

The Genetic Architecture of Carpal Tunnel Syndrome

A dissertation submitted to the University of Oxford in partial fulfilment of the
requirements for the degree of

Doctor of Philosophy



Akira Wiberg

Merton College

Trinity Term 2019

Abstract

Carpal tunnel syndrome (CTS) is a debilitating disease of the hand caused by compression of the median nerve within an anatomical tunnel in the wrist. It is a common disease with an estimated population prevalence of 5–10%. Symptoms include pain, paraesthesia, and numbness of the hand, and thenar weakness that ultimately leads to functional impairment. Many patients require surgery to relieve the pressure on the median nerve; although surgery is successful in the majority of patients, a significant sub-group experiences persistent or recurrent symptoms.

Despite being such a common disease, the pathophysiology of CTS is poorly understood, and even less is known about the genetic contribution to the disease. The heritability of CTS is estimated at 0.46, and approximately one-third of patients report a family history. CTS is a complex disease, where genetic and non-genetic factors interact to affect overall phenotypic expression. The current best method for interrogating a complex disease is with a genome-wide association study (GWAS).

I present the findings from the first ever genome-wide association of CTS, undertaken in over 12,000 CTS cases and nearly 400,000 controls from UK Biobank. The GWAS uncovered 16 genomic loci significantly associated with CTS. *In silico* analyses of the GWAS data using an array of bioinformatic tools discovered genetic variants at the GWAS-associated loci with strong evidence of functionality, and implicated several biologically plausible genes in CTS

pathogenesis, including *EFEMP1*, *ADAMTS17* and *ADAMTS10*. These genes were strongly enriched for biological pathways and gene ontologies pertaining to the extracellular matrix.

These analyses were followed by epidemiological investigations in UK Biobank, including Mendelian randomisation to assess the causal role of various putative risk factors on CTS aetiology. These studies identified a causal role for shorter height, greater BMI, and younger age at menarche in the development of CTS.

In the final results chapter, a polygenic risk score was constructed to validate the genetic variants discovered in the GWAS in a small cohort of deeply phenotyped CTS patients. RNA-sequencing performed in tenosynovial tissues collected from these patients demonstrated high expression of the candidate genes that were identified in the GWAS. Several GWAS loci demonstrated genotype-specific expression of these genes, providing evidence of functionality.

This work represents the first use of big data and computational genetics to delineate the genetic basis of CTS. I have provided evidence for a robust genetic architecture underlying CTS and have identified several biologically plausible candidate genes that contribute to its pathophysiology. The insights into the aetiology of CTS have revealed testable functional hypotheses and potential therapeutic avenues for further research.

Acknowledgements

First and foremost, I must give my sincerest thanks to my two supervisors, Dominic Furniss and David Bennett, for their invaluable guidance, wisdom and mentorship over the course of my doctoral studies. I am privileged to have had such generous and committed supervisors, both of whom are outstanding role models as clinician-scientists. Dominic has shown me that being a plastic surgeon and a scientist needn't be mutually exclusive, and I am grateful for his efforts to foster my development as both.

I am also indebted to Annina Schmid for her unfailingly sound scientific advice and for her instruction in the lab, and to Michael Ng for the many hours spent teaching and advising me on all matters computational. I must also thank Giorgios Baskozos for his help with RNA-Seq, and Michael Holmes for sharing his expertise on Mendelian randomisation.

Running the MGCTS study would not have been possible without the enthusiasm and tireless efforts of Edmarcia Araujo. I am also grateful to the hand surgeons at the Nuffield Orthopaedic Centre for their valuable contributions, and of course, all the patients who volunteered to participate in the study.

Undertaking this doctorate was made possible by grant funding in the form of a Clinical Research Training Fellowship from the Medical Research Council.

I have been a student at Merton College for nearly a decade now. My former tutor, David Paterson, was a formative influence in my scientific development, and originally inspired and encouraged me to pursue a DPhil one day. For that I am very grateful. I am also thankful to Julian Knight for generously giving his time, expertise and considered advice at several junctures during my DPhil. I count myself immensely fortunate to have spent a substantial portion of my life as a member of the beautiful, ancient, and *sui generis* academic community that is Merton. Alas, it is time to move on.

Finally, I would like to thank my parents, Anders and Yoko, for always having provided me with the opportunities and encouragement to pursue my academic ambitions; and Felicity, for all her kindness and support throughout this journey.

Declaration

Every piece of work that is described in this thesis was conducted solely by me, with the exception of the following:

Chapter 2: The genetic association study in the Estonian cohort was performed by Dr Reedik Mägi and Dr Kadri Künnapuu.

Chapter 3: Comparison of gene expression in human fibroblasts vs Schwann cells from publicly available RNA-Seq data was performed by Dr Giorgios Baskozos.

Chapter 4: The cross-phenotype association analysis between migraine and CTS was performed by Dr Michael Ng.

Chapter 5: Recruitment and phenotyping of the PiNS CTS cohort was performed by Dr Annina Schmid. Whole-genome genotyping and RNA-sequencing were performed by the Oxford Genomics Centre at the Wellcome Trust Centre for Human Genetics. Processing of raw RNA-Seq data was performed by Dr Giorgios Baskozos.

Akira Wiberg

May 2019

Abbreviations

BMI	Body mass index
bp	Base pair
CI	Confidence interval
CTD	Carpal tunnel decompression surgery
CTS	Carpal tunnel syndrome
DNA	Deoxyribonucleic acid
ECM	Extracellular matrix
EDT	Electrodiagnostic testing
EGCUT	Estonian Genome Center at the University of Tartu
eQTL	Expression quantitative trait locus/loci
FDR	False discovery rate
IV	Instrumental variable
GWAS	Genome-wide association study/studies
HR	Hazard ratio
HRT	Hormone replacement therapy
HWE	Hardy-Weinberg equilibrium
IENFD	Intra-epidermal nerve fibre density
Indel	Insertion-deletion polymorphism
kb	Kilobase
LD	Linkage disequilibrium
LDSC	LD score regression

LMM	Linear mixed model
MAF	Minor allele frequency
Mb	Megabase
MGCTS	Molecular Genetics of Carpal Tunnel Syndrome Study
MR	Mendelian randomisation
NCS	Nerve conduction study
NHS	National Health Service
OR	Odds ratio
PCA	Principal component analysis
PiNS	Pain in Neuropathy Study
PRS	Polygenic risk score/scores
QC	Quality control
Q-Q	Quantile-quantile plot
QST	Quantitative sensory testing
RA	Rheumatoid arthritis
RCT	Randomised control trial
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
SD	Standard deviation
SE	Standard error
SNAP	Sensory nerve action potential
SNP	Single nucleotide polymorphism
SSCT	Sub-synovial connective tissues
TSL	Thermal sensory limen
VAS	Visual Analogue Scale

Table of Contents

Chapter 1	16
1.1 The anatomy of the median nerve and carpal tunnel	16
1.2 Carpal tunnel syndrome	20
1.2.1 Definition	20
1.2.2 Epidemiology.....	21
1.2.3 Diagnosis	22
1.2.4 Treatment.....	24
1.2.5 Socioeconomic burden of CTS	26
1.3 Pathophysiology of CTS	28
1.3.1 History.....	28
1.3.2 Increased carpal tunnel pressures.....	29
1.3.3 Nerve Injury	30
1.3.4 The sub-synovial connective tissues (SSCT)	31
1.4 Aetiology	34
1.4.1 Anthropometric factors.....	35
1.4.2 Body mass index (BMI)	35
1.4.3 Occupational factors.....	36
1.4.4 Pregnancy, menopause and female hormones.....	36
1.4.5 Associated diseases.....	37
1.4.5.1 Diabetes.....	38
1.4.5.2 Rheumatoid arthritis	38
1.4.5.3 Hypothyroidism.....	39
1.4.6 Pain Comorbidities	39

1.5	The genetic aetiology of CTS	41
1.5.1	Population-based studies	41
1.5.2	Familial CTS	42
1.5.3	Candidate gene studies.....	43
1.6	CTS as a complex disease	44
1.7	Genome-wide association studies (GWAS).....	46
1.8	Scope of the thesis.....	50
Chapter 2		52
2.1	Introduction.....	52
2.1.1	Aims and Hypotheses	53
2.2	Methods	56
2.2.1	The UK Biobank Resource	56
2.2.2	Ethics.....	57
2.2.3	Quality control (QC) of genotypic data	57
2.2.3.1	Sample QC.....	58
2.2.3.2	SNP QC.....	60
2.2.3.3	Pre-association study exclusions.....	61
2.2.4	Case Ascertainment	62
2.2.5	Association Study	64
2.2.6	Replication in EGCUT.....	66
2.2.6.1	Description of the EGCUT cohort.....	66
2.2.6.2	Genotyping, QC, and association analysis.....	66
2.2.6.3	Genetic correlations between UK Biobank and EGCUT.....	67
2.3	Results	68
2.3.1	Sixteen novel loci associated with CTS risk	68
2.3.2	Replication in EGCUT cohort.....	76

2.4	Discussion.....	78
Chapter 3	86	
3.1	Introduction.....	86
3.1.1	Background	86
3.1.2	Aims	87
3.2	Methods	88
3.2.1	Functional annotation of associated loci.....	88
3.2.2	Gene-based analysis and gene-property analysis	88
3.2.3	Gene-set enrichment analyses	89
3.2.4	Summary-based Mendelian randomisation.....	89
3.2.5	Partitioned heritability analyses	90
3.2.6	Expression of candidate genes in fibroblasts vs Schwann cells	91
3.2.7	LD score correlation.....	91
3.2.8	Height and BMI in UK Biobank.....	92
3.2.9	Polygenic risk score (PRS).....	92
3.3	Results	94
3.3.1	Evidence of functional SNPs at the associated loci	94
3.3.2	Gene mapping	94
3.3.3	Enrichment for extracellular matrix (ECM) ontologies.....	98
3.3.4	Summary-based Mendelian randomisation in fibroblasts implicates a key gene in the TGF- β signalling pathway	101
3.3.5	Partitioned heritability analysis shows enrichment for regulatory regions, musculoskeletal and connective tissues.....	102
3.3.6	Top candidate genes are highly expressed in Schwann cells and fibroblasts	103
3.3.7	Genetic correlation with anthropometric phenotypes.....	105
3.3.8	Observed anthropometric measures mirror the genetic correlations	107
3.3.9	Polygenic risk score associates with disease severity	108

3.4	Discussion.....	111
Chapter 4	121	
4.1	Introduction.....	121
4.1.1	Mendelian Randomisation.....	122
4.1.2	Migraine and CTS.....	124
4.1.3	Aims	126
4.2	Methods	128
4.2.1	Nested case-control study of CTS and other disease phenotypes.....	128
4.2.1.1	Phenotyping	128
4.2.1.2	Case-control matching for four disease phenotypes.....	129
4.2.1.3	Odds ratios	130
4.2.2	Application of polygenic risk score (PRS) to other diseases and phenotypes.....	130
4.2.3	Sex-stratified GWAS	131
4.2.4	Mendelian randomisation analyses.....	132
4.2.4.1	Height.....	135
4.2.4.2	Body mass index (BMI)	135
4.2.4.3	Diabetes.....	136
4.2.4.4	Rheumatoid arthritis (RA).....	136
4.2.4.5	Hypothyroidism.....	137
4.2.4.6	Age at menarche.....	137
4.2.4.7	Age at menopause	137
4.2.5	Genetic correlations and enrichment between CTS and migraine	138
4.3	Results	140
4.3.1	Nested case-control studies in UK Biobank provide further evidence for CTS-disease associations	140
4.3.2	Polygenic risk score for CTS is raised in females with hypothyroidism.....	140
4.3.3	Sex differences in the genetic architecture of CTS.....	144

4.3.4	Two female-specific risk loci for CTS.....	148
4.3.5	Mendelian randomisation analyses suggest a causal role for height, BMI, and age at menarche in CTS aetiology.....	149
4.3.6	Weak evidence of genetic correlation between migraine and CTS.....	153
4.4	Discussion.....	156
Chapter 5		168
5.1	Introduction.....	168
5.1.1	From genotype to phenotype.....	169
5.1.2	Focus on the sub-synovial connective tissues (SSCT)	169
5.1.3	RNA-Sequencing.....	170
5.1.4	Aims	171
5.2	Methods	173
5.2.1	Description of Oxford cohorts.....	173
5.2.1.1	PiNS.....	173
5.2.1.2	MGCTS	175
5.2.2	PiNS phenotyping.....	175
5.2.2.1	Questionnaires.....	176
5.2.2.2	Electrodiagnostic tests	176
5.2.2.3	Quantitative sensory testing (QST)	177
5.2.2.4	Skin biopsy and confocal imaging	177
5.2.3	DNA extraction	178
5.2.4	Whole-genome genotyping	179
5.2.5	Quality control of genotyped data	179
5.2.6	Imputation.....	180
5.2.7	Polygenic risk score in the Oxford cohorts.....	181
5.2.8	PRS-phenotype analyses	182
5.2.9	RNA extraction from carpal tunnel tenosynovium.....	183

5.2.10	RNA-Seq of tenosynovial RNA	184
5.2.11	Genotype-specific gene expression in tenosynovium	186
5.2.12	Comparison of gene expression in tenosynovium vs index finger skin.....	188
5.2.13	Analysis of female sex hormone receptors	189
5.3	Results	190
5.3.1	Polygenic risk score analyses	190
5.3.2	PRS-phenotype association and correlation analyses.....	192
5.3.3	Candidate genes are highly expressed in tenosynovium	195
5.3.4	Genotype-specific expression analyses.....	197
5.3.5	Oestrogen and progesterone receptors are expressed in tenosynovium.....	202
5.4	Discussion.....	203
Chapter 6		213
6.1	Summary of key findings.....	213
6.2	Four postulates on the genetic aetiology of CTS	215
6.2.1	Postulate 1: Genetic variants linked to skeletal growth predispose to CTS.....	215
6.2.2	Postulate 2: Genetic variants linked to adiposity predispose to CTS.....	215
6.2.3	Postulate 3: Genetic variants acting in the SSCT predispose to CTS.....	216
6.2.4	Postulate 4: Genetic variants that predispose to CTS exert greater effects in females	217
6.2.5	Integration.....	217
6.3	Broader implications	219
6.3.1	CTS as a genetic disease	219
6.3.2	Other entrapment neuropathies	220
6.3.3	BMI – a modifiable risk factor?	221
6.4	Future directions	223
6.4.1	X-chromosome analysis	223
6.4.2	Functional studies.....	223

6.4.3	Migraine and CTS.....	225
6.5	Concluding remarks.....	227
	<i>Scientific outputs</i>	230
	<i>Bibliography</i>	232

“Man alone has a hand. He uses it as a tool, as a symbol, and as a weapon. A whole literature of legend, folklore, superstition, and myth has been built up around the human hand. As an organ of performance it serves as eyes for the blind, the mute talk with it, and it has become a symbol of salutation, supplication, and condemnation. The hand has played a part in the creative life of every known society, and it has come to be symbolic or representative of the whole person in art, in drama, and in the dance”.

EJ Alpenfels¹, The Anthropology and Social Significance
of the Human Hand

“Genetic architecture describes the characteristics of genetic variation that are responsible for heritable phenotypic variability... Defining the genetic architecture of a complex trait or disease is central to the scientific and clinical goals of human genetics, which are to understand disease aetiology and aid in disease screening, diagnosis, prognosis and therapy. Recent technological advances have enabled genome-wide association studies and emerging next-generation sequencing studies to begin to decipher the nature of the heritable contribution to traits and disease”.

NJ Timpson et al.², Genetic architecture: the shape of the
genetic contribution to human traits and disease

Chapter 1

Introduction

1.1 The anatomy of the median nerve and carpal tunnel

One of the most striking observations when comparing the human hand and the chimpanzee hand is the difference in the size of the thumb (**Fig 1.1**). Whereas the chimpanzee has a comparatively small thumb relative to its long, slender fingers, the human thumb is larger, more mobile, and crucially, fully opposable³.

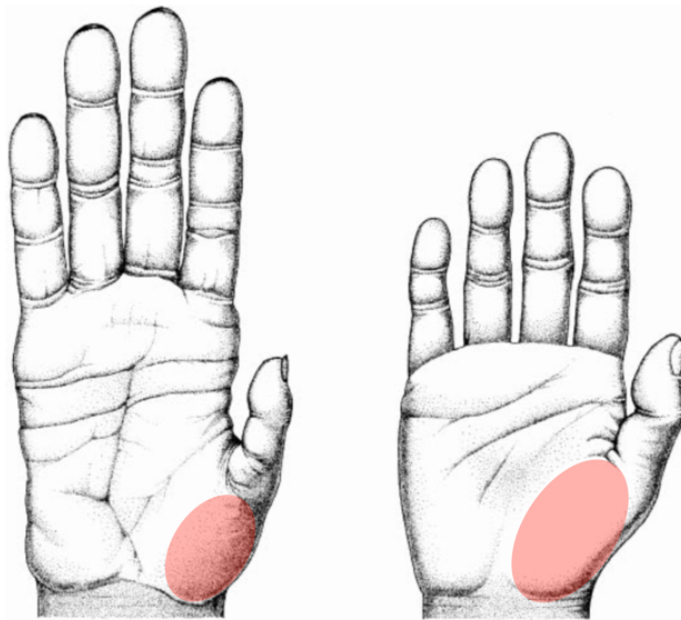


Fig. 1.1 Human vs chimpanzee hand. The chimpanzee hand (left) has longer and more slender fingers than the human hand (right), with a proportionally smaller thumb and less developed thenar muscles (in red). Adapted from R. W. Young³.

There is an intimate evolutionary relationship between human bipedalism and anatomical modifications to the hand. As the forelimbs ceased to be used for knuckle-walking, the newly emancipated hands were freed for the use of successively more complex tools, and central to the fine motor control required for tool use was thumb opposability. Most primates display some degree of functional independence of the thumb, but there is considerable variation in the extent and precision of opposition; humans exhibit by far the greatest proficiency.

Thumb opposition is made possible by the median nerve, which is formed from the union of the lateral and medial cords of the brachial plexus (anterior primary rami of the C₅ to T₁ spinal nerve roots), with the thenar muscles at the base of the thumb deriving their nerve supply from the C₈ and T₁ nerve roots. The median nerve's functions in the forearm include motor supply to the two pronator muscles and most of the flexor muscles of the anterior compartment of the forearm. Prior to entering the hand, the median nerve gives off a palmar cutaneous branch in the distal forearm, which supplies sensation to the central palm. In the hand, the median nerve divides into the digital cutaneous branches (proper palmar digital branch and the common palmar digital branch), that supply sensory innervation to the palmar aspects of the thumb, index, middle fingers and the radial half of the ring finger[†], and the muscular branch (also known as the thenar branch or recurrent branch), which supplies the thenar muscles (opponens pollicis, abductor pollicis brevis, and flexor pollicis brevis[‡]). It is these thenar muscles at the base of the thumb that permit the thumb to be opposed, and they are significantly better developed in the human compared to the chimpanzee – a reflection of our divergent evolutionary paths since our last common ancestor. Combined with the fact that most of our tactile interactions with the world occur through the palmar aspect of the thumb and index finger, the critical importance that the median nerve has played in human evolution becomes

[†] These digital branches also supply the first and second lumbrical muscles.

[‡] Flexor pollicis brevis is dually innervated by the median nerve and the ulnar nerve.

readily apparent. The anatomical territory innervated by the median nerve in the hand is shown in **Fig. 1.2**.

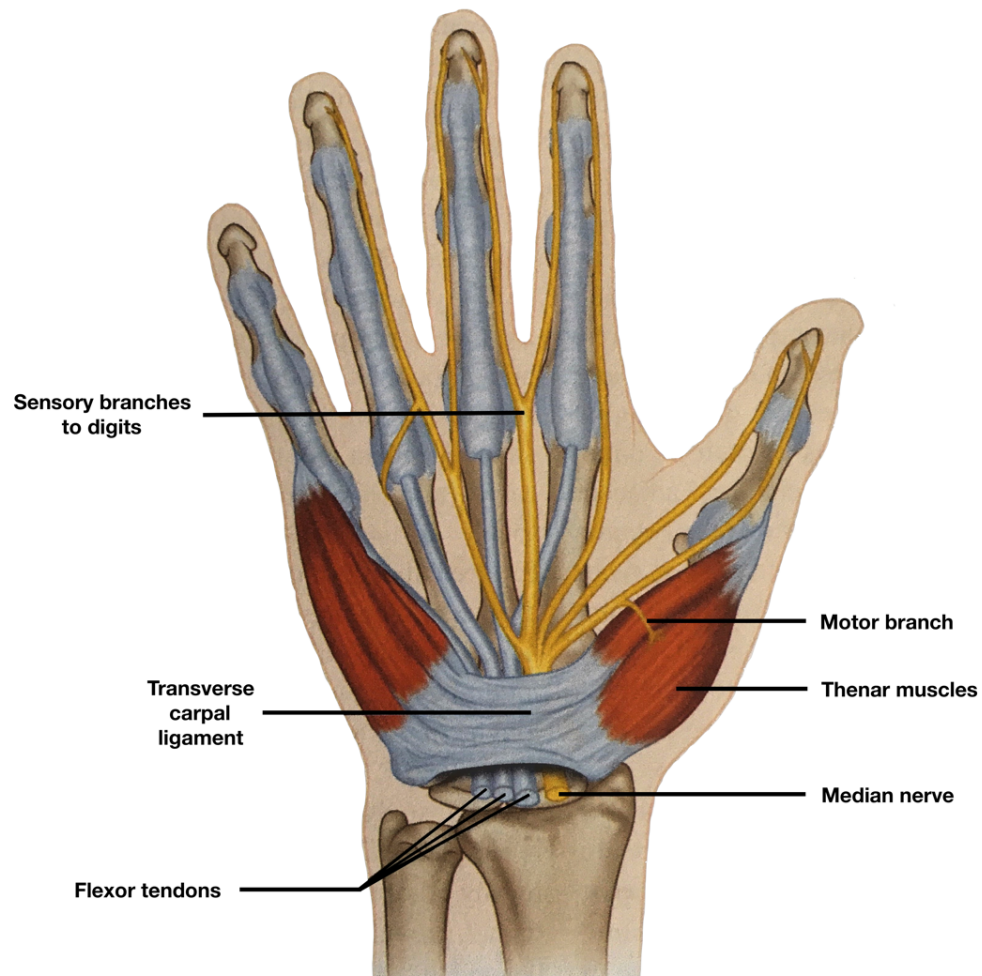


Fig. 1.2 Anatomy of the median nerve in the hand. The various branches of the median nerve in the hand are shown. Adapted from Duncan et al.⁴

The median nerve enters the hand by passing through the carpal tunnel, a narrow anatomical tunnel formed by the U-shaped configuration of the carpal bones that form its base. Attached to the scaphoid tubercle and the ridge of the trapezium on the radial side, and to the pisiform and hook of the hamate on the ulnar side, is the transverse carpal ligament (also known as the flexor retinaculum) – a thick, fibrous ligament - which forms the roof of the carpal tunnel. Traversing the carpal tunnel along with the median nerve are nine flexor tendons to the digits and thumb, meaning that the median nerve is vulnerable to compression within this anatomically crowded tunnel of confined dimensions (**Fig. 1.3**).

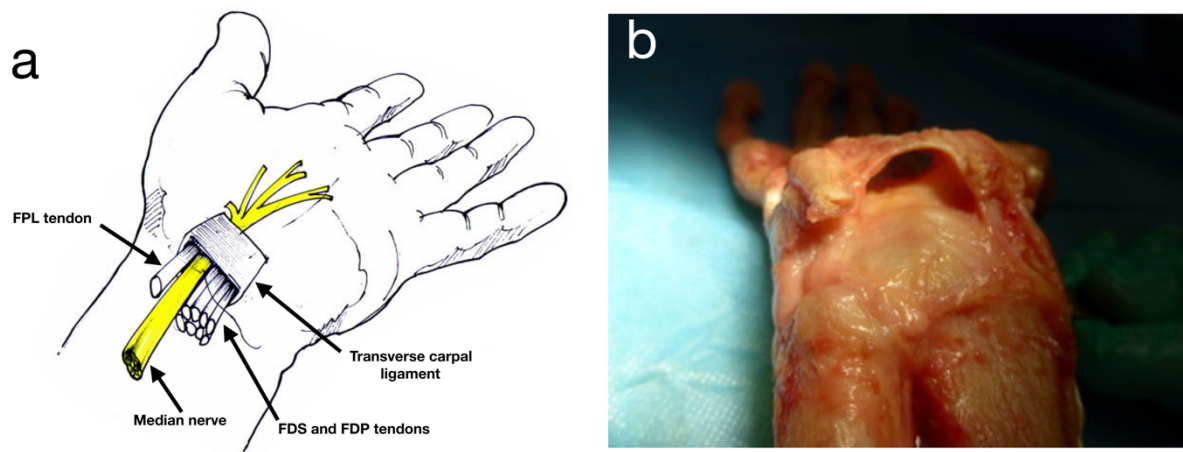


Fig. 1.3 The carpal tunnel. (a) The median nerve is seen entering the carpal tunnel in the wrist along with nine flexor tendons. (b) A cadaveric dissection of the wrist demonstrating the dimensions of the carpal tunnel. The transverse carpal ligament can be seen forming the roof of the tunnel. Adapted with permission from Mr Donald Sammut.

1.2 Carpal tunnel syndrome

1.2.1 Definition

The American Academy of Orthopedic Surgeons (AAOS) “Clinical Guideline on Diagnosis of Carpal Tunnel Syndrome”⁵ defines it as follows:

Carpal Tunnel Syndrome is a symptomatic compression neuropathy of the median nerve at the level of the wrist, characterized physiologically by evidence of increased pressure within the carpal tunnel and decreased function of the nerve at that level. Carpal Tunnel Syndrome can be caused by many different diseases, conditions and events. It is characterized by patients as producing numbness, tingling, hand and arm pain and muscle dysfunction. The disorder is not restricted by age, gender, ethnicity, or occupation and is associated with or caused by systemic disease and local mechanical and disease factors.

Carpal Tunnel Syndrome (CTS) is the commonest of the entrapment neuropathies[‡] – a subdivision of peripheral neuropathies characterised by peripheral nerve injury secondary to entrapment at specific anatomic locations, such as where a nerve courses through fibro-muscular or fibro-osseous structures, or is deformed by a fibrous band⁶. Although only small segments of a peripheral nerve are typically affected by entrapment neuropathies, they can have profound physical, psychological, and economic consequences for patients⁷.

[‡] Also known as compression neuropathies.

1.2.2 Epidemiology

CTS is a common disease; however, the reported incidence and prevalence of CTS vary widely depending on the diagnostic criteria used and on the population characteristics in different studies. The authors of one of the largest and most robust of such studies undertook a postal survey of hand symptoms in a randomly selected population cohort of 3,000 subjects in Sweden, followed by clinical examination and nerve conduction testing in those who reported CTS-like symptoms⁸. The authors determined that the population prevalence of clinically diagnosed CTS was 3.8%, whereas the prevalence of CTS diagnosed on nerve conduction testing was 4.9%. A British study employing similar methods in the general population estimated CTS prevalence to be between 7 to 16%, depending on different electrophysiological cut-offs used to define delayed median nerve conduction⁹, while a Dutch study estimated the prevalence of electrophysiologically confirmed CTS to be 5.8% in women and 0.6% in men¹⁰. Overall, it is believed that approximately one in ten individuals will develop clinical symptoms of CTS at some point in their life¹¹.

There is wide agreement that CTS is significantly more common in women than it is in men. Again, there is variation in the sex-specific incidence of CTS reported in the literature, with estimates of female:male prevalence ratios from the largest studies reported as 1.4:1⁸, 2:1¹², 2.9:1¹³, and 3.6:1¹⁴. There is also consensus that CTS is most common in patients over 40 years of age¹⁵, with the incidence peaking in the late 50s, and the late 70s, when the sex ratio is more equal; this bimodal distribution of CTS prevalence is more pronounced in women than it is in men¹².

1.2.3 Diagnosis

An accurate clinical history and assessment is considered the gold standard for the diagnosis of CTS⁷. Although the constellation of symptoms that define CTS differs from individual to individual, typically, patients first experience nocturnal paraesthesias and dysaesthesias in the median nerve-innervated territory of the hand, frequently describing having to “shake the hand” to relieve such symptoms. This is usually followed by sensory disturbance of the median-innervated digits (occasionally the whole hand), sometimes in conjunction with pain, which may radiate up the forearm and arm, towards the shoulder¹⁶. Later in the disease, there may be permanent sensory loss in the median-innervated digits, and weakness and atrophy of the thenar muscles, leading ultimately to functional impairment of the hand. Most patients present initially with symptoms affecting the dominant hand, although about 50% cases are bilateral at first presentation¹².

In addition to a convincing clinical history, a careful clinical examination is necessary to diagnose CTS. The most widely used provocative tests are Tinel’s sign (tapping over the carpal tunnel to provoke paraesthesia in the median nerve-innervated digits) and Phalen’s test (the provocation of median nerve paraesthesia by holding the wrist in 90 degrees of flexion for 60 seconds). While frequently performed in clinic, the diagnostic yield of these tests is widely debated, with sensitivity ranges from 42% to 85% for Phalen’s test, and 38% to 100% for Tinel’s test¹⁷. Clinical examination must also include sensibility testing (ideally quantitative sensory testing in the form of Von Frey’s test of punctate mechanical sensitivity using Semmes-Weinstein monofilaments, and Weber’s two-point discrimination test), and examination of the thenar muscles for weakness and atrophy. Finally, other potential causes of neurological disturbance in the hand, such as cervical radiculopathy, thoracic outlet syndrome, ulnar neuropathy, and generalised peripheral neuropathies, must be excluded.

There is controversy regarding the need for confirmatory tests in the form of electrophysiological testing⁷; however, there is broad consensus that these adjunctive electrophysiological tests should be undertaken in cases where there is diagnostic doubt⁷. Both sensory and motor nerve conduction studies (NCS) can be a sensitive tool for examining median nerve dysfunction, and is able to define the degree of demyelination and axonal loss¹⁸. Moreover, neurophysiological severity on the 7-point Canterbury scale has been shown to correlate with surgical prognosis¹⁹. However, NCS has an estimated false negative rate of 5-10%, with a series of 3,544 carpal tunnel decompressions demonstrating that 4.3% of patients had normal preoperative NCS¹⁶.

High-resolution ultrasonography is an imaging modality that can assist in the diagnosis of CTS. Morphological changes to the median nerve can be identified, such as a reduction in the median nerve volume at the site of compression, and an increase in cross-sectional area of the nerve just proximal to the lesion⁷. Such imaging has been demonstrated in a meta-analysis to have a sensitivity of 77% and a specificity of 87% for diagnosing CTS, compared with clinical diagnosis as a gold-standard²⁰. Magnetic resonance imaging (MRI) can also demonstrate morphological changes in the median nerve, while MR tractography has demonstrated fractional anisotropy and diffusion coefficient measures in the median nerve that correlate with NCS findings²¹. However, the utility of MRI is limited by cost, and it must be re-stated that these imaging modalities are merely diagnostic adjuncts, and that the gold standard for diagnosing CTS is an accurate clinical history and examination by an experienced clinician, which is itself far from infallible.

