 1.09 (1.05-1.14) | $8.10\times10^{-5}$ | 1.14 (1.09-1.18) | $6.05\times10^{-11}$ | 1.13 (1.08-1.18) | $4.19\times10^{-8}$ | 0.0 | 0.37 |
| rs17216707 | 20 | 52732362 | *CYP24A1* | T | C | 0.81 | 1.17 (1.15-1.20) | $3.41\times10^{-43}$ | 1.16 (1.12-1.20) | $9.70\times10^{-18}$ | 1.17 (1.13-1.21) | $5.69\times10^{-23}$ | 1.24 (1.13-1.37) | $5.90\times10^{-6}$ | 0.0 | 0.42 |
| rs2585442 | 20 | 52737123 | *CYP24A1* | G | C | 0.28 | 1.09 (1.06-1.11) | $2.06\times10^{-16}$ | 1.11 (1.08-1.14) | $7.70\times10^{-11}$ | 1.12 (1.09-1.15) | $8.81\times10^{-18}$ | 1.07 (0.98-1.17) | 0.11 | 0.0 | 0.58 |
| rs2762942 | 20 | 52788925 | *CYP24A1* | A | G | 0.95 | 1.13 (1.08-1.17) | $8.21\times10^{-9}$ | 1.11 (1.05-1.18) | $3.40\times10^{-4}$ | 1.16 (1.09-1.22) | $6.43\times10^{-7}$ | - | - | 0.0 | 0.36 |
| rs2823254 | 21 | 16783083 | *NRIP1** | A | T | 0.72 | 1.06 (1.04-1.08) | $9.04\times10^{-9}$ | 1.05 (1.02-1.09) | $3.00\times10^{-4}$ | 1.07 (1.04-1.10) | $1.20\times10^{-5}$ | 1.04 (0.99-1.09) | 0.16 | 0.0 | 0.57 |

| Variant | | | | | | | Meta-analysis | | UK Biobank | | FinnGen | | Biobank Japan | | Het | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **CHR** | **POS** | **Candidate gene** | **EA** | **NEA** | **EAF** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | **OR (95%CI)** | **P** | $I^2$ | **P** |
| rs56110144 | 21 | 37812848 | *CLDN14* | T | C | 0.81 | 1.08 (1.06-1.11) | $1.54\times10^{-11}$ | 1.06 (1.02-1.11) | $3.10\times10^{-3}$ | 1.04 (1-1.08) | 0.03 | 1.09 (1.04-1.14) | $3.75\times10^{-4}$ | 9.3 | 0.33 |
| rs219772 | 21 | 37835347 | *CLDN14* | A | T | 0.76 | 1.18 (1.15-1.21) | $1.79\times10^{-50}$ | 1.17 (1.13-1.21) | $2.50\times10^{-24}$ | 1.19 (1.15-1.23) | $5.55\times10^{-28}$ | - | - | 0.0 | 0.42 |

CHR=Chromosome; EA=Effect allele; EAF=Effect allele frequency; NEA=Non-effect allele; OR=Odds ratio; P=P-value; POS=Position based on NCBI Genome Build 37 (hg19); 95% CI=95% confidence interval. *denotes novel association.

Significant, independent signals in the trans-ancestry GWAS were directionally concordant in its constituent studies (Table 3.5). The three coding-region variants identified in the European ancestry KSD GWAS meta-analysis were significantly and independently associated with KSD in the trans-ancestry meta-analysis.

### 3.3.4 Heritability estimates

SNP-based heritability was ≈ 20% across all GWAS and meta-analyses conducted (Table 3.6). SNP-based heritability for the trans-ancestry GWAS meta-analysis could not be performed due to the absence of an appropriate reference panel for mixed-ancestry populations.

***Table 3.6. Heritability estimates for genome-wide association studies in UK Biobank, European ancestry genome-wide association study meta-analysis, and trans-ancestry meta-analysis.***

| | **UK Biobank** | **UK Biobank (Males)** | **UK Biobank (Females)** | **European ancestry** | **Trans-ancestry** |
|---|---|---|---|---|---|
| N SNPs in heritability analyses | 1,154,335 | 1,173,112 | 1,173,429 | 1,094,306 | 1,217,311 |
| N cases | 11,186 | 7,427 | 3,714 | 24,167 | 29,754 |
| N controls | 390,488 | 177,072 | 213,416 | 876,673 | 905,543 |
| $h^2$ SNP (SE) | 0.19 (0.02) | 0.20 (0.02) | 0.18 (0.04) | 0.18 (0.01) | - |
| LDSC intercept (SE) | 1.01 (0.01) | 1.01 (0.01) | 1.01 (0.01) | 1.05 (0.01) | 1.07 (0.01) |
| LDSC ratio (SE) | 0.09 (0.5) | 0.06 (0.06) | 0.12 (1.13) | 0.13 (0.03) | 0.17 (0.03) |
| Mean $\chi^2$ | 1.16 | 1.11 | 1.06 | 1.38 | 1.41 |

Estimates are calculated using LD-score regression (LDSC) at a population prevalence of 10%.
$h^2$ SNP=Mean SNP-based heritability (liability scale); N SNPs=Number of SNPs analysed; N cases=Number of cases; N controls=Number of controls; SE=Standard error.

### 3.3.5 Gene and gene set based analyses

Given the greater number of significant, independent association signals identified in the European and trans-ancestry KSD GWAS meta-analyses compared to the UK Biobank GWAS alone, these two meta-analyses were chosen for gene- and gene set-based analyses. Gene enrichment analysis associated SNPs from European ancestry GWAS meta-analysis (Table 3.7) with 128 protein-coding genes, and SNPs from the trans-ancestry GWAS meta-analysis (Table 3.8) with 127 protein-coding genes.

***Table 3.7. MAGMA gene association test for European ancestry kidney stone disease genome-wide association study meta-analysis.***

| Gene | CHR | N SNP | P-value |
|---|---|---|---|
| *DGKD* | 2 | 267 | $3.08 \times 10^{-15}$ |
| *MXD3* | 5 | 15 | $3.33 \times 10^{-15}$ |
| *DGKH* | 13 | 551 | $1.71 \times 10^{-14}$ |
| *SLC30A10* | 1 | 30 | $1.89 \times 10^{-14}$ |
| *SAG* | 2 | 144 | $2.05 \times 10^{-14}$ |
| *UGT8* | 4 | 156 | $2.50 \times 10^{-14}$ |
| *HIBADH* | 7 | 497 | $3.75 \times 10^{-14}$ |
| *TBX2* | 17 | 27 | $1.94 \times 10^{-13}$ |
| *PTGS2* | 1 | 21 | $2.40 \times 10^{-13}$ |
| *RAP1GAP* | 1 | 252 | $3.03 \times 10^{-13}$ |
| *AQP1* | 7 | 35 | $7.16 \times 10^{-13}$ |
| *ALPL* | 1 | 235 | $1.64 \times 10^{-12}$ |
| *ZNF512* | 2 | 81 | $2.92 \times 10^{-12}$ |
| *PDILT* | 16 | 118 | $5.09 \times 10^{-12}$ |
| *PRELID1* | 5 | 7 | $6.82 \times 10^{-12}$ |
| *PTGER1* | 19 | 10 | $7.91 \times 10^{-12}$ |
| *CASR* | 3 | 437 | $1.16 \times 10^{-11}$ |
| *RNLS* | 10 | 851 | $1.58 \times 10^{-11}$ |
| *LSM2* | 6 | 27 | $1.74 \times 10^{-11}$ |
| *MRPL33* | 2 | 13 | $2.97 \times 10^{-11}$ |
| *ASCC3* | 6 | 1111 | $4.00 \times 10^{-11}$ |
| *TMEM89* | 3 | 3 | $4.76 \times 10^{-11}$ |
| *KCNK5* | 6 | 144 | $5.64 \times 10^{-11}$ |
| *RBKS* | 2 | 234 | $9.77 \times 10^{-11}$ |
| *ATG16L1* | 2 | 123 | $1.04 \times 10^{-10}$ |
| *POU2AF1* | 11 | 103 | $1.06 \times 10^{-10}$ |
| *CDK12* | 17 | 232 | $1.42 \times 10^{-10}$ |
| *CLDN10* | 13 | 469 | $2.17 \times 10^{-10}$ |
| *TFAP2B* | 6 | 76 | $2.61 \times 10^{-10}$ |
| *FBXL20* | 17 | 352 | $4.85 \times 10^{-10}$ |
| *RAB29* | 1 | 21 | $4.98 \times 10^{-10}$ |
| *LMAN2* | 5 | 31 | $5.00 \times 10^{-10}$ |
| *RGS14* | 5 | 22 | $5.00 \times 10^{-10}$ |
| *SLC34A1* | 5 | 20 | $5.00 \times 10^{-10}$ |
| *MSH5* | 6 | 52 | $5.15 \times 10^{-10}$ |
| *YAF2* | 12 | 139 | $5.83 \times 10^{-10}$ |
| *MED1* | 17 | 90 | $6.74 \times 10^{-10}$ |
| *NSD1* | 5 | 280 | $7.24 \times 10^{-10}$ |
| *CYP24A1* | 20 | 92 | $8.04 \times 10^{-10}$ |
| *GCKR* | 2 | 44 | $1.04 \times 10^{-9}$ |
| *VARS* | 6 | 30 | $1.21 \times 10^{-9}$ |
| *TAX1BP1* | 7 | 215 | $1.69 \times 10^{-9}$ |
| *ZFPM1* | 16 | 355 | $1.83 \times 10^{-9}$ |
| *TNFAIP8L3* | 15 | 116 | $2.12 \times 10^{-9}$ |
| *CLDN14* | 21 | 404 | $2.40 \times 10^{-9}$ |
| *ZCRB1* | 12 | 22 | $3.29 \times 10^{-9}$ |
| *GIPR* | 19 | 35 | $4.45 \times 10^{-9}$ |
| *JAZF1* | 7 | 1038 | $5.39 \times 10^{-9}$ |
| *TRPV5* | 7 | 55 | $5.94 \times 10^{-9}$ |
| *FTO* | 16 | 1411 | $6.42 \times 10^{-9}$ |
| *VWA7* | 6 | 29 | $8.71 \times 10^{-9}$ |
| *AMPD3* | 11 | 213 | $1.55 \times 10^{-8}$ |
| *PPP2R1B* | 11 | 48 | $1.61 \times 10^{-8}$ |
| *TRIOBP* | 22 | 229 | $1.67 \times 10^{-8}$ |
| *TMEM171* | 5 | 43 | $1.68 \times 10^{-8}$ |
| *SIK2* | 11 | 194 | $1.69 \times 10^{-8}$ |
| *KLK15* | 19 | 18 | $1.82 \times 10^{-8}$ |
| *LY6G6F* | 6 | 9 | $1.84 \times 10^{-8}$ |
| *CCBL2* | 1 | 124 | $1.90 \times 10^{-8}$ |
| *MICA* | 6 | 253 | $1.91 \times 10^{-8}$ |
| *L3MBTL3* | 6 | 254 | $2.00 \times 10^{-8}$ |
| *GCAT* | 22 | 26 | $2.23 \times 10^{-8}$ |
| *STAP2* | 19 | 79 | $2.78 \times 10^{-8}$ |
| *CSNK2B* | 6 | 15 | $3.12 \times 10^{-8}$ |
| *NUCKS1* | 1 | 79 | $3.98 \times 10^{-8}$ |
| *AP4E1* | 15 | 239 | $4.23 \times 10^{-8}$ |
| *PLCL1* | 2 | 593 | $5.72 \times 10^{-8}$ |
| *PPHLN1* | 12 | 335 | $7.07 \times 10^{-8}$ |
| *FGFR4* | 5 | 25 | $7.47 \times 10^{-8}$ |
| *MICB* | 6 | 146 | $7.92 \times 10^{-8}$ |
| *RAB36* | 22 | 69 | $7.95 \times 10^{-8}$ |
| *EPRS* | 1 | 285 | $8.44 \times 10^{-8}$ |
| *CHAF1B* | 21 | 118 | $8.48 \times 10^{-8}$ |
| *AP1S3* | 2 | 250 | $9.10 \times 10^{-8}$ |
| *ABHD16A* | 6 | 25 | $1.00 \times 10^{-7}$ |
| *DXO* | 6 | 10 | $1.09 \times 10^{-7}$ |
| *PARD3* | 10 | 1674 | $1.09 \times 10^{-7}$ |
| *TNXB* | 6 | 192 | $1.20 \times 10^{-7}$ |
| *SLC12A3* | 16 | 246 | $1.23 \times 10^{-7}$ |
| *SF3B1* | 2 | 63 | $1.43 \times 10^{-7}$ |
| *DZIP1* | 13 | 105 | $1.53 \times 10^{-7}$ |
| *SLC15A2* | 3 | 69 | $1.53 \times 10^{-7}$ |
| *USP50* | 15 | 163 | $1.86 \times 10^{-7}$ |
| *BCR* | 22 | 460 | $2.25 \times 10^{-7}$ |
| *PRKAG2* | 7 | 1266 | $2.27 \times 10^{-7}$ |
| *LST1* | 6 | 11 | $2.90 \times 10^{-7}$ |
| *FOXL2NB* | 3 | 24 | $3.15 \times 10^{-7}$ |
| *QPCTL* | 19 | 26 | $3.34 \times 10^{-7}$ |
| *DDX41* | 5 | 7 | $3.46 \times 10^{-7}$ |
| *C2* | 6 | 121 | $3.64 \times 10^{-7}$ |
| *EPHB6* | 7 | 42 | $3.92 \times 10^{-7}$ |
| *KEL* | 7 | 53 | $4.14 \times 10^{-7}$ |
| *DDR1* | 6 | 63 | $4.35 \times 10^{-7}$ |
| *PKN1* | 19 | 104 | $4.93 \times 10^{-7}$ |
| *ITPK1* | 14 | 569 | $5.03 \times 10^{-7}$ |
| *TRPV6* | 7 | 18 | $5.39 \times 10^{-7}$ |
| *STH* | 17 | 1 | $5.65 \times 10^{-7}$ |
| *AQP4* | 18 | 38 | $5.71 \times 10^{-7}$ |
| *VPS13B* | 8 | 1781 | $6.23 \times 10^{-7}$ |
| *CRYAB* | 11 | 5 | $7.93 \times 10^{-7}$ |
| *CRHR1* | 17 | 1074 | $8.01 \times 10^{-7}$ |
| *TFEB* | 6 | 156 | $8.38 \times 10^{-7}$ |
| *PSORS1C1* | 6 | 212 | $9.71 \times 10^{-7}$ |
| *MAPT* | 17 | 785 | $9.85 \times 10^{-7}$ |
| *ATP6V1G2* | 6 | 8 | $1.02 \times 10^{-6}$ |
| *PDE3A* | 12 | 1100 | $1.05 \times 10^{-6}$ |
| *KANSL1* | 17 | 947 | $1.07 \times 10^{-6}$ |
| *RSPH14* | 22 | 211 | $1.10 \times 10^{-6}$ |
| *ARHGAP27* | 17 | 108 | $1.20 \times 10^{-6}$ |
| *SPPL2C* | 17 | 17 | $1.20 \times 10^{-6}$ |
| *PDE4D* | 5 | 3807 | $1.21 \times 10^{-6}$ |
| *PLEKHM1* | 17 | 156 | $1.29 \times 10^{-6}$ |
| *FAM209A* | 20 | 2 | $1.34 \times 10^{-6}$ |
| *ITIH1* | 3 | 29 | $1.43 \times 10^{-6}$ |
| *BRE* | 2 | 1126 | $1.44 \times 10^{-6}$ |
| *BAG6* | 6 | 38 | $1.59 \times 10^{-6}$ |
| *SLC14A2* | 18 | 1629 | $1.65 \times 10^{-6}$ |
| *NEU1* | 6 | 2 | $1.75 \times 10^{-6}$ |
| *SLC45A3* | 1 | 62 | $1.78 \times 10^{-6}$ |
| *HSPG2* | 1 | 379 | $1.82 \times 10^{-6}$ |
| *ITIH3* | 3 | 23 | $1.83 \times 10^{-6}$ |
| *FDXACB1* | 11 | 9 | $2.00 \times 10^{-6}$ |
| *TRPM7* | 15 | 444 | $2.08 \times 10^{-6}$ |
| *DNAH6* | 2 | 798 | $2.10 \times 10^{-6}$ |
| *CMIP* | 16 | 1139 | $2.15 \times 10^{-6}$ |
| *USP40* | 2 | 184 | $2.29 \times 10^{-6}$ |
| *FKBPL* | 6 | 5 | $2.43 \times 10^{-6}$ |
| *CUL2* | 10 | 204 | $2.50 \times 10^{-6}$ |

CHR=Chromosome; N SNP=Number of single nucleotide polymorphisms associated with gene included in analysis; P-value=P-value for association of gene.

***Table 3.8. MAGMA gene association test for trans-ancestry kidney stone disease genome-wide association study meta-analysis.***

| Gene | CHR | N SNP | P-value |
|---|---|---|---|
| *DGKH* | 13 | 620 | $4.09 \times 10^{-19}$ |
| *MXD3* | 5 | 16 | $2.16 \times 10^{-15}$ |
| *HIBADH* | 7 | 500 | $4.90 \times 10^{-15}$ |
| *SLC30A10* | 1 | 30 | $1.89 \times 10^{-14}$ |
| *UGT8* | 4 | 156 | $2.50 \times 10^{-14}$ |
| *SAG* | 2 | 152 | $7.69 \times 10^{-14}$ |
| *CASR* | 3 | 438 | $1.02 \times 10^{-13}$ |
| *PRELID1* | 5 | 8 | $1.54 \times 10^{-13}$ |
| *TBX2* | 17 | 27 | $1.94 \times 10^{-13}$ |
| *PTGS2* | 1 | 21 | $2.40 \times 10^{-13}$ |
| *PDILT* | 16 | 126 | $9.55 \times 10^{-13}$ |
| *DGKD* | 2 | 313 | $1.20 \times 10^{-12}$ |
| *AQP1* | 7 | 38 | $1.38 \times 10^{-12}$ |
| *ZNF512* | 2 | 81 | $2.92 \times 10^{-12}$ |
| *RAP1GAP* | 1 | 262 | $3.15 \times 10^{-12}$ |
| *PTGER1* | 19 | 10 | $7.91 \times 10^{-12}$ |
| *LSM2* | 6 | 31 | $1.83 \times 10^{-11}$ |
| *MRPL33* | 2 | 13 | $2.97 \times 10^{-11}$ |
| *RNLS* | 10 | 958 | $3.25 \times 10^{-11}$ |
| *TMEM89* | 3 | 3 | $4.76 \times 10^{-11}$ |
| *KCNK5* | 6 | 145 | $5.56 \times 10^{-11}$ |
| *RBKS* | 2 | 234 | $9.77 \times 10^{-11}$ |
| *ATG16L1* | 2 | 131 | $1.03 \times 10^{-10}$ |
| *POU2AF1* | 11 | 103 | $1.06 \times 10^{-10}$ |
| *ASCC3* | 6 | 1221 | $1.17 \times 10^{-10}$ |
| *CDK12* | 17 | 232 | $1.42 \times 10^{-10}$ |
| *NSD1* | 5 | 297 | $2.08 \times 10^{-10}$ |
| *CCBL2* | 1 | 182 | $2.74 \times 10^{-10}$ |
| *TFAP2B* | 6 | 77 | $2.82 \times 10^{-10}$ |
| *FBXL20* | 17 | 352 | $4.85 \times 10^{-10}$ |
| *ALPL* | 1 | 245 | $5.00 \times 10^{-10}$ |
| *LMAN2* | 5 | 31 | $5.00 \times 10^{-10}$ |
| *RGS14* | 5 | 22 | $5.00 \times 10^{-10}$ |
| *SLC34A1* | 5 | 21 | $5.00 \times 10^{-10}$ |
| *CLDN10* | 13 | 579 | $5.27 \times 10^{-10}$ |
| *L3MBTL3* | 6 | 292 | $5.93 \times 10^{-10}$ |
| *MED1* | 17 | 91 | $6.78 \times 10^{-10}$ |
| *CYP24A1* | 20 | 95 | $8.25 \times 10^{-10}$ |
| *MSH5* | 6 | 63 | $8.58 \times 10^{-10}$ |
| *GCKR* | 2 | 44 | $1.04 \times 10^{-9}$ |
| *ZFPM1* | 16 | 356 | $1.20 \times 10^{-9}$ |
| *ZCRB1* | 12 | 28 | $1.32 \times 10^{-9}$ |
| *TAX1BP1* | 7 | 216 | $1.69 \times 10^{-9}$ |
| *TNFAIP8L3* | 15 | 116 | $2.12 \times 10^{-9}$ |
| *YAF2* | 12 | 199 | $2.37 \times 10^{-9}$ |
| *VARS* | 6 | 37 | $2.75 \times 10^{-9}$ |
| *PPHLN1* | 12 | 417 | $3.73 \times 10^{-9}$ |
| *GIPR* | 19 | 37 | $4.71 \times 10^{-9}$ |
| *FTO* | 16 | 1472 | $4.82 \times 10^{-9}$ |
| *JAZF1* | 7 | 1070 | $5.79 \times 10^{-9}$ |
| *TRPV5* | 7 | 57 | $6.31 \times 10^{-9}$ |
| *VWA7* | 6 | 32 | $8.00 \times 10^{-9}$ |
| *CLDN14* | 21 | 478 | $1.05 \times 10^{-8}$ |
| *TRIOBP* | 22 | 232 | $1.25 \times 10^{-8}$ |
| *AP1S3* | 2 | 275 | $1.59 \times 10^{-8}$ |
| *PPP2R1B* | 11 | 48 | $1.61 \times 10^{-8}$ |
| *TMEM171* | 5 | 43 | $1.68 \times 10^{-8}$ |
| *SIK2* | 11 | 195 | $1.69 \times 10^{-8}$ |
| *GCAT* | 22 | 27 | $1.71 \times 10^{-8}$ |
| *AMPD3* | 11 | 223 | $1.76 \times 10^{-8}$ |
| *KLK15* | 19 | 18 | $1.82 \times 10^{-8}$ |
| *LY6G6F* | 6 | 9 | $1.84 \times 10^{-8}$ |
| *MICA* | 6 | 253 | $1.91 \times 10^{-8}$ |
| *RSPH14* | 22 | 221 | $2.11 \times 10^{-8}$ |
| *BCR* | 22 | 489 | $2.24 \times 10^{-8}$ |
| *SLC12A3* | 16 | 253 | $2.39 \times 10^{-8}$ |
| *STAP2* | 19 | 79 | $2.78 \times 10^{-8}$ |
| *NUCKS1* | 1 | 91 | $2.81 \times 10^{-8}$ |
| *PBRM1* | 3 | 341 | $3.02 \times 10^{-8}$ |
| *AP4E1* | 15 | 239 | $4.23 \times 10^{-8}$ |
| *LST1* | 6 | 12 | $4.85 \times 10^{-8}$ |
| *PLCL1* | 2 | 595 | $5.15 \times 10^{-8}$ |
| *RAB29* | 1 | 30 | $5.27 \times 10^{-8}$ |
| *CHAF1B* | 21 | 119 | $5.35 \times 10^{-8}$ |
| *GLT8D1* | 3 | 15 | $6.73 \times 10^{-8}$ |
| *FGFR4* | 5 | 25 | $7.47 \times 10^{-8}$ |
| *MICB* | 6 | 151 | $7.60 \times 10^{-8}$ |
| *SLC15A2* | 3 | 197 | $7.87 \times 10^{-8}$ |
| *RAB36* | 22 | 69 | $7.95 \times 10^{-8}$ |
| *EPRS* | 1 | 305 | $8.12 \times 10^{-8}$ |
| *SPCS1* | 3 | 4 | $9.07 \times 10^{-8}$ |
| *TNXB* | 6 | 209 | $1.35 \times 10^{-7}$ |
| *DXO* | 6 | 11 | $1.41 \times 10^{-7}$ |
| *SF3B1* | 2 | 63 | $1.43 \times 10^{-7}$ |
| *GNL3* | 3 | 16 | $1.44 \times 10^{-7}$ |
| *DZIP1* | 13 | 106 | $1.56 \times 10^{-7}$ |
| *PRKAG2* | 7 | 1335 | $1.56 \times 10^{-7}$ |
| *USP50* | 15 | 164 | $1.90 \times 10^{-7}$ |
| *NEK4* | 3 | 136 | $2.52 \times 10^{-7}$ |
| *PKN1* | 19 | 107 | $2.57 \times 10^{-7}$ |
| *ABHD16A* | 6 | 29 | $2.92 \times 10^{-7}$ |
| *PDE3A* | 12 | 1142 | $2.99 \times 10^{-7}$ |
| *FOXL2NB* | 3 | 25 | $3.23 \times 10^{-7}$ |
| *ITPK1* | 14 | 573 | $3.38 \times 10^{-7}$ |
| *DDX41* | 5 | 7 | $3.46 \times 10^{-7}$ |
| *QPCTL* | 19 | 29 | $3.60 \times 10^{-7}$ |
| *DDR1* | 6 | 65 | $4.57 \times 10^{-7}$ |
| *PARD3* | 10 | 1846 | $5.54 \times 10^{-7}$ |
| *STH* | 17 | 1 | $5.65 \times 10^{-7}$ |
| *AQP4* | 18 | 38 | $5.71 \times 10^{-7}$ |
| *CSNK2B* | 6 | 16 | $6.89 \times 10^{-7}$ |
| *TFEB* | 6 | 162 | $6.95 \times 10^{-7}$ |
| *PSORS1C1* | 6 | 248 | $7.63 \times 10^{-7}$ |
| *KANSL1* | 17 | 948 | $7.92 \times 10^{-7}$ |
| *CRYAB* | 11 | 5 | $7.93 \times 10^{-7}$ |
| *MAPT* | 17 | 791 | $7.98 \times 10^{-7}$ |
| *EPHB6* | 7 | 44 | $9.09 \times 10^{-7}$ |
| *CRHR1* | 17 | 1077 | $1.03 \times 10^{-6}$ |
| *VPS13B* | 8 | 1790 | $1.15 \times 10^{-6}$ |
| *ARHGAP27* | 17 | 108 | $1.20 \times 10^{-6}$ |
| *SPPL2C* | 17 | 17 | $1.20 \times 10^{-6}$ |
| *ITIH1* | 3 | 41 | $1.24 \times 10^{-6}$ |
| *PLEKHM1* | 17 | 156 | $1.29 \times 10^{-6}$ |
| *KEL* | 7 | 56 | $1.30 \times 10^{-6}$ |
| *FAM209A* | 20 | 2 | $1.34 \times 10^{-6}$ |
| *SLC14A2* | 18 | 1701 | $1.45 \times 10^{-6}$ |
| *BRE* | 2 | 1139 | $1.66 \times 10^{-6}$ |
| *NEU1* | 6 | 2 | $1.75 \times 10^{-6}$ |
| *ATP6V1G2* | 6 | 9 | $1.78 \times 10^{-6}$ |
| *SLC45A3* | 1 | 66 | $1.99 \times 10^{-6}$ |
| *FDXACB1* | 11 | 9 | $2.00 \times 10^{-6}$ |
| *STAB1* | 3 | 52 | $2.04 \times 10^{-6}$ |
| *ATL1* | 14 | 205 | $2.09 \times 10^{-6}$ |
| *ZFHX3* | 16 | 866 | $2.37 \times 10^{-6}$ |
| *PDE4D* | 5 | 4279 | $2.44 \times 10^{-6}$ |
| *CUL2* | 10 | 226 | $2.44 \times 10^{-6}$ |
| *CMIP* | 16 | 1207 | $2.44 \times 10^{-6}$ |

CHR=Chromosome; N SNP=Number of single nucleotide polymorphisms associated with gene included in analysis; P-value=P-value for association of gene.

Gene-set enrichment analysis using HPO GO gene sets and cell-type gene sets revealed enrichment of genes in 10 gene sets based on data from the KSD European ancestry GWAS meta-analysis (Table 3.9) and 10 gene sets based on data from the KSD trans-ancestry GWAS meta-analysis (Table 3.10). Eight of these gene sets were common to both European and trans-ancestry GWAS meta-analyses.

***Table 3.9. Gene set analysis in European ancestry kidney stone disease genome-wide association study meta-analysis.***

| Analysis | Gene set | N Genes | P-value | Significant genes |
|---|---|---|---|---|
| Human Phenotype Ontology Gene Ontology | Abnormal blood inorganic cation concentration | 153 | $4.43 \times 10^{-8}$ | *ALG12, ALPL, CASR, CLDN10, CYP24A1, GATA3, KCNJ1, NSD1, PDE4D, RREB1, SEC24C, SLC12A3, SLC30A10, SLC34A1* |
| | Hypocalciuria | 10 | $1.42 \times 10^{-7}$ | *CASR, CLDN10, SLC12A3* |
| | Hypermagnesemia | 7 | $1.40 \times 10^{-7}$ | *CASR, CLDN10, SLC12A3* |
| | Apical junction assembly | 73 | $2.71 \times 10^{-7}$ | *CLDN10, CLDN14, F11R, MYO1C, PARD3, ROCK2, VCL, WDR1* |
| | Fibroblast growth factor production | 6 | $3.12 \times 10^{-7}$ | *AIF1, HEG1, PTGS2, ROCK2, WNT11* |
| | Metanephric loop of Henle development | 6 | $6.20 \times 10^{-7}$ | *AQP1, PKD2, UMOD* |
| | Alkaline phosphatase activity | 6 | $6.58 \times 10^{-7}$ | *ALPI, ALPL, ALPP, ALPPL2* |
| | Anion homeostasis | 49 | $2.70 \times 10^{-6}$ | *ABCC6, ALPL, ATP6V1B1, FGFR4, SLC12A3, SLC17A6, SLC34A1, TFAP2B, UMOD* |
| | Urate metabolic process | 13 | $4.66 \times 10^{-6}$ | *ABCG2, GCKR, LRRC16A, SLC17A1, SLC17A3* |
| Cell-type | Lake adult kidney C3 proximal tubule epithelial cells S1 S2 | 207 | $9.26 \times 10^{-6}$ | *ANKRD11, BCAS3, CALD1, CLDN10, PARD3, PDE4D, PDE8A, PKHD1, SIK2, SLC17A1, SLC17A3, TFDP2, TLN2, WDR72* |

Gene set=Gene set name, ranked by descending p-value [461–463]; N Genes=Number of genes in gene set; P-value=P-value for association of gene set with kidney stone disease; Genes=Genes in gene set significantly associated with kidney stone disease, listed in alphabetical order.

***Table 3.10. Gene set analysis in trans-ancestry kidney stone disease genome-wide association study meta-analysis.***

| Analysis | Gene set | N Genes | P-value | Significant Genes |
|---|---|---|---|---|
| Human Phenotype Ontology Gene Ontology | Hypermagnesemia | 7 | $9.14 \times 10^{-8}$ | *CASR, CLDN10, SLC12A3* |
| | Hypocalciuria | 10 | $1.69 \times 10^{-7}$ | *CASR, CLDN10, SLC12A3* |
| | Abnormal blood inorganic cation concentration | 153 | $3.75 \times 10^{-7}$ | *ALG12, ALPL, CASR, CLDN10, CYP24A1, GATA3, KCNJ1, NSD1, PDE4D, RREB1, SEC24C, SLC12A3, SLC30A10, SLC34A1* |
| | Fibroblast growth factor production | 6 | $5.25 \times 10^{-7}$ | *AIF1, HEG1, PTGS2, ROCK2, WNT11* |
| | Anion homeostasis | 49 | $7.10 \times 10^{-7}$ | *ABCC6, ALPL, ATP6V1B1, FGFR4, SLC12A3, SLC17A6, SLC34A1, TFAP2B, UMOD* |
| | Metanephric loop of Henle development | 6 | $9.68 \times 10^{-7}$ | *AQP1, PKD2, UMOD* |
| | Apical junction assembly | 73 | $1.00 \times 10^{-6}$ | *CLDN10, CLDN14, F11R, MYO1C, PARD3, PKN2, WDR1* |
| Cell-type | Lake adult kidney C3 proximal tubule epithelial cells S1 S2 | 207 | $9.26 \times 10^{-6}$ | *ANKRD11, BCAS3, CALD1, CLDN10, PARD3, PDE4D, PDE8A, PKHD1, SIK2, SLC17A1, SLC17A3, TFDP2, TLN2, WDR72* |
| | Lake adult kidney C13 thick ascending limb | 121 | $1.97 \times 10^{-5}$ | *ADK, BCAS3, BRE, CASR, HIBADH, KANSL1, NR3C2, PKHD1, PLCL1, UMOD* |
| | Lake adult kidney C4 proximal tubule epithelial cells S2 | 170 | $5.81 \times 10^{-5}$ | *BRE, CALD1, CRIM1, DGKH, KANSL1, PARD3, PDE4D, PDE8A, PKHD1, PLCL1, SHROOM3, SLC17A1* |

Gene set=Gene set name, ranked by descending p-value [461–463]; N Genes=Number of genes in gene set; P-value=P-value for association of gene set with kidney stone disease; Genes=Genes in gene set significantly associated with kidney stone disease, listed in alphabetical order.

### 3.3.6 Exome-wide association study of kidney stone disease in the UK Biobank

#### 3.3.6.1 Single variant associations

After variant and sample QC, 5,116,334 variants and 392,594 individuals (11,225 cases and 381,369 controls) were included in an ExWAS. Following conditional analyses of 28 variants with $p<5\times10^{-8}$, 10 independent variants at nine loci across eight chromosomes were associated with KSD (Figure 3.9, Figure 3.10, Table 3.11). These variants had effect sizes ranging from 1.09 (95% CI=1.06-1.12, $p=1.18\times10^{-8}$, (22:23150049:C:T)) to 28.47 (95% CI=28.47-93.00, $p=2.96\times10^{-8}$, (19:36084555:C:A)).

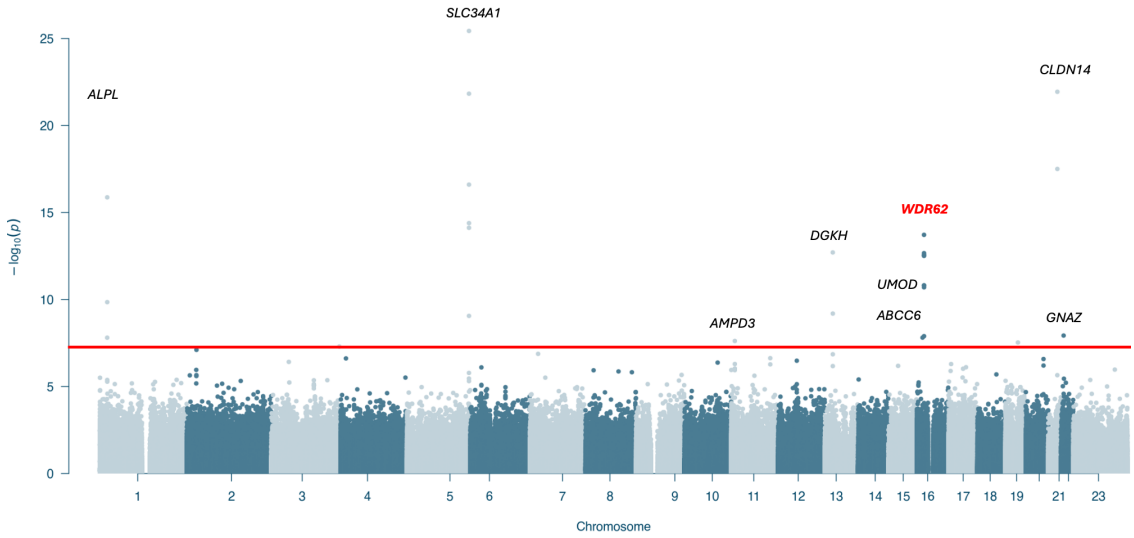



***Figure 3.9. Manhattan plot for an exome-wide association study of kidney stone disease in the UK Biobank.***

P-values (-log10) plotted against their respective positions on each chromosome. The horizontal red line indicates the genome-wide significance threshold of $5.00\times10^{-8}$. Loci are labelled with the primary candidate gene at each locus, as shown in Table 3.11. A novel locus is shown in red.

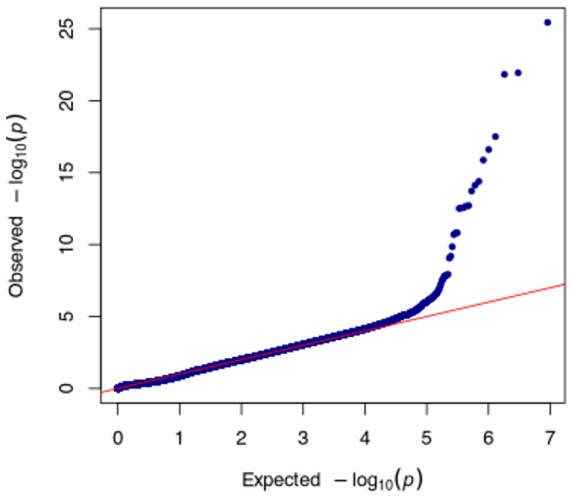



***Figure 3.10. Quantile-quantile plot of observed vs. expected p-values for exome-wide association study of kidney stone disease in the UK Biobank.***

An exome-wide association study of kidney stone disease was performed for 11,225 individuals with kidney stone disease and 381,369 controls from the UK Biobank.

***Table 3.11. Variants significantly associated with kidney stone disease from an exome-wide association study of kidney stone disease in the UK Biobank.***

| Variant ID | CHR | POS | Candidate Gene | EA | NEA | EAF | OR (95% CI) | P |
|---|---|---|---|---|---|---|---|---|
| 1:21478127:C:T | 1 | 21478127 | *ALPL* | T | C | 0.02 | 1.35 (1.22-1.49) | $2.95 \times 10^{-9}$ |
| 1:21570274:C:T | 1 | 21570274 | *ALPL* | T | C | 0.16 | 1.16 (1.12-1.2) | $1.34 \times 10^{-16}$ |
| 5:177371039:T:G | 5 | 177371039 | *SLC34A1* | G | T | 0.34 | 1.16 (1.13-1.2) | $3.68 \times 10^{-26}$ |
| 11:10484171:A:C | 11 | 10484171 | *AMPD3* | C | A | 0.09 | 1.14 (1.09-1.19) | $2.41 \times 10^{-8}$ |
| 13:42127603:A:T | 13 | 42127603 | *DGKH* | A | T | 0.18 | 1.13 (1.1-1.17) | $1.97 \times 10^{-13}$ |
| 16:16165739:G:A | 16 | 16165739 | *ABCC6* | G | A | 0.97 | 1.29 (1.18-1.41) | $1.79 \times 10^{-8}$ |
| 16:20350628:C:T | 16 | 20350628 | *UMOD* | T | C | 0.17 | 1.15 (1.11-1.19) | $1.92 \times 10^{-14}$ |
| 19:36084555:C:A | 19 | 36084555 | *WDR62* | A | C | $2.18 \times 10^{-5}$ | 28.47 (8.71-93.00) | $2.96 \times 10^{-8}$ |
| 21:36461453:G:A | 21 | 36461453 | *CLDN14* | G | A | 0.74 | 1.17 (1.13-1.2) | $1.15 \times 10^{-22}$ |
| 22:23150049:C:T | 22 | 23150049 | *GNAZ* | T | C | 0.33 | 1.09 (1.06-1.12) | $1.18 \times 10^{-8}$ |

Variants significantly associated with kidney stone disease from an exome-wide association study in the UK Biobank.
CHR=Chromosome; EA=Effect allele; EAF=Effect allele frequency; NEA=Non-effect allele; OR=Odds ratio; P=P-value; POS=Position based on NCBI Genome Build 38 (hg38); 95% CI=95% confidence interval.

As expected, the variant with the lowest minor allele frequency (19:36084555:C:A) had the largest effect estimate (Figure 3.11). This variant was found in 0.03% of cases compared to $1.32\times10^{-3}$% of controls. This rare, intronic variant is only present in 16 participants in the UK Biobank [484] and is not found in the *All of Us* cohort [442].

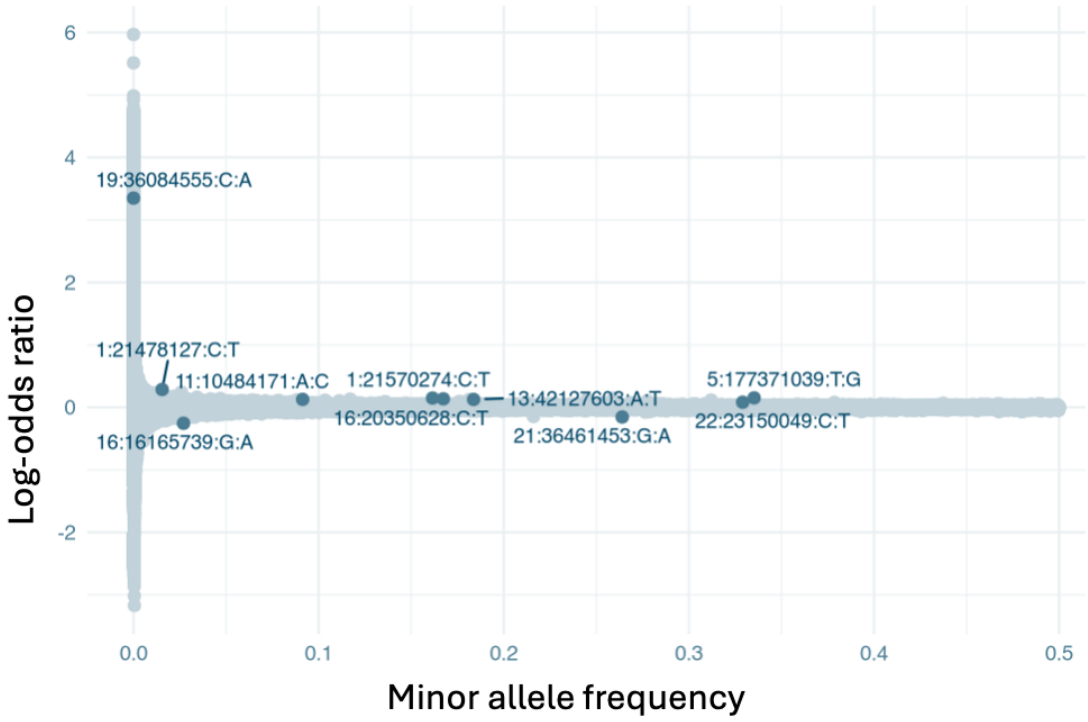



***Figure 3.11. Comparisons of rare variant effect sizes with common effects.***
Variant minor allele frequency versus absolute log-odds ratio for variants included in an exome-wide association study of kidney stone disease (KSD) in the UK Biobank. Variants independently and significantly ($p< 5 \times 10^{-8}$) associated with KSD are labelled and shown in dark blue.

### 3.3.6.2 Collapsing analyses

To elucidate genes whose impaired function is associated with KSD using whole-exome sequence data, I tested for associations between rare and damaging coding variants with KSD.

#### Protein truncating variants

In the analysis of rare PTVs with KSD, 11,168 genes were represented in the analyses and no genes were significantly associated ($p< 0.05/11,168 = 4.77\times10^{-6}$) with KSD .

**Missense variants**

In the analysis of rare, missense variants with KSD, 18,865 genes were represented in the analyses, thus SKAT-O p$< 0.05/18,856 = 2.68 \times 10^{-6}$ was considered significant. Missense variants in *ALPL* (p$=3.61 \times 10^{-7}$), *SLC34A1* (p$=1.15 \times 10^{-6}$), and *SLC26A2* (p$=2.58 \times 10^{-6}$) were significantly associated with KSD (Table 3.12).

In the LOVO analysis of missense variants in *ALPL* and of *SLC34A1*, omitting no single variant caused the significance threshold to reach p$> 0.05/53 = 9.43 \times 10^{-4}$ and p$> 0.05/67 = 7.46 \times 10^{-4}$, respectively. This suggests that the increased risk of KSD associated with rare, missense variants in *ALPL* and *SLC34A1* is likely due to the combined effect of multiple variants in each of these genes rather than being driven by a single variant.

In the LOVO analysis of missense variants in *SLC26A2*, excluding *5:149981314:T:C* caused the gene-based test p-value to increase from $2.58 \times 10^{-6}$ to 0.50, exceeding the significance threshold of p$> 0.05/57 = 8.77 \times 10^{-4}$. This indicates that this missense variant likely plays a key role in driving the association of *SLC26A2* with KSD. 5:149981314:C (p.T574I), the risk allele for KSD (OR=1.28, 95% CI=1.12-1.45, p$=2.25 \times 10^{-4}$, frequency 0.99) in the UK Biobank, results in an isoleucine substitution at position 574 of the solute carrier family 26 member 2 sulphate transporter protein, and is predicted benign in ClinVar [485].

**Non-synonymous variants**

In the analysis of rare, non-synonymous variants with KSD, 18,700 genes were represented in the analyses, thus SKAT-O p$< 0.05/18,700 = 2.67 \times 10^{-6}$ was considered significant. Non-synonymous variants in *ALPL* (p$=4.02 \times 10^{-7}$) and *SLC34A1* (p$=1.14 \times 10^{-6}$) were significantly associated with KSD (Table 3.12).

In the LOVO analysis of non-synonymous variants in *ALPL* and *SLC34A1*, omitting no single variant caused the significance threshold to reach p$> 0.05/54 = 9.26 \times 10^{-4}$ or p$> 0.05/70 = 7.14 \times 10^{-4}$, respectively.

### Protein truncating, missense, and non-synonymous variants

When the effects of rare PTV, missense, and non-synonymous variants were considered together, 17,962 genes were represented in the analyses, thus SKAT-O p$< 0.05/17,962 = 2.78 \times 10^{-6}$ was considered significant. Variants in *ALPL* (p$=4.02 \times 10^{-7}$) and *SLC34A1* (p$=1.14 \times 10^{-6}$) were significantly associated with KSD (Table 3.12).

Again, in the LOVO analysis of PTV, missense, and non-synonymous variants in *ALPL* and of *SLC34A1*, omitting no single variant caused the significance threshold to reach p$> 0.05/54 = 9.26 \times 10^{-4}$ or p$> 0.05/70 = 7.14 \times 10^{-4}$, respectively.

### CADD $\geq$ 25 variants

In the analysis of rare variants with a CADD score $\geq 25$ associated with KSD, 16,917 genes were represented in the analyses, thus SKAT-O p$< 0.05/16,917 = 2.96 \times 10^{-6}$ was considered significant. Variants in *SLC34A3* (p$=1.18 \times 10^{-6}$) were significantly associated with KSD (Table 3.12).

In the LOVO analysis of variants with a CADD score $\geq 25$ in *SLC34A3*, omitting no single variant caused the significance threshold to reach p$> 0.05/20 = 2.50 \times 10^{-3}$.

***Table 3.12. Summary of significant results from exome-wide, gene-based collapsing analyses of rare variants and kidney stone disease in the UK Biobank.***

| Variants in collapsing analysis | Gene | OR (95%CI) | SKAT-O P-value |
|---|---|---|---|
| Missense | *ALPL* | 1.13 (1.08-1.18) | $6.98 \times 10^{-8}$ |
| | *SLC34A1* | 1.18 (1.11-1.27) | $1.51 \times 10^{-6}$ |
| | *SLC26A2* | 1.28 (1.12-1.45) | $2.58 \times 10^{-6}$ |
| Non-synonymous | *ALPL* | 1.13 (1.08-1.18) | $4.02 \times 10^{-7}$ |
| | *SLC34A1* | 1.19 (1.11-1.26) | $1.14 \times 10^{-6}$ |
| PTV, missense, non-synonymous | *ALPL* | 1.13 (1.08-1.18) | $7.88 \times 10^{-8}$ |
| | *SLC34A1* | 1.19 (1.11-1.26) | $1.14 \times 10^{-6}$ |
| CADD $\geq$ 25 | *SLC34A3* | 1.71 (1.39-2.09) | $1.18 \times 10^{-6}$ |

All included variants had minor allele frequency < 1%.
CADD=Combined Annotation Dependent Depletion score; PTV=Protein-truncating variant; SKAT-O=Sequence kernel association test optimal.

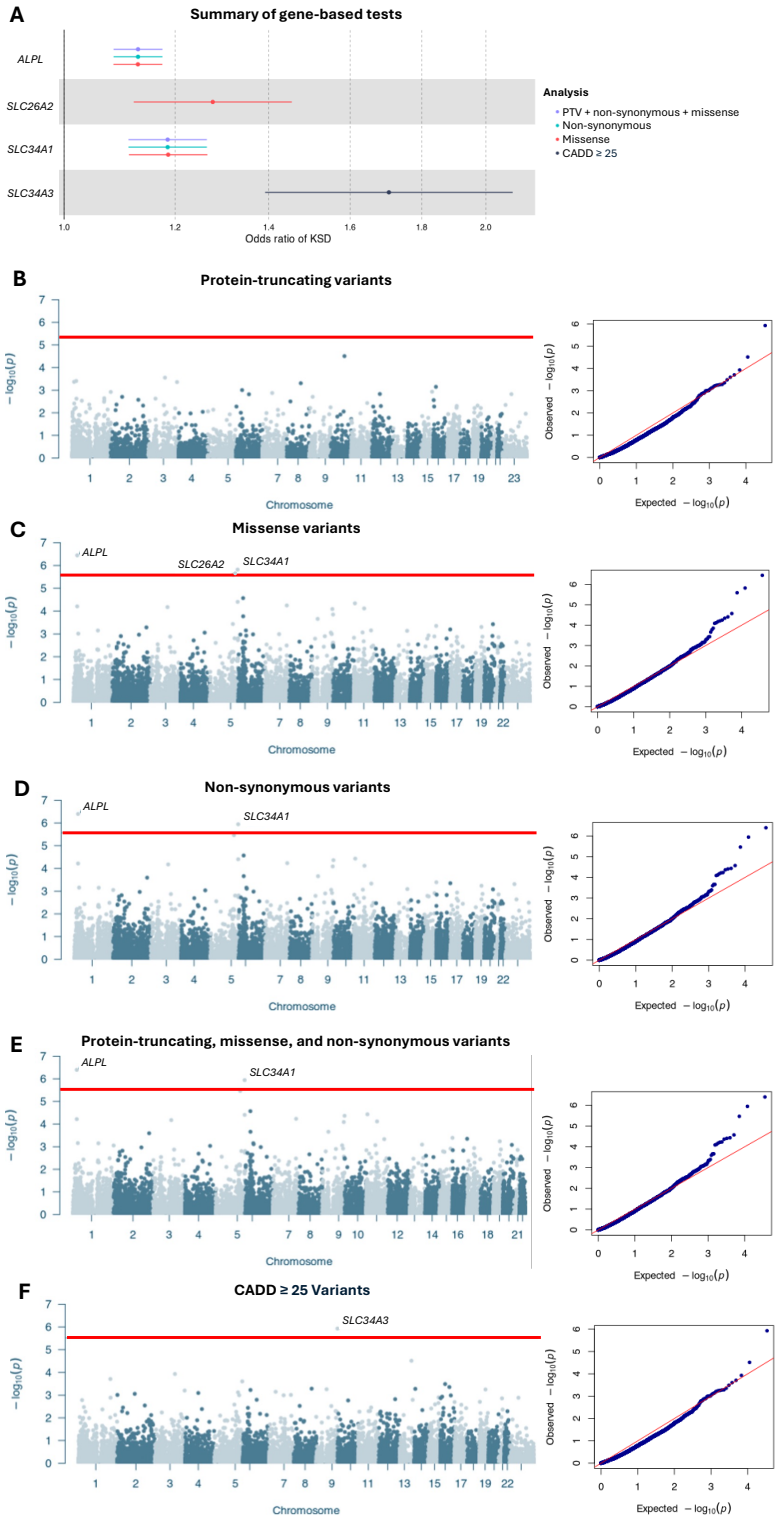



***Figure 3.12. Exome-wide gene-based association testing for kidney stone disease in 11,225 cases and 381,369 controls in the UK Biobank, based on rare variants.***

**A:** Significant results from gene-based association tests across all analyses; **B-F:** Left=Manhattan plots of results from gene-based testing on variant inclusion criteria. Right=Quantile-quantile plots of results from gene-based testing on variant inclusion criteria. **B:** Protein truncating variants, **C:** Missense variants, **D:** Non-synonymous variants, **E:** Protein-truncating, missense and non-synonymous variants, **F:** Variants with CADD score $\geq$ 25. The red lines indicate Bonferroni-based multiple testing correction thresholds.

CADD=Combined Annotation Dependent Depletion score; PTV=Protein-truncating variant.

## 3.4 Discussion and conclusions

My first aim for this thesis was to elucidate the genetic architecture of KSD by conducting GWAS and meta-analyses in European and trans-ancestral populations, as well as ExWAS in a European population. In this chapter, I present, to the best of my knowledge, the most extensive genetic association analyses of KSD performed to date, employing both genome- and exome-wide approaches. These efforts have not only identified a greater number of loci associated with KSD than previously recognised but have also pinpointed specific genetic variants and genes implicated in the disease. These findings hold significant potential to advance mechanistic research into the pathophysiology of KSD and drive the development of novel therapeutic approaches.

### 3.4.1 Genome-wide association studies

My first research objective in this chapter was to conduct combined-, male-, and female-sex GWAS in the UK Biobank, employing improved case ascertainment compared to previous studies, and to perform meta-analyses in European and trans-ancestry populations. Several notable findings emerged from these analyses.

The UK Biobank GWAS identified 17, 14, and six independent genome-wide significant variant associations in the combined-, male-, and female-sex studies, respectively. Notably, two of the loci (*FLOT1* and *ABCC6*) in combined-sex analyses have not previously been associated with KSD. *FLOT1* encodes a caveolae-associated, integral membrane protein that is expressed in renal tissue and is implicated in vesicular trafficking and signal transduction [486, 487]. *ABCC6* encodes the ATP binding cassette subfamily C member 6 transport protein. Biallelic mutations in *ABCC6* cause pseudoxanthoma elasticum which is associated with skin, cardiovascular and retinal calcifications and lower circulating pyrophosphate concentrations [49, 488]. Pyrophosphate inhibits calcium oxalate precipitation and this may represent one mechanism behind this gene's association with KSD.

Although there was a discrepancy in the number of independent signals associated

with KSD in the male- and female-sex GWAS, all signals demonstrated directional concordance in data from both sexes. The lower number of significant associations in the female-sex analysis may be due to reduced statistical power compared to the male-sex analysis. While KSD has historically been more prevalent in male than female patients, epidemiological evidence suggests that this disparity is narrowing [9]. Interestingly, I found no evidence for sex-specific loci associated with KSD, which may suggest that variations in KSD prevalence are driven by environmental or behavioural risk factors rather than underlying genetic differences.

My results from meta-analyses of European and trans-ancestry KSD GWAS provide important insights, identifying 33 and 38 novel loci associated with KSD, respectively, including three missense variants in protein-coding regions of the genome. One such locus is the missense variant rs116804195 (C>T, p.R200Q) in the claudin-19 gene (*CLDN19*). *CLDN19* encodes the claudin-19 protein which is expressed in the thick ascending limb of the loop of Henle. As described in the introduction to this thesis, claudin-19 enables paracellular renal tubular ion reabsorption [171, 489] and biallelic loss-of-function mutations in *CLDN19* are implicated in familial hypomagnesaemia with hypercalciuria and nephrocalcinosis. While monogenic mutations in *CLDN19* are an established cause of KSD and common polymorphisms in the *CLDN14* gene are linked to kidney stones [119, 158] and elevated urinary calcium levels [240], polymorphisms in *CLDN19* have not previously been significantly associated with KSD. Despite the rs116804195 missense variant having benign SIFT, PolyPhen, CADD, and REVEL predictions, is plausible that, with greater statistical power than previous GWAS, this study contributes to identifying an allelic series of *CLDN19* variants associated with KSD, which could open the door to new therapeutic avenues in the future.

A second notable missense variant identified in both the European and trans-ancestry GWAS meta-analyses is rs2231142 (G>T, p.Q141K) in *ABCG2*. The rs2231142-T allele, associated with increased risk for KSD, is linked to elevated serum urate concentrations, with a stronger observed in males than females [490]. This variant is also associated with statin responsiveness and has a higher allele frequency in East Asian populations (29.82%) compared to European populations (11.09%),

according to gnomAD data [491, 492], suggesting potential ethnic variability in KSD risk related to this variant. *ABCG2* encodes the breast cancer resistance protein (BCRP), a transporter expressed in the apical membrane of the proximal renal tubule [493]. As a member of the ATP-binding cassette (ABC) transporter family, *ABCG2* plays a critical role in excreting urate and various drugs. The association of this variant with increased serum urate levels implies that impaired renal excretion of uric acid may contribute to kidney stone formation. Given that rs2231142 affects statin, but not allopurinol, responsiveness, future therapeutic approaches leveraging these findings should take this distinction into account when considering potential for personalised medicine approaches for KSD.

### 3.4.2 Downstream bioinformatic analyses

In addition to identifying more independent genetic associations with KSD than previously reported, performing GWAS and meta-analyses with improved case ascertainment captured ≈20% of the estimated 45% heritability of KSD. This comprehensive approach facilitated gene-based analyses, mapping association signals to 136 distinct protein-coding genes. These annotations revealed enrichment in gene sets linked to homeostasis (HPO terms: hypermagnesemia, hypocalciuria, abnormal blood inorganic cation concentration, fibroblast growth factor production, anion homeostasis, urate metabolic process, and alkaline phosphatase activity), renal structure (HPO terms: metanephric loop of Henle development, apical junction assembly), and renal cell types (proximal tubule cells and thick ascending limb of the loop of Henle).

Several of these gene sets confirm our current understanding of the importance of calcium and phosphate metabolism in KSD pathophysiology such as "abnormal blood inorganic cation concentration", "fibroblast growth factor production", and "anion homeostasis". However, the importance of other enriched sets, such as "metanephric loop of Henle development", are arguably less well understood. These findings suggest that overlapping genetic and biological pathways may contribute to both KSD and renal structural disorders. Notably, significant associations of *PKD2*, *AQP1*, and *UMOD* genes with KSD hint at a potential mechanistic link between

renal cyst formation and kidney stone susceptibility.

*PKD2* encodes polycystin-2, a protein expressed in the kidney. Germline mutations in *PKD2* cause ADPKD in which ≈20% of patients develop kidney stones [126]. Interestingly, the *AQP1* gene, which encodes aquaporin-1 (AQP1), a water-transporting protein in the kidney, is expressed in both the proximal renal tubules and in the epithelium of most renal cysts in ADPKD [494, 495]. There is an inverse correlation between AQP1 expression and cyst size, implicating AQP1 in renal cystogenesis. These findings point to a possible functional role of aquaporins in both cyst formation and fluid balance, potentially influencing kidney stone formation in individuals with structural renal abnormalities.

Similarly, *UMOD*, which encodes the Tamm-Horsfall protein- an inhibitor of stone formation in its non-polymerised form (uromodulin) and a promoter of stone formation in its polymerised form (uromucoid)- is primarily expressed on the apical surface of tubular cells in the thick ascending limb of the loop of Henle [496]. Mutations in *UMOD* cause medullary cystic kidney disease type 2 [497]. Functional studies in mice deficient in Tamm-Horsfall protein have shown a dramatic increase in both the frequency and severity of renal calcium crystal formation, accompanied by upregulation of osteopontin, an inhibitor of calcium crystallisation, under high-calcium/oxalate conditions [498]. These observations suggest that additional to influencing the risk of medullary cystic kidney disease type 2, Tamm-Horsfall protein plays a critical role in susceptibility to KSD by modulating calcium crystal formation in the renal tubule.

The enrichment of these genes in KSD-associated loci underscores the importance of genetic factors involved in renal development, tubular integrity, and ion transport in modulating kidney stone risk. Moreover, it highlights the potential for shared genetic vulnerabilities in conditions such as ADPKD, medullary cystic kidney disease, and KSD, where disrupted renal architecture and ion handling may predispose individuals to stone formation.

Gene-set analyses also highlight the relevance of genes enriched in proximal tubular epithelial cells and the thick ascending limb of the loop of Henle to KSD. This may be considered unsurprising as studies investigating the cell types affected by loci

associated with kidney function in middle-aged adults localise 56% of kidney function heritability to loci regulating gene expression in kidney tubule epithelial cells [413]. However, my results provide a focus to consider whether loci identified through GWAS may demonstrate tissue-specific expression and to study their function at different locations along the nephron.

In summary, the gene-based analyses presented, based on the GWAS and meta-analyses, provide critical advances in deciphering the genetic architecture of KSD by uncovering novel loci, missense variants, and enriched genetic pathways. These findings underscore the possibility of shared biological mechanisms linking KSD and its risk factors, paving the way for future research aimed at developing targeted therapeutic strategies.

### 3.4.3 Exome-wide association study of kidney stone disease in the UK Biobank

To further explore the genetic architecture of KSD, I conducted one of the largest WES-based association studies to date, focusing on rare variants using collapsing analyses. The findings confirm the involvement of genes that have previously been associated with KSD and identify one novel association (*SLC26A2*).

The solute carrier family 26 member 2 (*SLC26A2*) gene encodes the sulphate transporter (solute carrier family 26 member 2) a sodium-independent sulphate/chloride antiporter. To my knowledge, variants in *SLC26A2* are not associated with KSD in genetic association studies. The missense variant (5:149981314:C, p.T574I) driving this association results in the substitution of the polar threonine amino acid with the medium-sized, hydrophobic isoleucine at position 574 of the protein. Although this change is predicted to be benign and the protein is not expressed in the kidney, it is expressed in the parathyroid gland, an important calcitropic organ implicated in calcium and phosphate homeostasis, which may warrant further investigation [485, 499].

Alkaline phosphatase (*ALPL*) encodes the tissue non-specific alkaline phosphatase

enzyme (ALPL), which is responsible for hydrolysing phosphomonoesters, including inorganic pyrophosphate, to inorganic phosphate [500]. This process facilitates bone mineralisation and the dephosphorylation of active vitamin B6 and phosphorylated osteopontin (OPN), among other functions [500]. Mutations in *ALPL* cause a variety of clinical presentations, ranging from stillbirth to fractures in late adulthood, due to allelic heterogeneity in the gene [500]. ALPL mutations lead to hypophosphatasia which impairs bone mineralisation due to reduced inorganic phosphate levels. Inorganic phosphate is required to bind with hydroxyapatite, a key mineral component of bone [501]. Patients with hypophosphatasia can be treated with asfotase-$\alpha$, a recombinant, bone-targeted enzyme replacement therapy.

ALPL also dephosphorylates OPN, a secreted glycoprotein expressed in the renal tubule that modulates mineral solubility in solution and is encoded by the secreted phosphoprotein 1 gene (*SPP1*) [502, 503]. In the kidney, OPN is a key protein component of the renal stone matrix and regulates the renal calcification process depending on its phosphorylation status. Phosphorylation of OPN alters its cell-membrane binding properties. Phosphorylated and unphosphorylated OPN have distinct physiological functions; when free in solution, OPN inhibits crystal growth, whereas when immobilised on surfaces, such as collagen, it promotes crystal aggregation [50, 504]. *SPP1*-null mice exhibit severe nephrocalcinosis [502–504].

Collapsing analyses of rare, missense *ALPL*-associated variants show associations with KSD, though no single variant was significantly associated in the LOVO analysis. This suggests that, at a population level, the KSD association results from the combined effects of multiple variants. At an individual level, it is plausible that a combination of *ALPL* variants alter ALPL activity, affect OPN dephosphorylation, and, in turn, modify an individual's risk of KSD. Further association and mechanistic studies are necessary to investigate this hypothesis.

As described in Chapter 1 Section 1.1.6.7, *SLC34A1* encodes the sodium-dependent phosphate transport protein 2a, NaPi-IIa, which is expressed in the proximal renal tubule. Variants in *SLC34A1* are implicated in hypophosphataemic hypercalciuric stone formation and infantile hypercalcaemia featuring nephrocalcinosis [160, 179, 181–183]. Similarly, LOVO in collapsing analyses of rare, missense

*SLC34A1*-associated variants showed that no single variant had a significant effect, suggesting that the KSD association results from the effects of multiple variants.

Rare variants in *SLC34A3* with a CADD score $\geq 25$ were also associated with KSD in ExWAS. While biallelic loss-of-function mutations in *SLC34A3* cause hereditary hypophosphataemic rickets with hypercalciuria [174–178] and previous studies have identified rare variants in *SLC34A3* that increase the risk of kidney stones [180], the LOVO analysis described in Section 3.3.6.2 did not identify any specific variant driving the association with KSD.

While I initially anticipated more striking discoveries from ExWAS, and to identify novel KSD-associated genes, I regard the ExWAS results not as a "failure" but as a valuable contribution that deepens our understanding of the complex genetic basis of KSD. My results point to the relevance of polygenic inheritance, where numerous common variants with small effect sizes contribute to the estimated $\approx 45\%$ heritability of KSD. This suggests that rare variants with large effects may play a limited role in overall disease risk at a population level. Although the study may have been underpowered to detect associations of rare variants with small to moderate effects, and additional collapsing models could be explored, my findings imply that rare variants account for only a small fraction of the heritability of KSD. Moreover, exome sequencing primarily captures SNPs and short variant sequences, leaving room for future research to investigate the potential role of larger structural variants, such as copy number variants, in KSD. Overall, the UK Biobank ExWAS results underscore the inherent challenges of rare variant analysis while affirming key genetic contributors to KSD risk.

### 3.4.4 Conclusion

In my first results chapter, I have carefully performed genetic association studies using combined-sex and sex-specific analyses, European and trans-ancestry populations, and common and rare genetic variants. The findings highlight the importance of these approaches in uncovering novel genetic loci and biological pathways associated with KSD while suggesting that sex-specific genetic loci and

rare variants in protein-coding genes do not account for a large proportion of the genetic architecture of KSD. However, they also emphasise the challenges inherent in interpreting association study results. As the name implies, loci identified through GWAS represent statistical associations rather than definitive proof of causation. Nevertheless, these results provide a critical foundation for further analyses to explore overlapping risk factors and mechanisms, which will be discussed in detail in Chapters 4, 5, and 6.

# 4 | The role of adiposity in kidney stone disease

## 4.1 Introduction

Obesity is complex. First, it is a complex condition, like KSD, arising from the interplay of environmental and genetic factors [313]. Research from the Genetic Investigation of Anthropometric Traits (GIANT) consortium has identified hundreds of common genetic variants associated with obesity, anthropometric measures, and adiposity traits through combined- and sex-stratified analyses [313, 505–507]. Second, beyond its multifactorial aetiology, obesity contributes to a complex range of physical comorbidities, including KSD [119], and serves as a focal point for research spanning biological, psychological, and social disciplines. The undeniable importance of obesity in the twenty-first century has sparked extensive debate and varying definitions and classifications. To provide clarity in this chapter, where I will explore the role of adiposity in KSD, I will begin by outlining several key definitions.

### 4.1.1 Definitions

#### 4.1.1.1 Overweight and obesity

The World Health Organisation (WHO) defines overweight as a "condition of excessive fat deposits" and obesity as "excessive fat deposits that can impair health" [508]. In the UK, the prevalence of obesity among adults has almost doubled in the last 50 years, with approximately 26% of adults aged ≥18 years and older living with obesity in 2023 [509, 510]. If global trends continue, it is projected that 57.4% of men and 60.3% of women will be overweight or obese by 2050 [511]. Obesity significantly increases the risk of numerous other medical conditions, with

projections from the Lancet Commission on Obesity suggesting that between 2010 and 2030, an additional 11 million obese adults in the UK could lead to 6-8.5 million new cases of diabetes, and 5.7-7.3 million new cases of heart disease and stroke [512].

Although direct measures of body fat, such as dual-energy X-ray absorptiometry and bioimpedance analysis, are ideal [513, 514], BMI is frequently used as a practical proxy due to its simplicity and strong correlation ($r^2$=0.7-0.9) with body fat [515–518]. Consequently, overweight and obesity are commonly classified based on BMI thresholds. In this chapter, "overweight" and "obesity" are used according to established classifications of BMI.

#### 4.1.1.2 Body mass index

BMI is a "surrogate marker of fatness" and is calculated as BMI=weight (kg) / height$^2$ (m$^2$). Overweight is defined as a BMI $\geq$25.0-29.9 kg/m$^2$, and obesity as a BMI $\geq$30.0-34.9 kg/m$^2$ in adult men and women, although BMI differs considerably by ethnicity [508]. The heritability of BMI is estimated at 31-90% depending on population and socioeconomic factors [519, 520]. BMI is widely used to identify individuals at risk of adverse health outcomes due to excess body fat. However, there is substantial inter-individual variability in the relationship between BMI and poor health outcomes [521] as BMI reflects both fat mass and fat-free mass without accounting for fat distribution [516–518]. Despite its limitations, BMI remains a useful tool for public health purposes.

#### 4.1.1.3 Adiposity

Adiposity refers to the degree of body fat accumulation [521]. Excess adiposity drives pathogenic responses in adipocytes and adipose tissue, leading to metabolic diseases such as type 2 diabetes, hypertension, dyslipidaemia, and metabolic syndrome (the co-occurrence of three or more of these conditions). Positive energy balance, resulting from the combination of an energy-dense diet and sedentary lifestyle, leads to adipocyte hypertrophy and visceral fat accumulation, particularly in those with

impaired subcutaneous adipose tissue adipogenesis [521].

Adipocyte hypertrophy contributes to mitochondrial and endoplasmic reticulum dysfunction, hormone dysregulation, impaired fatty acid storage, dyslipidaemia, and lipotoxicity in non-adipose tissues [521]. Insufficient subcutaneous adipogenesis exacerbates this process, increasing free fatty acid delivery to the liver, heart, and perivascular depots. Impaired vascularisation of growing adipose tissue causes hypoxia and immune dysregulation, resulting in oxidative stress and inflammatory responses. These processes, collectively termed "adiposopathy" or "sick fat", drive and exacerbate metabolic diseases, ultimately increasing the risk of cardiovascular disease and mortality [521, 522].

Direct measurements of fat distribution, using CT or magnetic resonance imaging (MRI), provide detailed insights but are limited by cost, logistics, and, in the case of CT, radiation exposure [523]. Anthropometric measurements, such as WHR, offer a practical alternative and correlate strongly ($r^2$=0.55-0.75) with intra-abdominal fat [515, 524]. In this chapter, "adiposity" refers to excess body fat and its associated metabolic consequences.

#### 4.1.1.4 Waist-to hip ratio

WHR, calculated as waist circumference divided by hip circumference, serves as a surrogate measure of central adiposity. It is influenced by factors such as gender, age, and body composition. The heritability of WHR is estimated as ≈30-60% , even after adjusting for BMI [505, 525]. While WHR provides an index of visceral and subcutaneous adipose tissue, evidence for ethnic group-specific cut-offs remains insufficient [526] and thresholds defining increased central adiposity are gender-specific, specifically WHR $\geq$ 0.85 for women and $\geq$0.90 for men [508, 527].

### 4.1.2 The relevance of adiposity to kidney stone disease

In Chapter 1 Section 1.1.6.3, I outlined the observational and MR evidence for obesity as a risk factor for KSD; increasing BMI and WHR are independent,

causal risk factors for KSD in males and females [98, 101, 119]. Despite these insights, the mechanisms underlying the relationship between adiposity and KSD are undefined, though could point to opportunities to improve therapeutic and preventive strategies for patients with KSD. Furthermore, it is unclear whether therapies affecting adipose tissues, for instance glucagon-like peptide-1 receptor agonists (GLP1-RA), might effectively prevent KSD, and in which patient cohorts such therapies might be most beneficial [315, 528–530].

### 4.1.3 Aim and Objectives

In this chapter my main research aim is to discover potential regions of the genome where adiposity and KSD have a shared genetic architecture and leverage findings to guide genomics-driven drug discovery in KSD. Specifically, my research objectives are to:

1. **Perform regional Mendelian randomisation**
    - Use genetic instruments for BMI and WHR to establish causal effects of adiposity on risk of KSD across different genomic regions.
2. **Perform regional colocalisation**
    - Identify genomic regions where genetic variants influence adiposity-related traits and KSD risk.
3. **Explore therapeutic implications of altering adiposity using drug-target Mendelian randomisation**
    - Highlight key genes or pathways implicated in adiposity traits and KSD as potential targets for future therapeutic interventions.

## 4.2 Methods

### 4.2.1 Data sources

#### 4.2.1.1 Adiposity traits

To perform MR and colocalisation analyses, I identified summary statistics from well-powered GWAS of BMI and WHR in individuals of European ancestry. Pulit et al. performed a meta-analysis of GWAS of BMI and of WHR, combining data from the UK Biobank and the GIANT consortium [313]. Analyses of 806,834 participants revealed 670 independent variants significantly associated with BMI at the GWAS significance threshold (p-value$< 5 \times 10^{-8}$) and 316 independent variants significantly associated with WHR from 697,734 participants.

#### 4.2.1.2 Kidney stone disease

Three KSD outcome datasets are considered in this chapter. Primarily, I used results from the European-ancestry KSD GWAS meta-analysis (Chapter 3 Section 3.3.2). As sensitivity analyses, I used the two studies that comprised the meta-analysis, specifically the combined-sex KSD GWAS data from the UK Biobank (Chapter 3 Section 3.3.1), and the publicly available summary statistics from the FinnGen study, release 12 [433].

### 4.2.2 Mendelian randomisation

To ascertain the regional effects of genetically-predicted BMI and WHR on odds of KSD, I employed a regional MR approach and systematically considered the effects of 1 Mbp genomic areas ±500 Kbp of lead independent variants associated with BMI or WHR at GWAS [313] as delineated in Chapter 2 Section 2.4.

### 4.2.3 Colocalisation

Where I identified potential causal effects of BMI or WHR arising from pre-specified genomic regions on the odds of KSD, I used colocalisation analyses to evaluate the probability of a shared causal variant underlying the two traits. Where there was evidence of causality and colocalisation between BMI and KSD or WHR and KSD at a genomic region, I performed multi-trait colocalisation to establish whether all three traits shared an underlying causal genetic variant. Colocalisation was performed as denoted in Chapter 2 Section 2.5.

### 4.2.4 Drug-target Mendelian randomisation

To identify potential therapeutic pathways that could be modulated to prevent KSD, I considered the gene(s) likely associated with candidate causal variants identified from multi-trait colocalisation in Chapter 4 Section 4.2.3 and undertook drug-target MR analyses based on the gene's genomic coordinates ±300 Kbp as described in Chapter 2 Section 2.4.

## 4.3 Results

### 4.3.1 Genetic variants increase the risk of kidney stone disease via alterations in body mass index or waist-to-hip ratio

Summary statistics from European ancestry GWAS meta-analysis of KSD, UK Biobank GWAS, and FinnGen GWAS, BMI, and WHR were used to identify 11 and 12 different 1 Mbp regions where increasing BMI and WHR, respectively, causally increases odds of KSD (Table 4.1).

***Table 4.1. Number of 1 megabase regions involved in evaluating the causal effect of adiposity on odds of kidney stone disease.***

| Analysis step | Dataset | Body mass index | Waist-to-hip ratio |
|---|---|---|---|
| Independent SNPs in GWAS | Meta-analysis | 670 | 316 |
| 1Mbp genomic regions with > 2 SNPs to assess in MR | Meta-analysis | 225 | 114 |
| | UK Biobank | 235 | 150 |
| | FinnGen | 230 | 147 |
| Significant 1Mbp regions in MR | Meta-analysis | 9 | 9 |
| | UK Biobank | 4 | 9 |
| | FinnGen | 6 | 8 |
| Significant 1Mbp regions in MR with evidence of colocalisation | Meta-analysis | 4 | 3 |
| | UK Biobank | 1 | 0 |
| | FinnGen | 3 | 4 |

GWAS=Genome-wide association study; SNP=Single nucleotide polymorphism; Mbp=Megabase pairs; MR=Mendelian randomisation.

Colocalisation analyses identified four 1 Mbp regions across four loci with shared, causal variants that explained liability to both higher BMI and KSD (Table 4.2, Figure 4.1), and four 1 Mbp regions across four loci with shared, causal variants explaining liability to both higher WHR and KSD (Table 4.3, Figure 4.2). Notably, overlap occurred at *TFAP2B*-associated regions (BMI-associated region: rs2206277 ±500 Kbp; WHR-associated region: rs987237 ±500 Kbp) and the *FTO*-associated regions (BMI-associated region: rs2072518 ±500 Kbp; WHR-associated regions: rs9923544 ±500 Kbp and rs9302652 ±500 Kbp) and rs1558902 was identified as a candidate causal variant underlying the effect of increasing BMI and of increasing WHR on risk of KSD at the *FTO*-associated region.

***Table 4.2. Genomic regions of 1 megabase where body mass index increases odds of kidney stone disease.***

| IV: 1 Mbp region from GWAS | | | | MR analysis | | | | | | | | | Colocalisation | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Index variant (Gene) | CHR | POS | Mean $R^2$ (SD) | Inverse variance weighted analysis | | | | Heterogeneity | | | Intercept | | N SNPs | PP H4 | Candidate SNP | SNP PP |
| | | | | Outcome | N SNPs | OR (95% CI) | P | Q | Q_df | Q P | $\beta$ (SE) | P | | | | |
| rs34811474 *(ANAPC4)* | 4 | 25408838 | $9.83\times10^{-5}$ ($7.72\times10^{-5}$) | Meta-analysis | 3 | 3.16 (1.68-5.96) | 0.01 | 1.47 | 2 | 0.48 | -0.02 (0.02) | 0.49 | 3352 | 0.78 | rs34811474 | 1.00 |
| | | | | UK Biobank | 3 | 2.13 (0.83-5.46) | 0.48 | 2.12 | 2 | 0.35 | -0.03 (0.03) | 0.44 | 3111 | 0.06 | | |
| | | | | FinnGen | 3 | 4.55 (1.88-11.02) | 0.03 | 0.13 | 2 | 0.94 | -0.01 (0.03) | 0.84 | 3254 | 0.39 | | |
| rs2206277 *(TFAP2B)* | 6 | 50798526 | $1.32\times10^{-4}$ ($1.23\times10^{-4}$) | Meta-analysis | 10 | 6.28 (3.94-10.02) | $1.25\times10^{-12}$ | 17.49 | 9 | 0.04 | 0.02 (0.01) | 0.14 | 2841 | 0.92 | rs2207139 | 0.45 |
| | | | | UK Biobank | 11 | 3.51 (1.97-6.27) | $2.47\times10^{-3}$ | 16.05 | 10 | 0.10 | 0.01 (0.02) | 0.68 | 2747 | 0.56 | | |
| | | | | FinnGen | 11 | 6.89 (4.05-11.72) | $1.33\times10^{-10}$ | 16.60 | 10 | 0.08 | 0.04 (0.02) | 0.04 | 2803 | 0.83 | rs2207139 | 0.46 |
| rs2072518 *(FTO)* | 16 | 54152730 | $2.29\times10^{-4}$ ($5.13\times10^{-4}$) | Meta-analysis | 17 | 2.18 (1.79-2.64) | $5.62\times10^{-13}$ | 9.29 | 16 | 0.90 | 0.01 (0.01) | 0.32 | 3451 | 0.99 | rs1558902 | 0.95 |
| | | | | UK Biobank | 14 | 1.62 (1.21-2.18) | 0.05 | 3.54 | 13 | 1.00 | -0.01 (0.01) | 0.46 | 3084 | 0.58 | | |
| | | | | FinnGen | 17 | 2.73 (2.11-3.53) | $4.10\times10^{-12}$ | 11.61 | 16 | 0.77 | 0.01 (0.01) | 0.07 | 3261 | 0.96 | rs1558902 | 0.95 |
| rs11672660 *(GIPR)* | 19 | 46180184 | $8.19\times10^{-5}$ ($1.00\times10^{-4}$) | Meta-analysis | 7 | 5.38 (3.36-8.62) | $1.91\times10^{-10}$ | 5.71 | 6 | 0.46 | -0.02 (0.01) | 0.13 | 3017 | 1.00 | rs1800437 | 0.64 |
| | | | | UK Biobank | 7 | 5.05 (2.47-10.33) | $2.09\times10^{-3}$ | 2.40 | 6 | 0.88 | -0.01 (0.01) | 0.35 | 2813 | 0.91 | rs1800437 | 0.70 |
| | | | | FinnGen | 7 | 5.64 (3.02-10.55) | $4.62\times10^{-6}$ | 5.21 | 6 | 0.52 | -0.02 (0.01) | 0.19 | 2845 | 0.99 | rs1800437 | 0.62 |

CHR=Chromosome; GWAS=Genome-wide association study; Mbp=Megabase pair; MR=Mendelian randomisation; N SNPs=Number of SNPs; OR=Odds ratio for kidney stone disease per genetically-instrumented standard deviation increase in trait; P=5% false discovery rate adjusted p-value; POS=Position based on NCBI Genome Build 37 (hg19); PP H4=Posterior probability of full colocalisation; Q=Q statistic; Q_df=Q degrees of freedom; $R^2$=Variance in exposure explained by instrumental variable; SD=Standard deviation; SE=Standard error; SNP PP=Posterior probability explained by SNP; 95% CI=95% confidence interval.

***Table 4.3. Genomic regions of 1 megabase where waist-to-hip ratio increases odds of kidney stone disease.***

| IV: 1 Mbp region from GWAS | | | | MR analysis | | | | | | | | | Colocalisation | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Index variant (Gene) | CHR | POS | Mean $R^2$ (SD) | Inverse variance weighted analysis | | | | Heterogeneity | | | Intercept | | N SNPs | PP H4 | Candidate SNP | SNP PP |
| | | | | Outcome | N SNPs | OR (95% CI) | P | Q | Q_df | Q P | $\beta$ (SE) | P | | | | |
| rs2894204 *(HLA-C)* | 6 | 31237061 | $8.17 \times 10^{-5}$ ($3.70 \times 10^{-5}$) | Meta-analysis | 15 | 4.65 (2.49-8.68) | $4.01 \times 10^{-5}$ | 41.24 | 14 | $1.63 \times 10^{-4}$ | -0.01 (0.02) | 0.50 | 9058 | 0.25 | | |
| | | | | UK Biobank | 16 | 4.36 (2.28-8.35) | $1.35 \times 10^{-3}$ | 28.79 | 15 | 0.02 | -0.01 (0.02) | 0.47 | 9110 | $7.81 \times 10^{-4}$ | | |
| | | | | FinnGen | 14 | 5.71 (2.71-12.01) | $3.36 \times 10^{-4}$ | 26.67 | 13 | 0.01 | -0.02 (0.03) | 0.51 | 8475 | 0.73 | rs2524139 | 0.43 |
| rs987237 *(TFAP2B)* | 6 | 50803050 | $7.92 \times 10^{-5}$ ($2.83 \times 10^{-5}$) | Meta-analysis | 3 | 52.28 (22.89-119.36) | $8.44 \times 10^{-19}$ | 1.19 | 2 | 0.55 | 0.03 (0.03) | 0.47 | 2845 | 0.92 | rs2207139 | 0.22 |
| | | | | UK Biobank | 4 | 10.19 (1.41-73.52) | 0.21 | 10.92 | 3 | 0.01 | -0.03 (0.09) | 0.78 | 2751 | 0.58 | | |
| | | | | FinnGen | 4 | 59.33 (20.22-174.07) | $1.53 \times 10^{-11}$ | 3.51 | 3 | 0.32 | $1.96 \times 10^{-3}$ (0.05) | 0.97 | 2806 | 0.83 | rs2207139 | 0.25 |
| rs9923544 *(FTO)* | 16 | 53801985 | $1.51 \times 10^{-4}$ ($2.45 \times 10^{-4}$) | Meta-analysis | 8 | 3.65 (2.22-6.02) | $1.78 \times 10^{-5}$ | 13.03 | 7 | 0.07 | -0.02 (0.01) | 0.05 | 2956 | 0.99 | rs1421085 | 0.50 |
| | | | | UK Biobank | 8 | 2.44 (1.43-4.16) | 0.03 | 2.17 | 7 | 0.95 | -0.02 (0.01) | 0.22 | 2662 | 0.58 | | |
| | | | | FinnGen | 8 | 5.22 (2.4-11.33) | $1.10 \times 10^{-3}$ | 16.57 | 7 | 0.02 | -0.03 (0.02) | 0.17 | 2754 | 0.96 | rs1558902 | 0.50 |
| rs9302652 *(FTO)* | 16 | 53865975 | $1.51 \times 10^{-4}$ ($2.45 \times 10^{-4}$) | Meta-analysis | 8 | 3.65 (2.22-6.02) | $1.78 \times 10^{-5}$ | 13.03 | 7 | 0.07 | -0.02 (0.01) | 0.05 | 2992 | 0.99 | rs1421085 | 0.50 |
| | | | | UK Biobank | 8 | 2.44 (1.43-4.16) | 0.03 | 2.17 | 7 | 0.95 | -0.02 (0.01) | 0.22 | 2688 | 0.58 | | |
| | | | | FinnGen | 8 | 5.22 (2.4-11.33) | $1.10 \times 10^{-3}$ | 16.57 | 7 | 0.02 | -0.03 (0.02) | 0.17 | 2788 | 0.96 | rs1558902 | 0.50 |

CHR=Chromosome; GWAS=Genome-wide association study; Mbp=Megabase pair; MR=Mendelian randomisation; N SNPs=Number of SNPs; OR=Odds ratio for kidney stone disease per genetically-instrumented standard deviation increase in trait; P=5% false discovery rate adjusted p-value; POS=Position based on NCBI Genome Build 37 (hg19); PP H4=Posterior probability of full colocalisation; Q=Q statistic; Q_df=Q degrees of freedom; $R^2$=Variance in exposure explained by instrumental variable; SD=Standard deviation; SE=Standard error; SNP PP=Posterior probability explained by SNP; 95% CI=95% confidence interval.

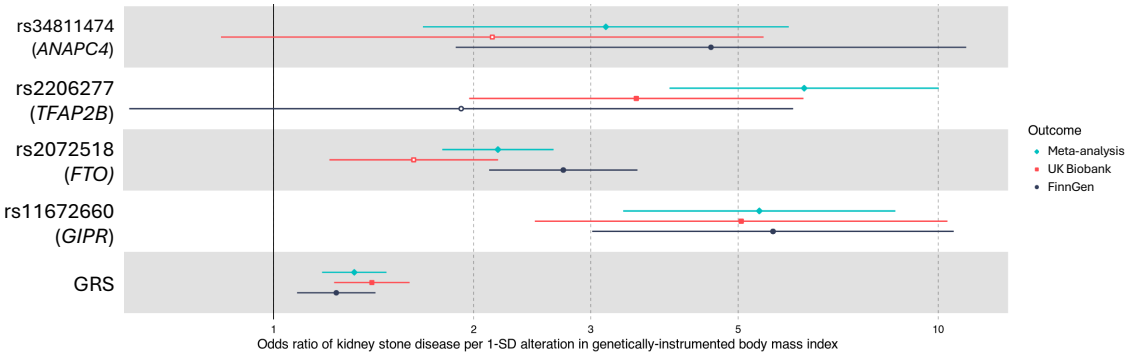



***Figure 4.1. Forest plot demonstrating regional Mendelian randomisation estimates for independent body mass index-associated variants (±500 kilobase pairs) on odds of kidney stone disease.***

Odds ratio is scaled per 1 standard deviation increase in body mass index ($\approx 4.80 kg/m^2$ [313]).

GRS=Pan-genome Mendelian randomisation estimate for body mass index on risk of kidney stone disease; KSD=kidney stone disease.

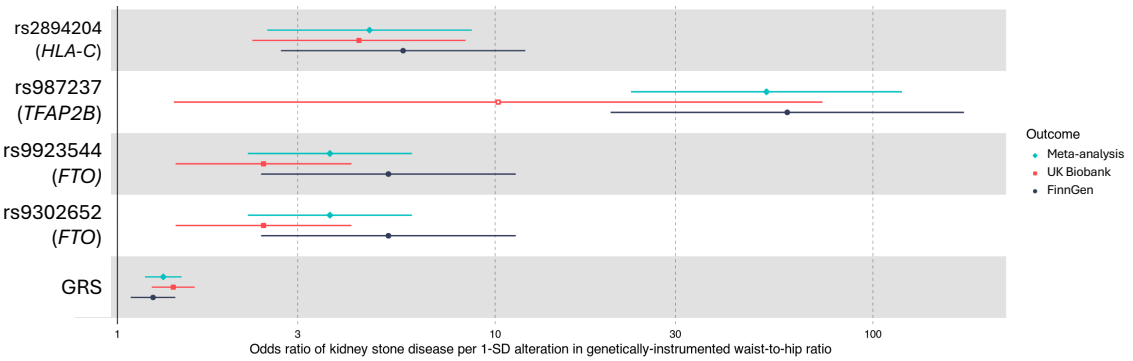



***Figure 4.2. Forest plot demonstrating regional Mendelian randomisation estimates for independent waist-to-hip ratio-associated variants (±500 kilobase pairs) on odds of kidney stone disease.***

Odds ratio is scaled per 1 standard deviation increase in waist-to-hip ratio (≈0.09) [313].

GRS=Pan-genome Mendelian randomisation estimate for body mass index on risk of kidney stone disease; KSD=Kidney stone disease.

BMI and WHR are highly correlated [531]. To increase confidence in putative causal variants and to account for the shared genetic architecture underlying these traits, I conducted multi-trait colocalisation analysis (HyPrColoc) for BMI, WHR, and KSD for the eight 1 Mbp regions that showed evidence of a causal effect and colocalisation in BMI-KSD or WHR-KSD analyses (Table 4.2 and Table 4.3). Full colocalisation of adiposity traits and KSD was detected at three of these eight regions (*TFAP2B*, *FTO*, *GIPR*; Table 4.4 and Table 4.5, Figure 4.3).

***Table 4.4. Genomic regions of 1 megabase with evidence of multi-trait colocalisation between body-mass index, waist-to-hip ratio, and kidney stone disease.***

| Index variant | CHR | POS | Gene | KSD study | PP H4 | Candidate SNP | SNP PP |
|---|---|---|---|---|---|---|---|
| rs2206277 | 6 | 50798526 | *TFAP2B* | Meta-analysis | 1.00 | rs2207139 | 0.70 |
| | | | | UK Biobank | 0.95 | | 0.71 |
| | | | | FinnGen | 1.00 | | 0.71 |
| rs987237 | 6 | 50803050 | *TFAP2B* | Meta-analysis | 1.00 | rs2207139 | 0.70 |
| | | | | UK Biobank | 0.95 | | 0.71 |
| | | | | FinnGen | 1.00 | | 0.71 |
| rs9923544 | 16 | 53801985 | *FTO* | Meta-analysis | 1.00 | rs1558902 | 0.95 |
| | | | | UK Biobank | 0.92 | | 0.95 |
| | | | | FinnGen | 1.00 | | 0.95 |
| rs9302652 | 16 | 53865975 | *FTO* | Meta-analysis | 1.00 | rs1558902 | 0.95 |
| | | | | UK Biobank | 0.92 | | 0.95 |
| | | | | FinnGen | 1.00 | | 0.95 |
| rs2072518 | 16 | 54152730 | *FTO* | Meta-analysis | 1.00 | rs1558902 | 0.95 |
| | | | | UK Biobank | 0.92 | | 0.95 |
| | | | | FinnGen | 1.00 | | 0.95 |
| rs11672660 | 19 | 46180184 | *GIPR* | Meta-analysis | 1.00 | rs1800437 | 0.81 |
| | | | | UK Biobank | 0.99 | | 0.85 |
| | | | | FinnGen | 1.00 | | 0.80 |

CHR=Chromosome; KSD=Kidney stone disease; POS=Position based on NCBI Genome Build 37 (hg19); PP=Posterior-probability; SNP=Single nucleotide polymorphism.

***Table 4.5. 95% credible sets for genomic regions with evidence for causal effects on kidney stone risk and colocalisation with adiposity traits.***

| Index variant | CHR | POS | Gene | KSD study | Candidate SNP | SNP PP |
|---|---|---|---|---|---|---|
| rs2206277 | 6 | 50798526 | *TFAP2B* | Meta-analysis | rs2207139 | 0.70 |
| | | | | | rs943005 | 0.13 |
| | | | | | rs4715213 | 0.09 |
| | | | | | rs12529728 | 0.04 |
| | | | | UK Biobank | rs2207139 | 0.71 |
| | | | | | rs943005 | 0.13 |
| | | | | | rs4715213 | 0.08 |
| | | | | | rs12529728 | 0.05 |
| | | | | FinnGen | rs2207139 | 0.71 |
| | | | | | rs943005 | 0.12 |
| | | | | | rs4715213 | 0.10 |
| | | | | | rs12529728 | 0.04 |
| rs987237 | 6 | 50803050 | *TFAP2B* | Meta-analysis | rs2207139 | 0.70 |
| | | | | | rs943005 | 0.13 |
| | | | | | rs4715213 | 0.09 |
| | | | | | rs12529728 | 0.04 |
| | | | | UK Biobank | rs2207139 | 0.71 |
| | | | | | rs943005 | 0.13 |
| | | | | | rs4715213 | 0.08 |
| | | | | | rs12529728 | 0.05 |
| | | | | FinnGen | rs2207139 | 0.71 |
| | | | | | rs943005 | 0.12 |
| | | | | | rs4715213 | 0.10 |
| | | | | | rs12529728 | 0.04 |
| rs9923544 | 16 | 53801985 | *FTO* | Meta-analysis | rs1558902 | 0.95 |
| | | | | UK Biobank | rs1558902 | 0.95 |
| | | | | FinnGen | rs1558902 | 0.95 |
| rs9302652 | 16 | 53865975 | *FTO* | Meta-analysis | rs1558902 | 0.95 |
| | | | | UK Biobank | rs1558902 | 0.95 |
| | | | | FinnGen | rs1558902 | 0.95 |
| rs2072518 | 16 | 54152730 | *FTO* | Meta-analysis | rs1558902 | 0.95 |
| | | | | UK Biobank | rs1558902 | 0.95 |
| | | | | FinnGen | rs1558902 | 0.95 |
| rs11672660 | 19 | 46180184 | *GIPR* | Meta-analysis | rs1800437 | 0.81 |
| | | | | | rs11672660 | 0.19 |
| | | | | UK Biobank | rs1800437 | 0.85 |
| | | | | | rs11672660 | 0.15 |
| | | | | FinnGen | rs1800437 | 0.80 |
| | | | | | rs11672660 | 0.20 |

CHR=Chromosome; KSD=Kidney stone disease; POS=Position based on NCBI Genome Build 37 (hg19); PP=Posterior-probability; SNP=Single nucleotide polymorphism.

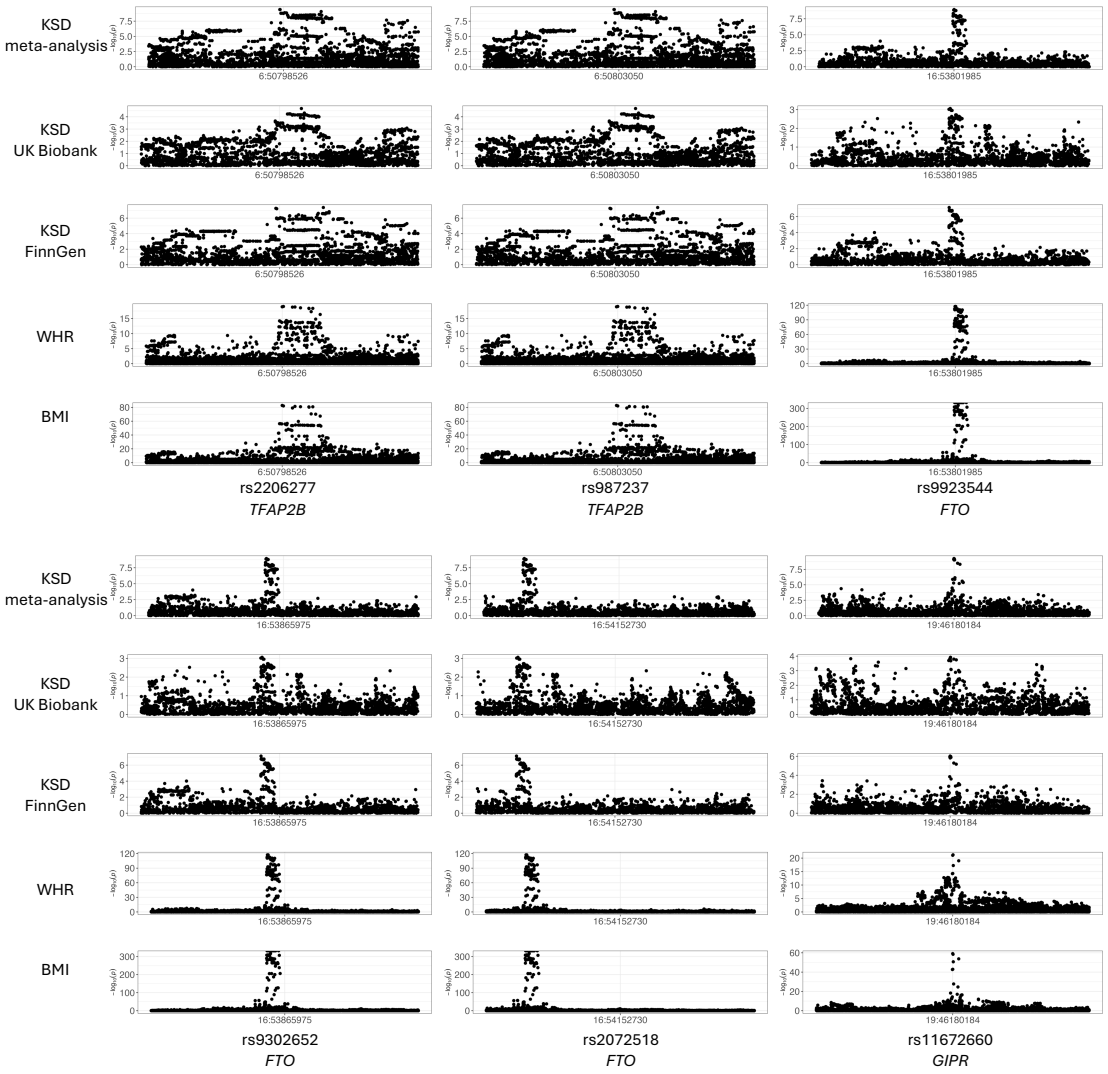



***Figure 4.3. Locus-zoom plots of 1 megabase pair regions with multi-trait colocalisation between body mass index, waist-to-hip ratio, and kidney stone disease.***

Genome-wide P values (-log10) are plotted against respective positions on each of chromosome. The central line annotated on each plot corresponds to the index variant from adiposity genome-wide association study. BMI=Body mass index; KSD=Kidney stone disease; WHR=Waist-to-hip ratio.

The three candidate causal variants comprising the majority of the 95% credible set for colocalisation at each locus include an intergenic variant ≈ 30 Kbp downstream of *TFAP2B* (rs2207139), an intronic *FTO* variant (rs1558902), and a missense *GIPR* variant (rs1800437, p.E354Q) (Table 4.6).

***Table 4.6. Associations of candidate causal variants identified from HyPrColoc with adiposity traits and kidney stone disease.***

| | Gene | *TFAP2B* | *FTO* | *GIPR* |
|---|---|---|---|---|
| **Candidate causal variant** | **CHR** | 6 | 16 | 19 |
| | **POS** | 50845490 | 53803574 | 46181392 |
| | **SNP** | rs2207139 | rs1558902 | rs1800437 |
| | **EA** | G | A | G |
| | **NEA** | A | T | C |
| | **EAF** | 0.81 | 0.41 | 0.23 |
| **BMI GWAS [313]** | **$\beta$ (SE)** | 0.04 ($2.10 \times 10^{-3}$) | 0.08 ($1.60 \times 10^{-3}$) | 0.04 ($2.10 \times 10^{-3}$) |
| | **P** | $4.58 \times 10^{-82}$ | Inf* | $3.32 \times 10^{-59}$ |
| **WHR GWAS [313]** | **$\beta$ (SE)** | 0.02 ($2.20 \times 10^{-3}$) | 0.04 ($1.70 \times 10^{-3}$) | 0.02 ($2.20 \times 10^{-3}$) |
| | **P** | $1.50 \times 10^{-10}$ | $3.99 \times 10^{-118}$ | $6.23 \times 10^{-22}$ |
| **KSD GWAS Meta-analysis** | **OR (95%CI)** | 1.07 (1.05-1.10) | 1.06 (1.04-1.08) | 1.07 (1.05-1.09) |
| | **P** | $3.73 \times 10^{-9}$ | $1.32 \times 10^{-9}$ | $6.28 \times 10^{-10}$ |
| **KSD UK Biobank GWAS** | **OR (95%CI)** | 1.06 (1.03-1.10) | 1.05 (1.02-1.08) | 1.07 (1.03-1.10) |
| | **P** | $6.30 \times 10^{-4}$ | $9.50 \times 10^{-4}$ | $1.20 \times 10^{-4}$ |
| **KSD FinnGen GWAS** | **OR (95%CI)** | 1.08 (1.05-1.11) | 1.07 (1.04-1.10) | 1.07 (1.04-1.10) |
| | **P** | $1.33 \times 10^{-6}$ | $1.93 \times 10^{-7}$ | $1.24 \times 10^{-6}$ |

BMI=Body mass index; CHR=Chromosome; GWAS=Genome-wide association study; KSD=Kidney stone disease; P=P-value; POS=Position based on NCBI Genome Build 37 (hg19); OR=Odds ratio; SE=Standard error; WHR=Waist-to-hip ratio; 95% CI=95% confidence interval; * P-value denoted as "0" in publicly available summary statistics [313].

*TFAP2B* encodes the transcription factor AP-2-$\beta$ (TFAP2B), a member of the activating protein-2 (AP-2) family, which functions as both a transcriptional activator and repressor to regulate the expression of genes essential for ectodermal tissue development, for example the neural crest and skin. It is also implicated in breast cancer and patent ductus arteriosus [532–534]. Members of the AP-2 family differ in their expression patterns and binding affinities for various genetic promoters. Specifically, TFAP2B is expressed in mesenchymal, corneal, and renal tissues [535], and mutations in the gene cause Char syndrome, an autosomal dominant disorder characterised by patent ductus arteriosus, facial dysmorphism, and hand anomalies

[536]. Variants in *TFAP2B* are associated with KSD in the European ancestry KSD GWAS meta-analysis (Chapter3 Section 3.3.2, Table 3.4, Figure 3.5), the trans-ancestry KSD GWAS meta-analysis (Chapter3 Section 3.3.3, Table 3.5, Figure 3.7), and in published GWAS meta-analyses comprising participants of European ancestry [243].

*FTO* encodes $\alpha$-ketoglutarate-dependent dioxygenase FTO, a fat mass and obesity-associated protein central to the genetic basis of adiposity [312]. Common variants associated with *FTO* were first identified in the 2007 GWAS for obesity traits [537, 538], and the gene has since become one of the most extensively studied loci in genetic research. While numerous studies link *FTO* with obesity, adiposity, energy expenditure, and food intake and demonstrate gene×environment interactions, evidence for mechanistic causality of variants within this ≈420 Kbp region remains limited [312]. The *FTO* locus influences gene expression across a range of tissues, and the genetic architecture of disease-associated loci likely involves considerable pleiotropy and allelic heterogeneity across each tissue [539]. Variants in *FTO* were associated with KSD in the European ancestry KSD GWAS meta-analysis (Chapter3 Section 3.3.2, Table 3.4, Figure 3.5) and the trans-ancestry KSD GWAS meta-analysis (Chapter3 Section 3.3.3, Table 3.5, Figure 3.7).

*GIPR* encodes the gastric inhibitory polypeptide receptor (GIPR), a G-protein coupled receptor primarily expressed in the gastrointestinal tract that is activated by gastric inhibitory polypeptide (GIP), an incretin hormone secreted by enteroendocrine K-cells in the duodenum and jejunum [540, 541]. The relationship between BMI, WHR, and GIPR is well-recognised. GIPR activation promotes pancreatic beta-cell function by potentiating insulin secretion and beta-cell proliferation, enhancing insulin sensitivity, and altering food intake behaviours to affect glycaemic control and body weight [542]. This pathway is targeted by medications used in obesity and diabetes, for example GIPR-agonists and dipeptidyl peptidase IV inhibitors (DPP4i), such as tirzepatide and sitagliptin, respectively. These prevent incretin hormone inactivation, lowering blood glucose levels and, in the case of GIPR-agonists, promoting weight loss. Variants in *GIPR* were associated with KSD in the European ancestry KSD GWAS meta-analysis

(Chapter3 Section 3.3.2, Table 3.4, Figure 3.5), the trans-ancestry KSD GWAS meta-analysis (Chapter3 Section 3.3.3, Table 3.5, Figure 3.7), and in published GWAS meta-analyses comprising participants of European ancestry [119, 243].

### 4.3.2 A candidate causal variant in *GIPR* has functional effects

The *GIPR* predicted causal variant, rs1800437 (p.E354Q), is a missense variant located in the constitutively-active transmembrane domain 6 of GIPR. It is predicted to be "tolerated" and "probably damaging" by SIFT and PolyPhen algorithms, respectively, with a scaled CADD score of 34 (likely deleterious) and is denoted to have "ambiguous" pathogenicity by alpha-missense [543]. The G allele results in a glutamine (amino acid code $Q$) to glutamate (amino acid code $E$) substitution at position 354 and is associated with higher BMI and WHR [313, 544] and greater risk of KSD in GWAS data (Table 4.6). rs1800437-G may account for 5-12% of KSD in the UK Biobank (Table 4.7).

***Table 4.7. Kidney stone population attributable risk and population attributable fraction for rs1800437-G in the UK Biobank.***

| Model | | Risk allele carrier | No risk allele | KSD prevalence | KSD prevalence carriers | KSD prevalence non-carriers | PAR | PAF (%) |
|---|---|---|---|---|---|---|---|---|
| Dominant | | GG or GC | CC | 0.029 | 0.029 | 0.025 | $3.54 \times 10^{-3}$ | 12.35 |
| | Cases | 11,296 | 390 | | | | | |
| | Controls | 381,092 | 15,141 | | | | | |
| Recessive | | GG | GC or CC | 0.029 | 0.029 | 0.027 | $1.36 \times 10^{-3}$ | 4.73 |
| | Cases | 7,799 | 3,887 | | | | | |
| | Controls | 257,697 | 138,536 | | | | | |

KSD=Kidney stone disease ; PAF=Population attributable fraction (calculated as (prevalence in total population- prevalence in individuals without risk allele)/ prevalence in total population); PAR=Population attributable risk (calculated as prevalence in total population- prevalence in individuals without risk allele).

Mice with the protective rs1800437-C allele (GIPR-Q354) are leaner and have better glucose tolerance than GIPR-E354 mice [545]. Studies in murine B-cells

show that the protective GIPR variant Q354, compared to E354, leads to slower receptor recycling back to the cell surface and increased receptor accumulation in the trans-Golgi network after GIP stimulation [545]. This altered post-activation trafficking of GIPR suggests that the p.E354Q missense variant has functional effects.

### 4.3.3 Modulating adiposity at *GIPR* may prevent kidney stone disease

Given the functional effects of rs1800437, the relevance of *GIPR* to adiposity traits and KSD, and the dramatic increase in GIPR-agonist prescribing in recent years [546, 547], I considered whether activating this pathway could reduce the risk of KSD (Table 4.8, Figure 4.4).

***Table 4.8. Drug target Mendelian randomisation; effects of modulating adiposity traits on odds of kidney stone disease.***

| Analysis | | | | IVW MR | | Heterogeneity | | | Intercept | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Exposure** | **N SNPs** | **$R^2$ (SD)** | **Outcome** | **OR (95% CI)** | **P** | **Q** | **Q_df** | **P** | **$\beta$** | **P** |
| BMI | 6 | $1.01 \times 10^{-4}$ ($1.21 \times 10^{-4}$) | Meta-analysis | 0.17 (0.10-0.27) | $5.24 \times 10^{-12}$ | 3.27 | 5 | 0.66 | 0.01 (0.01) | 0.27 |
| | | | UK Biobank | 0.18 (0.08-0.37) | $1.17 \times 10^{-5}$ | 1.43 | 5 | 0.92 | 0.01 (0.02) | 0.53 |
| | | | FinnGen | 0.16 (0.08-0.30) | $9.24 \times 10^{-8}$ | 3.73 | 5 | 0.59 | 0.01 (0.01) | 0.34 |
| WHR | 4 | $9.23 \times 10^{-5}$ ($4.52 \times 10^{-5}$) | Meta-analysis | 0.18 (0.04-0.92) | 0.05 | 16.44 | 3 | 0.00 | 0.05 (0.04) | 0.35 |
| | | | UK Biobank | 0.19 (0.05-0.83) | 0.04 | 6.01 | 3 | 0.11 | 0.03 (0.04) | 0.48 |
| | | | FinnGen | 0.17 (0.03-1.11) | 0.06 | 11.84 | 3 | 0.01 | 0.06 (0.05) | 0.29 |

CHR=Chromosome; IVW MR=Inverse-variance weighted Mendelian randomisation; P=5% false discovery rate adjusted p-value; N SNPs=Number of SNPs; Q=Q statistic; Q_df=Q degrees of freedom; $R^2$=Variance in exposure explained by instrumental variable; SD=Standard deviation; SE=Standard error.

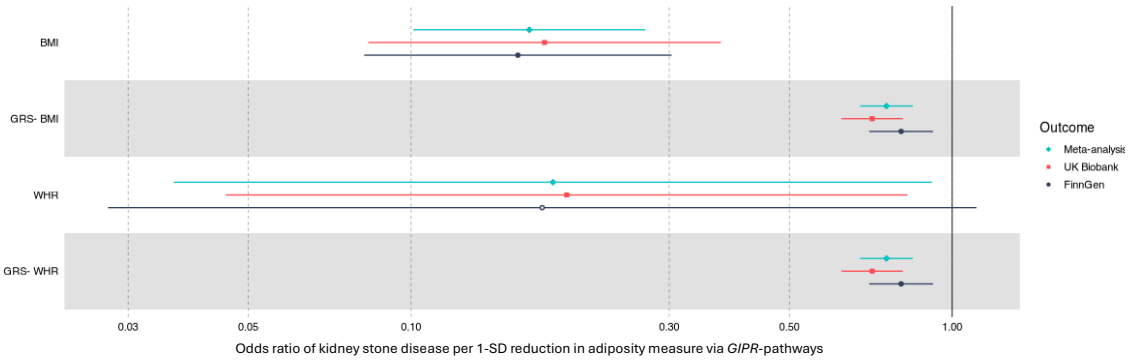



***Figure 4.4. Forest plot demonstrating drug-target Mendelian randomisation estimates for the effects of modulating body mass index and waist-to-hip ratio via GIPR.***

Non-significant results are indicated by hollow symbols. Gene position is defined via Ensembl ±300 kilbase pairs. Odds ratio is scaled per 1 standard deviation decrease in body mass index ($\approx 4.80 kg/m^2$) or waist-to-hip ratio (≈0.09) [313].

BMI=Body mass index; *GIPR*=Gastric inhibitory polypeptide receptor; GRS=Pan-genome Mendelian randomisation estimate for decreasing body mass index or waist-to-hip ratio on risk of kidney stone disease; KSD=Kidney stone disease; SD=Standard deviation; WHR=Waist-to-hip ratio.

Drug-target MR with an LD threshold of $r^2 < 0.1$ suggested that targeting *GIPR* to reduce BMI by 1 SD, ≈ 4.81m/kg$^2$ [313], may lower the risk of KSD by approximately 85% (Table 4.8, Figure 4.4). Targeting *GIPR* to reduce WHR by 1 SD, ≈ 0.90 [313], may lower the risk of KSD by approximately 80% (Table 4.8, Figure 4.4). However, when using a more stringent LD threshold of $r^2 < 0.01$ to define independent genetic instruments, there were insufficient genetic proxies available to perform comparable analyses. Given the magnitude of genetically-predicted effects from drug-target MR at the *GIPR* locus, I explored the potential value of modulating BMI or WHR via the *GIP* locus. Unfortunately, this analysis could not be undertaken due to a lack of sufficient genetic instruments.

## 4.4 Discussion and conclusions

The second aim in my thesis was to investigate effect of adiposity on the risk of KSD and identify potential disease-mediating genes to guide therapeutic approaches. In this chapter I localise the causal effects of higher BMI and WHR on the risk of KSD to *TFAP2B*, *FTO*, and *GIPR* and identify a candidate causal variant with

functional significance. Building on these findings, I exploit drug-target MR to demonstrate that medications targeting *GIPR*-related pathways may reduce the risk of KSD by up to 85%. Together, these findings enhance our understanding of the associations identified in GWAS and offer insights for future research. This work highlights that there are mechanisms by which the shared genetic architecture of KSD and adiposity traits at *GIPR* exert their effects and lays the groundwork for developing therapeutic approaches to benefit patients with KSD.

### 4.4.1 rs1800437- a candidate causal variant in *GIPR*

The relevance of adiposity to KSD is acknowledged, but the mechanisms driving adiposity-related KSD pathogenesis are poorly understood. Using an innovative MR and colocalisation workflow, I provide evidence that BMI, WHR, and KSD share a genetic architecture at the *GIPR* locus. Additionally, the *GIPR* missense variant, rs1800437 (p.E354Q), emerges as a likely shared causal variant.

The functional effects of p.E354Q on GIPR are characterised in adipocytes [545]. Studies in adipocytes stimulated by GIP demonstrate that the KSD risk allele (rs1800437-G) accelerates GIPR recycling from the cell membrane, through the trans-Golgi network, and back to the cell membrane. Mice carrying the G allele exhibit reductions in GIP sensitivity, glucose tolerance, glucose-stimulated insulin secretion, and responses to GIP stimulation [545].

While the role of p.E354Q in obesity is supported by evidence, its renal implications remain undefined. In vascular and adipose tissues, GIPR activation stimulates *SPP1* expression via endothelin-1-dependent and -independent pathways and reduced GIPR function is linked to diminished *SPP1* expression [548, 549]. *SPP1* encodes osteopontin, which, as discussed in the previous chapter, is a glycoprotein that is secreted in the renal tubule and is a component of the renal stone matrix but inhibits stone formation when in solution [50, 502–504, 550]. To date, the role of GIPR in the kidney has not been thoroughly investigated. Data from The Human Protein Atlas suggests that GIPR and osteopontin proteins are expressed in the renal tubule (Figure 4.5) [499].

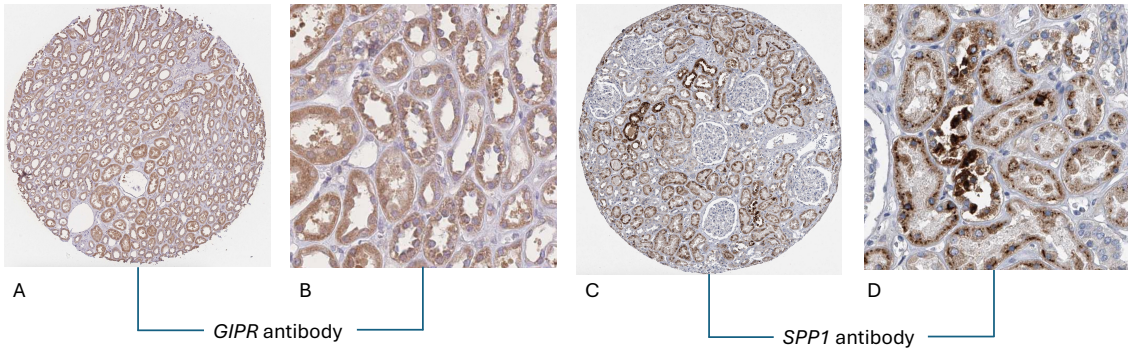



***Figure 4.5. Antibody staining from The Human Protein Atlas shows GIPR expression in normal human renal tissue.***

**A, B:** *GIPR* antibody staining from The Human Protein Atlas [499] shows expression in normal human renal tissue. **B**: Close-up of **A** showing renal tubules stained with *GIPR* antibody. **C, D:** *SPP1* antibody staining from The Human Protein Atlas [499] shows expression in normal human renal tissue. **C**: Close-up of **D** showing renal tubules stained with *SPP1* antibody. Image credit: Human Protein Atlas. Images available from v24.0.proteinatlas.org [551, 552].

Based on these findings, I hypothesise that *GIPR* modulates adiposity traits and is linked to KSD through GIPR-mediated signalling pathways. Specifically, I propose that decreased GIPR signal transduction leads to increased BMI via well-established effects on pancreatic and adipose tissues. Concurrently, a heightened risk of KSD may arise from impaired renal tubular secretion of osteopontin. Consequently, individuals carrying the GIPR-E354 variant (rs1800437-G allele) may have an elevated risk of KSD compared to individuals with GIPR-Q354 (rs1800437-C allele) due to reduced GIPR signalling, which diminishes osteopontin secretion by the renal tubules (Figure 4.6).

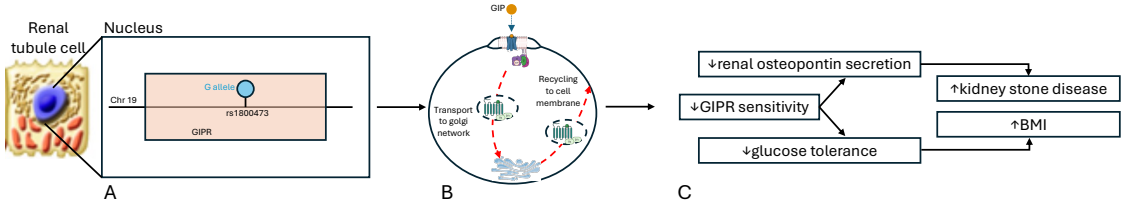



***Figure 4.6. A hypothesis for the role of GIPR in obesity and risk of kidney stone disease.***

**A, B:** The risk (G) allele of rs1800437 at the *GIPR* locus is associated with increased recycling of GIPR through the trans-Golgi network following GIP stimulation. GIPR activation by GIP enhances *SPP1* expression, resulting in the secretion of osteopontin protein in urine, which inhibits calcium crystal growth. **C:** I propose that the increased risk of kidney stone disease and higher BMI associated with the G allele of rs1800437 is mediated by accelerated GIPR recycling to the cell surface. This process reduces sensitivity to GIP, impairing glucose tolerance and decreasing renal tubular osteopontin secretion.

### 4.4.2 Implications for clinical practice

Bowker et al. identify rs1800437 as a putative causal variant linking fasting GIP levels, glycaemic-, and adiposity-related traits. Their findings distinguish this variant from rs7412, located ≈770 Kbp upstream, which drives causal relationships between coronary artery disease (CAD) and lipid traits [553]. This distinction is important given concerns that GIPR-mediated GIP elevation might increase CAD risk [553, 554]. However, in tandem, drug-target MR results (Section 4.3.3) and Bowker et al.'s findings suggest that GIPR agonism could reduce KSD risk without increasing CAD susceptibility. New insights into biological processes underlying complex diseases, as highlighted in this chapter, exemplify the potential value of genomics-driven drug repurposing and discovery in KSD [555].

Tirzepatide, marketed as "Mounjaro" and "Zepbound", is a long-acting dual agonist of GIP and Glucagon-like peptide-1 (GLP1) receptors. Clinical trials have demonstrated its efficacy in treating obesity, type 2 diabetes, obstructive sleep apnoea, and steatohepatitis [556–565]. In the UK, tirzepatide is approved for treating type 2 diabetes as either monotherapy or combination therapy, and for weight management in individuals with a BMI $\geq$30kg/m$^2$, or a BMI $\geq$27kg/m$^2$ with at least one weight-related co-morbidity, alongside dietary and lifestyle interventions [566].

The SURPASS trials (Study to Understand the Relationship between Tirzepatide and Type 2 Diabetes) evaluated the safety and efficacy of tirzepatide for type 2 diabetes compared to placebo and established anti-diabetic therapies [560, 562, 564, 567, 568]. Post-hoc analyses of the pooled SURPASS-1–5 population and the SURPASS-4 cohort demonstrated a protective effect of tirzepatide on renal function, evidenced by reduced albuminuria and improved estimated glomerular filtration rate [565, 569]. These renal benefits may involve GIP- or GLP1-mediated pathways acting directly or indirectly on the kidney, but further research is needed to determine the mechanisms underlying these effects.

Interestingly, data from a cohort of 215,970 individuals with type 2 diabetes initiating GLP1-RA compared to those starting sulphonylureas, DPP4i, or sodium-glucose co-transporter-2 inhibitors (SGLT2i), or continuing with "usual care" in the US Department of Veterans Affairs databases revealed an increased incidence of KSD in the GLP1-RA group in contrast to the DPP4i, SGLT2i, and "usual care" groups [570]. Sensitivity analyses using cumulative incidence functions demonstrated an increased risk of KSD in participants receiving GLP1-RA therapy compared with those receiving "usual care", with an incidence rate ratio of 1.17 (95% CI=1.12–1.22, p=$8.88 \times 10^{-8}$). Restricted mean survival time models revealed a survival time ratio of $9.98 \times 10^{-3}$ (95% CI=$9.96 \times 10^{-3}$–$9.99 \times 10^{-3}$, p=$5.37 \times 10^{-4}$), indicating a shorter average time free of KSD in the GLP1-RA cohort [570]. This observational finding is particularly intriguing given that, in the same study, Xie et al. also reported a reduced risk of acute and chronic kidney diseases in the GLP1-RA cohort, supporting the reno-protective effects observed in the SURPASS post-hoc analyses [565, 569]. However, they also identified a heightened risk of interstitial nephritis in the GLP1-RA group [570].Together, these findings suggest that GLP1-RA therapies have significant renal effects. However, the specific effects, particularly concerning KSD risk, and the underlying mechanisms remain to be elucidated. Despite these findings, the observational nature of Xie et al.'s study introduces limitations. Notably, there was no mention of baseline assessment of the presence or absence of KSD, raising the possibility that some reported KSD cases during follow-up were pre-existing, asymptomatic stones unevenly distributed across treatment groups. Furthermore, the study included a range of GLP1-RA therapies and reported

that tirzepatide was included, though only 0.03% of participants were prescribed tirzepatide (a dual agonist of GIP and GLP1 receptors). However, the study lacked subgroup analyses stratified by specific GLP1-RA agents, which limits the ability to disentangle potential differential effects mediated by GLP1 versus GIP pathways on KSD risk [570]. While these findings are noteworthy, they do not detract from the evidence presented in this chapter supporting the role of GIPR in modulating the risk of KSD.

At present, no large-scale clinical trials have directly investigated the impact of GIPR-agonist therapies on KSD risk or related proxies. This lack of trial evidence limits our understanding of their potential to mitigate KSD risk. Additionally, as tirzepatide only received approval for use in the UK in November 2023 [571], national health registries such as the Clinical Practice Research Datalink lack sufficient follow-up data to address this question through observational analyses. Further research, including well-designed clinical trials and detailed cohort studies, is essential to clarify the role of GIP- and GLP1-mediated pathways in modulating KSD risk.

### 4.4.3 Drug repurposing for kidney stone disease

As discussed earlier in this chapter, DPP4i, commonly referred to as "gliptins", are a class of drugs that increase incretin hormone levels to control glycaemia. By inhibiting dipeptidyl peptidase IV, an enzyme expressed in the kidney, small intestine, submaxillary gland, and liver, these therapies enhance beta-cell insulin secretion and reduce postprandial and fasting hyperglycaemia [572, 573]. Given their action on the GIPR pathway, one might hypothesise that DPP4i could provide therapeutic benefit for KSD. However, unlike GIPR agonists, DPP4i therapies are weight-neutral, and studies have not demonstrated their efficacy in reducing KSD risk [572, 574]. Several factors may explain this.

First, DPP4i may not effectively penetrate renal tissue, limiting their ability to influence a renal-intrinsic GIPR pathway. While they increase systemic GIP levels, their impact on local renal GIPR activation may be minimal. DPP4 exists in two

isoforms: a circulating form that cleaves GLP1 and a membranous form found in various cell types, including the brush border of the proximal convoluted tubule [575, 576]. Different DPP4i have distinct pharmacokinetic properties, which may influence their renal effects. For instance, linagliptin is not renally cleared, potentially reducing its direct impact on renal pathways compared to other DPP4i [575]. Further pharmacokinetic studies are needed to assess the renal tissue penetration of DPP4i and their potential influence on KSD.

Another possibility is that DPP4i do not achieve the threshold of GIPR activation required to impact KSD risk. The extent of GIPR activation depends on endogenous GIP levels, which different DPP4i modulate to varying degrees [577]. If the increase in GIP levels induced by DPP4i is insufficient to fully activate GIPR, any potential renal effects may not reach a biologically significant level. In contrast, direct GIPR agonists may provide stronger activation, potentially uncovering an effect on KSD risk.

SGLT2i lower blood glucose by promoting glycosuria and play a central role in diabetes management. Clinical trials demonstrate their cardio- and reno-protective benefits in both diabetic and non-diabetic patients [578]. Unlike DPP4i, SGLT2i show promise in reducing KSD risk. Observational studies and the SWEETSTONE trial support their anti-lithogenic effects. Paik et al. compared KSD risk in patients starting SGLT2i, DPP4i, or GLP1-RA and found the greatest risk reduction in those initiating SGLT2i, despite a median follow-up of only 192 days [574].

These findings underscore the importance of understanding disease pathophysiology and drug mechanisms of action. SGLT2i therapies modulate sodium transport in the proximal renal tubule, altering urine concentration, pH, citrate, urate, sodium, phosphate, and calcium levels, as well as systemic concentrations of 1,25-dihydroxyvitamin D, phosphate, and FGF23 [579–585]. Additionally, the SWEETSTONE trial demonstrated that empagliflozin significantly reduced calcium phosphate and uric acid urine relative supersaturation ratios, surrogate markers of stone recurrence, in non-diabetic adults with calcium and uric acid KSD, respectively, compared to placebo [586]. These mechanisms plausibly underpin the anti-lithogenic

effects of SGLT2i. Conversely, DPP4i therapies primarily target hyperglycaemia, which, evidence shows, does not contribute to KSD risk [119].

Since aligning a candidate therapy's mechanism of action with the underlying disease pathophysiology is crucial for reducing clinical risk, GIP analogues could be particularly relevant for KSD due to their dual effects on weight homeostasis and metabolic regulation. These therapies primarily influence weight homeostasis, a trait proven to have a causal role in KSD [119]. Furthermore, the shared genetic architecture between KSD and adiposity traits, as demonstrated in this chapter, underscores their therapeutic potential. A key question, in the era of personalised medicine, is whether therapeutic benefit may be seen in all individuals with KSD, irrespective of underlying BMI, or specifically in overweight or obese patients. Future research should evaluate the efficacy of GIP analogues in preventing KSD, and integrate experimental, clinical trial, and real-world evidence to optimise therapeutic strategies.

### 4.4.4 Conclusion

In my second results chapter, I have employed a novel approach to MR and colocalisation analyses, moved beyond genetic associations identified in GWAS and observational studies, established the causal effects of higher BMI and WHR on increasing the risk of KSD at specific genomic loci, and identified a functional, causal, missense variant in *GIPR*. Through drug-target MR, I have demonstrated that therapies targeting GIPR-related pathways could significantly reduce the risk of KSD, offering an exciting opportunity for future therapeutic development. Further research is needed to uncover the underlying mechanisms, identify reliable biomarkers, design clinical trials, and ultimately translate these findings into effective therapeutic strategies for patients with KSD.

# 5 | The role of mineral metabolism in kidney stone disease

## 5.1 Introduction

### 5.1.1 Disorders of calcium and phosphate homeostasis associated with kidney stone disease

As outlined in Chapter 1, disturbances in calcium and phosphate homeostasis contribute to KSD, particularly systemic disorders such as hyperparathyroidism. Rare, monogenic disorders affecting calcium and phosphate metabolism, including *SLC34A1*-, *SLC34A3*-, or *SLC9AR1*-related hereditary hypophosphataemic rickets with hypercalciuria [157, 178, 180, 226], *CASR*- or *GNA11*-associated autosomal dominant hypocalcaemia [53, 163], *CASR*-, *GNA11*-, or *AP2S1*- associated familial hypocalciuric hypercalcaemia (FHH), and *CYP24A1*- or *SLC34A1*-associated infantile hypercalcaemia [587], also contribute to the pathogenesis nephrolithiasis [588]. Although these conditions offer valuable insights into mechanisms underlying KSD, they account for only ≈3–15% of KSD cases [157, 179]. Most cases arise from complex, polygenic influences interacting with environmental factors [156, 589–591].

### 5.1.2 Causal effects of calcium and phosphate in kidney stone disease

GWAS of KSD, including those presented in Chapter 3, identify common genetic variants in proximity to genes involved in calcium and phosphate signalling, metabolism, and transport. Notable examples include *DGKD*, encoding diacylglycerol kinase $\delta$ (DGK$\delta$), a key component of the calcium-sensing receptor (CaSR) signalling pathway [592–594]; *CASR*, encoding CaSR; *SLC34A1*, encoding

the renal sodium-phosphate cotransporter 2A, NaPi-IIa; *GATA3*, encoding trans-acting T-cell-specific transcription factor GATA-3, implicated in parathyroid gland development [595]; *CYP24A1*, encoding 24-$\alpha$-hydroxylase; and *CLDN14*, encoding claudin-14, a tight-junction protein involved in renal calcium transport [158, 240, 242, 243, 296, 300]. Moreover, gene set analyses in GWAS meta-analyses of European and trans-ancestry populations highlight enrichment of pathways related to urinary calcium, blood inorganic cation levels, fibroblast growth factor production, and anion homeostasis (Chapter 3 Table 3.7 and Chapter 3 Table 3.8).

Beyond these genetic associations, MR studies that I conducted prior to my DPhil demonstrate a causal link between mineral homeostasis and KSD. Specifically, a 0.08 mmol/L increase in serum calcium concentrations causes a ≈48% increased risk of KSD, while a 0.016 mmol/L decrease in serum phosphate concentrations confers a ≈33% increased risk of KSD [119]. These findings suggest that even subtle perturbations in serum calcium and phosphate within the normal physiological range significantly influence stone formation.

### 5.1.3 The unmet need in kidney stone disease

Despite insights from monogenic and polygenic disorders of mineral metabolism, the precise renal and systemic mechanisms linking calcium and phosphate homeostasis to KSD remain poorly understood. These gaps hinder our ability to predict, prevent, and treat KSD effectively [596]. Identifying shared genetic architectures that drive KSD risk is essential for advancing targeted therapeutic strategies [302, 407].

Building on these genetic insights, multiomic and chromatin interaction studies can provide greater granularity by identifying regulatory elements, gene expression patterns, and epigenetic modifications that influence disease-relevant pathways. By integrating these methods, we can establish a mechanistic framework to understand how genetic and metabolic disturbances contribute to stone formation and identify opportunities for intervention.

### 5.1.4 Aim and Objectives

In this chapter, my main research aim is to investigate the causal effects of alterations in mineral homeostasis on the risk of KSD and to identify potential disease-mediating genetic loci to guide future therapeutic approaches. Specifically, the research objectives are to:

1. **Perform regional MR**
    - Use genetic instruments for serum calcium and phosphate to establish causal effects of these traits on KSD risk across different genomic regions.
2. **Perform regional colocalisation**
    - Identify genomic regions where genetic variants influence serum calcium, and serum phosphate, and KSD risk.
3. **Analyse single-cell multiomic and chromatin conformational capture data**
    - Explore gene expression and chromatin interactions to gain insights into the shared disease-causing genes underlying mineral metabolism traits and KSD.
4. **Critically appraise candidate loci and prioritise genes**
    - Investigate the biological roles of identified loci and prioritise candidate genes that may mediate the relationship between mineral metabolism traits and KSD.
5. **Explore therapeutic implications of altering serum calcium and phosphate concentrations using drug-target MR**
    - Identify key genes or pathways implicated in mineral metabolism traits and KSD as potential targets for future therapeutic interventions.

"Serum calcium concentration" refers to values adjusted for serum albumin levels.

## 5.2 Study design

Figure 5.1 outlines the study design for this chapter.

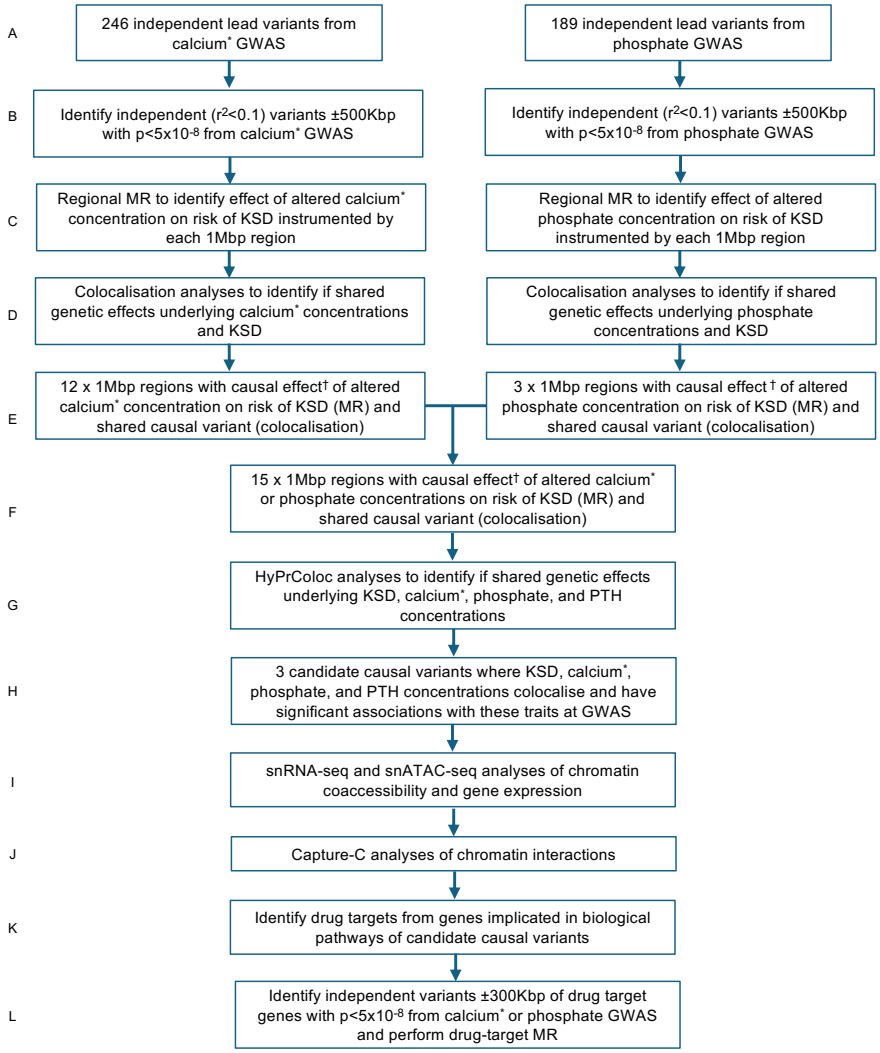



***Figure 5.1. Identifying genetic variants predisposing to increased risk of kidney stone disease via alterations in serum calcium and phosphate concentrations.***

**A, B:** Independent genetic variants ±500 kilobase pairs (Kbp) of lead independent variants from serum calcium or phosphate genome-wide association studies (GWAS) significantly associated with serum calcium or phosphate selected as instrumental variables (IVs). **C:** Mendelian randomisation (MR) to instrument the effects of altering biochemical exposures on risk of kidney stone disease (KSD) in European ancestry KSD GWAS meta-analysis, UK Biobank and FinnGen. **D:** Colocalisation analyses. **E:** Regions with significant MR results and evidence of colocalisation identified. **F,G,H:** HyPrColoc to assess multi-trait colocalisation and identify candidate causal variants. **I**: snRNA- and ATAC- seq data analysis to identify genes where chromatin accessibility at promoter correlates with gene expression. **J**: Capture-C analyses investigating physical chromatin interactions between candidate variants and genes prioritised in **I**. **K:** Drug-targets identified from genes associated with candidate causal variants. **L:** Drug-target MR to assess potential to prevent KSD, selecting IVs from ±300 Kbp of genes of interest. GWAS=Genome-wide association study; Kbp=Kilobase pair; KSD=Kidney stone disease; Mbp=Megabase pair; MR=Mendelian randomisation; PTH=Parathyroid hormone; ∗ Albumin-adjusted serum calcium concentration; †instrumental variable comprising ≥3 genetic variants.

## 5.3 Methods

### 5.3.1 Data sources

#### 5.3.1.1 Mineral metabolism traits

To perform MR analyses, I leveraged summary statistics for GWAS of serum calcium and phosphate conducted by Hannan et al. in participants of European ancestry from the UK Biobank (Figure 5.1-A) [119, 597]. In colocalisation analyses, I included PTH concentrations and leveraged data from a GWAS meta-analysis of serum PTH concentrations among 29,155 participants of European ancestry from 13 cohort studies [598].

#### 5.3.1.2 Kidney stone disease

For MR and colocalisation analyses, I considered three KSD outcome datasets; the European-ancestry KSD GWAS meta-analysis (Chapter 3 Section 3.3.2), the UK Biobank combined-sex KSD GWAS (Chapter 3 Section 3.3.1), and FinnGen study release 12 summary statistics [433].

### 5.3.2 Mendelian randomisation

To ascertain the regional effects of genetically-predicted serum calcium and phosphate concentrations on odds of KSD, I leveraged independent genetic variants ($r^2$ <0.1) ±500 Kbp of the lead independent variants from serum calcium or phosphate GWAS, significantly (p< $5 \times 10^{-8}$) associated with the respective biochemical traits as instrumental variables. MR was performed as described in Chapter 2 Section 2.4 (Figure 5.1-B and C). Results are scaled per 1 SD change in mineral metabolism trait which equates to 0.08 mmol/L change in serum calcium concentration and 0.016 mmol/L change in serum phosphate concentration.

To account for participant overlap between the UK Biobank GWAS of serum

calcium and phosphate and the KSD cohort, I conducted a power calculation using https://sb452.shinyapps.io/overlap/ and estimated the minimum and maximum effects I had 80% statistical power to detect [381, 467].

### 5.3.3 Colocalisation

I performed colocalisation analyses to assess the probability of a single shared causal variant influencing serum calcium levels and risk of KSD, or serum phosphate levels and risk of KSD (Figure 5.1-D, E) [407, 473].

Calcium and phosphate homeostasis are tightly regulated and influenced by PTH concentrations. To strengthen confidence in putative causal variants and identify genomic regions with a shared genetic architecture affecting serum calcium and phosphate concentrations, and KSD risk, I performed multi-trait colocalisation analysis for KSD and calcium, phosphate, and PTH concentrations for regions where there was evidence of colocalisation between serum calcium or phosphate concentrations and KSD (Figure 5.1-F, G, H) [302]. Full colocalisation methods are described in Chapter 2 Section 2.5.

### 5.3.4 Single-cell Multiomics

To deepen understanding of kidney biology and contextualise findings from MR analyses, I explored the relationships between gene expression and chromatin accessibility using snRNA-seq and snATAC-seq data from five healthy kidney samples published by Muto et al [417, 474]. snRNA-seq and snATAC-seq data preparation are detailed in Chapter 2 Section 2.6.

#### 5.3.4.1 *Cis*-regulatory interactions and gene activity scores

To elucidate how genetic variation influences KSD pathogenesis, I explored potential biological mechanisms underlying associations with serum calcium, phosphate levels, and KSD risk. I assessed the functional relevance of candidate causal variants by

evaluating their role in regulating gene expression within renal cell types. Using Cicero, outlined in Methods Chapter 2 Section 2.6.3 [417, 480], I leveraged paired multiomic data to identify regulatory regions where chromatin accessibility predicts transcriptional output across different cell types, providing insights into gene regulation in the context of KSD.

### 5.3.5 Chromatin conformational capture

After identifying genomic regions of interest through MR, colocalisation, and snRNA- and snATAC-seq analyses, I studied the three-dimensional structure of chromatin using Capture-C in RPTEC, HK2, and HEK293 cells to understand their relevance to regulating genes in the kidney. Detailed methods are provided in published protocols and in Methods Chapter 2 Section 2.7 [423, 480–483].

### 5.3.6 Drug-target Mendelian randomisation

I identified target genes associated with candidate causal variants using MR, multi-trait colocalisation, and multiomic analyses, as described above. To assess their potential as therapeutic targets to prevent KSD, I conducted drug-target MR analyses. Genetic variants associated with calcium or phosphate ($p < 5 \times 10^{-8}$) and located within ±300 Kbp of each target gene were selected for analyses (Figure 5.1-I, J). MR analyses were performed following the principles outlined in Chapter 2 Section 2.4.

## 5.4 Results

### 5.4.1 Genetic variants increase the risk of kidney stone disease via alterations in serum calcium or phosphate concentrations

The risk of bias due to sample overlap among participants in the UK Biobank for KSD, serum calcium, and phosphate GWAS was estimated to be 0.011 and 0.001, respectively (Table 5.1).

***Table 5.1. Estimate of bias for Mendelian randomisation with sample overlap in the UK Biobank.***

| Trait | GWAS sample size | Maximum number of instruments in MR analyses | Estimate of bias |
|---|---|---|---|
| Adjusted calcium | 308,679 | 24 | 0.011 |
| Phosphate | 308,110 | 20 | 0.001 |

The maximum number of instruments was used to derive the most pessimistic estimate of bias. Results are shown for a Type 1 error rate of 0.05.
GWAS=Genome-wide association study; MR=Mendelian randomisation.

Summary statistics from meta-analysis of European ancestry KSD GWAS, UK Biobank GWAS, and FinnGen GWAS, and serum calcium, and serum phosphate GWAS were used to identify 57 and 66 different 1 Mbp regions where increasing serum calcium and phosphate concentrations, respectively, causally alter odds of KSD (Table 5.2).

***Table 5.2. Number of 1 megabase regions involved in evaluating the causal effect of serum calcium and phosphate concentrations on odds of kidney stone disease.***

| Analysis step | Dataset | Serum calcium | Serum phosphate |
|---|---|---|---|
| Independent significant SNPs in GWAS | UK Biobank | 246 | 180 |
| 1Mbp genomic regions with > 2 SNPs to assess in MR | Meta-analysis | 114 | 73 |
| | UK Biobank | 116 | 68 |
| | FinnGen | 110 | 70 |
| Significant 1Mbp regions in MR | Meta-analysis | 62 | 24 |
| | UK Biobank | 46 | 26 |
| | FinnGen | 45 | 10 |
| Significant 1Mbp regions in MR with evidence of colocalisation | Meta-analysis | 16 | 5 |
| | UK Biobank | 9 | 3 |
| | FinnGen | 13 | 3 |

GWAS=Genome-wide association study; SNP=Single nucleotide polymorphism; Mbp=Megabase pairs; MR=Mendelian randomisation.

Colocalisation analyses identified 14 1 Mbp regions associated with eight candidate genes with shared causal variants contributing to liability for both higher serum calcium and an increased risk of KSD (Table 5.3, Figure 5.2). These regions align directionally with overall estimates of elevated serum calcium increasing KSD susceptibility. Three 1 Mbp regions across two loci with shared causal variants explained liability to both higher serum phosphate and reduced KSD risk, aligning with overall estimates of the protective effect of higher serum phosphate on KSD risk (Table 5.4, Figure 5.2). Notably, overlap was observed at *DGKD*-associated regions, where calcium-associated regions (rs838705 ±500 Kbp and rs838717±500 Kbp) and the phosphate-associated region (rs838718 ±500 Kbp) coincided. rs838717 emerged as a candidate causal variant underlying both the effect of increasing serum calcium and decreasing serum phosphate concentrations on KSD risk (Table 5.3 and Table 5.4).