1.2.4 Treatment

The treatment strategies for CTS can be broadly divided into non-surgical and surgical. The most commonly used non-surgical first-line treatment for CTS is wrist splinting. The rationale underlying this intervention is to provide mechanical respite for the median nerve. One randomised control trial of splinting vs surgery demonstrated successful symptom relief at three months in 54% in the splinting group compared to 80% in the surgery group²². Local corticosteroid injection into the carpal tunnel is another frequently used non-surgical treatment, which is believed to exert its effect by reducing the oedema within the carpal tunnel⁷. While the initial response rate is high (approximately 70%)¹⁶, so is the relapse rate. A randomised trial of steroid vs placebo injection into the carpal tunnel found that steroid was more effective than placebo in reducing symptom severity; however, three out of four patients proceeded to have surgery within one year²³. Thus, long-term benefits of steroid injections in the treatment of CTS have yet to be demonstrated, and the AAOS Clinical Practice Guidelines for the treatment of CTS has recommend surgery for failed nonsurgical treatment at any stage, or if the patients prefer surgery⁵.

Carpal tunnel decompression (CTD) surgery has therefore remained the mainstay of definitive treatment for CTS. The procedure involves the division of the transverse carpal ligament to surgically release the 'roof' of the carpal tunnel (thus creating more space for the median nerve and the flexor tendons), and this can be performed either as an open operation or endoscopically (**Fig. 1.4**). CTD achieves a permanent and complete cure in the majority of cases¹⁶, with high patient satisfaction and improvement in both symptoms and functional status²⁴. Furthermore, a systematic review of seven trials has found that CTD, when compared to non-surgical treatment, is more than twice as likely to result in symptom resolution and normalisation of nerve conduction findings²⁵.

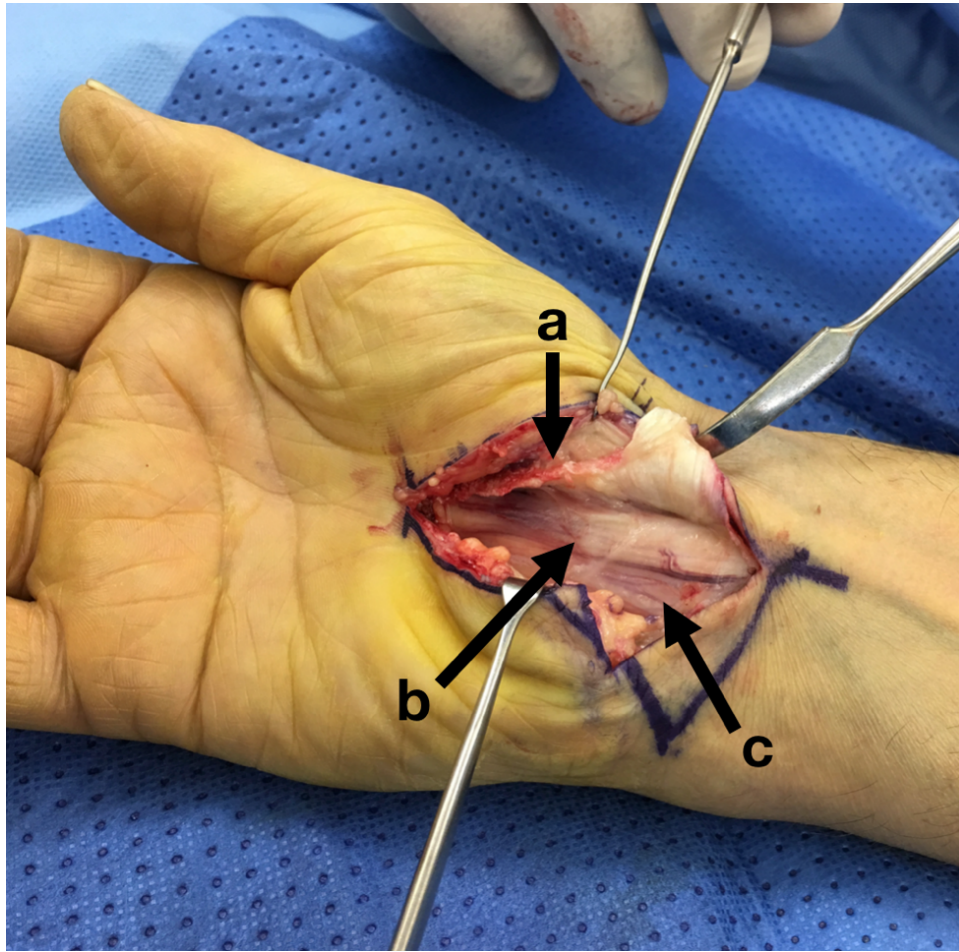


Fig. 1.4 Clinical photograph of carpal tunnel surgery. Intra-operative view of an extended carpal tunnel decompression. (a) The edge of the divided transverse carpal ligament. (b) The median nerve entering the carpal tunnel. The arrowhead indicates the proximal site of compression, where the nerve appears darker. (c) A flexor tendon adjacent to the median nerve, surrounded by oedematous and hyperaemic tenosynovium. Photograph provided by Professor Dominic Furniss.

However, surgery is not without its risks. Surgical complications of CTD have been reported to occur in 1-25% of cases²⁶. In addition to the potential complications of any surgical procedure (such as infection, bleeding and wound dehiscence), there are procedure-specific and potentially severe complications to CTD, including structural damage to the median or ulnar nerves, tendons and arteries. A literature review of 28,000 open and endoscopic CTD procedures found the overall incidence of such major complications to be 0.49% for open CTD²⁷. Other procedure-specific complications include weakened grip strength, scar pain when

pressure is applied to the heel of the hand (known as “pillar pain”), and complex regional pain syndrome⁷. A literature review of the long-term outcomes of carpal tunnel surgery reported a success rate of between 75% to 90%²⁸, while a survey of 4,000 patients who underwent CTD in the UK found that 75% considered the surgery an “unqualified success”, and 8% reported they were worse after surgery^{16,29}.

The phenomenon of recurrent CTS is widely reported in the literature, with revision CTD required in up to 12% of patients³⁰. Mackinnon has classified the pathology in these cases into: persistent CTS, recurrent CTS, or new symptoms; the first two categories are usually a result of incomplete release of the transverse carpal ligament at the initial operation, which may or may not be accompanied by perineural scarring leading to compression or tethering of the median nerve^{31,32}. Patients presenting with unsatisfactory outcomes following primary CTD are both diagnostically and surgically challenging, and unsatisfactory results are common³³; moreover, there is a relative paucity of literature on predicting patients who are more likely to need revision surgery.

Thus, while there is a considerable body of evidence to suggest that carpal tunnel decompression surgery offers the patient the best chance of a cure, it is by no means a panacea; there is unequivocally an unmet clinical need for being able to predict and to adequately treat those individuals who fare badly from carpal tunnel surgery.

1.2.5 Socioeconomic burden of CTS

Carpal tunnel syndrome exacts a considerable burden not only on individual sufferers, but also on society as a whole. The socioeconomic costs of CTS are twofold: the cost of treatment and the cost of lost productivity and wages. In the USA, approximately 500,000 carpal tunnel

operations are performed every year, with estimated treatment costs in excess of \$2 billion per annum – a sum unadjusted for inflation since 1995³⁴. Given that the incidence and prevalence have increased over the past 20 years³⁵, that figure is almost certainly considerably greater today. A 2007 study of cumulative earnings losses of a cohort of injured workers found that CTS claimants endured periods of time loss three times longer than claimants with upper extremity fractures, with earning losses in excess of \$45,000 per claimant over a six-year period³⁶. CTS is the most commonly performed elective hand operation, and its operative burden is projected to increase³⁷; thus, its economic ramifications for society will only continue to grow.

1.3 Pathophysiology of CTS

1.3.1 History

The first description of CTS is usually ascribed to James Paget, who in 1854 described two cases of median nerve compression within the carpal tunnel, one post-traumatic and one idiopathic³⁸. The first case series of CTS was published some decades later by Putnam in 1880, which included 37 mostly female patients. Strikingly, the description of symptoms in this series agrees almost perfectly with contemporary descriptions of typical CTS symptoms³⁹:

“..a disturbance of the subjective sensibility of the skin, giving rise to what is known popularly as numbness, recurring periodically, coming on especially at night or very early in the morning... in some cases simply letting the arm hang out of the bed or shaking it about for some moments would drive the numbness away”.

Despite these astute clinical observations, in terms of pathophysiology, Putnam postulated that the symptoms were secondary to alteration in vascular supply to the terminal filaments of sensory nerves to the affected digits, and coined the term “acroparaesthesia”, which became a popular diagnosis in the early 20th century. Other hypotheses that were prevalent at the time included compression of the lower trunk of the brachial plexus to explain the sensory symptoms, and lesions of the motor branch of the median nerve as it passed under the transverse carpal ligament (“median thenar neuritis”), to explain the motor symptoms of thenar atrophy. James Ramsay Hunt, who posited the latter hypothesis, entirely discounted his patients’ sensory symptoms⁴⁰, and a pathophysiological model to explain both the motor and sensory symptoms remained elusive. Marie and Foix⁴¹ are believed to be the first to put forward such a unifying model to explain both the sensory and motor dysfunction, by suggesting that

compression of the median nerve within the carpal tunnel was the culprit. Their paper gave a detailed pathoanatomic description of the compressed median nerve, and they also presciently suggested that “...*transection of the [transverse carpal] ligament could stop the development of these phenomena*”.

However, this theory went largely ignored for several decades, and it was not until the mid-20th century that the pathophysiology of CTS as a compression neuropathy of the median nerve within the carpal tunnel came to be widely accepted. This was in large part through a series of landmark papers by Brain and Phalen (after whom the diagnostic test for CTS is named), and the first ever description of an operation for idiopathic CTS by Cannon and Love in 1946⁴².

Despite the fact that over 150 years have elapsed since Paget’s first description of CTS, the pathophysiological basis of this disease remains poorly understood. While it is firmly established that CTS is a consequence of median nerve injury secondary to increased pressure within the carpal tunnel, there is still considerable uncertainty regarding the cause of this raised pressure, and how it results in nerve injury.

1.3.2 Increased carpal tunnel pressures

Gelberman et al. demonstrated that normal carpal tunnel pressures are 2-31 mmHg, and that these pressures are significantly higher in CTS patients (32-110 mmHg)⁴³. Importantly, these elevated carpal tunnel pressures rapidly return to normal after carpal tunnel decompression surgery. Furthermore, alterations in the posture of the wrist, and repetitive wrist and finger movements have also been shown to increase pressure within the carpal tunnel^{44,45}. It may seem counterintuitive that the carpal tunnel, which is essentially an elliptic cylinder with an opening both proximally and distally, could maintain such elevated pressures. It is the physical

properties of the contents of the carpal tunnel – i.e. the flexor tendons, median nerve, and connective tissues – that preclude the dissipation of elevated pressures down a pressure gradient (which one might expect if the contents of the tunnel were limited to, say, fat or water). Thus, in mechanical terms, the explanation for elevated pressures within the tunnel can be reduced down to two possibilities: either the carpal tunnel is too small, or its contents are too voluminous (or a combination of both), and I will discuss these in greater detail later in this chapter.

1.3.3 Nerve Injury

It has been shown that increased pressure within the carpal tunnel leads to localised ischaemia secondary to impairment of the neural microcirculation⁴⁶. In support of this hypothesis, recovery of localised pressure-induced median nerve dysfunction after removal of the pressure is abolished by the application of an upper-arm tourniquet in human volunteers⁴⁷.

In addition to impairing neural microcirculation, experimentally induced neural compression has been shown to increase vascular permeability of the epineural and endoneural microvasculature, leading to changes to the blood-nerve barrier and intraneural oedema formation⁴⁸. This oedema is postulated to increase the pressure exerted on the nerve (thus creating a ‘compartment syndrome’ of sorts within the endoneurial space, exacerbated by lack of lymphatics), further inhibiting the microcirculation to the nerve fascicles and creating a vicious cycle. If the oedema persists, it may lead to intra- and extraneural fibrotic changes⁴⁹, which is believed to impair the normal gliding of the median nerve relative to its surrounding structures. Such gliding is vital to maintain normal neural anatomy in the face of joint motion – the median nerve has been shown to move up to 9.6 mm during wrist flexion⁵⁰. Tethering of the median nerve within the carpal tunnel means that such movements increase strain within

the nerve⁵¹, which in turn leads to impaired microcirculation⁵², axonal transport⁵³, nerve conduction⁵⁴, and eventually Wallerian degeneration⁵⁵. Consistent with these experimental models, post-mortem examinations of the median nerves of CTS patients has demonstrated reduction in nerve fibre diameter and loss of the myelin sheath⁵⁶.

The effects of CTS on target organ innervation has also been studied. A combined small- and large-fibre sensory dysfunction on quantitative sensory testing, as well as histological evidence of small fibre loss and lengthened nodes of Ranvier in myelinated fibres in median-innervated skin of the index finger, has been shown⁵⁷. This study identified the loss of small unmyelinated fibres as a hitherto unrecognised component of CTS.

1.3.4 The sub-synovial connective tissues (SSCT)

The causes of increased pressure within the carpal tunnel and tethering of the median nerve are not clear from the literature. Efforts have focused on demonstrating changes in the sub-synovial connective tissue (SSCT) that surrounds the flexor tendons and median nerve in the carpal tunnel. The SSCT consists of multiple layers of collagenous fibres (consisting largely of collagen types I, III, and VI bundles) with elastin scattered in between⁵⁸, within which lies a rich network of blood and lymphatic vessels⁵⁹. The SSCT has been suggested to act as a “gliding mechanism” of the flexor tendons and median nerve in the carpal tunnel⁶⁰. Increased thickness of the SSCT has been demonstrated in CTS patients compared to controls⁶¹. Histological examination of this SSCT has failed to demonstrate inflammation as a major cause of carpal tunnel syndrome^{62–64}: Fuchs et al.⁶² demonstrated evidence of synovial inflammation in only 10% of 177 wrists with CTS, with the principal abnormality reported as being oedema and vascular sclerosis. Similarly, a study of 625 cases of CTS found evidence of acute or chronic inflammation in only 24 cases (4%)⁶³, while another study found evidence of inflammation of

the flexor tenosynovium in 17 wrists out of 166 (10%)⁶⁴; however, this study noted the high incidence of normal tenosynovial histology (in >50% of wrists), in stark contrast to previous reports.

While inflammation does not appear to be a key feature in CTS, a *non-inflammatory* fibrosis of the SSCT has been observed in several studies, consisting of increased collagen deposition in disorganised bundles⁶⁵. Furthermore, this collagen has been demonstrated to be type 3 collagen, which is immature collagen associated with the response to injury, and correlates with increased expression of transforming growth factor beta (TGF- β) – a pro-fibrotic cytokine^{66,67}. Increased expression of endothelial nitric oxide synthetase, increased fibroblast density, and vascular proliferation with vascular hypertrophy and intimal thickening have also been observed in the SSCT, again, consistent with an ischaemic insult^{67,68}. Further supporting the notion of aberrations in the microvasculature playing a role in CTS pathophysiology, Hirata et al. demonstrated increased expression of vascular endothelial growth factor (VEGF) and prostaglandin E2 in synovial biopsies from CTS patients, and found that the areas within the SSCT with increased VEGF expression closely matched the areas with marked alterations in the extracellular matrix, with proliferation of endothelial cells, vascular smooth muscle cells, and synovial lining cells, suggesting that VEGF induces the histological changes seen in the SSCT⁶⁹. The authors suggested that angiogenesis appears to take place as part of a regenerative reaction that results in fibrosis. A subsequent paper by the same group demonstrated increased expression of matrix metalloproteinase-2 (MMP2) in small arterioles in the SSCT, which correlated positively with the severity of reported pain in CTS patients⁷⁰. The observed histological changes to the microvasculature of the SSCT were supported by Jinrok et al., who additionally noted decreased elastin in the blood vessels walls in CTS patients compared to cadaveric controls⁶⁰.

It must be noted that there is a significant challenge inherent in performing molecular studies of CTS pathogenesis, insofar as CTS is a dynamic process in which the pathological findings evolve as the disease progresses in severity²⁹. This may in part explain the disparate molecular pathways that have been investigated, and the frequently contradictory findings in the literature. It is perhaps also why there is such difficulty in establishing a temporal relationship of causation between the changes seen in the SSCT and changes seen within the median nerve. **Fig. 1.5** is an attempt to summarise the current evidence for the molecular and histological changes that are implicated in CTS pathophysiology.

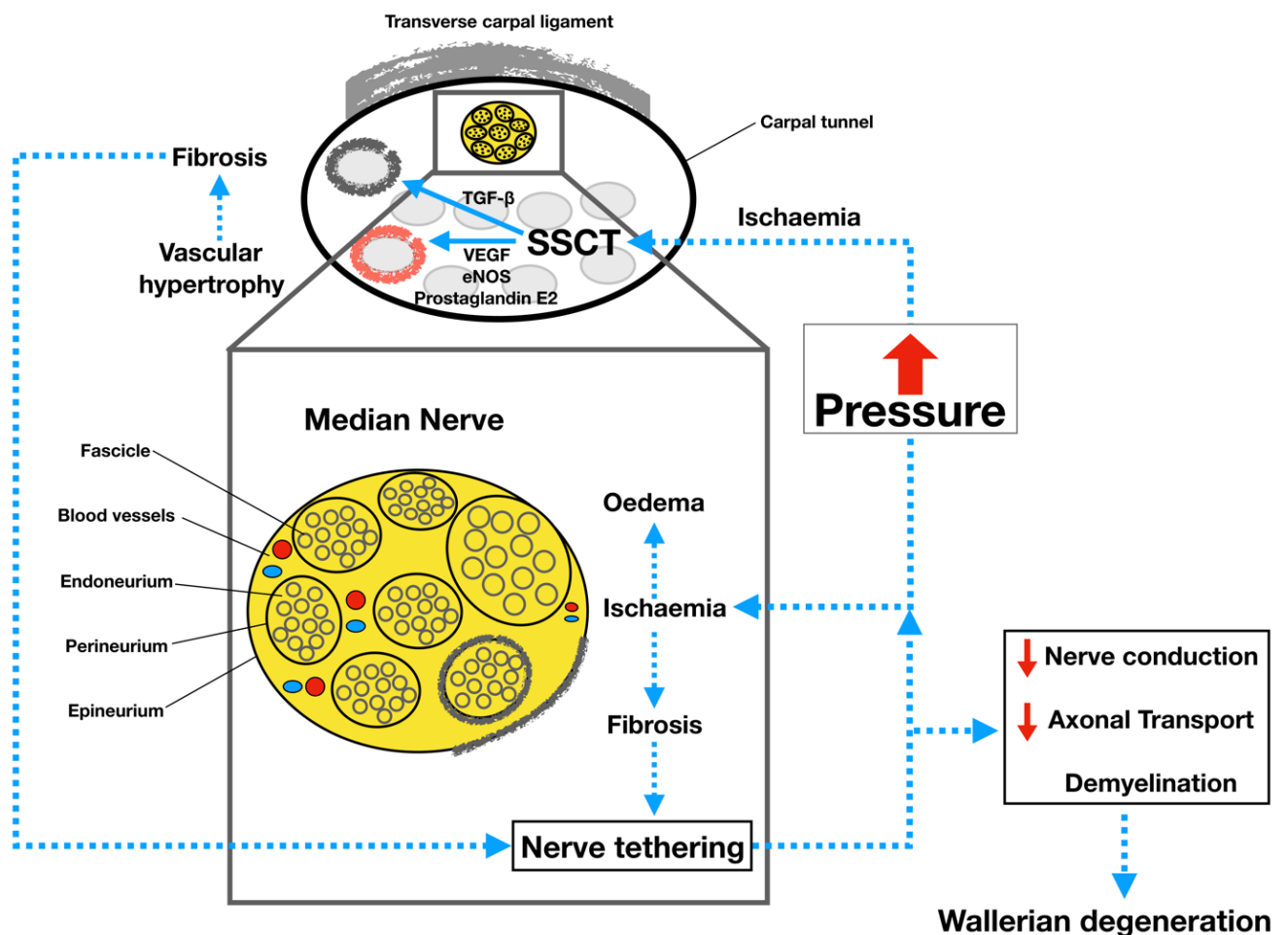


Fig. 1.5 Summary of CTS pathophysiology. Schematic summarising the current understanding of the sequence of events that lead from raised pressure within the carpal tunnel to ultimate nerve degeneration.

1.4 Aetiology

Much of the literature pertaining to CTS makes a distinction between “idiopathic” and “secondary” CTS^{7,71}. This is motivated by an attempt to distinguish CTS of an unknown cause (“idiopathic”, which accounts for the majority of cases), from CTS where there is believed to be definitive cause, such as rare anatomical variants, tumours, or infections within the carpal tunnel that result in excess pressure being exerted on the median nerve.

However, classified within this category of secondary CTS, one frequently encounters diseases that are commonly found in association with CTS – such as diabetes, hypothyroidism and rheumatoid arthritis – based on speculative mechanisms as to how they might increase the volume of contents within the carpal tunnel⁷¹. To further add to the confusion, other categories and sub-categories are added to the classification, such as “acute” CTS (where there is a single inciting event, such as a fracture of the distal radius or carpal dislocation) , “dynamic” CTS (related to overuse of the hand and wrist – frequently used interchangeably with “occupational” CTS), and “neuropathic” CTS (under which diabetes has been classified⁷²).

This classification of CTS into two or more separate disease entities based on putative aetiology and mechanistic conjecture is therefore problematic, not least because more than one of these factors will frequently coexist (diabetes and hypothyroidism, for instance, are both very common diseases). For this reason, classification of CTS into idiopathic and secondary will deliberately be avoided in this thesis. What follows, therefore, is a description of the diseases and other variables for which there is robust evidence that they are associated with an increased risk of developing CTS.

1.4.1 Anthropometric factors

Several studies have suggested that wrist dimensions contribute to CTS aetiology, with the observation that having a higher wrist index (wrist width/wrist depth, as measured on cross-section), increases the risk of developing CTS. There is some evidence that severity of disease correlates with a higher index⁷³. The prevailing hypothesis is that having a higher index (and thus a 'squarer' wrist shape) creates a carpal tunnel shape that predisposes the median nerve to greater compression. Johnson et al. reported a correlation between nerve conduction velocities and wrist dimensions, and suggested a ratio of 0.7 as a cut-off above which the hand is at risk of CTS⁷⁴, and subsequent authors have even recommended measurement of the wrist index as part of the physical examination for CTS⁷⁵. However, a study of over 1,000 subjects concluded that anthropometric measurements have limited value as diagnostic tools for CTS⁷⁶. On the subject of anthropometric measurements, height has also been evaluated as a potential risk factor for CTS, with one prospective study and one case-control study reporting a decreased risk for CTS in taller individuals^{77,78}.

1.4.2 Body mass index (BMI)

There is broad consensus that along with female sex and increased age, obesity (as measured by raised BMI) is a risk factor for CTS. A meta-analysis of 58 such studies reported that obese patients were more than twice as likely to develop CTS, with each one-unit increase in BMI increasing the risk of CTS by 7.4%⁷⁹. However, there is no mechanistic evidence for how raised BMI increases risk of CTS⁷, and it is not immediately clear if the effect might be mediated through co-morbidities of the obese (such as diabetes or metabolic syndrome). Kurt et al. attempted to establish causality between obesity and CTS – obese CTS patients were enrolled in weight-loss programmes in an attempt to reverse disease progression, but the study failed to

demonstrate any improvements in nerve conduction study parameters after 3 months (and successful weight loss), suggesting that factors other than raised BMI may be contributing to the higher prevalence of CTS in obese patients⁸⁰.

1.4.3 Occupational factors

There is a considerable body of evidence for an environmental influence on the development of CTS, and occupational risk factors for CTS have been widely reported for many decades. Perhaps the most robust of such studies, Kozak et al. performed a meta-analysis of ten published systematic reviews (covering 143 original studies) into the biomechanical risk factors for CTS in workers. The authors concluded that there was high quality evidence for risk factors such as repetitive movement and forceful exertion, moderate quality of evidence for exposure to vibration, low quality of evidence for wrist postures, and no evidence for computer use⁸¹. However, there remains significant controversy in this field¹⁶, especially with regard to whether CTS should be considered an occupational disease - notably in the USA, CTS is widely considered to be a “work-related injury”^{82,83}.

1.4.4 Pregnancy, menopause and female hormones

CTS symptoms are common in late pregnancy, and is transient in the majority of cases^{16,84}. The reported incidence of pregnancy-related CTS varies substantially in the literature, from 0.7% to 70%⁸⁴, and 7 to 43% when diagnosed electrophysiologically⁷. A widely cited aetiological factor is fluid retention in pregnancy⁸⁵, although definitive mechanistic evidence for this is lacking. Changes to the hormonal environment are also likely to play a role – not only in CTS in pregnancy, but perhaps in CTS in females as a whole – and there are several strands of evidence in support of such a hypothesis.

The evidence for CTS having a sex-specific incidence two to three times higher in females compared to males was presented earlier in this chapter. Strikingly, the number of diagnoses of CTS in females peaks between ages 50-59, and then gradually declines, whereas the number increases with age in men^{12,14}. It has been hypothesised, therefore, that hormonal changes around the time of menopause, such as hypooestrogenaemia, might be relevant to this observed trend³⁵. Another study compared the ages at menopause of postmenopausal women with CTS against the ages at menopause of females without CTS, and found that CTS cases had experienced menopause at significantly earlier ages than the control group⁸⁶. Adding further weight to this hormonal hypothesis of CTS aetiology are the observations that: (1) oral contraceptive use for ten years or more doubles CTS incidence⁸⁷, (2) females who have undergone bilateral oophorectomy have higher rates of CTS than controls (32% vs 10%)⁸⁸, (3) postmenopausal females given hormone replacement therapy were significantly less likely to develop CTS than females given placebo (HR 0.78, CI, 0.65-0.94)⁸⁹, and (4) the use of anastrozole (an aromatase inhibitor) for breast cancer was associated with a higher incidence of CTS (OR = 3.06, CI, 1.21-7.72)⁹⁰.

1.4.5 Associated diseases

An array of systemic diseases have been reported to show an association with increased CTS prevalence with varying degrees of evidence. Below, I will briefly discuss the associations between CTS and the most well-established of those co-morbid diseases, namely diabetes, rheumatoid arthritis, and hypothyroidism. Diseases not discussed in this section, but for which there is moderately robust evidence for an association with CTS, include: gout/pseudogout, amyloidosis, acromegaly, osteoarthritis, and vitamin B6 deficiency^{7,91,92}.

1.4.5.1 Diabetes

Strong evidence exists for the co-existence of both type 1 and type 2 diabetes with CTS. A large meta-analysis by Pourmemari et al.⁹³ found the pooled odds ratio for the association between CTS and diabetes was 1.97 (95% CI: 1.56–2.49). Four of the studies in the meta-analysis had adjusted their estimates for BMI, and the pooled odds ratios in these four studies was 1.68 (95% CI: 1.12–2.53), suggesting that diabetes is a risk factor independent of BMI. The lifetime risk of developing CTS in type 1 diabetes has been reported as 85% (95% CI: 72–97%)⁹⁴.

Although the mechanisms by which diabetes confers an increased risk of CTS remain conjectural, the subject of how diabetes affects peripheral nerves has been widely studied. Diabetics demonstrate a reduction in myelinated nerve fibres and endoneurial capillary densities, and lower axonal densities⁹⁵. There also appears to be an indirect metabolic effect on nerves: advanced glycation end-products (i.e. glycated proteins and lipids) can increase the levels of inflammatory cytokines in circulation (including VEGF)^{96,97}, which may contribute to aberrations in the microvascular circulation within the median nerve⁹⁸.

1.4.5.2 Rheumatoid arthritis

Rheumatoid arthritis (RA) is considered a relatively strong risk factor for CTS. Multivariate analysis in a large matched case-control study cohort (with >3,000 CTS cases) found that patients with RA are 2.23 times more likely to develop CTS⁹⁹, and Karadag et al. reported evidence of CTS in 17% of females with RA, compared to 4% in age-matched controls¹⁰⁰. It is widely posited in the literature that RA increases CTS risk through its propensity to cause tenosynovitis around the flexor tendons in the carpal tunnel, resulting in an elevation in carpal tunnel pressure. Again, no mechanistic evidence for this has been established, and it must also

be borne in mind that the multi-system nature of RA makes several mechanisms of peripheral neuropathy plausible.

1.4.5.3 Hypothyroidism

There is reasonably strong evidence for a link between hypothyroidism and CTS, with two studies reporting electrophysiologically diagnosed CTS in one-third of hypothyroid patients^{101,102}. One suggested mechanism for this association is the accumulation of mucopolysaccharides, hyaluronic acid glycosaminoglycans in the connective tissue surrounding the median nerve^{103,104}, whereas fluid accumulation and synovial oedema within the carpal tunnel has also been suggested as a cause¹⁰⁵. Interestingly, one study found electrodiagnostic evidence to suggest that hormone replacement therapy for three months could reverse the onset of CTS in patients with hypothyroidism¹⁰⁵. Although this study had several limitations, it raises an intriguing possibility in the context of the evidence for a role played by female hormones/hypooestrogenaemia in CTS pathogenesis.

1.4.6 Pain Comorbidities

Pain is a central feature in a sub-group of patients with CTS, and there is evidence to suggest a role played by central sensitisation[§] in CTS pain^{106,107}. In an experiment involving experimentally induced pain in patients with mild to moderate CTS at a site remote from the hand, central pain mechanisms were found to account for the observed pain hypersensitivity via increased pain facilitation and reduced endogenous pain inhibition¹⁰⁶. This mechanism is

[§] Central sensitisation is defined as an increase in pain response to nociceptive stimuli mediated by the amplification of signalling in the central nervous system.

believed to account for the finding that CTS patients frequently experience symptoms in an anatomical distribution beyond the classical median-innervated territory of the hand¹⁰⁸.

As with other pain conditions, there is a considerable body of literature on the psychological comorbidities of CTS, with depression and anxiety rates reported to be significantly higher in CTS patients compared to the general population^{109,110}. Nunez et al. found in a cohort of 54 CTS patients that sex, age and electrophysiological CTS severity did not correlate with pain intensity, whereas illness behaviour (depression and misinterpretation of nociception) did, accounting for 39% of the variation in CTS pain intensity¹¹¹. Shin et al. reported that depression levels and pain anxiety were significantly associated with CTS symptoms both pre- and post-operatively, and both parameters improved three months following carpal tunnel decompression surgery¹¹².