***Table 5.3. Genomic regions of 1 megabase where increasing serum calcium concentration affects odds of kidney stone disease with evidence of colocalisation.***

| **IV: 1 Mbp region from GWAS** | | | | **MR analysis** | | | | | | | | | **Colocalisation** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Index variant (Gene)** | **CHR** | **POS** | **Mean $R^2$ ($SD$)** | **Inverse variance weighted analysis** | | | | **Heterogeneity** | | | **Intercept** | | **N SNPs** | **PP H4** | **Candidate SNP** | **SNP PP** |
| | | | | **Outcome** | **N SNPs** | **OR (95% CI)** | **P** | **Q** | **Q_df** | **Q P** | **$\beta$ (SE)** | **P** | | | | |
| rs607518 *(ANXA9)* | 1.00 | 150954671 | $2.20\times10^{-4}$ ($9.56\times10^{-5}$) | Meta-analysis | 3 | 7.06 (4.12-12.11) | $1.75\times10^{-11}$ | 0.22 | 2 | 0.90 | -0.09 (-0.19) | 0.99 | 1870 | 0.91 | rs267738 | 0.36 |
| | | | | UK Biobank | 4 | 10.45 (5.28-20.70) | $3.24\times10^{-10}$ | 1.37 | 3 | 0.71 | -0.11 (-0.11) | 0.98 | 2231 | 0.98 | rs267738 | 0.44 |
| | | | | FinnGen | 3 | 2.37 (0.4-13.99) | 0.46 | 11.97 | 2 | $2.51\times10^{-3}$ | -0.4 (-0.11) | 0.97 | 1562 | 0.03 | | |
| rs3011 *(SLC30A10)* | 1 | 220085453 | $1.57\times10^{-4}$ ($2.9\times10^{-5}$) | Meta-analysis | 4 | 18.38 (9.57-35.30) | $4.25\times10^{-17}$ | 3.06 | 3 | 0.38 | 0.05 (-0.03) | 0.98 | 3392 | 0.99 | rs3011 | 0.19 |
| | | | | UK Biobank | 4 | 22.99 (10.65-49.60) | $3.00\times10^{-14}$ | 1.46 | 3 | 0.69 | 0.03 (-0.04) | 0.98 | 3344 | 0.98 | rs10863512 | 0.21 |
| rs2675966 *(SNORC)* | 2 | 233735543 | $3.74\times10^{-4}$ ($3.17\times10^{-4}$) | Meta-analysis | 4 | 3.87 (2.71-5.51) | $1.27\times10^{-12}$ | 2.43 | 3 | 0.49 | -0.01 (-0.02) | 0.99 | 3580 | 1.00 | rs11891546 | 1.00 |
| | | | | UK Biobank | 3 | 3.04 (1.77-5.23) | $2.72\times10^{-4}$ | 2.06 | 2 | 0.36 | -0.01 (-0.03) | 0.98 | 3617 | 0.96 | rs11891546 | 1.00 |
| | | | | FinnGen | 4 | 4.66 (2.9-7.49) | $3.05\times10^{-9}$ | 1.57 | 3 | 0.67 | $-3.69\times10^{-3}$ (-0.02) | 0.97 | 3087 | 0.99 | rs11891546 | 1.00 |
| rs838705 *(DGKD)* | 2 | 234273242 | $3.27\times10^{-4}$ ($3.77\times10^{-4}$) | Meta-analysis | 10 | 4.51 (3.38-6.03) | $3.31\times10^{-23}$ | 14.0 | 9 | 0.12 | -0.03 (-0.01) | 0.48 | 3303 | 1.00 | rs838717 | 1.00 |
| | | | | UK Biobank | 9 | 3.89 (2.61-5.79) | $3.94\times10^{-10}$ | 10.76 | 8 | 0.22 | -0.04 (-0.02) | 0.98 | 3409 | 1.00 | rs838717 | 1.00 |
| | | | | FinnGen | 11 | 4.93 (3.39-7.19) | $1.69\times10^{-15}$ | 15.29 | 10 | 0.12 | -0.02 (-0.02) | 0.97 | 2961 | 1.00 | rs838717 | 1.00 |
| rs838717 *(DGKD)* | | | | Meta-analysis | 10 | 4.51 (3.38-6.03) | $3.31\times10^{-23}$ | 14.0 | 9 | 0.12 | -0.03 (-0.01) | 0.48 | 3265 | 1.00 | rs838717 | 1.00 |

| IV: 1 Mbp region from GWAS | | | | MR analysis | | | | | | | | | Colocalisation | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Index variant (Gene) | CHR | POS | Mean $R^2$ ($SD$) | Inverse variance weighted analysis | | | | Heterogeneity | | | Intercept | | N SNPs | PP H4 | Candidate SNP | SNP PP |
| | | | | Outcome | N SNPs | OR (95% CI) | P | Q | Q_df | Q P | $\beta$ (SE) | P | | | | |
| | 2 | 234296444 | $3.27\times10^{-4}$ ($3.77\times10^{-4}$) | UK Biobank | 9 | 3.89 (2.61-5.79) | $3.94\times10^{-10}$ | 10.76 | 8 | 0.22 | -0.04 (-0.02) | 0.98 | 3372 | 1.00 | rs838717 | 1.00 |
| | | | | FinnGen | 11 | 4.93 (3.39-7.19) | $1.69\times10^{-15}$ | 15.29 | 10 | 0.12 | -0.02 (-0.02) | 0.97 | 2922 | 1.00 | rs838717 | 1.00 |
| rs79069677 *(AOPEP)* | 9 | 97530059 | $2.56\times10^{-4}$ ($1.85\times10^{-4}$) | Meta-analysis | 3 | 4.06 (1.68-9.79) | 0.01 | 6.87 | 2 | 0.03 | -0.04 (-0.02) | 0.98 | 2862 | 0.94 | rs925813 | 0.06 |
| | | | | UK Biobank | 3 | 2.48 (0.8-7.70) | 0.21 | 4.96 | 2 | 0.08 | -0.04 (-0.02) | 0.98 | 2999 | 0.50 | | |
| rs1858800 *(ZFHX3)* | 16 | 73024276 | $2.53\times10^{-4}$ ($2.4\times10^{-4}$) | Meta-analysis | 4 | 3.51 (1.73-7.15) | $1.63\times10^{-3}$ | 8.29 | 3 | 0.04 | $8.34\times10^{-4}$ (-0.04) | 0.99 | 2679 | 1.00 | rs1858800 | 1.00 |
| | | | | UK Biobank | 5 | 2.12 (0.74-6.06) | 0.27 | 14.7 | 4 | 0.01 | -0.01 (-0.05) | 0.98 | 2811 | 0.61 | | |
| | | | | FinnGen | 5 | 4.63 (2.67-8.04) | $6.97\times10^{-7}$ | 1.84 | 4 | 0.77 | $-4.03\times10^{-3}$ (-0.02) | 0.97 | 2325 | 0.97 | rs1858800 | 1.00 |
| rs4782358 *(ZFPM1)* | 16 | 88485506 | $2.66\times10^{-4}$ ($2.8\times10^{-4}$) | Meta-analysis | 5 | 0.31 (0.21-0.46) | $4.29\times10^{-8}$ | 4.46 | 4 | 0.35 | 0.02 (-0.02) | 0.98 | 4679 | 0.95 | rs12918968 | 0.57 |
| | | | | UK Biobank | 5 | 0.4 (0.23-0.67) | $2.53\times10^{-3}$ | 3.52 | 4 | 0.47 | $-2.45\times10^{-3}$ (-0.02) | 0.98 | 4809 | 0.29 | | |
| | | | | FinnGen | 4 | 0.22 (0.12-0.38) | $6.97\times10^{-7}$ | 1.79 | 3 | 0.62 | 0.04 (-0.04) | 0.97 | 4090 | 0.96 | rs12918968 | 0.90 |
| rs12447180 *(ZFPM1)* | 16 | 88517722 | $2.66\times10^{-4}$ ($2.8\times10^{-4}$) | Meta-analysis | 5 | 0.31 (0.21-0.46) | $4.29\times10^{-8}$ | 4.46 | 4 | 0.35 | 0.02 (-0.02) | 0.98 | 4570 | 0.95 | rs12918968 | 0.87 |
| | | | | UK Biobank | 5 | 0.4 (0.23-0.67) | $2.53\times10^{-3}$ | 3.52 | 4 | 0.47 | $-2.45\times10^{-3}$ (-0.02) | 0.98 | 4697 | 0.29 | | |
| | | | | FinnGen | 4 | 0.22 (0.12-0.38) | $6.97\times10^{-7}$ | 1.79 | 3 | 0.62 | 0.04 (-0.04) | 0.97 | 3949 | 0.96 | rs12918968 | 0.84 |
| rs3091842 *(MAFB)* | | | | Meta-analysis | 4 | 1.89 (1.29-2.76) | $2.88\times10^{-3}$ | 5.97 | 3 | 0.11 | $-2.03\times10^{-3}$ (-0.02) | 0.99 | 2761 | 0.94 | rs3091842 | 1.00 |

| IV: 1 Mbp region from GWAS | | | | MR analysis | | | | | | | | | Colocalisation | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Index variant (Gene) | CHR | POS | Mean $R^2$ ($SD$) | Inverse variance weighted analysis | | | | Heterogeneity | | | Intercept | | N SNPs | PP H4 | Candidate SNP | SNP PP |
| | | | | Outcome | N SNPs | OR (95% CI) | P | Q | Q_df | Q P | $\beta$ (SE) | P | | | | |
| | 20 | 39344272 | $4.6\times10^{-4}$ ($3.57\times10^{-4}$) | UK Biobank | 4 | 1.79 (1.14-2.81) | 0.03 | 1.65 | 3 | 0.65 | $4.64\times10^{-4}$ (-0.02) | 0.98 | 2697 | 0.08 | | |
| | | | | FinnGen | 4 | 1.95 (1.27-3.00) | 0.01 | 4.94 | 3 | 0.18 | $-2.71\times10^{-3}$ (-0.02) | 0.97 | 2361 | 0.79 | rs3091842 | 1.00 |
| rs2425431 *(MAFB)* | 20 | 39352714 | $4.6\times10^{-4}$ ($3.57\times10^{-4}$) | Meta-analysis | 4 | 1.89 (1.29-2.76) | $2.88\times10^{-3}$ | 5.97 | 3 | 0.11 | $-2.03\times10^{-3}$ (-0.02) | 0.99 | 2754 | 0.94 | rs3091842 | 1.00 |
| | | | | UK Biobank | 4 | 1.79 (1.14-2.81) | 0.03 | 1.65 | 3 | 0.65 | $4.64\times10^{-4}$ (-0.02) | 0.98 | 2690 | 0.08 | | |
| | | | | FinnGen | 4 | 1.95 (1.27-3.00) | 0.01 | 4.94 | 3 | 0.18 | $-2.71\times10^{-3}$ (-0.02) | 0.97 | 2351 | 0.79 | rs3091842 | 1.00 |
| rs209961 *(CYP24A1)* | 20 | 52715154 | $3.29\times10^{-4}$ ($3.08\times10^{-4}$) | Meta-analysis | 18 | 10.36 (8.54-12.56) | $1.45\times10^{-123}$ | 20.37 | 17 | 0.26 | -0.02 (-0.01) | 0.48 | 3561 | 1.00 | rs6127099 | 1.00 |
| | | | | UK Biobank | 17 | 8.46 (6.58-10.88) | $8.54\times10^{-61}$ | 15.74 | 16 | 0.47 | $3.29\times10^{-4}$ (-0.01) | 0.98 | 3647 | 0.98 | rs6127099 | 0.98 |
| | | | | FinnGen | 18 | 12.69 (9.88-16.31) | $1.74\times10^{-86}$ | 20.64 | 17 | 0.24 | -0.04 (-0.01) | 0.10 | 3236 | 1.00 | rs6127099 | 1.00 |
| rs2585442 *(CYP24A1)* | 20 | 52737123 | $3.29\times10^{-4}$ ($3.08\times10^{-4}$) | Meta-analysis | 18 | 10.36 (8.54-12.56) | $1.45\times10^{-123}$ | 20.37 | 17 | 0.26 | -0.02 (-0.01) | 0.48 | 3591 | 1.00 | rs6127099 | 1.00 |
| | | | | UK Biobank | 17 | 8.46 (6.58-10.88) | $8.54\times10^{-61}$ | 15.74 | 16 | 0.47 | $3.29\times10^{-4}$ (-0.01) | 0.98 | 3682 | 0.98 | rs6127099 | 0.98 |
| | | | | FinnGen | 18 | 12.69 (9.88-16.31) | $1.74\times10^{-86}$ | 20.64 | 17 | 0.24 | -0.04 (-0.01) | 0.10 | 3262 | 1.00 | rs6127099 | 1.00 |
| rs117268564 *(CYP24A1)* | 20 | 52738434 | $3.29\times10^{-4}$ ($3.08\times10^{-4}$) | Meta-analysis | 18 | 10.36 (8.54-12.56) | $1.45\times10^{-123}$ | 20.37 | 17 | 0.26 | -0.02 (-0.01) | 0.48 | 3594 | 1.00 | rs6127099 | 1.00 |
| | | | | UK Biobank | 17 | 8.46 (6.58-10.88) | $8.54\times10^{-61}$ | 15.74 | 16 | 0.47 | $3.29\times10^{-4}$ (-0.01) | 0.98 | 3685 | 0.98 | rs6127099 | 0.98 |

| IV: 1 Mbp region from GWAS | | | | MR analysis | | | | | | | | | Colocalisation | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Index variant (Gene) | CHR | POS | Mean $R^2$ ($SD$) | Inverse variance weighted analysis | | | | Heterogeneity | | | Intercept | | N SNPs | PP H4 | Candidate SNP | SNP PP |
| | | | | Outcome | N SNPs | OR (95% CI) | P | Q | Q_df | Q P | $\beta$ (SE) | P | | | | |
| | | | | FinnGen | 18 | 12.69 (9.88-16.31) | $1.74\times10^{-86}$ | 20.64 | 17 | 0.24 | -0.04 (-0.01) | 0.10 | 3264 | 1.00 | rs6127099 | 1.00 |
| rs35194449 *(CYP24A1)* | 20 | 52742047 | $3.29\times10^{-4}$ ($3.08\times10^{-4}$) | Meta-analysis | 18 | 10.36 (8.54-12.56) | $1.45\times10^{-123}$ | 20.37 | 17 | 0.26 | -0.02 (-0.01) | 0.48 | 3593 | 1.00 | rs6127099 | 1.00 |
| | | | | UK Biobank | 17 | 8.46 (6.58-10.88) | $8.54\times10^{-61}$ | 15.74 | 16 | 0.47 | $3.29\times10^{-4}$ (-0.01) | 0.98 | 3684 | 0.98 | rs6127099 | 0.98 |
| | | | | FinnGen | 18 | 12.69 (9.88-16.31) | $1.74\times10^{-86}$ | 20.64 | 17 | 0.24 | -0.04 (-0.01) | 0.10 | 3266 | 1.00 | rs6127099 | 1.00 |
| rs2762943 *(CYP24A1)* | 20 | 52790786 | $3.29\times10^{-4}$ ($3.08\times10^{-4}$) | Meta-analysis | 18 | 10.36 (8.54-12.56) | $1.45\times10^{-123}$ | 20.37 | 17 | 0.26 | -0.02 (-0.01) | 0.48 | 3599 | 1.00 | rs6127099 | 1.00 |
| | | | | UK Biobank | 17 | 8.46 (6.58-10.88) | $8.54\times10^{-61}$ | 15.74 | 16 | 0.47 | $3.29\times10^{-4}$ (-0.01) | 0.98 | 3760 | 0.98 | rs6127099 | 0.98 |
| | | | | FinnGen | 18 | 12.69 (9.88-16.31) | $1.74\times10^{-86}$ | 20.64 | 17 | 0.24 | -0.04 (-0.01) | 0.10 | 3340 | 1.00 | rs6127099 | 1.00 |

CHR=Chromosome; GWAS=Genome-wide association study; Mbp=Megabase pair; MR=Mendelian randomisation; N SNPs=Number of SNPs; OR=Odds ratio for kidney stone disease per genetically-instrumented standard deviation increase in trait; P=5% false discovery rate adjusted p-value; POS=Position based on NCBI Genome Build 37 (hg19); PP H4=posterior probability of full colocalisation; Q=Q statistic; Q_df=Q degrees of freedom; $R^2$=Variance in exposure explained by instrumental variable; SD=Standard deviation; SE=Standard error; SNP PP=Posterior probability explained by SNP; 95% CI=95% confidence interval.

***Table 5.4. Genomic regions of 1 megabase where decreasing serum phosphate concentration increases odds of kidney stone disease with evidence of colocalisation.***

| **IV: 1 Mbp region from GWAS** | | | | **MR analysis** | | | | | | | | | **Colocalisation** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Index variant (Gene)** | **CHR** | **POS** | **Mean $R^2$ ($SD$)** | **Inverse variance weighted analysis** | | | | **Heterogeneity** | | | **Intercept** | | **N SNPs** | **PP H4** | **Candidate SNP** | **SNP PP** |
| | | | | **Outcome** | **N SNPs** | **OR (95% CI)** | **P** | **Q** | **Q_df** | **Q P** | **$\beta$ (SE)** | **P** | | | | |
| rs28383573 *(APCS)* | 1.00 | 159558672 | $2.80\times10^{-4}$ ($1.05\times10^{-4}$) | Meta-analysis | 3 | 2.22 (1.28-3.83) | 0.02 | 3.45 | 2 | 0.18 | -0.04 (-0.03) | 0.98 | 3,268 | 0.87 | rs71628146 | 0.35 |
| | | | | UK Biobank | 3 | 2.08 (0.71-6.11) | 0.35 | 5.04 | 2 | 0.08 | -0.07 (-0.07) | 0.99 | 3213 | 0.35 | | |
| | | | | FinnGen | 3 | 2.3 (1.36-3.91) | 0.02 | 0.49 | 2 | 0.78 | -0.02 (-0.03) | 0.98 | 2741 | 0.31 | | |
| rs71628146 *(APCS)* | 1.00 | 159582487 | $2.80\times10^{-4}$ ($1.05\times10^{-4}$) | Meta-analysis | 3 | 2.22 (1.28-3.83) | 0.02 | 3.45 | 2 | 0.18 | -0.04 (-0.03) | 0.98 | 3,286 | 0.87 | rs71628146 | 0.35 |
| | | | | UK Biobank | 3 | 2.08 (0.71-6.11) | 0.35 | 5.04 | 2 | 0.08 | -0.07 (-0.07) | 0.99 | 3,232 | 0.35 | | |
| | | | | FinnGen | 3 | 2.3 (1.36-3.91) | 0.02 | 0.49 | 2 | 0.78 | -0.02 (-0.03) | 0.98 | 2754 | 0.31 | | |
| rs838718 *(DGKD)* | 2 | 234296650 | $2.13\times10^{-4}$ ($1.83\times10^{-4}$) | Meta-analysis | 4 | 0.05 (0.03-0.08) | $2.06\times10^{-34}$ | 2.9 | 3 | 0.41 | 0.01 (-0.02) | 0.98 | 3,268 | 0.99 | rs838717 | 0.96 |
| | | | | UK Biobank | 4 | 0.05 (0.02-0.14) | $4.88\times10^{-7}$ | 6.99 | 3 | 0.07 | $-1.56\times10^{-3}$ (-0.05) | 0.99 | 3,378 | 0.99 | rs838718 | 0.48 |
| | | | | FinnGen | 4 | 0.05 (0.03-0.10) | $3.45\times10^{-18}$ | 1.58 | 3 | 0.66 | 0.01 (-0.02) | 0.98 | 2,926 | 1.00 | rs838717 | 0.73 |
| rs12152922 *(SLC34A1)* | 5 | 176645599 | $2.98\times10^{-4}$ ($43.78\times10^{-4}$) | Meta-analysis | 9 | 0.09 (0.07-0.13) | $2.70\times10^{-38}$ | 14.56 | 8 | 0.07 | 0.01 (-0.01) | 0.98 | 2,409 | 1.00 | rs10051765 | 1.00 |
| | | | | UK Biobank | 9 | 0.07 (0.05-0.11) | $9.93\times10^{-40}$ | 7.54 | 8 | 0.48 | 0.03 (-0.01) | 0.99 | 2,582 | 1.00 | rs10051765 | 0.82 |
| | | | | FinnGen | 8 | 0.12 (0.07-0.18) | $1.91\times10^{-19}$ | 11.13 | 7 | 0.13 | $4.88\times10^{-3}$ (-0.02) | 0.98 | 2,128 | 1.00 | rs10051765 | 1.00 |
| rs10051765 *(SLC34A1)* | | | | Meta-analysis | 9 | 0.09 (0.07-0.13) | $2.70\times10^{-38}$ | 14.56 | 8 | 0.07 | 0.01 (-0.01) | 0.98 | 2,219 | 1.00 | rs10051765 | 1.00 |

| IV: 1 Mbp region from GWAS | | | | MR analysis | | | | | | | | Colocalisation | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Index variant (Gene) | CHR | POS | Mean $R^2$ ($SD$) | Inverse variance weighted analysis | | | | Heterogeneity | | | Intercept | | | | |
| | | | | Outcome | N SNPs | OR (95% CI) | P | Q | Q_df | Q P | $\beta$ (SE) | P | N SNPs | PP H4 | Candidate SNP | SNP PP |
| | 5 | 176799992 | $2.98\times10^{-4}$ ($3.78\times10^{-4}$) | UK Biobank | 9 | 0.07 (0.05-0.11) | $9.93\times10^{-40}$ | 7.54 | 8 | 0.48 | 0.03 (-0.01) | 0.99 | 2,359 | 1.00 | rs10051765 | 0.82 |
| | | | | FinnGen | 8 | 0.12 (0.07-0.18) | $1.91\times10^{-19}$ | 11.13 | 7 | 0.13 | $4.88\times10^{-3}$ (-0.02) | 0.98 | 2,022 | 1.00 | rs10051765 | 1.00 |
| | | | | FinnGen | 4 | 4.34 (2.25-8.38) | $1.83\times10^{-4}$ | 1.6 | 3 | 0.66 | -0.03 (-0.06) | 0.98 | 2,441 | 0.71 | rs2663546 | 0.02 |
| rs883622 *(FAM20A)* | 17 | 66442603 | $2.57\times10^{-4}$ ($2.36\times10^{-4}$) | Meta-analysis | 6 | 1.68 (0.9-3.13) | 0.22 | 15.82 | 5 | 0.01 | -0.04 (-0.01) | 0.53 | 3,189 | 0.97 | rs11867466 | 0.27 |
| | | | | UK Biobank | 6 | 2.35 (1.43-3.88) | $3.94\times10^{-3}$ | 2.74 | 5 | 0.74 | -0.02 (-0.02) | 0.99 | 3,438 | 0.61 | | |
| | | | | FinnGen | 6 | 1.21 (0.49-2.98) | 0.74 | 16.82 | 5 | 0 | -0.06 (-0.02) | 0.94 | 2,930 | 0.32 | | |
| rs11871728 *(FAM20A)* | 17 | 66703728 | $2.56\times10^{-4}$ ($2.32\times10^{-4}$) | Meta-analysis | 12 | 8.4 (4.33-16.29) | 0.01 | 31.64 | 11 | 0 | -0.04 (-0.01) | 0.03 | 3,165 | 0.97 | rs11867466 | 0.27 |
| | | | | UK Biobank | 12 | 2.35 (1.66-3.35) | $3.77\times10^{-5}$ | 5.49 | 11 | 0.91 | -0.02 (-0.01) | 0.99 | 3,261 | 0.61 | | |
| | | | | FinnGen | 12 | 11.36 (4.43-29.10) | 0.03 | 33.64 | 11 | 0 | -0.06 (-0.01) | 0.04 | 2,771 | 0.33 | | |

CHR=Chromosome; GWAS=Genome-wide association study; Mbp=Megabase pair; MR=Mendelian randomisation; N SNPs=Number of SNPs; OR=Odds ratio for kidney stone disease per genetically-instrumented standard deviation increase in trait; P=5% false discovery rate adjusted p-value; POS=Position based on NCBI Genome Build 37 (hg19); PP H4=posterior probability of full colocalisation; Q=Q statistic; Q_df=Q degrees of freedom; $R^2$=Variance in exposure explained by instrumental variable; SD=Standard deviation; SE=Standard error; SNP PP=Posterior probability explained by SNP; 95% CI=95% confidence interval.

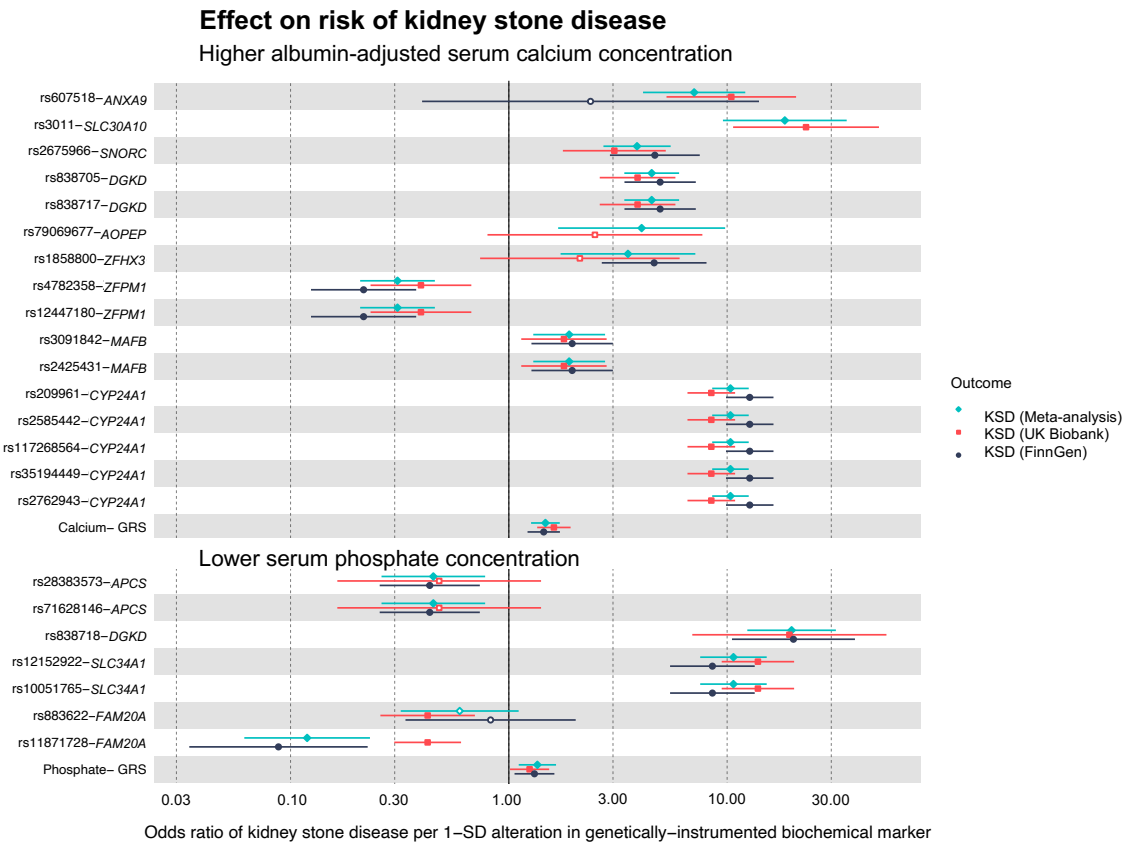



***Figure 5.2. Forest plot demonstrating regional Mendelian randomisation estimates for independent serum calcium and phosphate index-associated variants (±500 kilobase pairs) on odds of kidney stone disease.***

Non-significant results are indicated by hollow symbols.

GRS=Pan-genome Mendelian randomisation estimate for trait on risk of kidney stone disease; KSD=Kidney stone disease. Odds ratio is scaled per 1 standard deviation increase in serum calcium (0.08 mmol/L) or phosphate (0.016 mmol/L) concentrations.

For multi-trait colocalisation using `HyPrColoc`, I focused on the 17 1 Mbp regions with evidence of colocalisation and a causal effect of higher serum calcium or lower serum phosphate on increased KSD risk in MR (Table 5.3 and Table 5.4). Full colocalisation of serum calcium and phosphate concentrations with KSD and significant associations with these three traits at GWAS were detected in three regions in proximity to *DGKD*, *SLC34A1*, and *CYP24A1* (Table 5.5, Table 5.6, Figure 5.3). The candidate causal variant at the *ZFPM1* locus (rs12918968) does not reach replication significance threshold ($p < 5 \times 10^{-5}$) in UK Biobank KSD GWAS (Table 5.6). The candidate causal variant for the *MAFB*-associated locus (rs3091842) does not reach replication significance threshold in any of the three KSD GWAS datasets (Table 5.6). Considering this, I omitted these two loci from

subsequent studies.

***Table 5.5. Genomic regions of 1 megabase with evidence of multi-trait colocalisation between serum calcium, phosphate, and parathyroid hormone concentrations and kidney stone disease.***

| Index variant | CHR | POS | Gene | KSD study | PP H4 | Candidate SNP | SNP PP |
|---|---|---|---|---|---|---|---|
| rs838705 | 2 | 234273242 | *DGKD* | Meta-analysis | 1.00 | rs838717 | 1.00 |
| | | | | UK Biobank | 1.00 | | 1.00 |
| | | | | FinnGen | 1.00 | | 1.00 |
| rs838717 | 2 | 234296444 | *DGKD* | Meta-analysis | 1.00 | rs838717 | 1.00 |
| | | | | UK Biobank | 1.00 | | 1.00 |
| | | | | FinnGen | 1.00 | | 1.00 |
| rs838718 | 2 | 234296650 | *DGKD* | Meta-analysis | 1.00 | rs838717 | 1.00 |
| | | | | UK Biobank | 1.00 | | 1.00 |
| | | | | FinnGen | 1.00 | | 1.00 |
| rs12152922 | 5 | 176645599 | *SLC34A1* | Meta-analysis | 1.00 | rs10051765 | 1.00 |
| | | | | UK Biobank | 1.00 | | 1.00 |
| | | | | FinnGen | 1.00 | | 1.00 |
| rs10051765 | 5 | 176799992 | *SLC34A1* | Meta-analysis | 1.00 | rs10051765 | 1.00 |
| | | | | UK Biobank | 1.00 | | 1.00 |
| | | | | FinnGen | 1.00 | | 1.00 |
| rs4782358 | 16 | 88485506 | *ZFPM1* | Meta-analysis | 0.82 | rs12918968 | 0.84 |
| | | | | FinnGen | 0.82 | | 0.77 |
| rs12447180 | 16 | 88485506 | *ZFPM1* | Meta-analysis | 0.82 | rs12918968 | 0.84 |
| | | | | FinnGen | 0.82 | | 0.77 |
| rs3091842 | 20 | 39344272 | *MAFB* | Meta-analysis | 0.98 | rs3091842 | 1.00 |
| | | | | FinnGen | 0.96 | | 1.00 |
| rs2425431 | 20 | 39344272 | *MAFB* | Meta-analysis | 0.98 | rs3091842 | 1.00 |
| | | | | FinnGen | 0.96 | | 1.00 |
| rs209961 | 20 | 52715154 | *CYP24A1* | Meta-analysis | 1.00 | rs6127099 | 1.00 |
| | | | | UK Biobank | 1.00 | | 1.00 |
| | | | | FinnGen | 1.00 | | 1.00 |
| rs2585442 | 20 | 52737123 | *CYP24A1* | Meta-analysis | 1.00 | rs6127099 | 1.00 |
| | | | | UK Biobank | 1.00 | | 1.00 |
| | | | | FinnGen | 1.00 | | 1.00 |
| rs117268564 | 20 | 52738434 | *CYP24A1* | Meta-analysis | 1.00 | rs6127099 | 1.00 |
| | | | | UK Biobank | 1.00 | | 1.00 |
| | | | | FinnGen | 1.00 | | 1.00 |
| rs35194449 | 20 | 52742047 | *CYP24A1* | Meta-analysis | 1.00 | rs6127099 | 1.00 |
| | | | | UK Biobank | 1.00 | | 1.00 |
| | | | | FinnGen | 1.00 | | 1.00 |
| rs2762943 | 20 | 52790786 | *CYP24A1* | Meta-analysis | 1.00 | rs6127099 | 1.00 |
| | | | | UK Biobank | 1.00 | | 1.00 |
| | | | | FinnGen | 1.00 | | 1.00 |

KSD=Kidney stone disease; PP=Posterior-probability; SNP=Single nucleotide polymorphism.

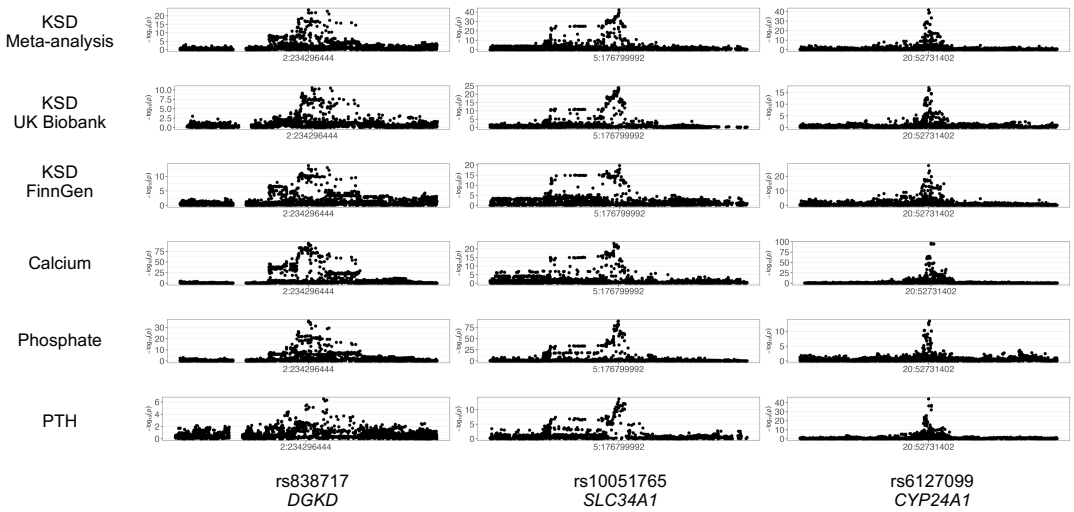



***Figure 5.3. Locus-zoom plots of 1 megabase pair regions with multi-trait colocalisation and causal effects between mineral metabolism traits and kidney stone disease.***

Genome-wide P values (-log10) are plotted against respective positions on each chromosome. The central line annotated on each plot corresponds to the index variant from calcium and phosphate genome-wide association studies.

KSD=Kidney stone disease; PTH=Parathyroid hormone.

***Table 5.6. Associations of candidate causal variants identified from HyPrColoc with mineral metabolism traits and kidney stone disease.***

| | | | | | | |
|---|---|---|---|---|---|---|
| **Candidate causal variant** | **Gene** | ***DGKD*** | ***SLC34A1*** | ***ZFPM1*** | ***MAFB*** | ***CYP24A1*** |
| | **SNP** | rs838717 | rs10051765 | rs12918968 | rs3091842 | rs6127099 |
| | **CHR** | 2 | 5 | 16 | 20 | 20 |
| | **POS** | 234296444 | 176799992 | 88520452 | 39344272 | 52731402 |
| | **EA** | G | C | C | A | A |
| | **NEA** | A | T | A | G | T |
| | **EAF** | 0.43 | 0.33 | 0.44 | 0.08 | 0.72 |
| **Serum calcium GWAS [597]** | ***β*** **(SE)** | 0.05 ($2.49\times10^{-3}$) | 0.03 ($2.62\times10^{-3}$) | -0.04 ($2.49\times10^{-3}$) | 0.11 ($6.20\times10^{-3}$) | 0.06 ($2.82\times10^{-3}$) |
| | **P** | $1.40\times10^{-94}$ | $5.30\times10^{-22}$ | $5.20\times10^{-56}$ | $2.90\times10^{-72}$ | $7.60\times10^{-97}$ |
| **Serum phosphate GWAS [597]** | ***β*** **(SE)** | -0.03 ($2.47\times10^{-3}$) | -0.05 ($2.59\times10^{-3}$) | 0.02 ($2.46\times10^{-3}$) | -0.05 ($6.14\times10^{-3}$) | 0.02 ($2.79\times10^{-3}$) |
| | **P** | $7.20\times10^{-36}$ | $9.80\times10^{-90}$ | $4.00\times10^{-10}$ | $2.60\times10^{-17}$ | $1.40\times10^{-13}$ |
| **Serum PTH GWAS** | ***β*** **(SE)** | 0.02 ($4.00\times10^{-3}$) | -0.04 ($5.00\times10^{-3}$) | 0.01 ($4.00\times10^{-3}$) | 0.04 (0.01) | -0.07 ($5.00\times10^{-3}$) |
| | **P** | $4.25\times10^{-6}$ | $2.34\times10^{-14}$ | $1.17\times10^{-3}$ | $1.59\times10^{-4}$ | $1.01\times10^{-44}$ |
| **KSD GWAS Meta-analysis** | **OR (95%CI)** | 1.10 (1.08-1.12) | 1.14 (1.12-1.16) | 1.05 (1.04-1.07) | 1.07 (1.04-1.11) | 1.16 (1.13-1.18) |
| | **P** | $3.08\times10^{-24}$ | $6.45\times10^{-43}$ | $1.37\times10^{-8}$ | $9.24\times10^{-5}$ | $2.06\times10^{-42}$ |
| **KSD UK Biobank GWAS** | **OR (95%CI)** | 1.10 (1.07-1.13) | 1.16 (1.13-1.19) | 1.04 (1.01-1.07) | 1.06 (1.00-1.14) | 1.14 (1.10-1.17) |
| | **P** | $2.60\times10^{-11}$ | $1.30\times10^{-24}$ | $3.50\times10^{-3}$ | 0.06 | $1.60\times10^{-16}$ |
| **KSD FinnGen GWAS** | **OR (95%CI)** | 1.10 (1.07-1.13) | 1.13 (1.10-1.15) | 1.07 (1.04-1.09) | 1.08 (1.03-1.13) | 1.17 (1.14-1.21) |
| | **P** | $1.71\times10^{-14}$ | $1.94\times10^{-20}$ | $4.61\times10^{-7}$ | $5.53\times10^{-4}$ | $4.22\times10^{-28}$ |

CHR=Chromosome; GWAS=Genome-wide association study; KSD=Kidney stone disease; P=P-value; PTH=Parathyroid hormone; POS=Psition based on NCBI Genome Build 37 (hg19); OR=Odds ratio; SE=Otandard error; 95% CI=95% confidence interval.

The 95% credible set of candidate causal variants at the remaining three loci comprised just one variant per locus. Each variant has SNP PP 1.00 (Table 5.5). rs838717 is an intronic variant in *DGKD* that is predicted to function as a transcription factor binding site. A G allele at rs838717 increases the odds of KSD

by 10% and is associated with 0.05 mmol/L higher serum calcium and 0.03 mmol/L lower serum phosphate concentrations (Table 5.6). rs10051765 is an intergenic variant ≈6 Kbp upstream of *SLC34A1*. A C allele at rs10051765 increases the odds of KSD by 13-16% and is associated with 0.03 mmol/L higher serum calcium and 0.05 mmol/L lower serum phosphate concentrations (Table 5.6). Finally, rs6127099 is an intergenic variant approximately 50 Kbp downstream of *CYP24A1* (Table 5.6). An A allele at rs6127099 increases the odds of KSD by 14-17% and is associated with 0.06 mmol/L higher serum calcium and 0.02 mmol/L higher serum phosphate concentrations (Table 5.6).

Variants in proximity to *DGKD*, *SLC34A1*, and *CYP24A1* were associated with KSD across UK Biobank KSD GWAS (Chapter 3 Section 3.3.1, Table 3.1, Figure 3.1), European ancestry KSD GWAS meta-analysis (Chapter 3 Section 3.3.2, Table 3.4, Figure 3.5), and trans-ancestry KSD GWAS meta-analysis (Chapter 3 Section 3.3.3, Table 3.5, Figure 3.7). These findings align with published GWAS and GWAS meta-analyses [119, 226, 241–243].

Together, these genetic epidemiological and colocalisation analyses underscore the importance of *DGKD*, *SLC34A1*, and *CYP24A1* to KSD and serum calcium and phosphate concentrations. They identify putative causal variants that contribute to a shared genetic architecture at each locus, highlighting an opportunity to move beyond genomic associations and explore their biological relevance.

## 5.4.2 The functional relevance of rs838717

### 5.4.2.1 The regulatory role of rs838717 in gene transcription as inferred through chromatin coaccessibility and RNA expression in renal cell types

To proxy gene activity, I used snATAC-seq data, and mapped chromatin accessibility on chromosome 2 to identify coaccessible regions linking rs838717±100 bp to promoters of protein-coding genes, defined using Ensembl v113 [466]. With these data, I revealed a significant and positive correlation between gene activity, related to rs838717±100 bp, and RNA expression for *INPP5D*, *NGEF*, *DGKD*, *UGT1A6*,

and *USP40* (Table 5.7, Figure 5.4 and Figure 5.5). This suggests that rs838717 may contribute to the regulation of these genes and play a potential functional role in renal cell types. Though there is low gene activity across all cell types, the correlation of *INPP5D* appears to be largely driven by gene expression in endothelial cells (Figure 5.4). *NGEF*, *DGKD*, *UGT1A6*, and *USP40* activity and expression are relatively high in proximal tubular cells in comparison to other renal cell types.

***Table 5.7. Significant Pearson correlation coefficients between gene activity and expression on chromosome 2 across renal cell types.***

| Gene | Maximum activity | Maximum expression | Pearson $r^2$ | Pearson P | Cell type | Mean activity | Mean expression |
|---|---|---|---|---|---|---|---|
| *INPP5D* | $2.08\times10^{-5}$ | 1.55 | 0.98 | $5.40\times10^{-4}$ | ENDO | $1.89\times10^{-5}$ | 1.41 |
| | | | | | FB | $7.99\times10^{-11}$ | 0.41 |
| | | | | | PT | $2.68\times10^{-11}$ | 0.24 |
| | | | | | DT | $7.69\times10^{-12}$ | 0.22 |
| | | | | | PEC | $3.57\times10^{-11}$ | 0.16 |
| | | | | | TAL | $3.24\times10^{-11}$ | 0.14 |
| *NGEF* | 0.02 | 1.64 | 0.94 | $1.40\times10^{-3}$ | PT | 0.02 | 1.49 |
| | | | | | ICA | $1.76\times10^{-6}$ | 0.65 |
| | | | | | PEC | $2.66\times10^{-9}$ | 0.51 |
| | | | | | FB | $7.17\times10^{-9}$ | 0.44 |
| | | | | | DT | $2.45\times10^{-7}$ | 0.4 |
| | | | | | ENDO | $1.83\times10^{-5}$ | 0.27 |
| | | | | | TAL | $2.82\times10^{-9}$ | 0.23 |
| *DGKD* | 0.04 | 1.8 | 0.91 | $1.90\times10^{-3}$ | GLOM | 0.04 | 1.64 |
| | | | | | PT | 0.03 | 1.38 |
| | | | | | FB | 0.04 | 1.27 |
| | | | | | PEC | 0.04 | 1.09 |
| | | | | | ICA | 0.02 | 0.79 |
| | | | | | ENDO | $2.36\times10^{-3}$ | 0.56 |
| | | | | | DT | $2.06\times10^{-3}$ | 0.56 |
| | | | | | TAL | $1.90\times10^{-3}$ | 0.37 |
| *UGT1A6* | 0.03 | 3.72 | 0.9 | 0.01 | PT | 0.03 | 3.38 |
| | | | | | DT | 0.02 | 0.74 |
| | | | | | ICA | 0.01 | 0.5 |
| | | | | | GLOM | $6.56\times10^{-9}$ | 0.38 |
| | | | | | PEC | $1.75\times10^{-7}$ | 0.38 |
| | | | | | TAL | $4.24\times10^{-4}$ | 0.19 |
| | | | | | ENDO | $5.34\times10^{-9}$ | 0.12 |
| *USP40* | 0.04 | 1.38 | 0.77 | 0.02 | PT | 0.03 | 1.25 |
| | | | | | ICA | 0.02 | 0.85 |
| | | | | | PEC | 0.04 | 0.82 |
| | | | | | GLOM | 0.01 | 0.68 |
| | | | | | DT | 0.02 | 0.55 |
| | | | | | FB | 0.02 | 0.46 |
| | | | | | TAL | 0.01 | 0.33 |
| | | | | | ENDO | $1.88\times10^{-3}$ | 0.20 |

Genes arranged in descending order of significance. Cell types arranged in descending order of gene expression. DT=Distal tubule; ENDO=Endothelial; FB=Fibroblast; GLOM=Glomerular; ICA=Intercalated A; P=P-value; PEC=Parietal epithelial; PT=Proximal tubule; TAL=Thick ascending limb of the loop of Henle.

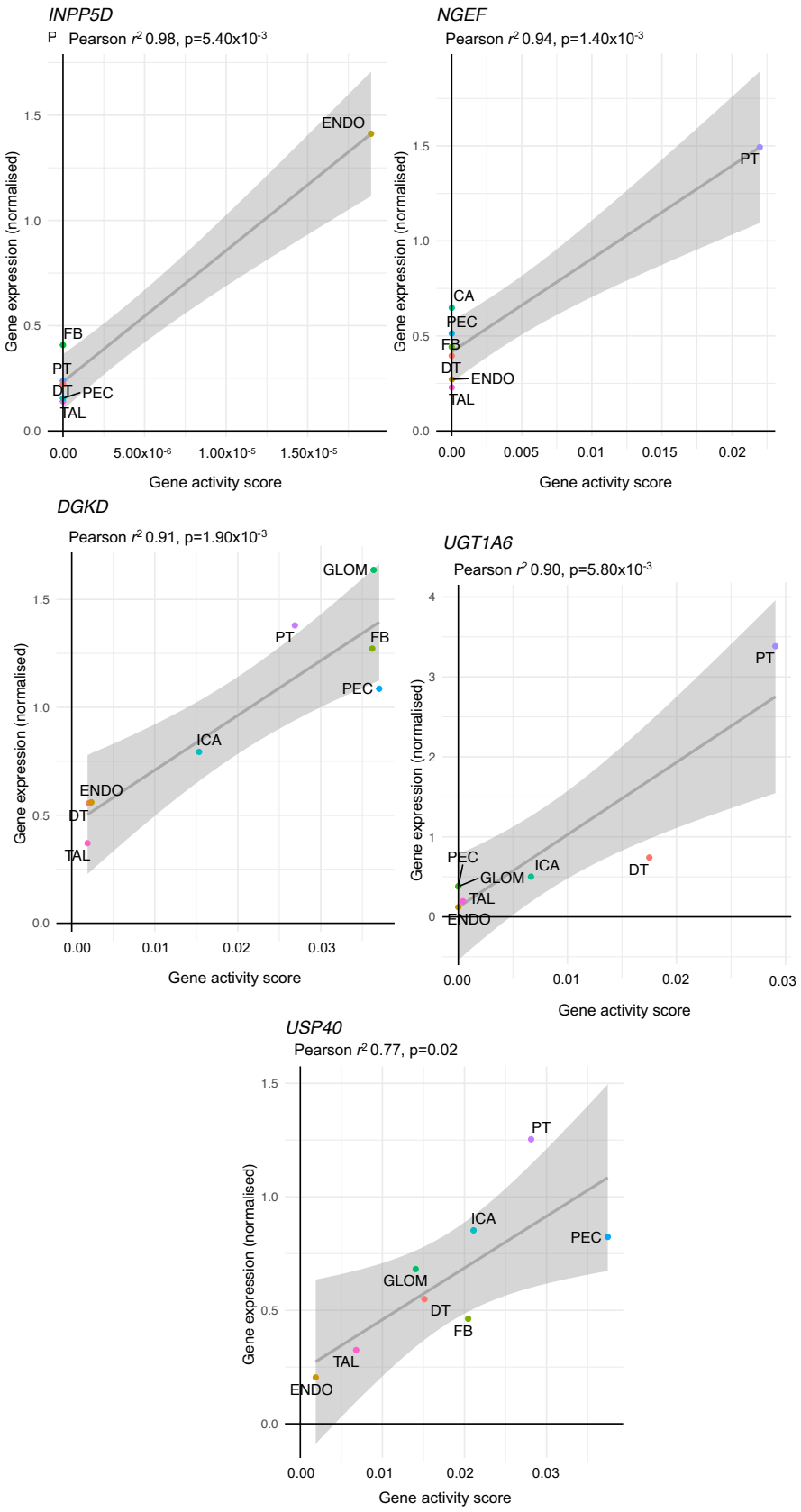



***Figure 5.4. Pearson correlation coefficients of gene activity, inferred from rs838717 coaccessibility, and RNA expression of INPP5D, NGEF, DGKD, UGT1A6, and USP40.***

Gene activity score reflects the weighted coaccessibility between rs838717 and gene promoter regions. Gene expression reflects normalised gene expression from snRNA-seq data.
DT=Distal tubule; ENDO=Endothelial; FB=Fibroblast; GLOM=Glomerular; ICA=Intercalated A; P=P-value; PEC=Parietal epithelial; PT=Proximal tubule; TAL=Thick ascending limb of the loop of Henle.

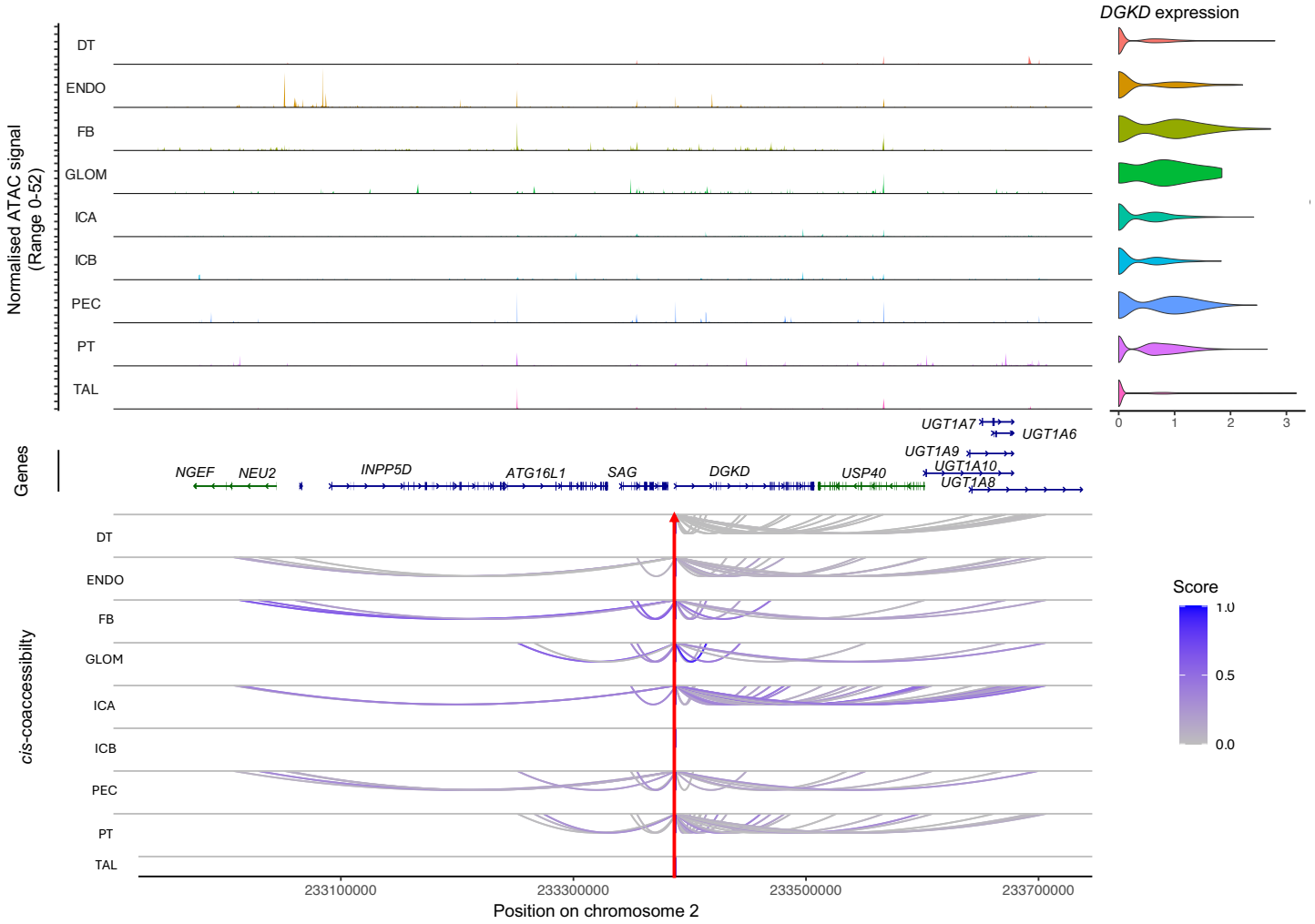



***Figure 5.5. Cis-coaccessibility of rs838717 and protein-coding gene promoters correlates with RNA expression in renal cell lines.***
**Top**: Normalised ATAC signal across renal cell types. Peaks represent increased chromatin accessibility. **Bottom**: Links denote *cis*-coaccessible regions, contributing to the gene activity score. Link colours indicate the magnitude of the Cicero coaccessibility score. The vertical red arrow marks the location of rs838717. **Violin plot**: *DGKD* gene expression in each renal cell type.
DT=distal tubule; ENDO=endothelial; FB=fibroblast; GLOM=glomerular; ICA=intercalated A; PEC=parietal epithelial; PT=proximal tubule; TAL=thick ascending limb of the loop of Henle. Coordinates are reported in genome build hg38.

#### 5.4.2.2 Regulatory potential of rs838717 inferred through chromatin interactions with *USP40* and *DGKD*

To investigate the physical proximity and potential regulatory role of rs838717 in relation to *INPP5D*, *NGEF*, *DGKD*, *UGT1A6*, and *USP40* expression, I analysed Capture-C data from RPTEC, HK2, and HEK293 cell lines. Using this Bayesian approach, I aimed to uncover chromatin interactions between rs838717 and promoter regions of these genes, investigating whether rs838717 physically connects to, and potentially regulates, these loci.

Chromatin interactions were observed between the genomic region encompassing rs838717 and the promoter regions of *USP40* and *DGKD* (Table 5.8, Figure 5.6). These findings reinforce the hypothesis that rs838717 exerts regulatory effects on these genes and may implicate them in the pathogenesis of KSD and in calcium and phosphate homeostasis. Interactions with the promoter regions of *NGEF* were present in RPTEC and HEK293 cell lines, and interactions with the promoter region of *INPP5D* present in HEK293 cells. There were no interactions with promoter region of *UGT1A6*. Together, these analyses provide mechanistic insights into the role of rs838717 in KSD pathophysiology and mineral metabolism.

***Table 5.8. Significant interactions of rs838717 with gene promoters in renal cell lines identified in Capture-C analysis.***

| Cell line | Gene | Promoter | | | MPPC score |
|---|---|---|---|---|---|
| | | ENSR ID | Coordinates (hg38) | Distance from rs838717 | |
| RPTEC | *NGEF* | 2_95FQF3 | 2:233691257-233691617 | 303.82 Kbp downstream | 0.99 |
| | | 00001361537 | 2:233690117-233691720 | 303.92 Kbp downstream | 0.99 |
| | | 00001361538 | 2:233691721-233692900 | 305.10 Kbp downstream | 0.99 |
| | | 00001361479 | 2:232885526-232885862 | 501.94 Kbp upstream | 0.21 |
| | *USP40* | 00001361523 | 2:233481632-233482258 | 94.46 Kbp downstream | 0.21 |
| | | 00001361529 | 2:233511748-233512491 | 124.69 Kbp downstream | 0.02 |
| | *DGKD* | 00001361521 | 2:233447206-233448040 | 60.24 Kbp downstream | 0.03 |
| | | 00001361511 | 2:233353955-233354582 | 33.22 Kbp upstream | 0.02 |
| | | 2_95FLRH | 2:233353955-233354504 | 33.29 Kbp upstream | 0.02 |
| HK2 | *DGKD* | 00001361511 | 2:233353955-233354582 | 33.22 Kbp upstream | 0.27 |
| | | 2_95FLRH | 2:233353955-233354504 | 33.29 Kbp upstream | 0.27 |
| | *USP40* | 00001361530 | 2:233512492-233514382 | 126.58 Kbp downstream | 0.02 |
| | | 00001361527 | 2:233491329-233491719 | 103.92 Kbp downstream | 0.01 |
| | | 00001361534 | 2:233565580-233566699 | 178.90 Kbp downstream | 0.01 |
| | | 2_95FP3P | 2:233565580-233566723 | 178.93 Kbp downstream | 0.01 |
| | | 00001361525 | 2:233489451-233490200 | 102.40 Kbp downstream | 0.01 |
| | | 00001361523 | 2:233481632-233482258 | 94.46 Kbp downstream | 0.01 |
| HEK293 | *NGEF* | 00001361479 | 2:232885526-232885862 | 501.94 Kbp upstream | 0.98 |
| | | 00001361487 | 2:233013246-233013561 | 374.24 Kbp upstream | 0.94 |
| | | 00001361485 | 2:232928144-232929060 | 458.74 Kbp upstream | 0.16 |
| | | 2_95FG89 | 2:232928144-232928979 | 458.82 Kbp upstream | 0.16 |
| | *INPP5D* | 00001361492 | 2:233074772-233076272 | 311.53 Kbp upstream | 0.98 |
| | | 00001361495 | 2:233163202-233164311 | 223.49 Kbp upstream | 0.04 |
| | *USP40* | 00001361525 | 2:233489451-233490200 | 102.40 Kbp downstream | 0.05 |
| | | 00001361534 | 2:233565580-233566699 | 178.90 Kbp downstream | 0.05 |
| | | 2_95FP3P | 2:233565580-233566723 | 178.93 Kbp downstream | 0.05 |
| | | 00001361529 | 2:233511748-233512491 | 124.69 Kbp downstream | 0.04 |
| | | 00001361530 | 2:233512492-233514382 | 126.58 Kbp downstream | 0.04 |
| | | 00001361527 | 2:233491329-233491719 | 103.92 Kbp downstream | 0.04 |
| | | 00001361533 | 2:233523242-233523400 | 135.60 Kbp downstream | 0.04 |
| | | 00001361523 | 2:233481632-233482258 | 94.46 Kbp downstream | 0.03 |
| | | 2_B3PCJJ | 2:233566724-233567281 | 179.48 Kbp downstream | 0.03 |
| | | 00001361535 | 2:233566700-233567600 | 179.80 Kbp downstream | 0.03 |
| | *DGKD* | 00001361521 | 2:233447206-233448040 | 60.24 Kbp downstream | 0.04 |
| | | 00001361511 | 2:233353955-233354582 | 33.21 Kbp upstream | 0.04 |
| | | 2_95FLRH | 2:233353955-233354504 | 33.29 Kbp upstream | 0.04 |
| | | 00001361520 | 2:233417751-233419251 | 31.45 Kbp downstream | 0.02 |

For each cell type, results are arranged in descending order of MPPC score.
ENSR ID=Ensembl stable ID for regulatory features; HEK293=Human embryonic kidney 293 cells; HK2=Human kidney 2 immortalised proximal tubule epithelial cells; MPPC=Marginal posterior probability of contact; RPTEC=Renal proximal tubule epithelial cells.

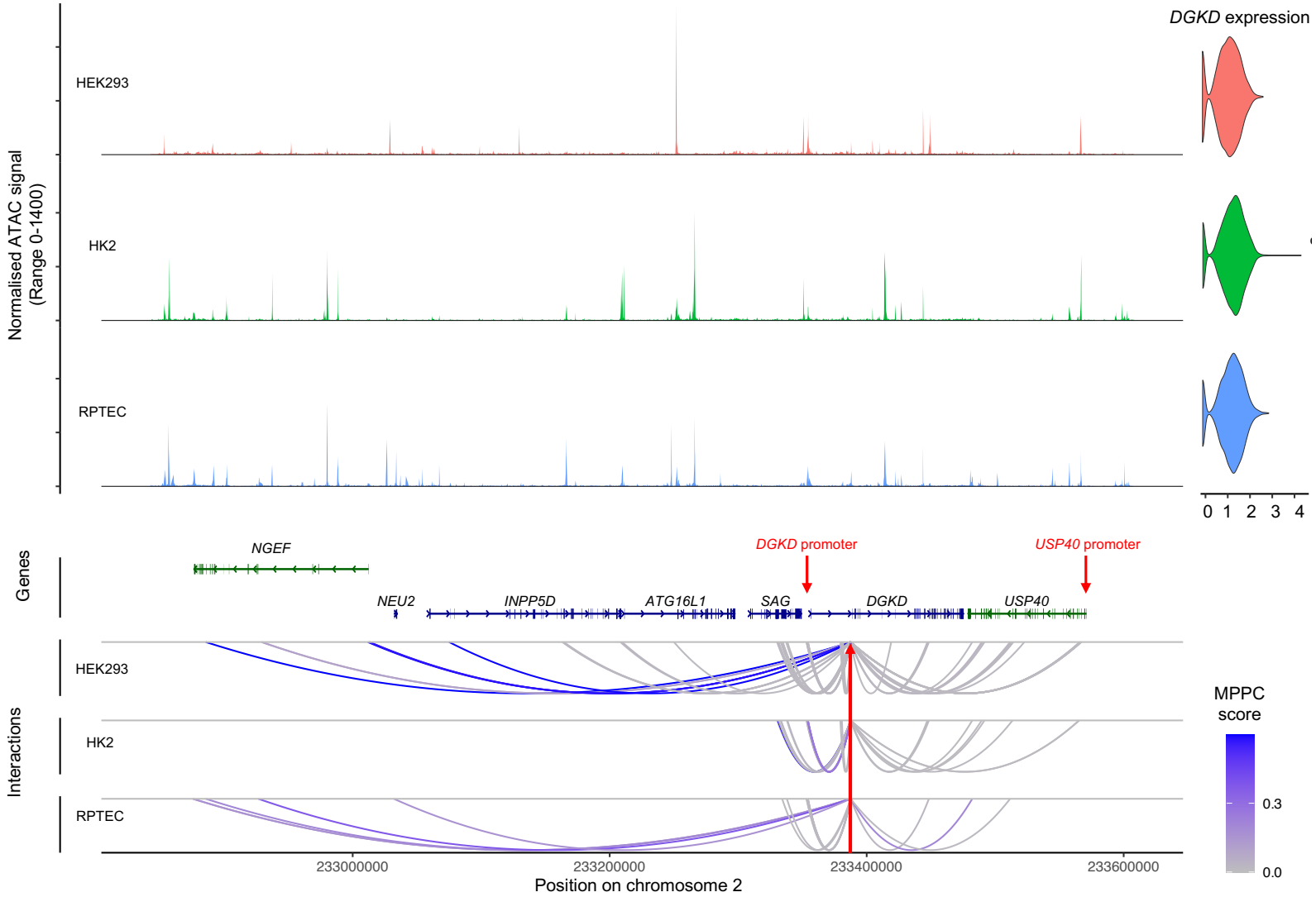



***Figure 5.6. Cature-C: chromatin interactions of rs838717 and gene promoters across three renal cell lines.***

**Top**: Normalised ATAC signal across renal cell types. Peaks represent increased chromatin accessibility. **Bottom**: Links represent significant chromatin interactions between rs838717 and promoter coordinates of protein-coding genes, defined using Ensembl v113 [466]. Colours of links reflect the magnitude of the marginal posterior probability of contact. The vertical red arrow denotes the location of rs838717. **Violin plot**: Expression of *DGKD* in each renal cell type.
HEK293=Human embryonic kidney 293 cells; HK2=Human kidney 2 immortalised proximal tubule epithelial cells; MPPC=Marginal posterior probability of contact; RPTEC=Renal proximal tubule epithelial cells. Coordinates are reported in genome build hg38.

#### 5.4.2.3 rs838717 regulates *DGKD* expression and increases kidney stone disease risk through the calcium-sensing receptor signalling pathway

Having established the role of rs838717 in regulating *USP40* and *DGKD* expression, I investigated the biological relevance of these genes to determine how the variant may contribute to an increased risk of KSD by altering calcium and phosphate homeostasis.

*USP40*, whose promoter is ≈181 Kbp downstream of rs838717, encodes ubiquitin carboxyl-terminal hydrolase 40 (USP40), a deubiquitinating enzyme regulating cellular proteins through the coordinated action of ubiquitin-conjugating and deubiquitinating enzymes. USP40 plays a key role in maintaining cellular homeostasis [599]. In zebrafish, USP40 knockdown disrupts glomerular barrier integrity by reducing nestin expression, an important intermediate filament protein critical for podocyte function [600]. In contrast, USP40 knockout in mice upregulates nestin expression without affecting glomerular permeability [600, 601]. In humans, USP40 protects against acute lung injury by attenuating lipopolysaccharide- and thrombin-induced microvascular endothelial barrier disruption and inflammatory responses [599]. Furthermore, genetic studies link USP40 variants to bone mineral density in South Asian populations and to serum bilirubin, calcium, phosphate, and urate levels in individuals of European ancestry [602–605].

*DGKD*, the gene in which rs838717 is located, encodes DGK$\delta$, a component of the CaSR signalling pathway, which is expressed in calcitropic tissues, including the parathyroid glands, bone, and renal cortex [158]. DGK$\delta$ phosphorylates diacylglycerol, facilitating CaSR-mediated membrane ruffling, activating protein kinase C signalling cascades, including mitogen-activated protein kinase pathways, and inducing intracellular calcium release [55, 606]. Reduced DGK$\delta$ expression impairs CaSR signal transduction *in vitro* [158]. As described in Chapter 1 Section 1.1.6.7, gain-of-function mutations in components of the CaSR signalling pathway cause autosomal dominant hypocalcaemia in which approximately 10% of patients present with hypercalciuria (Chapter 1 Table 1.2) [164, 607]. Conversely, loss-of-function mutations result in FHH (Chapter 1 Table 1.2). Both autosomal

dominant hypocalcaemia and FHH are associated with KSD.

The KSD-associated G allele of rs838717 associates with higher serum calcium and lower serum phosphate concentrations (Table 5.6). This phenotype is consistent with a, likely attenuated, mirror of FHH, suggesting that rs838717-G may impair CaSR signalling. These data indicate that rs838717 regulates *DGKD* expression by altering chromatin interactions at the gene promoter, providing a plausible mechanism for this association.

In summary, these studies corroborate that rs838717 exerts regulatory effects on *DGKD* expression, likely disrupting the CaSR signalling pathway. This disruption explains the observed changes in calcium and phosphate homeostasis and the increased risk of KSD. These findings support a causal role for rs838717 in KSD pathogenesis and highlight the importance of this locus in calcium and phosphate metabolism.

### 5.4.3 The functional relevance of rs10051765

#### 5.4.3.1 The regulatory role of rs10051765 in gene transcription via chromatin coaccessibility and RNA expression in renal cell types

By integrating chromatin coaccessibility and RNA expression data, I identified significant correlations between rs10051765±100 bp and the transcriptional activity of protein-coding genes on chromosome 5. Specifically, positive correlations were observed for *SLC34A1*, *FAM193B*, *B4GALT7*, and *PRELID1*, implicating rs10051765 in the regulation of these genes (Table 5.9, Figure 5.7, Figure 5.8). The inverse correlation with *GRK6* suggests that rs10051765 may act as a repressive regulatory element for this gene.

***Table 5.9. Significant Pearson correlation coefficients between gene activity and expression on chromosome 5 across renal cell types.***

| Gene | Maximum activity | Maximum expression | Pearson $r^2$ | Pearson P | Cell type | Mean activity | Mean expression |
|---|---|---|---|---|---|---|---|
| *SLC34A1* | 0.03 | 1.74 | 0.98 | $7.20\times10^{-6}$ | PT | 0.03 | 1.58 |
| | | | | | ICA | $3.74\times10^{-3}$ | 0.35 |
| | | | | | DT | 0.01 | 0.34 |
| | | | | | ICB | 0.01 | 0.26 |
| | | | | | FB | $2.22\times10^{-4}$ | 0.22 |
| | | | | | TAL | $8.91\times10^{-4}$ | 0.19 |
| | | | | | GLOM | $6.48\times10^{-8}$ | 0.18 |
| | | | | | PEC | $2.33\times10^{-7}$ | 0.16 |
| | | | | | ENDO | $1.08\times10^{-5}$ | 0.10 |
| *GRK6* | 0.05 | 1.62 | -0.83 | 0.01 | PT | 0.03 | 1.48 |
| | | | | | ICA | 0.03 | 1.40 |
| | | | | | DT | 0.03 | 1.11 |
| | | | | | ICB | 0.04 | 1.01 |
| | | | | | FB | 0.04 | 0.98 |
| | | | | | GLOM | 0.05 | 0.86 |
| | | | | | ENDO | 0.04 | 0.80 |
| | | | | | TAL | 0.04 | 0.77 |
| | | | | | PEC | 0.04 | 0.69 |
| *FAM193B* | 0.02 | 0.92 | 0.84 | 0.01 | GLOM | 0.02 | 0.83 |
| | | | | | FB | 0.01 | 0.63 |
| | | | | | ICA | 0.01 | 0.5 |
| | | | | | PEC | 0.01 | 0.46 |
| | | | | | PT | 0.01 | 0.38 |
| | | | | | DT | 0.01 | 0.37 |
| | | | | | TAL | 0.01 | 0.27 |
| | | | | | ENDO | 0.01 | 0.26 |
| | | | | | ICB | 0.01 | 0.24 |
| *B4GALT7* | 0.03 | 0.53 | 0.74 | 0.02 | ICA | 0.03 | 0.48 |
| | | | | | DT | 0.02 | 0.41 |
| | | | | | PT | 0.02 | 0.33 |
| | | | | | ICB | 0.03 | 0.31 |
| | | | | | FB | 0.02 | 0.3 |
| | | | | | GLOM | 0.01 | 0.29 |
| | | | | | TAL | 0.02 | 0.26 |
| | | | | | PEC | 0.01 | 0.21 |
| | | | | | ENDO | $3.77\times10^{-4}$ | 0.15 |
| *PRELID1* | 0.03 | 0.44 | 0.7 | 0.03 | PT | 0.03 | 0.40 |
| | | | | | GLOM | $3.77\times10^{-4}$ | 0.27 |
| | | | | | FB | 0.030 | 0.27 |
| | | | | | ICA | 0.02 | 0.26 |
| | | | | | DT | 0.02 | 0.24 |
| | | | | | TAL | 0.01 | 0.19 |
| | | | | | PEC | 0.01 | 0.18 |
| | | | | | ENDO | 0.01 | 0.17 |
| | | | | | ICB | 0.01 | 0.15 |

Genes arranged in descending order of significance. Cell types arranged in descending order of gene expression. DT=Distal tubule; ENDO=Endothelial; FB=Fibroblast; GLOM=Glomerular; ICA=Intercalated A; P=P-value; PEC=Parietal epithelial; PT=Proximal tubule; TAL=Thick ascending limb of the loop of Henle.

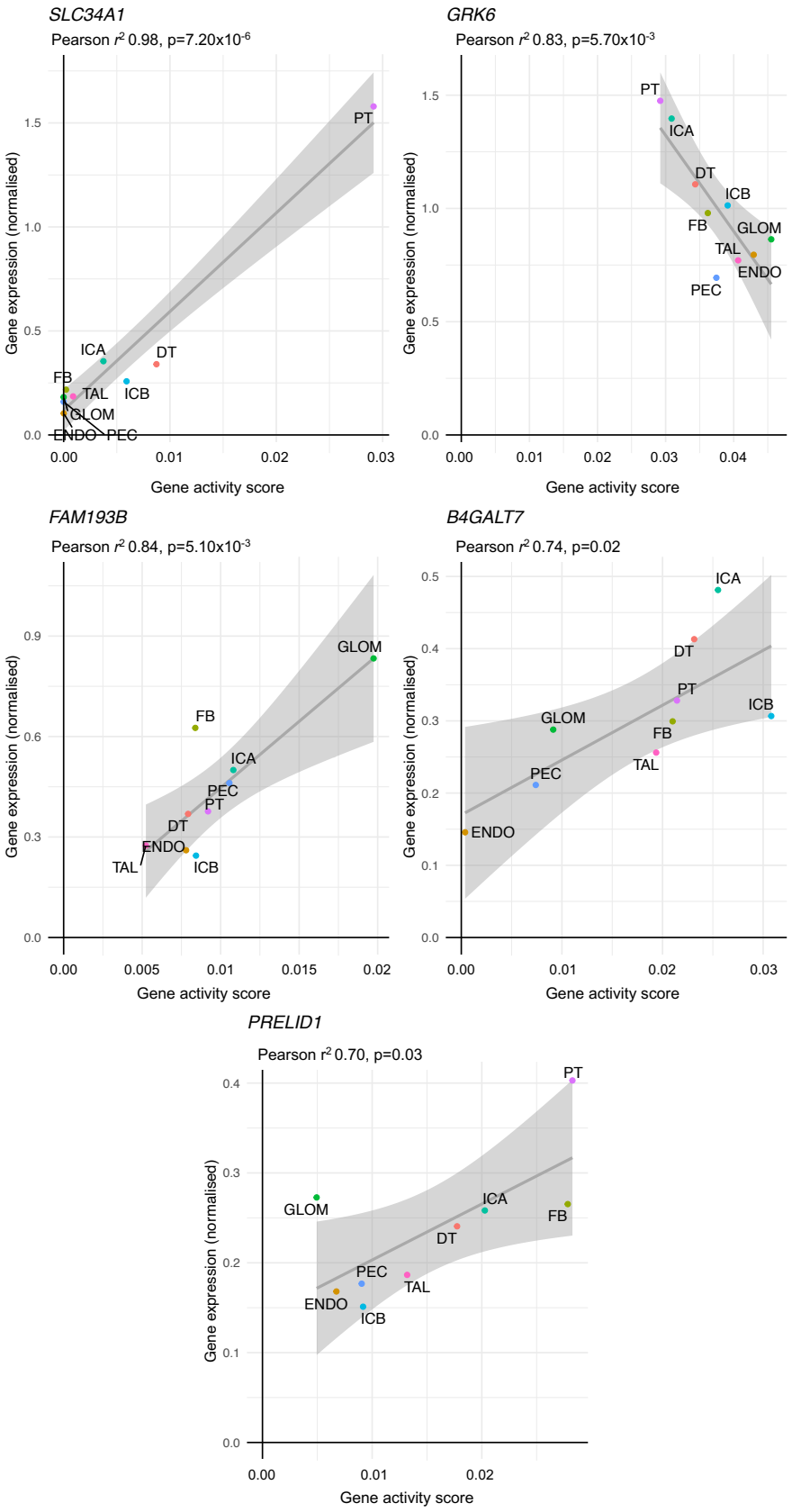



***Figure 5.7. Pearson correlation coefficients of gene activity, inferred from rs10051765 coaccessibility, and RNA expression of SLC34A1, GRK6, FAM193B, B4GALT7, and PRELID1.***

Gene activity score reflects the weighted coaccessibility between rs10051765 and gene promoter regions. Gene expression reflects normalised gene expression from snRNA-seq data.
DT=Distal tubule; ENDO=Endothelial; FB=Fibroblast; GLOM=Glomerular; ICA=Intercalated A; P=P-value; PEC=Parietal epithelial; PT=Proximal tubule; TAL=Thick ascending limb of the loop of Henle.

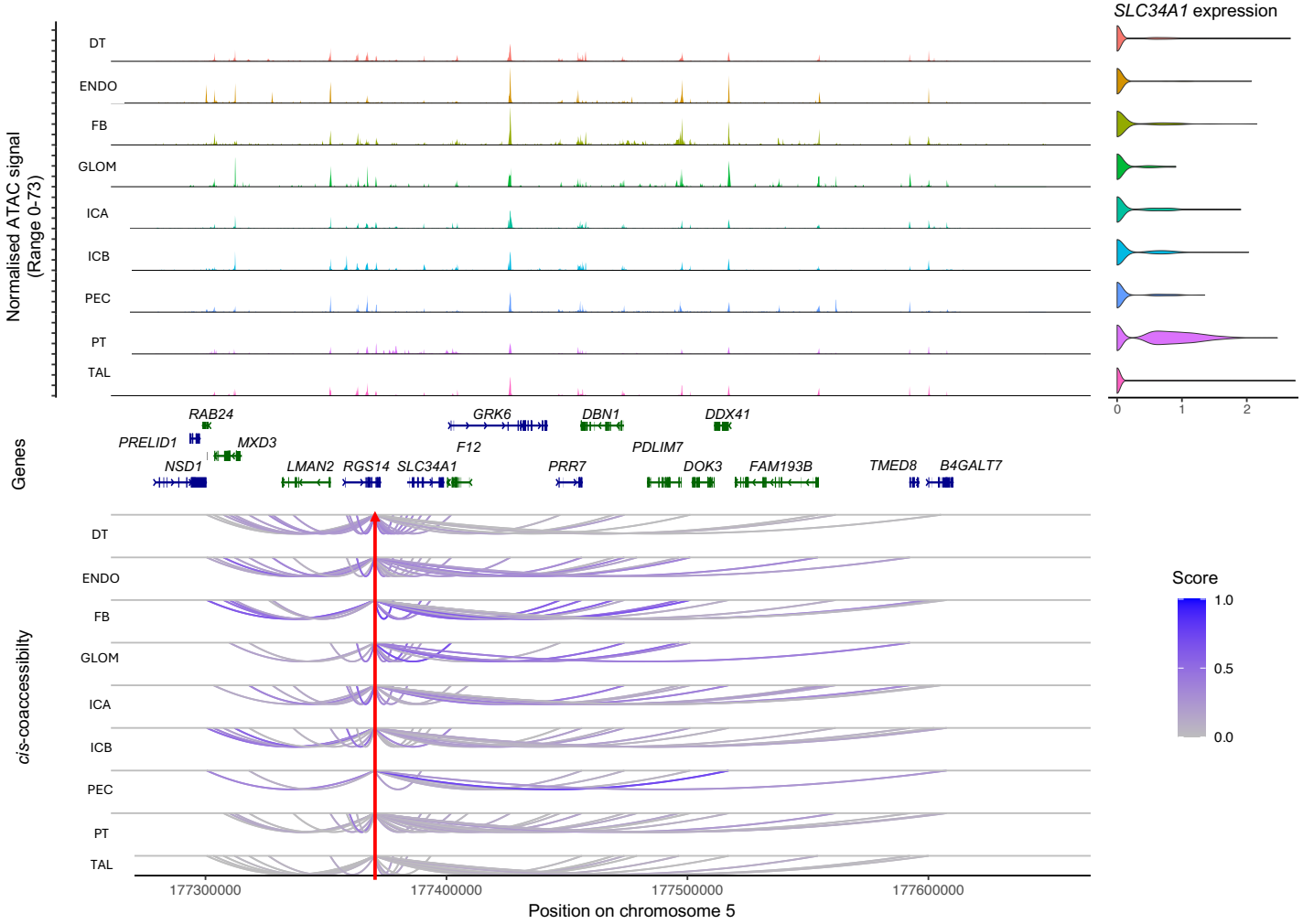



***Figure 5.8. Cis-coaccessibility of rs10051765 and protein-coding gene promoters correlates with RNA expression in renal cell lines.***
**Top**: Normalised ATAC signal across renal cell types. Peaks represent increased chromatin accessibility. **Bottom**: Links denote *cis*-coaccessible regions, contributing to the gene activity score. Link colours indicate the magnitude of the Cicero coaccessibility score. The vertical red arrow marks the location of rs10051765. **Violin plot**: *SLC34A1* gene expression in each renal cell type.
DT=distal tubule; ENDO=endothelial; FB=fibroblast; GLOM=glomerular; ICA=intercalated A; PEC=parietal epithelial; PT=proximal tubule; TAL=thick ascending limb of the loop of Henle. Coordinates are reported in genome build hg38.

#### 5.4.3.2 Regulatory potential of rs10051765 inferred through chromatin interactions with *SLC34A1* and *GRK6*

To further evaluate the regulatory potential of rs10051765 on *SLC34A1*, *GRK6*, *FAM193B*, *B4GALT7*, and *PRELID1* I analysed chromatin interactions using Capture-C data from three renal cell lines and identified interactions between the genomic region encompassing rs10051765 and the promoter regions of *SLC34A1* and *GRK6* across all three cell lines (Table 5.10, Figure 5.9). These findings confirm the results from chromatin coaccessibility analyses, highlighting a regulatory role for rs10051765 in modulating transcription of these genes. This may represent one mechanism by which this variant contributes to the pathogenesis of KSD and dysregulation of calcium and phosphate homeostasis. Interactions with the promoter regions of *FAM193B* were present only in HK2 and HEK293 cell lines, interactions with the promoter region of *PRELID1* were present in RPTEC and HEK293 cells, and interactions with the promoter region of *B4GALT7* only present in HEK293 cells.

***Table 5.10. Significant interactions of rs10051765 with gene promoters in renal cell lines identified in Capture-C analysis.***

| Cell line | Gene | Promoter | | | |
|---|---|---|---|---|---|
| | | **ENSR ID** | **Coordinates (hg38)** | **Distance from rs10051765** | **MPPC score** |
| RPTEC | *GRK6* | 00001384877 | 5:177402002-177403214 | 30.22 Kbp downstream | 0.02 |
| | *PRELID1* | 00001384853 | 5:177303866-177304698 | 68.29 Kbp upstream | 0.01 |
| | *SLC34A1* | 00001384871 | 5:177383802-177384461 | 11.47 Kbp downstream | 0.01 |
| HK2 | *FAM193B* | 00001384924 | 5:177539056-177540841 | 167.85 Kbp downstream | 0.03 |
| | | 00001384923 | 5:177538009-177539055 | 166.06 Kbp downstream | 0.01 |
| | *GRK6* | 00001384877 | 5:177402002-177403214 | 30.22 Kbp downstream | 0.02 |
| | | 00001384886 | 5:177425006-177426864 | 53.87 Kbp downstream | 0.01 |
| | | 5_B3B7PC | 5:177425273-177426853 | 53.86 Kbp downstream | 0.01 |
| | *SLC34A1* | 00001384871 | 5:177383802-177384461 | 11.47 Kbp downstream | 0.08 |
| HEK293 | *B4GALT7* | 5_94LH9T | 5:177598981-177600142 | 227.15 Kbp downstream | 0.04 |
| | | 00001384934 | 5:177598642-177600240 | 227.25 Kbp downstream | 0.04 |
| | | 00001384934 | 5:177598642-177600240 | 227.25 Kbp downstream | 0.03 |
| | *FAM193B* | 00001384927 | 5:177553256-177554336 | 181.35 Kbp downstream | 0.05 |
| | | 00001384924 | 5:177539056-177540841 | 167.85 Kbp downstream | 0.04 |
| | | 00001384923 | 5:177538009-177539055 | 166.06 Kbp downstream | 0.04 |
| | | 00001384921 | 5:177536380-177537880 | 164.89 Kbp downstream | 0.03 |
| | *GRK6* | 00001384888 | 5:177430221-177431721 | 58.73 Kbp downstream | 0.05 |
| | | 00001384889 | 5:177431722-177432243 | 59.25 Kbp downstream | 0.05 |
| | | 00001384877 | 5:177402002-177403214 | 30.22 Kbp downstream | 0.04 |
| | | 5_B3B7PC | 5:177425273-177426853 | 53.86 Kbp downstream | 0.04 |
| | | 00001384886 | 5:177425006-177426864 | 53.87 Kbp downstream | 0.04 |
| | | 00001384886 | 5:177425006-177426864 | 53.87 Kbp downstream | 0.04 |
| | | 00001384886 | 5:177425006-177426864 | 53.87 Kbp downstream | 0.03 |
| | | 5_B3B7PC | 5:177425273-177426853 | 53.86 Kbp downstream | 0.03 |
| | | 5_B3B7PC | 5:177425273-177426853 | 53.86 Kbp downstream | 0.03 |
| | *PRELID1* | 00001384853 | 5:177303866-177304698 | 68.29 Kbp upstream | 0.04 |
| | | 00001384854 | 5:177304699-177305626 | 67.37 Kbp upstream | 0.04 |
| | *SLC34A1* | 00001384873 | 5:177388157-177389657 | 16.67 Kbp downstream | 0.05 |
| | | 00001384871 | 5:177383802-177384461 | 11.47 Kbp downstream | 0.04 |

For each cell type, results are arranged in descending order of MPPC score.
ENSR ID=Ensembl stable ID for regulatory features; HEK293=Human embryonic kidney 293 cells; HK2=Human kidney 2 immortalised proximal tubule epithelial cells; MPPC=Marginal posterior probability of contact; RPTEC=Renal proximal tubule epithelial cells.

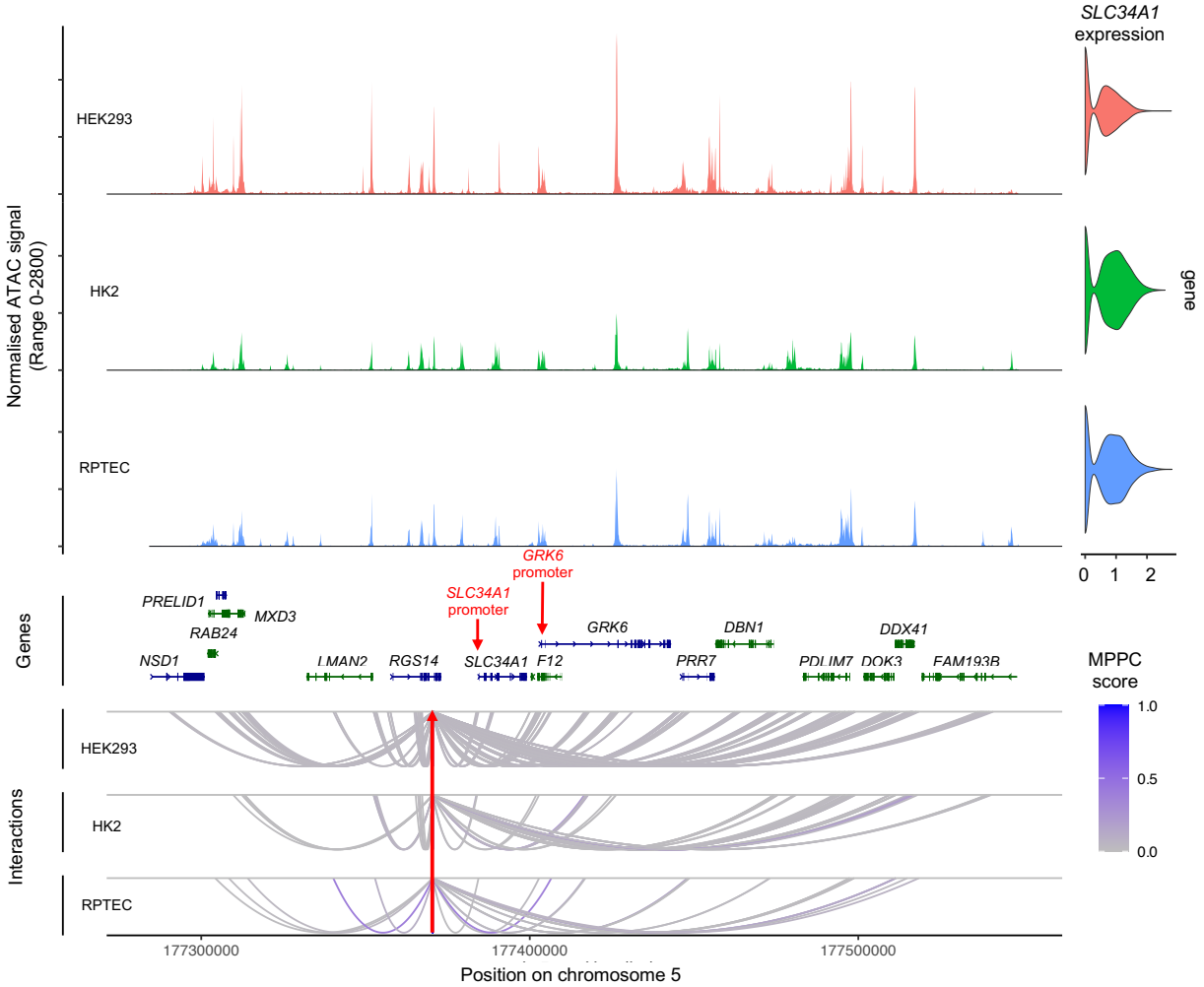



***Figure 5.9. Cature-C: chromatin interactions of rs10051765 and gene promoters across three renal cell lines.***

**Top**: Normalised ATAC signal across renal cell types. Peaks represent increased chromatin accessibility. **Bottom**: Links represent significant chromatin interactions between rs10051765 and promoter coordinates of protein-coding genes, defined using Ensembl v113 [466]. Colours of links reflect the magnitude of the marginal posterior probability of contact. The vertical red arrow denotes the location of rs10051765. **Violin plot**: Expression of *SLC34A1* in each renal cell type.
HEK293=human embryonic kidney 293 cells; HK2=human kidney 2 immortalised proximal tubule epithelial cells; MPPC=marginal posterior probability of contact; RPTEC=renal proximal tubule epithelial cells. Coordinates are reported in genome build hg38.

#### 5.4.3.3 rs10051765 regulates phosphate homeostasis and contributes to the pathogenesis of kidney stone disease

Thus far, analyses identify that chromatin accessibility at rs10051765 correlates with lower expression of *GRK6* and higher expression of *SLC34A1*, and that there are physical interactions of this locus with promoter regions of these three genes in renal cell types.

*GRK6*, whose promoter lies ≈30 Kbp downstream of rs10051765, encodes G protein-coupled receptor kinase 6 (GRK6), a member of the Ser/Thr protein kinase family that is widely expressed across tissues. GRK6 phosphorylates activated G-protein coupled receptors, initiating their downregulation by homologous receptor desensitisation [608]. In HEK293 cells, the GRK6A splice variant mediates constitutive phosphorylation of the $Na^+/H^+$ exchanger regulatory factor (NHERF1, encoded by *SLC9A3R1*) via a PDZ domain-mediated interaction [609]. NHERF1 is a cytoplasmic protein required to recruit transporters and signalling proteins to the plasma membrane. In renal proximal tubular cells, NHERF1 stabilises key proteins, including NaPi-IIa, PTH receptor 1, and phospholipase C at the brush border apical membrane, facilitating PTH signalling to NaPi-IIa [610].

Patients with heterozygous missense variants in NHERF1 exhibit decreased NaPi-IIa expression, KSD, hypophosphataemia, reduced tubular maximum reabsorption of phosphate to glomerular filtration rate (TmP/GFR), and normal serum PTH and FGF23 concentrations [611]. Reduced serum phosphate levels stimulate 1,25-dihydroxyvitamin D synthesis, promoting absorptive hypercalciuria. This, combined with elevated urinary phosphate excretion, can trigger calcium-phosphate crystal formation and increase the risk of KSD (Figure 5.10). Mice with targeted *SLC9A3R1* knockdown have aberrant Npt2a localisation and increased urinary phosphate excretion [612].

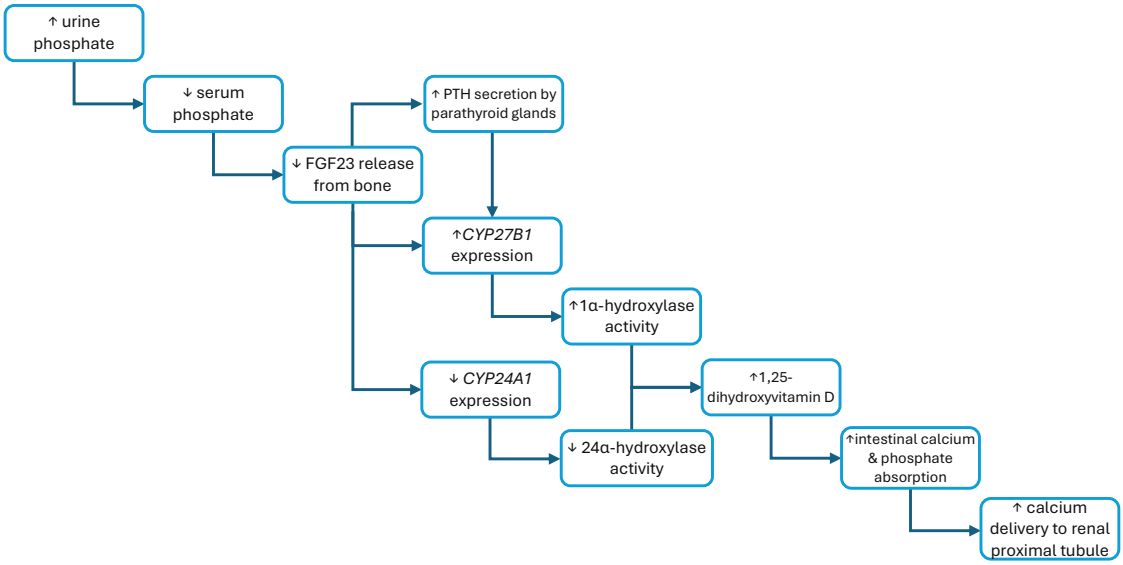



***Figure 5.10. Implications of increased NaPi-IIa-mediated phosphate leak.***

Reductions in NaPi-IIa-mediated renal phosphate resorption cause reductions in serum phosphate levels. This, in turn, reduces FGF23 release from bone. Lower FGF23 results in A) upregulated *CYP27B1* expression, increasing 1-$\alpha$-hydroxylase transcription promoting 1,25-dihydroxyvitamin D synthesis; and B) downregulated *CYP24A1* expression, decreasing 24-$\alpha$-hydroxylase activity and reducing the breakdown of 1,25-dihydroxyvitamin D to inactive 24,25-dihydroxyvitamin D. Higher 1,25-dihydroxyvitamin D enhances intestinal calcium absorption, contributing to increased serum calcium levels. This, in turn, results in increased calcium levels in the proximal renal tubule. FGF23=Fibroblast growth factor 23.

Furthermore, NHERF1 mutations cause autosomal dominant hypophosphatemia with nephrolithiasis or osteoporosis (NPHLOP2), a rare, genetic disease characterised by phosphate wasting in the proximal tubule. This condition results in hypercalciuria, recurrent urolithiasis and/or osteoporosis [611, 613]. Decreased NHERF1 expression is associated with ADPKD, with which ≈20% of patients experience KSD, as noted in my introductory chapter [128, 614]. Given this evidence, it is plausible that rs10051765 genotype modulates *GRK6* expression, thereby influencing NaPi-IIa localisation at the apical plasma membrane of proximal tubular cells. Such alterations could disrupt renal tubule-mediated phosphate transport, affect serum calcium levels, and increase the risk of KSD.

*SLC34A1*, encoding NaPi-IIa as previously mentioned, has its promoter ≈6 Kbp downstream from rs10051765, making it the gene closest to this variant. Impaired NaPi-IIa function increases renal phosphate excretion, reduces FGF23 levels, and activates 1-$\alpha$-hydroxylase in the kidney (Figure 5.10). This activation enhances the

conversion of inactive 25-hydroxyvitamin D to its active form, 1,25-dihydroxyvitamin D. Simultaneously, it inhibits 24-$\alpha$-hydroxylase, which prevents the breakdown of 1,25-dihydroxyvitamin D into its inactive form, 24,25-dihydroxyvitamin D [53].

Slc34a1$^{-/-}$ mice display hypophosphataemia, hypercalcaemia, and hypercalciuria, alongside reduced serum PTH levels [183]. In humans, homozygous mutations in *SLC34A1* cause autosomal recessive infantile hypercalcaemia type 2 (Table 1.2) [183, 615] and heterozygous mutations in *SLC34A1* lead to autosomal dominant hypophosphataemia with nephrolithiasis or osteoporosis (NPHLOP1), which differs from NPHLOP2 caused by mutations in *SLC9A3R1* [181]. Additionally, population studies link heterozygous *SLC34A1* variants to hypophosphataemic hypercalciuric stone formation [160, 179, 182].

The KSD-associated C allele of rs10051765 correlates with higher serum calcium and lower serum phosphate concentrations (Table 5.6). This phenotype resembles that observed in *SLC34A1*-related infantile hypercalcaemia type 2 and *SLC34A1*- or *SLC9A3R1*-related autosomal dominant hypophosphataemia with nephrolithiasis or osteoporosis, both characterised by increased urinary phosphate excretion. Chromatin interaction analyses reveal that rs10051765 interacts with the promoters of *GRK6* and *SLC34A1*, implicating these genes in phosphate homeostasis.

These findings suggest that rs10051765 drives NaPi-IIa-mediated phosphate wasting, altering serum calcium and phosphate levels in an attenuated manner akin to monogenic disorders of phosphate-wasting. This disruption of phosphate homeostasis plausibly underpins the variant's association with increased KSD risk, underscoring its causal role in KSD pathogenesis and its broader significance in calcium and phosphate metabolism.

### 5.4.4 The functional relevance of rs6127099

#### 5.4.4.1 The regulatory role of rs6127099 in gene transcription investigated through chromatin accessibility and RNA expression in renal cell types

The positive correlation of RNA expression with chromatin coaccessibility between rs6127099±100 bp and promoter regions of *CYP24A1* and *DOK5* highlighted rs6127099 as a potential regulator of their transcription (Table 5.11, Figure 5.12, Figure 5.11). These correlations are driven by gene activity scores in proximal tubule and thick ascending limb of the loop of Henle cells, respectively.

***Table 5.11. Significant Pearson correlation coefficients between gene activity and expression on chromosome 20 across renal cell types.***

| Gene | Maximum activity | Maximum expression | Pearson $r^2$ | Pearson P | Cell type | Mean activity | Mean expression |
|---|---|---|---|---|---|---|---|
| *CYP24A1* | $4.27\times10^{-3}$ | 0.31 | 0.96 | $4.00\times10^{-5}$ | PT | $3.88\times10^{-3}$ | 0.28 |
| | | | | | ICA | $1.73\times10^{-8}$ | 0.08 |
| | | | | | GLOM | $3.03\times10^{-9}$ | 0.06 |
| | | | | | DT | $3.16\times10^{-9}$ | 0.05 |
| | | | | | PEC | $2.41\times10^{-10}$ | 0.05 |
| | | | | | TAL | $1.85\times10^{-9}$ | 0.04 |
| | | | | | FB | $1.00\times10^{-10}$ | 0.02 |
| | | | | | ICB | $4.88\times10^{-11}$ | 0.02 |
| | | | | | ENDO | $4.13\times10^{-11}$ | 0.01 |
| *DOK5* | $4.57\times10^{-3}$ | 1.33 | 0.8 | 0.01 | TAL | $4.16\times10^{-3}$ | 1.21 |
| | | | | | DT | $4.82\times10^{-4}$ | 0.86 |
| | | | | | GLOM | $8.21\times10^{-8}$ | 0.62 |
| | | | | | ICA | $2.98\times10^{-10}$ | 0.57 |
| | | | | | PEC | $1.00\times10^{-3}$ | 0.52 |
| | | | | | PT | $1.19\times10^{-8}$ | 0.34 |
| | | | | | FB | $3.63\times10^{-10}$ | 0.29 |
| | | | | | ENDO | $1.05\times10^{-8}$ | 0.19 |
| | | | | | ICB | $4.67\times10^{-11}$ | 0.17 |

Genes arranged in descending order of significance. Cell types arranged in descending order of gene expression. DT=Distal tubule; ENDO=Endothelial; FB=Fibroblast; GLOM=Glomerular; ICA=Intercalated A; P=P-value; PEC=Parietal epithelial; PT=Proximal tubule; TAL=Thick ascending limb of the loop of Henle.

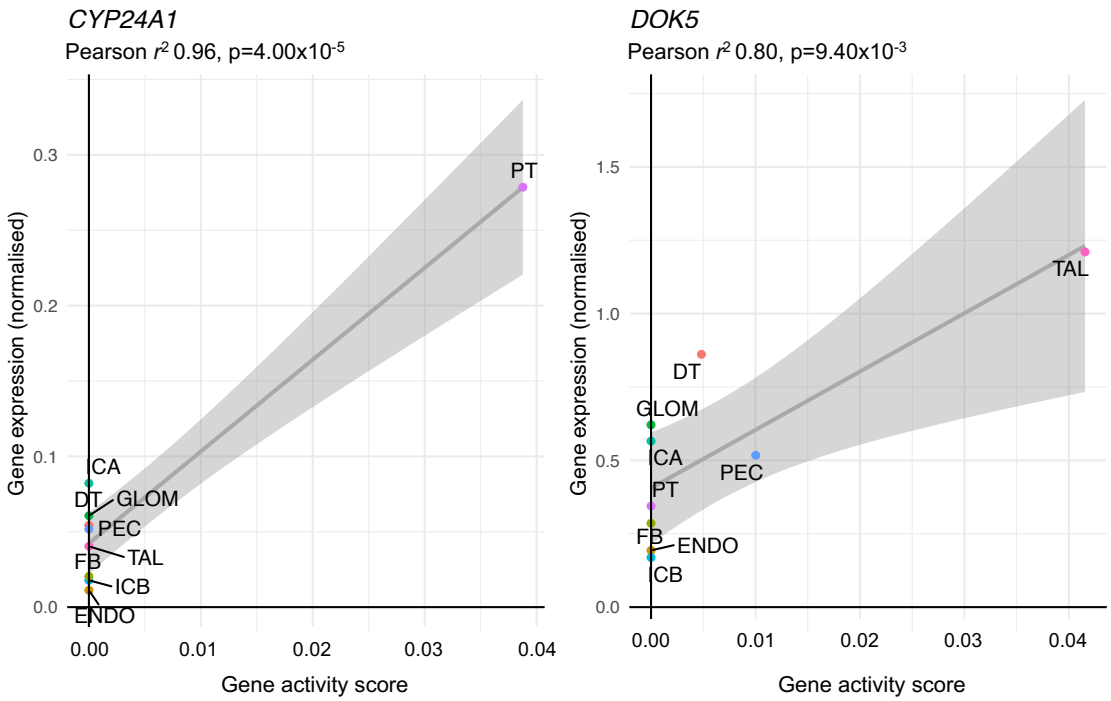



***Figure 5.11. Pearson correlation coefficients of gene activity, inferred from rs6127099 coaccessibility, and RNA expression of CYP24A1 and DOK5.***

Gene activity score reflects the weighted coaccessibility between rs6127099 and gene promoter regions. Gene expression reflects normalised gene expression from snRNA-seq data.

DT=Distal tubule; ENDO=Endothelial; FB=Fibroblast; GLOM=Glomerular; ICA=Intercalated A; P=P-value; PEC=Parietal epithelial; PT=Proximal tubule; TAL=Thick ascending limb of the loop of Henle.

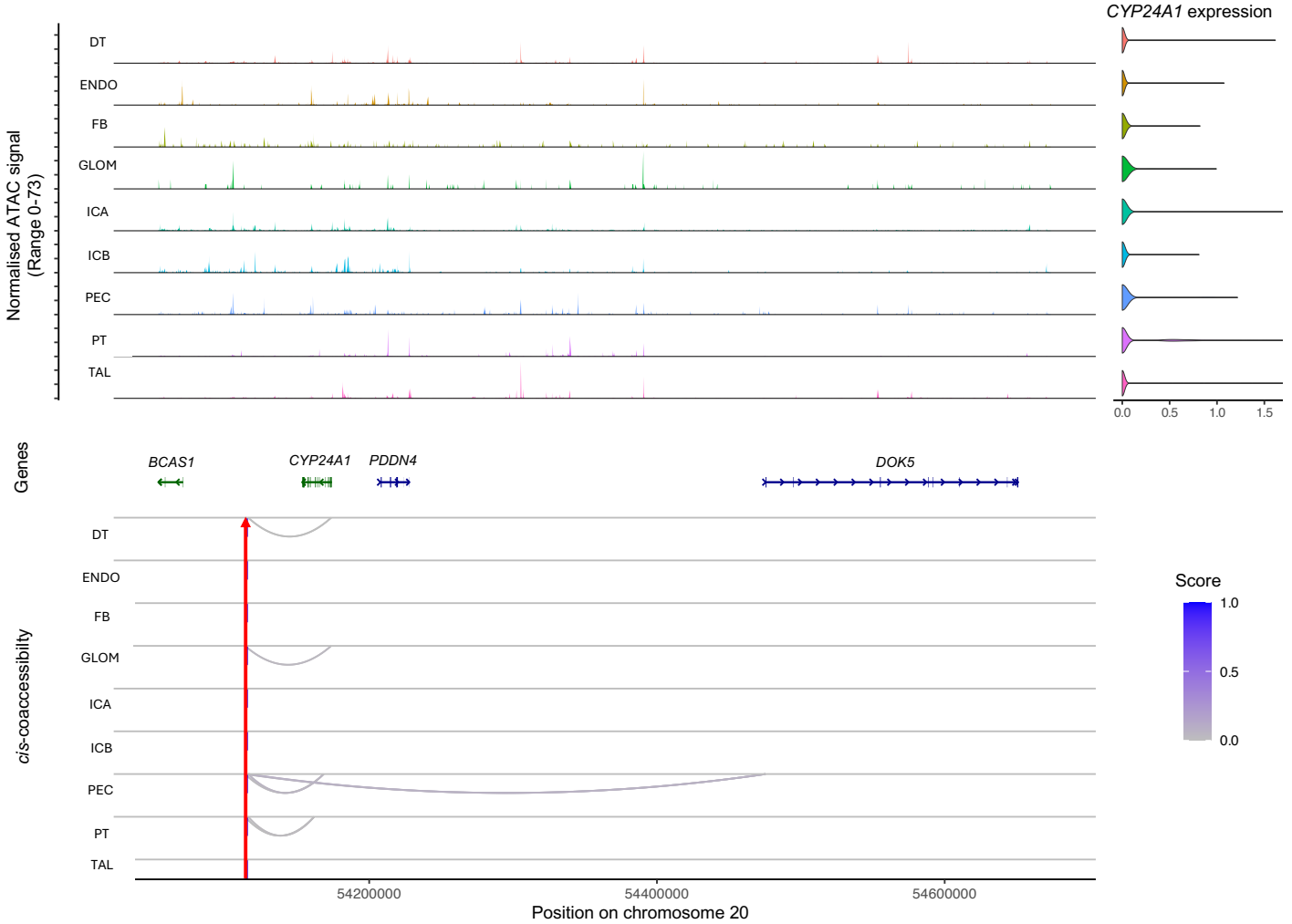



***Figure 5.12. Cis-coaccessibility of rs6127099 and protein-coding gene promoters correlates with RNA expression in renal cell lines.***
**Top**: Normalised ATAC signal across renal cell types. Peaks represent increased chromatin accessibility. **Bottom**: Links denote *cis*-coaccessible regions, contributing to the gene activity score. Link colours indicate the magnitude of the Cicero coaccessibility score. The vertical red arrow marks the location of rs6127099. **Violin plot**: *CYP24A1* gene expression in each renal cell type. DT=distal tubule; ENDO=endothelial; FB=fibroblast; GLOM=glomerular; ICA=intercalated A; PEC=parietal epithelial; PT=proximal tubule; TAL=thick ascending limb of the loop of Henle. Coordinates are reported in genome build hg38.

#### 5.4.4.2 Regulatory potential of rs6127099 inferred through chromatin interactions with *CYP24A1*

Analysing Capture-C data from RPTEC, HK2, and HEK293 cells revealed interactions between the genomic region corresponding to rs6127099 and the promoter region of *CYP24A1* (Table 5.12, Figure 5.13). These results, in tandem with findings from chromatin accessibility studies, implicate rs6127099 in *CYP24A1* regulation and which may be one mechanism contributing to higher risk of KSD and higher serum calcium and phosphate concentrations. No interactions were observed with the promoter region of *DOK5*.

***Table 5.12. Significant interactions of rs6127099 with gene promoters in renal cell lines identified in Capture-C analysis.***

| Cell line | Gene | Promoter | | | MPPC score |
|---|---|---|---|---|---|
| | | ENSR ID | Coordinates (hg38) | Distance from rs6127099 | |
| RPTEC | *CYP24A1* | 20_BDQ9N | 20:54173639-54175225 | 60.362 Kbp downstream | 0.13 |
| | | 00001364241 | 20:54173406-54175399 | 60.536 Kbp downstream | 0.13 |
| HK2 | *CYP24A1* | 20_BDQ9N | 20:54173639-54175225 | 60.362 Kbp downstream | 0.47 |
| | | 00001364241 | 20:54173406-54175399 | 60.536 Kbp downstream | 0.47 |

For each cell type, results are arranged in descending order of MPPC score.
ENSR ID=Ensembl stable ID for regulatory features; HEK293=Human embryonic kidney 293 cells; HK2=Human kidney 2 immortalised proximal tubule epithelial cells; MPPC=Marginal posterior probability of contact; RPTEC=Renal proximal tubule epithelial cells.
No significant interactions with promoter regions of *CYP24A1* or *DOK5* were identified in HEK293 cells.

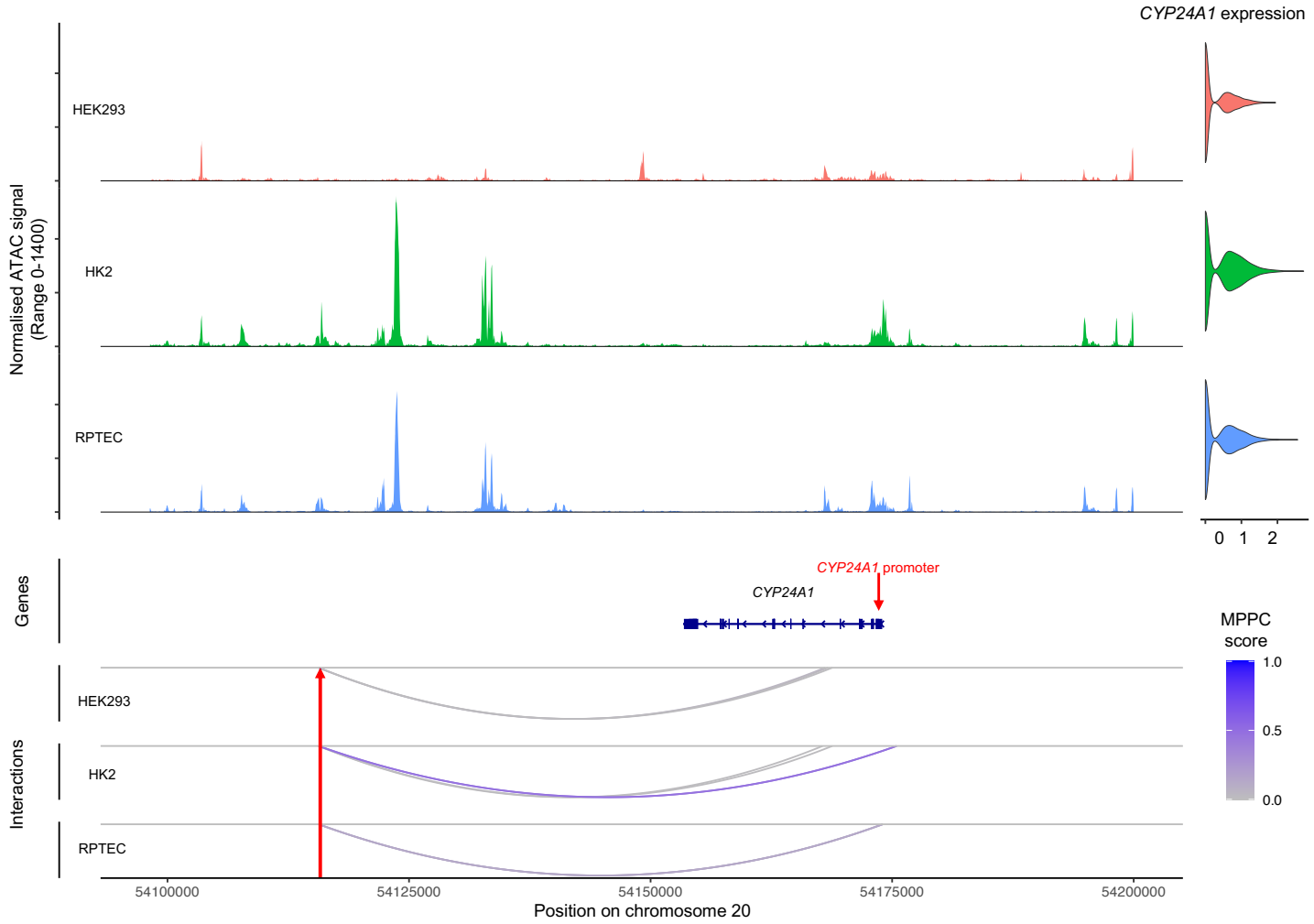



***Figure 5.13. Cature-C: chromatin interactions of rs6127099 and gene promoters across three renal cell lines.***

**Top**: Normalised ATAC signal across renal cell types. Peaks represent increased chromatin accessibility. **Bottom**: Links represent significant chromatin interactions between rs6127099 and promoter coordinates of protein-coding genes, defined using Ensembl v113 [466]. Colours of links reflect the magnitude of the marginal posterior probability of contact. The vertical red arrow denotes the location of rs6127099. **Violin plot**: Expression of *CYP24A1* in each renal cell type.
HEK293=Human embryonic kidney 293 cells; HK2=Human kidney 2 immortalised proximal tubule epithelial cells; MPPC=Marginal posterior probability of contact; RPTEC=Renal proximal tubule epithelial cells. Coordinates are reported in genome build hg38.

#### 5.4.4.3 rs6127099 regulates *CYP24A1* expression and increases kidney stone disease risk through impaired 1,25-dihydroxyvitamin D inactivation

*CYP24A1*, encoding 24-$\alpha$-hydroxylase, plays a crucial role in kidney-mediated inactivation of 1,25-dihydroxyvitamin D. 24-$\alpha$-hydroxylase dysfunction impairs 1,25-dihydroxyvitamin D breakdown, leading to elevated levels of the active hormone. This increase enhances intestinal and renal calcium absorption, leading to hypercalcaemia and heightened sensitivity to or toxicity from vitamin D [616]. Additionally, 1,25-dihydroxyvitamin D drives immune responses by inducing promyelocyte differentiation, activating monocytes/macrophages, and maturing dendritic cells into potent antigen-presenting cells [617–619]. These immune-modulating effects may explain the reported association of the rs6127099-T allele with a reduced risk of SARS-CoV-2 infection.

Homozygous or compound heterozygous mutations in *CYP24A1* cause infantile hypercalcaemia type 1, an autosomal recessive disorder. Affected individuals exhibit hypercalcaemia symptoms, such as weight loss, failure to thrive, dehydration, muscular hypotonia, nephrocalcinosis, and KSD [184, 185, 615, 620].

Common variants associated with *CYP24A1* influence serum calcium, PTH, 25-hydroxyvitamin D, creatinine, cystatin C, FGF23, urea, and urate levels. These variants also associate with risks of multiple sclerosis, KSD, and atopic dermatitis across diverse populations [413, 598, 621–628]. My analyses reveal that the rs6127099-A allele increases the risk of KSD and elevates serum calcium levels without a concomitant reduction in phosphate concentrations. This phenotype resembles a milder form of infantile hypercalcaemia type 1, indicating that the rs6127099-A allele likely disrupts 1,25-dihydroxyvitamin D inactivation.

Together, results of chromatin coaccessibility and Capture-C analyses, along with the KSD-related phenotype, collectively demonstrate that rs6127099 regulates *CYP24A1* expression. These findings reinforce its role as a key genetic driver of vitamin D metabolism, calcium-phosphate homeostasis, and KSD pathogenesis.

### 5.4.5 The clinical relevance of rs838717, rs10051765, and rs6127099 in treating kidney stone disease

These studies demonstrate the scientific relevance of rs838717, rs10051765, and rs6127099 in reduced CaSR signal transduction, renal phosphate wasting, and reduced vitamin D inactivation, respectively. To ascertain their clinical relevance, I calculated the population attributable fraction in the UK Biobank. These variants contribute substantially to the overall burden of KSD in populations of European ancestry; the *DGKD*-, *SLC34A1*-, and *CYP24A1*-associated variants account for ≈7%, ≈7%, and ≈14% of KSD cases, respectively, and ≈12-17% of KSD cases collectively (Table 5.13).

***Table 5.13. Kidney stone population attributable risk and population attributable fraction for candidate causal variants in the UK Biobank.***

| Analysis | | Risk allele carrier | No risk allele | KSD prevalence | KSD prevalence carriers | KSD prevalence non-carriers | PAR | PAF (%) |
|---|---|---|---|---|---|---|---|---|
| ***DGKD*- rs838717** | | GG or GA | AA | | | | | |
| Dominant | Cases | 8,211 | 3,475 | 0.027 | 0.030 | 0.029 | 0.002 | 6.97 |
| | Controls | 269,313 | 126,920 | | | | | |
| | | GG | GA or AA | | | | | |
| Recessive | Cases | 2,421 | 9,265 | 0.028 | 0.031 | 0.029 | 0.001 | 2.26 |
| | Controls | 74,626 | 321,607 | | | | | |
| ***SLC45A1*- rs10051765** | | CC or CT | TT | | | | | |
| Dominant | Cases | 6,709 | 4,777 | 0.026 | 0.030 | 0.028 | 0.002 | 7.06 |
| | Controls | 218,557 | 177,676 | | | | | |
| | | CC | CT or TT | | | | | |
| Recessive | Cases | 1,586 | 10,100 | 0.028 | 0.035 | 0.029 | 0.001 | 2.83 |
| | Controls | 43,494 | 352,739 | | | | | |
| ***CYP24A1*- rs6127099** | | AA or AT | TT | | | | | |
| Dominant | Cases | 10,913 | 773 | 0.025 | 0.029 | 0.029 | 0.004 | 14.02 |
| | Controls | 365,624 | 30,609 | | | | | |
| | | AA | AT or TT | | | | | |
| Recessive | Cases | 65,24 | 5,162 | 0.026 | 0.031 | 0.029 | 0.002 | 7.68 |
| | Controls | 206,222 | 190,011 | | | | | |
| **All** | | ≥1 risk allele | No risk alleles | | | | | |
| Dominant | Cases | 11,579 | 107 | 0.024 | 0.029 | 0.029 | 0.005 | 17.28 |
| | Controls | 391,825 | 4,408 | | | | | |
| | | Risk allele homozygosity at ≥1 variant | No homozygosity | | | | | |
| Recessive | Cases | 8,142 | 3,544 | 0.025 | 0.030 | 0.029 | 0.003 | 12.07 |
| | Controls | 259,089 | 137,144 | | | | | |

KSD=Kidney stone disease ; PAF=Population attributable fraction (calculated as (prevalence in total population- prevalence in individuals without risk allele)/ prevalence in total population); PAR=Population attributable risk (calculated as prevalence in total population- prevalence in individuals without risk allele).

Given the importance of these variants to the prevalence and pathophysiology of KSD, I sought to identify potential therapeutic pathways that could be modulated to prevent KSD using drug-target MR analyses. Modulating *CASR* and *CYP24A1* to reduce serum calcium concentrations by 1 SD (0.08 mmol/L) may decrease odds of KSD by ≈30% and ≈90%, respectively (Table 5.14, Figure 5.14); directionally concordant but statistically insignificant effects were detected for *CASR*-mediated effects using FinnGen outcome data (Table 5.14, Figure 5.14). Similar analyses of *DGKD* or *SLC34A1* modulation were not possible as there were insufficient genetic proxies. However, MR analyses relaxing the threshold of genetic independence of instrumental variables to $r^2$ <0.1, indicated that reducing serum calcium concentrations by 1 SD via *DGKD* may decrease risk of kidney stone disease by ≈70%, and increasing serum phosphate concentrations by 1 SD (0.016 mmol/L) via *SLC34A1* may decrease risk of KSD by over 90% (Table 5.14, Figure 5.14). Phenome wide association study data suggested that modulating *DGKD*, *CASR*, *CYP24A1*, or *SLC34A1* may result in target-mediated adverse effects including alterations in serum bilirubin, inflammatory bowel disease, migraine, atopic dermatitis, and eczematous phenotypes [472].

***Table 5.14. Drug target Mendelian randomisation; effects of modulating mineral metabolism traits on odds of kidney stone disease.***

| Decreasing serum calcium concentrations | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Analysis** | | | | **IVW MR** | | | **Heterogeneity** | | | **Intercept** | |
| **Gene** | $r^2$ | $R^2$ (SD) | **Outcome** | **N SNPs** | **OR (95%CI)** | **P** | **Q** | **Q_df** | **P** | $\beta$ (SE) | **P** |
| *DGKD* | 0.1 | $3.27\times10^{-4}$ ($3.77\times10^{-4}$) | Meta-analysis | 10 | 0.22 (0.17-0.30) | $2.74\times10^{-24}$ | 14.00 | 9 | 0.12 | $4.71\times10^{-3}$ (0.01) | 0.62 |
| | | | UK Biobank | 9 | 0.26 (0.17-0.38) | $4.50\times10^{-11}$ | 10.76 | 8 | 0.22 | 0.03 (0.02) | 0.24 |
| | | | FinnGen | 11 | 0.20 (0.14-0.30) | $1.62\times10^{-16}$ | 15.29 | 10 | 0.12 | -0.04 (0.03) | 0.25 |
| *CASR* | 0.01 | $9.51\times10^{-4}$ ($1.35\times10^{-3}$) | Meta-analysis | 14 | 0.72 (0.53-0.97) | 0.03 | 78.52 | 13 | $2.09\times10^{-11}$ | $8.17\times10^{-4}$ (0.01) | 0.94 |
| | | | UK Biobank | 13 | 0.67 (0.48-0.93) | 0.02 | 41.42 | 12 | $4.17\times10^{-5}$ | 0.03 (0.03) | 0.45 |
| | | | FinnGen | 11 | 0.76 (0.54-1.08) | 0.13 | 42.58 | 10 | $5.91\times10^{-6}$ | -0.04 (0.04) | 0.28 |
| | 0.1 | $7.89\times10^{-4}$ ($1.01\times10^{-3}$) | Meta-analysis | 43 | 0.74 (0.64-0.85) | $2.67\times10^{-5}$ | 142.07 | 42 | $8.59\times10^{-13}$ | 0.04 (0.02) | 0.06 |
| | | | UK Biobank | 44 | 0.70 (0.59-0.82) | $2.14\times10^{-5}$ | 111.10 | 43 | $6.17\times10^{-8}$ | 0.02 (0.01) | 0.15 |
| | | | FinnGen | 37 | 0.78 (0.67-0.91) | $2.41\times10^{-3}$ | 88.96 | 36 | $2.24\times10^{-6}$ | $-3.29\times10^{-4}$ (0.01) | 0.98 |
| *CYP24A1* | 0.01 | $5.37\times10^{-4}$ ($4.98\times10^{-4}$) | Meta-analysis | 4 | 0.09 (0.07-0.12) | $3.58\times10^{-64}$ | 2.86 | 3 | 0.41 | $-3.74\times10^{-3}$ (0.01) | 0.7 |
| | | | UK Biobank | 5 | 0.12 (0.08-0.19) | $3.46\times10^{-18}$ | 6.29 | 4 | 0.18 | 0.03 (0.03) | 0.37 |
| | | | FinnGen | 5 | 0.07 (0.05-0.11) | $8.75\times10^{-36}$ | 6.15 | 4 | 0.19 | -0.04 (0.04) | 0.33 |
| | 0.1 | $3.29\times10^{-4}$ ($3.08\times10^{-4}$) | Meta-analysis | 18 | 0.10 (0.08-0.12) | $3.43\times10^{-124}$ | 20.37 | 17 | 0.26 | 0.02 (0.02) | 0.31 |
| | | | UK Biobank | 17 | 0.12 (0.09-0.15) | $1.95\times10^{-61}$ | 15.74 | 16 | 0.47 | $-5.80\times10^{-4}$ (0.01) | 0.97 |
| | | | FinnGen | 18 | 0.08 (0.06-0.10) | $4.16\times10^{-87}$ | 20.64 | 17 | 0.24 | 0.04 (0.01) | $4.62\times10^{-3}$ |
| **Increasing serum phosphate concentrations** | | | | | | | | | | | |
| *SLC34A1* | 0.1 | $3.24\times10^{-4}$ ($3.95\times10^{-4}$) | Meta-analysis | 8 | 0.08 (0.06-0.11) | $1.09\times10^{-74}$ | 2.54 | 7 | 0.92 | 0.03 (0.01) | 0.04 |
| | | | UK Biobank | 8 | 0.07 (0.04-0.10) | $6.57\times10^{-42}$ | 4.38 | 7 | 0.74 | 0.01 (0.01) | 0.34 |
| | | | FinnGen | 7 | 0.10 (0.07-0.15) | $5.71\times10^{-35}$ | 1.44 | 6 | 0.96 | 0.02 (0.01) | 0.04 |

CHR=Chromosome; IVW MR=Inverse-variance weighted Mendelian randomisation; P=5% false discovery rate adjusted p-value; N SNPs=Number of SNPs; Q=Q statistic; Q_df=Q degrees of freedom; $R^2$=variance in exposure explained by instrumental variable; SD=Standard deviation; SE=Standard error 95% CI= 95% confidence interval.

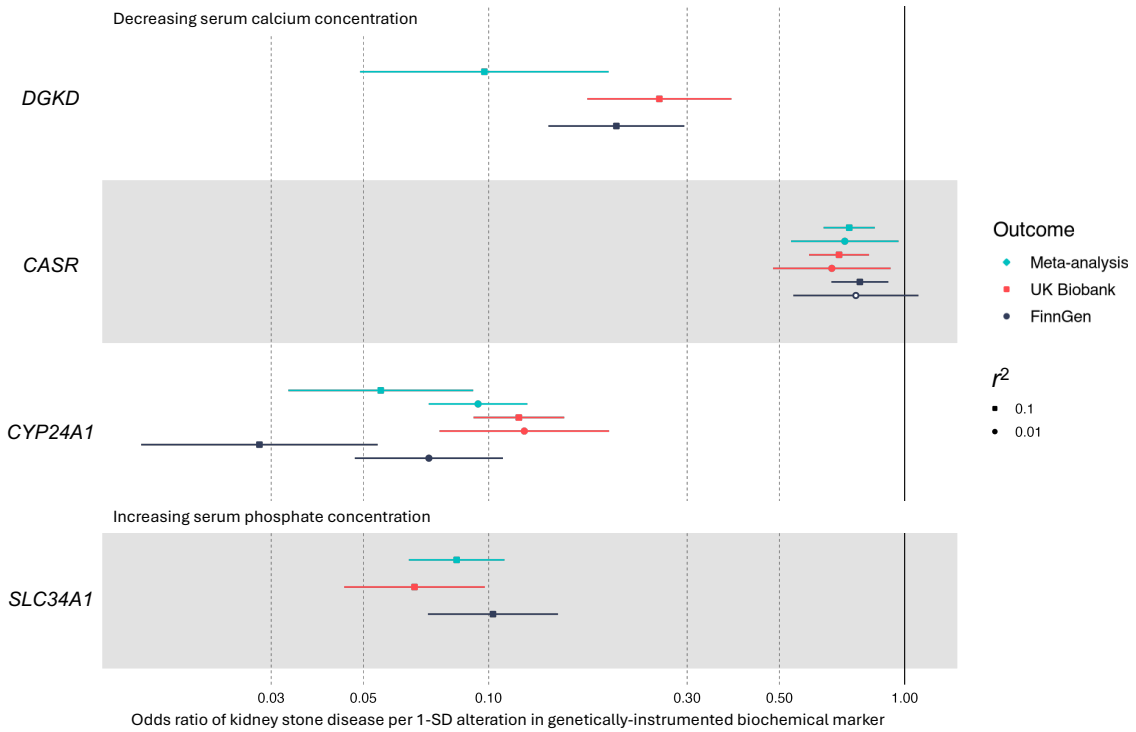



***Figure 5.14. Forest plot demonstrating drug-target Mendelian randomisation estimates for the effects of modulating serum calcium concentrations via DGKD, CASR, and CYP24A1, and serum phosphate concentrations via SLC34A1.***

Nonsignificant results are indicated by hollow symbols.

KSD=Kidney stone disease; SD=Standard deviation. Gene position defined by Ensembl ±300 kilbase pairs. Odds ratio is scaled per 1 standard deviation decrease in serum calcium (0.08 mmol/L) or increase in serum phosphate (0.016 mmol/L) concentrations.

## 5.5 Discussion and conclusions

In this chapter of my thesis, I aimed to investigate the causal effects of alterations in mineral homeostasis on the risk of KSD and to identify potential disease-mediating genetic loci to guide future therapeutic approaches by integrating evidence from genetic epidemiology, multiomic data, and chromatin interaction analyses. Together, these analyses have identified biological pathways commonly involved in KSD and uncovered novel therapeutic targets for its treatment. By moving beyond GWAS associations to pinpoint causal effects attributable to specific biological pathways, I have advanced our understanding of the pathogenesis of KSD. Using a systematic, region-specific *cis*-MR approach and colocalisation analyses, I have identified three genetic variants that likely drive KSD through disruptions in calcium and phosphate

homeostasis, collectively accounting for ≈12–17% of KSD cases. I also investigated their downstream effects using snRNA-seq, snATAC-seq, and Capture-C data to examine gene expression profiles and chromatin interactions. These findings suggest that the variants increase KSD risk via three distinct mechanisms: DGK$\delta$-mediated impaired CaSR signal transduction, NaPi-IIa-mediated increased renal phosphate excretion, and 24-$\alpha$-hydroxylase-mediated impaired inactivation of 1,25-dihydroxyvitamin D. Multiomic analyses of snRNA-seq data show that proximal tubular RNA expression of *DGKD*, *SLC34A1*, *GRK6*, and *CYP24A1*, the genes I have implicated in these three pathways, is relatively high, corroborating the importance of this cell type in the pathophysiology of KSD as identified in gene and gene set based analyses (Section 3.3.5, Table 3.9 and Table 3.10). Finally, drug-target MR analyses demonstrate the therapeutic potential of modulating these pathways to reduce KSD risk, offering actionable insights for future prevention and treatment strategies.

These findings have direct implications for clinical practice. Identifying patients harbouring these variants may enable clinicians to identify patients at risk of recurrent KSD, due to underlying genetic variants, and to implement lifestyle modifications, such as optimising fluid intake, to mitigate the risk associated with relative hypercalcaemia. Furthermore, targeted therapies could be selected based on the disrupted biological pathway in an individual patient. For instance, calcimimetics may benefit patients with DGK$\delta$-mediated CaSR-signalling perturbations, phosphate supplements could address NaPi-IIa-associated KSD, and triazole drugs or rifampicin might be employed to inhibit vitamin D activation in cases linked to impaired 24-$\alpha$-hydroxylase activity. My drug-target MR studies predict that modulating DGK$\delta$, CaSR, or 24-$\alpha$-hydroxylase to reduce serum calcium or targeting NaPi-IIa to increase serum phosphate could decrease KSD risk by up to ≈90%.

### 5.5.1 Implications of DGK$\delta$-mediated impaired calcium-sensing receptor signalling for kidney stone disease

Idiopathic hypercalciuria, present in up to 50% of KSD cases [64, 65], is a key metabolic abnormality with an important role in the pathogenesis of KSD. Mutations in the CaSR, which result in FHH or autosomal dominant hypocalcaemia, both of which associate with KSD, underscore the importance of CaSR signalling in maintaining calcium homeostasis. In addition to the CaSR itself, G-protein subunit $\alpha$11 ($G_{\alpha 11}$) and adaptor-related protein complex-2 $\sigma$-subunit (AP2$\sigma$) also regulate CaSR signal transduction, and germline mutations in these proteins are implicated in FHH and autosomal dominant hypocalcaemia [163, 164, 629]. Importantly, the studies described in this chapter implicate DGK$\delta$ in the pathogenesis of KSD. DGK$\delta$ is also a key regulator of CaSR signalling [158] and chromatin accessibility and Capture-C analyses demonstrate that the rs838717 risk variant modulates *DGKD* expression. The association of this variant with increased serum calcium and decreased serum phosphate concentrations closely mirrors the effects of impaired CaSR signalling seen in FHH, where altered calcium-sensing receptor signalling leads to elevated serum calcium levels and reduced renal calcium excretion. These findings provide strong evidence supporting the role of DGK$\delta$-mediated defects in CaSR signalling as a key mechanism driving KSD pathophysiology.

The discovery of these genetic insights into CaSR signalling opens new avenues for potential therapeutic interventions targeting the CaSR pathway. Calcimimetics (CaSR positive allosteric modulators) increase the sensitivity of the CaSR to activation by extracellular calcium, causing a rapid, dose-dependent decrease in PTH secretion followed by a lowering of plasma calcium levels [630]. Cinacalcet is a type II calcimimetic used to treat hypercalcemia associated with some cases of primary and secondary hyperparathyroidism and to ameliorate symptomatic hypercalcaemia caused by FHH [630, 631]. Conversely, calcilytics (CaSR negative allosteric modulators) decrease the sensitivity of the CaSR/$G_{\alpha 11}$ signalling pathway to activation by extracellular calcium, reset the set-point for systemic calcium homoeostasis, and decrease urinary calcium excretion [630].

While calcilytics are proposed as potential therapy to prevent idiopathic hypercalciuria-associated KSD by reducing urinary calcium excretion [630], my findings suggest that they may instead increase KSD risk by raising serum calcium concentrations. Increased calcium absorption in the thick ascending limb of the loop of Henle raises the calcium load in descending vasa recta, potentially elevating urine supersaturations sufficiently to cause crystal precipitation [632]. Higher serum and urine calcium levels can promote the formation of Randall's plaques, key to the fixed-particle theory of stone formation (Chapter 1 Section 1.1.4, Figure 1.3 and Figure 1.2). This mechanism may explain how KSD develops in the absence of overt hypercalcaemia or hypercalciuria.

In contrast, CaSR-active compounds such as type II calcimimetics, which decrease serum calcium concentrations, may offer a more targeted approach to reducing KSD risk in individuals with DGK$\delta$-mediated CaSR signalling defects. By restoring normal calcium sensing and reducing serum calcium concentrations, calcimimetics could potentially mitigate the risk of kidney stones in affected patients. However, rigorous *in vitro* and *in vivo* studies are needed to assess the therapeutic potential of both calcimimetics and calcilytics in reducing the risk of KSD in genetically predisposed individuals where CaSR signalling defects play a central role in disease pathogenesis.

### 5.5.2 Implications of NaPi-IIa-mediated increased renal phosphate excretion for kidney stone disease

Monogenic disorders affecting renal tubular phosphate transport, such as those associated with *SLC34A1* and *SLC34A3*, are well-established contributors to hypercalciuria and KSD (Figure 5.10) [176, 179, 244]. In this chapter, I identified rs10051765 as a causal, non-coding, variant linked to an increased risk of KSD, elevated serum calcium, and reduced serum phosphate concentrations. The proximity of rs10051765 to *SLC34A1* raises the possibility of a mechanistic relationship between the two. However, while nearest-gene mapping can provide initial insights, it does not establish causality, as the presence of LD clusters and distant regulatory effects

can obscure the underlying biological mechanisms [633–635].

To overcome these challenges, I integrated multiomic data to identify a correlation of chromatin accessibility at rs10051765 and the promoter regions of *SLC34A1* and *GRK6* with the expression of these genes. This suggests that rs10051765 may alter *SLC34A1* and *GRK6* expression. Capture-C analyses reveal direct chromatin interactions between rs10051765 and the promoters of both genes. These findings suggest that rs10051765 may have a dual mechanism of action, influencing both NaPi-IIa-mediated phosphate transport via direct regulatory effects on the *SLC34A1* promoter and modulating *GRK6* expression. As mentioned above, NHERF1 plays a key role in phosphate transport by affecting NaPi-IIa localisation at the apical plasma membrane of proximal tubular cells. NHERF1 phosphorylation enhances PTH-dependent phosphate transport via NaPi-IIa [636]. This interaction is mediated by NHERF1's PDZ2 domain and tail. Notably, my studies implicate *GRK6* in altered phosphate homeostasis and KSD pathogenesis. The protein it encodes, GRK6A, constitutively phosphorylates NHERF1 through interaction at the PDZ2 domain, facilitating its interaction with NaPi-IIa [609, 636].

These findings suggest that rs10051765 might contribute to KSD through a dual mechanism involving *SLC34A1*-mediated phosphate transport and *GRK6*-mediated regulation of phosphate homeostasis. This underscores the potential for targeted therapies aimed at correcting phosphate transport defects in KSD patients. For instance, phosphate supplementation could be considered to prevent FGF23 suppression and its downstream effects on vitamin D and calcium metabolism. However, Vitamin D supplementation may exacerbate hypercalcaemia and increase the risk of stone formation [183]. Oral phosphate supplements present further challenges, as they require multiple daily doses and can affect serum and urine sodium concentrations, potentially raising KSD risk [637].

Future research should explore the potential of phosphate-rich diets as an alternative and investigate the broader implications of targeting the NaPi-IIa pathway in patients with genetic defects in renal phosphate transport. These studies could inform more effective therapeutic strategies for patients with KSD and aberrations in phosphate metabolism.

### 5.5.3 Implications of 24-$\alpha$-hydroxylase-mediated impaired inactivation of 1,25-dihydroxyvitamin D for kidney stone disease

Impaired 24-$\alpha$-hydroxylase function plays a pivotal role in calcium and vitamin D homeostasis, and its dysfunction can lead to hypercalcaemia and vitamin D toxicity or hypersensitivity [185]. In this chapter, I demonstrate how rs6127099 contributes to KSD pathogenesis by elevating serum calcium and phosphate levels through its effects on *CYP24A1* expression and its direct interaction with the *CYP24A1* promoter. Notably, single-nucleotide variants in the 3' untranslated region of *CYP24A1*, which alter the mRNA's secondary structure, impair translation, and disrupt 1,25-dihydroxyvitamin D catabolism, have been identified in patients with hypercalcaemia, hypercalciuria, and nephrolithiasis, despite normal 25-hydroxyvitamin D levels [185]. These findings underscore the important role of non-coding genetic variants in disease pathophysiology, aligning with my observations regarding rs6127099.

rs6127099 is of considerable clinical relevance given that it may account for ≈14% of KSD cases. This highlights the need to further investigate its effects on 24-$\alpha$-hydroxylase expression and vitamin D metabolites. Importantly, vitamin D regulation extends beyond the kidneys, and extrarenal calcitriol synthesis is particularly significant following renal transplantation. This extrarenal calcitriol synthesis is mainly substrate driven as opposed to hormonally regulated by mineral metabolism hormones such as PTH [638, 639], thus we need a broader understanding of vitamin D metabolism in diverse physiological and pathological contexts to implement therapies which benefit patients with KSD ascribable to impaired inactivation of 1,25-dihydroxyvitamin D.

Identifying genetic variants that impair 24-$\alpha$-hydroxylase activity or enhance 1,25-dihydroxyvitamin D synthesis in patients with KSD could refine clinical management strategies. These may entail recommendations to avoid vitamin D supplementation or therapeutic interventions involving inhibitors of vitamin D

synthesis, such as fluconazole or ketoconazole, to manage hypercalcaemia and reduce the occurrence of kidney stones [186, 187, 640]. Integrating genetic insights into patient care will enable more personalised therapeutic approaches, helping to mitigate the risks posed by impaired 24-$\alpha$-hydroxylase function and offering targeted treatments based on an individual's underlying genetic predisposition to KSD.

### 5.5.4 Limitations

While I present an original approach to causal inference that moves beyond the genetic associations identified in GWAS and provide valuable insights into the pathogenesis of KSD with potential therapeutic implications, there are several limitations to acknowledge. One significant constraint is the absence of reliable GWAS and observational data on urinary calcium and phosphate excretion and 1,25-dihydroxyvitamin D levels, which precludes evaluating the effects of candidate causal variants on these traits. Using cell-line snRNA-seq and snATAC-seq data presents an additional limitation. These methods lack spatial information, restricting the mapping of specific cell types to their anatomical locations. This prevents interrogating localised changes to gene expression and cell–cell communication, both of which are critical to maintain cellular health and are often disrupted in disease. These are areas for future research, particularly to advance spatial genomics and improve our understanding of the mechanisms underlying KSD.

### 5.5.5 Conclusion

In conclusion, in this chapter of my thesis, I define three putative KSD-causing variants by integrating genetic epidemiological, multiomic, and chromatin interaction data. I demonstrate the central roles of DGK$\delta$-mediated impaired CaSR signal transduction, increased renal phosphate excretion, and disrupted 1,25-dihydroxyvitamin D inactivation in the pathogenesis of KSD. These findings reveal potential therapeutic approaches to prevent KSD, offering promising avenues for the development of personalised management strategies that could improve

clinical outcomes for patients with this condition.

# 6 | The role of renal cysts in kidney stone disease

## 6.1 Introduction

### 6.1.1 Gene and gene set associations between renal cysts and kidney stone disease

GWAS of KSD in the UK Biobank and meta-analysed European and trans-ancestry datasets (Chapter 3) reveal enrichment for gene sets linked to renal structural disorders, including "metanephric loop of Henle development". Notably, genes implicated in renal cystogenesis, such as *PKD2*, *AQP1*, and *UMOD*, show significant associations. In Chapter 3 Section 3.4.2, I discussed the relevance of these genes to renal cyst formation. Briefly, germline *PKD2* mutations disrupt nephrogenesis and cause ADPKD where approximately one in five patients develop KSD [126, 131, 132]; *AQP1*, encoding aquaporin-1, a water channel involved in water flux and ion currents, is expressed in renal cysts epithelium in ADPKD and its expression is inversely correlated with cyst size, alluding to a role in cystogenesis [494, 495]; and mutations in *UMOD*, encoding Tamm-Horsfall protein, cause medullary cystic kidney disease type 2 and alter calcium crystal regulation, influencing KSD susceptibility [497, 498]. These findings highlight a potential genetic link between kidney development, tubular structure, ion transport and KSD pathogenesis.

Whether this relationship extends to non-polycystic, "simple" renal cysts and KSD is unclear. Observational studies associate hypercalciuria, hypocitraturia, and a family history of KSD with simple renal cysts, but a shared genetic architecture has not been established [148, 149]. Identifying overlapping genetic pathways could provide mechanistic insights into KSD and inform targeted lifestyle or pharmacological interventions.

### 6.1.2 Aim and Objectives

This chapter aims to evaluate the relationship between non-polycystic renal cysts and KSD to identify shared genetic mechanisms underlying their pathophysiology. The specific objectives are to:

1. **Perform observational analyses**
    - Utilise the UK Biobank and *All of Us* datasets to investigate observational associations between non-polycystic renal cysts and KSD.
2. **Perform regional colocalisation**
    - Identify genomic regions where shared genetic variants influence both renal cyst and KSD risk.
3. **Perform Mendelian randomisation**
    - Investigate bidirectional causal effects between renal cysts and KSD.
4. **Leverage single-cell multiomics and chromatin conformational capture**
    - Explore gene expression and chromatin interactions to gain insights into the shared genetic architecture underlying renal cysts and KSD.
5. **Critically appraise candidate loci and prioritise genes**
    - Investigate the biological roles of identified loci and prioritise candidate genes that may mediate the relationship between renal cysts and KSD.

From hereon in, in this chapter, "cysts" will refer to non-polycystic "simple" cysts.

## 6.2 Methods

### 6.2.1 Cohorts

#### 6.2.1.1 Renal cysts

I used data from the UK Biobank and the *All of Us* cohort to investigate the relationship between renal cystic disease and KSD. I then performed a GWAS of renal cysts in the UK Biobank. Full cohort details are in Chapter 2 Section 2.1.1 and Chapter 2 Section 2.1.4.

#### 6.2.1.2 Kidney stone disease

The European-ancestry KSD GWAS meta-analysis (Chapter 3 Section 3.3.2), the UK Biobank combined-sex KSD GWAS (Chapter 3 Section 3.3.1), and FinnGen study release 12 summary statistics [433] were used in MR and colocalisation analyses.

### 6.2.2 Renal cyst case ascertainment

In the UK Biobank, renal cyst cases were identified using diagnostic codes from primary and secondary healthcare sources as of August 2022 (Appendix B), excluding individuals with polycystic kidney disease or medullary sponge kidney (Appendix C). Dr. Sarah Howles identified participants with renal cysts in the *All of Us* cohort through concept codes (Appendix D), with the same exclusions applied (Appendix E).

### 6.2.3 Observational analyses in the UK Biobank and *All of Us* study

I analysed the associations between simple renal cysts and KSD in both cohorts. To minimise confounding by indication, whereby people with KSD are more likely

to have imaging studies and incidental findings of renal cysts than people without KSD, I included only participants who had undergone abdominal ultrasound, CT, or MRI scans and had codes for non-polycystic renal cysts and/or KSD (Appendices F and G). For each cohort, I calculated the odds ratio of KSD based on cyst presence and performed a $\chi^2$ test of association to assess significance. All statistical analyses were conducted using R (version 4.2.0).

### 6.2.4 A genome-wide association study of renal cysts in the UK Biobank

To enable colocalisation and MR analyses, I conducted a combined-sex GWAS of renal cysts using REGENIE 3.1.1 (inclusion criteria: Appendix B, exclusion criteria: Appendix C) [434]. Full details on QC, test parameters, and GCTA are provided in Chapter 2 Section 2.2.1 and Chapter 2 Section 2.2.5 [457, 458].

### 6.2.5 Colocalisation and Mendelian randomisation analyses

Bidirectional MR was performed as described fully in Chapter 2 Section 2.4. Detailed methods for the colocalisation analyses are described in Chapter 2, Section 2.5. To facilitate analyses with FinnGen KSD GWAS as an exposure, I generated an IV using the publicly-available summary statistics [433], identifying variants significantly associated with KSD, and clumping using the "`clump_data`" function in TwoSampleMR [385, 470] with default parameters and a European population reference panel ($r^2$=0.001, window=10 Kbp).

#### 6.2.5.1 Statistical analyses of Mendelian randomisation estimates

Where appropriate, I performed a meta-analysis of MR estimates derived from UK Biobank and FinnGen analyses using the `meta` package in R [641]. I adopted a random-effects model when heterogeneity of results, evaluated through $I^2$ and Tau$^2$, was identified as >50% and a fixed-effect model when heterogeneity was <50%.

### 6.2.6 Single-cell multiomics

To investigate gene expression, chromatin accessibility, and cellular heterogeneity related to genomic regions identified to be implicated in renal cysts and KSD from colocalisation and MR analyses, I leveraged snRNA-seq and snATAC-seq data from five healthy kidney samples as described in Chapter 2 Section2.6. I studied the three-dimensional structure of chromatin at genomic regions of interest using Capture-C in RPTEC, HK2, and HEK293 cells as described in Methods Chapter 2 Section 2.7 [423, 480–483].

## 6.3 Results

### 6.3.1 Renal cysts are associated with kidney stone disease

After applying inclusion and exclusion criteria, I identified 20,310 eligible participants with renal cysts and/or KSD in the UK Biobank, and 113,258 in *All of Us*. In both datasets, individuals with renal cysts had more than 250% increased odds of having a diagnosis of KSD.

***Table 6.1. Associations of kidney stone disease in participants with and without renal cysts in All of Us and the UK Biobank.***

| **Cohort** | | **No KSD** | **KSD** | **Odds of KSD** | **OR** | $\chi^2$ | $\chi^2$ **P** |
|---|---|---|---|---|---|---|---|
| *All of Us* | No cysts | 90,697 | 13,192 | 0.15 | 2.95 | 2,132.23 | $< 2.20 \times 10^{-16}$ |
| | Cysts | 6,553 | 2,816 | 0.43 | | | |
| UK Biobank | No cysts | 18,734 | 986 | 0.05 | 2.68 | 61.52 | $4.37 \times 10^{-15}$ |
| | Cysts | 517 | 73 | 0.14 | | | |

KSD=Kidney stone disease; OR=Odds ratio; P=P-value.

### 6.3.2 Genome-wide association study of renal cysts in the UK Biobank

A combined-sex GWAS of 5,318 cases and 402,601 controls in the UK Biobank, identified 5 independent, GWAS significant associations with renal cysts at 5 loci across 5 chromosomes (Table 6.2, Figure 6.1, Figure 6.2).

***Table 6.2. Single nucleotide polymorphisms significantly associated with renal cysts from a genome-wide association study in the UK Biobank.***

| SNP | CHR | POS | Candidate Gene | EA | NEA | EAF | OR (95% CI) | P |
|---|---|---|---|---|---|---|---|---|
| rs11267336 | 1 | 202004254 | *ELF3* | A | ATGGT CACAT ACCTA AGTAT GTG | 0.55 | 1.12 (1.09-1.14) | $3.68\times10^{-8}$ |
| rs6751866 | 2 | 54573209 | *C2orf73* | A | G | 0.19 | 1.14 (1.11-1.22) | $3.11\times10^{-10}$ |
| rs60230975 | 4 | 167676881 | *SPOCK3* | T | G | 0.22 | 1.16 (1.11-1.21) | $2.08\times10^{-10}$ |
| 16:75457178_CA_C | 16 | 75457178 | *CFDP1* | CA | C | 0.38 | 1.12 (1.08-1.17) | $1.31\times10^{-8}$ |
| rs9895661 | 17 | 59456589 | *BCAS3* | T | C | 0.83 | 1.23 (1.17-1.30) | $1.60\times10^{-14}$ |

Variants significantly associated with renal cysts in a genome-wide association study in the UK Biobank.
CHR=Chromosome; EA=Effect allele; EAF=Effect allele frequency; NEA=Non-effect allele; OR=Odds ratio; P=P-value; POS=Position based on NCBI Genome Build 37 (hg19); 95% CI=95% confidence interval.

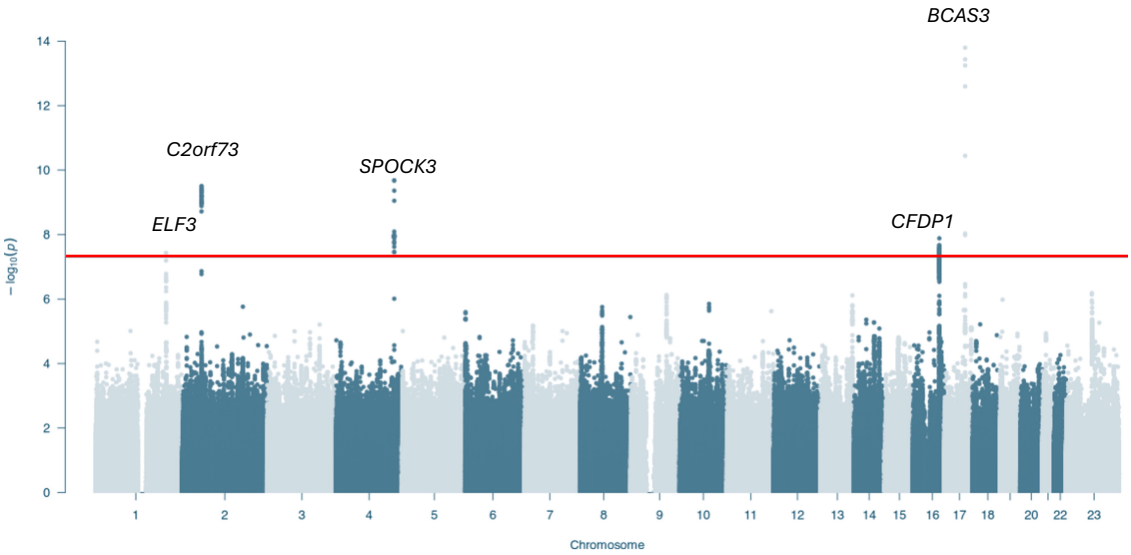



***Figure 6.1. Manhattan plot for a genome-wide association study of renal cysts in the UK Biobank.***

Genome-wide p-values (-log10) plotted against their respective positions on each chromosome. The horizontal red line indicates the genome-wide significance threshold of $5.00 \times 10^{-8}$. Loci are labelled with the primary candidate gene at each locus, as shown in Table 6.2.

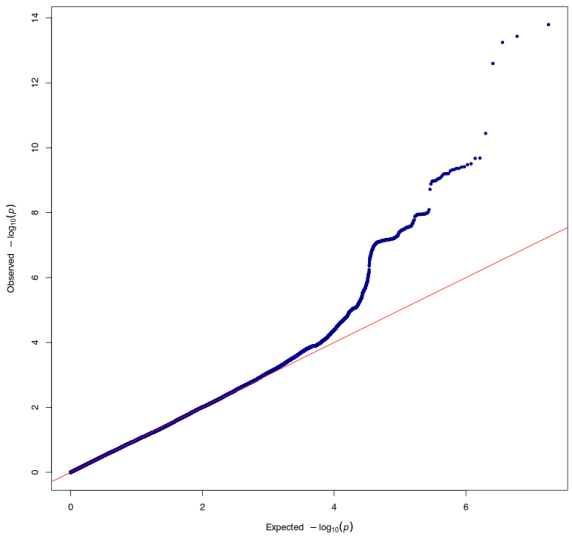



***Figure 6.2. Quantile-quantile plot of observed vs. expected p-values for a genome-wide association study of renal cysts in the UK Biobank.***

A genome-wide association study of renal cysts was performed for 5,318 cases with renal cysts and 402,601 controls from the UK Biobank. The $\lambda$ GC demonstrated some inflation (1.25), but the LDSC intercept of 1.01, with an attenuation ratio of 0.17 indicated that the inflation was largely due to polygenicity and the large sample size.

### 6.3.3 Renal cysts and kidney stone disease colocalise at the *BCAS3* locus

Regional colocalisation analyses at the five 1 Mbp regions surrounding the lead, independent variants identified from the GWAS of renal cysts in the UK Biobank (Table 6.2, Figure 6.1) shows colocalisation of KSD with renal cysts at one 1 Mbp region associated with *BCAS3*. At rs9895661±500 Kbp (PPH4 >0.91) there is strong evidence of colocalisation of these traits in analyses involving data from European-ancestry KSD GWAS meta-analysis, UK Biobank KSD GWAS, and FinnGen KSD GWAS (Table 6.3, Table 6.4, Figure 6.3).

***Table 6.3. Results from regional colocalisation of renal cysts and kidney stone disease.***

| Lead SNP (Gene) | CHR | POS | Study | N SNPs | PP H0 | PP H1 | PP H2 | PP H3 | PP H4 | Candidate SNP | SNP PP |
|---|---|---|---|---|---|---|---|---|---|---|---|
| rs11267336 *(ELF3)* | 1 | 202004254 | Meta-analysis | 2,387 | 0.06 | 0.89 | $2.71\times10^{-3}$ | 0.04 | 0.01 | | |
| | | | UK Biobank | 2,588 | 0.05 | 0.78 | 0.01 | 0.15 | 0.01 | | |
| | | | FinnGen | 2,154 | 0.09 | 0.82 | 0.01 | 0.07 | 0.01 | | |
| rs6751866 *(C2orf73)* | 2 | 54573209 | Meta-analysis | 3,173 | $1.06\times10^{-4}$ | 0.88 | $1.31\times10^{-5}$ | 0.11 | 0.01 | | |
| | | | UK Biobank | 3,262 | $1.11\times10^{-4}$ | 0.92 | $8.24\times10^{-6}$ | 0.07 | 0.01 | | |
| | | | FinnGen | 2,743 | $9.98\times10^{-5}$ | 0.76 | $2.95\times10^{-5}$ | 0.23 | 0.01 | | |
| rs60230975 *(SPOCK3)* | 4 | 167676881 | Meta-analysis | 3,780 | $2.71\times10^{-4}$ | 0.88 | $3.59\times10^{-5}$ | 0.12 | 0.01 | | |
| | | | UK Biobank | 3,813 | $1.62\times10^{-4}$ | 0.52 | $1.46\times10^{-4}$ | 0.47 | $4.61\times10^{-3}$ | | |
| | | | FinnGen | 3,145 | $3.45\times10^{-4}$ | 0.92 | $2.38\times10^{-5}$ | 0.06 | 0.01 | | |
| 16:75457178_CA_C *(CFDP1)* | 16 | 75457178 | Meta-analysis | 3,209 | $2.01\times10^{-4}$ | 0.14 | $5.68\times10^{-4}$ | 0.38 | 0.48 | | |
| | | | UK Biobank | 3,350 | $1.20\times10^{-3}$ | 0.81 | $1.67\times10^{-4}$ | 0.11 | 0.08 | | |
| | | | FinnGen | 2,718 | $7.31\times10^{-4}$ | 0.41 | $6.88\times10^{-4}$ | 0.38 | 0.2 | | |
| rs9895661 *(BCAS3)* | 17 | 59456589 | Meta-analysis | 1,450 | $1.40\times10^{-7}$ | $1.74\times10^{-7}$ | 0.04 | 0.05 | 0.91 | rs1000423 | 0.12 |
| | | | UK Biobank | 1,690 | $9.91\times10^{-10}$ | 0.02 | $3.25\times10^{-10}$ | 0.01 | 0.97 | rs9895661 | 0.43 |
| | | | FinnGen | 1,393 | $5.82\times10^{-21}$ | $1.36\times10^{-13}$ | $3.33\times10^{-10}$ | 0.01 | 0.99 | rs9895661 | 0.76 |

H0=The region is not associated with either trait; H1=The region is associated with renal cysts but not kidney stone disease; H2=The region is associated with kidney stone disease but not renal cysts; H3=The region is associated with both traits with two distinct causal variants; H4=The region is associated with both traits with a shared causal variant; PP=posterior probability of colocalisation.

***Table 6.4. 95% credible sets of colocalisation analyses between renal cysts and kidney stone disease at rs9895661± 500 kilobase pairs.***

| Study | SNP | SNP PP |
|---|---|---|
| Meta-analysis | rs1000423 | 0.12 |
| | rs2286526 | 0.10 |
| | rs2240736 | 0.07 |
| | rs34446110 | 0.06 |
| | rs2270114 | 0.06 |
| | rs729781 | 0.06 |
| | rs1058004 | 0.06 |
| | rs12952625 | 0.06 |
| | rs35619711 | 0.05 |
| | rs11871637 | 0.05 |
| | rs8078036 | 0.05 |
| | rs1057987 | 0.05 |
| | rs8073698 | 0.05 |
| | rs9891115 | 0.05 |
| | rs8074151 | 0.04 |
| | rs7215775 | 0.04 |
| | rs8068318 | 0.04 |
| UK Biobank | rs9895661 | 0.43 |
| | rs9905274 | 0.43 |
| | rs11657044 | 0.11 |
| FinnGen | rs9895661 | 0.76 |
| | rs9905274 | 0.13 |
| | rs11657044 | 0.09 |

PP=Posterior probability.