1.5 The genetic aetiology of CTS

Lozano-Calderon et al.¹¹³ synthesized 117 primary articles describing various risk factors for CTS, and assessed the strength of evidence linking each of these to CTS using the Bradford Hill criteria, which provides a method for quantitatively scoring the strength of evidence for causation (rather than simply association). By comparing the studies looking at occupational risk factors with those looking at biological risk factors (such as genetics, age and sex), the occupational risk factors had a score of 5.2, compared to the biological factors group which had a score of 12.2. When the studies in the biological factors group were subdivided into discrete categories, “genetics” was given the highest score of 14.2. This contrasted starkly with “repetitive hand use” (the highest scoring category in the occupational factors group), which had a score of 6.5.

1.5.1 Population-based studies

A genetic predisposition to CTS is well recognised. Familial clustering is consistently reported in studies of CTS patients, and a positive family history was reported in 27-39% of cases in a series of 711 patients¹¹⁴. The standardised incidence ratio (SIR) of CTS in siblings in Sweden has been calculated as 4.08, compared to an SIR among spouses of 2.06, suggesting a genetic influence on aetiology. However, a formal sibling recurrence risk for CTS has not been reported¹¹⁵. A large twin study from the UK reported a heritability of 0.46 (95% CI: 0.34–0.58), a figure that was essentially unchanged when modelling took account of known environmental risk factors, suggesting that genetic factors are the strongest risk factors for CTS¹¹⁶. Heritability (h^2) is defined as the proportion of the variance in a phenotypic trait in a population that is attributable to genetic variance¹¹⁷, and the heritability figure of 0.46 for CTS is greater than

that for diseases with well-established genetic contributions to their aetiology, such as breast cancer (0.31)¹¹⁸, Parkinson's disease (0.4)¹¹⁹, and ischaemic stroke (0.38)¹²⁰.

1.5.2 Familial CTS

Several large pedigrees segregating carpal tunnel syndrome have been reported, but genetic linkage has not been demonstrated¹²¹⁻¹²⁷. In some of these cases, a structural cause was found, such as thickening of the transverse carpal ligament¹²³. Alford et al. found that bilateral disease was commoner in familial cases, and speculated that the genetic predisposition may manifest in the form of a smaller carpal tunnel¹²⁷.

Molecular genetic investigations of CTS are scarce in the literature. There are rare variants in myelin genes that predispose to entrapment neuropathies including CTS as a consequence of myelin instability. The best example of this is hereditary neuropathy with liability to pressure palsy (HNPP) due to deletion of the *PMP22* gene¹²⁸. In addition, Lupski et al. described a family with Charcot-Marie Tooth Disease type 4C, where affected individuals were compound heterozygotes for mutations in the gene *SH3TC2*. Family members who were heterozygous carriers of a single mutation (R954X) in the gene had electrophysiological evidence of mild CTS¹²⁹. Deposition of amyloid in the transverse carpal ligament has been demonstrated in 8% of idiopathic CTS⁶⁴, and CTS is a common presenting feature of familial amyloid polyneuropathy¹³⁰; however, mutations in the *TTR* gene were not demonstrated in a cohort of 282 patients with idiopathic CTS¹³¹.

1.5.3 Candidate gene studies

A small number of candidate gene studies screening for association of CTS with single variants have been undertaken. The focus of these studies has been around genes that have previously been implicated in musculoskeletal disorders, such as Achilles tendinopathy and lumbar disc disease. Equivocal associations with CTS have been reported for *IL-6R*¹³², *COL5A1*¹³³, *COL11A1*¹³⁴ and *BGN*¹³⁵, but these studies are limited to small cohorts and the methodological quality is generally poor. Notwithstanding the limitations, the association of CTS with two collagen gene variants is of interest. Types V and XI collagen, implicated in the studies, regulate the assembly and diameter of collagen fibrils, so have a plausible role in altering the mechanical properties of the tendons or SSCT within the carpal tunnel.

1.6 CTS as a complex disease

Taken together, these data strongly suggest that both genetic and environmental factors are important in the aetiology of CTS. Thus, we can consider CTS to be a complex disease, defined as a disease in which multiple genetic and non-genetic factors interact to affect overall phenotypic expression. Unlike traditional Mendelian modes of inheritance, the manifestation of complex diseases are governed by the liability threshold model¹³⁶, whereby the severity (or indeed, expression) of the disease is determined by the genetic dosage of the risk factor that an individual carries. This threshold model is governed both by gene-gene interactions and gene-environment interactions.

This concept is perhaps best illustrated by comparing a classically Mendelian disease such as cystic fibrosis with a typical complex disease such as type 2 diabetes. In cystic fibrosis, a deletion of three nucleotides in the *CFTR* gene alters the structure of the chloride channel to the point that it is effectively rendered non-functional, with devastating consequences for homozygotes of the faulty gene. In contrast, all individuals will be born with varying genetic propensities for developing type 2 diabetes, in the form of genetic variants at over 140 genomic loci that have hitherto been discovered¹³⁷ (and undoubtedly many more that have yet to be discovered). If we imagine two individuals who have identical environmental exposures to diabetes (diet, exercise, and other variables that influence disease risk), one of them may go on to develop type 2 diabetes while the other will not, and this phenotypic variability between the two individuals will have been determined by the differences in the cumulative genetic dosage of the risk variants carried.

In complex diseases, these genetic variants typically take the form of single nucleotide polymorphisms (SNPs), which are substitutions of a single nucleotides that occur at specific

positions in the genome, where each variation is present to some appreciable degree within the population. Over 84 million SNPs have been identified across multiple human population groups, and a typical genome differs from the reference genome at 4-5 million SNPs¹³⁸. In addition to SNPs, other types of genetic variants include short insertion-deletion polymorphisms, and structural variants such as copy number variants (CNVs), insertions, deletions, duplications, inversions and translocations. However, SNPs and indels account for over 99.9% of human genetic variation.

The majority of these variants^{**} do not reside in protein-coding regions of the genomes, but rather in non-coding regions. Variants that affect a particular phenotype or disease risk will tend to reside in a genomic region with some regulatory properties, such as promoter, enhancer or insulator regions. Alternatively, they may influence mRNA stability if they reside within a microRNA binding site. These variants' effects on gene expression (or protein translation) are usually very subtle, which is the basis of the common disease/common variant (CDCV) hypothesis. This hypothesis posits that common diseases are a manifestation of multiple common^{††} variants across the genome that cumulatively contribute to the risk of phenotypic manifestation of a disease¹³⁹. It is for this reason why complex diseases are frequently referred to as “polygenic”.

^{**} Due to the fact that SNPs account for the vast majority of human genetic variation, the term “variants” will henceforth be used to refer specifically to SNPs (unless otherwise stated), and the two terms will be used interchangeably throughout this thesis.

^{††} By convention, common variants are defined as those that have a minor allele frequency of >5% in the population.

1.7 Genome-wide association studies (GWAS)

The now well-established method for interrogating the polygenic basis of complex diseases is the genome-wide association study (GWAS), and the last decade has seen a proliferation of these studies in the medical literature. This has been due to technological and bioinformatic advances that has led to the ability to efficiently assay genetic variation in a large number of individuals using high-throughput SNP arrays¹⁴⁰. GWAS lies at the very core of this thesis, so a brief overview of its principles is warranted.

Although analytic techniques employed in GWAS have evolved in sophistication, the fundamental concept of a GWAS remains relatively simple. SNP arrays typically genotype between 500,000 to 1,000,000 SNPs, thus capturing up to 1% of human genetic variation. Due to the tendency of SNPs that are close to one another to be inherited together, there are extensive local correlations between nearby SNPs – a phenomenon known as linkage disequilibrium (LD). Knowledge of LD structures in different ancestral populations (from large collaborations such as HapMap and the 1000 Genomes Project) can be leveraged to allow SNPs that are not captured by the arrays to be inferred in a process known as imputation, meaning that the majority of common variants in an individual can be deduced at the cost of only genotyping a modest fraction.

Most GWAS for disease phenotypes take the form of case-control studies, whereby hundreds of thousands of SNPs are genotyped in a large number of cases and controls. The frequency of the alleles at each of these SNPs is compared between the cases and controls, and statistical tests performed to determine the significance of any observed differences in allele frequency. Due to the fact that up to one million such tests are performed, a Bonferroni correction is applied to reduce the type 1 error rate. A conservative p-value of $< 5 \times 10^{-8}$ (i.e. 0.05/1 million) has

come to become accepted as the standard threshold for claiming that a SNP is genome-wide significant¹⁴⁰. **Fig. 1.6** is a schematic that illustrates the basic principles of GWAS.

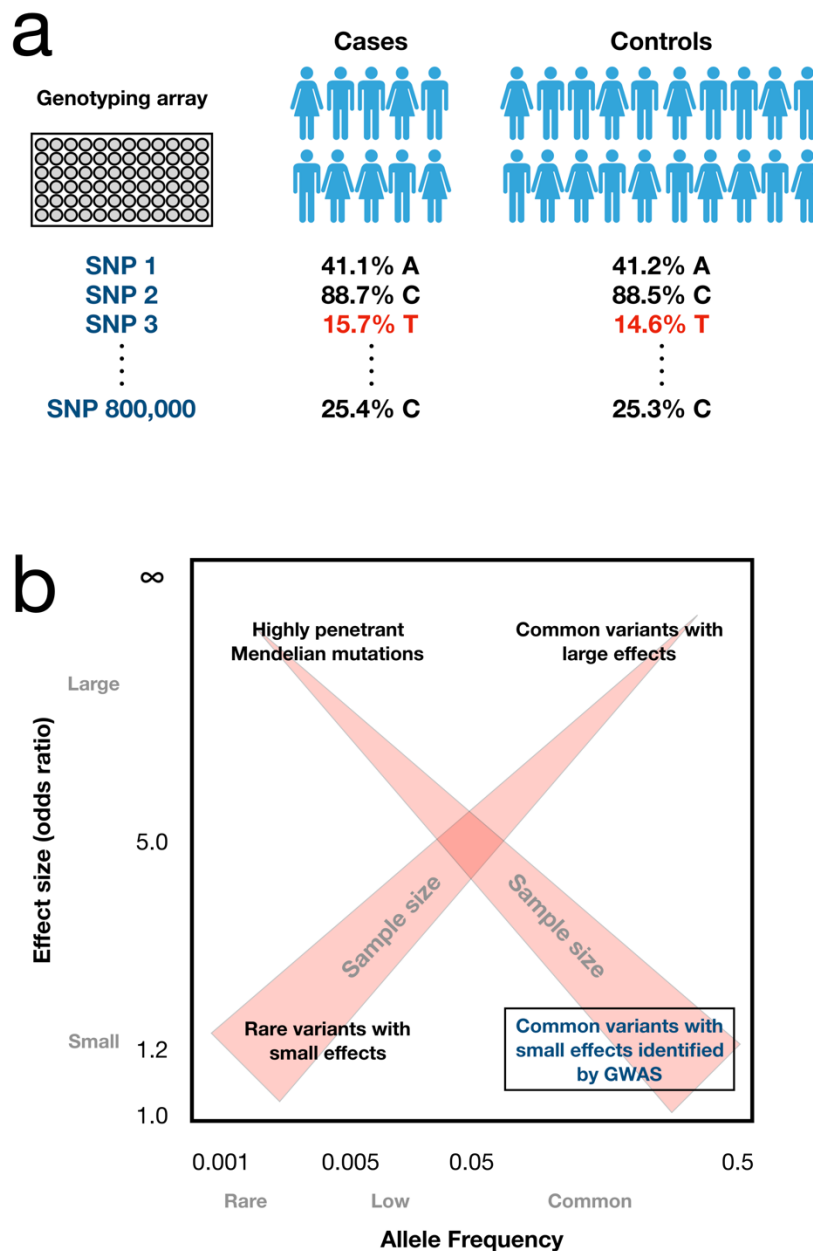


Fig 1.6 Illustration of GWAS. (a) Basic principles of GWAS, showing comparison of allele frequencies between cases and controls at 800,000 SNPs. The SNP highlighted in red shows a significant disparity in allele frequency. (b) GWAS typically identifies common variants with small effect size (bottom right corner). Adapted from McCarthy et al.¹³⁶

Each GWAS association (or “hit”) offers potential clues into disease mechanisms, as there is a high likelihood that the causal variant is modulating the expression of a gene at that genomic locus. One major caveat that must be mentioned is that there are typically tens, if not hundreds, of SNPs that are genome-wide significant at any associated GWAS locus, and due to LD between these SNPs, it is very challenging (if not impossible) to resolve which of those SNPs is the true causal variant responsible for increasing disease risk. This is further complicated by the fact that there are usually numerous genes in the vicinity of these GWAS hits, and there may be more than one causal SNP at a locus. Furthermore, proximity of a SNP to a gene does not necessarily imply causality; due to the conformation of DNA in three-dimensional space, a SNP that is a regulatory variant can feasibly modulate the expression of a gene (or several genes) up to 1 Mb away. A notable example of this phenomenon is the *FTO* locus, whereby intronic SNPs within the *FTO* gene were believed to regulate the expression of this gene to influence body mass index and obesity; it was subsequently found that these variants regulate the expression of *IRX3*, approximately 500 kb away¹⁴¹. Thus, associated SNPs in a GWAS are perhaps best conceptualised as location markers that have the potential to provide valuable clues to the underlying biology.

For these reasons, the elucidation of a mechanistic link between a causal variant and its functional consequences remains highly challenging, and this has been a frequent criticism of GWAS¹³⁶. Such criticisms are to some degree justifiable, as the discovery of molecular mechanisms through which these variants operate has lagged significantly behind their discovery. However, most geneticists are in agreement that GWAS provides a powerful hypothesis-generating strategy for linking genotype to phenotype, yielding valuable biological and therapeutic insights. Correspondingly, there are several notable GWAS ‘success stories’, such as the therapeutic targeting of several interleukin pathways discovered through GWAS, which have now become mainstay treatments for psoriasis and ankylosing spondylitis¹⁴².

Better understanding of disease biology and the development of novel therapies are not the sole end-goals of GWAS. One notable application of GWAS with significant potential clinical utility is the polygenic risk scores (PRS). PRS can quantify an individual's innate genetic liability for developing a particular trait or disease, and is typically calculated as a weighted sum^{**} of the risk alleles for a that trait carried by an individual. Thus, the ultimate aim of a PRS is to stratify a population by accounting for each individual's genetic susceptibility to a disease, and particularly to identify individuals with a significantly elevated disease risk who would benefit from preventative treatments¹⁴³. Although the applicability of PRS at the individual level is currently limited in providing sufficient risk stratification¹⁴⁴, the utility of these scores will increase as the number of disease-associated variants employed in their construction continues to rise. A recent study is worthy of mention: Khera et al¹⁴⁵ constructed a PRS to identify 8% of the population whose polygenic susceptibility place them at greater than three-fold increased risk of coronary artery disease compared to the general population. In comparison, 0.4% of the population carry a monogenic familial hypercholesterolaemia mutation that confers a comparable risk for coronary artery disease; thus, at a population level, the PRS is able to identify twenty times more individuals at this level of risk, highlighting the practical utility of the polygenic approach in identifying individuals who warrant early and aggressive preventative therapies.

^{**} The weight takes into account the measured effect of a risk allele on a trait, as measured by GWAS.

1.8 Scope of the thesis

This thesis is titled “The Genetic Architecture of Carpal Tunnel Syndrome”, so it would be prudent to define what exactly is meant by this term. Genetic architecture refers to the characteristics of genetic variation that are responsible for heritable phenotypic variability, or put more simply, “the landscape of genetic contributions to a given phenotype”². Specifically, genetic architecture comprises: (1) the number of genetic variants that contribute to a phenotype, (2) their allele frequencies in the population, (3) the magnitude of their effect sizes, and (4) how they interact with each other and with the environment. Together, these facets of genetic architecture contribute to the relationship between genotype and phenotype¹⁴⁶. In the current era, elucidating the genetic architecture of a disease is a fundamental step for shedding light on the biological mechanisms that influence disease predisposition on a genomic level, and informs the clinical goals of human genetics.

To this end, the chapters that follow have these broad aims:

1. Map the genetic architecture of CTS by performing a GWAS (Chapter 2)
2. Prioritise candidate genes and biological pathways that are important in CTS pathogenesis by performing a set of *in silico* analyses (Chapter 3)
3. Establish causality between CTS and putative risk factors using Mendelian randomisation (Chapter 4)
4. Correlate the polygenic risk for CTS with clinical phenotype by genotyping a cohort of CTS patients in Oxford (Chapter 5)
5. Assess whether there is genotype-specific expression of candidate genes by quantifying their expression in tenosynovium from the carpal tunnel in an RNA-Seq experiment (Chapter 5)

6. Synthesize all of the findings in the context of current knowledge of CTS pathophysiology, and summarise how the field has progressed in light of this research (Chapter 6)

Chapter 2

A genome-wide association study of carpal tunnel syndrome

2.1 Introduction

In the previous chapter, I outlined the weight of evidence in favour of CTS being a complex disease, with a significant genetic contribution to its aetiology. The current best method for locating genetic variants that predispose to a complex disease is the genome-wide association study (GWAS), and the principal aim of this chapter is to uncover the polygenic signal underlying CTS, by performing a GWAS in this disease.

From the literature presented in the previous chapter, we can surmise that there is considerable uncertainty with regard to the pathophysiology of CTS; moreover, the least well-studied aspect of CTS aetiology is unequivocally its genetic component. Thus far, no researchers have attempted to investigate the polygenic architecture of CTS, nor indeed, of any entrapment neuropathy. This is somewhat surprising given that CTS is an extremely common and societally costly disease³⁴ that is encountered not only by plastic and orthopaedic hand surgeons, but also by general practitioners and a range of physicians – from neurologists to rheumatologists and endocrinologists. One wonders therefore whether this indifference is due to the widely held belief (even amongst medical professionals) that CTS is largely an occupational, and therefore preventable, disease, or because there is a perception that CTS is a disease for which there is a

relatively simple surgical “cure” for all patients. Needless to say, the available evidence is in profound disagreement with both of these notions.

Mapping the genetic architecture of CTS through GWAS is the central aim of this thesis, but this is merely the first step. Being able to link genetic variants to the CTS phenotype allows the implementation of a multitude of bioinformatic analyses that can provide a wealth of information about its pathophysiology and its relationship to other complex diseases, including genetic epidemiological methods for inferring causation between CTS and its putative risk factors. With the advent of polygenic risk scores, it also opens the door to the realm of personalised medicine, which we will inevitably see a lot more of in the coming decades.

2.1.1 Aims and Hypotheses

The fundamental hypothesis underlying this GWAS is that CTS is indeed a complex disease with an underlying polygenic architecture. Provided that this is the case, the study is sufficiently powered to uncover genome-wide significant hits at a Bonferroni-corrected threshold of significance of $p < 5 \times 10^{-8}$.

One of the great strengths of GWAS is its ‘agnostic’, or hypothesis-generating nature. The first wave of GWA studies provided a novel conceptual framework to search for genetic variants underlying common disorders – the approach is free from any underlying biological assumptions, and instead, scans the entire genome for variants significantly associated with a disease or trait of interest.

Without undermining this agnostic philosophy underling GWAS, it is reasonable to generate broad, *a priori* hypotheses as to how genetic variants could feasibly contribute to increased risk

of developing CTS. The first of these hypotheses is ‘structural’ – namely that genetic variants alter the environment through which the median nerve transits, by altering the bony or soft tissue architecture of the carpal tunnel, or indeed the anatomy of the median nerve itself. The second hypothesis is that the genetic variants alter the physiology of the median nerve to render it more vulnerable to increased pressures, or by conferring vulnerability to a pain phenotype through central mechanisms (**Fig. 2.1**). These two hypotheses are of course not mutually exclusive; in fact, central to the very nature of complex disease is the idea that multiple common variants (affecting the expression of multiple genes) across the entire genome cumulatively contribute to the expression of a particular phenotype. Thus, these two broad hypotheses should be thought of not as dichotomous outcomes, but rather as two different conceptual frameworks under which genetic variants uncovered in this study might reasonably be categorised.

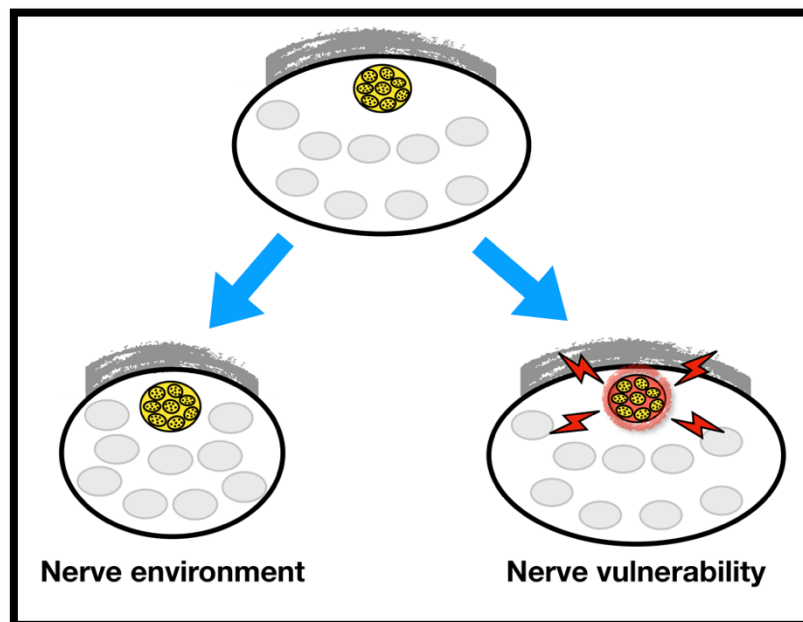


Fig 2.1 Illustration of GWAS hypotheses. Two possible mechanisms by which genetic variants might confer susceptibility to CTS: by altering the nerve environment (e.g. reducing the dimensions of the carpal tunnel) or making the nerve more vulnerable to increased pressures.

Demonstrating replicability of genome-wide associations from one study in a well-powered independent cohort is a central principle of GWAS¹⁴⁷. Replication can improve the credibility of association signals by reducing the probability that they are spurious (i.e. false positive) associations due to systematic biases in a study. To this end, I contacted several national and international cohort studies and biobanks whose participants have undergone whole-genome genotyping. The majority did not collect data on CTS, and those that did had very small numbers (in the order of a few hundred cases). The sole exception was the Estonian Genome Center at the University of Tartu (EGCUT), whose population-based biobank identified ~4,000 CTS cases. A collaboration was therefore established with EGCUT to attempt to replicate any associations that emerged in the UK Biobank GWAS of CTS.

To summarise, the aims of this chapter are to:

1. Map the genetic architecture of CTS by performing a GWAS using participants with CTS in the UK Biobank resource.
2. Attempt to replicate the findings from the UK Biobank GWAS in the EGCUT cohort.

2.2 Methods

2.2.1 The UK Biobank Resource

The UK Biobank prospective cohort was designed to support the investigation of risk factors for the major diseases of middle and old age, with the recruitment of approximately 500,000 participants of aged 40-69, between 2006-10¹⁴⁸. The participants have undergone whole-genome genotyping and have allowed linkage of these data with their medical records, lifestyle questionnaires, and physical, cognitive, and biochemical measures. In effect, the design of the study allows for the assessment of various different exposures to a range of health outcomes. Medical records are in the form of both routinely collected hospital data (such as ICD-10 diagnostic codes and OPCS operation codes), and self-reported diagnostic and surgical codes. The dataset is available on request to approved researchers around the world, making UK Biobank a rich resource for performing both epidemiological and genetic association studies.

While the size and scale of UK Biobank makes it uniquely powerful, certain limitations of this cohort must be discussed. Firstly, all participants are volunteers, meaning that the cohort is on the whole healthier than the population at large, due to sample selection bias (the so-called “healthy volunteer effect”)¹⁴⁹. UK Biobank participants were found to be less likely to be obese and to smoke, and had fewer self-reported health conditions compared to nationally representative data sources¹⁵⁰. Furthermore, secondary diagnostic codes such as ICD-10 and OPCS operative codes derived from Hospital Episode Statistics for England (HES) are only available from 1996 onwards¹⁵¹, and the UK Biobank resource is not currently linked to primary care data, meaning that disease prevalence estimates based on diagnostic codes available in UK Biobank are likely to be lower than in the general population. Finally, 94% of the UK

Biobank cohort self-report their ethnicity as “white”¹⁵², meaning that non-white participants are underrepresented in this cohort compared to the general UK population.

The genotypes for 488,377 UK Biobank participants are available. Genotyping was performed on two very similar genotyping arrays, sharing approximately 95% of marker content: The UK Biobank Lung Exome Variant Evaluation (BiLEVE) array (49,950 participants, 807,411 SNPs), and the UK Biobank Axiom Array (438,427 participants, 825,927 markers)¹⁵². The markers were chosen on the basis of known or putative associations with a range of diseases and for genome-wide coverage based on known LD relationships in Europeans, and include both SNPs and short indels, across a range of minor allele frequencies.

2.2.2 Ethics

Appropriate ethics approval was sought to use the UK Biobank dataset for this study. UK Biobank has approval from the North West Multi-Centre Research Ethics Committee (11/NW/0382), and this study was conducted under the remit of UK Biobank study ID 10948 (The Genetics of Carpal Tunnel Syndrome).

2.2.3 Quality control (QC) of genotypic data

Given that raw genotype data are innately imperfect, it is crucial to carefully curate the data prior to performing a GWAS, both on the sample (individual) level, and on the marker (SNP) level. GWAS will fail to generate reliable results without this quality control (QC) step, as there are multiple potential sources of error along the genotyping and data processing pipeline, such as sample mix-ups and contamination, poorly performing genotyping probes, and poor quality of DNA samples. The approach to QC in this study was to use a combination of QC already

performed by UK Biobank with additional levels of QC to impose greater stringency and obtain an even ‘cleaner’ set of data. The raw UK Biobank genotypic dataset included 488,377 individuals and 784,256 autosomal SNPs.

QC was performed using PLINK v1.9 and R v3.3.1. Firstly, nine individuals who had withdrawn consent from the study were excluded. Secondly, a crude level of SNP QC was applied prior to starting sample-level QC, by excluding SNPs with call rate <90%. This exclusion of poor call rate SNPs was performed by taking into account the two different genotyping platforms used by UK Biobank, by invoking the `--oblig-clusters` and `--oblig-missing` flags in PLINK. 6,683 SNPs were excluded.

2.2.3.1 Sample QC

Sample-level QC was initially performed by excluding individuals with abnormal levels of heterozygosity. For this step, I used the list of SNPs flagged in UK Biobank’s QC (Data Field 22004) as having principal component analysis (PCA)-adjusted heterozygosity >3 S.D. from the mean. 4,673 samples were flagged for exclusion. Next, individuals with discrepancies between their self-reported sex and their genetic sex were excluded, including those individuals with evidence of sex chromosome aneuploidy (configurations that are neither XX nor XY). These samples were identified in UK Biobank’s Data Field 22019, and 849 samples were flagged for exclusion. Finally, individuals with <98% sample call rate were flagged for exclusion; again, computations of call rate were performed by taking the two different genotyping platforms into account, leading to 7,004 individuals being flagged for exclusion.

In a GWAS, it is of critical importance to control for population stratification – an intrinsic source of systematic bias resulting from the presence of a systematic difference in allele

frequencies between subpopulations within a cohort. When population stratification exists, associations between genotypes and a trait of interest may be confounded, leading to false positive or false negative associations¹⁵³. Due to significant variations in the allele frequencies of genetic variants between humans of different ancestral origins, the vast majority of GWAS are undertaken in an ethnically homogenous dataset. Given that a large majority of UK Biobank participants are of white British ethnicity, all analyses were restricted to individuals of this ethnicity. UK Biobank, in Data Field 22019, flagged the individuals who were of white British ancestry based on self-reported questionnaire data and principal component analysis (PCA) of genotype data. Those 78,762 individuals who did not meet these criteria were therefore excluded in my QC pipeline.

Based on the above sample QC criteria, a total of 86,693 individuals flagged for exclusion were removed from the dataset. In order to ensure that the remaining individuals were all of white British ethnicity, the remaining 401,673 individuals were merged with publicly available data from the 1000 Genomes Project, and PCA was performed using flashpca¹⁵⁴. The first principal component was plotted against the second principal component, which clearly demonstrated the separation of ethnic groups, with individuals of European ancestries clustered at the apex of the graph (**Fig. 2.2**). The selected UK Biobank population post-sample QC overlapped nearly perfectly with the GBR (“British in England and Scotland”) population from 1000 Genomes, demonstrating that the QC had successfully excluded those individuals of non-white British ethnicity.

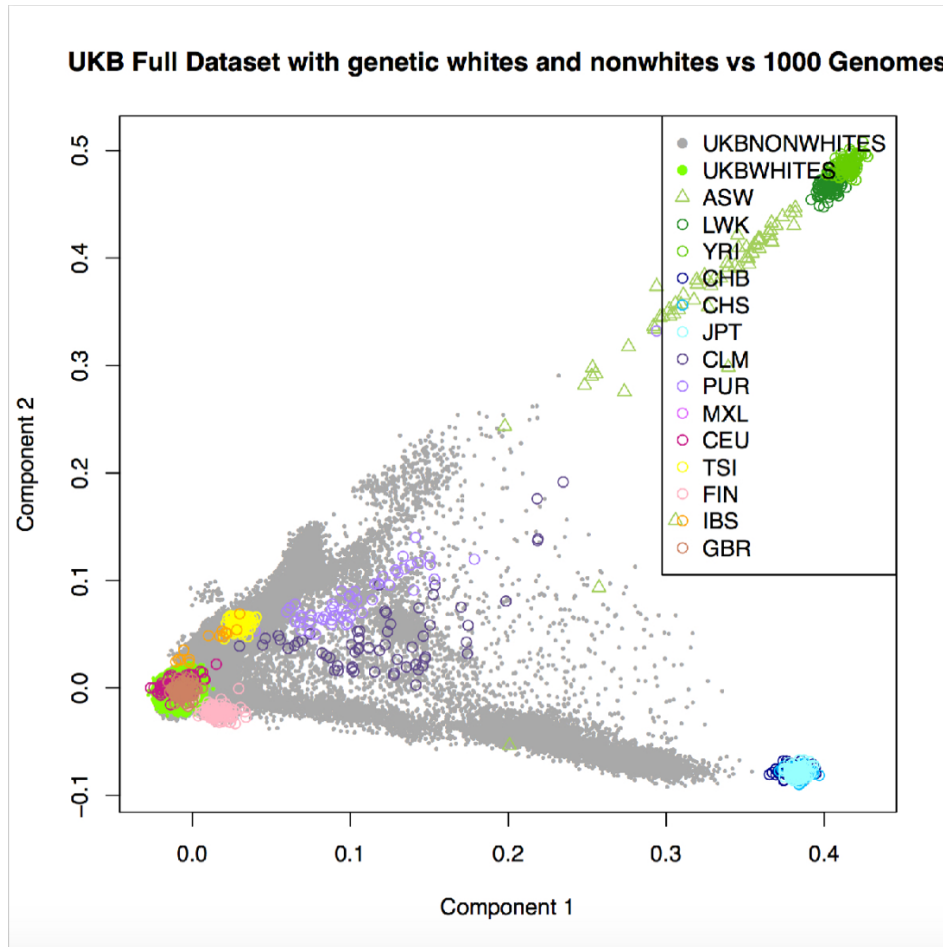


Fig. 2.2 Principal component analysis showing UK Biobank ancestries. UK Biobank white British samples (lime green) can be seen at the apex of the graph, perfectly overlapping the 1000 Genomes British samples (GBR, light brown). Grey dots = UK Biobank samples excluded from cases and controls.