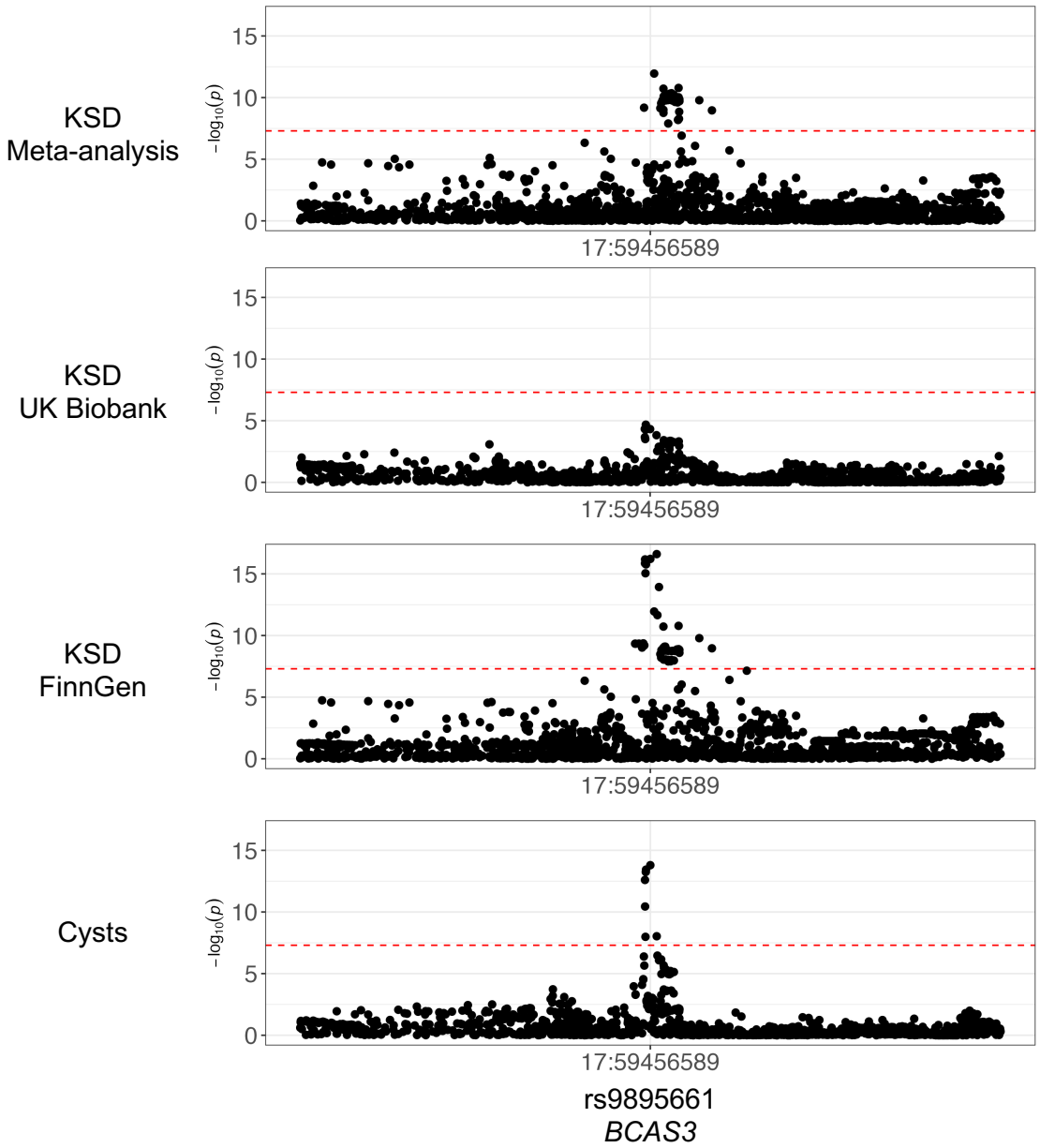



***Figure 6.3. Locus-zoom plots of 1 megabase pair regions with colocalisation between renal cysts and kidney stone disease.***

Genome-wide P values (-log10) plotted against respective positions on each chromosome. The central line and position annotate the index variant from the renal cyst genome-wide association study. The horizontal, dashed, red line corresponds to p=$5\times10^{-8}$.

KSD=Kidney stone disease.

The rs9895661±500 Kbp locus corresponds to *BCAS3*, which encodes Breast Carcinoma-Amplified Sequence 3 Microtubule-Associated Cell Migration Factor (BCAS3), a cytoskeletal protein expressed in renal glomerular and tubular cells. BCAS3 regulates cell polarity and directional endothelial migration by mediating actin cytoskeleton reorganisation [551]. Notably, the three variants comprising the

95% credible set at this locus, identified in UK Biobank and FinnGen analyses, are reported as expression quantitative trait loci (eQTL) that reduce *TBX2* and *TBX2-AS1* expression in renal tubular cells (Table 6.5) [624].

These three candidate causal variants, consistent across UK Biobank and FinnGen analyses, were excluded from the European-ancestry KSD GWAS meta-analysis summary statistics due to high heterogeneity ($I^2 > 75.00$) (Chapter 2, Section 2.2.4). However, they are significantly associated with KSD in the FinnGen GWAS and attain a replication threshold for significance ($p \leq 5 \times 10^{-8}$) in the UK Biobank GWAS (Table 6.5).

***Table 6.5. Associations of candidate causal variants identified from colocalisation analyses between renal cysts and kidney stone disease.***

| | | | | |
|---|---|---|---|---|
| **Candidate causal variant** | SNP | rs9895661 | rs9905274 | rs11657044 |
| | POS | 59456589 | 59450441 | 59450105 |
| | EA | T | C | C |
| | NEA | C | T | T |
| | EAF | 0.43 | 0.43 | 0.11 |
| **Cysts** | Odds ratio (95%CI) | 1.23 (1.17-1.30) | 1.23 (1.17-1.30) | 1.23 (1.19-1.28) |
| | P | $1.60\times10^{-14}$ | $3.67\times10^{-14}$ | $5.66\times10^{-14}$ |
| **KSD UK Biobank** | Odds ratio (95%CI) | 1.08 (1.04-1.12) | 1.08 (1.04-1.12) | 1.08 (1.04-1.14) |
| | P | $4.80\times10^{-5}$ | $2.10\times10^{-5}$ | $5.40\times10^{-5}$ |
| **KSD FinnGen** | Odds ratio (95%CI) | 1.15 (1.11-1.19) | 1.15 (1.11-1.19) | 1.15 (1.11-1.19) |
| | P | $6.11\times10^{-}17$ | $1.70\times10^{-16}$ | $1.57\times10^{-16}$ |
| ***TBX2*** | $\beta$ (SE) | -0.20 (-0.03) | -0.22 (-0.03) | -0.18 (-0.03) |
| | P | $5.21\times10^{-9}$ | $1.03\times10^{-10}$ | $9.45\times10^{-8}$ |

CHR=Chromosome; GWAS=Genome-wide association study; KSD=Kidney stone disease; P=P-value; POS=position based on NCBI Genome Build 37 (hg19); OR=Odds ratio; SE=Standard error; 95% CI=95% confidence interval.

### 6.3.4 Evidence for a bidirectional causal relationship between renal cysts and kidney stone disease

Bidirectional MR analyses assessing the effect of increased liability to renal cysts on KSD risk, and vice versa, yield inconsistent results (Table 6.6, Figure 6.4). After adjusting for multiple testing, analyses involving European-ancestry KSD GWAS meta-analysis data support an effect where greater liability to KSD increases the odds of developing renal cysts. In contrast, analyses with UK Biobank KSD GWAS data show no effect of greater liability to renal cysts on liability to KSD after adjusting the p-value for multiple testing (Table 6.6, Figure 6.4). Analyses using FinnGen KSD GWAS data suggest that greater liability to KSD increases the odds of renal cysts, while greater liability to renal cysts does not significantly affect the risk of KSD (Table 6.6, Figure 6.4).

***Table 6.6. Bidirectional Mendelian randomisation analyses between renal cysts and kidney stone disease.***

| **Analysis** | | | **MR Analysis*** | | | **MR heterogeneity** | | | **MR Intercept** | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Exposure** | **Outcome** | **Mean $R^2$ (SD)** | **N SNPs** | **Odds ratio (95%CI)** | **P** | **Q** | **Q_df** | **Q P** | **$\beta$ (SE)** | **P** |
| Cysts | KSD (Meta-analysis) | | 4 | 1.11 (0.98-1.26) | 0.11 | 6.11 | 3 | 0.11 | 0.03 (0.08) | 0.78 |
| Cysts | KSD (UK Biobank) | $9.87\times10^{-5}$ ($2.37\times10^{-5}$) | 5 | 1.20 (1.01-1.41) | 0.09 | 12.22 | 4 | 0.02 | -0.05 (0.05) | 0.44 |
| Cysts | KSD (FinnGen) | | 3 | 1.45 (0.99-2.12) | 0.09 | 25.87 | 2 | $2.42\times10^{-6}$ | -0.29 (0.06) | 0.12 |
| KSD (Meta-analysis) | Cysts | $5.81\times10^{-5}$ ($4.26\times10^{-5}$) | 74 | 1.20 (1.11-1.30) | $7.53\times10^{-6}$ | 97.79 | 73 | 0.03 | $-3.54\times10^{-3}$ ($8.78\times10^{-3}$) | 0.69 |
| KSD (UK Biobank) | Cysts | $1.40\times10^{-4}$ ($6.11\times10^{-5}$) | 18 | 1.09 (0.99-1.20) | 0.09 | 19.42 | 17 | 0.30 | 0.02 (0.02) | 0.45 |
| KSD (FinnGen) | Cysts | $1.14\times10^{-4}$ ($6.82\times10^{-5}$) | 19 | 1.26 (1.07-1.48) | $9.75\times10^{-3}$ | 53.57 | 18 | $2.14\times10^{-5}$ | $-6.47\times10^{-3}$ (0.03) | 0.81 |

MR analysis*=Mendelian randomisation- inverse weighted variance estimate where intercept p-value is >0.05, MR Egger estimate where intercept p-value is <0.05; N SNPs=Number of SNPs; OR=Odds ratio for outcome per unit increase in the log-odds of genetic liability to exposure trait; P=5% false discovery rate adjusted p-value; POS=Position based on NCBI Genome Build 37 (hg19); Q=Q statistic; Q_df=Q degrees of freedom; $R^2$=Variance in exposure explained by instrumental variable; SD=Standard deviation; SE=Standard error; 95% CI=95% confidence interval.

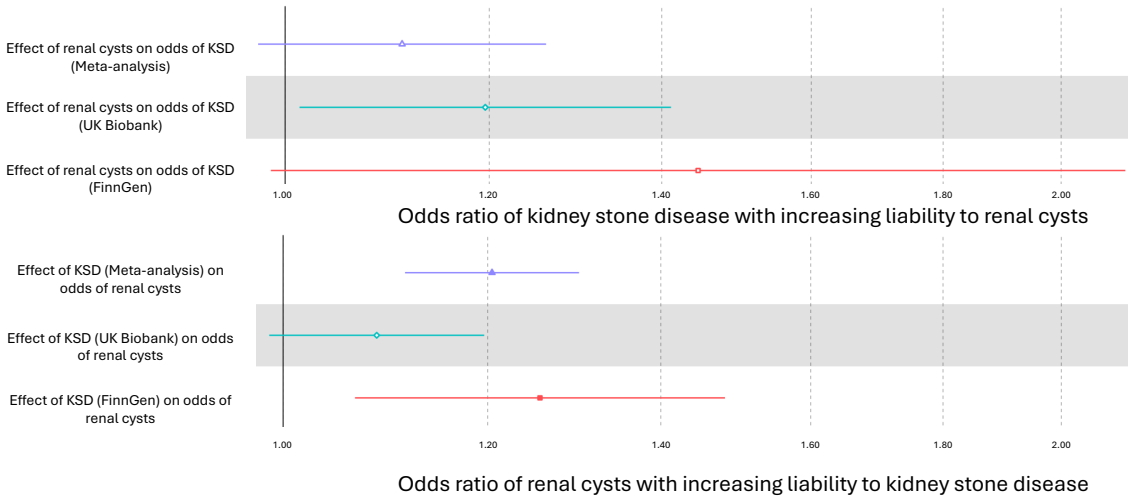



***Figure 6.4. Bidirectional Mendelian randomisation analyses between renal cysts and kidney stone disease.***

The odds ratio is scaled per unit increase in the log-odds of genetic liability to the exposure trait. This equates to a doubling in the genetic liability to the exposure trait.

KSD=Kidney stone disease.

The inconsistency in the results leveraging data from the European-ancestry KSD GWAS meta-analysis likely arises from the small number of SNPs in the exposure IV. Leave-one-out analysis indicates that rs9895661 drives the strongest effect in UK Biobank analyses (Figure 6.5). As discussed in the colocalisation results, this variant was excluded from the meta-analysis summary statistics due to its high heterogeneity ($I^2$=86.2).

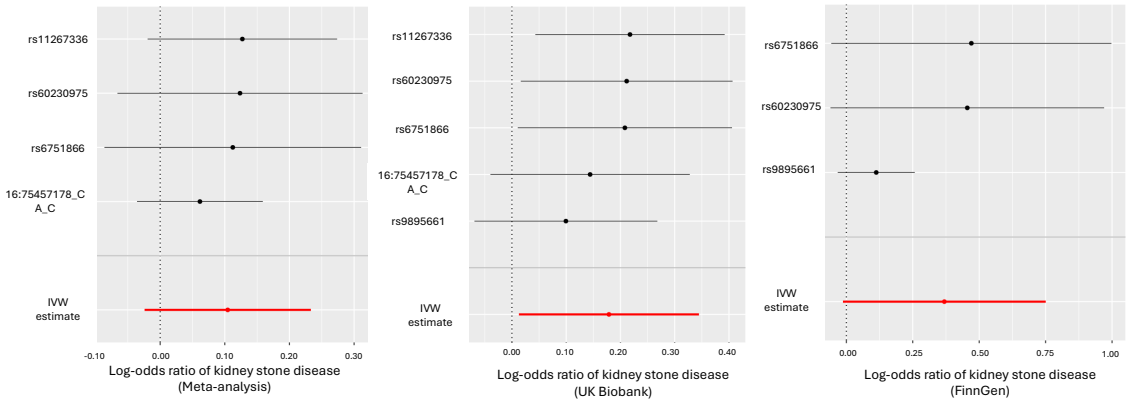



***Figure 6.5. Leave-one-out analyses for the effect of increasing genetic liability to renal cysts on the odds of kidney stone disease.***

The effect shown reflect the log-odds ratio of kidney stone disease after removing the corresponding variant from the instrumental variable for renal cysts.

Given the unreliability of the meta-analysis result, I meta-analysed the MR estimates from UK Biobank and FinnGen analyses. The meta-analysed results indicate a bidirectional causal relationship between liability to renal cysts and KSD (Figure 6.6).

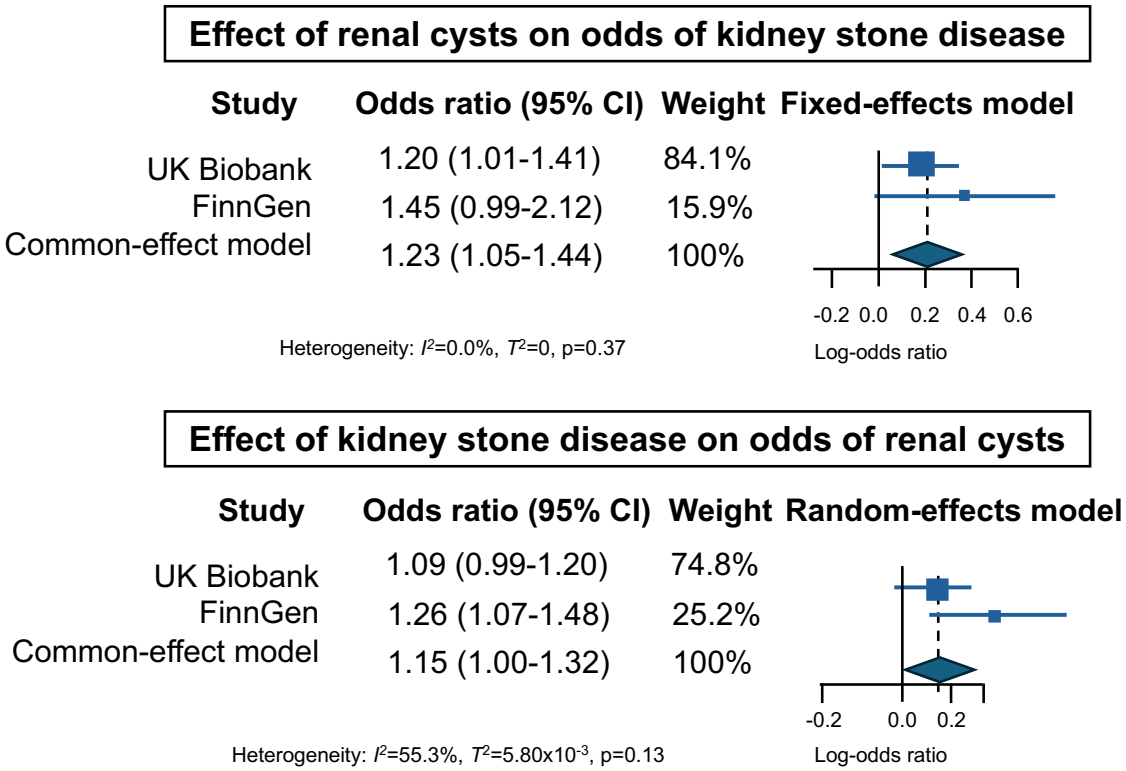



***Figure 6.6. Fixed-effect meta-analysis of bidirectional Mendelian randomisation estimates between liability to renal cysts and liability to kidney stone disease.***

The effect shown reflect the log-odds ratio of kidney stone disease.

### 6.3.5 The *BCAS3* locus mediates the causal effect of renal cysts on kidney stones

To determine whether the bidirectional causal effect of renal cysts on KSD risk is mediated by the *BCAS3* locus, I conducted *cis*-MR within a ±500 Kbp region of the candidate causal SNP with the highest SNP-posterior probability, rs9895661. Applying an $r^2 < 0.1$ threshold for independence, only one variant (rs9895661) met the criteria for use as an IV to proxy the effects of increasing liability to renal cysts. Wald ratio analysis identified a 42–94% increased risk of KSD per unit increase in liability to renal cysts (Table 6.7, Figure 6.7).

***Table 6.7. Regional bidirectional Mendelian randomisation between renal cysts and kidney stone disease mediated by rs9895661±500 kilobase pairs.***

| Exposure | Outcome | $R^2$ | N SNPs | OR (95%CI) | P |
|---|---|---|---|---|---|
| Cysts | KSD (UK Biobank) | $1.45\times10^{-3}$ | 1 | 1.42 (1.20-1.68) | $4.83\times10^{-5}$ |
| Cysts | KSD (FinnGen) | | | 1.94 (1.66-2.27) | $1.22\times10^{-16}$ |
| KSD (Meta-analysis) | Cysts | $4.78\times10^{-5}$ | 1 | 4.32 (2.27-8.22) | $8.58\times10^{-6}$ |
| KSD (UK Biobank ($p<5\times10^{-5}$)) | Cysts | $4.51\times10^{-5}$ | 1 | 14.81 (7.37-29.74) | $1.10\times10^{-13}$ |
| KSD (FinnGen) | Cysts | $1.44\times10^{-4}$ | 1 | 3.24 (2.17-4.84) | $1.40\times10^{-8}$ |

MR analysis∗=Mendelian randomisation; N SNPs=Number of SNPs; OR=Odds ratio for outcome per unit increase in log-odds of genetic liability to exposure trait; P=5% false discovery rate adjusted p-value; POS=Position based on NCBI Genome Build 37 (hg19); $R^2$=Variance in exposure explained by instrumental variable; 95% CI=95% Confidence interval.

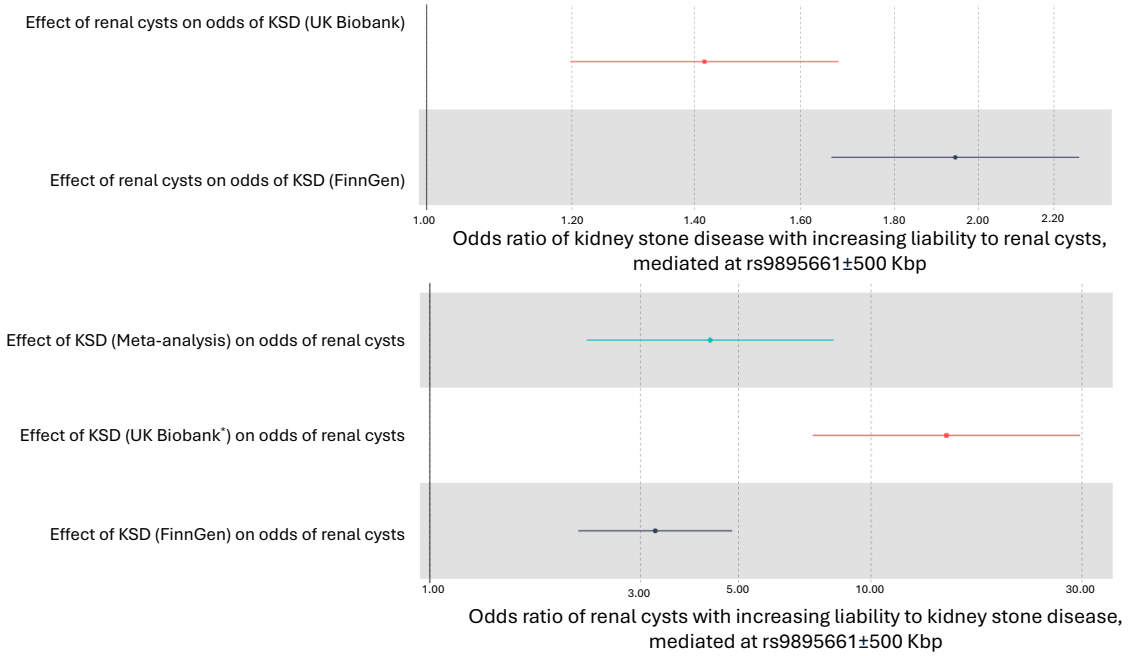



***Figure 6.7. Regional bidirectional Mendelian randomisation between renal cysts and kidney stone disease mediated by rs9895661±500 kilobase pairs.***

KSD=Kidney stone disease; UK Biobank *=Replication threshold for inclusion in instrumental variable ($p<5\times10^{-5}$) as no SNPs attained $p<5\times10^{-8}$ in UK Biobank KSD genome-wide association study in this region.

Similarly, applying an $r^2 < 0.1$ threshold for independence, only one variant met the criteria for use as an IV to proxy the effects of increasing liability to KSD in data from the European-ancestry KSD GWAS meta-analysis (rs1058004) and FinnGen data (rs2079742). Analyses support a causal effect of greater liability to KSD on increased liability to renal cysts, mediated by this rs9895661±500 *BCAS3*-associated locus (Table 6.7, Figure 6.7). In UK Biobank KSD GWAS data, no SNPs at rs9895661±500 Kbp reach genome-wide significance, precluding their use as IVs for MR. However, MR analyses using a KSD IV comprising one conditionally-independent SNP (rs9905274) meeting a replication threshold for significance ($p<5\times10^{-5}$) yield directionally concordant results (Table 6.7, Figure 6.7).

Together, evidence amassed from observational analyses that associate renal cysts and KSD, the evidence of these two traits colocalising at rs9895661±500 Kbp, and both conventional and *cis*-MR pointing towards a bidirectional causal effect between these two traits, implicating rs9895661±500 Kbp, suggest that this *BCAS3* locus is important in the pathogenesis of both KSD and renal cysts. However, dataset inconsistencies highlight the need for additional replication and functional studies to fully elucidate the mechanisms involved.

### 6.3.6 The functional relevance of rs9895661

#### 6.3.6.1 Characteristics of rs9895661 in chromatin co-accessibility and RNA expression in renal cell types

Mapping chromatin accessibility at rs9895661±100 bp with chromatin regions corresponding to promoter regions, defined using Ensembl v113 [642], using snATAC-seq revealed no correlation between gene activity and RNA expression for any regions that are co-accessible with rs9895661 (Table 6.8, Figure 6.8).

***Table 6.8. Pearson correlation coefficients between gene activity and expression on chromosome 17 across renal cell types.***

| Gene | Maximum activity | Maximum expression | Pearson $r^2$ | Pearson P | Cell type | Mean activity | Mean expression |
|---|---|---|---|---|---|---|---|
| *TBX4* | $1.58\times10^{-6}$ | 0.19 | -0.71 | 0.11 | FB | $6.86\times10^{-10}$ | 0.18 |
| | | | | | PT | $4.43\times10^{-10}$ | 0.13 |
| | | | | | PEC | $1.32\times10^{-9}$ | 0.13 |
| | | | | | ICA | $3.74\times10^{-10}$ | 0.12 |
| | | | | | DT | $3.57\times10^{-10}$ | 0.11 |
| | | | | | TAL | $1.43\times10^{-6}$ | 0.07 |
| *TBX2* | 1.10 | 1.26 | -0.38 | 0.35 | PT | 0.95 | 1.14 |
| | | | | | ICA | 1.00 | 1.04 |
| | | | | | ICB | 1.00 | 0.90 |
| | | | | | FB | 0.96 | 0.72 |
| | | | | | DT | 1.00 | 0.71 |
| | | | | | PEC | 0.98 | 0.43 |
| | | | | | ENDO | 1.00 | 0.14 |
| | | | | | TAL | 1.00 | 0.10 |

Genes arranged in descending order of significance. Cell types arranged in descending order of gene expression. DT=Distal tubule; ENDO=Endothelial; FB=Fibroblast; GLOM=Glomerular; ICA=Intercalated A; P=P-value; PEC=Parietal epithelial; PT=Proximal tubule; TAL=Thick ascending limb of the loop of Henle.

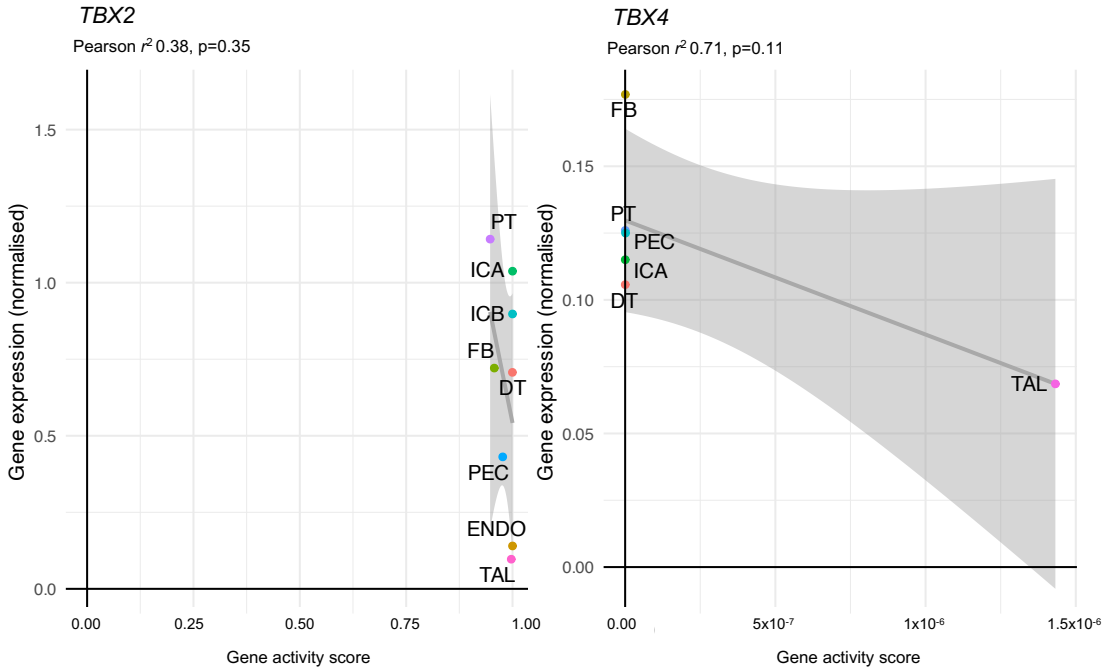



***Figure 6.8. Pearson correlation coefficients of gene activity, inferred from rs9895661 coaccessibility, and RNA expression of TBX2 and TBX4.***

Gene activity score reflects the weighted coaccessibility between rs9895661 and gene promoter regions. Gene expression reflects normalised gene expression from snRNA-seq data.

DT=Distal tubule; ENDO=Endothelial; FB=Fibroblast; GLOM=Glomerular; ICA=Intercalated A; P=P-value; PEC=Parietal epithelial; PT=Proximal tubule; TAL=Thick ascending limb of the loop of Henle.

#### 6.3.6.2 Regulatory potential of rs9895661 inferred through chromatin interactions with *TBX2-AS1*

Given the negative findings from chromatin co-accessibility and RNA expression analysis, I employed Capture-C analysis to investigate chromatin interactions in an unbiased manner. Thus, I extended analyses to include protein-coding genes and long non-coding RNAs to explore potential effects of rs9895661 on gene expression, chromatin remodelling, or cell differentiation more broadly. I identified that the chromatin region within which rs9895661 is located interacts with the *TBX2-AS1* promoter in RPTEC, HK2, and HEK293 cells (Table 6.9, Figure 6.9). This is the only promoter region with significant interactions across all cell lines.

***Table 6.9. Significant interactions of rs9895661 with gene promoters in renal cell lines identified in Capture-C analysis.***

| Cell line | Promoter | | | | MPPC score |
|---|---|---|---|---|---|
| | Gene | ENSR ID | Coordinates (hg38) | Distance from rs9895661 | |
| RPTEC | *MRPL38* | 00001344590 | 17:75904828-75906583 | 14525.6 Kbp downstream | 0.98 |
| | | 17_9355G9 | 17:75904874-75905601 | 14525.65 Kbp downstream | 0.98 |
| | *KRTAP1-6* | 00001341864 | 17:41027198-41027599 | 20351.63 Kbp upstream | 0.98 |
| | | 17_9L9NC | 17:41027198-41027550 | 20351.68 Kbp upstream | 0.98 |
| | *ZNF750* | 00001345573 | 17:82840012-82840999 | 21460.78 Kbp downstream | 0.67 |
| | *OGFOD3* | 00001345496 | 17:82403957-82405457 | 21024.73 Kbp downstream | 0.67 |
| | *ARHGDIA* | 00001345349 | 17:81869996-81872868 | 20490.77 Kbp downstream | 0.59 |
| | | 17_BL8NS | 17:81870964-81872866 | 20491.74 Kbp downstream | 0.59 |
| | *KRT36* | 00001341920 | 17:41489900-41491400 | 19887.83 Kbp upstream | 0.57 |
| | *FOXK2* | 00001345533 | 17:82584640-82586140 | 21205.41 Kbp downstream | 0.51 |
| | *MYO1D* | 00001341213 | 17:32659132-32659743 | 28719.49 Kbp upstream | 0.42 |
| | *CRLF3* | 00001341047 | 17:30809976-30811476 | 30567.75 Kbp upstream | 0.34 |
| | *BCAS3* | 00001343608 | 17:60676401-60677902 | 701.33 Kbp upstream | 0.32 |
| | | 17_BG35K | 17:60676709-60677899 | 701.33 Kbp upstream | 0.32 |
| | *TNS4* | 00001341822 | 17:40480865-40482365 | 20896.86 Kbp upstream | 0.32 |
| | *STAT3* | 00001342068 | 17:42317505-42319193 | 19060.04 Kbp upstream | 0.32 |
| | *FAM222B* | 00001340841 | 17:28812214-28813925 | 32565.3 Kbp upstream | 0.21 |
| | | 17_B7PDH | 17:28812214-28813415 | 32565.81 Kbp upstream | 0.21 |
| | *ARHGAP27P1 -BPTFP1 -KPNA2P3* | 17_9XP7W | 17:64778659-64779555 | 3399.43 Kbp downstream | 0.18 |
| | *CBX8* | 00001345063 | 17:79801523-79803173 | 18422.3 Kbp downstream | 0.14 |
| | *JUP* | 00001341976 | 17:41772259-41773759 | 19605.47 Kbp upstream | 0.12 |
| | *KRT19* | 00001341941 | 17:41528289-41529798 | 19849.43 Kbp upstream | 0.10 |
| | | 17_9LH3C | 17:41528298-41529798 | 19849.43 Kbp upstream | 0.10 |
| | *TBX2-AS1* | 17_BG6ZL | 17:61396303-61396919 | 17.08 Kbp downstream | 0.02 |
| | | 00001343634 | 17:61394052-61395436 | 14.82 Kbp downstream | 0.02 |
| HK2 | *HES7* | 00001339741 | 17:8124074-8125596 | 53253.63 Kbp upstream | 0.99 |
| | | 17_B3LJD | 17:8124074-8125596 | 53253.63 Kbp upstream | 0.99 |
| | *MYHAS* | 00001339909 | 17:10613251-10613490 | 50765.74 Kbp upstream | 0.99 |
| | *COPS3* | 00001340203 | 17:17280690-17282763 | 44096.47 Kbp upstream | 0.98 |
| | | 17_C3PB4 | 17:17280690-17281961 | 44097.27 Kbp upstream | 0.98 |
| | *GPS2* | 00001339564 | 17:7313260-7314807 | 54064.42 Kbp upstream | 0.98 |
| | *RTN4RL1* | 00001339001 | 17:2025324-2026599 | 59352.63 Kbp upstream | 0.96 |
| | | 17_9RP4 | 17:2025324-2026501 | 59352.73 Kbp upstream | 0.96 |
| | *PPM1D* | 17_BG2PC | 17:60599117-60600425 | 778.8 Kbp upstream | 0.93 |
| | *TUBD1* | 00001343539 | 17:59881138-59881255 | 1497.97 Kbp upstream | 0.77 |
| | *SPNS3* | 00001339247 | 17:4432450-4433973 | 56945.26 Kbp upstream | 0.57 |
| | | 17_93QHW | 17:4433464-4433950 | 56945.28 Kbp upstream | 0.57 |
| | *NUP85* | 00001344434 | 17:75204169-75205730 | 13824.94 Kbp downstream | 0.42 |
| | *PRR15L* | 00001342768 | 17:47957873-47959373 | 13419.86 Kbp upstream | 0.37 |
| | | 17_9P2XQ | 17:47957873-47958485 | 13420.74 Kbp upstream | 0.37 |
| | *OTOP3* | 00001344386 | 17:74935605-74935908 | 13556.38 Kbp downstream | 0.18 |
| | *TBCD* | 00001345584 | 17:82880004-82881504 | 21500.78 Kbp downstream | 0.12 |
| | *DUS1L* | 00001345428 | 17:82059131-82059231 | 20679.9 Kbp downstream | 0.11 |
| | | 00001343646 | 17:61451402-61452432 | 72.17 Kbp downstream | 0.05 |

| Cell line | Gene | ENSR ID | Coordinates (hg38) | Distance from rs9895661 | MPPC score |
|---|---|---|---|---|---|
| | *TBX4* | 17_9WDKB | 17:61451300-61452432 | 72.07 Kbp downstream | 0.05 |
| | | 17_BG7B8 | 17:61454956-61456474 | 75.73 Kbp downstream | 0.02 |
| | *TBX2-AS1* | 17_BG6ZL | 17:61396303-61396919 | 17.08 Kbp downstream | 0.01 |
| | *METTL2A* | 17_9WQZD | 17:62423343-62423907 | 1044.12 Kbp downstream | 1.00 |
| HEK293 | *STAT5B* | 00001342058 | 17:42276315-42277999 | 19101.23 Kbp upstream | 0.99 |
| | | 17_9LR3S | 17:42276381-42277728 | 19101.5 Kbp upstream | 0.99 |
| | *USP36* | 00001344998 | 17:78838784-78840284 | 17459.56 Kbp downstream | 0.99 |
| | | 00001345000 | 17:78840637-78842931 | 17461.41 Kbp downstream | 0.99 |
| | | 17_BKMGD | 17:78840660-78842004 | 17461.43 Kbp downstream | 0.99 |
| | *ST6GALNAC1* | 00001344725 | 17:76643650-76645199 | 15264.42 Kbp downstream | 0.99 |
| | *OGFOD3* | 00001345496 | 17:82403957-82405457 | 21024.73 Kbp downstream | 0.99 |
| | *PRPSAP2* | 00001340369 | 17:18859025-18860525 | 42518.7 Kbp upstream | 0.99 |
| | *CYGB* | 00001344689 | 17:76528983-76530483 | 15149.76 Kbp downstream | 0.99 |
| | *KCNH4* | 00001342044 | 17:42181071-42182632 | 19196.6 Kbp upstream | 0.99 |
| | | 17_BBD2L | 17:42181071-42181906 | 19197.32 Kbp upstream | 0.99 |
| | *SEPTIN9* | 00001344808 | 17:77318028-77319612 | 15938.8 Kbp downstream | 0.99 |
| | *NLE1* | 00001341303 | 17:35142223-35142999 | 26236.23 Kbp upstream | 0.99 |
| | *MED9* | 00001340211 | 17:17475602-17477035 | 43902.19 Kbp upstream | 0.99 |
| | *ASPSCR1* | 00001345401 | 17:82006653-82007445 | 20627.43 Kbp downstream | 0.99 |
| | *CCL18* | 00001341391 | 17:36063401-36064282 | 25314.95 Kbp upstream | 0.99 |
| | *EVPLL* | 00001340319 | 17:18376602-18377788 | 43001.44 Kbp upstream | 0.99 |
| | *TIMP2* | 00001345013 | 17:78925377-78926877 | 17546.15 Kbp downstream | 0.98 |
| | | 17_936BS9 | 17:78925377-78926477 | 17546.15 Kbp downstream | 0.98 |
| | *CXCL16* | 00001339311 | 17:4736370-4737870 | 56641.36 Kbp upstream | 0.98 |
| | *AATF* | 00001341468 | 17:37050927-37051145 | 24328.08 Kbp upstream | 0.98 |
| | *IGBP1C* | 00001343479 | 17:58692013-58692937 | 2686.29 Kbp upstream | 0.98 |
| | | 17_BFMHD | 17:58692013-58692892 | 2686.34 Kbp upstream | 0.98 |
| | *MFSD11* | 00001344759 | 17:76772002-76773107 | 15392.77 Kbp downstream | 0.98 |
| | *CHMP6* | 00001345166 | 17:80990108-80991851 | 19610.88 Kbp downstream | 0.98 |
| | *RENO1* | 00001345235 | 17:81345956-81347456 | 19966.73 Kbp downstream | 0.98 |
| | *MPRIP* | 00001340173 | 17:17170202-17170870 | 44208.36 Kbp upstream | 0.98 |
| | | 00001340174 | 17:17170871-17171730 | 44207.5 Kbp upstream | 0.98 |
| | *COPS3* | 00001340203 | 17:17280690-17282763 | 44096.47 Kbp upstream | 0.98 |
| | | 17_C3PB4 | 17:17280690-17281961 | 44097.27 Kbp upstream | 0.98 |
| | *RNF213* | 00001345123 | 17:80259362-80260880 | 18880.13 Kbp downstream | 0.98 |
| | | 17_BKX3R | 17:80259362-80260880 | 18880.13 Kbp downstream | 0.98 |
| | *CCDC47* | 00001343782 | 17:63761527-63762800 | 2382.3 Kbp downstream | 0.98 |
| | *VTN* | 00001340749 | 17:28370297-28371000 | 33008.23 Kbp upstream | 0.97 |
| | | 17_9F933 | 17:28370297-28371000 | 33008.23 Kbp upstream | 0.97 |
| | *USP36* | 00001344993 | 17:78799712-78800600 | 17420.48 Kbp downstream | 0.97 |
| | *PSMC5* | 00001343800 | 17:63831001-63831349 | 2451.77 Kbp downstream | 0.97 |
| | *CBX8* | 00001345059 | 17:79797067-79798320 | 18417.84 Kbp downstream | 0.97 |
| | | 17_936N6H | 17:79797067-79798567 | 18417.84 Kbp downstream | 0.97 |
| | *LGALS9B* | 00001340559 | 17:20467525-20469029 | 40910.2 Kbp upstream | 0.97 |
| | *C1QTNF1-AS1* | 00001014023 | 17:79027610-79028999 | 17648.38 Kbp downstream | 0.97 |
| | | 17_BKNGF | 17:79027610-79028910 | 17648.38 Kbp downstream | 0.97 |
| | *PDE6G* | 00001345301 | 17:81656532-81657799 | 20277.3 Kbp downstream | 0.97 |
| | *SPAG5-AS1* | 00001340786 | 17:28597300-28598800 | 32780.43 Kbp upstream | 0.97 |
| | *C1QTNF1* | 00001345037 | 17:79032695-79034195 | 17653.47 Kbp downstream | 0.97 |

| Cell line | Gene | ENSR ID | Coordinates (hg38) | Distance from rs9895661 | MPPC score |
|---|---|---|---|---|---|
| | *MED1* | 00001341664 | 17:39451253-39452762 | 21926.47 Kbp upstream | 0.96 |
| | *TEKT3* | 00001340020 | 17:15306991-15307599 | 46071.63 Kbp upstream | 0.96 |
| | *CTNS* | 00000547489 | 17:3638657-3640157 | 57739.07 Kbp upstream | 0.96 |
| | *PHF12* | 00001340872 | 17:28928280-28929780 | 32449.45 Kbp upstream | 0.96 |
| | *C17orf99* | 00001344910 | 17:78160075-78161050 | 16780.85 Kbp downstream | 0.95 |
| | | 00001344906 | 17:78144895-78146864 | 16765.67 Kbp downstream | 0.94 |
| | *ST6GALNAC2* | 00001344720 | 17:76585846-76586945 | 15206.62 Kbp downstream | 0.95 |
| | | 00001344721 | 17:76586946-76587399 | 15207.72 Kbp downstream | 0.95 |
| | *NOS2* | 00001340700 | 17:27793657-27795270 | 33583.96 Kbp upstream | 0.95 |
| | | 17_B7HX5 | 17:27793684-27794367 | 33584.86 Kbp upstream | 0.95 |
| | *THRA* | 00001341747 | 17:40057802-40058300 | 21320.93 Kbp upstream | 0.81 |
| | *ICAM2* | 00001343834 | 17:64020541-64022124 | 2641.31 Kbp downstream | 0.79 |
| | | 17_BGP39 | 17:64020597-64021434 | 2641.37 Kbp downstream | 0.79 |
| | *IGFBP4* | 00001341816 | 17:40441960-40443460 | 20935.77 Kbp upstream | 0.48 |
| | | 17_9L3FB | 17:40442512-40443460 | 20935.77 Kbp upstream | 0.48 |
| | *TLK2* | 00001343670 | 17:62476602-62478076 | 1097.37 Kbp downstream | 0.37 |
| | | 17_9WRK2 | 17:62477146-62477849 | 1097.92 Kbp downstream | 0.37 |
| | *PPM1D* | 00001343603 | 17:60598703-60600669 | 778.56 Kbp upstream | 0.32 |
| | | 17_BG2PC | 17:60599117-60600425 | 778.8 Kbp upstream | 0.32 |
| | *YPEL2* | 00001343513 | 17:59387350-59388850 | 1990.38 Kbp upstream | 0.19 |
| | *CHCT1* | 00001343592 | 17:60420402-60421752 | 957.48 Kbp upstream | 0.14 |
| | *AURKB* | 00001339774 | 17:8208215-8209715 | 53169.51 Kbp upstream | 0.12 |
| | *TBX2-AS1* | 00001343635 | 17:61395437-61396463 | 16.21 Kbp downstream | 0.07 |
| | | 00001343634 | 17:61394052-61395436 | 14.82 Kbp downstream | 0.06 |
| | | 17_BG6ZL | 17:61396303-61396919 | 17.08 Kbp downstream | 0.06 |
| | | 17_9WCZS | 17:61398507-61399333 | 19.28 Kbp downstream | 0.05 |
| | *BCAS3* | 00001343627 | 17:61157872-61159372 | 219.86 Kbp upstream | 0.06 |
| | | 00001343625 | 17:61155966-61157466 | 221.76 Kbp upstream | 0.05 |
| | | 00001343629 | 17:61245001-61245270 | 133.96 Kbp upstream | 0.05 |
| | *TBX4* | 00001343648 | 17:61454956-61456474 | 75.73 Kbp downstream | 0.06 |
| | | 17_BG7B8 | 17:61454956-61456474 | 75.73 Kbp downstream | 0.06 |
| | | 00001343646 | 17:61451402-61452432 | 72.17 Kbp downstream | 0.06 |
| | | 17_9WDKB | 17:61451300-61452432 | 72.07 Kbp downstream | 0.06 |

For each cell type, results are arranged in descending order of MPPC score.
ENSR ID=Ensembl stable ID for regulatory features; HEK293=Human embryonic kidney 293 cells; HK2=Human kidney 2 immortalised proximal tubule epithelial cells; MPPC=Marginal posterior probability of contact; RPTEC=Renal proximal tubule epithelial cells.

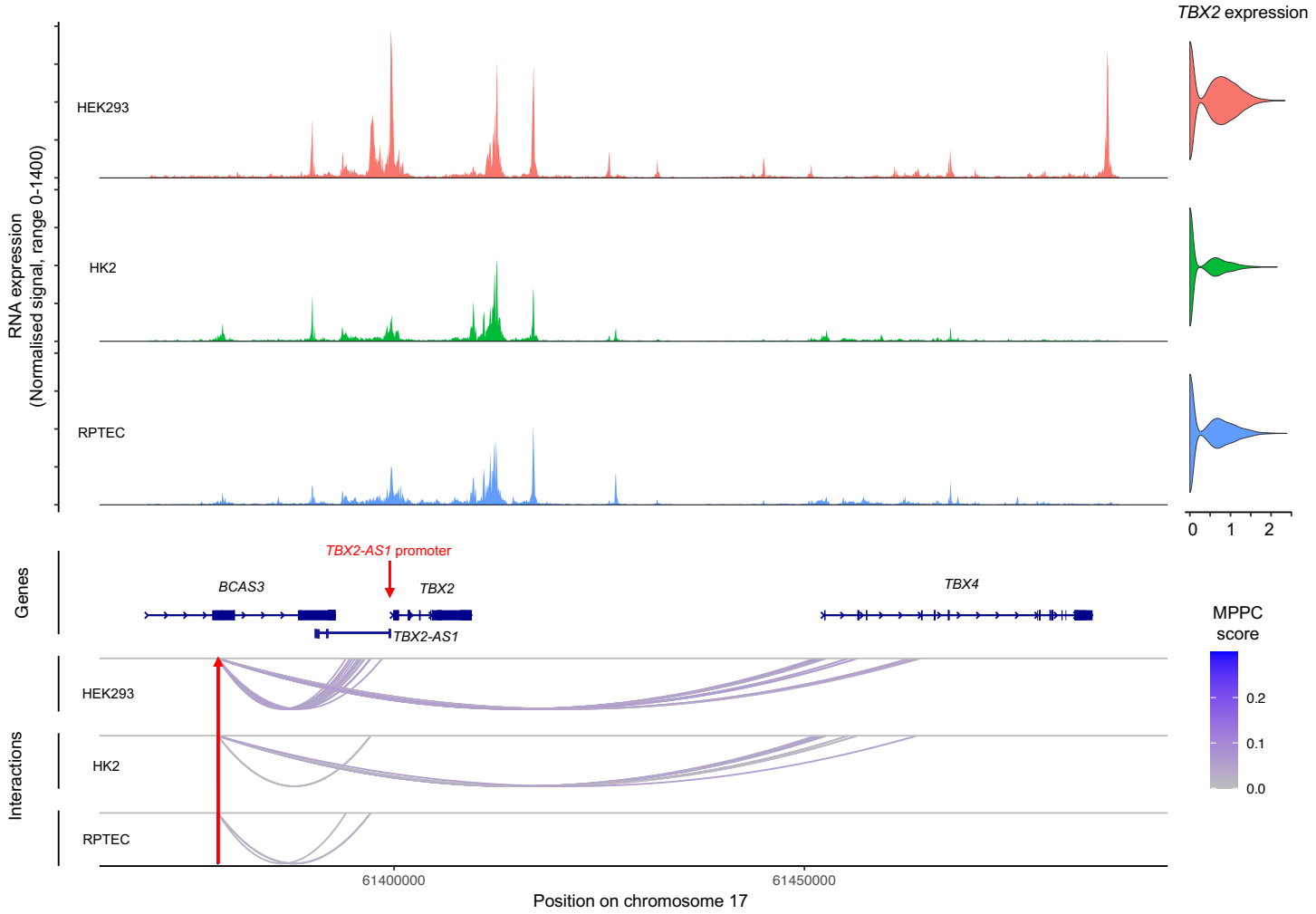



***Figure 6.9. CaptureC- chromatin interactions of rs9895661 and gene promoters across three renal cell lines.***

**Top**: Normalised RNA expression across renal cell types. Peaks represent higher expression. **Bottom**: Links represent significant chromatin interactions between rs9895661 and promoter coordinates of genes, defined using Ensembl v113 [466]. Colours of links reflect the magnitude of the marginal posterior probability of contact. The vertical red arrow denotes the location of rs9895661. **Violin plot**: Expression of *TBX2* in each renal cell type.
HEK293=Human embryonic kidney cells; HK2=Human kidney immortalised proximal tubule epithelial cells; MPPC=Marginal posterior probability of contact; RPTEC=Renal proximal tubule epithelial cells. Coordinates are reported in genome build hg38.

*TBX2-AS1* is a positionally conserved long non-coding RNA (pcRNA) [643]. pcRNAs preferentially localise at the boundaries of topologically associating domains (TADs) and chromatin loops, where they regulate gene expression by facilitating interactions between regulatory elements and their target genes. *TBX2-AS1* maps to the border of a TAD and functions as a bi-directionally transcribed topographical anchor point, forming chromatin loops that mediate transcriptional regulation [643, 644].

Co-expression analyses suggest that *TBX2-AS1* regulates *TBX2* in *cis* [645]. *TBX2* encodes a T-box transcription factor (TBX2) that plays a critical role in embryogenesis and the morphogenesis of limb, cardiac, renal, neural, and optic tissues [646–649]. It also promotes cell proliferation and epithelial-to-mesenchymal transition [650–652], processes frequently dysregulated in malignancies, including melanoma, breast, and pancreatic cancers [653–655], and implicated in the pathogenesis of polycystic kidney disease [656–659]. TBX2 primarily acts as a transcriptional repressor, targeting key genes such as *E-cadherin* and *β-catenin* [652]. Notably, the E-cadherin/*β*-catenin signalling pathway interacts with the actin cytoskeleton to regulate cell morphology, junction stability, adhesion, migration, and tissue remodelling [660, 661], and has been implicated in polycystic kidney disease.

It is plausible that rs9895661 interacts with the *TBX2-AS1* promoter, modulating *TBX2* expression and thereby altering E-cadherin/*β*-catenin signalling to promote epithelial-to-mesenchymal transition. This mechanism could increase susceptibility to renal cyst formation. Renal cyst epithelium exhibits loss of cell polarity and abnormal localisation of essential transporters, including $Na^+$-$K^+$-ATPase, which disrupts salt and fluid handling [662]. These abnormalities may alter renal tubular fluid composition, potentially explaining the increased risk of kidney stone formation in individuals with renal cysts [663].

## 6.4 Discussion and conclusions

In this chapter of my thesis, I sought to evaluate the relationship between non-polycystic renal cysts and KSD to identify shared genetic mechanisms underlying their pathophysiology. Through a combination of observational and

genetic epidemiological analyses and functional genomics, I provide evidence that not only are renal cysts associated with an increased risk of KSD, but also that there is a bidirectional, causal relationship between these traits. Furthermore, my findings suggest a shared genetic architecture at the *BCAS3* locus, implicating *TBX2-AS1* in the pathogenesis of both renal cysts and KSD. This provides a foundation for a novel hypothesis regarding the molecular mechanisms that may underlie the development of these two traits. By integrating a variety of statistical and scientific approaches, I identify the candidate causal variant rs9895661 and present a potential mechanism for its involvement in renal cyst formation and KSD.

### 6.4.1 The bidirectional, causal relationship between renal cysts and kidney stone disease

The link between renal cysts and KSD observed in this study is supported by both observational analyses and MR studies, which suggest a bidirectional causal relationship between these traits. My findings extend previous work by providing evidence that cysts may contribute to the development of KSD. In line with existing literature, we know that renal tubular microcrystal formation can trigger cystogenesis, likely as a protective response to crystal-induced obstruction [130]. This creates a feedback loop whereby KSD can exacerbate cyst formation. MR data in this chapter support the hypothesis that cysts have a causal influence on KSD, which has not been demonstrated in previous studies.

Renal cysts require both increased epithelial cell proliferation and altered trans-tubular fluid transport to develop and grow [664–667]. Cyst formation depends on a net influx of fluid secreted by the renal tubule to prevent cyst wall collapse, thus transforming the normal absorptive epithelium into a secretory phenotype [668]. This secretory transformation, often accompanied by dysregulated $Na^+K^+$-ATPase activity, further supports the idea that cysts alter renal fluid dynamics, creating an environment conducive to crystal precipitation, aggregation, and stone formation.

Moreover, disruptions to primary cilia are a common feature of renal cysts [496, 663]. Primary cilia play a central role in planar polarity, may initiate renal cyst

formation and contribute to altered fluid dynamics [663]. The loss or shortening of cilia leads to impaired orientation and movement, which in turn affects the integrity of the epithelial monolayer and its ability to regulate fluid transport and signalling. These alterations in epithelial function may lead to changes in tubular fluid composition that increase the risk of stone formation. An additional potential mechanism for the bidirectional relationship between cysts and KSD is altered tubular fluid flow. The mechanical disruption of normal kidney architecture due to cyst growth may exacerbate an individual's risk of KSD by creating an environment conducive to stone formation. As urine flows through the renal tubule, microcrystals can accumulate and cause obstructions, which may promote cyst formation. Conversely, cystic structures themselves could contribute to regions of high resistance, favouring the formation of kidney stones due to an accumulation of crystals, promoting crystal aggregation, in stagnant tubular fluid [669].

### 6.4.2 The role of *TBX2-AS1* in renal cyst formation

One of the key findings in this chapter comes from Capture-C data that reveal chromatin interactions between the rs9895661 region and the promoter of *TBX2-AS1*, a gene co-expressed with *TBX2*, which lies ≈3 Kbp downstream [645]. *TBX2* is expressed during both kidney organogenesis and in adulthood, and variants of this gene are associated with renal developmental abnormalities and medullary nephrocalcinosis [670]. Functional studies in *Xenopus* models demonstrate that Tbx2 plays a crucial role in the development of kidney structures. Specifically, loss of Tbx2 function results in an expanded boundary of each nephron component, leading to an enlarged pronephric duct [670]. These findings support the critical role of *TBX2* in the structural development of genitourinary organs, including the kidney, and suggests that alterations in its expression may contribute to renal pathology. Moreover, *TBX2* is a transcriptional repressor known to promote malignant transformation by repressing the transcription of tumour suppressor genes, which may lead to senescence bypass and cellular immortalisation and account for its association with melanoma, breast, and pancreatic cancers [653–655]. Additionally, it suppresses key proteins such as E-cadherin and $\beta$-catenin, which

are vital components of adherens junctions in epithelial cells. This is significant because E-cadherin is critical for cell-cell adhesion and maintaining planar polarity, both of which are essential for the proper function of renal tubular epithelium. The cytoplasmic domain of E-cadherin complexes with $\beta$-catenin, which links to the actin cytoskeleton and supports the maturation of adherens junctions, and regulates cell morphology, junction stability, cell-cell adhesion, cell migration, and tissue remodelling [671–674].

E-cadherin-$\beta$-catenin signalling is indispensable for proper nephron formation and kidney development [675–677]. Although $\beta$-catenin signalling is generally inactive in the adult kidney, it is reactivated in various renal pathologies, including fibrosis, obstructive nephropathy, diabetic nephropathy, and polycystic kidney disease [678–680]. As such, altered expression of *TBX2* and *TBX2-AS1*, as suggested by the rs9895661-T risk allele, could disrupt epithelial polarity and cell-cell adhesion, promoting cyst formation.

The evidence I present suggests that rs9895661 regulates the co-expression of *TBX2-AS1* and *TBX2*, and that lower expression of these genes likely results in abnormal epithelial function and cell-cell adhesion. This hypothesis is further supported by the known role of *TBX2* in regulating epithelial morphology and the potential disruption of these pathways in cystic renal epithelium. It is crucial, therefore, to investigate the effects of *TBX2* knockdown in renal cells on cyst formation, and to explore potential therapeutic avenues based on this hypothesis.

### 6.4.3 The clinical relevance of a shared genetic architecture between renal cysts and kidney stone disease

Findings from this chapter have important implications for diagnosing, monitoring, and treating patients with renal cysts and KSD. Current international guidelines classify cysts using the Bosniak grading system and do not recommend routine follow-up for simple (grade I or II) cysts [681–683]. However, with evidence that renal cysts causally increase the risk of KSD, clinicians should assess whether patients with incidentally identified cysts have serum or urinary biochemical abnormalities

that could further promote stone formation. Addressing these abnormalities could help prevent patients from experiencing pain, infection, and the need for surgical intervention.

There may be an argument to advise patients with renal cysts about their heightened risk of KSD and recommend lifestyle modifications, such as maintaining adequate hydration and adopting dietary changes, to lower their risk. My findings also emphasise the need to investigate whether personalised therapeutic strategies could help patients with both renal cysts and KSD by inhibiting cyst formation, for example with vasopressin receptor antagonists like tolvaptan. However, functional studies are required to evaluate the potential for targeted therapeutic approaches for patients with cysts and stones before these treatments can be recommended.

### 6.4.4 Limitations

Despite the novel insights presented in this chapter, I am aware of several limitations. Observational analyses were limited to participants with abdominal imaging to reduce confounding by indication. Patients with renal cysts may be more likely to undergo surveillance imaging, which may detect potentially clinically insignificant stones that may remain undetected in the general population. Similarly, patients with symptomatic renal stones are more likely to undergo cross-sectional imaging which may identify asymptomatic renal cysts that would otherwise go unnoticed in the general population. Despite my efforts to account for these factors, residual confounding remains a concern as lack of temporal data for imaging and diagnoses could influence the validity of the observed associations. Additionally, the lack of detailed data precluded validating whether stones identified through imaging were indeed stones, rather than calcifications within cyst walls. One could also argue that the observational data do not confirm that cysts and stones were present in the same kidney. In this chapter, I describe a causal genetic mechanism underlying a shared pathogenesis between these two traits. If inherited in a germline manner, the pathogenicity would affect both kidneys equally, suggesting that laterality may not be relevant. However, local phenotypic factors, such as the presence of tubular microcrystals, may also influence the risk of cysts and stones, potentially making

laterality an important consideration.

While the cyst GWAS data are valuable, they suffer from case-control imbalance and a limited sample size which may reduce the power to detect significant associations, particularly outside of European populations. With greater case ascertainment, it is plausible that variants in *PKD2*, *AQP1*, or *UMOD* may be identified and point to a shared genetic architecture between cysts and stones as alluded to in findings from gene and gene set analyses in Chapter 3. MR analyses are constrained by the small number of SNPs available as IVs, limiting the strength of causal inference. Similarly, Capture-C analyses are limited by my focus on gene promoters and I have not considered important chromatin interactions with other gene regulatory features, such as enhancers or transcription factor binding sites, that may underlie the shared pathogenesis of renal cysts and KSD. Despite these limitations, the findings presented here are thought-provoking and provide a strong rationale for further investigation into the genetic and molecular mechanisms underlying the relationship between renal cysts and KSD.

### 6.4.5 Conclusions

In conclusion, in this chapter I demonstrate that renal cysts and KSD are causally linked, with bidirectional effects between the two conditions. The shared genetic architecture at the *BCAS3* locus implicates *TBX2-AS1* and *TBX2* in the pathogenesis of both conditions. Altered tubule epithelial function, including disrupted transporter localisation and fluid handling, as well as changes in tubular fluid flow, may contribute to the increased risk of KSD in individuals with renal cysts. Given the evidence presented, I propose that *TBX2* and its associated pathways may play an important role in renal cyst formation and the development of KSD. These findings offer promising avenues for future research and may inform the development of therapeutic strategies targeting these pathways to mitigate the risk of KSD in patients with renal cysts.

# 7 | Discussion

## 7.1 Summary

The overarching goal of my thesis was to employ a range of genetic epidemiological and bioinformatic approaches, each with distinct sources of bias, to construct a robust body of evidence that deepens our understanding of the genetic and molecular basis of KSD. To this effect, I structured my work around the following research aims:

1. To determine the genetic architecture of KSD by performing GWAS and meta-analyses in European and trans-ancestral populations and ExWAS in a European population (Chapter 3).

2. To establish the effect of adiposity on the risk of KSD and identify potential disease-mediating genes to guide future therapeutic approaches (Chapter 4).

3. To investigate the impact of alterations in mineral homeostasis on risk of KSD and uncover genetic and molecular pathways that may influence KSD and offer therapeutic potential (Chapter 5).

4. To evaluate the relationship between renal cysts and risk of KSD to identify shared genetic and molecular mechanisms that may underpin a common pathophysiology (Chapter 6).

In Chapter 3, I established the foundations for the studies in later chapters. By conducting genetic association studies, integrating combined- and sex-specific analyses in European and trans-ancestry populations while examining both common and rare genetic variants, I identified novel loci linked to KSD. To my knowledge, these studies represent the largest genome- and exome-wide association analyses of KSD to date, demonstrating how increasing sample sizes reveal additional genetic associations. My subsequent gene and gene set analyses confirmed the importance of

mineral metabolism traits and suggested that renal structural disorders contribute to KSD pathogenesis. These findings provided the basis for the following chapters, which aimed to uncover the causal mechanisms underlying KSD.

In Chapter 4, I moved beyond association studies to investigate specific genetic mechanisms linking increased adiposity to a higher risk of KSD. Using a novel approach to MR and colocalisation analyses, I identified a shared genetic architecture between adiposity traits and KSD and reported that a functional, causal, missense variant in *GIPR* likely underlies this relationship. Based on existing evidence, I hypothesise that renal tubular osteopontin expression and secretion may mediate these effects. Moreover, I demonstrated the clinical relevance of these findings by highlighting the therapeutic potential of targeting GIPR-related pathways to reduce KSD risk. My results provide a strong rationale for further research to characterise these mechanisms, identify reliable biomarkers, design clinical trials, and ultimately translate these findings into effective therapeutic strategies for patients with KSD.

In Chapter 5, I examined the role of albumin-adjusted serum calcium and phosphate concentrations in KSD pathogenesis. By integrating MR and colocalisation analyses with multiomic and chromatin interaction data, I identified three biological pathways implicated in ≈12–17% of KSD cases. My findings highlight the role of DGK$\delta$ -mediated impaired CaSR signal transduction, NaPi-IIa-mediated increased renal phosphate excretion, and 24-$\alpha$-hydroxylase-mediated impaired inactivation of 1,25-dihydroxyvitamin D in the pathophysiology of KSD. These results not only enhance our understanding of the metabolic basis of KSD but also suggest potential therapeutic strategies to prevent and treat disease using personalised medicine.

In Chapter 6, I built on the novel gene set associations identified in Chapter 3 and provided the first, to my knowledge, evidence of a causal, bidirectional relationship between non-polycystic renal cysts and KSD. By leveraging snRNA-seq, snATAC-seq, and Capture-C data, I postulated that *TBX2*-related pathways provide a shared genetic architecture underlying both conditions. These findings

have direct clinical implications: clinicians should counsel patients with renal cysts on their increased risk of KSD, researchers should investigate the renal tubular effects of *TBX2* knockdown on cyst and stone formation, and future studies should explore therapeutic avenues targeting this pathway.

## 7.2 The relevance of results from this thesis

Rudolf Virchow's eponymous triad describes three factors contributing to venous thrombosis: hypercoagulability, stasis, and endothelial injury [684]. Inspired by this, I propose a lithogenic triad describing three key contributors to kidney stone formation: urine supersaturation, urine stasis, and tubular epithelial insult (Figure 7.1). These factors are interconnected: abnormal tubular epithelium alters urine composition and flow, urine stasis promotes crystal formation and epithelial injury, and urine supersaturation drives crystallisation and changes in tubular function. Together, they predispose individuals to KSD.

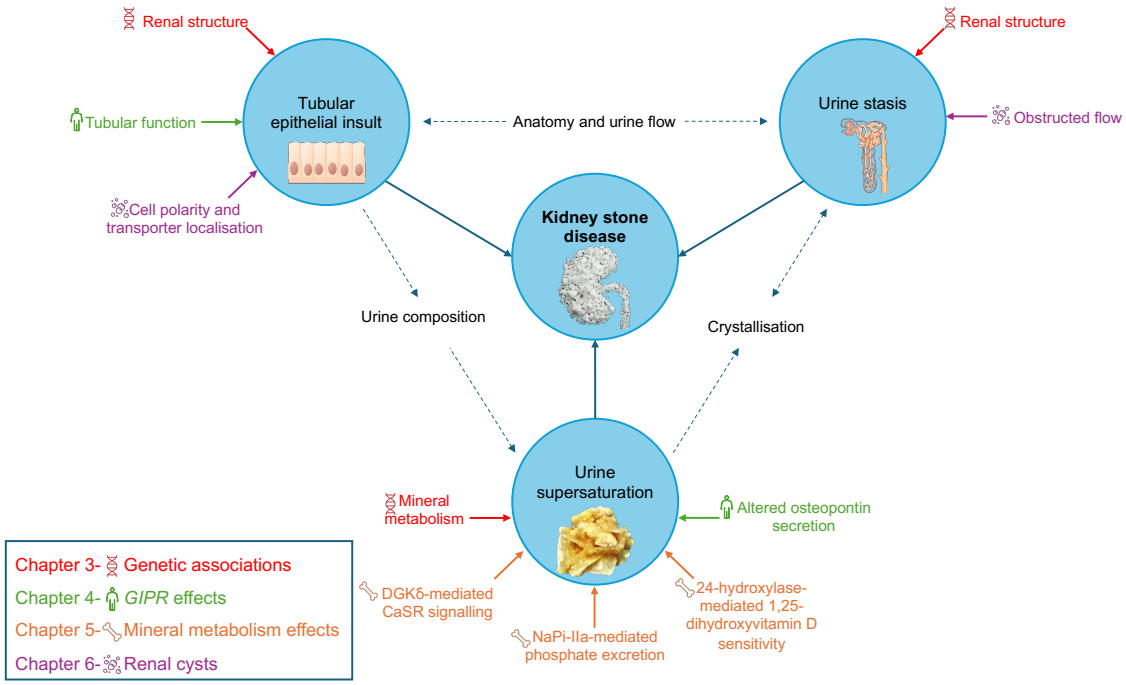



***Figure 7.1. A lithogenic triad of factors that increase the risk of kidney stone disease.***

CaSR=Calcium-sensing receptor; DGK$\delta$=Diacylglycerol kinase delta; GIPR=Gastric inhibitory polypeptide receptor; NaPi-IIa=Renal sodium-phosphate cotransporter 2A

My findings contribute to this hypothesis in several ways:

- The genetic associations with mineral metabolism and renal structural pathways, identified in Chapter 3, likely influence urine supersaturation, stasis, and epithelial function.
- The role of *GIPR* signalling, identified in Chapter 4, may mediate osteopontin secretion, affect tubular epithelial function, and promote urine supersaturation.
- The disturbances in calcium, phosphate, and vitamin D metabolism, uncovered in Chapter 5, likely contribute to urine supersaturation.
- Renal cysts, as discussed in Chapter 6, present an abnormal tubular epithelium and increase urine stasis, further promoting stone formation.

Taken together, these findings provide a new conceptual framework for understanding KSD pathogenesis. By recognising this lithogenic triad, we can shift research and therapeutic strategies towards targeting not just urine chemistry but also tubular function and urine flow dynamics.

## 7.3 Key limitations

Despite advancing our understanding of KSD genetics and pathophysiology, I recognise several limitations in my work.

### 7.3.1 Ancestral representation

Apart from the trans-ancestry GWAS meta-analysis in Chapter 3, my studies comprise data from European ancestry populations. This Eurocentric bias, driven by the limited availability of well-powered GWAS in non-European ancestry populations, restricts the generalisability of my findings [299]. My conducting studies in European cohorts, risks perpetuating this imbalance in genetic research.

### 7.3.2 Potential bias in MR and colocalisation studies

MR and colocalisation analyses rely on the accuracy and robustness of underlying GWAS data. Inconsistent results from analyses with UK Biobank KSD GWAS, FinnGen KSD GWAS, and European-ancestry KSD GWAS meta-analysis in Chapter 6, where I examined the effects of renal cysts on KSD risk, highlight the potential for bias. To mitigate these biases, I used the most robust datasets available, selected instruments strongly associated with each trait, and applied stringent statistical corrections. Nevertheless, both type I and type II errors remain possible.

### 7.3.3 Tissue specificity

The snRNA-seq, snATAC-seq, and Capture-C analyses that I performed focus exclusively on renal cell types. Consequently, I may have overlooked relevant genetic effects in other tissues which also play key roles in KSD pathogenesis such as the parathyroid, bone, or intestine.

To address the above limitations and to strengthen the reliability of my findings, I have adopted a multifaceted approach to triangulate evidence using multiple analytical approaches, including observational associations, genetic epidemiology, and chromatin studies, each with distinct and orthogonal sources of bias.

## 7.4 Future direction

Several important research avenues emerge from my findings. Future studies should investigate the functional impact of the genetic loci that I have identified. This includes characterising candidate genes and causal variants through functional genomic approaches, elucidating their molecular effects in renal and extrarenal tissues, and determining how they contribute to the pathophysiology of KSD. My analyses do not yet incorporate data on stone composition or urinary metabolic

and proteomic profiles. Future studies should integrate these data to consider their relevance to refining our understanding of disease mechanisms in KSD.

In future work, to translate my work from genetic discoveries to clinical applications, I aim to secure ethical approval to collect patient samples. I aim to expand research collaborations to enable me to incorporate *in vitro* and *in vivo* functional studies and to explore the clinical potential of the novel therapeutic strategies proposed in this thesis. Moreover, I am driven to clarify the role of GIP- and GLP-1-mediated pathways through well-designed cohort studies to support an RCT investigating their utility in modulating KSD risk. Given that KSD is a complex disease with heterogeneous causes across the population, a bespoke, targeted, and personalised approach is needed to effectively treat and prevent the disease in each patient. My findings suggest that new therapeutic approaches, such as calcimimetics, phosphate supplements, activated vitamin D antagonism, and targeting *GIPR* and *TBX2*-related pathways may be leveraged for these purposes. Future research should focus on preclinical validation and early-phase clinical trials to determine whether modulating these pathways can reduce KSD risk and facilitate advances in personalised medicine.

## 7.5 Conclusions

In conclusion, through a combination of conventional and genetic epidemiology, and functional genomic analyses, I have significantly advanced our understanding of the genetic architecture underlying KSD. By integrating a range of scientific approaches, I have identified causal effects of increased adiposity, altered calcium and phosphate homeostasis, and renal cysts on the risk of KSD and localised these effects to candidate genes and variants. These discoveries provide valuable insights into disease mechanisms and potential therapeutic targets.

Crucially, my work builds on traditional perspectives on KSD by proposing a lithogenic triad that accounts for urine supersaturation, stasis, and tubular epithelial dysfunction. This conceptual shift lays the foundation for future research and clinical strategies that move beyond urinary chemistry alone and incorporate a broader

understanding of renal physiology. By addressing key gaps in knowledge, my research sets the stage for future investigations that will refine these insights, translate them into patient care, and ultimately improve outcomes for individuals affected by this common, and frequently recurrent, condition.

# References


[1] Saeed R. Khan et al. "Kidney stones". In: *Nature reviews. Disease primers* 2 (Feb. 2016), p. 16008. DOI: 10.1038/nrdp.2016.8. (Visited on 10/13/2022).

[2] J. Shah and H.N. Whitfield. "Urolithiasis through the ages: UROLITHIASIS THROUGH THE AGES". en. In: *BJU International* 89.8 (Apr. 2002), pp. 801–810. DOI: 10.1046/j.1464-410X.2002.02769.x. (Visited on 10/13/2022).

[3] Hippocrates. "Complete works of Hippocrates". In: vol. VII. Paris, 1851, pp. 202–211.

[4] Athanasios Diamandopoulos, Pavlos Goudas, and Dimitrios Oreopoulos. "Thirty-six Hippocratic aphorisms of nephrologic interest." eng. In: *American journal of kidney diseases : the official journal of the National Kidney Foundation* 54.1 (July 2009). Place: United States, pp. 143–153. DOI: 10.1053/j.ajkd.2009.01.275.

[5] Mohammad Abufaraj et al. "Association Between Body Fat Mass and Kidney Stones in US Adults: Analysis of the National Health and Nutrition Examination Survey 2011–2018". en. In: *European Urology Focus* 8.2 (Mar. 2022), pp. 580–587. DOI: 10.1016/j.euf.2021.03.010. (Visited on 08/15/2023).

[6] Margaret S. Pearle et al. "Medical Management of Kidney Stones: AUA Guideline". In: *Journal of Urology* 192.2 (Aug. 2014). Publisher: WoltersKluwer, pp. 316–324. DOI: 10.1016/j.juro.2014.05.006. (Visited on 08/07/2023).

[7] Robert M. Geraghty et al. "Evaluation of the economic burden of kidney stone disease in the UK: a retrospective cohort study with a mean follow-up of 19 years". In: *BJU International* 125.4 (Apr. 2020). Publisher: John Wiley & Sons, Ltd, pp. 586–594. DOI: 10.1111/bju.14991. (Visited on 10/13/2022).

[8] J. Michael Soucie et al. “Demographic and geographic variability of kidney stones in the United States”. en. In: *Kidney International* 46.3 (Sept. 1994), pp. 893–899. DOI: 10.1038/ki.1994.347. (Visited on 10/14/2022).

[9] Mohammad Abufaraj et al. “Prevalence and Trends in Kidney Stone Among Adults in the USA: Analyses of National Health and Nutrition Examination Survey 2007–2018 Data”. en. In: *European Urology Focus* 7.6 (Nov. 2021), pp. 1468–1475. DOI: 10.1016/j.euf.2020.08.011. (Visited on 06/05/2024).

[10] Kyriaki Stamatelou and David S. Goldfarb. “Epidemiology of Kidney Stones”. en. In: *Healthcare* 11.3 (Feb. 2023), p. 424. DOI: 10.3390/healthcare11030424. (Visited on 08/15/2023).

[11] Whalley et al. “Lithogenic risk factors in normal black volunteers, and black and white recurrent stone formers”. en. In: *BJU International* 84.3 (1999). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1464-410x.1999.00170.x, pp. 243–248. DOI: 10.1046/j.1464-410x.1999.00170.x. (Visited on 10/14/2022).

[12] Sonja Lewandowski et al. “Oxalate and its handling in a low stone risk vs a stone-prone population group”. en. In: *World Journal of Urology* 23.5 (Nov. 2005), pp. 330–333. DOI: 10.1007/s00345-005-0030-6. (Visited on 10/14/2022).

[13] Allen L. Rodgers. “Race, ethnicity and urolithiasis: a critical review”. en. In: *Urolithiasis* 41.2 (Apr. 2013), pp. 99–103. DOI: 10.1007/s00240-012-0516-9. (Visited on 10/14/2022).

[14] Eric N. Taylor and Gary C. Curhan. “Differences in 24-Hour Urine Composition between Black and White Women”. en. In: *Journal of the American Society of Nephrology* 18.2 (Feb. 2007), pp. 654–659. DOI: 10.1681/ASN.2006080854. (Visited on 10/14/2022).