2.2.3.2 SNP QC

Sample QC was followed by SNP-based QC. First, all poorly genotyped SNPs with a call rate <98% were flagged for exclusion. This was followed by exclusion of SNP markers with significant deviations from Hardy-Weinberg equilibrium (HWE, a common indicator of genotyping error). The PLINK --nonfounders and --hardy commands were used to flag the 81,964 SNPs with HWE p-value < 1×10^{-4} . As statistical power is limited for SNPs that are very rare in the population, a minor allele frequency (MAF) threshold of >1% was imposed to keep only relatively common SNPs for subsequent analyses, leading to 130,707 SNPs being flagged

for exclusion. Based on these parameters, 230,562 SNPs were excluded from the dataset, leaving 547,011 SNPs.

2.2.3.3 Pre-association study exclusions

Following sample- and SNP-based QC, 401,667 individuals and 547,011 SNPs remained in the dataset. Two final QC steps were implemented: firstly, I excluded six samples which harboured an abnormal number of SNPs with a minor allele count of 1, and were visual outliers when autosomal heterozygosity was plotted against call rate. Secondly, 11 samples which were not present in UK Biobank's sample file accompanying the imputed genotype files (presumed to be individuals who had withdrawn their consent), were excluded. The final dataset following QC therefore consisted of 401,656 individuals and 547,011 SNPs in chromosomes 1-22. A summary of the QC pipeline is shown in **Figure 2.3**.

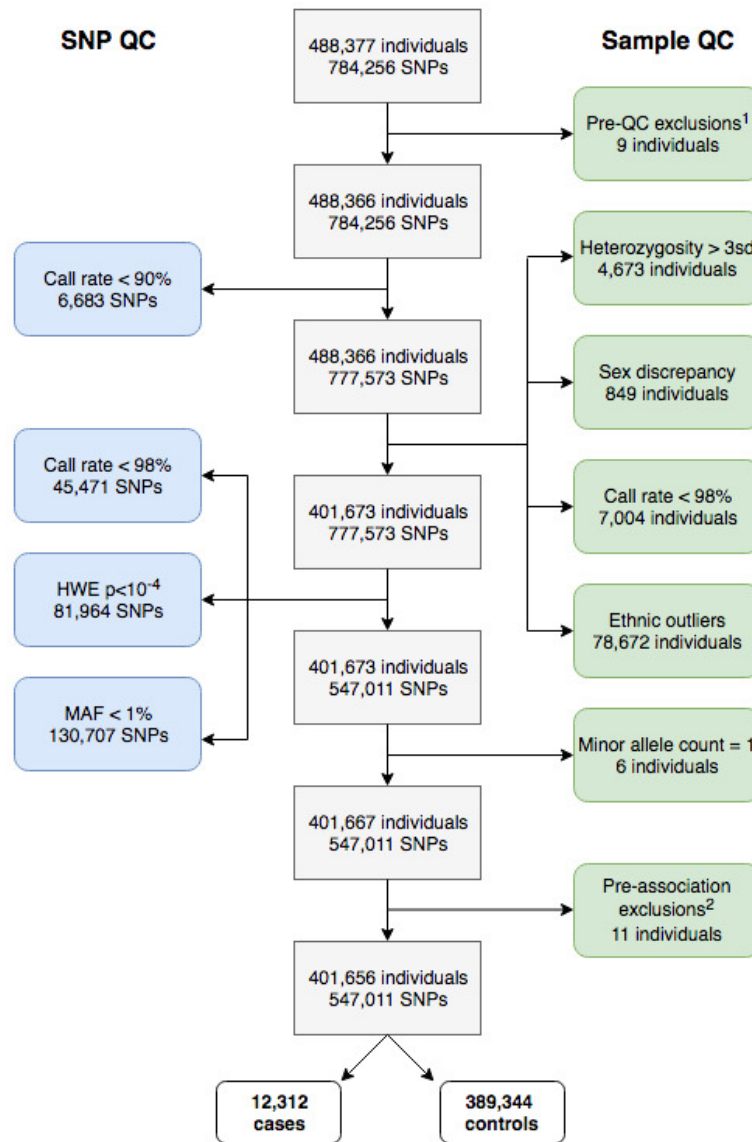


Fig. 2.3 Overview of QC pipeline. Excluded SNPs are in the blue panels and excluded samples are in the green panels. ¹Pre-QC exclusions: 3 individuals with invalid IDs and sex, and 8 individuals who had withdrawn from the study were excluded prior to QC. ²Pre-association exclusions: 11 individuals who were not present in UK Biobank's sample file accompanying the imputed genotype data were excluded prior to association.

2.2.4 Case Ascertainment

CTS cases were identified using diagnostic and operative codes from the UK Biobank showcase, downloaded in June 2017. Individuals were selected as CTS cases if they had any one of the following diagnostic codes (the codes are given in parentheses):

1. Primary ICD-10 codes for CTS (G560)
2. Secondary ICD-10 code for CTS (G560)
3. Primary OPCS code for carpal tunnel release (A651)
4. Primary OPCS code for revision carpal tunnel release (A692)
5. Secondary OPCS code for carpal tunnel release (A651)
6. Secondary OPCS code for revision carpal tunnel release (A692)
7. Self-reported non-cancer illness code for CTS (1541)
8. Self-reported operation code for carpal tunnel surgery (1501)

Excluding overlaps between the above groups, a total of 15,241 individuals had at least one CTS code and were classed as “cases”. A total of 12,312 CTS cases were present within the post-QC dataset of 401,656 individuals,. The remaining 389,344 individuals were classed as “controls” in the association study. **Table 2.1** shows the number of individuals with each diagnostic code.

Table 2.1. Diagnostic codes used for CTS case definition. The number of individuals with each of the diagnostic codes are shown.

Source of Data	UK Biobank Data Field	Code	Description	N
Primary ICD-10	41202	G560	Carpal Tunnel Syndrome	9,139
Secondary ICD-10	41204	G560	Carpal Tunnel Syndrome	648
Primary OPCS	41200	A651	Carpal Tunnel Release	9,071
		A692	Revision of Carpal Tunnel Release	97
Secondary OPCS	41210	A651	Carpal Tunnel Release	452
		A692	Revision of Carpal Tunnel Release	7
Non-cancer illness (self-report)	20002	1541	Carpal Tunnel Syndrome	965
Operation (self-report)	20004	1501	Carpal Tunnel Surgery	4,584
Total (excluding overlaps)				12,312

An independent validation study into the diagnostic accuracy of the above codes with respect to CTS was previously undertaken by other members in my research group¹⁵⁵. The hospital medical notes of several hundred UK Biobank participants with CTS diagnostic codes were interrogated, and the overall positive predictive value for individuals with CTS diagnoses being true CTS cases in UK Biobank was determined to be 94%.

2.2.5 Association Study

The association study used a combination of the 547,011 genotyped post-QC SNPs, and approximately 8.4 million imputed SNPs supplied by UK Biobank. For imputation, UK Biobank performed phasing on the autosomes using SHAPEIT3, using the 1000 Genomes Phase 3 dataset as the reference panel. Imputation was performed using a reference panel constituting both HRC and a merged UK 10K/1000 Genomes Phase 3 panel, resulting in a dataset with approximately 92 million autosomal SNPs, large structural variants and short indels¹⁵². UK Biobank imputation files were released in the BGEN v1.2 format.

A stringent quality control was performed on the imputed SNPs, by imposing an INFO score cut-off of ≥ 0.9 , and restricting imputed SNPs to those with a MAF ≥ 0.01 , as per the genotyped SNPs. This resulted in approximately 8.4 million imputed SNPs for the association study, in addition to the 547,011 genotyped SNPs.

The association study was undertaken using a linear mixed non-infinitesimal model implemented in BOLT-LMM v2.3¹⁵⁶. Such mixed-model association methods have become a widely adopted method for performing association analyses for very large datasets such as UK Biobank. The principal advantage conferred by this method (over more traditional linear regression methods) is that it is robust to potential confounding due to relatedness, meaning

that individuals need not be excluded on the basis of kinship. This has significant power implications for UK Biobank, where removing samples based on relatedness would exclude a substantial proportion of the cohort¹⁵². The methodology is also highly computationally efficient even for very large datasets, and increases the power to detect associations by conditioning on polygenic predictions from genome-wide SNPs, which reduces the noise in an association test¹⁵⁷.

A minimally adjusted model was chosen in this association study to avoid potential collider bias¹⁵⁸, which has the potential to induce spurious associations. Thus, there was no adjustment for heritable covariates known to associate with CTS (such as diabetes, rheumatoid arthritis, etc.), and only two covariates were adjusted for in the association analysis: sex and genotyping platform (i.e. whether the UK Biobank Axiom or BiLEVE array were used for genotyping individuals).

Following the association study, conditional analysis was undertaken at each locus containing genome-wide significant SNPs, in order to identify whether more than one independent association signal was present at that locus. QCTOOL v2 was used to compute the allelic dosage of the top-associated (index) SNP at each locus for each individual, and an association analysis was repeated on that chromosome, using the allelic dosage of the index SNP as an additional covariate.

A quantile-quantile (Q-Q) plot (**Fig. 2.4**) and Manhattan plot (**Fig. 2.5**) were created in R to visualise the degree of genomic inflation and the association peaks, and regional plots were created for each of the associated loci with a ± 500 kb window using LocusZoom¹⁵⁹, to scrutinise the genomic landscape around each association signal and visualise the LD relationship between the associated SNPs (**Fig. 2.6**).

2.2.6 Replication in EGCUT

2.2.6.1 Description of the EGCUT cohort

The Estonian Biobank is a population-based biobank of the Estonian Genome Center at the University of Tartu (EGCUT). All 51,886 participants have signed a broad informed consent form, allowing periodical linking to national registries, electronic medical record databases and hospital information systems¹⁶⁰. Whole genome sequencing data are available for the 2,535 individuals who were selected randomly by county of birth. Participants were genotyped on Illumina genotyping arrays: (1) Global Screening Array (N=33,277), (2) HumanCoreExome (N=7,832), (3) HumanOmniExpress (N=8,137), and (4) 370K (N=2,640).

2.2.6.2 Genotyping, QC, and association analysis

Illumina's GenomeStudio v2010.3 was used for genotype calling. After variant calling, sample QC was performed using PLINK (call rate >95%, no sex mismatches, heterozygosity < mean $\pm$ 3 S.E.). SNP QC was also performed in PLINK (HWE $p > 1 \times 10^{-6}$, call rate >95%), with an additional layer of QC for markers on the GSA array: a cluster separation score >0.4 and an Illumina GenomeStudio GenTrain score >0.6. Indels, variants with MAF <1%, and G/C or A/T polymorphisms were removed prior to imputation.

Pre-phasing of genotyped data was performed using Eagle¹⁶¹ v2.3 with default parameters, and imputation performed with BEAGLE5 v4.1, using a population-specific imputation reference panel¹⁶². The four imputed arrays were merged into a single dataset using BCFtools, and imputation information measures (INFO scores) were added. Monomorphic and multi-allelic imputed variants were excluded.

Case ascertainment for CTS performed using a mixture of self-reported and physician-confirmed CTS diagnoses identified a total of 4,438 CTS cases and 45,958 controls.

The GWAS was performed using an additive model in SNPTEST¹⁶³ using the following covariates: sex, year of birth, and the first 10 principal components. The summary statistics for the index SNPs at loci that were genome-wide significant in the UK Biobank discovery GWAS were extracted from the EGCUT summary statistics to allow comparison across the two studies.

2.2.6.3 Genetic correlations between UK Biobank and EGCUT

In order to assess the degree of genetic correlation between the CTS GWAS in UK Biobank and EGCUT, bivariate LD score correlation (LDSC) analysis¹⁶⁴ was performed. This method derives the heritability of a phenotype by regressing a SNP's association statistic onto its LD score (the sum of squared correlations between the minor allele count of a SNP and the minor allele count of every other SNP). LDSC analysis determines whether there is correlation between the traits in two different GWA studies by using their summary statistics. Genetic correlations that are statistically and quantitatively significant suggest that the phenotypic correlation is not attributable solely to environmental confounders. LD score analysis was also used to compute a value for the LD score intercept and SNP-based heritability of CTS in the UK Biobank GWAS¹⁶⁵. Finally, a linear regression was performed in order to compute the degree of correlation (r^2) between the odds ratios for each of the associated SNPs in the UK Biobank and EGCUT GWAS.

2.3 Results

2.3.1 Sixteen novel loci associated with CTS risk

There were 422 genome-wide significant ($p < 5 \times 10^{-8}$) associations at 16 loci across nine chromosomes. Conditional analysis demonstrated no residual signal at the associated loci.

The λ_{GC} of 1.15 demonstrated evidence of genomic inflation, as demonstrated in the Q-Q plot (Fig. 2.4). However, the LDSC regression intercept was 1.015, with an attenuation ratio ((LD intercept – 1)/(mean $\chi^2 - 1$)) of 0.073. This indicates that the genomic inflation was largely due to polygenicity and large sample size, in keeping with other GWAS of similar size. The SNP-based heritability estimate in this GWAS calculated by LD score regression was 2.4% (SE = 0.17%).

Fig. 2.5 shows all of the SNP-phenotype associations in the form of a Manhattan plot.

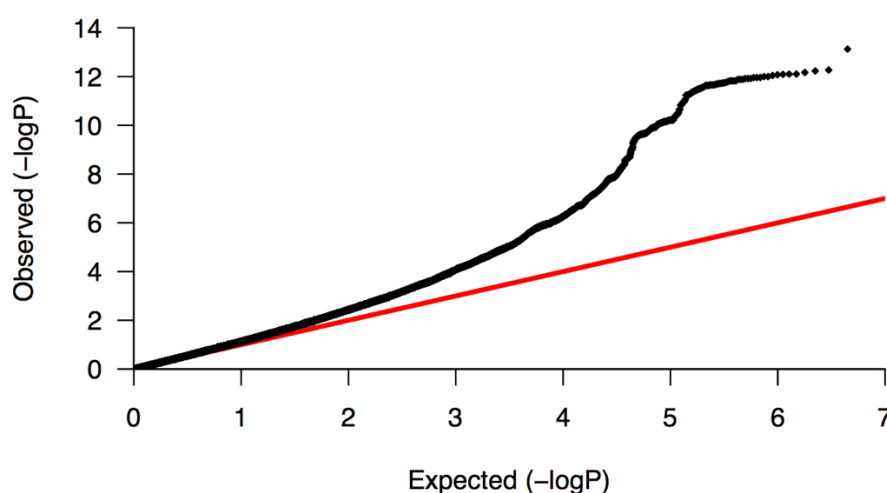


Fig. 2.4 Quantile-quantile (Q-Q) plot of associated SNPs. The degree of genomic inflation is in keeping with polygenicity and the large sample size.

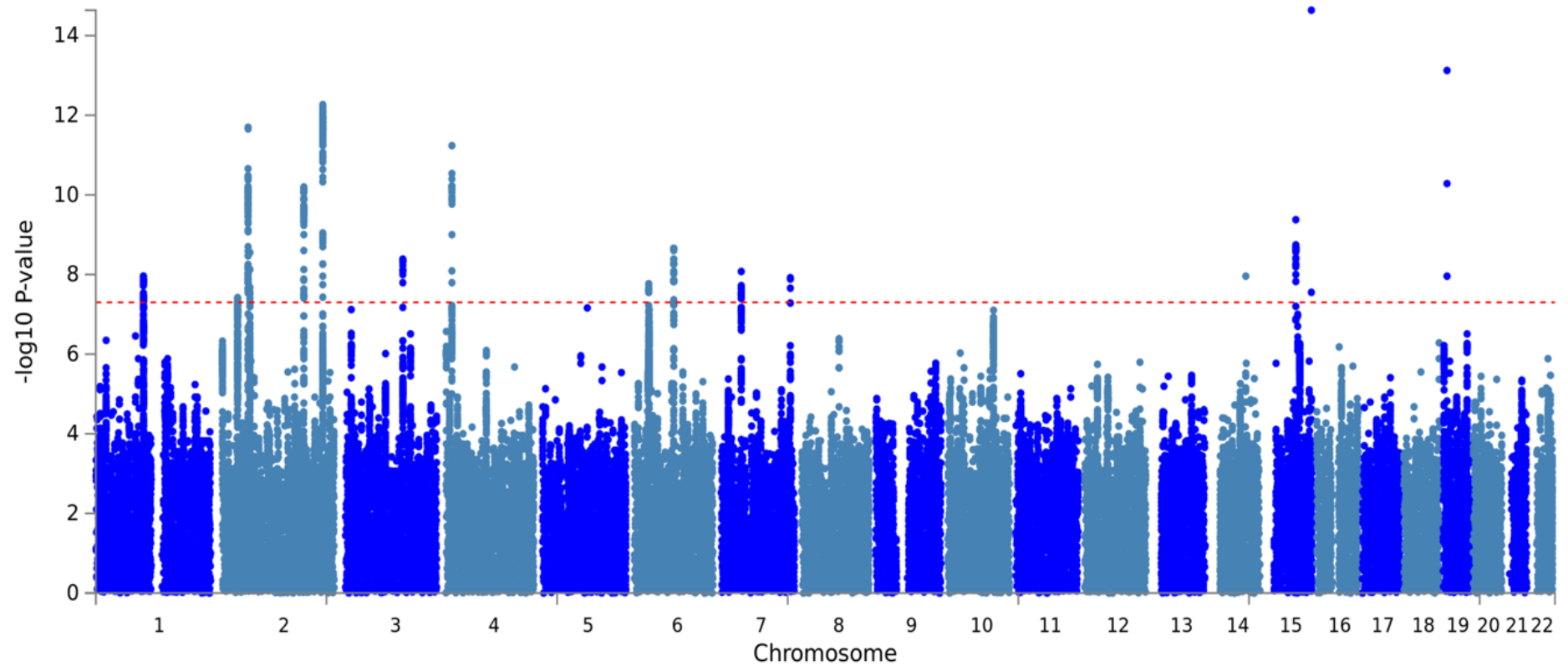


Fig. 2.5 SNP associations with CTS. Manhattan plot showing $-\log_{10}$ p-values for SNP associations. The horizontal red line represents $p = 5 \times 10^{-8}$, and any SNPs lying above this line are genome-wide significant.

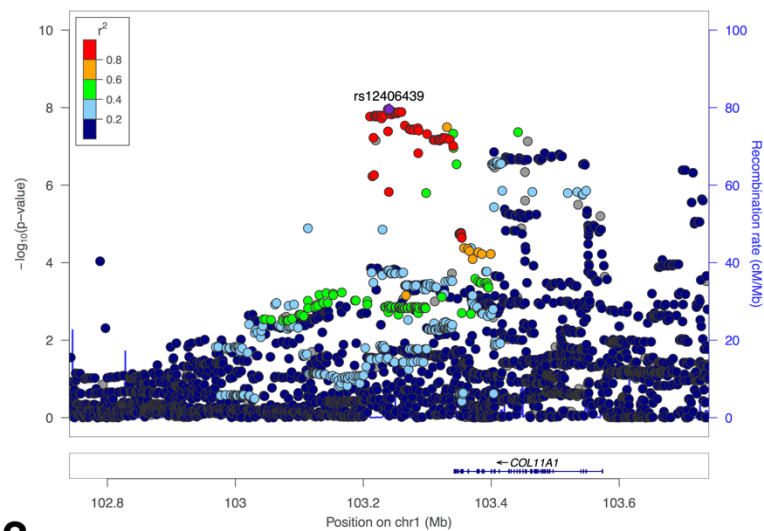
The most significantly associated SNP (index SNP) at each locus is shown in **Table 2.2**. The FUMA method employed to map candidate genes to the associated loci will be discussed in detail in the following chapter. The index SNPs at each locus were of high quality (three genotyped; minimum INFO score for the imputed SNPs ranged between 0.924 to 0.999), and all were common SNPs, with minor allele frequencies ranging from 3.6% to 49.9%. Odds ratio for the effect allele ranged from 1.08-1.31. Regional plots of the 16 associated loci are shown in **Fig. 2.6**.

Table 2.2. SNPs significantly associated with Carpal Tunnel Syndrome.

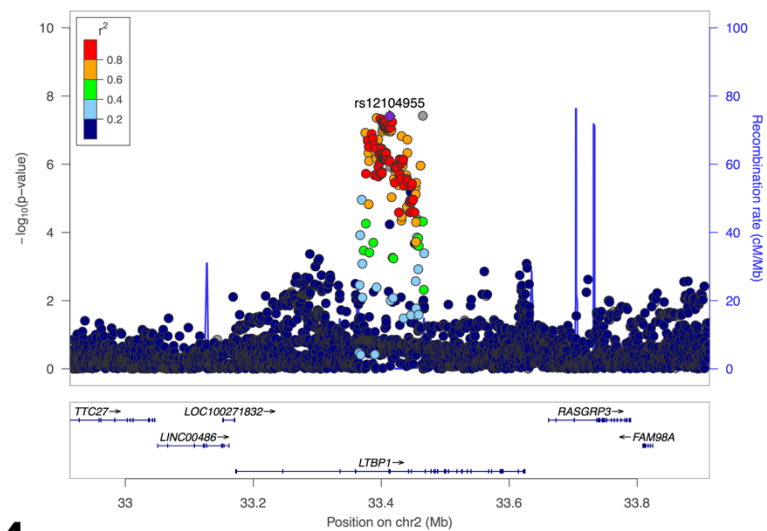
Chromosome	Position ^a	rsID	Effect Allele	EAF Cases ^b	EAF Controls ^c	INFO ^d	OR (95% CI)	P-value	Mapped genes ^e
1	103240351	rs12406439	T	0.605	0.587	0.995021	1.08 (1.05-1.11)	1.10×10 ⁻⁸	<i>COL11A1</i>
2	33413303	rs12104955	C	0.517	0.499	0.994032	1.07 (1.05-1.10)	3.90×10 ⁻⁸	<i>LTBP1</i>
2	56096892	rs3791679	G	0.244	0.225	G	1.11 (1.08-1.15)	2.00×10 ⁻¹²	<i>EFEMP1</i>
2	60175475	rs1025128	C	0.582	0.563	0.988458	1.08 (1.05-1.11)	2.80×10 ⁻⁹	-
2	176900271	rs847139	C	0.805	0.787	0.985197	1.11 (1.07-1.14)	7.20×10 ⁻¹¹	<i>KIAA1715, EVX2</i>
2	218128152	rs1863190	T	0.78	0.76	0.997145	1.12 (1.08-1.15)	5.40×10 ⁻¹³	<i>DIRC3</i>
3	124450081	rs4678145	G	0.88	0.868	0.999293	1.12 (1.08-1.16)	4.10×10 ⁻⁹	<i>ITGB5, UMPS, KALRN, MUC13</i>
4	13221747	rs6843953	T	0.154	0.138	0.99611	1.14 (1.10-1.18)	5.80×10 ⁻¹²	-
6	31440651	rs3828889	C	0.748	0.732	0.996908	1.09 (1.06-1.12)	1.70×10 ⁻⁸	-
6	85715955	rs62422907	G	0.899	0.887	0.991783	1.13 (1.09-1.18)	2.20×10 ⁻⁹	-
7	44145178	rs55841377	C	0.789	0.773	0.992012	1.09 (1.06-1.13)	8.40×10 ⁻⁹	<i>AEBP1, PLD2, MYL7, GCK</i>
7	150542515	rs6977081	G	0.685	0.668	0.97855	1.08 (1.05-1.11)	1.20×10 ⁻⁸	<i>AOC1, TMEM176A, KCNH2, TMEM176B</i>
14	76245906	rs72725608	C	0.051	0.044	G	1.20 (1.13-1.27)	1.10×10 ⁻⁸	<i>IFT43, TTLL5, TGFB3</i>
15	67034812	rs1866745	A	0.369	0.35	0.993487	1.09 (1.06-1.12)	4.20×10 ⁻¹⁰	<i>SMAD6</i>
15	100692953	rs72755233	A	0.128	0.112	G	1.18 (1.13-1.22)	2.30×10 ⁻¹⁵	<i>ADAMTS17</i>
19	8670147	rs62621197	T	0.045	0.036	0.924188	1.31 (1.22-1.40)	7.50×10 ⁻¹⁴	<i>ADAMTS10, MYO1F</i>

^aBased on NCBI Genome Build 37 (hg19). ^bThe effect allele frequency in cases. ^cThe effect allele frequency in controls. ^dThe SNP INFO score for imputed SNPs; G = genotyped SNP. ^eGenes mapped to these loci based on positional mapping in FUMA (see Methods in Chapter 3).