[15] Kymora B. Scotland et al. “Social Determinants of Kidney Stone Disease: The Impact of Race, Income and Access on Urolithiasis Treatment and Outcomes”. eng. In: *Urology* 163 (May 2022), pp. 190–195. DOI: 10.1016/j.urology.2021.08.037.

[16] Saeed Khan R. “Modulators of urinary stone formation”. en. In: *Frontiers in Bioscience* 9.1-3 (2004), p. 1450. DOI: 10.2741/1347. (Visited on 10/14/2022).

[17] Rosemary L. Ryall, Magali C. Chauvet, and Phulwinder K. Grover. “Intracrystalline proteins and urolithiasis: a comparison of the protein content and ultrastructure of urinary calcium oxalate monohydrate and dihydrate crystals”. en. In: *BJU International* 96.4 (Sept. 2005), pp. 654–663. DOI: 10.1111/j.1464-410X.2005.05701.x. (Visited on 10/04/2024).

[18] Rosemary Lyons Ryall. “Macromolecules and Urolithiasis: Parallels and Paradoxes”. en. In: *Nephron Physiology* 98.2 (Oct. 2004), p37–p42. DOI: 10.1159/000080262. (Visited on 10/04/2024).

[19] M. D. McKee, A. Nancl, and S. R. Khan. “Ultrastructural immunodetection of osteopontin and osteocalcin as major matrix components of renal calculi”. en. In: *Journal of Bone and Mineral Research* 10.12 (Dec. 1995), pp. 1913–1929. DOI: 10.1002/jbmr.5650101211. (Visited on 10/04/2024).

[20] Lance Wheeler. *Cystine Crystals in Canine Urine Sediment*. Aug. 2018. URL: https://commons.wikimedia.org/wiki/File:Cystine_Crystals_in_Canine_Urine_Sediment.jpg (visited on 02/06/2025).

[21] R Lehmann. *Formed elements of urine*. Jan. 2021. URL: https://commons.wikimedia.org/wiki/File:Formed_elements_of_urine.png (visited on 02/06/2025).

[22] Ajay Kumar Chaurasiya. *Calcium oxalate crystals in urine microscopy*. Sept. 2017. URL: https://commons.wikimedia.org/wiki/File:Calcium_oxalate_crystals_in_Urine_Microscopy.jpg (visited on 02/06/2025).

[23] Doruk Salancı. *Calcium oxalate crystals (urine)*. Aug. 2012. URL: https://commons.wikimedia.org/w/index.php?curid=20865307#file (visited on 02/06/2025).

[24] Doruk Salancı. *Uric acid crystals (urine)*. Jan. 2014. URL: https://commons.wikimedia.org/wiki/File:Uric_acid_crystals_(urine)_-_%C3%9Crik_asit_kristalleri_(idrar)_-_03.png (visited on 02/06/2025).

[25] Doruk Salancı. *Struvite crystals (urine)*. Aug. 2012. URL: https://commons.wikimedia.org/wiki/File:Struvite_crystals_(urine)_-_Str%C3%BCvit_kristalleri_(idrar)_-_01.png (visited on 02/06/2025).

[26] Roswitha Siener, Linda Netzer, and Albrecht Hesse. "Determinants of Brushite Stone Formation: A Case-Control Study". en. In: *PLoS ONE* 8.11 (Nov. 2013). Ed. by Jeff M. Sands, e78996. DOI: 10.1371/journal.pone.0078996. (Visited on 10/04/2024).

[27] A-Jin Lee et al. "Differential identification of urine crystals with morphologic characteristics and solubility test". In: *Journal of Clinical Laboratory Analysis* 36.11 (Sept. 2022), e24707. DOI: 10.1002/jcla.24707. (Visited on 03/03/2025).

[28] Gheun-Ho Kim and Jae-Bum Jun. "Altered Serum Uric Acid Levels in Kidney Disorders". In: *Life* 12.11 (Nov. 2022), p. 1891. DOI: 10.3390/life12111891. (Visited on 02/21/2024).

[29] Katie S. Eyre et al. "Utility of blood tests in screening for metabolic disorders in kidney stone disease". en. In: *BJU International* 127.5 (2021). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/bju.15250, pp. 538–543. DOI: 10.1111/bju.15250. (Visited on 10/14/2022).

[30] "NICE Guideline – Renal and ureteric stones: assessment and management". en. In: *BJU International* 123.2 (2019). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/bju.14654, pp. 220–232. DOI: 10.1111/bju.14654. (Visited on 09/13/2024).

[31] Andreas Skolarikos et al. "Metabolic Evaluation and Recurrence Prevention for Urinary Stone Patients: EAU Guidelines". en. In: *European Urology* 67.4 (Apr. 2015), pp. 750–763. DOI: 10.1016/j.eururo.2014.10.029. (Visited on 01/07/2025).

[32] Fredric L. Coe, Andrew Evan, and Elaine Worcester. "Kidney stone disease". eng. In: *The Journal of Clinical Investigation* 115.10 (Oct. 2005), pp. 2598–2608. DOI: 10.1172/JCI26662.

[33] Kay Thomas et al. "Cystinuria-a urologist's perspective". eng. In: *Nature Reviews. Urology* 11.5 (May 2014), pp. 270–277. DOI: 10.1038/nrurol.2014.51.

[34] Josep Chillarón et al. “Pathophysiology and treatment of cystinuria”. en. In: *Nature Reviews Nephrology* 6.7 (July 2010). Number: 7 Publisher: Nature Publishing Group, pp. 424–434. DOI: 10.1038/nrneph.2010.69. (Visited on 02/19/2024).

[35] Maria C. Martins et al. “Cystine: a promoter of the growth and aggregation of calcium oxalate crystals in normal undiluted human urine”. eng. In: *The Journal of Urology* 167.1 (Jan. 2002), pp. 317–321. DOI: 10.1016/s0022-5347(05)65461-4.

[36] Lael Reinstatler et al. “Conversion from Cystine to Noncystine Stones: Incidence and Associated Factors”. en. In: *Journal of Urology* 200.6 (Dec. 2018), pp. 1285–1289. DOI: 10.1016/j.juro.2018.07.047. (Visited on 10/31/2024).

[37] J. Asplin et al. “Supersaturation and stone composition in a network of dispersed treatment sites”. eng. In: *The Journal of Urology* 159.6 (June 1998), pp. 1821–1825. DOI: 10.1016/S0022-5347(01)63164-1.

[38] F. L. Coe et al. “Proportional reduction of urine supersaturation during nephrolithiasis treatment”. eng. In: *The Journal of Urology* 166.4 (Oct. 2001), pp. 1247–1251.

[39] B. Finlayson and F. Reid. “The expectation of free and fixed particles in urinary stone disease”. eng. In: *Investigative Urology* 15.6 (May 1978), pp. 442–448.

[40] W. G. Robertson, M. Peacock, and B. E. C. Nordin. “Calcium Oxalate Crystalluria and Urine Saturation in Recurrent Renal Stone-Formers”. en. In: *Clinical Science* 40.5 (May 1971), pp. 365–374. DOI: 10.1042/cs0400365. (Visited on 10/14/2022).

[41] Saeed R. Khan and Benjamin K. Canales. “Unified theory on the pathogenesis of Randall’s plaques and plugs”. eng. In: *Urolithiasis* 43 Suppl 1.0 1 (Jan. 2015), pp. 109–123. DOI: 10.1007/s00240-014-0705-9.

[42] A. Evan et al. “Randall’s plaque: Pathogenesis and role in calcium oxalate nephrolithiasis”. en. In: *Kidney International* 69.8 (Apr. 2006), pp. 1313–1318. DOI: 10.1038/sj.ki.5000238. (Visited on 01/21/2025).

[43] Andrew P. Evan. “Physiopathology and etiology of stone formation in the kidney and the urinary tract”. en. In: *Pediatric Nephrology* 25.5 (May 2010), pp. 831–841. DOI: 10.1007/s00467-009-1116-y. (Visited on 01/21/2025).

[44] Fredric L. Coe et al. “Three pathways for human kidney stone formation”. en. In: *Urological Research* 38.3 (June 2010), pp. 147–160. DOI: 10.1007/s00240-010-0271-8. (Visited on 01/21/2025).

[45] Marshall L. Stoller et al. “The Primary Stone Event: A New Hypothesis Involving a Vascular Etiology”. en. In: *Journal of Urology* 171.5 (May 2004), pp. 1920–1924. DOI: 10.1097/01.ju.0000120291.90839.49. (Visited on 11/24/2024).

[46] Xiangling Wang et al. “Distinguishing characteristics of idiopathic calcium oxalate kidney stone formers with low amounts of Randall’s plaque”. eng. In: *Clinical journal of the American Society of Nephrology: CJASN* 9.10 (Oct. 2014), pp. 1757–1763. DOI: 10.2215/CJN.01490214.

[47] Michael P. Linnes et al. “Phenotypic characterization of kidney stone formers by endoscopic and histological quantification of intrarenal calcification”. en. In: *Kidney International* 84.4 (Oct. 2013), pp. 818–825. DOI: 10.1038/ki.2013.189. (Visited on 10/07/2024).

[48] H. Fleisch. “Inhibitors and promoters of stone formation”. eng. In: *Kidney International* 13.5 (May 1978), pp. 361–371. DOI: 10.1038/ki.1978.54.

[49] H. Fleisch and S. Bisaz. “The inhibitory effect of pyrophosphate on calcium oxalate precipitation and its relation to urolithiasis”. en. In: *Experientia* 20.5 (May 1964), pp. 276–277. DOI: 10.1007/BF02151805. (Visited on 08/15/2024).

[50] Eiji Konya et al. “The Role of Osteopontin on Calcium Oxalate Crystal Formation”. In: *European Urology* 43.5 (May 2003), pp. 564–571. DOI: 10.1016/S0302-2838(03)00088-5. (Visited on 09/26/2024).

[51] Alexander Randall. “The etiology of primary renal calculus”. In: *International Abstract of Surgery*. Vol. 71. 1940, pp. 209–240.

[52] Fadil M. Hannan et al. “The calcium-sensing receptor in physiology and in calcitropic and noncalcitropic diseases”. en. In: *Nature Reviews Endocrinology*

15.1 (Jan. 2019). Number: 1 Publisher: Nature Publishing Group, pp. 33–51. DOI: 10.1038/s41574-018-0115-0. (Visited on 06/07/2023).

[53] Fadil M Hannan, Valerie N Babinsky, and Rajesh V Thakker. "Disorders of the calcium-sensing receptor and partner proteins: insights into the molecular basis of calcium homeostasis". In: *Journal of Molecular Endocrinology* 57.3 (Oct. 2016), R127–R142. DOI: 10.1530/JME-16-0124. (Visited on 02/21/2024).

[54] E. M. Brown and R. J. MacLeod. "Extracellular calcium sensing and extracellular calcium signaling". eng. In: *Physiological Reviews* 81.1 (Jan. 2001), pp. 239–297. DOI: 10.1152/physrev.2001.81.1.239.

[55] Caroline M. Gorvin et al. "AP2$\sigma$ Mutations Impair Calcium-Sensing Receptor Trafficking and Signaling, and Show an Endosomal Pathway to Spatially Direct G-Protein Selectivity". eng. In: *Cell Reports* 22.4 (Jan. 2018), pp. 1054–1066. DOI: 10.1016/j.celrep.2017.12.089.

[56] Michael P. Grant et al. "Agonist-driven maturation and plasma membrane insertion of calcium-sensing receptors dynamically control signal amplitude". eng. In: *Science Signaling* 4.200 (Nov. 2011), ra78. DOI: 10.1126/scisignal.2002208.

[57] Theun de Groot, René J. M. Bindels, and Joost G. J. Hoenderop. "TRPV5: an ingeniously controlled calcium channel". In: *Kidney International* 74.10 (Nov. 2008), pp. 1241–1246. DOI: 10.1038/ki.2008.320. (Visited on 02/23/2024).

[58] R. Todd Alexander and Henrik Dimke. "Effects of parathyroid hormone on renal tubular calcium and phosphate handling". en. In: *Acta Physiologica* 238.1 (May 2023), e13959. DOI: 10.1111/apha.13959. (Visited on 03/05/2025).

[59] Daniela Riccardi and Giovanna Valenti. "Localization and function of the renal calcium-sensing receptor". eng. In: *Nature Reviews. Nephrology* 12.7 (July 2016), pp. 414–425. DOI: 10.1038/nrneph.2016.59.

[60] Lakshmi Kantham et al. "The calcium-sensing receptor (CaSR) defends against hypercalcemia independently of its regulation of parathyroid hormone secretion". eng. In: *American Journal of Physiology. Endocrinology and Metabolism* 297.4 (Oct. 2009), E915–923. DOI: 10.1152/ajpendo.00315.2009.

[61] Susan L. Murray and Myles Wolf. “Calcium and Phosphate Disorders: Core Curriculum 2024”. en. In: *American Journal of Kidney Diseases* 83.2 (Feb. 2024), pp. 241–256. DOI: 10.1053/j.ajkd.2023.04.017. (Visited on 03/05/2025).

[62] on behalf of the ERA–EDTA Working Group on Chronic Kidney Disease–Mineral and Bone Disorders and the European Renal Nutrition Working Group et al. “The role of phosphate in kidney disease”. en. In: *Nature Reviews Nephrology* 13.1 (Jan. 2017), pp. 27–38. DOI: 10.1038/nrneph.2016.164. (Visited on 10/29/2024).

[63] P. Raggi et al. “The ADVANCE study: a randomized study to evaluate the effects of cinacalcet plus low-dose vitamin D on vascular calcification in patients on hemodialysis”. en. In: *Nephrology Dialysis Transplantation* 26.4 (Apr. 2011), pp. 1327–1339. DOI: 10.1093/ndt/gfq725. (Visited on 10/29/2024).

[64] Fredric L. Coe, Elaine M. Worcester, and Andrew P. Evan. “Idiopathic hypercalciuria and formation of calcium renal stones”. eng. In: *Nature Reviews. Nephrology* 12.9 (Sept. 2016), pp. 519–533. DOI: 10.1038/nrneph.2016.101.

[65] Scott E. Liebman, Jeremy G. Taylor, and David A. Bushinsky. “Idiopathic hypercalciuria”. eng. In: *Current Rheumatology Reports* 8.1 (Feb. 2006), pp. 70–75. DOI: 10.1007/s11926-006-0029-z.

[66] G. C. Curhan and E. N. Taylor. “24-h uric acid excretion and the risk of kidney stones”. English. In: *Kidney International* 73.4 (Feb. 2008). Publisher: Elsevier, pp. 489–496. DOI: 10.1038/sj.ki.5002708. (Visited on 02/21/2024).

[67] Elaine M. Worcester and Fredric L. Coe. “New Insights Into the Pathogenesis of Idiopathic Hypercalciuria”. en. In: *Seminars in Nephrology*. Nephrolithiasis 28.2 (Mar. 2008), pp. 120–132. DOI: 10.1016/j.semnephrol.2008.01.005. (Visited on 10/17/2022).

[68] F. L. Coe and D. A. Bushinsky. “Pathophysiology of hypercalciuria”. eng. In: *The American Journal of Physiology* 247.1 Pt 2 (July 1984), F1–13. DOI: 10.1152/ajprenal.1984.247.1.F1.

[69] F. L. Coe, J. H. Parks, and E. S. Moore. “Familial idiopathic hypercalciuria”. eng. In: *The New England Journal of Medicine* 300.7 (Feb. 1979), pp. 337–340. DOI: 10.1056/NEJM197902153000703.

[70] U. A. Liberman et al. “Metabolic and calcium kinetic studies in idiopathic hypercalciuria”. en. In: *Journal of Clinical Investigation* 47.12 (Dec. 1968), pp. 2580–2590. DOI: 10.1172/JCI105940. (Visited on 10/29/2024).

[71] P. Bataille et al. “Diet, vitamin D and vertebral mineral density in hypercalciuric calcium stone formers”. eng. In: *Kidney International* 39.6 (June 1991), pp. 1193–1205. DOI: 10.1038/ki.1991.151.

[72] Fredric L. Coe et al. “Effects of low-calcium diet on urine calcium excretion, parathyroid function and serum 1,25(OH)2D3 levels in patients with idiopathic hypercalciuria and in normal subjects”. In: *The American Journal of Medicine* 72.1 (Jan. 1982), pp. 25–32. DOI: 10.1016/0002-9343(82)90567-8. (Visited on 02/21/2024).

[73] Fredric L. Coe, Joan H. Parks, and John R. Asplin. “The Pathogenesis and Treatment of Kidney Stones”. In: *New England Journal of Medicine* 327.16 (Oct. 1992). Publisher: Massachusetts Medical Society _eprint: https://www.nejm.org/doi/pdf/10.1056/NEJM199210153271607, pp. 1141–1152. DOI: 10.1056/NEJM199210153271607. (Visited on 10/07/2024).

[74] Marcella D. Walker and Shonni J. Silverberg. “Primary hyperparathyroidism”. en. In: *Nature Reviews Endocrinology* 14.2 (Feb. 2018). Publisher: Nature Publishing Group, pp. 115–125. DOI: 10.1038/nrendo.2017.104. (Visited on 10/16/2024).

[75] Michael W. Yeh et al. “Incidence and prevalence of primary hyperparathyroidism in a racially mixed population”. eng. In: *The Journal of Clinical Endocrinology and Metabolism* 98.3 (Mar. 2013), pp. 1122–1129. DOI: 10.1210/jc.2012-4022.

[76] Edward M. Brown. “Role of the calcium-sensing receptor in extracellular calcium homeostasis”. eng. In: *Best Practice & Research. Clinical Endocrinology & Metabolism* 27.3 (June 2013), pp. 333–343. DOI: 10.1016/j.beem.2013.02.006.

[77] F. Cetani et al. “Parathyroid expression of calcium-sensing receptor protein and in vivo parathyroid hormone-Ca(2+) set-point in patients with primary hyperparathyroidism”. eng. In: *The Journal of Clinical Endocrinology and Metabolism* 85.12 (Dec. 2000), pp. 4789–4794. DOI: 10.1210/jcem.85.12.7028.

[78] S. Khosla et al. “Calcium infusion suggests a "set-point" abnormality of parathyroid gland function in familial benign hypercalcemia and more complex disturbances in primary hyperparathyroidism”. eng. In: *The Journal of Clinical Endocrinology and Metabolism* 76.3 (Mar. 1993), pp. 715–720. DOI: 10.1210/jcem.76.3.8445032.

[79] Silverberg Sj et al. “A 10-year prospective study of primary hyperparathyroidism with or without parathyroid surgery”. en. In: *The New England journal of medicine* 341.17 (Oct. 1999). Publisher: N Engl J Med. DOI: 10.1056/NEJM199910213411701. (Visited on 10/16/2024).

[80] Lars Rejnmark, Peter Vestergaard, and Leif Mosekilde. “Nephrolithiasis and renal calcifications in primary hyperparathyroidism”. eng. In: *The Journal of Clinical Endocrinology and Metabolism* 96.8 (Aug. 2011), pp. 2377–2385. DOI: 10.1210/jc.2011-0569.

[81] W. Saliba and B. El-Haddad. “Secondary Hyperparathyroidism: Pathophysiology and Treatment”. en. In: *The Journal of the American Board of Family Medicine* 22.5 (Sept. 2009), pp. 574–581. DOI: 10.3122/jabfm.2009.05.090026. (Visited on 01/21/2025).

[82] Kithmini Nadeeshani Gamage et al. “The role of fluid intake in the prevention of kidney stone disease: A systematic review over the last two decades”. In: *Turkish Journal of Urology* 46.Suppl 1 (Nov. 2020), S92–S103. DOI: 10.5152/tud.2020.20155. (Visited on 10/07/2024).

[83] Thomas J. Littlejohns et al. “Fluid Intake and Dietary Factors and the Risk of Incident Kidney Stones in UK Biobank: A Population-based Prospective Cohort Study”. en. In: *European Urology Focus* 6.4 (July 2020), pp. 752–761. DOI: 10.1016/j.euf.2019.05.002. (Visited on 09/09/2024).

[84] Loris Borghi et al. “Comparison of Two Diets for the Prevention of Recurrent Stones in Idiopathic Hypercalciuria”. en. In: *New England Journal of Medicine*

346.2 (Jan. 2002), pp. 77–84. DOI: 10.1056/NEJMoa010369. (Visited on 03/05/2024).

[85] Kiriaki K. Stamatelou et al. "Time trends in reported prevalence of kidney stones in the United States: 1976–199411.See Editorial by Goldfarb, p. 1951." en. In: *Kidney International* 63.5 (May 2003), pp. 1817–1823. DOI: 10.1046/j.1523-1755.2003.00917.x. (Visited on 10/13/2022).

[86] R. A. Hiatt et al. "Randomized controlled trial of a low animal protein, high fiber diet in the prevention of recurrent calcium oxalate kidney stones". eng. In: *American Journal of Epidemiology* 144.1 (July 1996), pp. 25–33. DOI: 10.1093/oxfordjournals.aje.a008851.

[87] Bertrand Dussol et al. "A randomized trial of low-animal-protein or high-fiber diets for secondary prevention of calcium nephrolithiasis". eng. In: *Nephron. Clinical Practice* 110.3 (2008), pp. c185–194. DOI: 10.1159/000167271.

[88] Stef Robijn et al. "Hyperoxaluria: a gut–kidney axis?" English. In: *Kidney International* 80.11 (Dec. 2011). Publisher: Elsevier, pp. 1146–1158. DOI: 10.1038/ki.2011.287. (Visited on 02/27/2024).

[89] H. Andersson and R. Jagenburg. "Fat-reduced diet in the treatment of hyperoxaluria in patients with ileopathy". eng. In: *Gut* 15.5 (May 1974), pp. 360–366. DOI: 10.1136/gut.15.5.360.

[90] Frederic L. Coe, Murray J. Favus, and Jon R. Davids. "Disorders of bone and mineral metabolism". In: *Journal of Pediatric Orthopaedics* 12.6 (1992). ISBN: 0271-6798 Publisher: LWW, p. 815.

[91] Francesco P. Cappuccio et al. "A prospective study of hypertension and the incidence of kidney stones in men". en-US. In: *Journal of Hypertension* 17.7 (July 1999), pp. 1017–1022. (Visited on 10/13/2022).

[92] Andrew D. Rule, Amy E. Krambeck, and John C. Lieske. "Chronic Kidney Disease in Kidney Stone Formers". en. In: *Clinical Journal of the American Society of Nephrology* 6.8 (Aug. 2011), pp. 2069–2075. DOI: 10.2215/CJN.10651110. (Visited on 10/13/2022).

[93] Andrew D. Rule et al. "Kidney stones and the risk for chronic kidney disease". eng. In: *Clinical journal of the American Society of Nephrology: CJASN* 4.4 (Apr. 2009), pp. 804–811. DOI: 10.2215/CJN.05811108.

[94] Yee Wong et al. “Metabolic Syndrome and Kidney Stone Disease: A Systematic Review of Literature”. eng. In: *Journal of Endourology* 30.3 (Mar. 2016), pp. 246–253. DOI: 10.1089/end.2015.0567.

[95] Hao Ping et al. “New-onset metabolic risk factors and the incidence of kidney stones: a prospective cohort study”. en. In: *BJU International* 124.6 (2019). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/bju.14805, pp. 1028–1033. DOI: 10.1111/bju.14805. (Visited on 10/13/2022).

[96] Jodi A. Antonelli et al. “Use of the National Health and Nutrition Examination Survey to Calculate the Impact of Obesity and Diabetes on Cost and Prevalence of Urolithiasis in 2030”. en. In: *European Urology* 66.4 (Oct. 2014), pp. 724–729. DOI: 10.1016/j.eururo.2014.06.036. (Visited on 10/13/2022).

[97] Knut Westlund. “Urolithiasis and coronary heart disease”. en. In: *American Journal of Epidemiology* 97.3 (Mar. 1973), pp. 167–172. DOI: 10.1093/oxfordjournals.aje.a121497. (Visited on 10/07/2024).

[98] Robert M. Geraghty et al. “Risk of Metabolic Syndrome in Kidney Stone Formers: A Comparative Cohort Study with a Median Follow-Up of 19 Years”. eng. In: *Journal of Clinical Medicine* 10.5 (Mar. 2021), p. 978. DOI: 10.3390/jcm10050978.

[99] M. I. Schmidt et al. “Clustering of dyslipidemia, hyperuricemia, diabetes, and hypertension and its association with fasting insulin and central and overall obesity in a general population. Atherosclerosis Risk in Communities Study Investigators”. eng. In: *Metabolism: Clinical and Experimental* 45.6 (June 1996), pp. 699–706. DOI: 10.1016/s0026-0495(96)90134-1.

[100] Eric N. Taylor, Meir J. Stampfer, and Gary C. Curhan. “Obesity, Weight Gain, and the Risk of Kidney Stones”. In: *JAMA* 293.4 (Jan. 2005), pp. 455–462. DOI: 10.1001/jama.293.4.455. (Visited on 10/17/2022).

[101] Adam O. Kadlec et al. “Metabolic Syndrome and Urinary Stone Composition: What Factors Matter Most?” English. In: *Urology* 80.4 (Oct. 2012). Publisher: Elsevier, pp. 805–810. DOI: 10.1016/j.urology.2012.05.011. (Visited on 11/27/2024).

[102] Nicola Dalbeth et al. “Effect of body mass index on serum urate and renal uric acid handling responses to an oral inosine load: experimental intervention study in healthy volunteers”. en. In: *Arthritis Research & Therapy* 22.1 (Dec. 2020), p. 259. DOI: 10.1186/s13075-020-02357-y. (Visited on 10/26/2022).

[103] S. Yamashita et al. “Studies on the impaired metabolism of uric acid in obese subjects: marked reduction of renal urate excretion and its improvement by a low-calorie diet”. eng. In: *International Journal of Obesity* 10.4 (1986), pp. 255–264.

[104] Nicola Abate et al. “The metabolic syndrome and uric acid nephrolithiasis: Novel features of renal manifestation of insulin resistance”. English. In: *Kidney International* 65.2 (Feb. 2004). Publisher: Elsevier, pp. 386–392. DOI: 10.1111/j.1523-1755.2004.00386.x. (Visited on 10/17/2022).

[105] Mary Ann Cameron et al. “Urine Composition in Type 2 Diabetes: Predisposition to Uric Acid Nephrolithiasis”. en. In: *Journal of the American Society of Nephrology* 17.5 (May 2006). Publisher: American Society of Nephrology Section: Clinical Nephrology, pp. 1422–1428. DOI: 10.1681/ASN.2005121246. (Visited on 10/17/2022).

[106] Naim M. Maalouf et al. “Association of urinary pH with body weight in nephrolithiasis”. en. In: *Kidney International* 65.4 (Apr. 2004), pp. 1422–1425. DOI: 10.1111/j.1523-1755.2004.00522.x. (Visited on 10/17/2022).

[107] Naim M. Maalouf et al. “Metabolic Basis for Low Urine pH in Type 2 Diabetes”. en-US. In: *Clinical Journal of the American Society of Nephrology* 5.7 (July 2010), p. 1277. DOI: 10.2215/CJN.08331109. (Visited on 11/27/2024).

[108] I. Alexandru Bobulescu et al. “Effect of renal lipid accumulation on proximal tubule Na+/H+ exchange and ammonium secretion”. In: *American Journal of Physiology-Renal Physiology* 294.6 (June 2008). Publisher: American Physiological Society, F1315–F1322. DOI: 10.1152/ajprenal.00550.2007. (Visited on 11/27/2024).

[109] Zahra Jafari-Giv et al. “Association of body mass index with serum calcium and phosphate levels”. en. In: *Diabetes & Metabolic Syndrome: Clinical*

*Research & Reviews* 13.2 (Mar. 2019), pp. 975–980. DOI: 10.1016/j.dsx.2018.12.017. (Visited on 11/24/2024).

[110] Fabio Cesar Miranda Torricelli et al. "Dyslipidemia and Kidney Stone Risk". In: *The Journal of Urology* 191.3 (Mar. 2014), pp. 667–672. DOI: 10.1016/j.juro.2013.09.022. (Visited on 11/27/2024).

[111] Orhan Efe, Ashish Verma, and Sushrut S. Waikar. "Urinary oxalate as a potential mediator of kidney disease in diabetes mellitus and obesity:" en. In: *Current Opinion in Nephrology and Hypertension* 28.4 (July 2019), pp. 316–320. DOI: 10.1097/MNH.0000000000000515. (Visited on 11/24/2024).

[112] Ruhul Amin et al. "Reduced active transcellular intestinal oxalate secretion contributes to the pathogenesis of obesity-associated hyperoxaluria". en. In: *Kidney International* 93.5 (May 2018), pp. 1098–1107. DOI: 10.1016/j.kint.2017.11.011. (Visited on 10/24/2022).

[113] Gang Jee Ko, Hamid Rabb, and Heitham T. Hassoun. "Kidney-lung crosstalk in the critically ill patient". eng. In: *Blood Purification* 28.2 (2009), pp. 75–83. DOI: 10.1159/000218087.

[114] Loris Borghi et al. "Essential arterial hypertension and stone disease". English. In: *Kidney International* 55.6 (June 1999). Publisher: Elsevier, pp. 2397–2406. DOI: 10.1046/j.1523-1755.1999.00483.x. (Visited on 10/18/2022).

[115] F P Cappuccio, P Strazzullo, and M Mancini. "Kidney stones and hypertension: population based study of an independent clinical association." In: *BMJ : British Medical Journal* 300.6734 (May 1990), pp. 1234–1236. (Visited on 10/18/2022).

[116] Noelia Díaz-Morales et al. "Renal sympathetic activity: A key modulator of pressure natriuresis in hypertension". en. In: *Biochemical Pharmacology* 208 (Feb. 2023), p. 115386. DOI: 10.1016/j.bcp.2022.115386. (Visited on 11/27/2024).

[117] Bharathy Ponnuchamy and Raouf A. Khalil. "Cellular mediators of renal vascular dysfunction in hypertension". eng. In: *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology* 296.4 (Apr. 2009), R1001–1018. DOI: 10.1152/ajpregu.90960.2008.

[118] Christopher Hartman et al. “Does Hypertension Impact 24-Hour Urine Parameters in Patients With Nephrolithiasis?” English. In: *Urology* 85.3 (Mar. 2015). Publisher: Elsevier, pp. 539–543. DOI: 10.1016/j.urology.2014.12.013. (Visited on 11/27/2024).

[119] Catherine E. Lovegrove et al. “Central Adiposity Increases Risk of Kidney Stone Disease through Effects on Serum Calcium Concentrations”. en-US. In: *Journal of the American Society of Nephrology* 34.12 (Dec. 2023), p. 1991. DOI: 10.1681/ASN.0000000000000238. (Visited on 12/15/2023).

[120] Ahmad Jayedi et al. “Central fatness and risk of all cause mortality: systematic review and dose-response meta-analysis of 72 prospective cohort studies”. en. In: *BMJ* (Sept. 2020), p. m3324. DOI: 10.1136/bmj.m3324. (Visited on 01/20/2025).

[121] Amanda M. Barberio et al. “Central body fatness is a stronger predictor of cancer risk than overall body size”. en. In: *Nature Communications* 10.1 (Jan. 2019). Publisher: Nature Publishing Group, p. 383. DOI: 10.1038/s41467-018-08159-w. (Visited on 01/20/2025).

[122] Tiziana Ronti, Graziana Lupattelli, and Elmo Mannarino. “The endocrine function of adipose tissue: an update”. en. In: *Clinical Endocrinology* 64.4 (Apr. 2006), pp. 355–365. DOI: 10.1111/j.1365-2265.2006.02474.x. (Visited on 01/20/2025).

[123] C. Y. Pak and L. H. Arnold. “Heterogeneous nucleation of calcium oxalate by seeds of monosodium urate”. eng. In: *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, N.Y.)* 149.4 (Sept. 1975), pp. 930–932. DOI: 10.3181/00379727-149-38929.

[124] Michael R. Wiederkehr and Orson W. Moe. “Uric Acid Nephrolithiasis: A Systemic Metabolic Disorder”. In: *Clinical reviews in bone and mineral metabolism* 9.3-4 (Dec. 2011), pp. 207–217. DOI: 10.1007/s12018-011-9106-6. (Visited on 02/21/2024).

[125] Antonia Fabris et al. “Medullary sponge kidney: state of the art”. In: *Nephrology Dialysis Transplantation* 28.5 (May 2013), pp. 1111–1119. DOI: 10.1093/ndt/gfs505. (Visited on 12/19/2024).

[126] Melina Messing et al. “Trigger Warning: How Modern Diet, Lifestyle, and Environment Pull the Trigger on Autosomal Dominant Polycystic Kidney Disease Progression”. en. In: *Nutrients* 16.19 (Sept. 2024), p. 3281. DOI: 10.3390/nu16193281. (Visited on 12/19/2024).

[127] Le-Ting Zhou et al. “Association between Kidney Stones and CKD: A Bidirectional Mendelian Randomization Study”. en-US. In: *Journal of the American Society of Nephrology* 35.12 (Dec. 2024), p. 1746. DOI: 10.1681/ASN.0000000000000453. (Visited on 02/10/2025).

[128] V. E. Torres et al. “Renal stone disease in autosomal dominant polycystic kidney disease”. eng. In: *American Journal of Kidney Diseases: The Official Journal of the National Kidney Foundation* 22.4 (Oct. 1993), pp. 513–519. DOI: 10.1016/s0272-6386(12)80922-x.

[129] Vinusha Kalatharan et al. “Stone Prevalence in Autosomal Dominant Polycystic Kidney Disease: A Systematic Review and Meta-Analysis”. en. In: *Canadian Journal of Kidney Health and Disease* 7 (Jan. 2020), p. 2054358120934628. DOI: 10 . 1177 / 2054358120934628. (Visited on 02/04/2025).

[130] Jacob A. Torres et al. “Crystal deposition triggers tubule dilation that accelerates cystogenesis in polycystic kidney disease”. eng. In: *The Journal of Clinical Investigation* 129.10 (July 2019), pp. 4506–4522. DOI: 10.1172/JCI128503.

[131] Zhaoxia Sun et al. “A genetic screen in zebrafish identifies cilia genes as a principal cause of cystic kidney”. en. In: *Development* 131.16 (Aug. 2004), pp. 4085–4093. DOI: 10.1242/dev.01240. (Visited on 02/04/2025).

[132] Qihong Zhang, Patrick D. Taulman, and Bradley K. Yoder. “Cystic Kidney Diseases: All Roads Lead to the Cilium”. en. In: *Physiology* 19.4 (Aug. 2004), pp. 225–230. DOI: 10.1152/physiol.00003.2004. (Visited on 02/04/2025).

[133] Eva Sammels et al. “Polycystin-2 activation by inositol 1,4,5-trisphosphate-induced Ca2+ release requires its direct association with the inositol 1,4,5-trisphosphate receptor in a signaling microdomain”. eng. In: *The Journal of Biological Chemistry* 285.24 (June 2010), pp. 18794–18805. DOI: 10.1074/jbc.M109.090662.

[134] Yiqiang Cai et al. “Calcium dependence of polycystin-2 channel activity is modulated by phosphorylation at Ser812”. eng. In: *The Journal of Biological Chemistry* 279.19 (May 2004), pp. 19987–19995. DOI: 10.1074/jbc.M312031200.

[135] Ivana Y. Kuo et al. “Cyst formation following disruption of intracellular calcium signaling”. In: *Proceedings of the National Academy of Sciences* 111.39 (Sept. 2014). Publisher: Proceedings of the National Academy of Sciences, pp. 14283–14288. DOI: 10.1073/pnas.1412323111. (Visited on 02/11/2025).

[136] Kyproulla Demetriou et al. “Autosomal dominant polycystic kidney disease—type 2. Ultrasound, genetic and clinical correlations”. In: *Nephrology Dialysis Transplantation* 15.2 (Feb. 2000), pp. 205–211. DOI: 10.1093/ndt/15.2.205. (Visited on 02/04/2025).

[137] V. E. Torres et al. “The Association of Nephrolithiasis and Autosomal Dominant Polycystic Kidney Disease”. English. In: *American Journal of Kidney Diseases* 11.4 (Apr. 1988). Publisher: Elsevier, pp. 318–325. DOI: 10.1016/S0272-6386(88)80137-9. (Visited on 02/04/2025).

[138] José Luiz Nishiura, Samara Rodrigues Moreira Eloi, and Ita Pfeferman Heilberg. “Pain determinants of pain in autosomal dominant polycystic kidney disease”. en. In: *Jornal Brasileiro de Nefrologia* 35.3 (2013), pp. 242–243. DOI: 10.5935/0101-2800.20130038. (Visited on 02/04/2025).

[139] Abdullah Ozkok et al. “Clinical characteristics and predictors of progression of chronic kidney disease in autosomal dominant polycystic kidney disease: a single center experience”. en. In: *Clinical and Experimental Nephrology* 17.3 (June 2013), pp. 345–351. DOI: 10.1007/s10157-012-0706-3. (Visited on 02/04/2025).

[140] Berenice Y. Gitomer et al. “Kidney Stone Disease and Progression Risk in Autosomal Dominant Polycystic Kidney Disease: A Post Hoc Analysis of OVERTURE”. In: *Kidney360* 5.9 (July 2024), pp. 1364–1366. DOI: 10.34067/KID.0000000000000524. (Visited on 02/04/2025).

[141] Isabella Pisani et al. “Ultrasound to address medullary sponge kidney: a retrospective study”. In: *BMC Nephrology* 21 (Oct. 2020), p. 430. DOI: 10.1186/s12882-020-02084-1. (Visited on 02/04/2025).

[142] Talha H. Imam, Haris Patail, and Hassan Patail. “Medullary Sponge Kidney: Current Perspectives”. eng. In: *International Journal of Nephrology and Renovascular Disease* 12 (2019), pp. 213–218. DOI: 10.2147/IJNRD.S169336.

[143] Anja Pfau and Felix Knauf. “Update on Nephrolithiasis: Core Curriculum 2016”. eng. In: *American Journal of Kidney Diseases: The Official Journal of the National Kidney Foundation* 68.6 (Dec. 2016), pp. 973–985. DOI: 10.1053/j.ajkd.2016.05.016.

[144] M. M. Murshidi and Z. A. Suwan. “Simple renal cysts”. eng. In: *Archivos Espanoles De Urologia* 50.8 (Oct. 1997), pp. 928–931.

[145] S. P. Laucks and M. S. McLachlan. “Aging and simple cysts of the kidney”. eng. In: *The British Journal of Radiology* 54.637 (Jan. 1981), pp. 12–14. DOI: 10.1259/0007-1285-54-637-12.

[146] Michele Bisceglia et al. “Renal Cystic Diseases: A REVIEW”. en. In: *Advances in Anatomic Pathology* 13.1 (Jan. 2006), pp. 26–56. DOI: 10.1097/01.pap.0000201831.77472.d3. (Visited on 02/04/2025).

[147] L. Baert and A. Steg. “Is the Diverticulum of the Distal and Collecting Tubules a Preliminary Stage of the Simple Cyst in the Adult?” en. In: *Journal of Urology* 118.5 (Nov. 1977), pp. 707–710. DOI: 10.1016/S0022-5347(17)58167-7. (Visited on 02/04/2025).

[148] Victor Garcia-Nieto et al. “Are simple renal cysts in childhood associated with kidney stones?” eng. In: *Nephro-Urology Monthly* 4.4 (2012), pp. 596–598. DOI: 10.5812/numonthly.2682.

[149] Chang-Chi Chang et al. “Prevalence and Clinical Characteristics of Simple Renal Cyst”. en-US. In: *Journal of the Chinese Medical Association* 70.11 (Nov. 2007), p. 486. DOI: 10.1016/S1726-4901(08)70046-7. (Visited on 02/04/2025).

[150] Daniel G. Fuster and Orson W. Moe. “Incomplete Distal Renal Tubular Acidosis and Kidney Stones”. eng. In: *Advances in Chronic Kidney Disease* 25.4 (July 2018), pp. 366–374. DOI: 10.1053/j.ackd.2018.05.007.

[151] M. Henry Gault et al. “Comparison of Patients with Idiopathic Calcium Phosphate and Calcium Oxalate Stones:” en. In: *Medicine* 70.6 (Nov. 1991), pp. 345–359. DOI: 10.1097/00005792-199111000-00001. (Visited on 01/21/2025).

[152] Jordi Guimerà et al. “Prevalence of distal renal tubular acidosis in patients with calcium phosphate stones”. en. In: *World Journal of Urology* 38.3 (Mar. 2020), pp. 789–794. DOI: 10.1007/s00345-019-02804-9. (Visited on 01/21/2025).

[153] Marta Alonso-Varela et al. “Distal renal tubular acidosis. Clinical manifestations in patients with different underlying gene mutations”. eng. In: *Pediatric Nephrology (Berlin, Germany)* 33.9 (Sept. 2018), pp. 1523–1529. DOI: 10.1007/s00467-018-3965-8.

[154] R. J. Caruana and V. M. Buckalew. “The syndrome of distal (type 1) renal tubular acidosis. Clinical and laboratory findings in 58 cases”. eng. In: *Medicine* 67.2 (Mar. 1988), pp. 84–99. DOI: 10.1097/00005792-198803000-00002.

[155] David S. Goldfarb et al. “A Twin Study of Genetic Influences on Nephrolithiasis in Women and Men”. In: *Kidney International Reports* 4.4 (Nov. 2018), pp. 535–540. DOI: 10.1016/j.ekir.2018.11.017. (Visited on 06/07/2023).

[156] Martin Resnick, Durward B. Pridgen, and Harold O. Goodman. “Genetic Predisposition to Formation of Calcium Oxalate Renal Calculi”. In: *New England Journal of Medicine* 278.24 (June 1968). Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJM196806132782403, pp. 1313–1318. DOI: 10.1056/NEJM196806132782403. (Visited on 02/23/2024).

[157] Manuel A Anderegg et al. “Prevalence and characteristics of genetic disease in adult kidney stone formers”. In: *Nephrology Dialysis Transplantation* (Mar. 2024), gfae074. DOI: 10.1093/ndt/gfae074. (Visited on 07/03/2024).

[158] Sarah A. Howles et al. “Genetic variants of calcium and vitamin D metabolism in kidney stone disease”. en. In: *Nature Communications* 10.1 (Nov. 2019).

Number: 1 Publisher: Nature Publishing Group, p. 5175. DOI: 10.1038/s41467-019-13145-x. (Visited on 09/05/2022).

[159] Ria Schönauer et al. "Systematic assessment of monogenic etiology in adult-onset kidney stone formers undergoing urological intervention-evidence for genetic pretest probability". eng. In: *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics* 190.3 (Sept. 2022), pp. 279–288. DOI: 10.1002/ajmg.c.31991.

[160] Ankana Daga et al. "Whole exome sequencing frequently detects a monogenic cause in early onset nephrolithiasis and nephrocalcinosis". In: *Kidney international* 93.1 (Jan. 2018), pp. 204–213. DOI: 10.1016/j.kint.2017.06.025. (Visited on 02/21/2024).

[161] Dorothea Szczawinska et al. "A Homozygous CaSR Mutation Causing a FHH Phenotype Completely Masked by Vitamin D Deficiency Presenting as Rickets". In: *The Journal of Clinical Endocrinology & Metabolism* 99.6 (June 2014), E1146–E1153. DOI: 10.1210/jc.2013-3593. (Visited on 10/16/2024).

[162] S. H. Pearce et al. "A familial syndrome of hypocalcemia with hypercalciuria due to mutations in the calcium-sensing receptor". eng. In: *The New England Journal of Medicine* 335.15 (Oct. 1996), pp. 1115–1122. DOI: 10.1056/NEJM199610103351505.

[163] M. Andrew Nesbit et al. "Mutations in AP2S1 cause familial hypocalciuric hypercalcemia type 3". en. In: *Nature Genetics* 45.1 (Jan. 2013). Number: 1 Publisher: Nature Publishing Group, pp. 93–97. DOI: 10.1038/ng.2492. (Visited on 06/07/2023).

[164] M. Andrew Nesbit et al. "Mutations affecting G-protein subunit $\alpha$11 in hypercalcemia and hypocalcemia". eng. In: *The New England Journal of Medicine* 368.26 (June 2013), pp. 2476–2486. DOI: 10.1056/NEJMoa1300253.

[165] S. H. Pearce et al. "Calcium-sensing receptor mutations in familial hypocalciuric hypercalcaemia with recurrent pancreatitis". eng. In: *Clinical Endocrinology* 45.6 (Dec. 1996), pp. 675–680. DOI: 10.1046/j.1365-2265.1996.750891.x.

[166] Daniela Riccardi and Edward M. Brown. "Physiology and pathophysiology of the calcium-sensing receptor in the kidney". eng. In: *American Journal of*

*Physiology. Renal Physiology* 298.3 (Mar. 2010), F485–499. DOI: 10.1152/ajprenal.00608.2009.

[167] Berenice Y. Reed et al. "Mapping a Gene Defect in Absorptive Hypercalciuria to Chromosome 1q23.3-q241". In: *The Journal of Clinical Endocrinology & Metabolism* 84.11 (Nov. 1999), pp. 3907–3913. DOI: 10.1210/jcem.84.11.6155. (Visited on 02/21/2024).

[168] P. Scott et al. "Suggestive evidence for a susceptibility gene near the vitamin D receptor locus in idiopathic calcium stone formation". eng. In: *Journal of the American Society of Nephrology: JASN* 10.5 (May 1999), pp. 1007–1013. DOI: 10.1681/ASN.V1051007.

[169] Berenice Y. Reed et al. "Identification and characterization of a gene with base substitutions associated with the absorptive hypercalciuria phenotype and low spinal bone density". eng. In: *The Journal of Clinical Endocrinology and Metabolism* 87.4 (Apr. 2002), pp. 1476–1485. DOI: 10.1210/jcem.87.4.8300.

[170] D. B. Simon et al. "Paracellin-1, a renal tight junction protein required for paracellular Mg2+ resorption". eng. In: *Science (New York, N.Y.)* 285.5424 (July 1999), pp. 103–106. DOI: 10.1126/science.285.5424.103.

[171] Jianghui Hou et al. "Claudin-16 and claudin-19 interaction is required for their assembly into tight junctions and for renal reabsorption of magnesium". In: *Proceedings of the National Academy of Sciences* 106.36 (Sept. 2009). Publisher: Proceedings of the National Academy of Sciences, pp. 15350–15355. DOI: 10.1073/pnas.0907724106. (Visited on 03/11/2024).

[172] Felix Claverie-Martin. "Familial hypomagnesaemia with hypercalciuria and nephrocalcinosis: clinical and molecular characteristics". In: *Clinical Kidney Journal* 8.6 (Dec. 2015), pp. 656–664. DOI: 10.1093/ckj/sfv081. (Visited on 02/21/2024).

[173] Shaomin Peng et al. "Effects of proinflammatory cytokines on the claudin-19 rich tight junctions of human retinal pigment epithelium". eng. In: *Investigative Ophthalmology & Visual Science* 53.8 (July 2012), pp. 5016–5028. DOI: 10.1167/iovs.11-8311.

[174] M. Tieder et al. “Hereditary hypophosphatemic rickets with hypercalciuria”. eng. In: *The New England Journal of Medicine* 312.10 (Mar. 1985), pp. 611–617. DOI: 10.1056/NEJM198503073121003.

[175] Yue Chi et al. “A compound heterozygous mutation in SLC34A3 causes hereditary hypophosphatemic rickets with hypercalciuria in a Chinese patient”. eng. In: *Bone* 59 (Feb. 2014), pp. 114–121. DOI: 10.1016/j.bone.2013.11.008.

[176] Debayan Dasgupta et al. “Mutations in SLC34A3/NPT2c Are Associated with Kidney Stones and Nephrocalcinosis”. In: *Journal of the American Society of Nephrology : JASN* 25.10 (Oct. 2014), pp. 2366–2375. DOI: 10.1681/ASN.2013101085. (Visited on 02/21/2024).

[177] Sezer Acar et al. “Clinical and genetic characteristics of 15 families with hereditary hypophosphatemia: Novel Mutations in PHEX and SLC34A3”. eng. In: *PloS One* 13.3 (2018), e0193388. DOI: 10.1371/journal.pone.0193388.

[178] Vickie Braithwaite, John M. Pettifor, and Ann Prentice. “Novel SLC34A3 mutation causing hereditary hypophosphataemic rickets with hypercalciuria in a Gambian family”. In: *Bone* 53.1 (Mar. 2013), pp. 216–220. DOI: 10.1016/j.bone.2012.12.003. (Visited on 02/21/2024).

[179] Jan Halbritter et al. “Fourteen monogenic genes account for 15% of nephrolithiasis/nephrocalcinosis”. eng. In: *Journal of the American Society of Nephrology: JASN* 26.3 (Mar. 2015), pp. 543–551. DOI: 10.1681/ASN.2014040388.

[180] Omid Sadeghi-Alavijeh et al. “Rare variants in SLC34A3 explain missing heritability of urinary stone disease”. In: *medRxiv* (Jan. 2022), p. 2022.12.02.22283024. DOI: 10.1101/2022.12.02.22283024.

[181] Dominique Prié et al. “Nephrolithiasis and osteoporosis associated with hypophosphatemia caused by mutations in the type 2a sodium-phosphate cotransporter”. eng. In: *The New England Journal of Medicine* 347.13 (Sept. 2002), pp. 983–991. DOI: 10.1056/NEJMoa020028.

[182] Daniela Anne Braun et al. “Prevalence of Monogenic Causes in Pediatric Patients with Nephrolithiasis or Nephrocalcinosis”. eng. In: *Clinical journal*

*of the American Society of Nephrology: CJASN* 11.4 (Apr. 2016), pp. 664–672. DOI: 10.2215/CJN.07540715.

[183] Karl P. Schlingmann et al. "Autosomal-Recessive Mutations in SLC34A1 Encoding Sodium-Phosphate Cotransporter 2A Cause Idiopathic Infantile Hypercalcemia". eng. In: *Journal of the American Society of Nephrology: JASN* 27.2 (Feb. 2016), pp. 604–614. DOI: 10.1681/ASN.2014101025.

[184] Karl P. Schlingmann et al. "Mutations in CYP24A1 and Idiopathic Infantile Hypercalcemia". In: *New England Journal of Medicine* 365.5 (Aug. 2011). Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa1103864, pp. 410–421. DOI: 10.1056/NEJMoa1103864. (Visited on 03/01/2024).

[185] Nicole Ball et al. "3' Untranslated Region Structural Elements in CYP24A1 Are Associated With Infantile Hypercalcemia Type 1". en. In: *Journal of Bone and Mineral Research* 38.3 (2023). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jbmr.4769, pp. 414–426. DOI: 10.1002/jbmr.4769. (Visited on 03/05/2024).

[186] A. Molin et al. "CYP24A1 Mutations in a Cohort of Hypercalcemic Patients: Evidence for a Recessive Trait". eng. In: *The Journal of Clinical Endocrinology and Metabolism* 100.10 (Oct. 2015), E1343–1352. DOI: 10.1210/jc.2014-4387.

[187] Judith Sayers et al. "Successful treatment of hypercalcaemia associated with a CYP24A1 mutation with fluconazole". In: *Clinical Kidney Journal* 8.4 (Aug. 2015), pp. 453–455. DOI: 10.1093/ckj/sfv028. (Visited on 03/05/2024).

[188] H. E. Williams and L. H. Smith. "Hyperoxaluria in L-glyceric aciduria: possible pathogenic mechanism". eng. In: *Science (New York, N.Y.)* 171.3969 (Jan. 1971), pp. 390–391. DOI: 10.1126/science.171.3969.390.

[189] S. D. Cramer et al. "The gene encoding hydroxypyruvate reductase (GRHPR) is mutated in patients with primary hyperoxaluria type II". eng. In: *Human Molecular Genetics* 8.11 (Oct. 1999), pp. 2063–2069. DOI: 10.1093/hmg/8.11.2063.

[190] Travis J. Riedel et al. "Structural and Biochemical Studies of Human 4-hydroxy-2-oxoglutarate Aldolase: Implications for Hydroxyproline

Metabolism in Primary Hyperoxaluria". en. In: *PLOS ONE* 6.10 (Oct. 2011). Publisher: Public Library of Science, e26021. DOI: 10.1371/journal.pone.0026021. (Visited on 03/21/2024).

[191] Heon Yung Gee et al. "Mutations in SLC26A1 Cause Nephrolithiasis". eng. In: *American Journal of Human Genetics* 98.6 (June 2016), pp. 1228–1234. DOI: 10.1016/j.ajhg.2016.03.026.

[192] Aleksandra Sindić et al. "Renal physiology of SLC26 anion exchangers". eng. In: *Current Opinion in Nephrology and Hypertension* 16.5 (Sept. 2007), pp. 484–490. DOI: 10.1097/MNH.0b013e3282e7d7d0.

[193] Nicolas Cornière et al. "Dominant negative mutation in oxalate transporter SLC26A6 associated with enteric hyperoxaluria and nephrolithiasis". en. In: *Journal of Medical Genetics* 59.11 (Nov. 2022). Publisher: BMJ Publishing Group Ltd Section: Novel disease loci, pp. 1035–1043. DOI: 10.1136/jmedgenet-2021-108256. (Visited on 04/08/2024).

[194] Pierre Cochat et al. "Primary hyperoxaluria Type 1: indications for screening and guidance for diagnosis and treatment". In: *Nephrology Dialysis Transplantation* 27.5 (May 2012), pp. 1729–1736. DOI: 10.1093/ndt/gfs078. (Visited on 02/27/2024).

[195] Sander F. Garrelfs et al. "Lumasiran, an RNAi Therapeutic for Primary Hyperoxaluria Type 1". eng. In: *The New England Journal of Medicine* 384.13 (Apr. 2021), pp. 1216–1226. DOI: 10.1056/NEJMoa2021712.

[196] Michelle A. Baum et al. "PHYOX2: a pivotal randomized study of nedosiran in primary hyperoxaluria type 1 or 2". eng. In: *Kidney International* 103.1 (Jan. 2023), pp. 207–217. DOI: 10.1016/j.kint.2022.07.025.

[197] David S. Goldfarb et al. "Nedosiran in primary hyperoxaluria subtype 3: results from a phase I, single-dose study (PHYOX4)". eng. In: *Urolithiasis* 51.1 (Apr. 2023), p. 80. DOI: 10.1007/s00240-023-01453-3.

[198] Bernd Hoppe et al. "Safety, pharmacodynamics, and exposure-response modeling results from a first-in-human phase 1 study of nedosiran (PHYOX1) in primary hyperoxaluria". eng. In: *Kidney International* 101.3 (Mar. 2022), pp. 626–634. DOI: 10.1016/j.kint.2021.08.015.

[199] David S. Goldfarb. “The Search for Monogenic Causes of Kidney Stones”. en. In: *Journal of the American Society of Nephrology : JASN* 26.3 (Mar. 2015). Publisher: American Society of Nephrology, p. 507. DOI: 10.1681/ASN.2014090847. (Visited on 02/19/2024).

[200] Lisa M. Guay-Woodford. “Hereditary Nephropathies and Developmental Abnormalities of the Urinary Tract”. en. In: *Goldman’s Cecil Medicine*. Elsevier, 2012, pp. 800–805. DOI: 10.1016/B978-1-4377-1604-7.00130-5. (Visited on 02/20/2024).

[201] H. Anne Simmonds. “2,8-Dihydroxyadenine lithiasis”. In: *Clinica Chimica Acta* 160.2 (Oct. 1986), pp. 103–108. DOI: 10.1016/0009-8981(86)90129-4. (Visited on 02/23/2024).

[202] B. J. Roessler et al. “Human X-linked phosphoribosylpyrophosphate synthetase superactivity is associated with distinct point mutations in the PRPS1 gene”. eng. In: *The Journal of Biological Chemistry* 268.35 (Dec. 1993), pp. 26476–26481.

[203] E. Zoref, A. De Vries, and O. Sperling. “Mutant feedback-resistant phosphoribosylpyrophosphate synthetase associated with purine overproduction and gout. Phosphoribosylpyrophosphate and purine metabolism in cultured fibroblasts”. eng. In: *The Journal of Clinical Investigation* 56.5 (Nov. 1975), pp. 1093–1099. DOI: 10.1172/JCI108183.

[204] H. A. Jinnah et al. “The spectrum of inherited mutations causing HPRT deficiency: 75 new cases and a review of 196 previously reported cases”. eng. In: *Mutation Research* 463.3 (Oct. 2000), pp. 309–326. DOI: 10.1016/s1383-5742(00)00052-1.

[205] Andrea Mancikova et al. “Functional analysis of novel allelic variants in URAT1 and GLUT9 causing renal hypouricemia type 1 and 2”. en. In: *Clinical and Experimental Nephrology* 20.4 (Aug. 2016), pp. 578–584. DOI: 10.1007/s10157-015-1186-z. (Visited on 02/21/2024).

[206] Teiichi Mizuno et al. “An Autopsy Case of the Lesch-Nyhan Syndrome: Normal HGPRT Activity in Liver and Xanthine Calculi in Various Tissues”. en. In: *Neuropädiatrie* 7.03 (Aug. 1976). Publisher: © 1976 by Thieme

Medical Publishers, Inc., pp. 351–355. DOI: 10.1055/s-0028-1091635. (Visited on 02/21/2024).

[207] Guillaume Bollée. "Adenine phosphoribosyltransferase deficiency: Leave no stone unturned". en. In: *World Journal of Clinical Urology* 3.3 (2014), p. 218. DOI: 10.5410/wjcu.v3.i3.218. (Visited on 02/23/2024).

[208] Guillaume Bollée et al. "Phenotype and Genotype Characterization of Adenine Phosphoribosyltransferase Deficiency". In: *Journal of the American Society of Nephrology : JASN* 21.4 (Apr. 2010), pp. 679–688. DOI: 10.1681/ASN.2009080808. (Visited on 02/21/2024).

[209] Vidar Orn Edvardsson, Amrik Sahota, and Runolfur Palsson. "Adenine Phosphoribosyltransferase Deficiency". eng. In: *GeneReviews®*. Ed. by Margaret P. Adam et al. Seattle (WA): University of Washington, Seattle, 1993. (Visited on 02/23/2024).

[210] *Adenine Phosphoribosyltransferase Deficiency and 2,8-Dihydroxyadenine Lithiasis*. en. URL: https://ommbid.mhmedical.com/content.aspx?bookId=2709&sectionId=225089735 (visited on 02/23/2024).

[211] Vidar O. Edvardsson et al. "Comparison of the effect of allopurinol and febuxostat on urinary 2,8-dihydroxyadenine excretion in patients with Adenine phosphoribosyltransferase deficiency (APRTd): A clinical trial". eng. In: *European Journal of Internal Medicine* 48 (Feb. 2018), pp. 75–79. DOI: 10.1016/j.ejim.2017.10.007.

[212] Vidar O. Edvardsson et al. "Hereditary causes of kidney stones and chronic kidney disease". eng. In: *Pediatric Nephrology (Berlin, Germany)* 28.10 (Oct. 2013), pp. 1923–1942. DOI: 10.1007/s00467-012-2329-z.

[213] Y Hidaka et al. "Human adenine phosphoribosyltransferase deficiency. Demonstration of a single mutant allele common to the Japanese." In: *Journal of Clinical Investigation* 81.3 (Mar. 1988), pp. 945–950. (Visited on 02/23/2024).

[214] Hrafnhildur Linnet Runolfsdottir et al. "Kidney Disease in Adenine Phosphoribosyltransferase Deficiency". In: *American journal of kidney diseases : the official journal of the National Kidney Foundation* 67.3 (Mar.

2016), pp. 431–438. DOI: 10.1053/j.ajkd.2015.10.023. (Visited on 02/23/2024).

[215] K. Ichida et al. "Mutation of human molybdenum cofactor sulfurase gene is responsible for classical xanthinuria type II". eng. In: *Biochemical and Biophysical Research Communications* 282.5 (Apr. 2001), pp. 1194–1200. DOI: 10.1006/bbrc.2001.4719.

[216] K Ichida et al. "Identification of two mutations in human xanthine dehydrogenase gene responsible for classical type I xanthinuria." In: *Journal of Clinical Investigation* 99.10 (May 1997), pp. 2391–2397. (Visited on 02/22/2024).

[217] T. Kojima et al. "Biochemical studies on the purine metabolism of four cases with hereditary xanthinuria". eng. In: *Clinica Chimica Acta; International Journal of Clinical Chemistry* 137.2 (Feb. 1984), pp. 189–198. DOI: 10.1016/0009-8981(84)90179-7.

[218] H. A. Simmonds, S Reiter, and T Nishino. "Hereditary xanthinuria". In: *The Metabolic and Molecular Basis of Inherited Disease*. 7th Edition. New York: McGraw-Hill, 1995, pp. 1781–1797.

[219] Maha S. Zaki et al. "Molybdenum cofactor and isolated sulphite oxidase deficiencies: Clinical and molecular spectrum among Egyptian patients". eng. In: *European journal of paediatric neurology: EJPN: official journal of the European Paediatric Neurology Society* 20.5 (Sept. 2016), pp. 714–722. DOI: 10.1016/j.ejpn.2016.05.011.

[220] Jochen Reiss and Rita Hahnewald. "Molybdenum cofactor deficiency: Mutations in GPHN, MOCS1, and MOCS2". eng. In: *Human Mutation* 32.1 (Jan. 2011), pp. 10–18. DOI: 10.1002/humu.21390.

[221] Nina Arikyants et al. "Xanthinuria type I: a rare cause of urolithiasis". en. In: *Pediatric Nephrology* 22.2 (Feb. 2007), pp. 310–314. DOI: 10.1007/s00467-006-0267-3. (Visited on 11/19/2024).

[222] Fiona E. Karet et al. "Localization of a Gene for Autosomal Recessive Distal Renal Tubular Acidosis with Normal Hearing (rdRTA2) to 7q33-34". In: *American Journal of Human Genetics* 65.6 (Dec. 1999), pp. 1656–1665. (Visited on 03/21/2024).

[223] F. E. Karet et al. “Mutations in the gene encoding B1 subunit of H+-ATPase cause renal tubular acidosis with sensorineural deafness”. eng. In: *Nature Genetics* 21.1 (Jan. 1999), pp. 84–90. DOI: 10.1038/5022.

[224] Sven Enerbäck et al. “Acidosis and Deafness in Patients with Recessive Mutations in FOXI1”. en-US. In: *Journal of the American Society of Nephrology* 29.3 (Mar. 2018), p. 1041. DOI: 10.1681/ASN.2017080840. (Visited on 03/21/2024).

[225] N. Rungroj et al. “Distal renal tubular acidosis caused by tryptophan-aspartate repeat domain 72 (WDR72) mutations”. In: *Clinical genetics* 94.5 (2018). ISBN: 0009-9163 Publisher: Wiley Online Library, pp. 409–418.

[226] Sarah A. Howles and Rajesh V. Thakker. “Genetics of kidney stone disease”. en. In: *Nature Reviews Urology* 17.7 (July 2020). Number: 7 Publisher: Nature Publishing Group, pp. 407–421. DOI: 10.1038/s41585-020-0332-x. (Visited on 08/15/2023).

[227] H. Zhang et al. “WDR72 Mutations Associated with Amelogenesis Imperfecta and Acidosis”. eng. In: *Journal of Dental Research* 98.5 (May 2019), pp. 541–548. DOI: 10.1177/0022034518824571.

[228] Paulo Marcio Yamaguti et al. “Amelogenesis imperfecta in familial hypomagnesaemia and hypercalciuria with nephrocalcinosis caused by CLDN19 gene mutations”. eng. In: *Journal of Medical Genetics* 54.1 (Jan. 2017), pp. 26–37. DOI: 10.1136/jmedgenet-2016-103956.

[229] Hiroko Soda et al. “Carbonic anhydrase II deficiency in a Japanese patient produced by a nonsense mutation (TAT →TAG) at Tyr-40 in exon 2, (Y40X)”. en. In: *Human Mutation* 5.4 (1995). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/humu.1380050415, pp. 348–350. DOI: 10.1002/humu.1380050415. (Visited on 03/21/2024).

[230] P J Venta et al. “Carbonic anhydrase II deficiency syndrome in a Belgian family is caused by a point mutation at an invariant histidine residue (107 His—-Tyr): complete structure of the normal human CA II gene.” In: *American Journal of Human Genetics* 49.5 (Nov. 1991), pp. 1082–1090. (Visited on 10/08/2024).

[231] Audrey G. H. Janssen et al. "Disease-causing dysfunctions of barttin in Bartter syndrome type IV". eng. In: *Journal of the American Society of Nephrology: JASN* 20.1 (Jan. 2009), pp. 145–153. DOI: 10.1681/ASN.2008010102.

[232] Rosa Vargas-Poussou et al. "Functional characterization of a calcium-sensing receptor mutation in severe autosomal dominant hypocalcemia with a Bartter-like syndrome". eng. In: *Journal of the American Society of Nephrology: JASN* 13.9 (Sept. 2002), pp. 2259–2266. DOI: 10.1097/01.asn.0000025781.16723.68.

[233] Sumiyo Watanabe et al. "Association between activating mutations of calcium-sensing receptor and Bartter's syndrome". eng. In: *Lancet (London, England)* 360.9334 (Aug. 2002), pp. 692–694. DOI: 10.1016/S0140-6736(02)09842-2.

[234] Haifa Hichri et al. "From Lowe syndrome to Dent disease: correlations between mutations of the OCRL1 gene and clinical and biochemical phenotypes". eng. In: *Human Mutation* 32.4 (Apr. 2011), pp. 379–388. DOI: 10.1002/humu.21391.

[235] Hannsjörg W. Seyberth and Karl P. Schlingmann. "Bartter- and Gitelman-like syndromes: salt-losing tubulopathies with loop or DCT defects". In: *Pediatric Nephrology (Berlin, Germany)* 26.10 (Oct. 2011), pp. 1789–1802. DOI: 10.1007/s00467-011-1871-4. (Visited on 02/21/2024).

[236] Radovan Bogdanović et al. "A novel CLCN5 mutation in a boy with Bartter-like syndrome and partial growth hormone deficiency". en. In: *Pediatric Nephrology* 25.11 (Nov. 2010), pp. 2363–2368. DOI: 10.1007/s00467-010-1615-x. (Visited on 02/27/2025).

[237] Caroline M. Gorvin et al. "Receptor-mediated endocytosis and endosomal acidification is impaired in proximal tubule epithelial cells of Dent disease patients". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 110.17 (Apr. 2013), pp. 7014–7019. DOI: 10.1073/pnas.1302063110.

[238] Detlef Bockenhauer et al. "Renal Phenotype in Lowe Syndrome: A Selective Proximal Tubular Dysfunction". In: *Clinical Journal of the American Society*

*of Nephrology : CJASN* 3.5 (Sept. 2008), pp. 1430–1436. DOI: 10.2215/CJN.00520108. (Visited on 02/21/2024).

[239] Zenobia B. Mehta, Grzegorz Pietka, and Martin Lowe. "The cellular and physiological functions of the Lowe syndrome protein OCRL1". eng. In: *Traffic (Copenhagen, Denmark)* 15.5 (May 2014), pp. 471–487. DOI: 10.1111/tra.12160.

[240] Gudmar Thorleifsson et al. "Sequence variants in the CLDN14 gene associate with kidney stones and bone mineral density". en. In: *Nature Genetics* 41.8 (Aug. 2009). Number: 8 Publisher: Nature Publishing Group, pp. 926–930. DOI: 10.1038/ng.404. (Visited on 02/23/2024).

[241] Yuji Urabe et al. "A Genome-Wide Association Study of Nephrolithiasis in the Japanese Population Identifies Novel Susceptible Loci at 5q35.3, 7p14.3, and 13q14.1". en. In: *PLOS Genetics* 8.3 (Mar. 2012). Publisher: Public Library of Science, e1002541. DOI: 10.1371/journal.pgen.1002541. (Visited on 02/23/2024).

[242] Asmundur Oddsson et al. "Common and rare variants associated with kidney stones and biochemical traits". In: *Nature Communications* 6 (Aug. 2015), p. 7975. DOI: 10.1038/ncomms8975. (Visited on 06/13/2023).

[243] Xingjie Hao et al. "Integrative genome-wide analyses identify novel loci associated with kidney stones and provide insights into its genetic architecture". en. In: *Nature Communications* 14.1 (Nov. 2023). Number: 1 Publisher: Nature Publishing Group, p. 7498. DOI: 10.1038/s41467-023-43400-1. (Visited on 02/23/2024).

[244] Omid Sadeghi-Alavijeh et al. "Rare variants in the sodium-dependent phosphate transporter gene SLC34A3 explain missing heritability of urinary stone disease". In: *Kidney International* 104.5 (Nov. 2023), pp. 975–984. DOI: 10.1016/j.kint.2023.06.019. (Visited on 01/17/2024).

[245] Konrad J. Karczewski et al. "Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK Biobank exomes". en. In: *Cell Genomics* 2.9 (Sept. 2022), p. 100168. DOI: 10.1016/j.xgen.2022.100168. (Visited on 01/06/2025).

[246] Catherine E. Lovegrove et al. "Natural history of small asymptomatic kidney and residual stones over a long-term follow-up: systematic review over 25 years." eng. In: *BJU international* 129.4 (Apr. 2022). Place: England, pp. 442–456. DOI: 10.1111/bju.15522.

[247] Julian Yaxley and William Yaxley. "Obstructive uropathy – acute and chronic medical management". In: *World Journal of Nephrology* 12.1 (Jan. 2023), pp. 1–9. DOI: 10.5527/wjn.v12.i1.1. (Visited on 10/07/2024).

[248] Wayne Brisbane, Michael R. Bailey, and Mathew D. Sorensen. "An overview of kidney stone imaging techniques". eng. In: *Nature Reviews. Urology* 13.11 (Nov. 2016), pp. 654–662. DOI: 10.1038/nrurol.2016.154.

[249] Courtney A. Coursey et al. "ACR Appropriateness Criteria® Acute Onset Flank Pain–Suspicion of Stone Disease". en. In: *Ultrasound Quarterly* 28.3 (Sept. 2012), pp. 227–233. DOI: 10.1097/RUQ.0b013e3182625974. (Visited on 01/21/2025).

[250] Charles Y. C. Pak et al. "Ambulatory evaluation of nephrolithiasis: Classification, clinical presentation and diagnostic criteria". In: *The American Journal of Medicine* 69.1 (July 1980), pp. 19–30. DOI: 10.1016/0002-9343(80)90495-7. (Visited on 03/05/2024).

[251] Calyani Ganesan et al. "Analysis of Primary Hyperparathyroidism Screening Among US Veterans With Kidney Stones". en. In: *JAMA Surgery* 155.9 (Sept. 2020), p. 861. DOI: 10.1001/jamasurg.2020.2423. (Visited on 01/21/2025).

[252] Sumit Sharma et al. "Prevalence and Predictors of Primary Hyperparathyroidism Among Patients with Urolithiasis". en. In: *Endocrine Practice* 23.11 (Nov. 2017), pp. 1311–1315. DOI: 10.4158/EP171759.OR. (Visited on 01/21/2025).

[253] *Nephrocalcinosis or nephrolithiasis (Version 4.19)*. URL: https://panelapp.genomicsengland.co.uk/panels/149/ (visited on 02/10/2025).

[254] Robert Geraghty et al. "Role of Genetic Testing in Kidney Stone Disease: A Narrative Review". en. In: *Current Urology Reports* 25.12 (Dec. 2024), pp. 311–323. DOI: 10.1007/s11934-024-01225-5. (Visited on 02/05/2025).

[255] L. Borghi et al. “Urinary volume, water and recurrences in idiopathic calcium nephrolithiasis: a 5-year randomized prospective study”. eng. In: *The Journal of Urology* 155.3 (Mar. 1996), pp. 839–843.

[256] Antonio Nouvenne et al. “Effects of a low-salt diet on idiopathic hypercalciuria in calcium-oxalate stone formers: a 3-mo randomized controlled trial”. en. In: *The American Journal of Clinical Nutrition* 91.3 (Mar. 2010), pp. 565–570. DOI: 10.3945/ajcn.2009.28614. (Visited on 12/17/2024).

[257] J Shuster. “Soft drink consumption and urinary stone recurrence: A randomized prevention trial”. en. In: *Journal of Clinical Epidemiology* 45.8 (Aug. 1992), pp. 911–916. DOI: 10.1016/0895-4356(92)90074-W. (Visited on 12/17/2024).

[258] Bekir Aras et al. “Can lemon juice be an alternative to potassium citrate in the treatment of urinary calcium stones in patients with hypocitraturia? A prospective randomized study”. en. In: *Urological Research* 36.6 (Dec. 2008), pp. 313–317. DOI: 10.1007/s00240-008-0152-6. (Visited on 12/17/2024).

[259] P. Barcelo et al. “Randomized Double-Blind Study of Potassium Citrate in Idiopathic Hypocitraturic Calcium Nephrolithiasis”. en. In: *Journal of Urology* 150.6 (Dec. 1993), pp. 1761–1764. DOI: 10.1016/S0022-5347(17)35888-3. (Visited on 12/17/2024).

[260] Ural Oğuz and Ali Unsal. “The Efficacy of Medical Prophylaxis in Children with Calcium Oxalate Urolithiasis After Percutaneous Nephrolithotomy”. en. In: *Journal of Endourology* 27.1 (Jan. 2013), pp. 92–95. DOI: 10.1089/end.2012.0243. (Visited on 12/17/2024).

[261] Kemal Sarica et al. “Effect of Potassium Citrate Therapy on Stone Recurrence and Regrowth after Extracorporeal Shockwave Lithotripsy in Children”. en. In: *Journal of Endourology* 20.11 (Nov. 2006), pp. 875–879. DOI: 10.1089/end.2006.20.875. (Visited on 12/17/2024).

[262] Tarkan Soygür, Ayşegül Akbay, and Sadettin Küpeli. “Effect of Potassium Citrate Therapy on Stone Recurrence and Residual Fragments after Shockwave Lithotripsy in Lower Caliceal Calcium Oxalate Urolithiasis: A Randomized Controlled Trial”. en. In: *Journal of Endourology* 16.3 (Apr.

2002), pp. 149–152. DOI: 10.1089/089277902753716098. (Visited on 12/17/2024).

[263] Piyaratana Tosukhowong et al. “Citraturic, alkalinizing and antioxidative effects of limeade-based regimen in nephrolithiasis patients”. en. In: *Urological Research* 36.3-4 (Aug. 2008), pp. 149–155. DOI: 10.1007/s00240-008-0141-9. (Visited on 12/17/2024).

[264] Rebecca Phillips et al. “Citrate salts for preventing and treating calcium containing kidney stones in adults”. eng. In: *The Cochrane Database of Systematic Reviews* 2015.10 (Oct. 2015), p. CD010057. DOI: 10.1002/14651858.CD010057.pub2.

[265] C. Ahlstrand et al. “Clinical Experience with Long-term Bendroflumethiazide Treatment in Calcium Oxalate Stone Formers”. en. In: *British Journal of Urology* 56.3 (1984). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1464-410X.1984.tb05383.x, pp. 255–262. DOI: 10.1111/j.1464-410X.1984.tb05383.x. (Visited on 12/17/2024).

[266] M. Ala-Opas et al. “Unprocessed bran and intermittent thiazide therapy in prevention of recurrent urinary calcium stones”. eng. In: *Scandinavian Journal of Urology and Nephrology* 21.4 (1987), pp. 311–314. DOI: 10.3109/00365598709180789.

[267] M. Arrabal-Martín et al. “Extracorporeal renal lithotripsy: evolution of residual lithiasis treated with thiazides”. eng. In: *Urology* 68.5 (Nov. 2006), pp. 956–959. DOI: 10.1016/j.urology.2006.06.007.

[268] B. Baggio et al. “Hydrochlorothiazide and allopurinol vs. placebo on urinary excretion of stone promoters and inhibitors”. In: *Current Therapeutic Research-Clinical and Experimental* 34.1 (1983). ISBN: 0011-393X Publisher: EXCERPTA MEDICA INC 245 WEST 17TH STREET, NEW YORK, NY 10011, pp. 145–151.

[269] L. Borghi et al. “Randomized prospective study of a nonthiazide diuretic, indapamide, in preventing calcium stone recurrences”. eng. In: *Journal of Cardiovascular Pharmacology* 22 Suppl 6 (1993), S78–86.

[270] B. Ettinger et al. “Chlorthalidone reduces calcium oxalate calculous recurrence but magnesium hydroxide does not”. eng. In: *The Journal of Urology* 139.4 (Apr. 1988), pp. 679–684. DOI: 10.1016/s0022-5347(17)42599-7.

[271] K. Kohri et al. “Allopurinol and thiazide effects on new urinary stone formed after discontinued therapy in patients with urinary stones”. eng. In: *Urology* 36.4 (Oct. 1990), pp. 309–314. DOI: 10.1016/0090-4295(90)80235-f.

[272] E. Laerum. “Metabolic effects of thiazide versus placebo in patients under long-term treatment for recurrent urolithiasis”. eng. In: *Scandinavian Journal of Urology and Nephrology* 18.2 (1984), pp. 143–149. DOI: 10.3109/00365598409182182.

[273] Even Lærum and Stig Larsen. “Thiazide Prophylaxis of Urolithiasis”. en. In: *Acta Medica Scandinavica* 215.4 (1984). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0954-6820.1984.tb05023.x, pp. 383–389. DOI: 10.1111/j.0954-6820.1984.tb05023.x. (Visited on 12/17/2024).

[274] M. Ohkawa et al. “Thiazide treatment for calcium urolithiasis in patients with idiopathic hypercalciuria”. eng. In: *British Journal of Urology* 69.6 (June 1992), pp. 571–576. DOI: 10.1111/j.1464-410x.1992.tb15624.x.

[275] Dieter Scholz, Paul Otto Schwille, and Alfred Sigel. “Double-Blind Study with Thiazide in Recurrent Calcium Lithiasis”. In: *The Journal of Urology* 128.5 (Nov. 1982), pp. 903–907. DOI: 10.1016/S0022-5347(17)53269-3. (Visited on 03/05/2024).

[276] H. Wolf, P. Brocks, and C. Dahl. “Do thiazides prevent recurrent idiopathic renal calcium oxalate stones?” eng. In: *Proceedings of the European Dialysis and Transplant Association. European Dialysis and Transplant Association* 20 (1983), pp. 477–480.

[277] P. Brocks et al. “Do thiazides prevent recurrent idiopathic renal calcium stones?” eng. In: *Lancet (London, England)* 2.8238 (July 1981), pp. 124–125. DOI: 10.1016/s0140-6736(81)90302-0.

[278] Nasser A. Dhayat et al. “Hydrochlorothiazide and Prevention of Kidney-Stone Recurrence”. In: *New England Journal of Medicine* 388.9 (Mar. 2023). Publisher: Massachusetts Medical Society _eprint:

https://doi.org/10.1056/NEJMoa2209275, pp. 781–791. DOI: 10.1056/NEJMoa2209275. (Visited on 03/05/2024).

[279] Dor Golomb et al. "Long-Term Adherence to Medications in Secondary Prevention of Urinary Tract Stones". eng. In: *Journal of Endourology* 33.6 (June 2019), pp. 469–474. DOI: 10.1089/end.2019.0040.

[280] B. Ettinger et al. "Randomized trial of allopurinol in the prevention of calcium oxalate calculi". eng. In: *The New England Journal of Medicine* 315.22 (Nov. 1986), pp. 1386–1389. DOI: 10.1056/NEJM198611273152204.

[281] Dan-Feng Li et al. "Use of thiazide diuretics for the prevention of recurrent kidney calculi: a systematic review and meta-analysis". eng. In: *Journal of Translational Medicine* 18.1 (Feb. 2020), p. 106. DOI: 10.1186/s12967-020-02270-7.

[282] Margaret S. Pearle, Claus G. Roehrborn, and Charles Y.c. Pak. "Meta-Analysis of Randomized Trials for Medical Prevention of Calcium Oxalate Nephrolithiasis". In: *Journal of Endourology* 13.9 (Nov. 1999). Publisher: Mary Ann Liebert, Inc., publishers, pp. 679–685. DOI: 10.1089/end.1999.13.679. (Visited on 03/05/2024).

[283] Ibrahim Jour, Angela Lam, and Ben Turney. "Urological stone disease: a 5-year update of stone management using Hospital Episode Statistics". eng. In: *BJU international* 130.3 (Sept. 2022), pp. 364–369. DOI: 10.1111/bju.15728.

[284] Meng Zhe and Zheng Hang. "Nephrolithiasis as a risk factor of chronic kidney disease: a meta-analysis of cohort studies with 4,770,691 participants". eng. In: *Urolithiasis* 45.5 (Oct. 2017), pp. 441–448. DOI: 10.1007/s00240-016-0938-x.

[285] Kristina L. Penniston and Stephen Y. Nakada. "Development of an Instrument to Assess the Health Related Quality of Life of Kidney Stone Formers". en. In: *Journal of Urology* 189.3 (Mar. 2013), pp. 921–930. DOI: 10.1016/j.juro.2012.08.247. (Visited on 09/19/2024).

[286] M. J Khoury. "The emergence of epidemiology in the genomics age". en. In: *International Journal of Epidemiology* 33.5 (May 2004), pp. 936–944. DOI: 10.1093/ije/dyh278. (Visited on 10/02/2024).

[287] Eric S. Lander and Nicholas J. Schork. "Genetic Dissection of Complex Traits". en. In: *Focus* 4.3 (Aug. 2006), pp. 442–458. DOI: 10.1176/foc.4.3.442. (Visited on 10/03/2024).

[288] International Human Genome Sequencing Consortium et al. "Initial sequencing and analysis of the human genome". en. In: *Nature* 409.6822 (Feb. 2001), pp. 860–921. DOI: 10.1038/35057062. (Visited on 10/02/2024).

[289] †The International HapMap Consortium. "The International HapMap Project". en. In: *Nature* 426.6968 (Dec. 2003), pp. 789–796. DOI: 10.1038/nature02168. (Visited on 10/02/2024).

[290] The International HapMap 3 Consortium. "Integrating common and rare genetic variation in diverse human populations". en. In: *Nature* 467.7311 (Sept. 2010), pp. 52–58. DOI: 10.1038/nature09298. (Visited on 10/02/2024).

[291] Declan Butler. "Epidemiology set to get fast-track treatment". en. In: *Nature* 414.6860 (Nov. 2001), pp. 139–139. DOI: 10.1038/35102713. (Visited on 10/02/2024).

[292] Cathie Sudlow et al. "UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age". In: *PLoS medicine* 12.3 (2015). ISBN: 1549-1676 Publisher: Public Library of Science, e1001779.

[293] Andres Metspalu. "The Estonian Genome Project". en. In: *Drug Development Research* 62.2 (June 2004), pp. 97–101. DOI: 10.1002/ddr.10371. (Visited on 10/02/2024).

[294] Mitja I. Kurki et al. "FinnGen provides genetic insights from a well-phenotyped isolated population". en. In: *Nature* 613.7944 (Jan. 2023). Number: 7944 Publisher: Nature Publishing Group, pp. 508–518. DOI: 10.1038/s41586-022-05473-8. (Visited on 03/30/2023).

[295] Nicholas J. Timpson et al. "Genetic architecture: the shape of the genetic contribution to human traits and disease". en. In: *Nature Reviews Genetics* 19.2 (Feb. 2018). Publisher: Nature Publishing Group, pp. 110–124. DOI: 10.1038/nrg.2017.101. (Visited on 12/19/2024).

[296] Emil Uffelmann et al. "Genome-wide association studies". en. In: *Nature Reviews Methods Primers* 1.1 (Aug. 2021). Number: 1 Publisher: Nature

Publishing Group, pp. 1–21. DOI: 10.1038/s43586-021-00056-9. (Visited on 02/23/2024).

[297] Andries T. Marees et al. "A tutorial on conducting genome-wide association studies: Quality control and statistical analysis". en. In: *International Journal of Methods in Psychiatric Research* 27.2 (June 2018), e1608. DOI: 10.1002/mpr.1608. (Visited on 10/02/2024).

[298] Nilanjan Chatterjee, Jianxin Shi, and Montserrat García-Closas. "Developing and evaluating polygenic risk prediction models for stratified disease prevention". en. In: *Nature Reviews Genetics* 17.7 (July 2016). Number: 7 Publisher: Nature Publishing Group, pp. 392–406. DOI: 10.1038/nrg.2016.27. (Visited on 01/06/2023).

[299] Alicia R. Martin et al. "Clinical use of current polygenic risk scores may exacerbate health disparities". en. In: *Nature Genetics* 51.4 (Apr. 2019), pp. 584–591. DOI: 10.1038/s41588-019-0379-x. (Visited on 11/28/2023).

[300] Catherine E Lovegrove et al. "Causal inference in health and disease: a review of the principles and applications of Mendelian randomization". In: *Journal of Bone and Mineral Research* (Aug. 2024), zjae136. DOI: 10.1093/jbmr/zjae136. (Visited on 09/24/2024).

[301] Stephen Burgess and Jeremy A. Labrecque. "Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates". en. In: *European Journal of Epidemiology* 33.10 (Oct. 2018), pp. 947–952. DOI: 10.1007/s10654-018-0424-6. (Visited on 04/04/2023).

[302] Christopher N. Foley et al. "A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits". en. In: *Nature Communications* 12.1 (Feb. 2021). Number: 1 Publisher: Nature Publishing Group, p. 764. DOI: 10.1038/s41467-020-20885-8. (Visited on 12/06/2022).

[303] Olivier J. Wouters, Martin McKee, and Jeroen Luyten. "Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018". en. In: *JAMA* 323.9 (Mar. 2020), p. 844. DOI: 10.1001/jama.2020.1166. (Visited on 01/20/2025).

[304] Katarzyna Smietana, Marcin Siatkowski, and Martin Møller. "Trends in clinical success rates". en. In: *Nature Reviews Drug Discovery* 15.6 (June 2016), pp. 379–380. DOI: 10.1038/nrd.2016.85. (Visited on 01/20/2025).

[305] Isabella Wy Mak, Nathan Evaniew, and Michelle Ghert. "Lost in translation: animal models and clinical trials in cancer treatment". eng. In: *American Journal of Translational Research* 6.2 (2014), pp. 114–118.

[306] Junhee Seok et al. "Genomic responses in mouse models poorly mimic human inflammatory diseases". en. In: *Proceedings of the National Academy of Sciences* 110.9 (Feb. 2013), pp. 3507–3512. DOI: 10.1073/pnas.1222878110. (Visited on 01/20/2025).

[307] Tal Shay et al. "Conservation and divergence in the transcriptional programs of the human and mouse immune systems". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 110.8 (Feb. 2013), pp. 2946–2951. DOI: 10.1073/pnas.1222738110.

[308] Matthew R Nelson et al. "The support of human genetic evidence for approved drug indications". en. In: *Nature Genetics* 47.8 (Aug. 2015), pp. 856–860. DOI: 10.1038/ng.3314. (Visited on 01/20/2025).

[309] Kyuto Sonehara and Yukinori Okada. "Genomics-driven drug discovery based on disease-susceptibility genes". In: *Inflammation and Regeneration* 41.1 (Mar. 2021), p. 8. DOI: 10.1186/s41232-021-00158-7. (Visited on 01/20/2025).

[310] Abdel Abdellaoui et al. "15 years of GWAS discovery: Realizing the promise". In: *American Journal of Human Genetics* 110.2 (Feb. 2023), pp. 179–194. DOI: 10.1016/j.ajhg.2022.12.011. (Visited on 09/24/2024).

[311] Ruth J. F. Loos. "15 years of genome-wide association studies and no signs of slowing down". en. In: *Nature Communications* 11.1 (Nov. 2020). Number: 1 Publisher: Nature Publishing Group, p. 5900. DOI: 10.1038/s41467-020-19653-5. (Visited on 08/11/2023).

[312] Ruth J. F. Loos and Giles S. H. Yeo. "The genetics of obesity: from discovery to biology". en. In: *Nature Reviews Genetics* 23.2 (Feb. 2022). Number: 2 Publisher: Nature Publishing Group, pp. 120–133. DOI: 10.1038/s41576-021-00414-z. (Visited on 08/08/2023).

[313] Sara L. Pulit et al. “Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry”. In: *Human molecular genetics* 28.1 (2019). ISBN: 0964-6906 Publisher: Oxford University Press, pp. 166–174.

[314] I Sadaf Farooqi and Stephen O’Rahilly. “20 YEARS OF LEPTIN: Human disorders of leptin action”. In: *Journal of Endocrinology* 223.1 (Oct. 2014), T63–T70. DOI: 10.1530/JOE-14-0480. (Visited on 01/06/2025).

[315] John P.H. Wilding et al. “Once-Weekly Semaglutide in Adults with Overweight or Obesity”. en. In: *New England Journal of Medicine* 384.11 (Mar. 2021), pp. 989–1002. DOI: 10.1056/NEJMoa2032183. (Visited on 10/04/2024).

[316] Xavier Pi-Sunyer et al. “A Randomized, Controlled Trial of 3.0 mg of Liraglutide in Weight Management”. In: *New England Journal of Medicine* 373.1 (July 2015). Publisher: Massachusetts Medical Society _eprint: https://www.nejm.org/doi/pdf/10.1056/NEJMoa1411892, pp. 11–22. DOI: 10.1056/NEJMoa1411892. (Visited on 01/06/2025).

[317] NICE (National Institute for Health and Care Excellence). *Liraglutide for managing overweight and obesity*. eng. Publisher: NICE. Dec. 2020. URL: https://www.nice.org.uk/guidance/ta664 (visited on 01/06/2025).

[318] NICE (National Institute for Health and Care Excellence). *Semaglutide for managing overweight and obesity*. eng. Publisher: NICE. Mar. 2023. URL: https://www.nice.org.uk/guidance/ta875 (visited on 01/06/2025).

[319] Mark I. McCarthy et al. “Genome-wide association studies for complex traits: consensus, uncertainty and challenges”. en. In: *Nature Reviews Genetics* 9.5 (May 2008). Publisher: Nature Publishing Group, pp. 356–369. DOI: 10.1038/nrg2344. (Visited on 09/24/2024).

[320] Peter M. Visscher et al. “10 Years of GWAS Discovery: Biology, Function, and Translation”. In: *American Journal of Human Genetics* 101.1 (July 2017), pp. 5–22. DOI: 10.1016/j.ajhg.2017.06.005. (Visited on 08/11/2023).

[321] Joshua D. Backman et al. “Exome sequencing and analysis of 454,787 UK Biobank participants”. en. In: *Nature* 599.7886 (Nov. 2021). Number: 7886

Publisher: Nature Publishing Group, pp. 628–634. DOI: 10.1038/s41586-021-04103-z. (Visited on 06/13/2023).

[322] Montgomery Slatkin. "Linkage disequilibrium - understanding the evolutionary past and mapping the medical future". In: *Nature reviews. Genetics* 9.6 (June 2008), pp. 477–485. DOI: 10.1038/nrg2361. (Visited on 11/28/2023).

[323] Alkes L. Price et al. "New approaches to population stratification in genome-wide association studies". en. In: *Nature Reviews Genetics* 11.7 (July 2010), pp. 459–463. DOI: 10.1038/nrg2813. (Visited on 10/02/2024).

[324] B. Devlin and Kathryn Roeder. "Genomic Control for Association Studies". en. In: *Biometrics* 55.4 (Dec. 1999), pp. 997–1004. DOI: 10.1111/j.0006-341X.1999.00997.x. (Visited on 10/02/2024).

[325] David G. Clayton et al. "Population structure, differential bias and genomic control in a large-scale, case-control association study". en. In: *Nature Genetics* 37.11 (Nov. 2005). Publisher: Nature Publishing Group, pp. 1243–1246. DOI: 10.1038/ng1653. (Visited on 10/02/2024).

[326] Hyun Min Kang et al. "Variance component model to account for sample structure in genome-wide association studies". en. In: *Nature Genetics* 42.4 (Apr. 2010). Publisher: Nature Publishing Group, pp. 348–354. DOI: 10.1038/ng.548. (Visited on 10/02/2024).

[327] G. McVean. "Linkage Disequilibrium, Recombination and Selection". en. In: *Handbook of Statistical Genetics*. Ed. by D. J. Balding, M. Bishop, and C. Cannings. 1st ed. Wiley, Aug. 2007, pp. 909–944. DOI: 10.1002/9780470061619.ch27. (Visited on 10/03/2024).

[328] Gregor Mendel. "Experiments in plant hybridization (1865)". In: *Verhandlungen des naturforschenden Vereins Brünn.) Available online: www. mendelweb. org/Mendel. html (accessed on 1 January 2013)* (1996).

[329] Ilona Miko. "Gregor Mendel and the principles of inheritance". In: *Nature Education* 1.1 (2008), p. 134.

[330] Sagiv Shifman et al. "Linkage disequilibrium patterns of the human genome across populations". In: *Human Molecular Genetics* 12.7 (Apr. 2003), pp. 771–776. DOI: 10.1093/hmg/ddg088. (Visited on 10/02/2024).

[331] William Astle and David J. Balding. "Population Structure and Cryptic Relatedness in Genetic Association Studies". In: *Statistical Science* 24.4 (Nov. 2009). DOI: 10.1214/09-STS307. (Visited on 10/03/2024).

[332] Cristopher V. Van Hout et al. "Exome sequencing and characterization of 49,960 individuals in the UK Biobank". en. In: *Nature* 586.7831 (Oct. 2020). Number: 7831 Publisher: Nature Publishing Group, pp. 749–756. DOI: 10.1038/s41586-020-2853-0. (Visited on 06/13/2023).

[333] Elizabeth T. Cirulli et al. "Genome-wide rare variant analysis for thousands of phenotypes in over 70,000 exomes from two cohorts". en. In: *Nature Communications* 11.1 (Jan. 2020). Publisher: Nature Publishing Group, p. 542. DOI: 10.1038/s41467-020-14288-y. (Visited on 12/12/2024).

[334] Sean J. Jurgens et al. "Analysis of rare genetic variation underlying cardiometabolic diseases and traits among 200,000 individuals in the UK Biobank". en. In: *Nature Genetics* 54.3 (Mar. 2022). Publisher: Nature Publishing Group, pp. 240–250. DOI: 10.1038/s41588-021-01011-w. (Visited on 12/12/2024).

[335] Joseph D. Szustakowski et al. "Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank". en. In: *Nature Genetics* 53.7 (July 2021). Number: 7 Publisher: Nature Publishing Group, pp. 942–948. DOI: 10.1038/s41588-021-00885-0. (Visited on 06/13/2023).

[336] Alexander I. Young. "Solving the missing heritability problem". en. In: *PLOS Genetics* 15.6 (June 2019). Publisher: Public Library of Science, e1008222. DOI: 10.1371/journal.pgen.1008222. (Visited on 12/12/2024).

[337] Teri A. Manolio et al. "Finding the missing heritability of complex diseases". en. In: *Nature* 461.7265 (Oct. 2009). Publisher: Nature Publishing Group, pp. 747–753. DOI: 10.1038/nature08494. (Visited on 12/12/2024).

[338] Pierrick Wainschtein et al. "Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data". en. In: *Nature Genetics* 54.3 (Mar. 2022). Publisher: Nature Publishing Group, pp. 263–273. DOI: 10.1038/s41588-021-00997-7. (Visited on 12/12/2024).

[339] Jonathan C. Cohen et al. "Sequence Variations in *PCSK9,* Low LDL, and Protection against Coronary Heart Disease". en. In: *New England Journal of*

*Medicine* 354.12 (Mar. 2006), pp. 1264–1272. DOI: 10.1056/NEJMoa054013. (Visited on 12/12/2024).

[340] Noura S. Abul-Husn et al. "A Protein-Truncating *HSD17B13* Variant and Protection from Chronic Liver Disease". en. In: *New England Journal of Medicine* 378.12 (Mar. 2018), pp. 1096–1106. DOI: 10.1056/NEJMoa1712191. (Visited on 12/12/2024).

[341] Frederick E. Dewey et al. "Genetic and Pharmacologic Inactivation of ANGPTL3 and Cardiovascular Disease". en. In: *New England Journal of Medicine* 377.3 (July 2017), pp. 211–221. DOI: 10.1056/NEJMoa1612790. (Visited on 12/12/2024).

[342] Robert A. Scott et al. "A genomic approach to therapeutic target validation identifies a glucose-lowering *GLP1R* variant protective for coronary heart disease". en. In: *Science Translational Medicine* 8.341 (June 2016). DOI: 10.1126/scitranslmed.aad3744. (Visited on 12/12/2024).

[343] Noura S. Abul-Husn et al. "Genetic identification of familial hypercholesterolemia within a single U.S. health care system". en. In: *Science* 354.6319 (Dec. 2016), aaf7000. DOI: 10.1126/science.aaf7000. (Visited on 12/12/2024).

[344] Kandamurugu Manickam et al. "Exome Sequencing–Based Screening for *BRCA1/2* Expected Pathogenic Variants Among Adult Biobank Participants". en. In: *JAMA Network Open* 1.5 (Sept. 2018), e182140. DOI: 10.1001/jamanetworkopen.2018.2140. (Visited on 12/12/2024).

[345] Clare Bycroft et al. "The UK Biobank resource with deep phenotyping and genomic data". en. In: *Nature* 562.7726 (Oct. 2018). Number: 7726 Publisher: Nature Publishing Group, pp. 203–209. DOI: 10.1038/s41586-018-0579-z. (Visited on 06/13/2023).

[346] Christian Fuchsberger et al. "The genetic architecture of type 2 diabetes". en. In: *Nature* 536.7614 (Aug. 2016). Publisher: Nature Publishing Group, pp. 41–47. DOI: 10.1038/nature18642. (Visited on 12/12/2024).

[347] Paul L. Auer et al. "Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits". en. In: *Nature Genetics*

46.6 (June 2014). Publisher: Nature Publishing Group, pp. 629–634. DOI: 10.1038/ng.2962. (Visited on 12/12/2024).

[348] Nathan Pankratz et al. "Meta-analysis of rare and common exome chip variants identifies S1PR4 and other loci influencing blood cell traits". en. In: *Nature Genetics* 48.8 (Aug. 2016). Publisher: Nature Publishing Group, pp. 867–876. DOI: 10.1038/ng.3607. (Visited on 12/12/2024).

[349] John D. Eicher et al. "Platelet-Related Variants Identified by Exomechip Meta-analysis in 157,293 Individuals". English. In: *The American Journal of Human Genetics* 99.1 (July 2016). Publisher: Elsevier, pp. 40–55. DOI: 10.1016/j.ajhg.2016.05.005. (Visited on 12/12/2024).

[350] Gina M. Peloso et al. "Association of Low-Frequency and Rare Coding-Sequence Variants with Blood Lipids and Coronary Heart Disease in 56,000 Whites and Blacks". English. In: *The American Journal of Human Genetics* 94.2 (Feb. 2014). Publisher: Elsevier, pp. 223–232. DOI: 10.1016/j.ajhg.2014.01.009. (Visited on 12/12/2024).

[351] Praveen Surendran et al. "Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension". en. In: *Nature Genetics* 48.10 (Oct. 2016). Publisher: Nature Publishing Group, pp. 1151–1161. DOI: 10.1038/ng.3654. (Visited on 12/12/2024).

[352] Malay Kumar Basu et al. "Exome Sequencing Identifies Abnormalities in Glycosylation and ANKRD36C in Patients with Immune-Mediated Thrombotic Thrombocytopenic Purpura". en. In: *Thrombosis and Haemostasis* 121.04 (Apr. 2021), pp. 506–517. DOI: 10.1055/s-0040-1719030. (Visited on 12/12/2024).

[353] Xingchen Zhou et al. "Whole Exome Sequencing in Psoriasis Patients Contributes to Studies of Acitretin Treatment Difference". en. In: *International Journal of Molecular Sciences* 18.2 (Jan. 2017), p. 295. DOI: 10.3390/ijms18020295. (Visited on 12/12/2024).

[354] Eirini Marouli et al. "Rare and low-frequency coding variants alter human adult height". en. In: *Nature* 542.7640 (Feb. 2017). Publisher: Nature Publishing Group, pp. 186–190. DOI: 10.1038/nature21039. (Visited on 12/12/2024).

[355] Valérie Turcot et al. “Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity”. en. In: *Nature Genetics* 50.1 (Jan. 2018). Publisher: Nature Publishing Group, pp. 26–41. DOI: 10.1038/s41588-017-0011-x. (Visited on 12/12/2024).

[356] Alexander R. Chang et al. “Exome Sequencing of a Clinical Population for Autosomal Dominant Polycystic Kidney Disease”. In: *JAMA* 328.24 (Dec. 2022), pp. 2412–2421. DOI: 10.1001/jama.2022.22847. (Visited on 11/12/2024).

[357] Kushani Jayasinghe et al. “Cost-Effectiveness of Targeted Exome Analysis as a Diagnostic Test in Glomerular Diseases”. English. In: *Kidney International Reports* 6.11 (Nov. 2021). Publisher: Elsevier, pp. 2850–2861. DOI: 10.1016/j.ekir.2021.08.028. (Visited on 11/12/2024).

[358] Matthias Wuttke et al. “Imputation-powered whole-exome analysis identifies genes associated with kidney function and disease in the UK Biobank”. en. In: *Nature Communications* 14.1 (Mar. 2023). Number: 1 Publisher: Nature Publishing Group, p. 1287. DOI: 10.1038/s41467-023-36864-8. (Visited on 03/30/2023).

[359] Mark D. Elliott et al. “Clinical and Genetic Characteristics of CKD Patients with High-Risk APOL1 Genotypes”. en. In: *Journal of the American Society of Nephrology : JASN* 34.5 (Feb. 2023), p. 909. DOI: 10.1681/ASN.0000000000000094. (Visited on 11/12/2024).

[360] George Davey Smith and Shah Ebrahim. “‘Mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease?*”. In: *International Journal of Epidemiology* 32.1 (Feb. 2003), pp. 1–22. DOI: 10.1093/ije/dyg070. (Visited on 03/05/2024).

[361] Zoe Fewell, George Davey Smith, and Jonathan A. C. Sterne. “The Impact of Residual and Unmeasured Confounding in Epidemiologic Studies: A Simulation Study”. In: *American Journal of Epidemiology* 166.6 (Sept. 2007), pp. 646–655. DOI: 10.1093/aje/kwm165. (Visited on 11/28/2023).

[362] Jan P. Vandenbroucke et al. “Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration”. In:

*International Journal of Surgery* 12.12 (Dec. 2014), pp. 1500–1524. DOI: 10.1016/j.ijsu.2014.07.014. (Visited on 11/28/2023).

[363] George Davey Smith and Shah Ebrahim. "Data dredging, bias, or confounding: They can all get you into the BMJ and the Friday papers". en. In: *BMJ* 325.7378 (Dec. 2002). Publisher: British Medical Journal Publishing Group Section: Editorial, pp. 1437–1438. DOI: 10.1136/bmj.325.7378.1437. (Visited on 11/28/2023).

[364] Emily C. Zabor, Alexander M. Kaizer, and Brian P. Hobbs. "Randomized Controlled Trials". In: *Chest* 158.1 Suppl (July 2020), S79–S87. DOI: 10.1016/j.chest.2020.03.013. (Visited on 11/28/2023).

[365] Eduardo Hariton and Joseph J. Locascio. "Randomised controlled trials—the gold standard for effectiveness research". en. In: *BJOG : an international journal of obstetrics and gynaecology* 125.13 (Dec. 2018). Publisher: NIH Public Access, p. 1716. DOI: 10.1111/1471-0528.15199. (Visited on 11/28/2023).

[366] Maria K Sobczyk et al. "Systematic comparison of Mendelian randomisation studies and randomised controlled trials using electronic databases". In: *BMJ Open* 13.9 (Sept. 2023), e072087. DOI: 10.1136/bmjopen-2023-072087. (Visited on 11/28/2023).

[367] Sonja A. Swanson et al. "Nature as a Trialist?: Deconstructing the Analogy Between Mendelian Randomization and Randomized Trials". en. In: *Epidemiology* 28.5 (Sept. 2017), pp. 653–659. DOI: 10.1097/EDE.0000000000000699. (Visited on 11/28/2023).

[368] George Davey Smith et al. "Mendel's laws, Mendelian randomization and causal inference in observational data: substantive and nomenclatural issues". en. In: *European Journal of Epidemiology* 35.2 (Feb. 2020), pp. 99–111. DOI: 10.1007/s10654-020-00622-7. (Visited on 11/28/2023).

[369] George Davey Smith and Gibran Hemani. "Mendelian randomization: genetic anchors for causal inference in epidemiological studies". In: *Human Molecular Genetics* 23.R1 (Sept. 2014), R89–R98. DOI: 10.1093/hmg/ddu328. (Visited on 08/08/2023).

[370] Stephen R. Cole and Constantine E. Frangakis. "The Consistency Statement in Causal Inference: A Definition or an Assumption?" en. In: *Epidemiology* 20.1 (Jan. 2009), pp. 3–5. DOI: 10.1097/EDE.0b013e31818ef366. (Visited on 09/30/2024).

[371] George Davey Smith. "Epigenesis for epidemiologists: does evo-devo have implications for population health research and practice?" In: *International Journal of Epidemiology* 41.1 (Feb. 2012), pp. 236–247. DOI: 10.1093/ije/dys016. (Visited on 11/28/2023).

[372] Jeremy A Labrecque and Sonja A Swanson. "Interpretation and Potential Biases of Mendelian Randomization Estimates With Time-Varying Exposures". In: *American Journal of Epidemiology* 188.1 (Jan. 2019), pp. 231–238. DOI: 10.1093/aje/kwy204. (Visited on 11/28/2023).

[373] Tom G Richardson et al. "Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study". In: *The BMJ* 369 (May 2020), p. m1203. DOI: 10.1136/bmj.m1203. (Visited on 01/22/2024).

[374] Qingyuan Zhao et al. "Two-sample instrumental variable analyses using heterogeneous samples". In: *Statistical Science* 34.2 (2019). ISBN: 0883-4237 Publisher: JSTOR, pp. 317–333.

[375] Eleanor Sanderson et al. "Mendelian randomization". In: *Nature reviews. Methods primers* 2 (Feb. 2022), p. 6. DOI: 10.1038/s43586-021-00092-5. (Visited on 11/28/2023).

[376] Stephen Burgess et al. "Guidelines for performing Mendelian randomization investigations: update for summer 2023". In: *Wellcome Open Research* 4 (Aug. 2023), p. 186. DOI: 10.12688/wellcomeopenres.15555.3. (Visited on 11/28/2023).

[377] Lina Chen et al. "Alcohol intake and blood pressure: a systematic review implementing a Mendelian randomization approach". eng. In: *PLoS medicine* 5.3 (Mar. 2008), e52. DOI: 10.1371/journal.pmed.0050052.

[378] Stephen Burgess and Simon G. Thompson. "Use of allele scores as instrumental variables for Mendelian randomization". In: *International*

*journal of epidemiology* 42.4 (2013). ISBN: 1464-3685 Publisher: Oxford University Press, pp. 1134–1144.

[379] Jess Buxton and Jon Turney. *The rough guide to genes & cloning*. Penguin, 2007.

[380] Stephen Burgess, Simon G. Thompson, and CRP CHD Genetics Collaboration. "Avoiding bias from weak instruments in Mendelian randomization studies". eng. In: *International Journal of Epidemiology* 40.3 (June 2011), pp. 755–764. DOI: 10.1093/ije/dyr036.

[381] Stephen Burgess, Neil M. Davies, and Simon G. Thompson. "Bias due to participant overlap in two-sample Mendelian randomization". en. In: *Genetic Epidemiology* 40.7 (Nov. 2016), pp. 597–608. DOI: 10.1002/gepi.21998. (Visited on 04/04/2023).

[382] Philip C Haycock et al. "Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies1". In: *The American Journal of Clinical Nutrition* 103.4 (Apr. 2016), pp. 965–978. DOI: 10.3945/ajcn.115.118216. (Visited on 11/28/2023).

[383] Stephen Burgess, Dylan S Small, and Simon G Thompson. "A review of instrumental variable estimators for Mendelian randomization". en. In: *Statistical Methods in Medical Research* 26.5 (Oct. 2017). Publisher: SAGE Publications Ltd STM, pp. 2333–2355. DOI: 10.1177/0962280215597579. (Visited on 12/13/2023).

[384] Gibran Hemani, Jack Bowden, and George Davey Smith. "Evaluating the potential role of pleiotropy in Mendelian randomization studies". In: *Human Molecular Genetics* 27.R2 (Aug. 2018), R195–R208. DOI: 10.1093/hmg/ddy163. (Visited on 01/22/2024).

[385] Gibran Hemani et al. "The MR-Base platform supports systematic causal inference across the human phenome". en. In: *eLife* 7 (May 2018), e34408. DOI: 10.7554/eLife.34408. (Visited on 04/04/2023).

[386] Jack Bowden et al. "Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic". en. In: *International Journal of Epidemiology* (Sept. 2016), dyw220. DOI: 10.1093/ije/dyw220. (Visited on 06/16/2024).

[387] Jack Bowden, George Davey Smith, and Stephen Burgess. "Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression". In: *International Journal of Epidemiology* 44.2 (Apr. 2015), pp. 512–525. DOI: 10.1093/ije/dyv080. (Visited on 12/13/2023).

[388] Stephen Burgess and Simon G. Thompson. "Interpreting findings from Mendelian randomization using the MR-Egger method". In: *European Journal of Epidemiology* 32.5 (2017), pp. 377–389. DOI: 10.1007/s10654-017-0255-x. (Visited on 12/13/2023).

[389] Jack Bowden et al. "Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator". eng. In: *Genetic Epidemiology* 40.4 (May 2016), pp. 304–314. DOI: 10.1002/gepi.21965.

[390] Fernando Pires Hartwig, George Davey Smith, and Jack Bowden. "Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption". In: *International Journal of Epidemiology* 46.6 (Dec. 2017), pp. 1985–1998. DOI: 10.1093/ije/dyx102. (Visited on 12/13/2023).

[391] S. Burgess and S. G. Thompson. "Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects". en. In: *American Journal of Epidemiology* 181.4 (Feb. 2015), pp. 251–260. DOI: 10.1093/aje/kwu283. (Visited on 12/05/2023).

[392] Eleanor Sanderson et al. "An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings". en. In: *International Journal of Epidemiology* 48.3 (June 2019), pp. 713–727. DOI: 10.1093/ije/dyy262. (Visited on 12/05/2023).

[393] Alice R. Carter et al. "Mendelian randomisation for mediation analysis: current methods and challenges for implementation". en. In: *European Journal of Epidemiology* 36.5 (May 2021), pp. 465–478. DOI: 10.1007/s10654-021-00757-1. (Visited on 08/08/2023).

[394] Stephen Burgess et al. "Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways".

In: *International Journal of Epidemiology* 44.2 (Apr. 2015), pp. 484–495. DOI: 10.1093/ije/dyu176. (Visited on 12/09/2023).

[395] Iyas Daghlas and Dipender Gill. "Mendelian randomization as a tool to inform drug development using human genetics". en. In: *Cambridge Prisms: Precision Medicine* 1 (2023), e16. DOI: 10.1017/pcm.2023.5. (Visited on 12/09/2023).

[396] Amand F. Schmidt et al. "Genetic drug target validation using Mendelian randomisation". en. In: *Nature Communications* 11.1 (June 2020). Number: 1 Publisher: Nature Publishing Group, p. 3255. DOI: 10.1038/s41467-020-16969-0. (Visited on 12/09/2023).

[397] Verena Zuber et al. "Combining evidence from Mendelian randomization and colocalization: Review and comparison of approaches". en. In: *The American Journal of Human Genetics* 109.5 (May 2022), pp. 767–782. DOI: 10.1016/j.ajhg.2022.04.001. (Visited on 11/21/2022).

[398] Brian A Ference. "How to use Mendelian randomization to anticipate the results of randomized trials". en. In: *European Heart Journal* 39.5 (Feb. 2018), pp. 360–362. DOI: 10.1093/eurheartj/ehx462. (Visited on 11/28/2023).

[399] Daniel B. Rosoff et al. "Evaluating the Cardiovascular Impact of Genetically Proxied PCSK9 and HMGCR Inhibition in East Asian and European Populations: A Drug-Target Mendelian Randomization Study". en. In: *Circulation: Genomic and Precision Medicine* 17.1 (Feb. 2024). DOI: 10.1161/CIRCGEN.122.004224. (Visited on 10/04/2024).

[400] Marc S. Sabatine et al. "Cardiovascular safety and efficacy of the PCSK9 inhibitor evolocumab in patients with and without diabetes and the effect of evolocumab on glycaemia and risk of new-onset diabetes: a prespecified analysis of the FOURIER randomised controlled trial". In: *The lancet Diabetes & endocrinology* 5.12 (2017). ISBN: 2213-8587 Publisher: Elsevier, pp. 941–950.

[401] Mary D Fortune et al. "Statistical colocalization of genetic risk variants for related autoimmune diseases in the context of common controls". en. In: *Nature Genetics* 47.7 (July 2015), pp. 839–846. DOI: 10.1038/ng.3330. (Visited on 10/03/2024).

[402] Marijana Vujkovic et al. “Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis”. en. In: *Nature Genetics* 52.7 (July 2020). Number: 7 Publisher: Nature Publishing Group, pp. 680–691. DOI: 10.1038/s41588-020-0637-y. (Visited on 11/28/2022).

[403] Wenan Chen et al. “Fine Mapping Causal Variants with an Approximate Bayesian Method Using Marginal Test Statistics”. en. In: *Genetics* 200.3 (July 2015), pp. 719–736. DOI: 10.1534/genetics.115.176107. (Visited on 10/03/2024).

[404] Vincent Plagnol et al. “Statistical independence of the colocalized association signals for type 1 diabetes and RPS26 gene expression on chromosome 12q13”. In: *Biostatistics* 10.2 (2009). ISBN: 1468-4357 Publisher: Oxford University Press, pp. 327–334.

[405] Chris Wallace et al. “Statistical colocalization of monocyte gene expression and genetic risk variants for type 1 diabetes”. en. In: *Human Molecular Genetics* 21.12 (June 2012), pp. 2815–2824. DOI: 10.1093/hmg/dds098. (Visited on 10/04/2024).

[406] Chris Wallace. “Statistical Testing of Shared Genetic Control for Potentially Related Traits”. en. In: *Genetic Epidemiology* 37.8 (Dec. 2013), pp. 802–813. DOI: 10.1002/gepi.21765. (Visited on 10/04/2024).

[407] Claudia Giambartolomei et al. “Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics”. en. In: *PLOS Genetics* 10.5 (May 2014). Publisher: Public Library of Science, e1004383. DOI: 10.1371/journal.pgen.1004383. (Visited on 11/21/2022).

[408] Jon Wakefield. “Bayes factors for genome-wide association studies: comparison with $P$ -values”. en. In: *Genetic Epidemiology* 33.1 (Jan. 2009), pp. 79–86. DOI: 10.1002/gepi.20359. (Visited on 10/04/2024).

[409] Chris Wallace. “Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses”. en. In: *PLOS Genetics* 16.4 (Apr. 2020). Ed. by Michael P. Epstein, e1008720. DOI: 10.1371/journal.pgen.1008720. (Visited on 10/04/2024).

[410] Chris Wallace. “A more accurate method for colocalisation analysis allowing for multiple causal variants”. en. In: *PLOS Genetics* 17.9 (Sept. 2021). Ed. by Heather J. Cordell, e1009440. DOI: 10.1371/journal.pgen.1009440. (Visited on 10/04/2024).

[411] Gao Wang et al. “A Simple New Approach to Variable Selection in Regression, with Application to Genetic Fine Mapping”. en. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82.5 (Dec. 2020), pp. 1273–1300. DOI: 10.1111/rssb.12388. (Visited on 10/04/2024).

[412] Alev Baysoy et al. “The technological landscape and applications of single-cell multi-omics”. en. In: *Nature Reviews Molecular Cell Biology* 24.10 (Oct. 2023). Publisher: Nature Publishing Group, pp. 695–713. DOI: 10.1038/s41580-023-00615-w. (Visited on 11/28/2024).

[413] Gabriel B. Loeb et al. “Variants in tubule epithelial regulatory elements mediate most heritable differences in human kidney function”. en. In: *Nature Genetics* 56.10 (Oct. 2024). Publisher: Nature Publishing Group, pp. 2078–2092. DOI: 10.1038/s41588-024-01904-6. (Visited on 12/10/2024).

[414] Darren A. Cusanovich et al. “A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility”. In: *Cell* 174.5 (Aug. 2018). Publisher: Elsevier, 1309–1324.e18. DOI: 10.1016/j.cell.2018.06.052. (Visited on 12/10/2024).

[415] Sandy L. Klemm, Zohar Shipony, and William J. Greenleaf. “Chromatin accessibility and the regulatory epigenome”. en. In: *Nature Reviews Genetics* 20.4 (Apr. 2019). Publisher: Nature Publishing Group, pp. 207–220. DOI: 10.1038/s41576-018-0089-8. (Visited on 12/10/2024).

[416] Parker C. Wilson, Nicolas Ledru, and Benjamin D. Humphreys. “Epigenomics and the kidney:” en. In: *Current Opinion in Nephrology and Hypertension* 29.3 (May 2020), pp. 280–285. DOI: 10.1097/MNH.0000000000000602. (Visited on 12/10/2024).

[417] Yoshiharu Muto et al. “Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney”. en. In:

*Nature Communications* 12.1 (Apr. 2021), p. 2190. DOI: 10.1038/s41467-021-22368-w. (Visited on 09/09/2024).

[418] A. Marieke Oudelaar and Douglas R. Higgs. "The relationship between genome structure and function". en. In: *Nature Reviews Genetics* 22.3 (Mar. 2021). Publisher: Nature Publishing Group, pp. 154–168. DOI: 10.1038/s41576-020-00303-x. (Visited on 12/09/2024).

[419] Boyan Bonev and Giacomo Cavalli. "Organization and function of the 3D genome". en. In: *Nature Reviews Genetics* 17.11 (Nov. 2016). Publisher: Nature Publishing Group, pp. 661–678. DOI: 10.1038/nrg.2016.112. (Visited on 12/09/2024).

[420] Jim R. Hughes et al. "Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment". en. In: *Nature Genetics* 46.2 (Feb. 2014). Publisher: Nature Publishing Group, pp. 205–212. DOI: 10.1038/ng.2871. (Visited on 12/09/2024).

[421] Elzo de Wit and Wouter de Laat. "A decade of 3C technologies: insights into nuclear organization". eng. In: *Genes & Development* 26.1 (Jan. 2012), pp. 11–24. DOI: 10.1101/gad.179804.111.

[422] Job Dekker et al. "Capturing chromosome conformation". eng. In: *Science (New York, N.Y.)* 295.5558 (Feb. 2002), pp. 1306–1311. DOI: 10.1126/science.1067799.

[423] Damien J. Downes et al. "Capture-C: a modular and flexible approach for high-resolution chromosome conformation capture". en. In: *Nature Protocols* 17.2 (Feb. 2022). Publisher: Nature Publishing Group, pp. 445–475. DOI: 10.1038/s41596-021-00651-w. (Visited on 12/09/2024).

[424] Haikuo Li, Dian Li, and Benjamin D. Humphreys. "Chromatin conformation and histone modification profiling across human kidney anatomic regions". en. In: *Scientific Data* 11.1 (July 2024). Publisher: Nature Publishing Group, p. 797. DOI: 10.1038/s41597-024-03648-8. (Visited on 12/10/2024).

[425] Blue B. Lake et al. "An atlas of healthy and injured cell states and niches in the human kidney". en. In: *Nature* 619.7970 (July 2023), pp. 585–594. DOI: 10.1038/s41586-023-05769-3. (Visited on 12/10/2024).

[426] Rory Collins. "UK biobank: protocol for a large-scale prospective epidemiological resource". In: (2007). Publisher: UK Biobank.

[427] Lyle J Palmer. "UK Biobank: bank on it". en. In: *The Lancet* 369.9578 (June 2007), pp. 1980–1982. DOI: 10.1016/S0140-6736(07)60924-6. (Visited on 09/26/2024).

[428] Sjoerd Van Alten et al. "Reweighting UK Biobank corrects for pervasive selection bias due to volunteering". en. In: *International Journal of Epidemiology* 53.3 (Apr. 2024), dyae054. DOI: 10.1093/ije/dyae054. (Visited on 09/26/2024).

[429] Bart Klijs et al. "Representativeness of the LifeLines Cohort Study". en. In: *PLOS ONE* 10.9 (Sept. 2015). Ed. by Raghib Ietheqad Ali, e0137203. DOI: 10.1371/journal.pone.0137203. (Visited on 09/26/2024).

[430] Anna Fry et al. "Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population". en. In: *American Journal of Epidemiology* 186.9 (Nov. 2017), pp. 1026–1034. DOI: 10.1093/aje/kwx246. (Visited on 09/26/2024).

[431] Emily Banks et al. "Characteristics of Australian cohort study participants who do and do not take up an additional invitation to join a long-term biobank: The 45 and Up Study". en. In: *BMC Research Notes* 5.1 (Dec. 2012), p. 655. DOI: 10.1186/1756-0500-5-655. (Visited on 09/26/2024).

[432] UK Biobank. *Data providers and dates of data availability*. Sept. 2024. URL: https://biobank.ctsu.ox.ac.uk/crystal/exinfo.cgi?src=Data_providers_and_dates (visited on 09/26/2024).

[433] FinnGen. *FinnGen: Documentation of R12 release*. 2024. URL: https://finngen.gitbook.io/documentation/ (visited on 09/26/2024).

[434] Joelle Mbatchou et al. "Computationally efficient whole-genome regression for quantitative and binary traits". en. In: *Nature Genetics* 53.7 (July 2021). Number: 7 Publisher: Nature Publishing Group, pp. 1097–1103. DOI: 10.1038/s41588-021-00870-7. (Visited on 06/14/2023).

[435] Akiko Nagai et al. "Overview of the BioBank Japan Project: Study design and profile". en. In: *Journal of Epidemiology* 27.3 (Mar. 2017), S2–S8. DOI: 10.1016/j.je.2016.12.005. (Visited on 09/26/2024).

[436] Toshihiro Tanaka, Yuki Nagata, and Akira Takemoto. “Integrating biomedical and clinical data with BioBank Japan”. en. In: *Nature Cardiovascular Research* 1.7 (July 2022), pp. 597–598. DOI: 10.1038/s44161-022-00105-w. (Visited on 09/26/2024).

[437] Masato Akiyama et al. “Genome-wide association study identifies 112 new loci for body mass index in the Japanese population”. eng. In: *Nature Genetics* 49.10 (Oct. 2017), pp. 1458–1467. DOI: 10.1038/ng.3951.

[438] Chizu Tanikawa et al. “GWAS identifies two novel colorectal cancer loci at 16q24.1 and 20q13.12”. eng. In: *Carcinogenesis* 39.5 (May 2018), pp. 652–660. DOI: 10.1093/carcin/bgy026.

[439] Yun Li et al. “MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes”. eng. In: *Genetic Epidemiology* 34.8 (Dec. 2010), pp. 816–834. DOI: 10.1002/gepi.20533.

[440] The 1000 Genomes Project Consortium et al. “A global reference for human genetic variation”. en. In: *Nature* 526.7571 (Oct. 2015), pp. 68–74. DOI: 10.1038/nature15393. (Visited on 09/26/2024).

[441] “The “All of Us” Research Program”. In: *New England Journal of Medicine* 381.7 (Aug. 2019). Publisher: Massachusetts Medical Society _eprint: https://www.nejm.org/doi/pdf/10.1056/NEJMsr1809937, pp. 668–676. DOI: 10.1056/NEJMsr1809937. (Visited on 11/12/2024).

[442] Alexander G. Bick et al. “Genomic data in the All of Us Research Program”. en. In: *Nature* 627.8003 (Mar. 2024). Publisher: Nature Publishing Group, pp. 340–346. DOI: 10.1038/s41586-023-06957-x. (Visited on 11/12/2024).

[443] J. Marc Overhage et al. “Validation of a common data model for active safety surveillance research”. eng. In: *Journal of the American Medical Informatics Association: JAMIA* 19.1 (2012), pp. 54–60. DOI: 10.1136/amiajnl-2011-000376.

[444] Po-Ru Loh et al. “Mixed-model association for biobank-scale datasets”. en. In: *Nature Genetics* 50.7 (July 2018). Number: 7 Publisher: Nature Publishing Group, pp. 906–908. DOI: 10.1038/s41588-018-0144-6. (Visited on 08/08/2023).

[445] Shaun Purcell and Christopher C Chang. *PLINK*. URL: www.cog-genomics.org/plink/2.0/.

[446] Christopher C Chang et al. "Second-generation PLINK: rising to the challenge of larger and richer datasets". In: *GigaScience* 4.1 (Dec. 2015), s13742–015–0047–8. DOI: 10.1186/s13742-015-0047-8. (Visited on 03/30/2023).

[447] *HapMap 3 - Wellcome Sanger Institute*. URL: https://www.sanger.ac.uk/resources/downloads/human/hapmap3.html (visited on 07/05/2023).

[448] Brendan Bulik-Sullivan et al. "An atlas of genetic correlations across human diseases and traits". In: *Nature genetics* 47.11 (2015). ISBN: 1061-4036 Publisher: Nature Publishing Group US New York, pp. 1236–1241.

[449] Brendan K. Bulik-Sullivan et al. "LD Score regression distinguishes confounding from polygenicity in genome-wide association studies". In: *Nature genetics* 47.3 (2015). ISBN: 1061-4036 Publisher: Nature Publishing Group US New York, pp. 291–295.

[450] Shoichiro Tsugane and Tomotaka Sobue. "Baseline survey of JPHC study design and participation rate". In: *Journal of epidemiology* 11.6sup (2001). ISBN: 0917-5040 Publisher: Japan Epidemiological Association, pp. 24–29.

[451] Nobuyuki Hamajima et al. "The Japan Multi-institutional Collaborative Cohort Study (JMICC Study) to detect gene-environment interactions for cancer". In: *Asian Pacific Journal of Cancer Prevention* 8.2 (2007). ISBN: 1513-7368 Publisher: Asian Pacific Organization for Cancer Prevention, pp. 317–323.

[452] Shinichi Kuriyama et al. "The Tohoku medical megabank project: design and mission". In: *Journal of epidemiology* 26.9 (2016). ISBN: 0917-5040 Publisher: Japan Epidemiological Association, pp. 493–511.

[453] Wei Zhou et al. "Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies". en. In: *Nature Genetics* 50.9 (Sept. 2018). Publisher: Nature Publishing Group, pp. 1335–1341. DOI: 10.1038/s41588-018-0184-y. (Visited on 02/20/2025).

[454] Cristen J. Willer, Yun Li, and Gonçalo R. Abecasis. "METAL: fast and efficient meta-analysis of genomewide association scans". In: *Bioinformatics*

26.17 (2010). ISBN: 1460-2059 Publisher: Oxford University Press, pp. 2190–2191.

[455] Hao Zhao et al. “CrossMap: a versatile tool for coordinate conversion between genome assemblies”. In: *Bioinformatics* 30.7 (Apr. 2014), pp. 1006–1007. DOI: 10.1093/bioinformatics/btt730. (Visited on 03/30/2023).

[456] Makoto Hirata et al. “Overview of BioBank Japan follow-up data in 32 diseases”. In: *Journal of epidemiology* 27.Supplement_III (2017). ISBN: 0917-5040 Publisher: Japan Epidemiological Association, S22–S28.

[457] Jian Yang et al. “Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits”. en. In: *Nature Genetics* 44.4 (Apr. 2012). Number: 4 Publisher: Nature Publishing Group, pp. 369–375. DOI: 10.1038/ng.2213. (Visited on 04/04/2023).

[458] Jian Yang et al. “GCTA: a tool for genome-wide complex trait analysis”. eng. In: *American Journal of Human Genetics* 88.1 (Jan. 2011), pp. 76–82. DOI: 10.1016/j.ajhg.2010.11.011.

[459] Bruce S. Weir. *Genetic data analysis. Methods for discrete population genetic data.* Sinauer Associates, Inc. Publishers, 1990.

[460] Christiaan A. de Leeuw et al. “MAGMA: Generalized Gene-Set Analysis of GWAS Data”. en. In: *PLOS Computational Biology* 11.4 (Apr. 2015). Publisher: Public Library of Science, e1004219. DOI: 10.1371/journal.pcbi.1004219. (Visited on 02/07/2023).

[461] Aravind Subramanian et al. “Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles”. In: *Proceedings of the National Academy of Sciences* 102.43 (Oct. 2005). Publisher: Proceedings of the National Academy of Sciences, pp. 15545–15550. DOI: 10.1073/pnas.0506580102. (Visited on 02/07/2023).

[462] Arthur Liberzon et al. “Molecular signatures database (MSigDB) 3.0”. In: *Bioinformatics* 27.12 (June 2011), pp. 1739–1740. DOI: 10.1093/bioinformatics/btr260. (Visited on 02/07/2023).

[463] Arthur Liberzon et al. “The Molecular Signatures Database (MSigDB) hallmark gene set collection”. In: *Cell systems* 1.6 (Dec. 2015), pp. 417–425. DOI: 10.1016/j.cels.2015.12.004. (Visited on 02/07/2023).

[464] Broad Institute. “Picard toolkit”. In: *{Broad Institute, GitHub repository}* (2019).

[465] Pablo Cingolani et al. “A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff”. In: *Fly* 6.2 (Apr. 2012). Publisher: Taylor & Francis _eprint: https://doi.org/10.4161/fly.19695, pp. 80–92. DOI: 10.4161/fly.19695. (Visited on 06/13/2023).

[466] Peter W Harrison et al. “Ensembl 2024”. en. In: *Nucleic Acids Research* 52.D1 (Jan. 2024), pp. D891–D899. DOI: 10.1093/nar/gkad1049. (Visited on 12/11/2024).

[467] Marie-Jo A Brion, Konstantin Shakhbazov, and Peter M Visscher. “Calculating statistical power in Mendelian randomization studies”. en. In: *International Journal of Epidemiology* 42.5 (Oct. 2013), pp. 1497–1501. DOI: 10.1093/ije/dyt179. (Visited on 04/04/2023).

[468] Yoav Benjamini and Yosef Hochberg. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 57.1 (1995). Publisher: [Royal Statistical Society, Wiley], pp. 289–300. (Visited on 08/08/2023).

[469] Ju-Hyun Park et al. “Estimation of effect size distribution from genome-wide association studies and implications for future discoveries”. In: *Nature genetics* 42.7 (July 2010), pp. 570–575. DOI: 10.1038/ng.610. (Visited on 12/07/2023).

[470] Gibran Hemani, Kate Tilling, and George Davey Smith. “Orienting the causal relationship between imprecisely measured traits using GWAS summary data”. en. In: *PLOS Genetics* 13.11 (Nov. 2017). Ed. by Jun Li, e1007081. DOI: 10.1371/journal.pgen.1007081. (Visited on 04/04/2023).

[471] Edward Mountjoy et al. “An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci”. en. In: *Nature Genetics* 53.11 (Nov. 2021). Number: 11 Publisher: Nature Publishing

Group, pp. 1527–1533. DOI: 10.1038/s41588-021-00945-5. (Visited on 07/06/2023).

[472] Maya Ghoussaini et al. "Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics". In: *Nucleic Acids Research* 49.D1 (Jan. 2021), pp. D1311–D1320. DOI: 10.1093/nar/gkaa840. (Visited on 07/06/2023).

[473] Jack Bowden et al. "A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization: A framework for two-sample summary data MR". en. In: *Statistics in Medicine* 36.11 (May 2017), pp. 1783–1802. DOI: 10.1002/sim.7221. (Visited on 04/04/2023).

[474] Yoshiharu Muto et al. "Defining cellular complexity in human autosomal dominant polycystic kidney disease by multimodal single cell analysis". eng. In: *Nature Communications* 13.1 (Oct. 2022), p. 6497. DOI: 10.1038/s41467-022-34255-z.

[475] Yuhan Hao et al. "Dictionary learning for integrative, multimodal and scalable single-cell analysis". en. In: *Nature Biotechnology* 42.2 (Feb. 2024). Publisher: Nature Publishing Group, pp. 293–304. DOI: 10.1038/s41587-023-01767-y. (Visited on 12/10/2024).

[476] Tim Stuart et al. "Single-cell chromatin state analysis with Signac". en. In: *Nature Methods* 18.11 (Nov. 2021), pp. 1333–1341. DOI: 10.1038/s41592-021-01282-5. (Visited on 12/10/2024).

[477] Christopher S. McGinnis, Lyndsay M. Murrow, and Zev J. Gartner. "DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors". In: *Cell Systems* 8.4 (Apr. 2019). Publisher: Elsevier, 329–337.e4. DOI: 10.1016/j.cels.2019.03.003. (Visited on 12/10/2024).

[478] Ilya Korsunsky et al. "Fast, sensitive and accurate integration of single-cell data with Harmony". en. In: *Nature Methods* 16.12 (Dec. 2019). Publisher: Nature Publishing Group, pp. 1289–1296. DOI: 10.1038/s41592-019-0619-0. (Visited on 12/10/2024).

[479] Yong Zhang et al. “Model-based Analysis of ChIP-Seq (MACS)”. In: *Genome Biology* 9.9 (Sept. 2008), R137. DOI: 10.1186/gb-2008-9-9-r137. (Visited on 12/10/2024).

[480] Hannah A. Pliner et al. “Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data”. en. In: *Molecular Cell* 71.5 (Sept. 2018), 858–871.e8. DOI: 10.1016/j.molcel.2018.06.044. (Visited on 12/11/2024).

[481] Chris Eijsbouts and Chris Wallace. *Peaky- Capture Hi-C calling algorithm*. en. Sept. 2020. URL: https://cqgd.github.io/pky/index.html (visited on 01/22/2025).

[482] Jason M. Torres et al. “Inferring causal genes at type 2 diabetes GWAS loci through chromosome interactions in islet cells”. eng. In: *Wellcome Open Research* 8 (2023), p. 165. DOI: 10.12688/wellcomeopenres.18653.2.

[483] Damien J. Downes et al. *An integrated platform to systematically identify causal variants and genes for polygenic human traits*. en. Oct. 2019. DOI: 10.1101/813618. URL: http://biorxiv.org/lookup/doi/10.1101/813618 (visited on 01/22/2025).

[484] Shuwei Li et al. *Whole-genome sequencing of half-a-million UK Biobank participants*. en. Dec. 2023. DOI: 10.1101/2023.12.06.23299426. URL: http://medrxiv.org/lookup/doi/10.1101/2023.12.06.23299426 (visited on 01/09/2025).

[485] Melissa J. Landrum et al. “ClinVar: public archive of relationships among sequence variation and human phenotype”. en. In: *Nucleic Acids Research* 42.D1 (Jan. 2014), pp. D980–D985. DOI: 10.1093/nar/gkt1113. (Visited on 01/10/2025).

[486] Tanja Babuke and Ritva Tikkanen. “Dissecting the molecular function of reggie/flotillin proteins”. In: *European Journal of Cell Biology*. MPI-CBG and IIMCB Trilateral Workshop "Cell Biology of Intracellular Transport Processes" for Young Scientists from Germany, Czech Republic and Poland 86.9 (Sept. 2007), pp. 525–532. DOI: 10.1016/j.ejcb.2007.03.003. (Visited on 08/15/2024).

[487] Antje Banning, Ana Tomasovic, and Ritva Tikkanen. "Functional Aspects of Membrane Association of Reggie/Flotillin Proteins". In: *Current Protein and Peptide Science* 12.8 (Dec. 2011), pp. 725–735. DOI: 10.2174/138920311798841708.

[488] Emmanuel Letavernier et al. "Pseudoxanthoma Elasticum, Kidney Stones and Pyrophosphate: From a Rare Disease to Urolithiasis and Vascular Calcifications". en. In: *International Journal of Molecular Sciences* 20.24 (Dec. 2019), p. 6353. DOI: 10.3390/ijms20246353. (Visited on 08/15/2024).

[489] Caroline Prot-Bertoye and Pascal Houillier. "Claudins in Renal Physiology and Pathology". In: *Genes* 11.3 (Mar. 2020), p. 290. DOI: 10.3390/genes11030290. (Visited on 03/21/2024).

[490] K. Yamagishi et al. "The rs2231142 variant of the ABCG2 gene is associated with uric acid levels and gout among Japanese people". en. In: *Rheumatology* 49.8 (Aug. 2010), pp. 1461–1465. DOI: 10.1093/rheumatology/keq096. (Visited on 01/10/2025).

[491] Siwei Chen et al. "A genomic mutational constraint map using variation in 76,156 human genomes". In: *Nature* 625.7993 (Jan. 2024), pp. 92–100. DOI: 10.1038/s41586-023-06045-0.

[492] Lili Zhang et al. "Association of Functional Polymorphism rs2231142 (Q141K) in the ABCG2 Gene With Serum Uric Acid and Gout in 4 US Populations: The PAGE Study". In: *American Journal of Epidemiology* 177.9 (May 2013), pp. 923–932. DOI: 10.1093/aje/kws330. (Visited on 01/10/2025).

[493] M. Huls et al. "The breast cancer resistance protein transporter ABCG2 is expressed in the human kidney proximal tubule apical membrane". en. In: *Kidney International* 73.2 (Jan. 2008), pp. 220–225. DOI: 10.1038/sj.ki.5002645. (Visited on 01/10/2025).

[494] Olivier Devuyst. "The Expression of Water Channels AQP1 and AQP2 in a Large Series of ADPKD Kidneys". en. In: *Nephron* 78.1 (1998), pp. 116–116. DOI: 10.1159/000044893. (Visited on 01/10/2025).

[495] Weiling Wang et al. "Aquaporin-1 retards renal cyst development in polycystic kidney disease by inhibition of Wnt signaling". In: *The FASEB Journal* 29.4

(Apr. 2015), pp. 1551–1563. DOI: 10.1096/fj.14-260828. (Visited on 01/10/2025).

[496] Frank Zaucke et al. "Uromodulin is expressed in renal primary cilia and UMOD mutations result in decreased ciliary uromodulin expression". In: *Human Molecular Genetics* 19.10 (May 2010), pp. 1985–1997. DOI: 10.1093/hmg/ddq077. (Visited on 01/10/2025).

[497] Matthias T. F. Wolf et al. "Mutations of the *Uromodulin* gene in MCKD type 2 patients cluster in exon 4, which encodes three EGF-like domains". In: *Kidney International* 64.5 (Nov. 2003), pp. 1580–1587. DOI: 10.1046/j.1523-1755.2003.00269.x. (Visited on 01/10/2025).

[498] Lan Mo et al. "Tamm-Horsfall protein is a critical renal defense factor protecting against calcium oxalate crystal formation". eng. In: *Kidney International* 66.3 (Sept. 2004), pp. 1159–1166. DOI: 10.1111/j.1523-1755.2004.00867.x.

[499] Max Karlsson et al. "A single–cell type transcriptomics map of human tissues". en. In: *Science Advances* 7.31 (July 2021), eabh2169. DOI: 10.1126/sciadv.abh2169. (Visited on 09/13/2024).

[500] Daniel Liedtke et al. "Tissue-Nonspecific Alkaline Phosphatase-A Gatekeeper of Physiological Conditions in Health and a Modulator of Biological Environments in Disease". eng. In: *Biomolecules* 10.12 (Dec. 2020), p. 1648. DOI: 10.3390/biom10121648.

[501] Michael P. Whyte. "Hypophosphatasia — aetiology, nosology, pathogenesis, diagnosis and treatment". en. In: *Nature Reviews Endocrinology* 12.4 (Apr. 2016). Publisher: Nature Publishing Group, pp. 233–246. DOI: 10.1038/nrendo.2016.14. (Visited on 01/10/2025).

[502] Jack G. Kleinman et al. "Expression of osteopontin, a urinary inhibitor of stone mineral crystal growth, in rat kidney". en. In: *Kidney International* 47.6 (June 1995), pp. 1585–1596. DOI: 10.1038/ki.1995.222. (Visited on 09/09/2024).

[503] Jason R. Stubbs et al. "Critical Role of Osteopontin in Maintaining Urinary Phosphate Solubility in CKD". en. In: *Kidney360* 3.9 (Sept. 2022), pp. 1578–1589. DOI: 10.34067/KID.0007352021. (Visited on 09/09/2024).

[504] Yuansheng Xie et al. “Expression, roles, receptors, and regulation of osteopontin in the kidney”. en. In: *Kidney International* 60.5 (Nov. 2001), pp. 1645–1657. DOI: 10.1046/j.1523-1755.2001.00032.x. (Visited on 09/09/2024).

[505] Dmitry Shungin et al. “New genetic loci link adipose and insulin biology to body fat distribution.” eng. In: *Nature* 518.7538 (Feb. 2015). Place: England, pp. 187–196. DOI: 10.1038/nature14132.

[506] Adam E. Locke et al. “Genetic studies of body mass index yield new insights for obesity biology”. In: *Nature* 518.7538 (2015). ISBN: 1476-4687 Publisher: Nature Publishing Group, pp. 197–206.

[507] Loic Yengo et al. “Meta-analysis of genome-wide association studies for height and body mass index in∼ 700000 individuals of European ancestry”. In: *Human molecular genetics* 27.20 (2018). ISBN: 0964-6906 Publisher: Oxford University Press, pp. 3641–3649.

[508] World Health Organisation (WHO). *Obesity and overweight*. Jan. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight (visited on 01/20/2025).

[509] Our World in Data. *Obesity in adults*. Jan. 2025. URL: https://ourworldindata.org/grapher/share-of-adults-defined-as-obese (visited on 01/13/2025).

[510] *Obesity Profile: short statistical commentary May 2024*. en. URL: https://www.gov.uk/government/statistics/update-to-the-obesity-profile-on-fingertips/obesity-profile-short-statistical-commentary-may-2024 (visited on 01/13/2025).

[511] Marie Ng et al. “Global, regional, and national prevalence of adult overweight and obesity, 1990–2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021”. en. In: *The Lancet* (Mar. 2025), S0140673625003551. DOI: 10.1016/S0140-6736(25)00355-1. (Visited on 03/04/2025).

[512] Y. Claire Wang et al. “Health and economic burden of the projected obesity trends in the USA and the UK”. English. In: *The Lancet* 378.9793 (Aug. 2011).

Publisher: Elsevier, pp. 815–825. DOI: 10.1016/S0140-6736(11)60814-3. (Visited on 01/13/2025).

[513] Eszter Völgyi et al. "Assessing Body Composition With DXA and Bioimpedance: Effects of Obesity, Physical Activity, and Age". en. In: *Obesity* 16.3 (2008). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1038/oby.2007.94, pp. 700–705. DOI: 10.1038/oby.2007.94. (Visited on 01/20/2025).

[514] Prospective Studies Collaboration. "Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies". In: *The Lancet* 373.9669 (Mar. 2009). Publisher: Elsevier, pp. 1083–1096. DOI: 10.1016/S0140-6736(09)60318-4. (Visited on 01/20/2025).

[515] J C Seidell et al. "Assessment of intra-abdominal and subcutaneous abdominal fat: relation between anthropometry and computed tomography". en. In: *The American Journal of Clinical Nutrition* 45.1 (Jan. 1987), pp. 7–13. DOI: 10.1093/ajcn/45.1.7. (Visited on 01/20/2025).

[516] Chathuranga Ranasinghe et al. "Relationship between Body mass index (BMI) and body fat percentage, estimated by bioelectrical impedance, in a group of Sri Lankan adults: a cross sectional study". en. In: *BMC Public Health* 13.1 (Dec. 2013), p. 797. DOI: 10.1186/1471-2458-13-797. (Visited on 01/20/2025).

[517] Mukadas O. Akindele, Julie S. Phillips, and Ehimario U. Igumbor. "The relationship between body fat percentage and body mass index in overweight and obese individuals in an urban african setting". In: *Journal of Public Health in Africa* (Aug. 2016). DOI: 10.4081/jphia.2016.515. (Visited on 01/20/2025).

[518] Kathryn E Bradbury et al. "Association between physical activity and body fat percentage, with adjustment for BMI: a large cross-sectional analysis of UK Biobank". en. In: *BMJ Open* 7.3 (Mar. 2017), e011843. DOI: 10.1136/bmjopen-2016-011843. (Visited on 01/20/2025).

[519] J. Min, D. T. Chiu, and Y. Wang. "Variation in the heritability of body mass index based on diverse twin studies: a systematic review". en. In: *Obesity*

*Reviews* 14.11 (Nov. 2013), pp. 871–882. DOI: 10.1111/obr.12065. (Visited on 01/13/2025).

[520] Karri Silventoinen et al. "Heritability of body size and muscle strength in young adulthood: a study of one million Swedish men". eng. In: *Genetic Epidemiology* 32.4 (May 2008), pp. 341–349. DOI: 10.1002/gepi.20308.

[521] Harold E. Bays et al. "Obesity, adiposity, and dyslipidemia: A consensus statement from the National Lipid Association". In: *Journal of Clinical Lipidology* 7.4 (July 2013), pp. 304–383. DOI: 10.1016/j.jacl.2013.04.001. (Visited on 01/20/2025).

[522] Harold E. Bays. "Adiposopathy". en. In: *Journal of the American College of Cardiology* 57.25 (June 2011), pp. 2461–2473. DOI: 10.1016/j.jacc.2011.02.038. (Visited on 01/20/2025).

[523] World Health Organisation (WHO). "Obesity: preventing and managing the global epidemic. Report of a WHO consultation". eng. In: *World Health Organization Technical Report Series* 894 (2000), pp. i–xii, 1–253.

[524] M Ashwell, T J Cole, and A K Dixon. "Obesity: new insight into the anthropometric classification of fat distribution shown by computed tomography." en. In: *BMJ* 290.6483 (June 1985), pp. 1692–1694. DOI: 10.1136/bmj.290.6483.1692. (Visited on 01/20/2025).

[525] K. M. Rose et al. "Genetic and behavioral determinants of waist-hip ratio and waist circumference in women twins." eng. In: *Obesity research* 6.6 (Nov. 1998). Place: United States, pp. 383–392. DOI: 10.1002/j.1550-8528.1998.tb00369.x.

[526] J. Stevens, E. G. Katz, and R. R. Huxley. "Associations between gender, age and waist circumference". en. In: *European Journal of Clinical Nutrition* 64.1 (Jan. 2010). Publisher: Nature Publishing Group, pp. 6–15. DOI: 10.1038/ejcn.2009.101. (Visited on 01/20/2025).

[527] World Health Organization. "Waist circumference and waist-hip ratio : report of a WHO expert consultation, Geneva, 8-11 December 2008". en. In: (2011). Place: Geneva Publisher: World Health Organization Section: vi, 39 p. (Visited on 01/20/2025).

[528] Melanie Davies et al. “Semaglutide 2·4 mg once a week in adults with overweight or obesity, and type 2 diabetes (STEP 2): a randomised, double-blind, double-dummy, placebo-controlled, phase 3 trial”. In: *The Lancet* 397.10278 (Mar. 2021). Publisher: Elsevier, pp. 971–984. DOI: 10.1016/S0140-6736(21)00213-0. (Visited on 01/20/2025).

[529] W. Timothy Garvey et al. “Two-year effects of semaglutide in adults with overweight or obesity: the STEP 5 trial”. en. In: *Nature Medicine* 28.10 (Oct. 2022), pp. 2083–2091. DOI: 10.1038/s41591-022-02026-4. (Visited on 01/20/2025).

[530] Domenica Rubino et al. “Effect of Continued Weekly Subcutaneous Semaglutide vs Placebo on Weight Loss Maintenance in Adults With Overweight or Obesity: The STEP 4 Randomized Clinical Trial”. en. In: *JAMA* 325.14 (Apr. 2021), p. 1414. DOI: 10.1001/jama.2021.3224. (Visited on 01/20/2025).

[531] Norbert Stefan. “Causes, consequences, and treatment of metabolically unhealthy fat distribution”. English. In: *The Lancet Diabetes & Endocrinology* 8.7 (July 2020). Publisher: Elsevier, pp. 616–627. DOI: 10.1016/S2213-8587(20)30110-8. (Visited on 01/16/2025).

[532] J. A. Williamson et al. “Chromosomal mapping of the human and mouse homologues of two new members of the AP-2 family of transcription factors”. eng. In: *Genomics* 35.1 (July 1996), pp. 262–264. DOI: 10.1006/geno.1996.0351.

[533] J. M. Bosher, T. Williams, and H. C. Hurst. “The developmentally regulated transcription factor AP-2 is involved in c-erbB-2 overexpression in human mammary carcinoma”. eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 92.3 (Jan. 1995), pp. 744–747. DOI: 10.1073/pnas.92.3.744.

[534] Mieke Raap et al. “Lobular carcinoma in situ and invasive lobular breast cancer are characterized by enhanced expression of transcription factor AP-2$\beta$”. en. In: *Laboratory Investigation* 98.1 (Jan. 2018). Publisher: Nature Publishing Group, pp. 117–129. DOI: 10.1038/labinvest.2017.106. (Visited on 01/15/2025).

[535] Markus Moser, Josef Rüschoff, and Reinhard Buettner. "Comparative analysis of AP-2$\alpha$ and AP-2$\beta$ gene expression during murine embryogenesis". en. In: *Developmental Dynamics* 208.1 (Jan. 1997), pp. 115–124. DOI: 10.1002/(SICI)1097-0177(199701)208:1<115::AID-AJA11>3.0.CO;2-5. (Visited on 01/15/2025).

[536] Masahiko Satoda et al. "Mutations in TFAP2B cause Char syndrome, a familial form of patent ductus arteriosus". en. In: *Nature Genetics* 25.1 (May 2000), pp. 42–46. DOI: 10.1038/75578. (Visited on 01/15/2025).

[537] Timothy M. Frayling et al. "A Common Variant in the *FTO* Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity". en. In: *Science* 316.5826 (May 2007), pp. 889–894. DOI: 10.1126/science.1141634. (Visited on 01/15/2025).

[538] Tuomo Rankinen et al. "The Human Obesity Gene Map: The 2005 Update". en. In: *Obesity* 14.4 (Apr. 2006), pp. 529–644. DOI: 10.1038/oby.2006.71. (Visited on 01/15/2025).

[539] Débora R. Sobreira et al. "Extensive pleiotropism and allelic heterogeneity mediate metabolic effects of *IRX3* and *IRX5*". en. In: *Science* 372.6546 (June 2021), pp. 1085–1091. DOI: 10.1126/science.abf1008. (Visited on 01/15/2025).

[540] Jens Juul Holst. "The incretin system in healthy humans: The role of GIP and GLP-1". en. In: *Metabolism* 96 (July 2019), pp. 46–55. DOI: 10.1016/j.metabol.2019.04.014. (Visited on 09/09/2024).

[541] W. Creutzfeldt and R. Ebert. "The Incretin Concept". en. In: *Advances in Metabolic Disorders*. Vol. 11. Elsevier, 1988, pp. 333–367. DOI: 10.1016/B978-0-12-027311-9.50014-7. (Visited on 09/09/2024).

[542] Yutaka Seino, Mitsuo Fukushima, and Daisuke Yabe. "GIP and GLP-1, the two incretin hormones: Similarities and differences". en. In: *Journal of Diabetes Investigation* 1.1-2 (Feb. 2010), pp. 8–23. DOI: 10.1111/j.2040-1124.2010.00022.x. (Visited on 09/09/2024).

[543] Mihaly Varadi et al. "AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with

high-accuracy models". eng. In: *Nucleic Acids Research* 50.D1 (Jan. 2022), pp. D439–D444. DOI: 10.1093/nar/gkab1061.