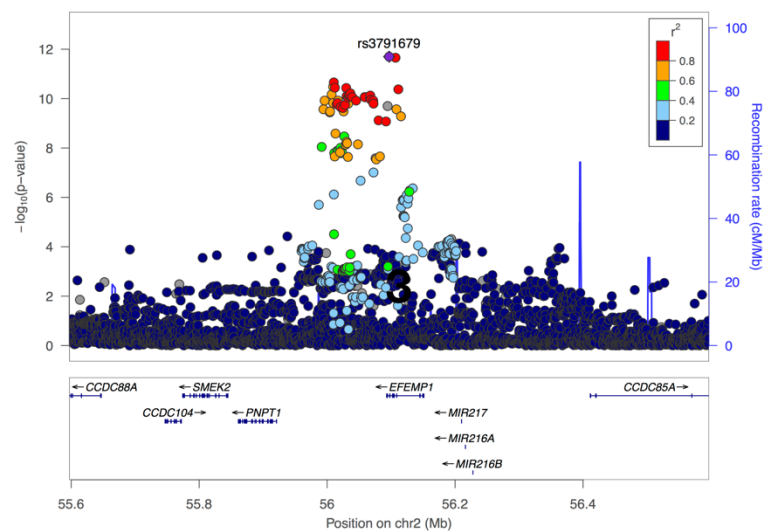
1



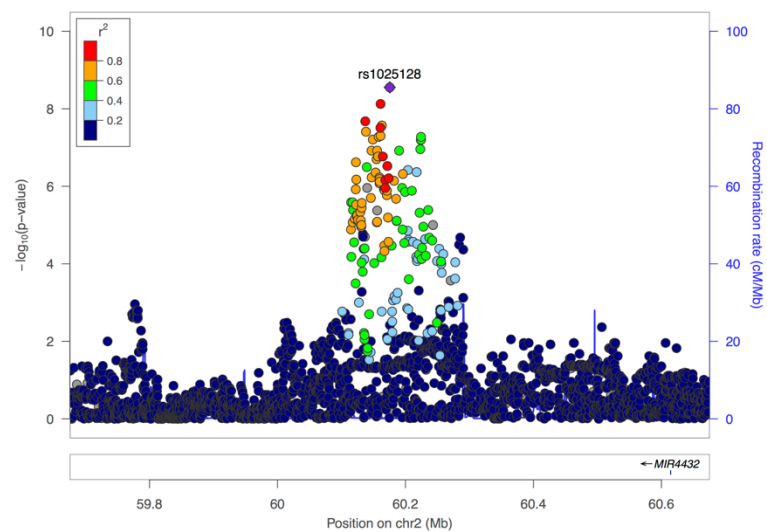
2



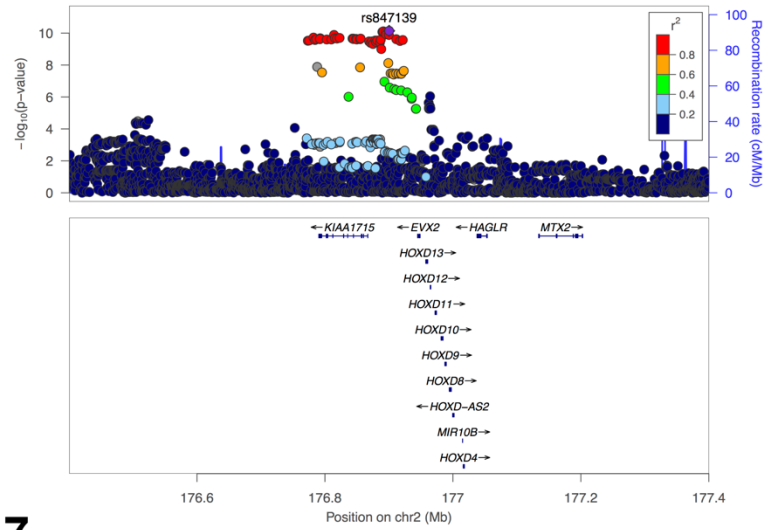
3



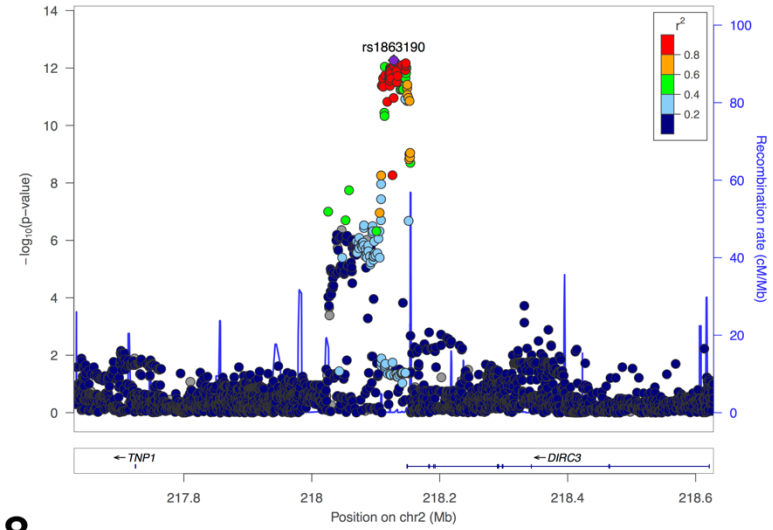
4



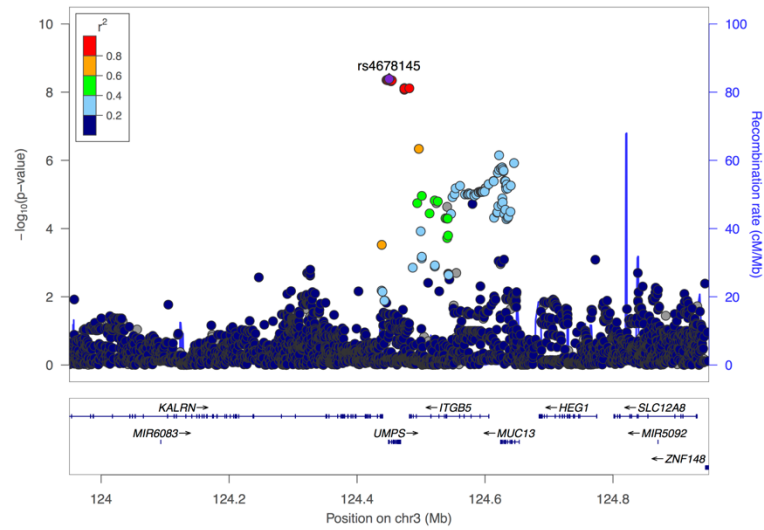
5



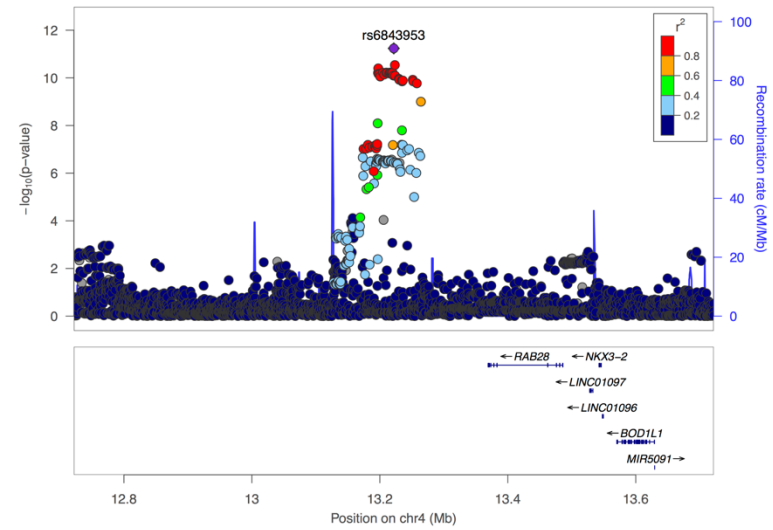
6



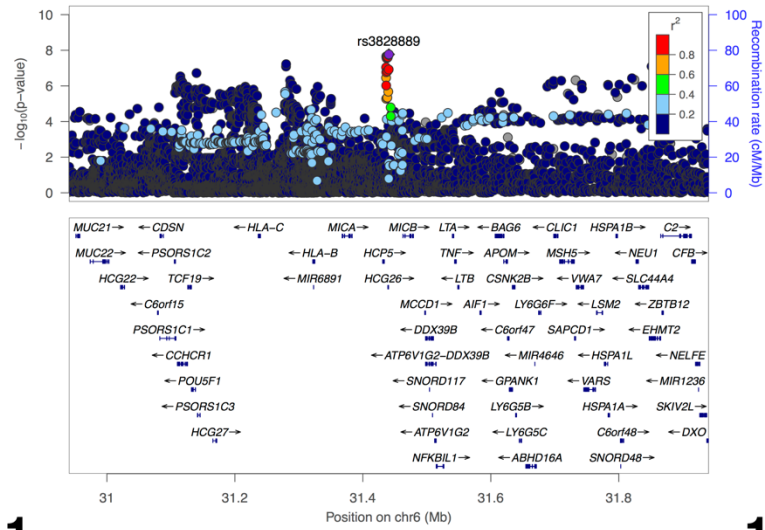
7



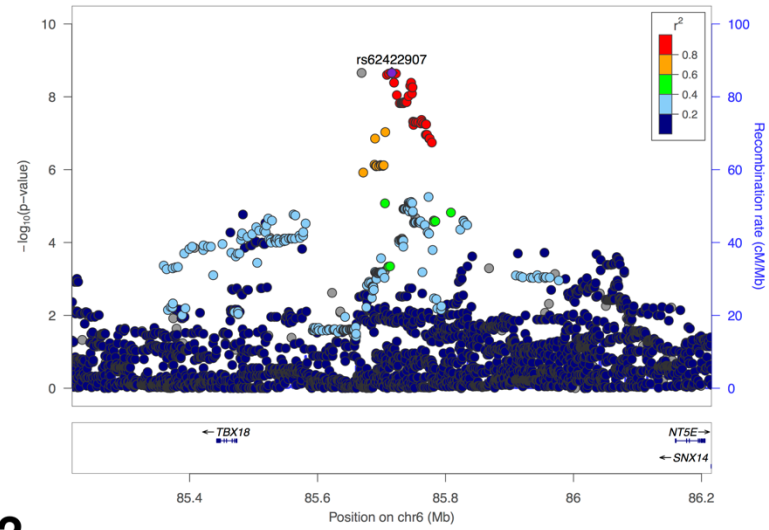
8



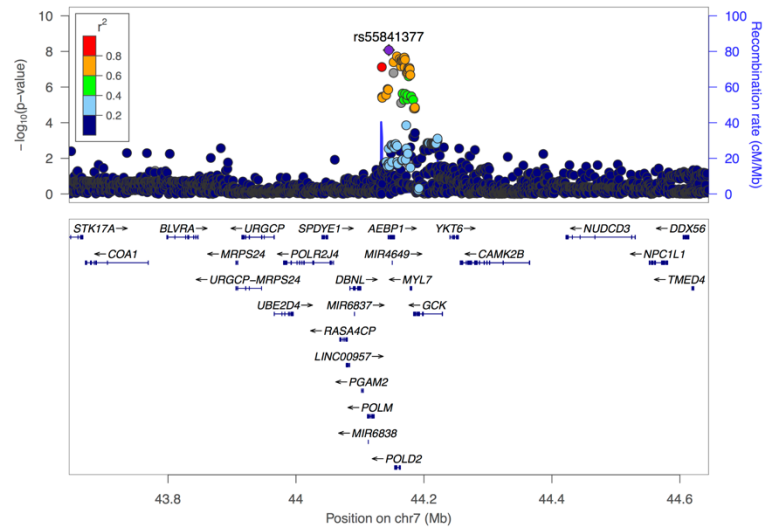
9



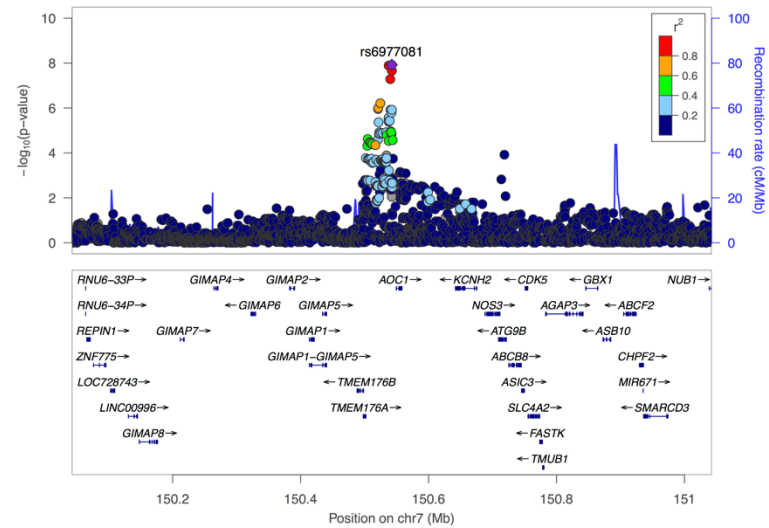
10



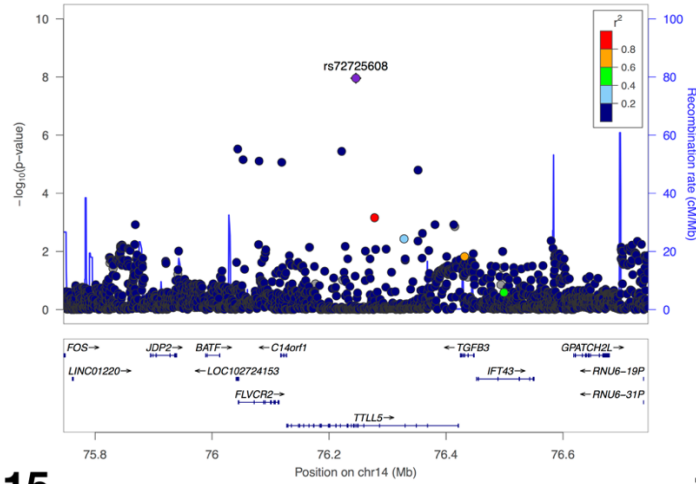
11



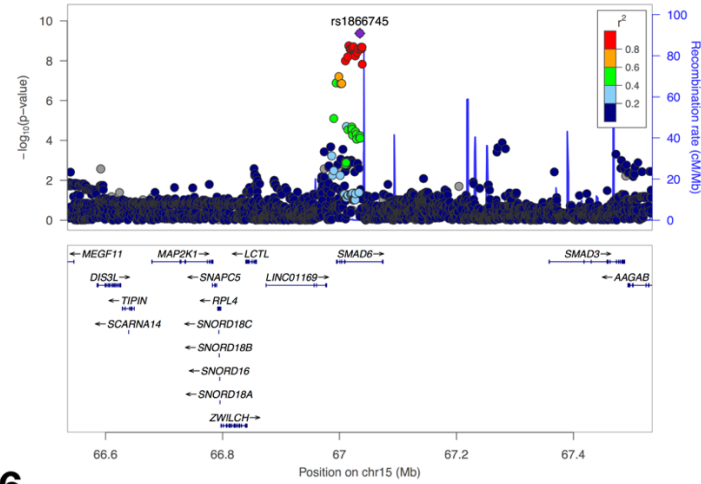
12



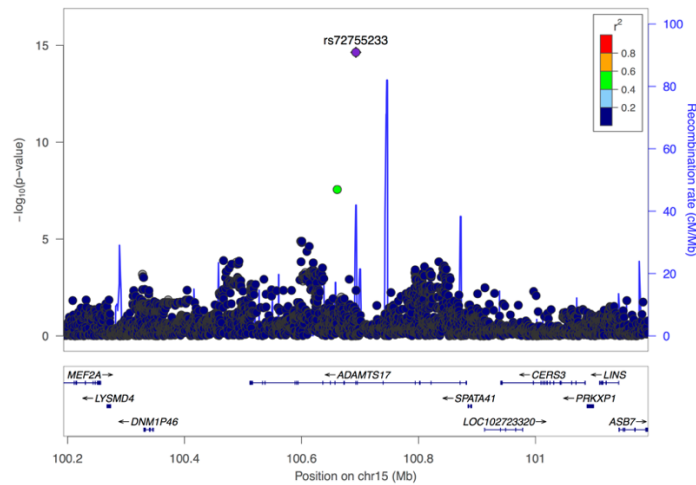
13



14



15



16

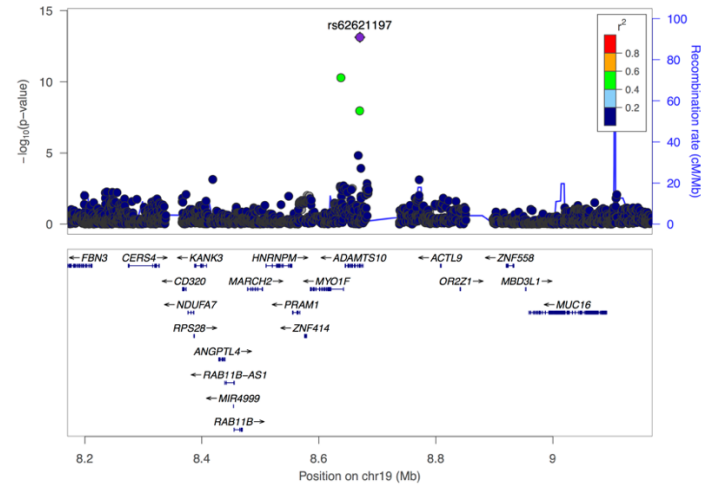


Fig. 2.6. Regional association plots for the 16 associated loci. Loci are numbered sequentially in ascending chromosomal and positional order. SNP position is shown on the x-axis, and the strength of association on the y-axis. The LD relationships between the lead SNP and its surrounding SNPs are indicated by colour. Genes within 500 kb of the index SNP are shown on the lower panel, and positions are shown in relation to Human Genome build hg19. The sparsity of SNPs in LD with the index SNPs at loci 15 (*ADAMTS17*) and 16 (*ADAMTS10*) are consistent with regional plots of these SNPs in a previous study¹⁶⁶.

2.3.2 Replication in EGCUT cohort

The 16 index SNPs at the associated loci in the UK Biobank GWAS were extracted from the summary statistics from the EGCUT CTS GWAS. The p-values and odds ratios for each of these SNPs, in addition to whether the alleles are directionally concordant in the two studies have been summarised in **Table 2.3**.

One out of the 16 SNPs replicated at a Bonferroni-corrected significance threshold of $p < 0.0031$ ($0.05/16$), and a further two SNPs were nominally significant at $p \leq 0.05$. Thirteen out of 16 SNPs demonstrated directional concordance of effects across the two studies.

LDSC found that the genetic correlation (r_g) between the UK Biobank and EGCUT GWAS of CTS was 0.904 (SE 0.0031), Z score = 5.76, $p = 8.19 \times 10^{-9}$, and linear regression of the odds ratios for the SNPs between the two studies demonstrated a statistically significant positive correlation ($r^2 = 0.69$, F statistic: 31.3, $P = < 0.0001$) (**Fig. 2.7**).

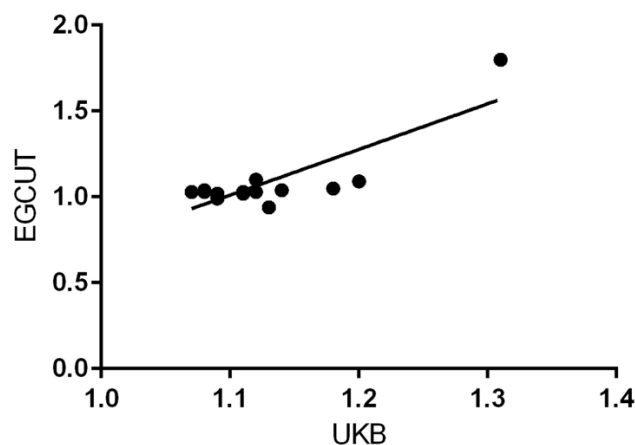


Fig. 2.7. UK Biobank vs EGCUT odds ratio correlation. The odds ratios for each of the 16 associated SNPs is shown for the two studies. A statistically significant correlation of 0.69 was found.

Table 2.3 UK Biobank vs EGCUT summary association statistics. For each of the 16 index SNPs discovered in the UK Biobank CTS GWAS, the corresponding allele frequencies, odds ratios and p-values are shown for the EGCUT CTS GWAS. The p-value that was significant in EGCUT after Bonferroni correction is shown in ***bold italic***. The final column indicates whether the allelic effects are directionally concordant across the two studies.

UK BIOBANK						EGCUT					
rsID	Effect Allele	EAF Cases	EAF Controls	OR (95% CI)	P-value	Effect Allele	EAF Cases	EAF Controls	OR (95% CI)	P-value	Directionally concordant?
rs12406439	T	0.605	0.587	1.08 (1.05-1.11)	1.10×10 ⁻⁸	T	0.627	0.617	1.04 (0.99-1.09)	0.050	Yes
rs12104955	C	0.517	0.499	1.07 (1.05-1.10)	3.90×10 ⁻⁸	C	0.455	0.447	1.03 (0.99-1.08)	0.116	Yes
rs3791679	G	0.244	0.225	1.11 (1.08-1.15)	2.00×10 ⁻¹²	G	0.276	0.271	1.02 (0.98-1.08)	0.327	Yes
rs1025128	C	0.582	0.563	1.08 (1.05-1.11)	2.80×10 ⁻⁹	C	0.676	0.668	1.03 (0.99-1.09)	0.105	Yes
rs847139	C	0.805	0.787	1.11 (1.07-1.14)	7.20×10 ⁻¹¹	C	0.820	0.815	1.03 (0.97-1.09)	0.169	Yes
rs1863190	T	0.780	0.760	1.12 (1.08-1.15)	5.40×10 ⁻¹³	T	0.708	0.701	1.03 (0.98-1.08)	0.264	Yes
rs4678145	G	0.880	0.868	1.12 (1.08-1.16)	4.10×10 ⁻⁹	G	0.914	0.905	1.10 (1.02-1.20)	0.014	Yes
rs6843953	T	0.154	0.138	1.14 (1.10-1.18)	5.80×10 ⁻¹²	T	0.096	0.0924	1.04 (0.97-1.13)	0.338	Yes
rs3828889	C	0.748	0.732	1.09 (1.06-1.12)	1.70×10 ⁻⁸	C	0.692	0.694	0.99 (0.95-1.04)	0.757	No
rs62422907	G	0.899	0.887	1.13 (1.09-1.18)	2.20×10 ⁻⁹	G	0.911	0.916	0.94 (0.86-1.01)	0.088	No
rs55841377	C	0.789	0.773	1.09 (1.06-1.13)	8.40×10 ⁻⁹	C	0.841	0.838	1.02 (0.95-1.08)	0.500	Yes
rs6977081	G	0.685	0.668	1.08 (1.05-1.11)	1.20×10 ⁻⁸	G	0.610	0.603	1.03 (0.98-1.08)	0.239	Yes
rs72725608	C	0.051	0.044	1.20 (1.13-1.27)	1.10×10 ⁻⁸	C	0.0335	0.031	1.09 (0.96-1.24)	0.165	Yes
rs1866745	A	0.369	0.35	1.09 (1.06-1.12)	4.20×10 ⁻¹⁰	A	0.324	0.325	0.997 (0.95-1.24)	0.848	No
rs72755233	A	0.128	0.112	1.18 (1.13-1.22)	2.30×10 ⁻¹⁵	A	0.145	0.139	1.05 (0.99-1.12)	0.086	Yes
rs62621197	T	0.045	0.036	1.31 (1.22-1.40)	7.50×10 ⁻¹⁴	T	0.049	0.042	1.18 (1.06-1.31)	<i>0.002</i>	Yes

2.4 Discussion

This GWAS of CTS in UK Biobank demonstrates that there is a robust polygenic architecture underlying CTS, confirming the evidence from the literature that there is a significant genetic contribution to its aetiology. In keeping with other published GWAS, the SNP-based heritability (i.e. the proportion of the liability-scale genetic variance explained by all the SNPs) of 2.4% is considerably lower than the 46% heritability estimated from twin studies¹¹⁶, and should ideally be corroborated using alternative methods such as single-component GREML¹⁶⁷. However, this discrepancy between the heritability estimates from twin studies and SNP-based heritability estimates from GWAS illustrates the concept of ‘missing heritability’ – GWAS variants typically explain only a small proportion of the heritability estimates derived from twin studies¹⁶⁸. This has been a subject of considerable debate in the genetics community for well over a decade¹⁶⁹, but it is becoming increasingly clear from association studies using whole genome sequencing data that the missing heritability of complex traits is probably accounted for by rare- to very rare variants¹⁷⁰. Thus, a typical MAF cut-off of >1%, as imposed in this study, means that the cumulative contribution of such rare variants goes undetected.

A corollary to this concept is that even if such variants were sampled, the SNP-based heritability estimate is fundamentally a function of statistical power. If, for example, the number of cases in this study could be doubled to 24,000, a significantly greater number of association signals are likely to be uncovered. Evidence for this can be seen in the Manhattan plot (**Fig. 2.5**), which demonstrates several very plausible-looking association signals that just fail to meet the genome-wide significance threshold, some of which would pass the threshold of significance as the sample size increases. Such loci are illustrated particularly well on chromosomes 3 and 10 (**Fig. 2.8**).

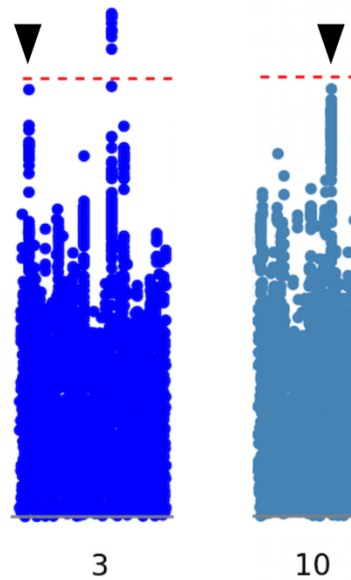


Fig. 2.8. Illustration of sub-threshold association signals. Zooming in on chromosomes 3 and 10 on the Manhattan plot (Fig 2.5) demonstrates two loci that just fail to reach genome-wide significance at $p = 5 \times 10^{-8}$ (black arrowheads). These signals are likely to be genuine associations with CTS.

A discussion of the individual associated loci will be deferred until the next chapter, where functional annotation of these loci will be discussed in conjunction with results from various of bioinformatic analyses.

In the Estonian replication study, 13 out of 16 loci discovered in the UK Biobank discovery GWAS were directionally concordant in the EGCUT replication cohort, but only one SNP (rs62621197) achieved significance at a Bonferroni-corrected p-value threshold. There is significant genetic correlation between the two cohorts, as evidenced by the correlation coefficient of 0.90 on LDSC and the positive correlation coefficient between the allelic odds ratios between the two studies. However, there are substantial limitations to this replication study that warrant discussion at some length, especially as they illustrate some of the vulnerabilities inherent in GWAS replication.

Firstly, there is a significant discrepancy between the mean age of the CTS cases in the two cohorts (69 in UK Biobank vs 50 in EGCUT) (**Table 2.4**).

Table 2.4. Demographic characteristics of UK Biobank vs EGCUT. The cohorts have been separated by sex and by CTS status. Standard deviations are shown in parentheses.

		UKB Cases	UKB Controls	UKB All	EGCUT Cases	EGCUT Controls	EGCUT All
Males	N	3,738	180,761	184,499	783	16,390	17,173
	Prevalence	0.02			0.046		
	Age	69.0 (7.4)	67.0 (7.4)	67.1 (8.1)	49.8 (14.9)	41.9 (18.0)	42.3 (17.9)
	Height (cm)	173.8 (6.7)	175.9 (6.7)	175.9 (6.8)	176.6 (6.8)	178.8 (7.3)	178.7 (7.3)
	Weight (kg)	90.35 (15.9)	86.1 (14.2)	86.2 (14.3)	88.4 (16.3)	84.1 (15.6)	84.2 (15.7)
Females	N	8,574	208,583	217,157	3,655	29,568	33,223
	Prevalence	0.039			0.11		
	Age	68.3 (7.1)	66.6 (8.0)	66.6 (7.9)	49.5 (13.2)	43.9 (17.6)	44.5 (17.2)
	Height (cm)	160.7 (6.2)	162.7 (6.2)	162.6 (6.2)	163.4 (6.1)	165.0 (6.5)	164.8 (6.5)
	Weight (kg)	75.3 (15.9)	71.4 (13.8)	71.5 (13.9)	76.1 (16.8)	70.3 (14.8)	71.0 (15.2)

Fig. 2.9, which shows the age distributions of the EGCUT cases and controls, demonstrates a near-normal distribution of CTS cases around the mean age of 50. Given that non-pregnancy-related CTS is very rare below the age of 40, it seems anomalous that the overall prevalence of CTS in the EGCUT cohort is substantially higher than in UK Biobank (8.5% vs 3.1%; Fisher's exact test $p < 1 \times 10^{-5}$), even though the EGCUT cases are on average nearly 20 years younger than the UK Biobank cases. UK Biobank participants were recruited at ages 40-69 (the peak age for CTS incidence), and the prevalence was still under 5%. While it must be remembered that the UK Biobank CTS prevalence is likely to be an underestimate due to reasons stated at the start of this chapter, the high CTS prevalence in EGCUT does raise two questions: (1) the robustness of the phenotyping methodology for CTS cases in EGCUT, and (2) whether the EGCUT CTS cases include a significant number of pregnancy-related CTS cases (which is usually a transient phenomenon). In support of (2), there is a large discrepancy in the

proportion of female CTS cases between the two cohorts (70% in UK Biobank vs 82% in EGCUT; Fisher's exact test $p < 1 \times 10^{-5}$).

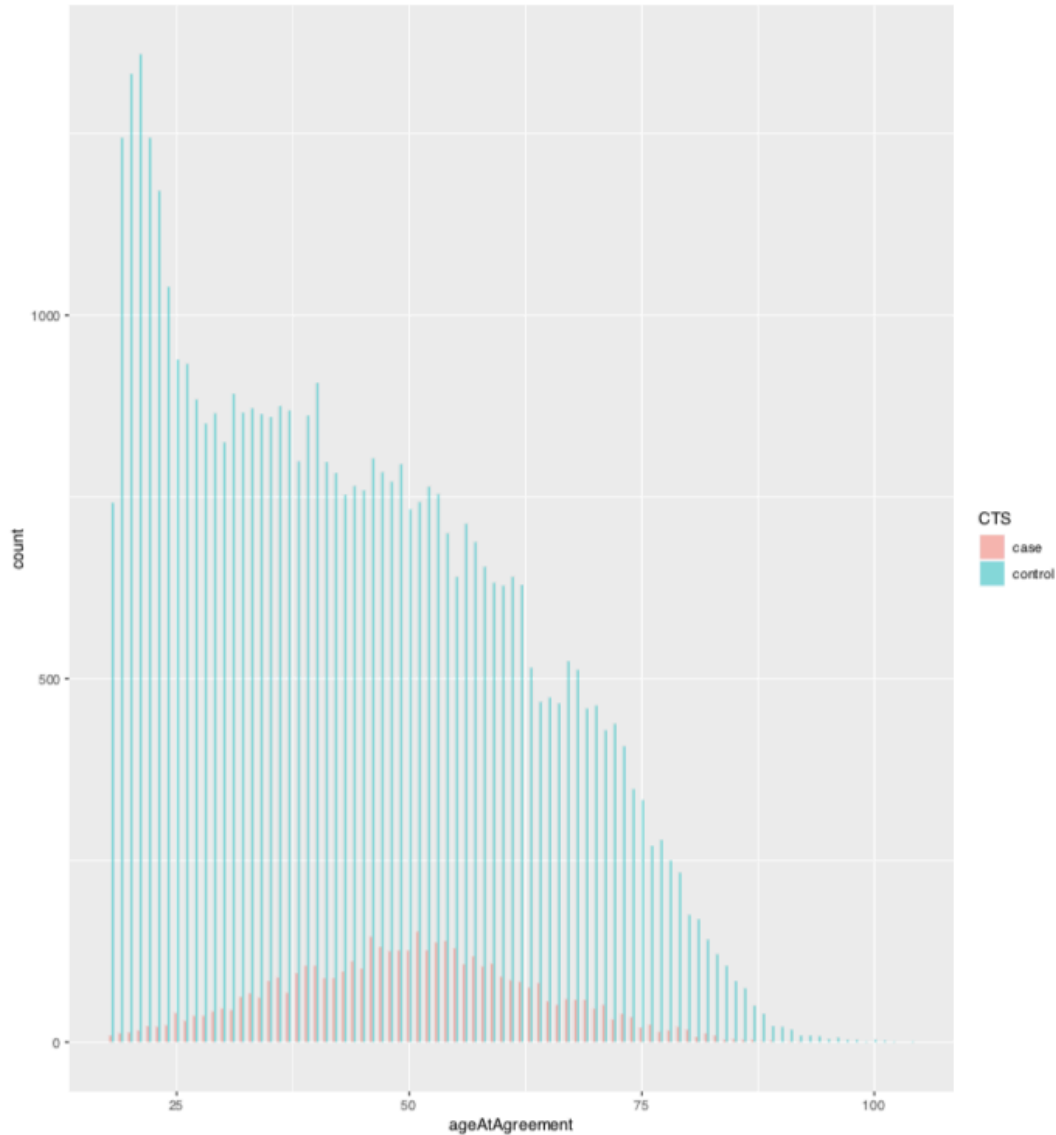


Fig. 2.9. Age distribution of EGCUT cases and controls. This histogram demonstrates that EGCUT CTS cases (in pink) have a mean age of 50 and are approximately normally distributed about this mean. In contrast, the mean age of the UK Biobank cases was 69 years.

In terms of case ascertainment in the UK Biobank GWAS, several steps were taken to ensure the highest quality of phenotyping possible. A validation study of CTS diagnoses had been performed previously¹⁵⁵, which involved scrutinising hundreds of medical records of putative

CTS cases. This validation study determined that the positive predictive value of a CTS ‘case’ being a true case is 94%¹⁵⁵. Furthermore, by including surgical codes in the case ascertainment, it meant that 94% of the UK Biobank CTS cases in this association study had at least one operation code, placing this cohort on the phenotypically more severe end of the CTS disease spectrum. Finally, by relying on hospital diagnostic codes, 80% of the UK Biobank CTS cases had at least one ICD-10 or OPCS code for CTS, meaning that the reliance on self-reported questionnaire diagnoses was very low. Very few GWAS that draw their cases from large population-based cohorts such as UK Biobank have gone to these lengths to ensure the validity of their phenotyping methodology.

EGCUT, in contrast, were unable to supply information with regard to how their CTS cohort was phenotyped, beyond stating that a mixture of physician-reported and self-reported questionnaire answers were used. The critical importance of robust phenotyping in genetic association studies cannot be overstated (poor quality input inevitably leads to poor quality output), and it is questionable to what degree such steps were undertaken.

However, the most critical reason for the relative lack of replication of the 16 SNPs in the EGCUT cohort the lack of statistical power. Despite having >4,000 cases, this still only represents approximately one-third of the cases included in the UK Biobank discovery cohort. With controls included, the EGCUT cohort is only 12% of the size of the UK Biobank cohort. Based on the number of Estonian cases and controls and the allelic odds ratios from the UK Biobank GWAS, it is possible to calculate the power to detect an association¹⁷¹ at the Bonferroni-corrected significance threshold of $p=0.0031$, using allelic effect estimates from the UK Biobank GWAS and assuming a population disease prevalence of 5%. It is apparent from **Table 2.5** that this replication study was significantly underpowered at the majority of loci;

only at 4/16 loci was there >80% power to detect an association at this threshold, and that is assuming that the majority of the 4,087 cases are true cases of CTS.

Table 2.5. Power calculations in EGCUT. The power to detect an association at a Bonferroni-corrected significance threshold at each of the 16 index SNPs is shown.

SNP	Power (4087 cases)
rs12406439	0.61
rs12104955	0.51
rs3791679	0.88
rs1025128	0.63
rs847139	0.56
rs1863190	0.92
rs4678145	0.39
rs6843953	0.67
rs3828889	0.68
rs62422907	0.50
rs55841377	0.39
rs6977081	0.63
rs72725608	0.45
rs1866745	0.73
rs72755233	0.99
rs62621197	0.97

In fact, it has been found that the effect estimates for associated SNPs in a GWAS are prone to inflation – the so-called winner’s curse bias¹⁷². This quirk of statistics arises from the fact that when multiple statistical tests are performed, as in a GWAS, SNPs that have a larger than average effect size are more likely to be found ‘significant’ than SNPs with lower effect size¹⁷³. This phenomenon is a bias ingrained in essentially all GWA studies, which use the same population for both gene discovery and effect size estimation¹⁷⁴. Thus, if we factor in winner’s curse, the power estimates become even more diminished.

In summary, while the LD score correlation of 90% between the two GWAS provides strong evidence for shared genetic architecture for CTS between the two cohorts, the replication results were not altogether convincing. There are some fundamental limitations to this replication experiment, pertaining to case ascertainment, discrepancies in age and sex distribution and disease prevalence, and most importantly, lack of statistical power. It is for this reason that a formal meta-analysis of UK Biobank and EGCUT was not undertaken, notwithstanding the fact that the association p-values would actually have become more significant at 13/16 loci.

Unlike well-studied and well-funded disease phenotypes such as diabetes, cardiovascular disease and cancer, there are no large international consortia investigating CTS or entrapment neuropathies that are able to provide bountiful resources for replication. While the possibility of false positive associations from a single-stage GWAS in UK Biobank must be acknowledged, this deficiency is mitigated by several factors, including the robustness of the phenotyping methodology, the low reliance on self-reported diagnostic codes, and the large sample size of approximately 400,000 – a number in significant excess of most published two-stage GWAS.

Consider this thought experiment: one hypothetical solution to the lack of a suitable replication cohort could be to artificially split the UK Biobank samples into a discovery and replication cohort. A discovery GWAS could be undertaken in the interim UK Biobank genotype data release of c. 150,000 samples, and all nominally significant hits ($p < 1 \times 10^{-5}$) could be taken through to a replication GWAS in the remaining c. 350,000 samples, followed by a meta-analysis. However, there are significant disadvantages to this approach. Skol et al¹⁷¹ elegantly demonstrated that undertaking the standard strategy of a two-stage GWAS such as the one described results in significantly lower statistical power for detecting associations compared to the alternative strategy of jointly analysing the data from the two stages.

Although this discussion is hypothetical, it does illustrate certain shortcomings of the traditional two-stages GWAS design, which was motivated by the financial constraints imposed by the prohibitively expensive cost of genotyping in the early days of GWAS. In the two-stage strategy, the discovery cohort is whole-genome genotyped, whereas the replication cohort (which almost without exception is larger than the discovery cohort) need only be genotyped at the nominally significant SNPs discovered in the first stage. Despite the fact that genotyping costs are now significantly lower, the concept of the two-stage GWAS has become somewhat of a dogma in the genetics community^{§§}, even though its *raison d'être* has little relevance in the era of publicly available UK Biobank-scale datasets. A comprehensive review article by Kraft et al.¹⁴⁷ from 2009 on the subject of GWAS replication made the following (rather prescient) observation:

“We should note that the primary motivation of multi-stage designs is not to increase power by testing fewer hypotheses in the second stage samples, and hence paying a smaller penalty for multiple testing at the second stage. Rather the primary goal of multi-stage designs is to save genotyping costs or maximize power given a fixed genotyping budget. If genotyping costs were not an issue, then the multistage approach is less powerful than simply testing all markers in the entire available sample. As genotyping costs decrease, and as more samples have been genotyped as part of previous GWA analyses, single-stage analyses become more common”.

The chapters that follow provide both internal and external validation for the reliability of the results generated from this single-stage GWAS.

^{§§} Much like the p-value threshold of 5×10^{-8} .

Chapter 3

Interrogating the genetic architecture of carpal tunnel syndrome

3.1 Introduction

3.1.1 Background

GWAS has been successful in linking thousands of genetic variants to a spectrum of disease phenotypes, but as discussed in Chapter 1, an enormous challenge remains in linking those risk loci to a functional understanding of the biology underlying disease risk¹⁷⁵. Consequently, there is a considerable mismatch between the genomic information generated in GWAS and the biological knowledge revealed¹⁷⁶. The first hurdle encountered post-GWAS is the difficulty inherent in resolving the true causal variant(s) at a locus, out of a multitude of SNPs in high LD with each other that have statistically indistinguishable associations with the phenotype. The next hurdle lies in untangling how the disease-associated variants (the vast majority of which lie in non-coding regions of the genome) alter the regulation of one or more target genes through an array of mechanisms, including regulation of transcription, splicing or histone modifications. Experimental methods to detect functional genomic regions such as ChIP-seq, DNase-Seq, and ATAC-seq, are invariably labour-intensive and costly¹⁷⁵.

Expression quantitative trait loci (eQTL) are SNPs that modulate gene expression, and these have been shown to be significantly enriched for GWAS associations¹⁷⁷. Several large-scale collaborative projects have attempted to identify the functional elements in the human genome, such as the ENCODE (Encyclopedia of DNA Elements)¹⁷⁸ Project, RegulomeDB¹⁷⁹, and the GTEx (Genotype tissue expression) Project¹⁸⁰. The aim of the latter is to characterise the regulatory sequence variants across 53 different human tissues, obtained from post-mortem sections. Such large, publicly available datasets have greatly facilitated the functional annotation of GWAS associations.

An array of bioinformatic tools have emerged that integrate the functional data from such repositories with other useful information, such as LD patterns from population reference panels, results from previously published GWAS, and knowledge about biological pathways and gene ontologies. Such tools can greatly aid in the interpretation of GWAS, and many have the considerable advantage of needing only the summary association statistics from a GWAS to implement. There has been an increasing trend in recent years to make such summary association statistics publicly available, which mitigates logistic and privacy concerns by precluding the need for access to individual-level data.

3.1.2 Aims

The previous chapter concerned the execution of the CTS GWAS, and in broad terms, the focus of this chapter will lie in its interpretation. To this end, the aim here is to undertake a multitude of *in silico* analyses using the GWAS association summary statistics, in order to: (1) identify candidate causal SNPs and genes, (2) generate testable functional hypotheses, and (3) obtain insights into CTS pathophysiology at the biological pathway level.

3.2 Methods

3.2.1 Functional annotation of associated loci

Firstly, the SNPs at the 16 loci were scrutinised for functionality, looking in particular for coding variants and SNPs in functional regions. To facilitate this process, the summary statistics from the CTS GWAS were analysed using FUMA¹⁸¹ v1.3.3. Independent SNPs were selected on the basis of genome-wide significance and being independent from each other ($r^2 < 0.3$) within a 1 Mb window. The reference population panel selected was the UK Biobank release 2 White British population, corresponding to the samples used in the GWAS. Given the high degree of linkage disequilibrium (LD) in the extended major histocompatibility (MHC) locus on chromosome 6, this region was excluded from the FUMA analysis. FUMA provides annotations on potential functionality of SNPs, including ANNOVAR¹⁸² categories, chromatin states, combined annotation dependent depletion (CADD¹⁸³) scores, and RegulomeDB¹⁷⁹ scores. Positional gene mapping was also performed in FUMA, which is used to map genes to the genome-wide significant SNPs, based on physical position in the genome and potential functionality of the relevant SNPs.

3.2.2 Gene-based analysis and gene-property analysis

A gene-based analysis was implemented in MAGMA¹⁸⁴ v1.06 – a tool for gene analysis that uses information about SNPs located within 18,633 protein-coding genes to derive a p-value for association with the phenotype of interest. This creates a list of genes that are significantly associated with the phenotype, with a Bonferroni correction applied to control for multiple testing ($p < 2.68 \times 10^{-6}$). A gene-property analysis was also performed in MAGMA, which determines if tissue-specific differential expression levels are predictive of the association of a

gene with CTS. This analysis was performed across 53 different tissues taken from the GTEx¹⁸⁰ v6 database.

3.2.3 Gene-set enrichment analyses

A gene-set enrichment analysis was performed in order to gain insight into the biological pathways relevant to the prioritised genes. This was undertaken using the GENE2FUNC tool in FUMA, using (1) the genes that were prioritised with FUMA positional mapping, and (2) the genes that met genome-wide significance in the MAGMA gene-based analysis. The parameters applied for this analysis were: multiple testing correction using the Benjamini-Hochberg false discovery rate (FDR) method (with alpha set at 0.05), gene expression data taken from GTEx v7 RNA-Seq data, minimum number of overlapped genes per ontology or pathway = 5. The genes were examined for overrepresentation in different biological ontologies and gene sets. A similar gene-set enrichment analysis was also performed using XGR¹⁷⁶, a bioinformatic tool which incorporates ontology, gene annotations and knowledge about biological networks to aid in the interpretation of genomic summary data. The FUMA-mapped genes were analysed using XGR's gene-based enrichment analysis (selecting default parameters and the MSigDB 'canonical pathways' ontology), to identify enriched ontology terms within the CTS GWAS summary statistics.

3.2.4 Summary-based Mendelian randomisation

Summary-based Mendelian randomisation (SMR)¹⁸⁵ was performed to test for pleiotropic associations between expression levels of genes and the CTS phenotype. This method relies on the principle of Mendelian randomisation (which will be discussed in detail in Chapter 4): if the expression level of a gene is influenced by an eQTL variant, then individuals who carry

different genotypes at the variant will demonstrate different levels of gene expression¹⁸⁶. Given that genotype at the eQTL for any individual is fixed at birth, it can be used as an instrumental variable to quantify the magnitude of the “exposure” (i.e. expression levels of the gene linked to the eQTL), and the CTS summary statistics can be used to measure the relationship between the instrumental variable and the “outcome” (i.e. CTS status). By regressing the genotype-CTS relationship on the genotype-gene expression relationship, it is possible to calculate whether the gene in question is likely to have a causal role in the development of CTS.

The eQTL data from transformed fibroblasts (from GTEx v7) was selected for this analysis on the basis that fibroblasts are the cell type with the most relevance to putative mechanisms of CTS aetiology in the sub-synovial connective tissues. Associations were tested between CTS status and the expression level of each gene (using the top-associated eQTL as an instrumental variable). To correct for multiple testing across 4,323 genes, a Bonferroni correction was applied, with the threshold for statistical significance defined as $P_{\text{SMR}} < 2.68 \times 10^{-6}$. In conjunction with the SMR analysis, a HEIDI¹⁸⁵ (heterogeneity in dependent instruments) analysis was performed to test for heterogeneity in SMR estimates at SNPs in LD with the top-associated eQTL (which allows pleiotropy to be distinguished from linkage).

3.2.5 Partitioned heritability analyses

Stratified LDSC regression was undertaken in order to perform partitioned heritability analysis of CTS, with the aim of determining whether SNPs that account for the heritability of CTS cluster in functional genomic regions. Twenty-four functional categories in the ‘full baseline model’ were selected for this analysis – these categories are derived from publicly available annotations that are not limited to a particular cell type¹⁸⁷. Stratified LD scores were calculated using 1000 Genomes European-ancestry sample data, which includes only SNPs in HapMap3

with $MAF > 0.05$. The enrichment metric is calculated by dividing the proportion of heritability captured by a type of functional annotation by the proportion of SNPs within that category – it therefore provides a metric to reflect whether a particular functional annotation is over- or under-represented within a set of GWAS summary statistics.

LDSC-SEG¹⁸⁸ (LD score regression applied to expressed genes) was also performed, with the aim of identifying tissues and cell types relevant to CTS. Publicly available RNA-Seq data from GTEx and the Franke Lab¹⁸⁹ have been divided into 205 different cell and tissue types, and these were classified into nine distinct categories of related phenotypes, as per Finucane et al¹⁸⁸, to look for overrepresentation of any of the categories within CTS.

3.2.6 Expression of candidate genes in fibroblasts vs Schwann cells

Gene expression levels for the top three candidate genes emerging from the above analyses were compared in two cell types with biological relevance to CTS: fibroblasts versus Schwann cells. These RNA-Seq data of cultured human peripheral Schwann cells and fibroblasts from Weiss et al.¹⁹⁰ are publicly available, and were downloaded from GEO (GSE90711). STAR aligner¹⁹¹ was used to map the reads to the GRCh38.88 human genome build, and counts were assigned to ENSEMBL gene annotations using HTSeq¹⁹². DESeq2¹⁹³ was used for differential expression analysis; counts were normalised and transformed using the regularised log2 transformation $\pm$ the standard error of the mean.

3.2.7 LD score correlation

LDSC^{165,194} was used to compute genetic correlations (r_g) between CTS and phenotypes that are known to be associated with an increased risk of CTS: anthropometric traits, diabetes,

rheumatoid arthritis, obesity and elevated urate (a proxy phenotype for gout). This analysis was implemented in LD Hub¹⁹⁵. Publicly available GWAS meta-analysis results for the relevant phenotype group were selected: “Anthropometric traits”, “Autoimmune diseases”, “Glycaemic traits”, “Uric acid”, and “Bone mineral density” (as a skeletal phenotype). All but “rheumatoid arthritis” were removed from the “Autoimmune diseases” trait group, and “body fat” was removed from the “Anthropometric traits” group, as this phenotype contained data from GWAS meta-analyses that included individuals of non-European ancestry. A total of 36 traits across these groups were therefore selected, and a conservative Bonferroni-corrected p-value threshold of 0.0014 (0.05/36) was used to define a correlation as statistically significant. In order to implement the analysis, the CTS GWAS summary statistics were pruned to harmonise them with the list of SNPs provided on LD Hub (with MHC region SNPs removed), leaving a total of 1,171,431 SNPs.

3.2.8 Height and BMI in UK Biobank

Height and body mass index (BMI) are two anthropometric measures that are available for the vast majority of UK Biobank participants. Given the probable relevance of these variables to CTS aetiology, mean height and BMI were calculated for cases and controls. T-tests were performed to ascertain differences in the mean between CTS cases and controls.

3.2.9 Polygenic risk score (PRS)

A polygenic risk score (PRS) was calculated using the weighted genetic risk score method, as described by de Jager et al¹⁹⁶. This method takes into account the discrepancies in the allelic odds ratio amongst individual SNPs, by weighting the risk allele dosage at each SNP. A PRS for

CTS was calculated for each individual in UK Biobank using the sixteen genome-wide significant index SNPs, using the following formula:

$$wGRS = \sum_{i=1}^n W_i X_i$$

where i is the index SNP, n is the number of index SNPs (16 in this GWAS), W_i is the weight for the SNP (the natural logarithm of the allelic odds ratio), and X_i is the allelic dosage (the number of effect alleles – i.e. 0, 1 or 2). The allelic dosage was calculated using QCTOOL v2. The PRS was calculated in R, and t-tests performed to ascertain whether there were significant differences in mean PRS between sub-groups within the UK Biobank cohort. PRS was calculated for: (1) All CTS cases, (2) all controls, (3) CTS cases with at least one operation code (either OPCS or self-reported), and (4) CTS cases with no operation codes. The whole UK Biobank cohort was also divided into quintiles based on PRS, and the prevalence of CTS determined for each quintile.

3.3 Results

3.3.1 Evidence of functional SNPs at the associated loci

In terms of association p-values, the two most significantly associated SNPs in the GWAS were missense variants. rs72755233 on chromosome 15 (OR = 1.18, $p = 2.3 \times 10^{-15}$) results in a threonine-to-isoleucine substitution in *ADAMTS17*, while rs62621197 on chromosome 19 (OR = 1.31, $p = 7.5 \times 10^{-14}$) results in an arginine-to-glutamine substitution in *ADAMTS10*. Within the non-coding variants that were significantly associated with CTS, there was strong evidence of functionality: 30/422 genome-wide significant SNPs had a Combined Annotation-Dependent Depletion (CADD)¹⁸³ score of greater than 12.37, which is the threshold above which SNPs have a high probability of being deleterious¹⁹⁷. These 30 SNPs included the two missense variants above, and also rs3791679 (OR = 1.11, $p = 2.0 \times 10^{-12}$), the fourth most significantly associated SNP in the GWAS, which resides within an intronic enhancer region of *EFEMP1*¹⁹⁸. A further 3 out of 422 SNPs had a RegulomeDB score of 2b or more – suggesting they are likely to affect protein binding. The SNPs with likely functional consequences are summarised in **Table 3.1**. The top three candidate genes on the basis of functional annotation were *ADAMTS17*, *ADAMTS10* and *EFEMP1*.

3.3.2 Gene mapping

Table 3.2 shows the genes that were mapped to the GWAS summary statistics based on the gene-based association analysis in MAGMA. Genes mapped by FUMA positional mapping (based on ANNOVAR annotations and physical position in the genome) were shown in Chapter 2, **Table 2.2**. The two analyses mapped 17 and 25 genes, respectively, with an overlap of 7 genes (**Fig. 3.1**).

Table 3.1 Genome-wide significant SNPs with likely functional consequences. SNPs listed either have a CADD score > 12.37 or a RegulomeDB score of 2b or more. CADD > 12.37 SNPs are ranked by descending score. RegulomeDB ≥ 2b SNPs are ranked by p-value.

Genome-wide significant SNPs with CADD score > 12.37											
SNP	Chr	Position	Allele	MAF	P-value	Beta	SE	Nearest gene	Distance from nearest gene	Position	CADD score
rs5002640	2	218146239	G	0.2553	1.20×10 ⁻¹²	-0.0031	0.0004	AC009492.1	1222	intergenic	22.3
rs17315753	2	176784435	A	0.2132	2.70×10 ⁻¹⁰	0.0030	0.0005	KIAA1715	4184	intergenic	21.2
rs62621197	19	8670147	C	0.03195	7.50×10 ⁻¹⁴	-0.0079	0.0011	ADAMTS10	0	exonic	21
rs72755233	15	100692953	G	0.1189	2.30×10 ⁻¹⁵	-0.0048	0.0006	ADAMTS17	0	exonic	20.6
rs17181956	2	218146080	C	0.1049	1.20×10 ⁻¹¹	0.0042	0.0006	AC009492.1	1381	intergenic	19.94
rs56874501	2	176888041	A	0.2037	1.00×10 ⁻⁹	0.0029	0.0005	KIAA1715	20473	intergenic	19.62
rs6752931	2	56004219	G	0.2132	3.60×10 ⁻¹⁰	-0.0030	0.0005	PNPT1	83173	intergenic	18.85
rs72725608	14	76245906	T	0.0442	1.10×10 ⁻⁸	-0.0054	0.0009	TTLL5	0	intronic	18.72
rs6857769	4	13252148	G	0.1261	1.40×10 ⁻¹⁰	-0.0037	0.0006	HSP90AB2P	82889	intergenic	18.51
rs11891328	2	176889698	G	0.211	8.40×10 ⁻¹¹	0.0030	0.0005	KIAA1715	22130	intergenic	17.02
rs73227609	4	13204806	G	0.1265	6.10×10 ⁻¹¹	-0.0037	0.0006	HSP90AB2P	130231	intergenic	16.39
rs1025128	2	60175475	G	0.4358	2.80×10 ⁻⁹	-0.0023	0.0004	RP11-444A22.1	0	ncRNA_intronic	15.86
rs847181	2	176918910	G	0.2888	3.60×10 ⁻⁸	0.0023	0.0004	EVX2	23289	intergenic	15.26
rs4146922	2	56067182	T	0.2073	1.00×10 ⁻¹⁰	-0.0031	0.0005	EFEMP1	25919	intergenic	15.26
rs62422913	6	85733547	T	0.1095	1.50×10 ⁻⁸	0.0035	0.0006	RP3-435K13.1	57203	intergenic	15.2
rs3791679	2	56096892	A	0.2296	2.00×10 ⁻¹²	-0.0032	0.0005	EFEMP1	0	intronic	14.94
rs58680090	2	56080379	T	0.2093	7.60×10 ⁻¹⁰	-0.0029	0.0005	EFEMP1	12722	intergenic	14.74
rs11636457	15	67021996	G	0.3414	2.20×10 ⁻⁹	-0.0024	0.0004	SMAD6	0	intronic	14.6
rs76850558	6	85715743	C	0.1146	2.20×10 ⁻⁹	0.0036	0.0006	RP3-435K13.1	39399	intergenic	14.56
rs1099941	2	176907320	C	0.2897	3.90×10 ⁻⁸	0.0023	0.0004	EVX2	34879	intergenic	14.38
rs10205394	2	218150948	G	0.1995	1.10×10 ⁻¹¹	0.0033	0.0005	AC009492.1:DIRC3	00:00	ncRNA_intronic	14.34

rs1344732	2	56029071	C	0.2571	7.30×10 ⁻⁹	-0.0026	0.0004	<i>EFEMP1</i>	64030	intergenic	14.09
rs4674140	2	218143061	A	0.144	5.60×10 ⁻¹²	0.0038	0.0005	<i>AC009492.1</i>	4400	intergenic	13.95
rs62422907	6	85715955	G	0.1146	2.20×10 ⁻⁹	0.0036	0.0006	<i>RP3-435K13.1</i>	39611	intergenic	13.94
rs55841377	7	44145178	C	0.2246	8.40×10 ⁻⁹	0.0026	0.0005	<i>AEBP1</i>	0	intronic	13.83
rs73978124	2	176818825	G	0.2137	2.20×10 ⁻¹⁰	0.0030	0.0005	<i>KIAA1715</i>	0	intronic	13.38
rs10804254	2	218058184	T	0.3255	1.80×10 ⁻⁸	0.0023	0.0004	<i>AC010887.1</i>	16878	intergenic	13.31
rs7563032	2	56068441	C	0.2072	1.10×10 ⁻¹⁰	-0.0031	0.0005	<i>EFEMP1</i>	24660	intergenic	13.1
rs62175240	2	218134786	G	0.1435	3.30×10 ⁻¹²	0.0038	0.0005	<i>AC009492.1</i>	12675	intergenic	13.06
rs11676136	2	218140435	T	0.1441	5.70×10 ⁻¹²	0.0038	0.0005	<i>AC009492.1</i>	7026	intergenic	12.46
Genome-wide significant SNPs with RegulomeDB score of 2b or more											
SNP	Chr	Position	Allele	MAF	P-value	Beta	SE	Nearest gene	Distance from nearest gene	Position	RegulomeDB score
rs35283599	15	67036674	T	0.3458	2.50×10 ⁻⁹	-0.0024	0.0004	SMAD6	0	intronic	2b
rs2194747	2	218126181	A	0.2899	5.40×10 ⁻⁹	-0.0025	0.0004	RN7SKP43	9582	intergenic	2b
rs6739641	2	56020962	C	0.2576	1.60×10 ⁻⁸	-0.0025	0.0004	EFEMP1	72139	intergenic	2b

The gene-property analysis performed in MAGMA across the 53 different GTEx tissue types demonstrated statistically significant enrichments in two tissues: tibial artery and coronary artery. The next highest (albeit non-significant) enrichments were in: tibial nerve, aorta, and transformed fibroblasts (**Figure 3.2**).

Table 3.2 Genes mapped to CTS in the MAGMA gene-based association test. Genes are ranked by descending association p-value.

Gene	Chr	MAGMA P-value
<i>LOXL4</i>	10	9.47×10^{-9}
<i>R3HCC1L</i>	10	7.08×10^{-8}
<i>COL11A1</i>	1	1.07×10^{-7}
<i>SMAD6</i>	15	1.59×10^{-7}
<i>PTPN9</i>	15	1.66×10^{-7}
<i>HAS1</i>	19	1.87×10^{-7}
<i>ZBTB38</i>	3	2.13×10^{-7}
<i>AFF1</i>	4	3.80×10^{-7}
<i>ADAMTS17</i>	15	4.88×10^{-7}
<i>PBX3</i>	9	4.90×10^{-7}
<i>KIAA1715</i>	2	5.83×10^{-7}
<i>XRCC6</i>	22	7.30×10^{-7}
<i>AOC1</i>	7	1.10×10^{-6}
<i>CCDC134</i>	22	1.69×10^{-6}
<i>AEBP1</i>	7	2.32×10^{-6}
<i>EFEMP1</i>	2	2.40×10^{-6}
<i>SIN3A</i>	15	2.55×10^{-6}

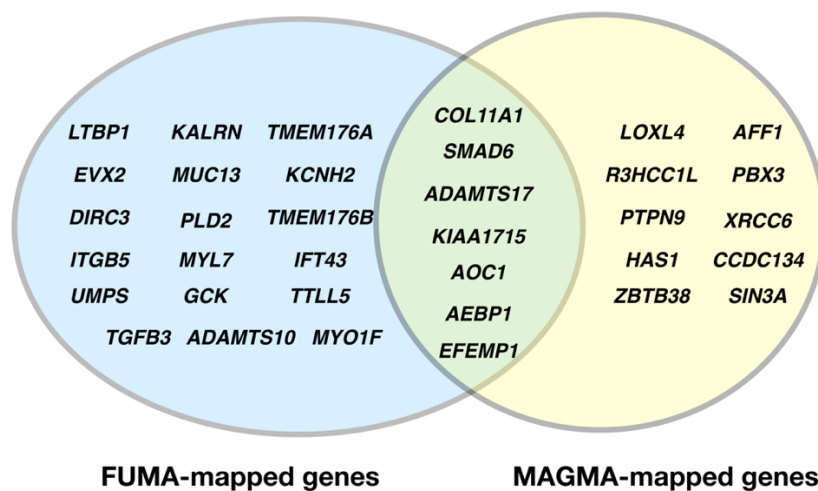


Fig 3.1 MAGMA and FUMA-mapped genes. The seven overlapped genes are shown in the intersection of this Venn diagram.

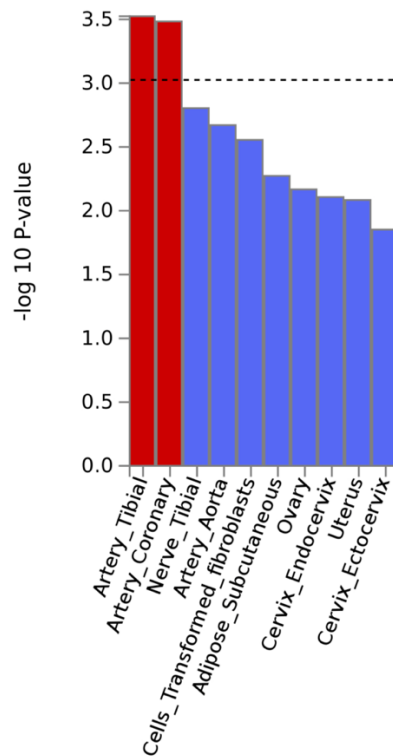


Fig 3.2 Gene-property analysis in MAGMA. This tests the relationship between highly expressed genes in a specific tissue and the genetic associations from the GWAS. Gene-property analysis is performed using average expression of genes per tissue type as a gene covariate. The dotted line indicates the Bonferroni-corrected p-value threshold, and the tissues that meet this significance threshold are highlighted in red. The top ten enriched tissues are shown here.

3.3.3 Enrichment for extracellular matrix (ECM) ontologies

There was a strong enrichment for ECM and extracellular space-related ontologies, respectively, on performing gene-set analysis on the 25 FUMA-mapped genes (adjusted $p = 2.7 \times 10^{-8}$) and the 17 MAGMA-mapped genes (adjusted $p = 1.6 \times 10^{-3}$). There was also a significant enrichment for GWAS catalog-reported genes for waist circumference and height (**Table 3.3**). The gene-based enrichment analysis performed in XGR on the 25 FUMA-mapped genes corroborated this strong enrichment for ECM-related genes, with the top enrichment seen in the ontology called “genes encoding core extracellular matrix including ECM glycoproteins, collagens and proteoglycans” (Z-score = 3.24, $p=0.0016$) (**Table 3.4**).

Table 3.3 Gene-based enrichment analysis in FUMA. This analysis was performed using the GENE2FUNC tool in FUMA, using the 25 positionally mapped genes in FUMA and the 17 genes from the MAGMA gene-based analysis. N = number of genes in the ontology; n = number of overlapped genes.

25 genes from FUMA positional mapping					
GO Cellular components					
GeneSet	N	n	P-value	Adjusted P-value	Genes
GO_EXTRACELLULAR_MATRIX	424	7	4.67×10^{-11}	2.71×10^{-8}	<i>COL11A1, TGFB3, ADAMTS17, ADAMTS10, LTBP1, EFEMP1, AEBP1</i>
GO_PROTEINACEOUS_EXTRACELLULAR_MATRIX	355	5	3.50×10^{-8}	6.77×10^{-6}	<i>COL11A1, ADAMTS17, ADAMTS10, LTBP1, EFEMP1</i>
GO_EXTRACELLULAR_SPACE	1367	5	8.30×10^{-5}	4.81×10^{-3}	<i>TGFB3, EFEMP1, MUC13, AEBP1, AOC1</i>
GO_CYTOSKELETON	1965	5	5.95×10^{-4}	2.30×10^{-2}	<i>TTLL5, MYO1F, KALRN, MYL7, GCK</i>
GWAS Catalog Reported Genes					
Gene Set	N	n	P-value	Adjusted P-value	Genes
Waist circumference adjusted for body mass index	126	5	7.01×10^{-11}	8.57×10^{-8}	<i>ADAMTS17, MYO1F, ADAMTS10, LTBP1, EFEMP1</i>
Height	522	7	2.42×10^{-10}	1.48×10^{-7}	<i>TTLL5, ADAMTS17, MYO1F, ADAMTS10, LTBP1, EFEMP1, TMEM176A</i>
17 genes from MAGMA gene-based analysis					
GO Cellular components					
Gene set	N	n	P-value	Adjusted P-value	Genes
GO_EXTRACELLULAR_SPACE	1367	5	2.93×10^{-5}	1.55×10^{-3}	<i>LOXL4, EFEMP1, ZBTB38, AEBP1, AOC1</i>
GWAS Catalog Reported Genes					
Gene set	N	n	P-value	Adjusted P-value	Genes
Height	522	5	1.12×10^{-7}	4.54×10^{-5}	<i>SIN3A, PTPN9, ADAMTS17, EFEMP1, ZBTB38</i>

Table 3.4 Gene-based enrichment analysis in XGR. XGR Gene-based Enrichment Analysis function to classify the 25 genes mapped by FUMA into different functional categories. The Z-scores, p-values, FDR, and the overlapped genes for each of the ontologies are shown.

Term Name	Z-Score	P-value	FDR	Number of overlapped genes	Genes
Ensemble of genes encoding core extracellular matrix including ECM glycoproteins, collagens and proteoglycans	3.24	0.0016	0.0032	4	<i>AEBP1, COL11A1, EFEMP1, LTBP1</i>
Ensemble of genes encoding extracellular matrix and extracellular matrix-associated proteins	2.79	0.0016	0.0032	8	<i>ADAMTS10, ADAMTS17, AEBP1, COL11A1, EFEMP1, LTBP1, MUC13, TGFB3</i>
Genes encoding structural ECM glycoproteins	2.89	0.0035	0.0047	3	<i>AEBP1, EFEMP1, LTBP1</i>
Ensemble of genes encoding ECM-associated proteins including ECM-affiliated proteins, ECM regulators and secreted factors	0.96	0.1	0.1	4	<i>ADAMTS10, ADAMTS17, MUC13, TGFB3</i>

3.3.4 Summary-based Mendelian randomisation in fibroblasts implicates a key gene in the TGF- β signalling pathway

The SMR analysis using eQTL data for fibroblasts implicated two genes that met the conservative SMR p-value significance threshold: *LTBP1* on chromosome 2p22.3, and *MAN2C1* at 15q24.2, suggesting that there is an association between the expression levels of these genes and CTS status (**Table 3.5**). HEIDI analysis excluded SMR associations due to linkage (i.e. the scenario whereby one causal variant affects gene expression whereas another variant in close LD independently affects CTS risk), as both genes passed the HEIDI test with $p > 0.025$ ($0.05/\text{number of } P_{\text{SMR}}\text{-significant genes}$). *LTBP1* is key component of the TGF- β signalling pathway, and was a FUMA-mapped gene at one of the genome-wide significant loci on chromosome 2. *MAN2C1* (which was implicated neither in FUMA positional mapping nor in the MAGMA gene-based test) encodes Mannosidase Alpha Class 2C Member 1, which belongs to the α -mannosidase family and is believed to be involved in glycoprotein maturation¹⁹⁹, and has been linked to prostate cancer tumorigenesis and metastasis in nasopharyngeal carcinoma²⁰⁰.