[544] Unnur Styrkarsdottir et al. "Obesity Variants in the GIPR Gene Are not Associated With Risk of Fracture or Bone Mineral Density". In: *The Journal of Clinical Endocrinology & Metabolism* 109.8 (Aug. 2024), e1608–e1615. DOI: 10.1210/clinem/dgad734. (Visited on 09/23/2024).

[545] Lucie Yammine et al. "Spatiotemporal regulation of GIPR signaling impacts glucose homeostasis as revealed in studies of a common GIPR variant". In: *Molecular Metabolism* 78 (Dec. 2023), p. 101831. DOI: 10.1016/j.molmet.2023.101831. (Visited on 09/23/2024).

[546] Bennett Institute for Applied Data Science. *OpenPrescribing.net*. URL: https://openprescribing.net/ (visited on 01/16/2025).

[547] Jonathan H. Watanabe et al. "Trends in glucagon-like peptide 1 receptor agonist use, 2014 to 2022". en. In: *Journal of the American Pharmacists Association* 64.1 (Jan. 2024), pp. 133–138. DOI: 10.1016/j.japh.2023.10.002. (Visited on 01/16/2025).

[548] Emma Ahlqvist et al. "Link Between GIP and Osteopontin in Adipose Tissue and Insulin Resistance". en. In: *Diabetes* 62.6 (June 2013), pp. 2088–2094. DOI: 10.2337/db12-0976. (Visited on 09/09/2024).

[549] Lisa M. Berglund et al. "Glucose-Dependent Insulinotropic Polypeptide Stimulates Osteopontin Expression in the Vasculature via Endothelin-1 and CREB". en. In: *Diabetes* 65.1 (Jan. 2016), pp. 239–254. DOI: 10.2337/db15-0122. (Visited on 09/09/2024).

[550] Kiyoko Kaneko et al. "Comparison of matrix proteins in different types of urinary stone by proteomic analysis using liquid chromatography-tandem mass spectrometry". eng. In: *International Journal of Urology: Official Journal of the Japanese Urological Association* 19.8 (Aug. 2012), pp. 765–772. DOI: 10.1111/j.1442-2042.2012.03005.x.

[551] The Human Protein Atlas. *Tissue expression of SPP1 - Staining in kidney - The Human Protein Atlas*. Jan. 2025. URL: https://www.proteinatlas.org/ENSG00000118785-SPP1/tissue/kidney#img (visited on 01/16/2025).

[552] The Human Protein Atlas. *Tissue expression of GIPR - Staining in kidney - The Human Protein Atlas*. Jan. 2025. URL: https://www.proteinatlas.org/ENSG00000010310-GIPR/tissue/kidney#img (visited on 01/16/2025).

[553] Nicholas Bowker et al. "Genetically Predicted Glucose-Dependent Insulinotropic Polypeptide (GIP) Levels and Cardiovascular Disease Risk Are Driven by Distinct Causal Variants in the GIPR Region". In: *Diabetes* 70.11 (Aug. 2021), pp. 2706–2719. DOI: 10.2337/db21-0103. (Visited on 01/16/2025).

[554] Amra Jujić et al. "Glucose-dependent insulinotropic peptide and risk of cardiovascular events and mortality: a prospective study". en. In: *Diabetologia* 63.5 (May 2020), pp. 1043–1054. DOI: 10.1007/s00125-020-05093-9. (Visited on 03/03/2025).

[555] Jill M. Pulley et al. "Accelerating Precision Drug Development and Drug Repurposing by Leveraging Human Genetics". eng. In: *Assay and Drug Development Technologies* 15.3 (Apr. 2017), pp. 113–119. DOI: 10.1089/adt.2016.772.

[556] Ania M. Jastreboff et al. "Tirzepatide Once Weekly for the Treatment of Obesity". en. In: *New England Journal of Medicine* 387.3 (July 2022), pp. 205–216. DOI: 10.1056/NEJMoa2206038. (Visited on 01/16/2025).

[557] Ania M. Jastreboff et al. "Tirzepatide for Obesity Treatment and Diabetes Prevention". en. In: *New England Journal of Medicine* (Nov. 2024), NEJMoa2410819. DOI: 10.1056/NEJMoa2410819. (Visited on 01/16/2025).

[558] Atul Malhotra et al. "Tirzepatide for the Treatment of Obstructive Sleep Apnea and Obesity". In: *New England Journal of Medicine* 391.13 (Oct. 2024). Publisher: Massachusetts Medical Society _eprint: https://www.nejm.org/doi/pdf/10.1056/NEJMoa2404881, pp. 1193–1205. DOI: 10.1056/NEJMoa2404881. (Visited on 01/16/2025).

[559] Rohit Loomba et al. "Tirzepatide for Metabolic Dysfunction–Associated Steatohepatitis with Liver Fibrosis". In: *New England Journal of Medicine* 391.4 (July 2024). Publisher: Massachusetts Medical Society _eprint: https://www.nejm.org/doi/pdf/10.1056/NEJMoa2401943, pp. 299–310. DOI: 10.1056/NEJMoa2401943. (Visited on 01/16/2025).

[560] Juan P. Frías et al. “Tirzepatide versus Semaglutide Once Weekly in Patients with Type 2 Diabetes”. In: *New England Journal of Medicine* 385.6 (Aug. 2021). Publisher: Massachusetts Medical Society _eprint: https://www.nejm.org/doi/pdf/10.1056/NEJMoa2107519, pp. 503–515. DOI: 10.1056/NEJMoa2107519. (Visited on 01/16/2025).

[561] W. Timothy Garvey et al. “Tirzepatide once weekly for the treatment of obesity in people with type 2 diabetes (SURMOUNT-2): a double-blind, randomised, multicentre, placebo-controlled, phase 3 trial”. English. In: *The Lancet* 402.10402 (Aug. 2023). Publisher: Elsevier, pp. 613–626. DOI: 10.1016/S0140-6736(23)01200-X. (Visited on 01/16/2025).

[562] Julio Rosenstock et al. “Efficacy and safety of a novel dual GIP and GLP-1 receptor agonist tirzepatide in patients with type 2 diabetes (SURPASS-1): a double-blind, randomised, phase 3 trial”. English. In: *The Lancet* 398.10295 (July 2021). Publisher: Elsevier, pp. 143–155. DOI: 10.1016/S0140-6736(21)01324-6. (Visited on 01/16/2025).

[563] Tim Heise et al. “Effects of subcutaneous tirzepatide versus placebo or semaglutide on pancreatic islet function and insulin sensitivity in adults with type 2 diabetes: a multicentre, randomised, double-blind, parallel-arm, phase 1 clinical trial”. eng. In: *The Lancet. Diabetes & Endocrinology* 10.6 (June 2022), pp. 418–429. DOI: 10.1016/S2213-8587(22)00085-7.

[564] Stefano Del Prato et al. “Tirzepatide versus insulin glargine in type 2 diabetes and increased cardiovascular risk (SURPASS-4): a randomised, open-label, parallel-group, multicentre, phase 3 trial”. English. In: *The Lancet* 398.10313 (Nov. 2021). Publisher: Elsevier, pp. 1811–1824. DOI: 10.1016/S0140-6736(21)02188-7. (Visited on 01/16/2025).

[565] Ellen M. Apperloo et al. “Tirzepatide Associated With Reduced Albuminuria in Participants With Type 2 Diabetes: Pooled Post Hoc Analysis From the Randomized Active- and Placebo-Controlled SURPASS-1–5 Clinical Trials”. In: *Diabetes Care* (Jan. 2025), p. dc241773. DOI: 10.2337/dc24-1773. (Visited on 01/16/2025).

[566] Joint Formulary Committee. “Tirzepatide”. en-GB. In: *British National Formulary*. 2025. (Visited on 01/16/2025).

[567] Bernhard Ludvik et al. “Once-weekly tirzepatide versus once-daily insulin degludec as add-on to metformin with or without SGLT2 inhibitors in patients with type 2 diabetes (SURPASS-3): a randomised, open-label, parallel-group, phase 3 trial”. en. In: *The Lancet* 398.10300 (Aug. 2021), pp. 583–598. DOI: 10.1016/S0140-6736(21)01443-4. (Visited on 01/16/2025).

[568] Dominik Dahl et al. “Effect of Subcutaneous Tirzepatide vs Placebo Added to Titrated Insulin Glargine on Glycemic Control in Patients With Type 2 Diabetes: The SURPASS-5 Randomized Clinical Trial”. en. In: *JAMA* 327.6 (Feb. 2022), p. 534. DOI: 10.1001/jama.2022.0078. (Visited on 01/16/2025).

[569] Hiddo J. L. Heerspink et al. “Effects of tirzepatide versus insulin glargine on kidney outcomes in type 2 diabetes in the SURPASS-4 trial: post-hoc analysis of an open-label, randomised, phase 3 trial”. English. In: *The Lancet Diabetes & Endocrinology* 10.11 (Nov. 2022). Publisher: Elsevier, pp. 774–785. DOI: 10.1016/S2213-8587(22)00243-1. (Visited on 01/16/2025).

[570] Yan Xie, Taeyoung Choi, and Ziyad Al-Aly. “Mapping the effectiveness and risks of GLP-1 receptor agonists”. en. In: *Nature Medicine* (Jan. 2025). Publisher: Nature Publishing Group, pp. 1–12. DOI: 10.1038/s41591-024-03412-w. (Visited on 01/21/2025).

[571] Medicines and Healthcare products Regulatory Agency. *MHRA authorises diabetes drug Mounjaro (tirzepatide) for weight management and weight loss*. en. Aug. 2023. URL: https://www.gov.uk/government/news/mhra-authorises-diabetes-drug-mounjaro-tirzepatide-for-weight-management-and-weight-loss (visited on 01/20/2025).

[572] Bernd Richter et al. “Dipeptidyl peptidase-4 (DPP-4) inhibitors for type 2 diabetes mellitus”. en. In: *Cochrane Database of Systematic Reviews* 2010.1 (Apr. 2008). Ed. by Cochrane Metabolic and Endocrine Disorders Group. DOI: 10.1002/14651858.CD006739.pub2. (Visited on 01/20/2025).

[573] T. Karagiannis et al. “Dipeptidyl peptidase-4 inhibitors for treatment of type 2 diabetes mellitus in the clinical setting: systematic review and meta-analysis”. en. In: *BMJ* 344.mar12 1 (Mar. 2012), e1369–e1369. DOI: 10.1136/bmj.e1369. (Visited on 01/20/2025).

[574] Julie M. Paik et al. “Sodium-Glucose Cotransporter 2 Inhibitors and Nephrolithiasis Risk in Patients With Type 2 Diabetes”. en. In: *JAMA Internal Medicine* 184.3 (Mar. 2024), p. 265. DOI: 10.1001/jamainternmed.2023.7660. (Visited on 01/20/2025).

[575] Rodrigo Daza-Arnedo et al. “Dipeptidyl Peptidase-4 Inhibitors and Diabetic Kidney Disease: A Narrative Review”. English. In: *Kidney Medicine* 3.6 (Nov. 2021). Publisher: Elsevier, pp. 1065–1073. DOI: 10.1016/j.xkme.2021.07.007. (Visited on 03/06/2025).

[576] Vivian A. Fonseca. “New developments in diabetes management: medications of the 21st century”. eng. In: *Clinical Therapeutics* 36.4 (Apr. 2014), pp. 477–484. DOI: 10.1016/j.clinthera.2014.01.018.

[577] Jixin Zhong, Xiaoquan Rao, and Sanjay Rajagopalan. “An emerging role of dipeptidyl peptidase 4 (DPP4) beyond glucose control: Potential implications in cardiovascular disease”. English. In: *Atherosclerosis* 226.2 (Feb. 2013). Publisher: Elsevier, pp. 305–314. DOI: 10.1016/j.atherosclerosis.2012.09.012. (Visited on 03/06/2025).

[578] The EMPA-KIDNEY Collaborative Group. “Empagliflozin in Patients with Chronic Kidney Disease”. en. In: *New England Journal of Medicine* 388.2 (Jan. 2023), pp. 117–127. DOI: 10.1056/NEJMoa2204233. (Visited on 03/06/2025).

[579] Sanjay Kalra. “Sodium glucose co-transporter-2 (SGLT2) inhibitors: a review of their basic and clinical pharmacology”. In: *Diabetes Therapy* 5 (2014). ISBN: 1869-6953 Publisher: Springer, pp. 355–366.

[580] Clifford J. Bailey, Caroline Day, and Srikanth Bellary. “Renal Protection with SGLT2 Inhibitors: Effects in Acute and Chronic Kidney Disease”. en. In: *Current Diabetes Reports* 22.1 (Jan. 2022), pp. 39–52. DOI: 10.1007/s11892-021-01442-z. (Visited on 01/20/2025).

[581] Jenny E. Blau et al. “Canagliflozin triggers the FGF23/1,25-dihydroxyvitamin D/PTH axis in healthy volunteers in a randomized crossover study”. en. In: *JCI Insight* 3.8 (Apr. 2018), e99123. DOI: 10.1172/jci.insight.99123. (Visited on 01/20/2025).

[582] Dusan Harmacek et al. “Empagliflozin Changes Urine Supersaturation by Decreasing pH and Increasing Citrate”. en. In: *Journal of the American*

*Society of Nephrology* 33.6 (June 2022), pp. 1073–1075. DOI: 10.1681/ASN.2021111515. (Visited on 01/20/2025).

[583] Živka Dika et al. "SGLT2 Inhibitors and Their Effect on Urolithiasis: Current Evidence and Future Directions". en. In: *Journal of Clinical Medicine* 13.19 (Oct. 2024), p. 6017. DOI: 10.3390/jcm13196017. (Visited on 01/20/2025).

[584] S. K. Yarovoi, E. N. Kareva, and O. V. Dzhalilov. "Effects of oral hypoglycemic drugs on lithogenic properties of urine in nephrolithiasis patients with concurrent type 2 diabetes". rus. In: *Urologiia (Moscow, Russia: 1999)* 3 (July 2018), pp. 63–69.

[585] Stefan Scherr et al. "SGLT2 Inhibitor Use for Treatment of Hypocitraturia in a Distal Renal Tubular Acidosis". In: *Kidney Medicine* 6.7 (July 2024), p. 100839. DOI: 10.1016/j.xkme.2024.100839. (Visited on 01/20/2025).

[586] Manuel A. Anderegg et al. "Empagliflozin in nondiabetic individuals with calcium and uric acid kidney stones: a randomized phase 2 trial". en. In: *Nature Medicine* (Jan. 2025). DOI: 10.1038/s41591-024-03330-x. (Visited on 01/20/2025).

[587] Caroline M. Gorvin. "Genetic causes of neonatal and infantile hypercalcaemia". In: *Pediatric Nephrology (Berlin, Germany)* 37.2 (2022), pp. 289–301. DOI: 10.1007/s00467-021-05082-z. (Visited on 06/07/2023).

[588] Jan Halbritter. "Genetics of kidney stone disease—Polygenic meets monogenic". In: *Néphrologie & Thérapeutique*. Hôpital Necker, 2021Paris, 27 au 29 avril 2021 17 (Apr. 2021), S88–S94. DOI: 10.1016/j.nephro.2020.02.003. (Visited on 01/17/2024).

[589] G C Curhan et al. "Family history and risk of kidney stones." en. In: *Journal of the American Society of Nephrology* 8.10 (Oct. 1997), pp. 1568–1573. DOI: 10.1681/ASN.V8101568. (Visited on 03/21/2024).

[590] John A. Sayer. "Progress in Understanding the Genetics of Calcium-Containing Nephrolithiasis". eng. In: *Journal of the American Society of Nephrology: JASN* 28.3 (Mar. 2017), pp. 748–759. DOI: 10.1681/ASN.2016050576.

[591] Carla G. Monico and Dawn S. Milliner. "Genetic determinants of urolithiasis". eng. In: *Nature Reviews. Nephrology* 8.3 (Dec. 2011), pp. 151–162. DOI: 10.1038/nrneph.2011.211.

[592] Chihiro Hisatsune et al. "Amplification of Ca2+ signaling by diacylglycerol-mediated inositol 1,4,5-trisphosphate production". eng. In: *The Journal of Biological Chemistry* 280.12 (Mar. 2005), pp. 11723–11730. DOI: 10.1074/jbc.M409535200.

[593] Mary Jane Knight et al. "Zinc binding drives sheet formation by the SAM domain of diacylglycerol kinase $\delta$". eng. In: *Biochemistry* 49.44 (Nov. 2010), pp. 9667–9676. DOI: 10.1021/bi101261x.

[594] Bryan T. Harada et al. "Regulation of enzyme localization by polymerization: polymer formation by the SAM domain of diacylglycerol kinase delta1". eng. In: *Structure (London, England: 1993)* 16.3 (Mar. 2008), pp. 380–387. DOI: 10.1016/j.str.2007.12.017.

[595] Katherine U. Gaynor et al. "A Missense GATA3 Mutation, Thr272Ile, Causes the Hypoparathyroidism, Deafness, and Renal Dysplasia Syndrome". In: *The Journal of Clinical Endocrinology & Metabolism* 94.10 (Oct. 2009), pp. 3897–3904. DOI: 10.1210/jc.2009-0717. (Visited on 01/27/2025).

[596] Giovanni Gambaro et al. "Metabolic diagnosis and medical prevention of calcium nephrolithiasis and its systemic manifestations: a consensus statement". eng. In: *Journal of Nephrology* 29.6 (Dec. 2016), pp. 715–734. DOI: 10.1007/s40620-016-0329-y.

[597] Fadil Hannan and Anibha Mahajan. *Summary statistics for mineral metabolism GWAS*. Artwork Size: 1290026401 Bytes Pages: 1290026401 Bytes. 2023. DOI: 10.25446/OXFORD.23735709. URL: https://portal.sds.ox.ac.uk/articles/dataset/Summary_statistics_for_mineral_metabolism_GWAS/23735709 (visited on 08/08/2023).

[598] C Robinson-Cohen et al. "Genetic Variants Associated with Circulating Parathyroid Hormone". en. In: *Journal of the American Society of Nephrology : JASN* 28.5 (May 2017). Publisher: J Am Soc Nephrol. DOI: 10.1681/ASN.2016010069. (Visited on 07/05/2023).

[599] Jiaxing Miao et al. "The deubiquitinase USP40 preserves endothelial integrity by targeting the heat shock protein HSP90$\beta$". en. In: *Experimental & Molecular Medicine* 56.2 (Feb. 2024). Publisher: Nature Publishing Group, pp. 395–407. DOI: 10.1038/s12276-024-01160-y. (Visited on 01/24/2025).

[600] Hisashi Takagi et al. "USP40 gene knockdown disrupts glomerular permeability in zebrafish". en. In: *American Journal of Physiology-Renal Physiology* 312.4 (Apr. 2017), F702–F715. DOI: 10.1152/ajprenal.00197.2016. (Visited on 01/24/2025).

[601] Shohei Takahashi et al. "USP40 deubiquitinates HINT1 and stabilizes p53 in podocyte damage". en. In: *Biochemical and Biophysical Research Communications* 614 (July 2022), pp. 198–206. DOI: 10.1016/j.bbrc.2022.05.043. (Visited on 01/24/2025).

[602] Nasa Sinnott-Armstrong et al. "Genetics of 35 blood and urine biomarkers in the UK Biobank". en. In: *Nature Genetics* 53.2 (Feb. 2021). Publisher: Nature Publishing Group, pp. 185–194. DOI: 10.1038/s41588-020-00757-z. (Visited on 01/24/2025).

[603] Yao Yao et al. "Evaluate the effects of serum urate level on bone mineral density: a genome-wide gene–environment interaction analysis in UK Biobank cohort". en. In: *Endocrine* 73.3 (Sept. 2021), pp. 702–711. DOI: 10.1007/s12020-021-02760-8. (Visited on 01/24/2025).

[604] Dan He et al. "A longitudinal genome-wide association study of bone mineral density mean and variability in the UK Biobank". en. In: *Osteoporosis International* 34.11 (Nov. 2023), pp. 1907–1916. DOI: 10.1007/s00198-023-06852-1. (Visited on 01/24/2025).

[605] Alison R. Barton et al. "Whole-exome imputation within UK Biobank powers rare coding variant association and fine-mapping analyses". eng. In: *Nature Genetics* 53.8 (Aug. 2021), pp. 1260–1269. DOI: 10.1038/s41588-021-00892-1.

[606] Daniel Schlam and Johnathan Canton. "Every day I'm rufflin': Calcium sensing and actin dynamics in the growth factor-independent membrane ruffling of professional phagocytes". en. In: *Small GTPases* 8.2 (Apr. 2017), pp. 65–70. DOI: 10.1080/21541248.2016.1197873. (Visited on 01/22/2025).

[607] Sian E Piret et al. "Identification of a G-Protein Subunit-$\alpha$11 Gain-of-Function Mutation, Val340Met, in a Family With Autosomal Dominant Hypocalcemia Type 2 (ADH2)". en. In: *Journal of Bone and Mineral Research* 31.6 (June 2016), pp. 1207–1214. DOI: 10.1002/jbmr.2797. (Visited on 01/22/2025).

[608] Eric Reiter and Robert J. Lefkowitz. "GRKs and $\beta$-arrestins: roles in receptor silencing, trafficking and signaling". English. In: *Trends in Endocrinology & Metabolism* 17.4 (May 2006). Publisher: Elsevier, pp. 159–165. DOI: 10.1016/j.tem.2006.03.008. (Visited on 01/26/2025).

[609] Randy A. Hall et al. "G Protein-coupled Receptor Kinase 6A Phosphorylates the Na+/H+ Exchanger Regulatory Factor via a PDZ Domain-mediated Interaction*". In: *Journal of Biological Chemistry* 274.34 (Aug. 1999), pp. 24328–24334. DOI: 10.1074/jbc.274.34.24328. (Visited on 01/26/2025).

[610] Marie Courbebaisse et al. "A New Human NHERF1 Mutation Decreases Renal Phosphate Transporter NPT2a Expression by a PTH-Independent Mechanism". en. In: *PLOS ONE* 7.4 (Apr. 2012). Publisher: Public Library of Science, e34764. DOI: 10.1371/journal.pone.0034764. (Visited on 01/26/2025).

[611] Zoubida Karim et al. "NHERF1 mutations and responsiveness of renal parathyroid hormone". eng. In: *The New England Journal of Medicine* 359.11 (Sept. 2008), pp. 1128–1135. DOI: 10.1056/NEJMoa0802836.

[612] S. Shenolikar et al. "Targeted disruption of the mouse NHERF-1 gene promotes internalization of proximal tubule sodium-phosphate cotransporter type IIa and renal phosphate wasting". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 99.17 (Aug. 2002), pp. 11470–11475. DOI: 10.1073/pnas.162232699.

[613] Rebecca J Gordon et al. "Digenic Heterozygous Mutations in SLC34A3 and SLC34A1 Cause Dominant Hypophosphatemic Rickets with Hypercalciuria". en. In: *The Journal of Clinical Endocrinology & Metabolism* 105.7 (July 2020), pp. 2392–2400. DOI: 10.1210/clinem/dgaa217. (Visited on 01/26/2025).

[614] Jian-Qing Wang, Fu Qin, and Liang Zhu. "Expression of $Na^+$ /$H^+$ exchanger regulatory factor 1 in autosomal-dominant polycystic kidney disease". en.

In: *Journal of International Medical Research* 43.5 (Oct. 2015), pp. 629–638. DOI: 10.1177/0300060515581182. (Visited on 01/26/2025).

[615] Ewa Pronicka et al. "Biallelic mutations in CYP24A1 or SLC34A1 as a cause of infantile idiopathic hypercalcemia (IIH) with vitamin D hypersensitivity: molecular study of 11 historical IIH cases". en. In: *Journal of Applied Genetics* 58.3 (Aug. 2017), pp. 349–353. DOI: 10.1007/s13353-017-0397-2. (Visited on 08/10/2023).

[616] Andrew Dauber et al. "Genetic Defect in CYP24A1, the Vitamin D 24-Hydroxylase Gene, in a Patient with Severe Infantile Hypercalcemia". In: *The Journal of Clinical Endocrinology and Metabolism* 97.2 (Feb. 2012), E268–E274. DOI: 10.1210/jc.2011-1972. (Visited on 01/26/2025).

[617] Lorenzo Piemonti et al. "Vitamin D3 Affects Differentiation, Maturation, and Function of Human Monocyte-Derived Dendritic Cells". en. In: *The Journal of Immunology* 164.9 (May 2000), pp. 4443–4451. DOI: 10.4049/jimmunol.164.9.4443. (Visited on 01/26/2025).

[618] K Nakamura et al. "1,25-dihydroxyvitamin D3 differentiates normal neutrophilic promyelocytes to monocytes/macrophages in vitro". In: *Blood* 87.7 (Apr. 1996), pp. 2693–2701. DOI: 10.1182/blood.V87.7.2693.bloodjournal8772693. (Visited on 01/26/2025).

[619] Takayuki Takahashi, Kishiko Nakamura, and Sumiko Iho. "Differentiation of Myeloid Cells and 1, 25-Dihydroxyvitamin D3". In: *Leukemia & Lymphoma* 27.1-2 (Jan. 1997). Publisher: Taylor & Francis _eprint: https://doi.org/10.3109/10428199709068268, pp. 25–33. DOI: 10.3109/10428199709068268. (Visited on 01/26/2025).

[620] Sarah M Azer et al. "24-Hydroxylase Deficiency Due to CYP24A1 Sequence Variants: Comparison With Other Vitamin D-mediated Hypercalcemia Disorders". In: *Journal of the Endocrine Society* 5.9 (July 2021), bvab119. DOI: 10.1210/jendso/bvab119. (Visited on 08/10/2023).

[621] Saori Sakaue et al. "A cross-population atlas of genetic associations for 220 human phenotypes". eng. In: *Nature Genetics* 53.10 (Oct. 2021), pp. 1415–1424. DOI: 10.1038/s41588-021-00931-x.

[622] Kira J. Stanzick et al. “Discovery and prioritization of variants and genes for kidney function in >1.2 million individuals”. eng. In: *Nature Communications* 12.1 (July 2021), p. 4350. DOI: 10.1038/s41467-021-24491-0.

[623] Joana A. Revez et al. “Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration”. eng. In: *Nature Communications* 11.1 (Apr. 2020), p. 1647. DOI: 10.1038/s41467-020-15421-7.

[624] Hongbo Liu et al. “Epigenomic and transcriptomic analyses define core cell types, genes and targetable mechanisms for kidney disease”. eng. In: *Nature Genetics* 54.7 (July 2022), pp. 950–962. DOI: 10.1038/s41588-022-01097-w.

[625] International Multiple Sclerosis Genetics Consortium. “Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility”. eng. In: *Science (New York, N.Y.)* 365.6460 (Sept. 2019), eaav7188. DOI: 10.1126/science.aav7188.

[626] Tanya J. Major et al. “A genome-wide association analysis reveals new pathogenic pathways in gout”. eng. In: *Nature Genetics* 56.11 (Nov. 2024), pp. 2392–2406. DOI: 10.1038/s41588-024-01921-5.

[627] Nao Tanaka et al. “Eight novel susceptibility loci and putative causal variants in atopic dermatitis”. eng. In: *The Journal of Allergy and Clinical Immunology* 148.5 (Nov. 2021), pp. 1293–1306. DOI: 10.1016/j.jaci.2021.04.019.

[628] Cassianne Robinson-Cohen et al. “Genetic Variants Associated with Circulating Fibroblast Growth Factor 23”. eng. In: *Journal of the American Society of Nephrology: JASN* 29.10 (Oct. 2018), pp. 2583–2592. DOI: 10.1681/ASN.2018020192.

[629] F. M. Hannan et al. “Identification of 70 calcium-sensing receptor mutations in hyper- and hypo-calcaemic patients: evidence for clustering of extracellular domain mutations at calcium-binding sites”. en. In: *Human Molecular Genetics* 21.12 (June 2012), pp. 2768–2778. DOI: 10.1093/hmg/dds105. (Visited on 01/27/2025).

[630] E. F. Nemeth and W. G. Goodman. “Calcimimetic and Calcilytic Drugs: Feats, Flops, and Futures”. en. In: *Calcified Tissue International* 98.4 (Apr. 2016), pp. 341–358. DOI: 10.1007/s00223-015-0052-z. (Visited on 08/10/2023).

[631] Fadil M Hannan, Mie K Olesen, and Rajesh V Thakker. "Calcimimetic and calcilytic therapies for inherited disorders of the calcium-sensing receptor signalling pathway". en. In: *British Journal of Pharmacology* 175.21 (Nov. 2018), pp. 4083–4094. DOI: 10.1111/bph.14086. (Visited on 01/27/2025).

[632] Andrew P. Evan et al. "Randall's plaque of patients with nephrolithiasis begins in basement membranes of thin loops of Henle". en. In: *The Journal of Clinical Investigation* 111.5 (Mar. 2003). Publisher: American Society for Clinical Investigation, pp. 607–616. DOI: 10.1172/JCI17038. (Visited on 01/17/2024).

[633] Richard She and Daniel F. Jarosz. "Mapping causal variants with single nucleotide resolution reveals biochemical drivers of phenotypic change". In: *Cell* 172.3 (Jan. 2018), 478–490.e15. DOI: 10.1016/j.cell.2017.12.015. (Visited on 01/27/2025).

[634] Kyle Kai-How Farh et al. "Genetic and Epigenetic Fine-Mapping of Causal Autoimmune Disease Variants". In: *Nature* 518.7539 (Feb. 2015), pp. 337–343. DOI: 10.1038/nature13835. (Visited on 01/27/2025).

[635] Daniel J. M. Crouch and Walter F. Bodmer. "Polygenic inheritance, GWAS, polygenic risk scores, and the search for functional variants". In: *Proceedings of the National Academy of Sciences* 117.32 (Aug. 2020). Publisher: Proceedings of the National Academy of Sciences, pp. 18924–18933. DOI: 10.1073/pnas.2005634117. (Visited on 01/27/2025).

[636] Qiangmin Zhang et al. "Parathyroid hormone initiates dynamic NHERF1 phosphorylation cycling and conformational changes that regulate NPT2A-dependent phosphate transport". eng. In: *The Journal of Biological Chemistry* 294.12 (Mar. 2019), pp. 4546–4571. DOI: 10.1074/jbc.RA119.007421.

[637] Joint Formulary Committee. "Phosphate | Drugs | BNF content published by NICE". en-GB. In: *British National Formulary*. (Visited on 01/27/2025).

[638] Charles Ginsberg and Joachim H. Ix. "New Insights into Vitamin D Metabolism in Kidney Disease and Transplant". English. In: *American Journal of Kidney Diseases* 84.4 (Oct. 2024). Publisher: Elsevier, pp. 400–402. DOI: 10.1053/j.ajkd.2024.06.003. (Visited on 01/27/2025).

[639] Hanne Skou Jørgensen et al. “Vitamin D Metabolites Before and After Kidney Transplantation in Patients Who Are Anephric”. English. In: *American Journal of Kidney Diseases* 84.4 (Oct. 2024). Publisher: Elsevier, 427–436.e1. DOI: 10.1053/j.ajkd.2024.03.025. (Visited on 01/27/2025).

[640] Emma Davidson Peiris and Raghav Wusirika. “A Case Report of Compound Heterozygous CYP24A1 Mutations Leading to Nephrolithiasis Successfully Treated with Ketoconazole”. eng. In: *Case Reports in Nephrology and Dialysis* 7.3 (2017), pp. 167–171. DOI: 10.1159/000485243.

[641] Sara Balduzzi, Gerta Rücker, and Guido Schwarzer. “How to perform a meta-analysis with R: a practical tutorial”. eng. In: *Evidence-Based Mental Health* 22.4 (Nov. 2019), pp. 153–160. DOI: 10.1136/ebmental-2019-300117.

[642] Fergal J Martin et al. “Ensembl 2023”. In: *Nucleic Acids Research* 51.D1 (Jan. 2023), pp. D933–D941. DOI: 10.1093/nar/gkac958. (Visited on 01/30/2024).

[643] Paulo P. Amaral et al. “Genomic positional conservation identifies topological anchor point RNAs linked to developmental loci”. en. In: *Genome Biology* 19.1 (Mar. 2018), p. 32. DOI: 10.1186/s13059-018-1405-5. (Visited on 02/03/2025).

[644] Sabina Wansleben et al. “T-box transcription factors in cancer biology”. In: *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1846.2 (Dec. 2014), pp. 380–391. DOI: 10.1016/j.bbcan.2014.08.004. (Visited on 02/03/2025).

[645] Bieke Decaesteker et al. “TBX2 is a neuroblastoma core regulatory circuitry component enhancing MYCN/FOXM1 reactivation of DREAM targets”. en. In: *Nature Communications* 9.1 (Nov. 2018). Publisher: Nature Publishing Group, p. 4866. DOI: 10.1038/s41467-018-06699-9. (Visited on 02/03/2025).

[646] Steven H. Fong et al. “Wnt signalling mediated by Tbx2b regulates cell migration during formation of the neural plate”. In: *Development* 132.16 (Aug. 2005), pp. 3587–3596. DOI: 10.1242/dev.01933. (Visited on 02/03/2025).

[647] Gun-Sik Cho et al. “Role of Tbx2 in defining the territory of the pronephric nephron”. In: *Development* 138.3 (Feb. 2011), pp. 465–474. DOI: 10.1242/dev.061234. (Visited on 02/03/2025).

[648] Zachary Harrelson et al. “Tbx2 is essential for patterning the atrioventricular canal and for morphogenesis of the outflow tract during heart development”. In: *Development* 131.20 (Oct. 2004), pp. 5041–5052. DOI: 10.1242/dev.01378. (Visited on 02/03/2025).

[649] Takayuki Suzuki et al. “Tbx Genes Specify Posterior Digit Identity through Shh and BMP Signaling”. English. In: *Developmental Cell* 6.1 (Jan. 2004). Publisher: Elsevier, pp. 43–53. DOI: 10.1016/S1534-5807(03)00401-5. (Visited on 02/03/2025).

[650] Gun-Sik Cho et al. “Tbx2 regulates anterior neural specification by repressing FGF signaling pathway”. In: *Developmental Biology* 421.2 (Jan. 2017), pp. 183–193. DOI: 10.1016/j.ydbio.2016.11.020. (Visited on 02/03/2025).

[651] Bo Zhu et al. “TBX2 blocks myogenesis and promotes proliferation in rhabdomyosarcoma cells”. en. In: *International Journal of Cancer* 135.4 (2014). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ijc.28721, pp. 785–797. DOI: 10.1002/ijc.28721. (Visited on 02/03/2025).

[652] Bin Wang et al. “The T Box Transcription Factor TBX2 Promotes Epithelial-Mesenchymal Transition and Invasion of Normal and Malignant Breast Epithelial Cells”. en. In: *PLOS ONE* 7.7 (July 2012). Publisher: Public Library of Science, e41355. DOI: 10.1371/journal.pone.0041355. (Visited on 02/03/2025).

[653] Jacqueline J. L. Jacobs et al. “Senescence bypass screen identifies TBX2, which represses Cdkn2a (p19ARF) and is amplified in a subset of human breast cancers”. en. In: *Nature Genetics* 26.3 (Nov. 2000). Publisher: Nature Publishing Group, pp. 291–299. DOI: 10.1038/81583. (Visited on 02/03/2025).

[654] Keith W. Vance et al. “Tbx2 Is Overexpressed and Plays an Important Role in Maintaining Proliferation and Suppression of Senescence in Melanomas”. In: *Cancer Research* 65.6 (Mar. 2005), pp. 2260–2268. DOI: 10.1158/0008-5472.CAN-04-3045. (Visited on 02/03/2025).

[655] Eija H. Mahlamäki et al. “Frequent amplification of 8q24, 11q, 17q, and 20q-specific genes in pancreatic cancer”. en. In: *Genes, Chromosomes and Cancer* 35.4 (2002). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/gcc.10122, pp. 353–358. DOI: 10.1002/gcc.10122. (Visited on 02/03/2025).

[656] Xuewen Song et al. “Systems biology of autosomal dominant polycystic kidney disease (ADPKD): computational identification of gene expression pathways and integrated regulatory networks”. eng. In: *Human Molecular Genetics* 18.13 (July 2009), pp. 2328–2343. DOI: 10.1093/hmg/ddp165.

[657] Gisela Schieren et al. “Gene profiling of polycystic kidneys”. eng. In: *Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association - European Renal Association* 21.7 (July 2006), pp. 1816–1824. DOI: 10.1093/ndt/gfl071.

[658] Seung Wan Chea and Kyu-Beck Lee. “TGF-beta mediated epithelial-mesenchymal transition in autosomal dominant polycystic kidney disease”. eng. In: *Yonsei Medical Journal* 50.1 (Feb. 2009), pp. 105–111. DOI: 10.3349/ymj.2009.50.1.105.

[659] H. Okada et al. “Progressive renal fibrosis in murine polycystic kidney disease: an immunohistochemical observation”. eng. In: *Kidney International* 58.2 (Aug. 2000), pp. 587–597. DOI: 10.1046/j.1523-1755.2000.00205.x.

[660] François Fagotto. “Looking beyond the Wnt pathway for the deep nature of $\beta$-catenin”. en. In: *EMBO reports* 14.5 (May 2013), pp. 422–433. DOI: 10.1038/embor.2013.45. (Visited on 02/03/2025).

[661] Frans Van Roy. “Beyond E-cadherin: roles of other cadherin superfamily members in cancer”. en. In: *Nature Reviews Cancer* 14.2 (Feb. 2014), pp. 121–134. DOI: 10.1038/nrc3647. (Visited on 02/03/2025).

[662] P. D. Wilson et al. “Reversed polarity of Na(+) -K(+) -ATPase: mislocation to apical plasma membranes in polycystic kidney disease epithelia”. en. In: *American Journal of Physiology-Renal Physiology* 260.3 (Mar. 1991), F420–F430. DOI: 10.1152/ajprenal.1991.260.3.F420. (Visited on 02/03/2025).

[663] E D Avner, W E Sweeney, and W J Nelson. “Abnormal sodium pump distribution during renal tubulogenesis in congenital murine polycystic kidney disease.” en. In: *Proceedings of the National Academy of Sciences* 89.16 (Aug. 1992), pp. 7447–7451. DOI: 10.1073/pnas.89.16.7447. (Visited on 02/03/2025).

[664] Ellis D. Avner. “Epithelial polarity and differentiation in polycystic kidney disease”. en. In: *Journal of Cell Science* 1993.Supplement_17 (Dec. 1993), pp. 217–222. DOI: 10.1242/jcs.1993.Supplement_17.30. (Visited on 02/03/2025).

[665] E. D. Avner, J. A. McAteer, and A. P. Evan. “Models of Cysts and Cystic Kidneys”. In: *The Cystic Kidney*. Ed. by Kenneth D. Gardner and Jay Bernstein. Dordrecht: Springer Netherlands, 1990, pp. 55–98. DOI: 10.1007/978-94-009-0457-6_4.

[666] L. W. Welling. “Pathogenesis of Cysts and Cystic Kidneys”. In: *The Cystic Kidney*. Ed. by Kenneth D. Gardner and Jay Bernstein. Dordrecht: Springer Netherlands, 1990, pp. 99–116. DOI: 10.1007/978-94-009-0457-6_5.

[667] Patricia D. Wilson and Ann C. Sherwood. “Tubulocystic epithelium”. en. In: *Kidney International* 39.3 (Mar. 1991), pp. 450–463. DOI: 10.1038/ki.1991.56. (Visited on 02/03/2025).

[668] Noel S. Murcia, William E. Sweeney Jr, and Ellis D. Avner. “New insights into the molecular pathophysiology of polycystic kidney disease”. In: *Kidney international* 55.4 (1999). ISBN: 0085-2538 Publisher: Elsevier, pp. 1187–1197.

[669] Gabrielle G. Gilmer et al. “Flow resistance along the rat renal tubule”. In: *American Journal of Physiology-Renal Physiology* 315.5 (Nov. 2018). Publisher: American Physiological Society, F1398–F1405. DOI: 10.1152/ajprenal.00219.2018. (Visited on 02/03/2025).

[670] Ning Liu et al. “Functional variants in TBX2 are associated with a syndromic cardiovascular and skeletal developmental disorder”. en. In: *Human Molecular Genetics* 27.14 (July 2018), pp. 2454–2465. DOI: 10.1093/hmg/ddy146. (Visited on 02/03/2025).

[671] Yicheng Chen et al. "Inhibition of canonical WNT/$\beta$-catenin signaling is involved in leflunomide (LEF)-mediated cytotoxic effects on renal carcinoma cells". eng. In: *Oncotarget* 7.31 (Aug. 2016), pp. 50401–50416. DOI: 10.18632/oncotarget.10409.

[672] Jen X. Xu et al. "Polycystin-1 and G$\alpha$12 regulate the cleavage of E-cadherin in kidney epithelial cells". In: *Physiological Genomics* 47.2 (Feb. 2015). Publisher: American Physiological Society, pp. 24–32. DOI: 10.1152/physiolgenomics.00090.2014. (Visited on 02/03/2025).

[673] Andrew J. Streets et al. "Functional Analysis of PKD1 Transgenic Lines Reveals a Direct Role for Polycystin-1 in Mediating Cell-Cell Adhesion". en. In: *Journal of the American Society of Nephrology* 14.7 (July 2003), pp. 1804–1815. DOI: 10.1097/01.ASN.0000076075.49819.9B. (Visited on 02/03/2025).

[674] Ryan J. Russo et al. "Impaired formation of desmosomal junctions in ADPKD epithelia". en. In: *Histochemistry and Cell Biology* 124.6 (Dec. 2005), pp. 487–497. DOI: 10.1007/s00418-005-0055-3. (Visited on 02/03/2025).

[675] Jing Yu et al. "A Wnt7b-dependent pathway regulates the orientation of epithelial cell division and establishes the cortico-medullary axis of the mammalian kidney". eng. In: *Development (Cambridge, England)* 136.1 (Jan. 2009), pp. 161–171. DOI: 10.1242/dev.022087.

[676] Kai M. Schmidt-Ott and Jonathan Barasch. "WNT/$\beta$-catenin signaling in nephron progenitors and their epithelial progeny". English. In: *Kidney International* 74.8 (Oct. 2008). Publisher: Elsevier, pp. 1004–1008. DOI: 10.1038/ki.2008.322. (Visited on 02/03/2025).

[677] Diana M. Iglesias et al. "Canonical WNT signaling during kidney development". In: *American Journal of Physiology-Renal Physiology* 293.2 (Aug. 2007). Publisher: American Physiological Society, F494–F500. DOI: 10.1152/ajprenal.00416.2006. (Visited on 02/03/2025).

[678] Kameswaran Surendran, Sean P. McCaul, and Theodore C. Simon. "A role for Wnt-4 in renal fibrosis". en. In: *American Journal of Physiology-Renal Physiology* 282.3 (Mar. 2002), F431–F441. DOI: 10.1152/ajprenal.0009.2001. (Visited on 02/03/2025).

[679] Kameswaran Surendran, Susan Schiavi, and Keith A. Hruska. "Wnt-Dependent $\beta$-Catenin Signaling Is Activated after Unilateral Ureteral Obstruction, and Recombinant Secreted Frizzled-Related Protein 4 Alters the Progression of Renal Fibrosis". en. In: *Journal of the American Society of Nephrology* 16.8 (Aug. 2005), pp. 2373–2384. DOI: 10.1681/ASN.2004110949. (Visited on 02/03/2025).

[680] Weichun He et al. "Wnt/$\beta$-Catenin Signaling Promotes Renal Interstitial Fibrosis". en. In: *Journal of the American Society of Nephrology* 20.4 (Apr. 2009), pp. 765–776. DOI: 10.1681/ASN.2008060566. (Visited on 02/03/2025).

[681] B Ljungberg et al. *Renal Cell Carcinoma*. en. 2024. URL: https://uroweb.org/guidelines/renal-cell-carcinoma (visited on 02/04/2025).

[682] Steven C. Campbell et al. "Renal Mass and Localized Renal Cancer: Evaluation, Management, and Follow-Up: AUA Guideline: Part I". EN. In: *The Journal of Urology* (Aug. 2021). Publisher: Wolters KluwerPhiladelphia, PA. DOI: 10.1097/JU.0000000000001911. (Visited on 02/04/2025).

[683] Thomas F. Whelan. "Guidelines on the management of renal cyst disease". In: *Canadian Urological Association Journal* 4.2 (Apr. 2010), pp. 98–99. (Visited on 02/04/2025).

[684] D. R. Kumar et al. "Virchow's Contribution to the Understanding of Thrombosis and Cellular Biology". en. In: *Clinical Medicine & Research* 8.3-4 (Dec. 2010), pp. 168–172. DOI: 10.3121/cmr.2009.866. (Visited on 02/03/2025).

# A | Inclusion criteria for kidney stone disease association analyses in the UK Biobank

| Coding classification | Code | Definition |
|---|---|---|
| International Classification of Diseases (ICD) | ICD9- 7880 | Renal colic |
| | ICD9- 5920 | Calculus of kidney |
| | ICD9- 5920A | Calculus of kidney |
| | ICD9- 5921 | Calculus of ureter |
| | ICD9- 5929 | Urinary calculus, unspecified |
| | ICD10-N20.0 | Calculus of kidney |
| | ICD10-N20.1 | Calculus of ureter |
| | ICD10-N20.2 | Calculus of kidney with calculus of ureter |
| | ICD10-N20.9 | Urinary calculus, unspecified |
| | ICD10-N23 | Unspecified renal colic |
| Classification of Interventions and Procedures (OPCS) | OPCS3- 563.1 | Removal of renal calculus : nephrolithotomy |
| | OPCS3- 563.2 | Removal of renal calculus : pyelolithotomy |
| | OPCS3- 563.3 | Removal of renal calculus : removal without incision |
| | OPCS3- 580 | Ureterolithotomy |
| | OPCS3- 587.2 | Cystoscopic operation on ureter : lithotomy |
| | OPCS4- M06.1 | Open removal of calculus from kidney |
| | OPCS4- M07.1 | Ureteroscopic laser fragmentation of calculus of kidney |
| | OPCS4- M07.2 | Ureteroscopic extraction of calculus of kidney NEC |
| | OPCS4- M09 | Therapeutic endoscopic operations on calculus of kidney |
| | OPCS4- M09.1 | Endoscopic ultrasound fragmentation of calculus of kidney |
| | OPCS4- M09.2 | Endoscopic electrohydraulic shockwave fragmentation of calculus of kidney |
| | OPCS4- M09.3 | Endoscopic laser fragmentation of calculus of kidney |
| | OPCS4- M09.4 | Endoscopic extraction of calculus of kidney NEC |
| | OPCS4- M09.8 | Other specified |
| | OPCS4- M09.9 | Unspecified |
| | OPCS4- M14 | Extracorporeal fragmentation of calculus of kidney |
| | OPCS4- M14.1 | Extracorporeal shock wave lithotripsy of calculus of kidney |
| | OPCS4- M14.8 | Other specified |
| | OPCS4- M14.9 | Unspecified |
| | OPCS4- M16·4 | Percutaneous nephrolithotomy |
| | OPCS4- M26.1 | Nephroscopic laser fragmentation of calculus of ureter |
| | OPCS4- M26.2 | Nephroscopic fragmentation of calculus of ureter NEC |
| | OPCS4- M26.3 | Nephroscopic extraction of calculus of ureter |
| | OPCS4- M27.1 | Ureteroscopic laser fragmentation of calculus of ureter |
| | OPCS4- M27.2 | Ureteroscopic fragmentation of calculus of ureter NEC |
| | OPCS4- M27.3 | Ureteroscopic extraction of calculus of ureter |
| | OPCS4- M28 | Other endoscopic removal of calculus from ureter |

| Coding classification | Code | Definition |
|---|---|---|
| | OPCS4- M28·1 | Endoscopic laser fragmentation of calculus of ureter |
| | OPCS4- M28·2 | Endoscopic fragmentation of calculus of ureter |
| | OPCS4- M28·3 | Endoscopic extraction of calculus of ureter |
| | OPCS4- M28.4 | Endoscopic catheter drainage of calculus of ureter |
| | OPCS4- M28.5 | Endoscopic drainage of calculus of ureter by dilation of ureter |
| | OPCS4- M28.8 | Other specified |
| | OPCS4- M28.9 | Unspecified |
| | OPCS4- M31 | Extracorporeal fragmentation of calculus of ureter |
| | OPCS4- M31.1 | Extracorporeal shockwave lithotripsy of calculus of ureter |
| | OPCS4- M31.8 | Other specified |
| | OPCS4- M31.9 | Unspecified |
| Self-reported operation | 1197 | Percutaneous/open kidney stone surgery/lithotripsy |
| Death codes | N20.0 | Calculus of kidney |
| | N20.1 | Calculus of ureter |
| | N20.2 | Calculus of kidney with calculus of ureter |
| | N20.9 | Urinary calculus, unspecified |
| | N23 | Unspecified renal colic |
| Primary care (Read) | V2- 14D3. | H/O: urinary stone |
| | V2- 1A54. | Ureteric colic |
| | V2- 4G6.. | O/E - ureteric calculus |
| | V2- 7B07. | Percutaneous renal stone surgery |
| | V2- 7B070 | Nephroscopy and ultrasound lithotripsy of renal calculus |
| | V2- 7B071 | Nephroscopy and electrohydraulic lithotripsy of renal calculus |
| | V2- 7B072 | Nephroscopy and laser lithotripsy of renal calculus |
| | V2- 7B073 | Percutaneous nephrolithotomy without disintegration |
| | V2- 7B074 | Endoscopic extraction of calculus of kidney nec |
| | V2- 7B07y | Other specified percutaneous renal stone surgery |
| | V2- 7B07z | Percutaneous renal stone surgery NOS |
| | V2- 7B0B. | Extracorporeal shockwave lithotripsy for renal calculus |
| | V2- 7B0B0 | Extracorporeal shockwave lithotripsy for renal calculus of unspecified size |
| | V2- 7B0B1 | Extracorporeal shockwave lithotripsy for renal calculus less than 2 cm in diameter |
| | V2- 7B0B2 | Extracorporeal shockwave treatment for renal calculus of 2 cm or more in diameter |
| | V2- 7B0By | Other specified extracorporeal shockwave lithotripsy for renal calculus |
| | V2- 7B0Bz | Extracorporeal shockwave lithotripsy for renal calculus NOS |
| | V2- 7B170 | Nephroscopic laser lithotripsy of ureteric calculus |
| | V2- 7B171 | Other nephroscopic fragmentation of ureteric calculus |
| | V2- 7B172 | Nephroscopic extraction of ureteric calculus |
| | V2- 7B18. | Ureteroscopic operations for ureteric calculus |
| | V2- 7B180 | Ureteroscopic laser lithotripsy of ureteric calculus |
| | V2- 7B181 | Other ureteroscopic fragmentation of ureteric calculus |
| | V2- 7B182 | Ureteroscopic extraction of ureteric calculus |
| | V2- 7B19. | Cystoscopic removal of ureteric calculus |
| | V2- 7B190 | Cystoscopic laser lithotripsy of ureteric calculus |
| | V2- 7B191 | Other cystoscopic fragmentation of ureteric calculus |
| | V2- 7B193 | Cystoscopic catheter drainage for ureteric calculus |
| | V2- 7B194 | Cystoscopic dilation of ureter for drainage of calculus |
| | V2- 7B19y | Other specified cystoscopic removal of ureteric calculus |
| | V2- 7B19z | Cystoscopic removal of ureteric calculus NOS |
| | V2- 7B1C. | Extracorporeal shockwave lithotripsy of ureteric calculus |

| Coding classification | Code | Definition |
|---|---|---|
| | V2- 7B1C0 | Extracorporeal shockwave lithotripsy of unspecified ureteric calculus |
| | V2- 7B1C1 | Extracorporeal shockwave therapy for stone in upper ureter |
| | V2- 7B1C2 | Extracorporeal shockwave lithotripsy for stone in mid-ureter |
| | V2- 7B1C3 | Extracorporeal shockwave lithotripsy for stone in lower ureter |
| | V2- 7B1Cy | Other specified extracorporeal shockwave lithotripsy of ureteric calculus |
| | V2- 7B1Cz | Extracorporeal shockwave lithotripsy of ureteric calculus NOS |
| | V2- C3411 | Uric acid nephrolithiasis |
| | V2- K112. | Hydronephrosis with renal and ureteral calculous obstruction |
| | V2- K12.. | Calculus of kidney and ureter |
| | V2- K120. | Calculus of kidney |
| | V2- K1200 | Staghorn calculus |
| | V2- K120z | Renal calculus NOS |
| | V2- K121. | Calculus of ureter |
| | V2- K122. | Calculus of kidney with calculus of ureter |
| | V2- K12z. | Urinary calculus NOS |
| | V2- Kyu3. | [X]Urolithiasis |
| | V2- R080. | [D]Renal colic |
| | V2- R0800 | [D]Renal colic, unspecified |
| | V2- R0801 | [D]Ureteric colic |
| | V2- R080z | [D]Renal colic NOS |
| | V3- 14D3. | H/O: urinary stone |
| | V3- 4G6.. | O/E - ureteric calculus |
| | V3- 7B07. | Percutaneous nephrolithotomy |
| | V3- 7B070 | Nephroscopy and ultrasound lithotripsy of renal calculus |
| | V3- 7B071 | Nephroscopy and electrohydraulic lithotripsy of renal calculus |
| | V3- 7B072 | Endoscopic laser fragmentation of renal calculus |
| | V3- 7B073 | Percutaneous nephrolithotomy without disintegration |
| | V3- 7B07y | Other specified percutaneous renal stone surgery |
| | V3- 7B07z | Percutaneous renal stone surgery NOS |
| | V3- 7B0B. | Extracorporeal shockwave lithotripsy for renal calculus |
| | V3- 7B0B0 | Extracorporeal shockwave lithotripsy for renal calculus of unspecified size |
| | V3- 7B0B1 | Extracorporeal shockwave lithotripsy for renal calculus less than 2 cm in diameter |
| | V3- 7B0B2 | Extracorporeal shockwave treatment for renal calculus of 2 cm or more in diameter |
| | V3- 7B0By | Other specified extracorporeal shockwave lithotripsy for renal calculus |
| | V3- 7B0Bz | Extracorporeal shockwave lithotripsy for renal calculus NOS |
| | V3- 7B170 | Nephroscopic laser fragmentation of ureteric calculus |
| | V3- 7B171 | Other nephroscopic fragmentation of ureteric calculus |
| | V3- 7B172 | Nephroscopic removal of ureteric calculus |
| | V3- 7B18. | Ureteroscopic operations for ureteric calculus |
| | V3- 7B180 | Ureteroscopic laser lithotripsy of ureteric calculus |
| | V3- 7B181 | Other ureteroscopic fragmentation of ureteric calculus |
| | V3- 7B182 | Ureteroscopic extraction of ureteric calculus |
| | V3- 7B19. | Cystoscopic operation for ureteric calculus |
| | V3- 7B190 | Cystoscopic laser lithotripsy of ureteric calculus |
| | V3- 7B191 | Other cystoscopic fragmentation of ureteric calculus |
| | V3- 7B193 | Cystoscopic catheter drainage for ureteric calculus |
| | V3- 7B194 | Cystoscopic dilation of ureter for drainage of calculus |
| | V3- 7B19y | Other specified cystoscopic removal of ureteric calculus |
| | V3- 7B19z | Cystoscopic removal of ureteric calculus NOS |
| | V3- 7B1C. | Extracorporeal shockwave lithotripsy of ureteric calculus |

| Coding classification | Code | Definition |
|---|---|---|
| | V3- 7B1C0 | Extracorporeal shockwave lithotripsy of unspecified ureteric calculus |
| | V3- 7B1C1 | Extracorporeal shockwave therapy for stone in upper ureter |
| | V3- 7B1C2 | Extracorporeal shockwave lithotripsy for stone in mid-ureter |
| | V3- 7B1C3 | Extracorporeal shockwave lithotripsy for stone in lower ureter |
| | V3- 7B1Cy | Other specified extracorporeal shockwave lithotripsy of ureteric calculus |
| | V3- 7B1Cz | Extracorporeal shockwave lithotripsy of ureteric calculus NOS |
| | V3- C3411 | Renal stone - uric acid |
| | V3- K112. | Hydronephrosis with renal and ureteral calculous obstruction |
| | V3- K1200 | Staghorn calculus |
| | V3- K120z | Renal calculus NOS |
| | V3- K121. | Calculus of ureter |
| | V3- K12z. | Urinary calculus NOS |
| | V3- R080. | [D]Renal colic |
| | V3- R0800 | [D]Renal colic, unspecified |
| | V3- R0801 | [D]Ureteric colic |
| | V3- R080z | [D]Renal colic NOS |
| | V3- X30Pl | Urolithiasis |
| | V3- X30Pm | Urinary calculus |
| | V3- X30Pn | Nephrolithiasis NOS |
| | V3- Xa07P | C/O - ureteric pain |
| | V3- Xa6m6 | Ureteroscopic operation for ureteric calculus |
| | V3- Xa8P3 | Cystoscopic extraction of ureteric calculus without disintegration |
| | V3- XE0dj | Calculus of kidney and ureter |
| | V3- XE0dk | Kidney calculus |
| | V3- XE0G7 | Ureteroscopic removal of ureteric calculus |
| | V3- XE0G8 | Cystoscopic extraction of ureteric calculus |
| | V3- XE2Pu | Ureteric colic |
| | V3- XM0CQ | C/O - ureteric colic |

# B | Inclusion criteria for renal cyst association analyses in the UK Biobank

| Coding classification | Code | Definition |
|---|---|---|
| International Classification of Diseases (ICD) | ICD9 - 75318 | Cystic kidney disease, other |
| | ICD9 - 75310 | Cystic kidney disease, unspecified |
| | ICD9 - 5932 | Cyst of kidney, acquired |
| | ICD9 - 75314 | Medullary cystic disease, juvenile |
| | ICD9 - 75315 | Medullary cystic disease, adult |
| | ICD10 - Q610 | Congenital single renal cyst |
| | ICD10 - Q615 | Medullary cystic kidney |
| | ICD10 - Q618 | Other cystic kidney diseases |
| | ICD10 - Q619 | Cystic kidney disease, unspecified |
| | ICD10 - N281 | Cyst of kidney, acquired |
| Primary care (Read) | ReadV2 - K132. | Acquired cyst of kidney |
| | ReadV2 - K132z | Acquired cyst of kidney NOS |
| | ReadV2 - K1320 | Single acquired kidney cyst |
| | ReadV2 - K1321 | Multiple acquired kidney cysts |
| | ReadV2 - 7B030 | Open deroofing of renal cyst |
| | ReadV2 - 7B054 | Therapeutic open aspiration of renal cyst |
| | ReadV2 - K1323 | Acquired renal cyst with neoplastic change |
| | ReadV2 - K1324 | Acquired renal cyst without neoplastic change |
| | ReadV2 - K13G. | Simple renal cyst |
| | ReadV2 - PD10. | Congenital renal cyst, single |
| | ReadV2 - PD1.. | Congenital cystic kidney disease |
| | ReadV2 - PD1y. | Other specified congenital cystic kidney disease |
| | ReadV2 - PD1yz | Other congenital cystic kidney disease NOS |
| | ReadV2 - PD1z. | Congenital cystic kidney disease NOS |
| | ReadV2 - Pyu70 | [X]Other cystic kidney diseases |
| | ReadV2 - PD12. | Medullary cystic disease |
| | ReadV2 - PD122 | Nephronophthisis - medullary cystic disease |
| | ReadV2 - PD12y | Medullary cystic disease OS |
| | ReadV2 - PD12z | Medullary cystic disease NOS |
| | ReadV3 - K132. | Acquired renal cyst |
| | ReadV3 - K132z | Acquired cyst of kidney NOS |
| | ReadV3 - K1320 | Single acquired kidney cyst |
| | ReadV3 - K1321 | Multiple acquired kidney cysts |
| | ReadV3 - 7B030 | Open deroofing of renal cyst |
| | ReadV3 - 7B054 | Therapeutic open aspiration of renal cyst |
| | ReadV3 - X30J9 | Acquired renal cyst with neoplastic change |
| | ReadV3 - X30JA | Acquired renal cyst without neoplastic change |

| Coding classification | Code | Definition |
|---|---|---|
| | ReadV3 - XaecJ | Simple renal cyst |
| | ReadV3 - PD10. | Congenital renal cyst, single |
| | ReadV3 - XE1LM | Congenital cystic disease of kidney |
| | ReadV3 - PD1y. | Other specified congenital cystic kidney disease |
| | ReadV3 - PD1yz | Other congenital cystic kidney disease NOS |
| | ReadV3 - PD1z. | Congenital cystic kidney disease NOS |
| | ReadV3 - Pyu70 | [X]Other cystic kidney diseases |
| | ReadV3 - PD12. | Medullary cystic disease |
| | ReadV3 - PD120 | Nephronophthisis |
| | ReadV3 - XE2bK | Medullary cystic disease, adult type |
| | ReadV3 - X30JB | Medullary sponge kidney |
| | ReadV3 - X785r | Autosomal dominant medullary cystic disease |
| | ReadV3 - PD12y | Medullary cystic disease OS |
| | ReadV3 - PD12z | Medullary cystic disease NOS |

# C | Exclusion criteria for renal cyst association analyses in the UK Biobank

| Coding classification | Code | Definition |
|---|---|---|
| International Classification of Diseases (ICD) | ICD9 - 75311 | Polycystic kidney, infantile type |
| | ICD9 - 75312 | Polycystic kidney, adult type |
| | ICD9 - 75313 | Polycystic kidney, unspecified |
| | ICD9 - 75315 | Renal dysplasia |
| | ICD9 - 75316 | Multicystic renal dysplasia |
| | ICD9 - 75317 | Medullary sponge kidney |
| | ICD10 - Q611 | Polycystic kidney, infantile type |
| | ICD10 - Q612 | Polycystic kidney, adult type |
| | ICD10 - Q613 | Polycystic kidney, unspecified |
| | ICD10 - Q614 | Renal dysplasia |
| Primary care (Read) | ReadV2 - 12F1. | FH: Polycystic kidney |
| | ReadV2 - 7B033 | Rovsing's operation for polycystic kidney |
| | ReadV2 - PD11. | Polycystic kidney disease |
| | ReadV2 - PD110 | Polycystic kidneys, infantile type |
| | ReadV2 - PD111 | Polycystic kidneys, adult type |
| | ReadV2 - PD11z | Polycystic kidney disease NOS |
| | ReadV2 - PD1y0 | Fibrocystic kidney disease |
| | ReadV3 - 12F1. | FH: Polycystic kidney |
| | ReadV3 - 7B033 | Rovsing's operation for polycystic kidney |
| | ReadV3 - PD11. | Polycystic kidney disease |
| | ReadV3 - PD110 | Autosomal recessive polycystic kidney disease |
| | ReadV3 - PD110 | Infantile polycystic kidney disease |
| | ReadV3 - PD111 | Polycystic kidneys - adult type |
| | ReadV3 - PD111 | Autosomal dominant polycystic kidney disease |
| | ReadV3 - X785l | Fibrocystic renal degeneration |
| | ReadV3 - X785m | Cystic kidney disease NEC |
| | ReadV3 - XE1LM | Sponge kidney |
| | ReadV3 - XE1LM | Fibrocystic kidney |
| | ReadV3 - XE1LN | Polycystic kidney disease NOS |
| | ReadV3 - XE1LO | Fibrocystic kidney disease |

# D | Inclusion criteria for renal cyst association analyses in *All of Us*

| Concept ID | Code | Definition |
|---|---|---|
| 44832487 | ICD9 - 593.2 | Cyst of kidney acquired |
| 44835056 | ICD9 - 753.1 | Cystic kidney disease unspecified |
| 44822258 | ICD9 - 753.11 | Congenital single renal cyst |
| 44821204 | ICD9 - 753.19 | Other specified cystic kidney disease |
| 35209298 | ICD10 - N28.1 | Cyst of kidney acquired |
| 1572109 | ICD10 - Q61.0 | Congenital renal cyst |
| 35211023 | ICD10 - Q61.8 | Other specified cystic kidney disease |
| 35211024 | ICD10 - Q61.9 | Cystic kidney disease unspecified |
| 36716270 | SNOMED - 722223000 | Cyst of kidney |

# E | Exclusion criteria for renal cyst association analyses in *All of Us*

| Concept ID | Code | Definition |
|---|---|---|
| 44824574 | ICD9 - 753.12 | Polycystic kidney unspecified type |
| 44820018 | ICD9 - 753.13 | Polycystic kidney autosomal dominant |
| 44828106 | ICD9 - 753.14 | Polycystic kidney autosomal recessive |
| 1572110 | ICD10 - Q61.1 | Polycystic kidney infantile type |
| 35211019 | ICD10 - Q61.2 | Polycystic kidney adult type |
| 35211020 | ICD10 - Q61.3 | Polycystic kidney unspecified |
| 201675 | SNOMED - 28770003 | Polycystic kidney disease infantile type |
| 35623051 | SNOMED - 765330003 | Autosomal dominant polycystic kidney disease |

# F | Imaging codes for renal cyst observational association analyses in the UK Biobank

| Read V2 code | Read V3 code | Definition |
|---|---|---|
| 5679. | 5679. | CT of abdomen |
| 567B. | 567B. | CAT scan - whole body |
| 5855. | 5855. | US scan of abdomen |
| 58550 | XaY40 | Ultrasound scan of lower abdomen |
| 58551 | XaY44 | Ultrasound scan of upper abdomen |
| 5859. | 5859. | US scan of gallbladder |
| 5859. | 5859. | US scan of gallbladder |
| 585A. | 585A. | Liver US scan |
| 585A. | 585A. | US scan of liver |
| 585A. | 585A. | Liver US scan |
| 585A. | 585A. | US scan of liver |
| 58DB. | X70ly | Ultrasound scan of aorta |
| R134. | R134. | [D]Gastrointestinal X-ray or scan abnormality |
| R135. | R135. | [D]Genitourinary X-ray or scan abnormality |
| R135z | R135z | [D]Genitourinary X-ray or scan abnormality NOS |
| R136. | R136. | [D]Abdominal area X-ray or scan abnormality |
| R1480 | R1480 | [D]Abnormal liver scan |
| 5696. | XaKab | Magnetic resonance imaging of abdomen abnormal |
| 569J. | X70oj | Magnetic resonance imaging of abdomen |
| 7P001 | XaLZN | Magnetic resonance imaging of whole body |
| 7P055 | X70oj | Magnetic resonance imaging of abdomen |
| 7P0N0 | X70om | Magnetic resonance imaging of kidneys |

# G | Imaging codes for renal cyst observational association analyses in *All of Us*

| Concept ID | Code | Definition |
|---|---|---|
| 42740713 | CPT4 - 0066T | Computed tomographic (CT) colonography (ie, virtual colonoscopy); screening (Deprecated) |
| 42740714 | CPT4 - 0067T | Computed tomographic (CT) colonography (ie, virtual colonoscopy); diagnostic (Deprecated) |
| 2211426 | CPT4 - 72195 | Magnetic resonance (eg, proton) imaging, pelvis; without contrast material(s) |
| 2211510 | CPT4 - 74150 | Computed tomography, abdomen; without contrast material |
| 2211511 | CPT4 - 74160 | Computed tomography, abdomen; with contrast material(s) |
| 2211512 | CPT4 - 74170 | Computed tomography, abdomen; without contrast material, followed by contrast material(s) and further sections |
| 40756931 | CPT4 - 74176 | Computed tomography, abdomen and pelvis; without contrast material |
| 40756910 | CPT4 - 74177 | Computed tomography, abdomen and pelvis; with contrast material(s) |
| 40756956 | CPT4 - 74178 | Computed tomography, abdomen and pelvis; without contrast material in one or both body regions, followed by contrast material(s) and further sections in one or both body regions |
| 2211514 | CPT4 - 74181 | Magnetic resonance (eg, proton) imaging, abdomen; without contrast material(s) |
| 2211515 | CPT4 - 74182 | Magnetic resonance (eg, proton) imaging, abdomen; with contrast material(s) |
| 2211516 | CPT4 - 74183 | Magnetic resonance (eg, proton) imaging, abdomen; without contrast material(s), followed by with contrast material(s) and further sequences |
| 2211517 | CPT4 - 74185 | Magnetic resonance angiography, abdomen, with or without contrast material(s) |
| 40756913 | CPT4 - 74261 | Computed tomographic (CT) colonography, diagnostic, including image postprocessing; without contrast material |
| 40757081 | CPT4 - 74262 | Computed tomographic (CT) colonography, diagnostic, including image postprocessing; with contrast material(s) including non-contrast images, if performed |
| 40757125 | CPT4 - 74263 | Computed tomographic (CT) colonography, screening, including image postprocessing |
| 2211740 | CPT4 - 76700 | Ultrasound, abdominal, real time with image documentation; complete |
| 2211741 | CPT4 - 76705 | Ultrasound, abdominal, real time with image documentation; limited (eg, single organ, quadrant, follow-up) |

| Concept ID | Code | Definition |
|---|---|---|
| 2211742 | CPT4 - 76770 | Ultrasound, retroperitoneal (eg, renal, aorta, nodes), real time with image documentation; complete |
| 2211743 | CPT4 - 76775 | Ultrasound, retroperitoneal (eg, renal, aorta, nodes), real time with image documentation; limited |
| 2211767 | CPT4 - 76856 | Ultrasound, pelvic (nonobstetric), real time with image documentation; complete |
| 2006920 | ICD9 - 88.74 | Diagnostic ultrasound of digestive system |
| 2006921 | ICD9 - 88.75 | Diagnostic ultrasound of urinary system |
| 2006932 | ICD9 - 88.76 | Diagnostic ultrasound of abdomen and retroperitoneum |
| 2792320 | ICD10 - BU3CY0Z | Magnetic Resonance Imaging (MRI) of Uterus and Ovaries using Other Contrast, Unenhanced and Enhanced |
| 3009344 | LOINC - 30704-1 | US Abdomen limited |
| 3051514 | LOINC - 36103-0 | CT Abdomen limited WO contrast |
| 3049940 | LOINC - 36424-0 | CT Abdomen WO contrast |
| 40771797 | LOINC - 69276-4 | US Abdominal Aorta |
| 21491943 | LOINC - 79374-5 | US Abdominal Aorta for screening |
| 4023136 | SNOMED - 105377009 | Ultrasonography of liver |
| 4061009 | SNOMED - 169070004 | CT of abdomen |
| 4058336 | SNOMED - 169071000 | CT of pelvis |
| 4060500 | SNOMED - 169072007 | CAT scan - whole body |
| 4082524 | SNOMED - 241460001 | Ultrasound angiography of aorta |
| 4082868 | SNOMED - 241480000 | US scan of adrenals |
| 4082869 | SNOMED - 241481001 | US scan of kidney and ureters |
| 4083230 | SNOMED - 241621009 | MRI of abdomen |
| 4083232 | SNOMED - 241625000 | MRI of pancreas |
| 4086677 | SNOMED - 24848001 | Pelvic echography |
| 4097527 | SNOMED - 2690005 | MRI of pelvis |
| 4116191 | SNOMED - 302801006 | US scan of genitourinary system |
| 4125530 | SNOMED - 303917008 | US urinary tract |
| 4147846 | SNOMED - 306005 | Echography of kidney |
| 36713202 | SNOMED - 395771000119107 | CT of pelvis without contrast |
| 37109313 | SNOMED - 396101000119105 | CT of abdomen without contrast |
| 4171400 | SNOMED - 419084009 | CT of urinary tract |
| 4303252 | SNOMED - 419739000 | US scan and biopsy of liver |
| 4138003 | SNOMED - 425666002 | Computed tomography of colon |
| 40479585 | SNOMED - 441538004 | CT of small intestine |
| 4261497 | SNOMED - 45036003 | Ultrasonography of abdomen |
| 37109428 | SNOMED - 480621000119104 | PET CT of whole body |
| 46271499 | SNOMED - 709631008 | CT guided biopsy of retroperitoneum |
| 42535859 | SNOMED - 734358005 | CT angiography of abdominal and/or pelvic blood vessel with contrast |

# Scientific outputs

## Publications

- **Lovegrove CE**, Howles SA, Furniss D, et al. Causal inference in health and disease: a review of the principles and applications of Mendelian randomization. Journal of Bone and Mineral Research. 2024 Aug 21; 10.1093/jbmr/zjae136.
- **Lovegrove CE**, Bešević J, Wiberg A, et al. Central Adiposity Increases Risk of Kidney Stone Disease through Effects on Serum Calcium Concentrations. Journal of the American Society of Nephrology. 2023 Dec; 10.1681/ASN.0000000000000238.
- Geraghty R, **Lovegrove CE** , Howles S, et al. Role of Genetic Testing in Kidney Stone Disease: A Narrative Review. Curr Urol Rep. 2024 Dec; 10.1007/s11934-024-01225-5.

## Presentations at scientific meetings

### International

#### Oral presentations

- Society for Endocrinology, British Endocrine Society Annual Conference, Glasgow, UK. (13-15 November 2023) “Mendelian randomisation and colocalization analyses reveal novel drug targets for the prevention of kidney stone disease by modulating serum calcium and phosphate concentrations”.
- Society for Endocrinology, British Endocrine Society Annual Conference, Harrogate, UK. (14-16 November 2022) "Central adiposity raises serum calcium concentrations and increases risk of kidney stone disease"
- Royal Society of Medicine Section of Urology Winter Meeting 2023, Livigno,

Italy. (20-27 January 2023) "Preventing stones: nature or nurture?"

- Society for Endocrinology, British Endocrine Society Annual Conference, Harrogate, UK. (14-16 November 2022) "Central adiposity raises serum calcium concentrations and increases risk of kidney stone disease"

## National

### Oral presentations

- BAUS Endourology Annual Conference, Harrogate. (13-15th September 2023) "Mendelian randomisation and colocalisation analyses reveal novel drug targets for the prevention of kidney stone disease"

- BAUS Endourology Annual Conference, Southampton. (8-9th September 2022) "SLC34A1-associated hypercalcaemia and hypophosphataemia in kidney stone disease"

- BAUS Annual Conference, Birmingham. (13-15th June 2022) "Central adiposity increases risk of kidney stone disease in part through effects on serum calcium concentrations"

### Moderated poster presentations

- BAUS Endourology Annual Conference, Edinburgh. (10-11th September 2024) "Causal effects of renal cysts on kidney stone disease"

- BAUS Annual Conference, Birmingham. (24-26th June 2024) "Genetic variants associated with *DGKD*, *CYP24A1*, and *SLC34A1* predisposing to increased risk of kidney stone disease via effects on serum calcium or phosphate concentrations"

- BAUS Annual Conference, Birmingham. (24-26th June 2024) "Causal effects of renal cysts on kidney stone disease"

## Local

### Oral presentations

- University of Oxford Clinical Academic Graduate School Annual Symposium, Oxford (September 2024) "Causal effects of renal cysts on kidney stone disease"
- "Perturbations in CASR-signalling alter serum calcium concentrations to increase risk of kidney stone disease through shared genetic mechanisms", Oxford MRC-DTP Symposium, June 2023

### Moderated poster presentations

- University of Oxford Clinical Academic Graduate School Annual Symposium, Oxford (September 2024) "Genetic variants associated with *DGKD*, *CYP24A1*, and *SLC34A1* predisposing to increased risk of kidney stone disease via effects on serum calcium or phosphate concentrations"
- Nuffield Department of Surgical Sciences Student Symposium, University of Oxford. (1st July 2024) "Genetic variants associated with *DGKD*, *CYP24A1*, and *SLC34A1* predisposing to increased risk of kidney stone disease via effects on serum calcium or phosphate concentrations"
- University of Oxford Medical Research Council Annual Symposium, Oxford (June 2024) "Causal effects of renal cysts on kidney stone disease"