Table 3.5 Summary-based Mendelian Randomisation (SMR). The analysis was performed using eQTL data from GTEx v7 transformed fibroblasts.

Chr	Gene	Top SNP	P_{GWAS}	P_{eQTL}	P_{SMR}	P_{HEIDI}
2	<i>LTBP1</i>	rs7571401	8.10×10^{-7}	2.30×10^{-47}	3.03×10^{-6}	0.530
15	<i>MAN2C1</i>	rs10220738	2.20×10^{-6}	6.48×10^{-63}	5.25×10^{-6}	0.968

3.3.5 Partitioned heritability analysis shows enrichment for regulatory regions, musculoskeletal and connective tissues

The partitioned heritability analysis across 24 functional genomic categories demonstrated the greatest enrichment for evolutionarily conserved genomic regions across 29 mammalian species (10.8-fold enrichment, $p = 0.0013$), suggesting that a significant portion of the variants associated with CTS are under negative selection pressure²⁰¹. There was also statistically significant enrichment for several histone markers, strongly implying regulatory activity (Fig. 3.3).

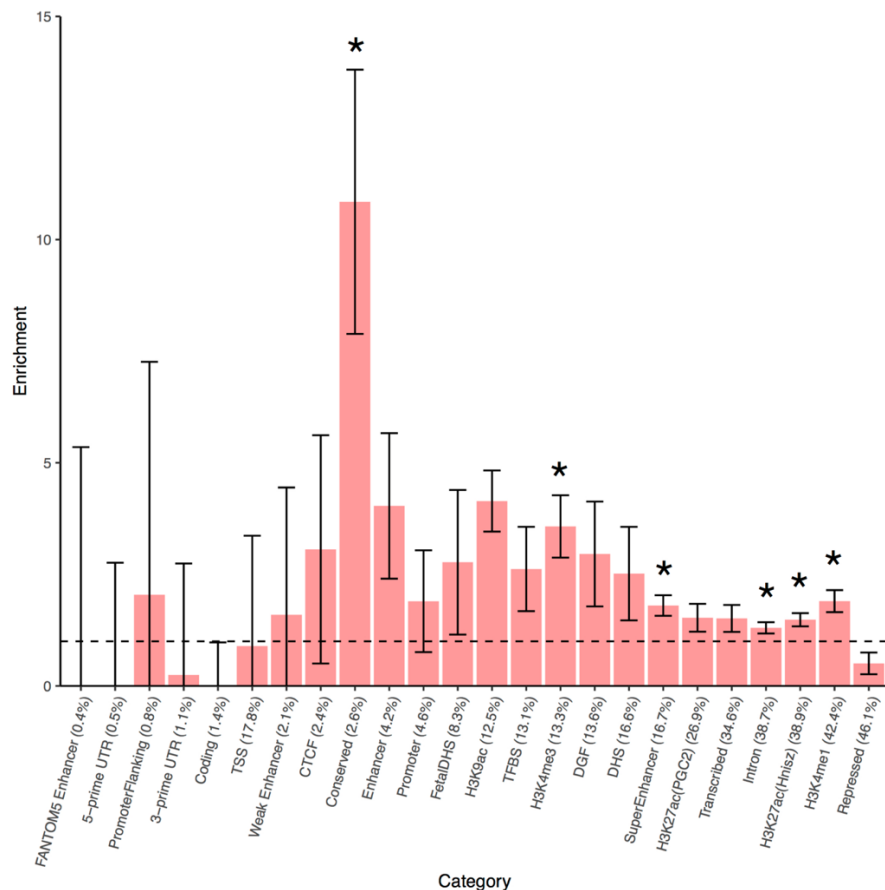


Fig.3.3 Partitioned CTS heritability across genomic functional categories. Enrichment estimates across the 24 main functional annotation categories for CTS heritability. The enrichment statistic is the proportion of heritability found in each functional group divided by the proportion of SNPs in each group. The dashed line indicates an enrichment proportion of 1 (i.e. no enrichment). Error bars represent jackknife standard errors around the enrichment estimates, and an asterisk indicates significance at $p < 0.05$ after Bonferroni correction for the 24 annotations tested.

The LD-SEG analysis identified various tissues and cell types that are enriched for CTS heritability (**Table 3.6**), with positive regression coefficients indicating a contribution of these cell and tissue types to CTS heritability. The top enrichment was seen in osteoblasts ($p=5.4\times 10^{-4}$) and smooth muscle myocytes ($p=2.9\times 10^{-3}$). By subdividing the 205 cell and tissue types tested into the nine phenotypic categories used by Finucane et al¹⁸⁸, there was a disproportionate enrichment for cell and tissue types within the ‘Musculoskeletal/Connective’ tissue subgroup, with eight out of 19 significantly enriched tissue types ($p < 0.05$) within this group.

3.3.6 Top candidate genes are highly expressed in Schwann cells and fibroblasts

The top three candidate genes from this GWAS – *ADAMTS17*, *ADAMTS10* and *EFEMP1* – were all expressed in cultured human peripheral nerve fibroblasts and Schwann cells according to publicly available RNA-Seq from Weiss et al¹⁹⁰. Notably, *ADAMTS17* showed significantly greater expression in Schwann cells than in fibroblasts, with regularised log2 counts more than twice as high in Schwann cells compared to fibroblasts. *EFEMP1* and *ADAMTS10* demonstrated greater expression in fibroblasts than in Schwann cells (**Fig. 3.4**).

Table 3.6 Partitioned CTS heritability across different cell and tissue types. This analysis was implemented using LDSC-SEG across 205 different tissue and cell types from GTEx and Franke Lab gene expression datasets. The cell and tissue types with coefficient $p < 0.05$ are shown (with regression coefficients and standard errors). Additionally, the cell and tissue types have been classified into nine distinct tissue categories, as per Finucane et al¹⁸⁸.

Name	Coefficient	Coefficient SE	Coefficient p-value	Tissue category
A11.329.629.Osteoblasts	2.72×10^{-9}	8.32×10^{-10}	0.0005	Musculoskeletal/Connective
A11.620.520.Myocytes..Smooth.Muscle	2.14×10^{-9}	7.76×10^{-10}	0.0029	Musculoskeletal/Connective
A11.872.190.260.Embryoid.Bodies	2.56×10^{-9}	9.31×10^{-10}	0.0030	Other
A10.165.114.830.750.Subcutaneous.Fat	1.56×10^{-9}	6.24×10^{-10}	0.0063	Adipose
A10.165.450.300.Cicatrix	1.55×10^{-9}	6.27×10^{-10}	0.0068	Musculoskeletal/Connective
Artery_Tibial	1.42×10^{-9}	6.11×10^{-10}	0.0101	Cardiovascular
A11.329.171.Chondrocytes	1.94×10^{-9}	8.39×10^{-10}	0.0105	Musculoskeletal/Connective
Artery_Aorta	1.45×10^{-9}	6.38×10^{-10}	0.0116	Cardiovascular
A02.165.Cartilage	1.80×10^{-9}	8.19×10^{-10}	0.0139	Musculoskeletal/Connective
A05.360.319.679.690.Myometrium	1.52×10^{-9}	6.93×10^{-10}	0.0143	Musculoskeletal/Connective
A08.186.211.653.Mesencephalon	1.45×10^{-9}	6.88×10^{-10}	0.0178	CNS
A10.690.467.Muscle..Smooth	1.46×10^{-9}	7.48×10^{-10}	0.0257	Musculoskeletal/Connective
A11.872.580.Mesenchymal.Stem.Cells	1.37×10^{-9}	7.37×10^{-10}	0.0311	Other
A11.118.637.555.567.562.B.Lymphocytes	1.52×10^{-9}	8.26×10^{-10}	0.0327	Blood/Immune
A10.165.450.300.425.Keloid	1.10×10^{-9}	6.08×10^{-10}	0.0349	Musculoskeletal/Connective
Adipose_Subcutaneous	1.14×10^{-9}	6.46×10^{-10}	0.0393	Adipose
A10.165.114.830.500.750.Subcutaneous.Fat..Abdomin	1.12×10^{-9}	6.36×10^{-10}	0.0398	Adipose
A10.615.789.Serous.Membrane	1.42×10^{-9}	8.26×10^{-10}	0.0423	Other
A03.556.500.760.464.Parotid.Gland	1.21×10^{-9}	7.33×10^{-10}	0.0492	Digestive

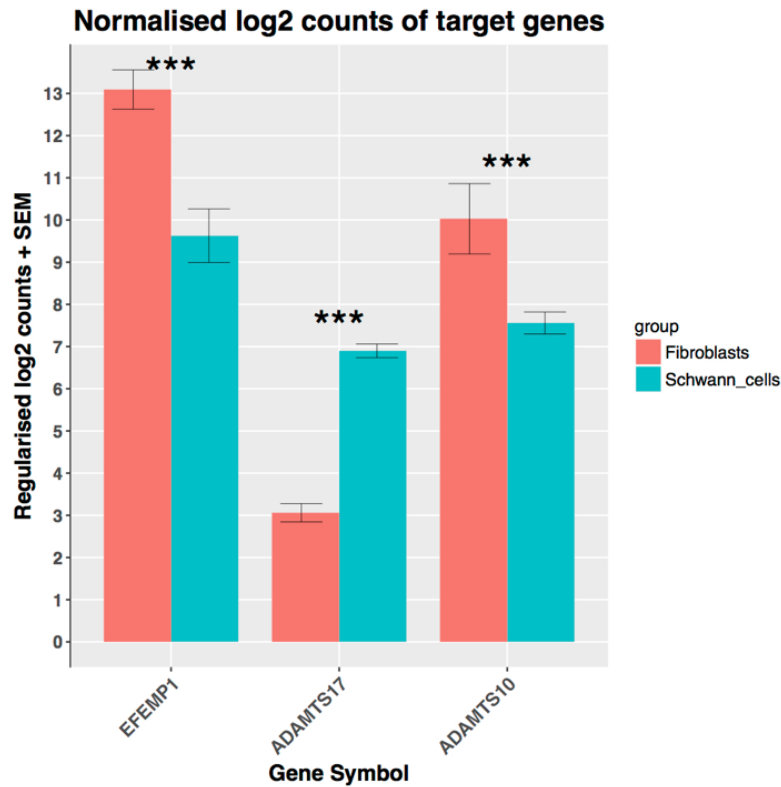


Fig. 3.4 Expression of candidate genes in human fibroblasts and Schwann cells. Data taken from Weiss et al.¹⁹⁰ Error bars represent the standard error of the mean of the regularised log2 counts. P-value was determined using Wald test and was FDR adjusted. ***p < 0.01.

3.3.7 Genetic correlation with anthropometric phenotypes

The LD score correlation analysis demonstrated that several anthropometric phenotypes were significantly correlated with CTS: most notably, body mass index (BMI) and several other measures of adiposity were positively correlated with CTS ($r_g=0.346$, $p=5.8 \times 10^{-23}$ for “BMI”), and there were strong negative correlations between two height phenotypes and CTS ($r_g=-0.217$, $p=3.7 \times 10^{-9}$ for “Height_2010” and $r_g=-0.223$, $p=2.1 \times 10^{-6}$ for “Extreme height”). Lumbar spine bone mineral density, another anthropometric phenotype, had a significant positive correlation with CTS ($r_g=0.183$, $p=0.0002$). In terms of disease phenotypes, rheumatoid arthritis ($r_g=0.113$, $p=9.8 \times 10^{-3}$) and type 2 diabetes ($r_g=0.141$, $p=0.011$) showed nominally significant correlations that did not meet the Bonferroni-corrected

significance threshold for multiple testing. However, HbA1c, a measure of long-term glycaemic control, was significantly correlated with CTS ($r_g=0.282$, $p=7.1\times10^{-5}$). Serum urate, a proxy measure for gout, was non-significant ($r_g=-0.099$, $p=0.067$). The results are summarised in

Table 3.7.

Table 3.7 Genetic correlation between CTS and other phenotypes. The analysis was performed using LD score (LDSC) regression implemented in LD Hub. The traits are shown along with the PMID of the study from which the data were derived, the trait category, and the correlation coefficient (r_g). Traits are ranked by p-value, and the top 14 traits (in bold italic) meet a Bonferroni-corrected significant threshold of $p<0.0014$.

Trait	PMID	Category	r_g	se	p
<i>Body mass index</i>	20935630	anthropometric	0.3464	0.0351	5.83×10^{-23}
<i>Overweight</i>	23563607	anthropometric	0.3613	0.0376	6.92×10^{-22}
<i>Obesity class 1</i>	23563607	anthropometric	0.3268	0.0382	1.09×10^{-17}
<i>Waist-to-hip ratio</i>	25673412	anthropometric	0.2465	0.035	1.83×10^{-12}
<i>Obesity class 2</i>	23563607	anthropometric	0.3176	0.0482	4.44×10^{-11}
<i>Height_2010</i>	20881960	anthropometric	-0.2168	0.0368	3.67×10^{-9}
<i>Waist circumference</i>	25673412	anthropometric	0.208	0.037	1.84×10^{-8}
<i>Extreme bmi</i>	23563607	anthropometric	0.2726	0.0548	6.72×10^{-7}
<i>Extreme height</i>	23563607	anthropometric	-0.2235	0.0472	2.13×10^{-6}
<i>Obesity class 3</i>	23563607	anthropometric	0.2928	0.0681	1.69×10^{-5}
<i>HbA1C</i>	20858683	glycemic	0.2824	0.0711	7.07×10^{-5}
<i>Lumbar spine bone mineral density</i>	22504420	bone	0.1826	0.0483	0.0002
<i>Childhood obesity</i>	22484627	anthropometric	0.2054	0.0575	0.0004
<i>Hip circumference</i>	25673412	anthropometric	0.1235	0.037	0.0009
Lumbar Spine bone mineral density	26367794	bone	0.156	0.0498	0.0017
Child birth weight	23202124	anthropometric	-0.2127	0.0694	0.0022
Fasting glucose main effect	22581228	glycemic	0.1493	0.0522	0.0042
Difference in height between adolescence and adulthood; age 14	23449627	anthropometric	-0.2452	0.0943	0.0093
Rheumatoid Arthritis	24390342	autoimmune	0.1132	0.0439	0.0098
Type 2 Diabetes	22885922	glycemic	0.1407	0.0556	0.0114
Difference in height between childhood and adulthood; age 8	23449627	anthropometric	-0.169	0.0709	0.0172
Birth weight	27680694	anthropometric	-0.0997	0.0422	0.0181
Child birth length	25281659	anthropometric	-0.1402	0.0727	0.0538
Serum Urate overweight	25811787	uric_acid	-0.0988	0.0539	0.0666
Femoral Neck bone mineral density	26367794	bone	0.0966	0.0536	0.0715

Fasting insulin main effect	22581228	glycemic	0.1117	0.0669	0.0951
Height; Females at age 10 and males at age 12	23449627	anthropometric	-0.0849	0.0561	0.1303
Femoral neck bone mineral density	22504420	bone	0.0624	0.0454	0.1693
Fasting proinsulin	20081858	glycemic	0.1263	0.0938	0.1779
Extreme waist-to-hip ratio	23563607	anthropometric	0.1121	0.089	0.2076
Forearm Bone mineral density	26367794	bone	0.171	0.1389	0.2185
Infant head circumference	22504419	anthropometric	-0.0945	0.0799	0.2368
HOMA-IR	20081858	glycemic	0.0869	0.0807	0.2815
Sitting height ratio	25865494	anthropometric	-0.064	0.0609	0.2934
2hr glucose adjusted for BMI	20081857	glycemic	-0.0732	0.1064	0.4915
HOMA-B	20081858	glycemic	0.0051	0.0704	0.9428

3.3.8 Observed anthropometric measures mirror the genetic correlations

Height and BMI for the UK Biobank cohort are shown in **Table 3.8**. On average, both male and female CTS cases were ≥ 2 cm shorter than, and had a mean BMI ≥ 2.1 kg/m² greater than controls.

Table 3.8 Height and BMI in the UK Biobank cohort. Standing height and BMI (plus S.D.) are given for the whole post-QC UK Biobank cohort for whom height and BMI values were available (400,806 for height and 400,386 for BMI, out of 401656). “All” refers to males and females combined, and breakdowns are given by sex and CTS status. Mean differences and t-tests correspond to CTS cases vs controls.

Height (cm)					
	CTS cases and controls	CTS cases	CTS controls	t-test P	Mean height difference (95% CI)
All	168.7 (9.2)	164.7 (8.7)	168.9 (9.2)	< 0.0001	4.2 (4.0-4.3)
N	400,806	12,265	388,541		
Males	175.9 (6.8)	173.8 (6.7)	175.9 (6.7)	< 0.0001	2.1 (1.9-2.4)
N	184,066	3,712	180,354		
Females	162.6 (6.2)	160.7 (6.2)	162.7 (6.2)	< 0.0001	2.0 (1.8-2.0)
N	216,740	8,553	208,187		

BMI (kg/m ²)					
	CTS cases and controls	CTS cases	CTS controls	T-test P	Mean BMI difference (95% CI)
All	27.4 (4.8)	29.4 (5.7)	27.4 (4.7)	< 0.0001	2.0 (1.96-2.13)
N	400,386	12,239	388,147		
Males	27.9 (4.2)	29.9 (4.8)	27.8 (4.2)	< 0.0001	2.1 (1.95-2.22)
N	183,867	3,706	180,161		
Females	27.0 (5.1)	29.2 (6.0)	27.0 (5.1)	< 0.0001	2.2 (2.11-2.33)
N	216,519	8,533	207,986		

3.3.9 Polygenic risk score associates with disease severity

The PRS based on the 16 genome-wide significant SNPs was significantly higher in UK Biobank CTS cases compared to controls (1.620 vs 1.566, $p < 0.0001$). This was an expected (and circular) finding, given that the 16 SNPs were derived from a GWAS of CTS cases vs controls within UK Biobank. **Fig 3.5** shows the PRS in UK Biobank by quintile, and the corresponding CTS prevalence for each quintile. The top quintile had a CTS prevalence of 4.1%, compared to a bottom quintile prevalence of 2.2%, corresponding to a relative risk of 1.88 for developing CTS if an individual's PRS is in the top rather than bottom quintile (95% CI: 1.78-2.00, $p < 0.0001$).

Of note, the UK Biobank CTS cases who have undergone surgery ($n=11,626$) had a significantly higher PRS than CTS cases who have not undergone surgery ($n=686$) (1.622 vs 1.586, $p=1.5 \times 10^{-4}$), which suggests that PRS correlates with disease severity, based on the assumption that those who have undergone surgery are phenotypically more severe than those who have not. Results from the PRS analyses are summarised in **Table 3.9**.

Polygenic risk score and CTS prevalence by PRS quintile

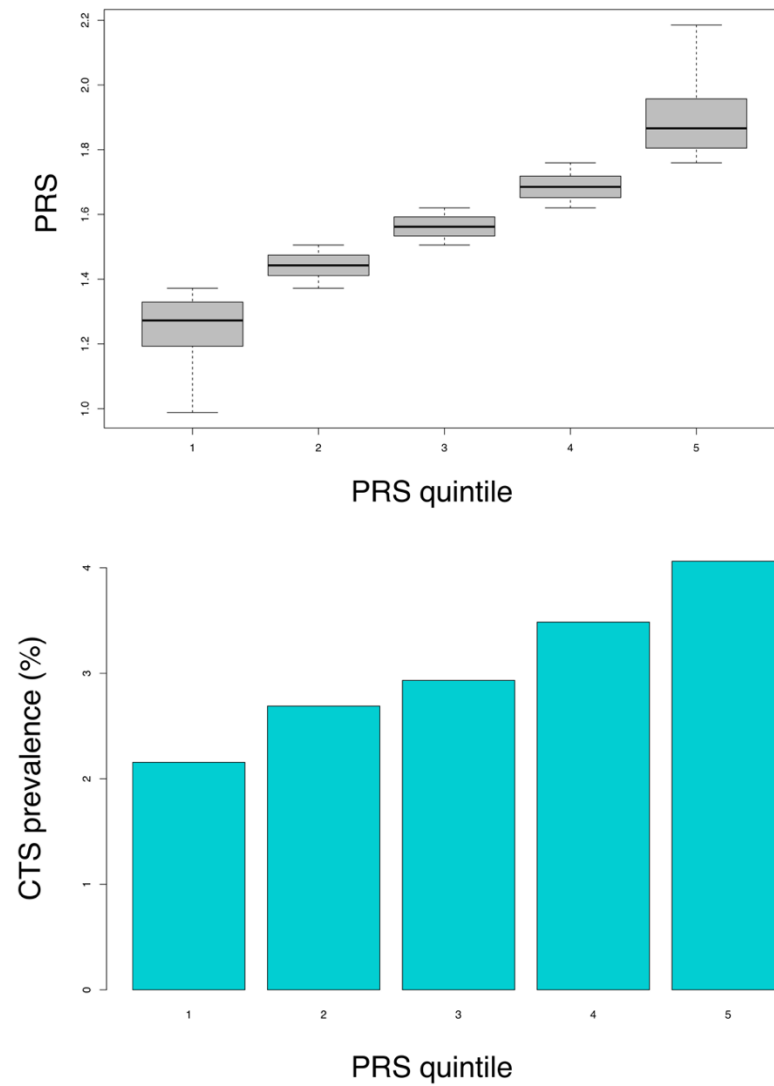


Fig. 3.5 PRS by quintile in UK Biobank. The whole UK Biobank post-QC cohort was grouped according to PRS quintile (upper box plot), and CTS prevalence calculated for each quintile (lower bar chart).

Table 3.9 PRS analyses in UK Biobank. Polygenic risk scores were compared between: (1) UK Biobank CTS cases and controls, (2) UK Biobank CTS cases with an operation code and UK Biobank CTS cases without an operation code.

Group	UKB Cases	UKB Controls	p-value
N	12,312	389,344	
PRS (S.D.)	1.62 (0.24)	1.57 (0.23)	<0.0001

Group	UKB surgical cases	UKB Non-surgical cases	p-value
N	11,626	686	
PRS (S.D.)	1.62 (0.24)	1.57 (0.23)	1.45×10 ⁻⁴

3.4 Discussion

For the analyses in this chapter, I employed a suite of bioinformatic tools in order to contextualise the GWAS findings from the previous chapter, and there were several informative findings. The 422 genome-wide significant SNPs across 16 loci demonstrated strong evidence of functionality, with several histone markers showing enrichment in CTS-associated regions of the genome. The two most significantly associated SNPs were missense variants in *ADAMTS17* and *ADAMTS10*, respectively, with another strong functional candidate residing in an enhancer region of *EFEMP1*. These genes are expressed in both fibroblasts and in Schwann cells cultured from human peripheral nerves, giving them biological plausibility in the context of CTS pathophysiology. Other implicated genes in the GWAS were disproportionately represented in gene ontologies related to ECM architecture. Partitioned heritability analysis across different tissues demonstrated an overrepresentation of musculoskeletal cell and tissue types associated with CTS heritability. Consistently, LD score correlation analysis showed strong correlations with anthropometric phenotypes, most notably with adiposity (positive correlation) and height (negative correlation). These genetic correlations were mirrored in observed associations between higher BMI (≥ 2.1 kg/m²) and shorter stature (≥ 2 cm) in UK Biobank CTS cases. A weighted polygenic risk score (PRS) based on the 16 genome-wide significant SNPs found a higher PRS in operated CTS compared to unoperated cases, consistent with the notion that carrying a greater number of the risk alleles leads to a more severe phenotype (and also providing a degree of internal validation of the phenotyping and GWAS methodology). Finally, there was also external validation of the PRS in the Estonian replication cohort, with the finding of a higher PRS in CTS cases compared to controls.

It is highly unusual in a GWAS to find that the two most significantly associated SNPs in the entire study are both missense variants, and even more unusual that both variants confer

coding changes to two closely related genes on two different chromosomes - *ADAMTS17* and *ADAMTS10*. The ADAMTS family of proteins belong to a family of multidomain, secreted extracellular proteolytic enzymes closely related to the matrix metalloproteinases (MMPs) and have important roles in ECM maintenance by cleaving procollagens and proteoglycans. The substrates of ADAMTS10 and ADAMTS17 have yet to be characterised^{202,203}, but closely related ADAMTS proteins participate in various cellular processes, including cell adhesion and migration²⁰⁴. Despite the uncertain functions of these two ADAMTS proteins, rs72755233 is known to result in a non-conservative threonine to isoleucine substitution within the protease domain of ADAMTS17²⁰⁵, which is predicted by SIFT²⁰⁶ to be deleterious. rs62621197 results in an arginine to glutamine substitution directly adjacent to the furin cleavage site of ADAMTS10, which is predicted to reduce the enzyme's activation efficiency^{205,207}. It is quite probable, therefore, that these two index SNPs are the true causal variants that confer an increased risk to an individual developing CTS. The regional plots for these two loci (Chapter 2, **Fig. 2.6**) demonstrate a scarcity of other candidate SNPs in LD with these variants in the vicinity.

As noted in Chapter 1, there is considerable evidence for an association between anthropometric measurements and CTS – short stature, short hand length, and a greater wrist index have all been linked to an increased risk of CTS^{76,208}. Having shorter hands and a ‘squarer’ wrist shape has been postulated to produce an anatomical configuration that increases the pressure within the carpal tunnel²⁰⁹. The potential relevance of these anthropometric correlates of CTS to the ADAMTS proteins is illustrated by two Mendelian diseases. Mutations in *ADAMTS10* can result in autosomal recessive Weill-Marchesani syndrome (WMS), a condition characterised by brachydactyly, joint stiffness, eye anomalies and short stature²¹⁰. Mutations in *ADAMTS17* can manifest as a WMS-like syndrome, in which affected individuals have short stature and eye anomalies²¹¹. Strikingly, CTS in the context of WMS has been previously

reported²¹², including cases of paediatric CTS²¹³ – an exceptionally rare clinical finding. These observations give credence to the notion that aberrations in skeletal growth and development as a result of altered activity of these two *ADAMTS* gene variants could feasibly contribute to CTS susceptibility.

Short stature *per se* emerged from the LD score correlation analyses as a phenotypic trait that is positively correlated with CTS. Consistent with this relationship, the four index SNPs that were most significantly associated with CTS in the GWAS (rs72755233, rs62621197, rs1863190, and rs3791679) have all been reported previously as GWAS loci for height^{205,214,215}, and with the exception of rs1863190, the CTS risk allele is associated with shorter stature in previous GWAS of height. Evidence for the short stature-CTS relationship can be found within the UK Biobank cohort, where participants with CTS of both sexes are on average 2 cm shorter than controls (**Table 3.9**).

Further support for this anthropometric hypothesis of CTS predisposition was found in the partitioned heritability analyses across different cell and tissue types, which implicated a disproportionately high number of cell and tissue types classed as ‘musculoskeletal’, with osteoblasts having the top enrichment. Of potential relevance to the latter, LDSC uncovered a statistically significant genetic correlation between CTS and bone mineral density. The question of whether there is a causal relationship between height and CTS risk will be explored in detail in the following chapter.

Aside from the strong genetic relationship between CTS and anthropometric variables, the other key finding from the various gene-based and gene-set analyses undertaken was the implication of biological pathways and gene ontologies concerned with extracellular matrix (ECM) architecture. As outlined in Chapter 1, the subsynovial connective tissues (SSCT) within

the carpal tunnel have long been the focus of investigations into CTS pathology^{134,216}, and the SSCTs from CTS patients show evidence of thickening and fibrosis^{67,217}. These changes to the SSCT are believed to interfere with neural biomechanics²¹⁸ and impair the neural microcirculation⁵², which in turn results in impaired axonal transport⁵³, altered nerve conduction⁵⁴, and eventually demyelination²¹⁹. Consequently, it seems probable that aberrations in elastic and connective tissue components within the ECM lie at the core of this pathological cascade, and several plausible genes were implicated in the GWAS to lend credibility to this idea.

rs3791679 was the fourth most significantly associated GWAS SNP, and resides in an enhancer region of *EFEMP1*¹⁹⁸ (EGF-Containing Fibulin-Like Extracellular Matrix Protein 1). It is a plausible causal variant at this locus – it has a high CADD score of 14.94, and is an eQTL for *EFEMP1* in two GTEx tissues (human skin and thyroid tissue). Moreover, several other SNPs in high LD with rs3791679 are also eQTLs for *EFEMP1*, and these have been summarised in **Table 3.10**. The protein product of *EFEMP1* is Fibulin-3, an ECM glycoprotein which is a key constituent of basement membranes, elastic fibres and matrix components²²⁰. In addition to being implicated in ECM architecture²¹⁴, *EFEMP1* is also important in bone and cartilage development pathways; thus, it has potentially contributed to the observed enrichments between CTS and height, osteoblasts and bone mineral density. In fact, *EFEMP1* variants (including rs3791679) have been associated with adult height²¹⁴, but also a range of disease phenotypes including various cancers²²¹, inguinal hernia²²², and age-related macular degeneration²²³. *EFEMP1* is crucial for the integrity of elastic tissues, as evidenced by disrupted elastic tissues and propensity to develop hernias in *EFEMP1* knockout mice²²⁴. From the eQTL data, it would be reasonable to hypothesize that the CTS risk allele at rs3791679 leads to an overexpression of *EFEMP1*, thereby altering the elastic properties of the SSCT; such a change could contribute either to increased tethering and/or impaired gliding of the median nerve

within the carpal tunnel, which, in turn, would predispose it to the vicious cycle of ischaemia and fibrosis. What makes this gene particularly interesting is that a microRNA, miR-338-5p, has been shown to inhibit *EFEMP1* synthesis in glioblastoma-derived cells²²⁵, illustrating that it has the potential to act as a modifiable therapeutic target in CTS.

Table 3.10 eQTLs for functional SNPs at 2p16.1 (*EFEMP1* locus). eQTLs were obtained from GTEx v7. The lead SNP at this locus is rs3791679. The other SNPs were chosen on the basis of being genome-wide significant and having likely functional consequences (CADD score >12.37 or a Regulome DB score of 2b or more). The LD relationship between rs3791679 and the other SNPs was calculated using 1000 Genomes GBR data.

SNP	Gene	p-value	Normalised Effect Size	Tissue	r ² with rs3791679
rs3791679	<i>EFEMP1</i>	2.0×10 ⁻¹³	0.32	Thyroid	1
	<i>EFEMP1</i>	1.5×10 ⁻⁶	0.25	Skin – Not Sun Exposed	
	<i>CLHC1</i>	4.4×10 ⁻⁵	0.29	Whole Blood	
	<i>EFEMP1</i>	6.1×10 ⁻⁵	0.19	Skin – Sun Exposed	
rs6752931	<i>EFEMP1</i>	1.1×10 ⁻¹⁶	0.34	Thyroid	0.76
rs4146922	<i>EFEMP1</i>	7.2×10 ⁻¹⁶	0.35	Thyroid	0.82
	<i>CLHC1</i>	1.1×10 ⁻⁵	0.32	Whole Blood	
	<i>RPS27A</i>	2.2×10 ⁻⁵	-0.21	Colon - Sigmoid	
rs58680090	<i>EFEMP1</i>	1.3×10 ⁻¹⁶	0.36	Thyroid	0.86
	<i>CLHC1</i>	2.6×10 ⁻⁵	0.31	Whole Blood	
	<i>RPS27A</i>	3.7×10 ⁻⁵	-0.21	Colon - Sigmoid	
rs1344732	<i>EFEMP1</i>	3.1×10 ⁻¹³	0.28	Thyroid	0.67
rs7563032	<i>EFEMP1</i>	1.1×10 ⁻¹⁵	0.35	Thyroid	0.83
	<i>CLHC1</i>	2.2×10 ⁻⁵	0.31	Whole Blood	
	<i>RPS27A</i>	3.3×10 ⁻⁵	-0.21	Colon - Sigmoid	
rs6739641	<i>EFEMP1</i>	4.0×10 ⁻¹³	0.28	Thyroid	0.67
	<i>CLHC1</i>	4.5×10 ⁻⁵	0.26	Whole Blood	

Another matrix gene that was strongly implicated in the GWAS was *COL11A1*. The top GWAS hit at the 1p21.1 locus is a variant located approximately 0.1 Mb away from *COL11A1* (rs12406439, p=1.1×10⁻⁸, OR=1.05), which is the only gene within ± 0.5 Mb of this SNP

(Chapter 2, **Figure 2.5**), making *COL11A1* a strong candidate gene for influencing CTS susceptibility. *COL11A1* encodes $\alpha 1(XI)$, one of the three constituent chains of Type XI collagen's triple helix²²⁶. Aberrations in Type XI collagen have been linked to a number of musculoskeletal conditions such as lumbar disc herniation²²⁷ and Achilles tendinopathy²²⁸. A candidate gene study has linked a variant in this gene to CTS, demonstrating that TT homozygotes at rs3753841 (a missense variant) in *COL11A1* were significantly overrepresented among CTS patients compared to controls²²⁹. The index SNP at this locus in the GWAS, rs12406439, is approximately 140 kb away from rs3753841, and the two SNPs are in moderately high LD ($r^2=0.69$, 1000 Genomes Phase 3, GBR data). As a matter of fact, rs3753841 is a genotyped SNP on the UK Biobank genotyping array, and its p-value in the association study is nominally significant (5.4×10^{-5}), strongly suggesting an association with the CTS phenotype. Much like *EFEMP1*, therefore, there is fairly compelling evidence that the abnormal function or expression of *COL11A1* contributes to the genetic aetiology of CTS.

Transforming growth factor-beta (TGF- β) plays a central role in the regulation of ECM gene expression²³⁰. The role played by the Transforming growth factor-beta (TGF- β)/Smad signalling pathway in the development of fibrosis in the SSCT in CTS has been reported, including the finding that the inhibition of TGF- β 1 in fibroblasts derived from the SSCT of CTS patients results in the inhibition of Smad activity and the down-regulation of several fibrosis-related genes^{66,231}. Given these previous findings, it is particularly striking that three of the 16 GWAS loci had TGF- β /Smad signalling pathway genes mapped to them in the FUMA analysis: latent-transforming growth factor beta-binding protein 1 (*LTBP1*) on chromosome 2, transforming growth factor β -3 (*TGFB3*) on chromosome 14, and *SMAD6* on chromosome 15. *LTBP1* is an extracellular glycoprotein that plays a critical role in the storage of TGF- β in the matrix, thus regulating its availability²³². It is noteworthy that *LTBP1* was one of only two genes that met statistical significance in the SMR analysis using eQTL data from transformed

fibroblasts, lending further credence to the importance of this signalling pathway in CTS pathogenesis. Local corticosteroid injection is an effective therapy for CTS²³³ (at least in the short term), and it has been demonstrated in hepatic stellate cells and myofibroblasts that dexamethasone specifically upregulates TGF- β 3 (but not TGF- β 1 or TGF- β 2)²³⁴. A testable hypothesis in a future study is therefore whether steroid injection into the SSCT of CTS patients alters the expression of these three genes in the TGF- β /Smad signalling pathway.

It would be naïve to make the assumption that the above genes that are linked to ECM architecture increase CTS susceptibility by acting solely in the SSCT, as there are two other tissues that need to be considered: nerves and blood vessels. The gene-property analysis in MAGMA (**Fig. 3.1**) found that the GTEx tissues most strongly enriched in the CTS GWAS were, in descending order of significance: tibial artery, coronary artery, tibial nerve, aorta, transformed fibroblasts. The evidence for the role of vascular proliferation and hypertrophy in the fibrosis seen in the SSCT was reviewed in Chapter 1, as was the role of aberrations in the epineural and endoneurial microvasculature contributing to fibrotic changes in the median nerve. It is well known that the ECM of a blood vessel is more than just a scaffold, and contributes substantially to the diverse functions of blood vessels^{235,236}. Therefore, the possibility of variants in genes such as *COL11A1* and *EFEMP1* contributing to CTS susceptibility by altering blood vessel structure or function within the SSCT or median nerve cannot be dismissed.

Peripheral nerves, like blood vessels, are functionally highly dependent on the ECM. ECM glycoproteins affect nearly all aspects of nervous system development and function, from neural stem cell differentiation to the formation of axonal tracts²³⁷. Schwann cells form a basement membrane composed of ECM which is required for myelination²³⁸, and each fascicle of a peripheral nerve is encircled by flattened cellular layers of the perineurium, which is

associated with ECM that contains collagen fibrils²³⁹. In a study of experimental nerve injury in the mouse, it was found that *ADAMTS17* is upregulated in the distal nerve stump following nerve crush, and this upregulation was dependent on the expression of Neuregulin-1, a growth factor that plays a critical role in remyelination and nerve repair²⁴⁰. The Weiss et al. RNA-Seq data demonstrated that *ADAMTS17* is more highly expressed in human cultured Schwann cells than in fibroblasts (**Fig. 3.4**), which raises the intriguing hypothesis that variants in *ADAMTS17* may impede neural recovery following treatment for CTS.

On the subject of peripheral nerves, it is noteworthy that the GWAS did not implicate genes that are associated with CTS in the context of inherited neuropathies. Haplo-insufficiency of the *PMP22* gene manifests as Hereditary neuropathy with liability to pressure palsy (HNPP)¹²⁸, a disorder in which CTS is a common manifestation²⁴¹. Reduced expression of *PMP22*, a myelin protein, results in myelin instability and conduction block in response to external pressure²⁴². Similarly, *SH3TC2* is a plasma membrane protein expressed in the Schwann cells of peripheral nerves, and mutations in this gene are associated with Charcot-Marie-Tooth Disease, which also confers susceptibility to CTS²⁴³. These monogenic disorders illustrate how genetically conferred vulnerability of peripheral nerves could be relevant to the genetic landscape of CTS. As such, both *PMP22* and *SH3TC2* are genes that might reasonably have been expected to be linked to common genetic variants that would be discovered in a GWAS of CTS. Examining all of the SNPs located within ± 1 Mb of the genes (*PMP22*: 17:148982150-149063174; *SH3TC2*: 5:148982150-149063174), the most significantly associated SNP at the *PMP22* locus was rs71361436 ($p=2.7 \times 10^{-4}$), and rs17709363 ($p=9.7 \times 10^{-4}$) at the *SH3TC2* locus. Therefore, the evidence for enrichment of these two genes in the CTS GWAS is weak, contrasting markedly with the over-representation of growth- and extracellular matrix-related genes.

An array of bioinformatic tools has been central to the various analyses described above, so it would be prudent to conclude this chapter by discussing the limitations inherent in such methods. Central to these limitations is the fact that the regulation of gene expression is highly tissue- and context-specific. Ideally, data linking SNPs to gene expression should be generated experimentally for every human tissue, in both diseased and non-diseased conditions¹⁷⁶. There is potentially limited utility of *in silico* analyses based on eQTL data from tissues that are irrelevant to the disease phenotype being studied. In the case of CTS, this limitation is illustrated by the Summary-based Mendelian randomisation (SMR) analysis, which requires eQTL data of a particular tissue type to be specified. Such eQTL data are publicly available for the 53 GTEx tissue types and for whole blood, but not for the tissue types that are posited to be specifically relevant to CTS (i.e. tenosynovium/SSCT, median nerve, and bone). As a compromise measure, “transformed fibroblast” eQTL data were chosen for the analysis, on the basis that fibroblasts are a major cellular constituent of tenosynovium. However, it goes without saying that transformed fibroblasts will be phenotypically different from the fibroblasts that are found in carpal tunnel tenosynovium *in vivo*, which in turn will differ from fibroblasts found in diseased carpal tunnel tenosynovium.

Another limitation that applies to bioinformatic tools such as FUMA and XGR is that their pipelines rely heavily on GWAS and transcriptomic data that have been published and made publicly available. This means that the phenotypes and tissue types that are available for genetic correlation studies or pathway/ontological analyses are invariably biased towards diseases and tissues that have been extensively investigated. This skew is problematic for diseases such as CTS, which might be expected to show strong correlations with other relatively under-investigated diseases such as cubital tunnel syndrome or cervical radiculopathy. Consequently, it is important that these *in silico* tools are not used in isolation; rather, a combination of such tools (that employ different methods and are predicated on different

assumptions) should be used, and the results should be interpreted cautiously in the context of what is known about the biology of the phenotype in question.

In conclusion, the summary association statistics derived from the GWAS has allowed the interrogation of common variants associated with CTS. At the start of the previous chapter, two broad *a priori* hypotheses were presented that could explain how common genetic variants could feasibly increase the risk of an individual developing CTS: either variants alter the environment within the carpal tunnel through which the median nerve transits, or the variants render the nerve fibres within the median nerve more vulnerable to compression or sensitisation to pain. The multitude of *in silico* analyses performed in this chapter appear to converge to favour the first hypothesis over the second: genetic factors related to the extracellular matrix and musculoskeletal growth (with implications for nerve environment and structure) appear to be the dominant genetic determinants of CTS susceptibility. It must be restated, however, that the two hypotheses are merely broad conceptual frameworks to help differentiate ‘structure’ from ‘function’, and are not mutually exclusive; in fact, the highly polygenic nature of complex traits are incongruous with such crude dichotomies.

In the following chapter, I will make further use of the GWAS results in an attempt to investigate causality between CTS and a number of proposed risk factors. Chapter 5 will see further applications of the polygenic risk score to a CTS cohort recruited in Oxford, and I will also investigate gene expression in surgical tissues resected from the carpal tunnel.

Chapter 4

The epidemiology of carpal tunnel syndrome and its associated phenotypes

4.1 Introduction

Chapter 1 saw an exposition of the evidence for an association between CTS and several other phenotypes, the most well-established of which are: sex, age, anthropometric measures (wrist dimensions and height), adiposity as measured by body mass index (BMI), diabetes, rheumatoid arthritis, and hypothyroidism. Although the evidence for these phenotypes as putative risk factors for CTS is relatively robust, these associations are by no means conclusive, with several studies presenting conflicting evidence – for instance, Hendriks et al.²⁴⁴ found in a retrospective case-control study of 997 CTS cases and 594 controls that the association between diabetes and CTS disappeared once adjustments were made for age, sex and BMI. Moreover, it is difficult, if not impossible, to establish causality (rather than mere association) between these phenotypes and CTS due to potential confounding – a perennial problem in epidemiological studies.

This can be illustrated by considering the case of BMI and its relationship with CTS. A large meta-analysis concluded that obesity increased the risk of CTS more than twofold²⁴⁵, and this association was indeed faithfully replicated in the UK Biobank cohort (Chapter 3, **Table 3.8**). While this relationship may suggest that raised BMI is causal in CTS aetiology, the reverse of

this interpretation is also conceivable – hand pain and dysfunction as a result of CTS might prevent an individual from undertaking their activities of daily living and taking regular exercise, which would result in raised BMI. Perhaps more problematic than this reverse causation is the presence of several plausible confounding variables that are potentially independently associated with both raised BMI and increased risk of CTS, such as diet, occupation and socioeconomic status.

The gold standard experimental study in epidemiology for inferring causality between a risk factor (exposure) and a disease (outcome) is the randomised control trial (RCT). In the case of BMI and CTS, one could hypothetically devise such an RCT: randomise a large group of individuals without CTS to either to a high-calorie diet or a normal diet, follow the two cohorts over several years, track their weight to confirm more weight gain in the high-calorie group, and ascertain prevalence of CTS in both groups over time. Needless to say, such an RCT would not be feasible on ethical and practical grounds. There is, however, an alternative way to conduct this RCT.

4.1.1 Mendelian Randomisation

A Mendelian randomisation (MR) experiment is frequently described as a “natural RCT”. Due to the independent assortment of chromosomes during the meiotic segregation of alleles, all individuals are randomly allocated a number of genetic variants for any given phenotype, at the moment of conception. This concept was elegantly described by R. A. Fisher²⁴⁶:

"Genetics is indeed in a peculiarly favoured condition in that Providence has shielded the geneticist from many of the difficulties of a reliably controlled comparison. The different genotypes possible from the same mating have been beautifully randomised by the meiotic process. A more perfect

control of conditions is scarcely possible, than that of different genotypes appearing in the same litter."

Individuals in the population can therefore be grouped according to possession of genetic variants that modify an exposure (i.e. risk factor), and the MR approach uses genetic variants as instrumental variables (IVs) to assess the causal effect of an exposure on an outcome²⁴⁷ (**Fig. 4.1**). Given that the IVs are determined by parental genotypes and are independent of environmental factors, MR is less vulnerable to unmeasured confounding and reverse causation, which are major limitations in observational studies. MR is predicated on the satisfaction of three different assumptions regarding each the genetic variants used as IVs^{248,249}: (1) they are associated with the risk factor of interest, (2) they are not associated with any confounders of the risk factor-outcome association, and (3) they affect the outcome only through their effect on the risk factor.

Single-sample MR measures both the exposure and the outcome in the same population, whereas two-sample MR allows those measurements to be made in two separate study populations²⁴⁸. The advantage of the latter approach is that it allows summary association statistics from two different GWAS to be used in MR analyses, and it can therefore exploit the growing availability of summary data from large GWA studies, thus conferring greater statistical power. In this chapter, I will make use of publicly available summary association statistics from large GWAS of phenotypes causally implicated in CTS, to perform a set of MR studies of these phenotypes and CTS.

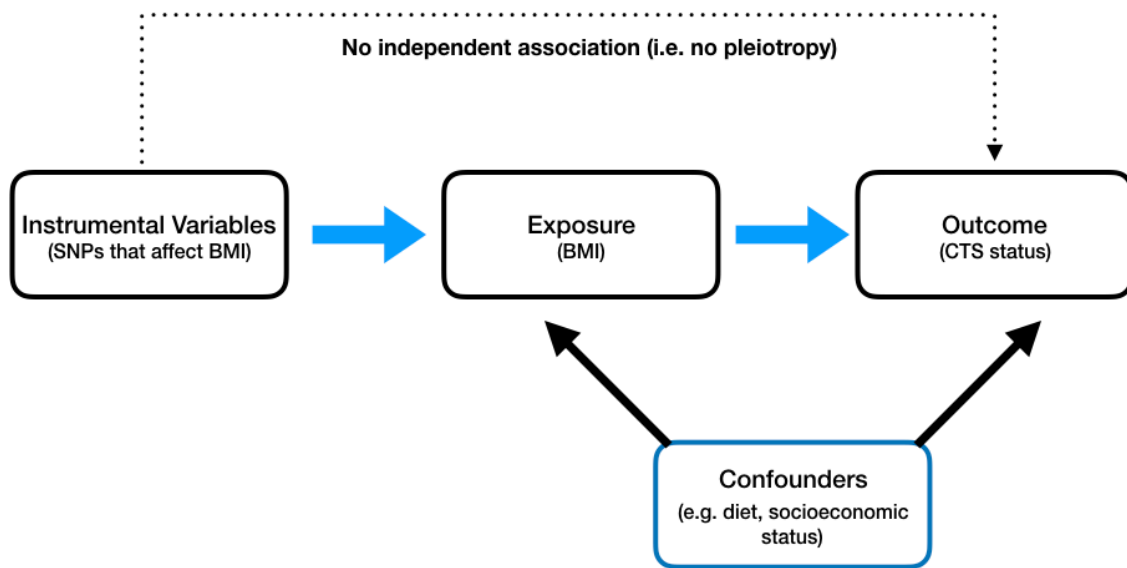


Fig. 4.1 Schematic of Mendelian randomisation analysis. This illustrates the principles of a MR analysis between BMI and CTS status, using genome-wide significant SNPs taken from a GWAS of BMI as instrumental variables. One of the fundamental assumptions of MR analysis is the absence of pleiotropy, i.e. that the genetic instruments only affect the outcome (CTS status) by acting through the exposure (BMI).

4.1.2 Migraine and CTS

The advantage of hypothesis-generating research such as GWAS is that it allows the study of hitherto unexplored associations between different phenotypes. While the principal focus of this chapter will be around the relationship between CTS and its well-established associated and comorbid phenotypes, I will also take the opportunity to study the genetic associations between CTS and migraine headaches, which has hitherto not been investigated.

The proposed association between CTS and migraine is currently almost entirely conjectural. On the surface, CTS and migraine appear to be two completely unrelated pathologies: one is an entrapment neuropathy in the wrist, and the other is a neurological disorder characterised by severe headache accompanied by a host of neurological symptoms such as visual

disturbance²⁵⁰. However, both are diseases that cause debilitating pain, both disproportionately affect females, and both are known to have significant pain and psychological and psychiatric comorbidities²⁵¹. Intriguingly, both diseases can be treated by nerve decompression of peripheral nerves, and the story of surgical migraine treatment deserves some scrutiny.

The idea of a surgical treatment for migraines was arrived at by serendipity – in 1999, it was reported that several patients who underwent plastic surgical procedures for forehead rejuvenation noticed that their migraine headaches had improved, or disappeared entirely²⁵². Since the early 2000s, Dr Bahman Guyuron has developed and promoted the theory that migraine headaches are triggered by nerve compression at specific points in the head and neck²⁵³. Several such “trigger points” that are amenable to surgical decompression have been discovered, including the ‘frontal’ (supraorbital and supratrochlear nerves), ‘temporal’ (zygomaticotemporal nerve), and ‘occipital’ (greater occipital nerve), and several anatomical studies have confirmed the entrapment of head and neck sensory nerves in bone, muscle, fascia, and vessels^{254,255}. The efficacy of such treatments has been purported in several studies, including a prospective RCT in which 92% of patients reported a >50% improvement in the frequency, intensity and duration their migraine headaches²⁵⁶, with 88% reporting improvements at 5 years post-intervention²⁵⁷. A 2009 prospective RCT with sham surgery²⁵⁸ found that 83% of patients demonstrated symptom improvement (with complete symptom elimination in 57%), compared to 56% of patients in the sham surgery arm. However, several of these (and similar) studies have come under criticism in the neurological and headache literature^{259–261}, with the suggestion that there were methodological flaws in these studies and that placebo mechanisms and the natural history of migraine explain the outcomes decompression surgery. While nerve decompression surgery for migraine is now widely practised (predominantly by plastic surgeons in the USA)²⁵⁵, its efficacy and rationale remain deeply controversial within the wider academic community.

Based on the similarities in surgical treatment for CTS and migraine headaches, Law et al.²⁶² hypothesized that there might be a pathophysiological link between CTS and migraine. The researchers scrutinised the cross-sectional 2010 National Health Interview Survey data from 25,880 respondents in the USA, and calculated that CTS prevalence in patients with migraine headache was 8% compared with 3% in those without migraine headache (adjusted OR = 2.67; 95% CI, 2.22–3.22). A crucial shortcoming of this study was that their case ascertainment of “migraine” patients included anybody who answered “yes” to the following question: “During the past 3 months, did you have severe headache or migraine?” Thus, the study did not specifically address the epidemiological association between migraine headache *per se* and CTS, which significantly undermines the study’s conclusion that it demonstrates an association between CTS and migraine headache.

In this chapter, I will investigate the epidemiological association between migraine *specifically* and CTS in UK Biobank, and will make use of a large meta-analysis of migraine GWAS to investigate whether there are any genetic correlations between the two diseases.

4.1.3 Aims

The broad aim of this chapter is to harness the power of the epidemiological and genetic data in the UK Biobank resource to explore the relationships between CTS and the aforementioned associated phenotypes and comorbid diseases.

The specific aims of this chapter are as follows:

1. Investigate epidemiological associations between CTS and three associated comorbid diseases (diabetes, rheumatoid arthritis, hypothyroidism) and one putative comorbid disease (migraine). I will further investigate whether individuals who have these diseases are genetically predisposed to CTS by having elevated polygenic risk scores (PRS) for CTS compared to controls.
2. Investigate whether there are sex differences in the genetic architecture of CTS by performing sex-stratified GWAS of CTS using UK Biobank data.
3. Perform Mendelian randomisation analyses for causal inference between CTS and a range of phenotypes for which GWAS association statistics are publicly available: height, BMI, rheumatoid arthritis, hypothyroidism, and two female reproductive phenotypes (age at menarche and age at menopause).
4. Investigate possible genetic correlations between migraine and CTS by performing LD score correlation and cross-phenotype association analyses.

4.2 Methods

4.2.1 Nested case-control study of CTS and other disease phenotypes

A series of nested case-control studies were undertaken in order to confirm epidemiological associations in the UK Biobank dataset between CTS and three previously established comorbid diseases (diabetes, rheumatoid arthritis and hypothyroidism), and one putative disease association – migraine. The fundamental question being asked in these analyses was whether having a diagnosis of one of these comorbid diseases increases the odds ratio for the association with CTS.

4.2.1.1 Phenotyping

Phenotyping for these nested case-control studies was performed in a similar manner to phenotyping for CTS, as outlined in Chapter 2. The UK Biobank showcase was scrutinised for both hospital and self-reported diagnostic codes for diabetes, rheumatoid arthritis, hypothyroidism and migraine. The hierarchical ICD-10 diagnostic codes and the self-reported diagnostic codes used for phenotyping individuals with these diseases are summarised in **Table 4.1**, including the total numbers in each group. The cohort used for case and control ascertainment was the post-QC UK Biobank cohort – i.e. 401,656 individuals of white British ancestry.

Table 4.1 Diagnostic codes used to phenotype individuals with four diseases associated with CTS. All sub-codes of the ICD-10 codes shown below were included.

Diabetes	Rheumatoid Arthritis	Hypothyroidism	Migraine
ICD-10 codes			
E10: Insulin-dependent diabetes mellitus	M05: Seropositive rheumatoid arthritis	E02: Subclinical iodine-deficiency hypothyroidism	G43: Migraine
E11: Non-insulin-dependent diabetes mellitus	M06: Seronegative rheumatoid arthritis	E03: Other hypothyroidism	
E13: Other specified diabetes mellitus			
E14: Unspecified diabetes mellitus			
Non-cancer illness code (self-reported)			
1220: Diabetes	1464: Rheumatoid arthritis	1226: Hypothyroidism	1265: Migraine
1222: Type 1 Diabetes			
1223: Type 2 Diabetes			
N = 24,588	N = 6,711	N = 23,838	N = 14,409

4.2.1.2 Case-control matching for four disease phenotypes

Matching of cases to controls was performed at a 1:5 ratio, using the R package MatchIt²⁶³. The nearest neighbour matching method was selected, and the matching variables were “sex” and “year of birth” (UK Biobank data field 34). The association between obesity and diabetes is well-known, and hypothyroidism²⁶⁴ and rheumatoid arthritis²⁶⁵ are also linked to raised BMI. Given the body of evidence implicating BMI as a risk factor for CTS, it is a potentially significant confounder in these nested case-control studies. Therefore, for each of the four disease phenotypes, two sets of case-control matching were performed, both including and excluding BMI (UK Biobank data field 21001; measured at the participants’ initial assessment visit). Thus, the matching variables were:

- (1) Sex (exact match), year of birth (nearest match)

(2) Sex (exact match), year of birth (nearest match), BMI (nearest match)

A total of eight case-control datasets were therefore generated for the four disease phenotypes.

The cobalt R package was used to assess covariate balance following the matching.

4.2.1.3 Odds ratios

For each of the eight case-control datasets, three analyses were performed: (1) for the whole cohort, and sub-group analyses for (2) males only, and (3) females only. Odds ratios were calculated as a 2×2 contingency table, with the disease phenotype (e.g. diabetes) as the exposure and CTS status as the outcome. Thus, a total of 24 odds ratios were computed. A Bonferroni correction was applied to the significance threshold to account for multiple testing across four independent disease phenotypes, i.e. $p < 0.0125$ ($0.05/4$). No additional correction was applied to account for the BMI-dependent and BMI-independent matching criteria used for each phenotype, nor for the sex-stratified analyses, as these results are not independent of one another and will be highly correlated.

4.2.2 Application of polygenic risk score (PRS) to other diseases and phenotypes

LD score correlation analyses were performed in Chapter 3, which suggested genetic correlations between CTS and both diabetes and rheumatoid arthritis (both were nominally significant but did not survive Bonferroni-correction). To further explore this relationship, and to ascertain whether individuals with the above four diseases have a greater genetic liability for CTS than individuals without those diseases, PRS for CTS was calculated for cases and controls for each of the four phenotypes (e.g. diabetics vs non-diabetics). The same method was used to compute a weighted PRS across the 16 risk loci as previously described in Chapter

3. As with the odds ratio computations, a total of 24 comparisons of PRS between cases and controls were performed (four phenotypes matched with and without BMI, with three analyses for each group – the whole cohort, males only, and females only). Two-tailed t-tests were performed in R, and the same Bonferroni-corrected threshold of $p < 0.0125$ was used to define statistical significance.

4.2.3 Sex-stratified GWAS

Two sex-stratified GWAS were undertaken, using only the male and female CTS cases and controls, respectively, from the CTS GWAS described in Chapter 2. The reasons for performing these GWAS were threefold: (1) to ascertain whether there are differences in the genetic architecture of CTS between the sexes, (2) discover any sex-specific genome-wide significant association loci, and (3) to create a set of GWAS summary association statistics specific to female CTS patients that can be used in Mendelian randomisation analyses of CTS and female reproductive phenotypes.

The post-QC UK Biobank dataset from Chapter 2 was reused for this analysis, and the cohort was split according to genetically determined sex:

(1) Male CTS GWAS: 3,738 cases; 180,761 controls

(2) Female CTS GWAS: 8,574 cases; 208,583 controls

Given that the ratio of female:male CTS cases is $> 2:1$, there is expected to be a significant discrepancy in statistical power between the two GWAS. Therefore, heterogeneity statistics (I^2 and heterogeneity p-value) were calculated for the 16 index SNPs between these two GWAS, using the MetaGenyo²⁶⁶ web tool.

A third GWAS was performed as a sensitivity analysis, whereby the number of female cases and controls in the GWAS were matched to the numbers of cases and controls in the male-only GWAS:

(3) Matched Female CTS GWAS: 3,705 cases; 179,469 controls

The rationale behind this sensitivity analysis was that it would negate the effects of discrepancies in statistical power, and a fair comparison can therefore be made between the strength of association signals in the two GWAS. Matching was performed 1:1 using the R package MatchIt²⁶³ as described above, using the following matching variables and parameters: “sex” (exact match), “genotyping platform” (exact match) and “year of birth” (nearest match). Based on this, cases/controls were matched almost exactly 1:1 between the male and female GWAS.

The three association analyses were performed in BOLT-LMM v2.3¹⁵⁶, with no conditioning on covariates. Association summary statistics were scrutinised locus-by-locus, and Manhattan plots produced in FUMA¹⁸¹. Regional association plots were produced using LocusZoom¹⁵⁹ for novel genome-wide significant loci not previously found in the original (mixed-sex) GWAS. Furthermore, a comparison was made of association p-values and odds ratios for the 16 genome-wide significant SNPs discovered in the original GWAS in each of the above three GWAS. eQTLs were analysed using GTEx¹⁸⁰.

4.2.4 Mendelian randomisation analyses

To assess the potential causal roles of various phenotypes on CTS, Mendelian randomisation analyses were undertaken. Seven phenotypes with relevance to CTS (height, BMI, diabetes,

rheumatoid arthritis, hypothyroidism, age at menarche and age at menopause) were selected as exposures, with CTS status as the outcome.

A two-sample MR approach was used, whereby the genetic effect estimates for the genome-wide significant SNPs in the GWAS for the seven phenotypes are compared to the genetic effect estimates for those same SNPs in the CTS GWAS. A literature search was performed to find the most recent and comprehensive GWAS or GWAS meta-analyses for the seven phenotypes in order to identify the instrumental variable (IV) SNPs. Studies were limited to those of European-ancestry individuals, with the exception of rheumatoid arthritis, explained in greater detail below. All SNPs selected as IVs were scrutinised to ensure that they were genome-wide significant in their respective GWAS, and that they were independent. The linkage disequilibrium (LD) relationships between SNPs in close proximity were examined, so that if a pair of SNPs had an $r^2 > 0.1$, the less statistically significant of the two was excluded.

Candidate IV SNPs that passed this level of QC were then extracted from the CTS GWAS summary association statistics (or the female-only CTS GWAS summary statistics in the case of the two female reproductive phenotypes). In most cases, not all of the candidate IV SNPs were present in the CTS summary statistics (owing to different imputation panels used), and in these instances, SNPs in high LD ($r^2 > 0.9$) with the IV SNPs were selected as proxy SNPs. Strand alignment between the two sets of summary statistics was checked manually, as was concordance between the allele frequencies. If any of the IV SNPs were genome-wide significant in the CTS GWAS (which could potentially bias the MR estimate), a sensitivity analysis was performed by running a second MR with the exclusion of those SNPs.

MR analyses were performed using the MendelianRandomization package²⁶⁷ for R. This software calculates the Wald ratios (β_{IV}) and corresponding 95% CIs for each SNP. The ratios

are calculated by dividing the per-allele log-OR of CTS (β_{ZY}) by the per-allele difference in the exposure phenotype^{***} for each SNP (β_{ZX}). If the outcome is a binary disease trait, the Wald ratio can be interpreted as the log-OR^{†††} for disease per unit increase in the exposure due to the SNP. Three different MR analyses were performed for each phenotype:

(1) *The Inverse-variance weighted (IVW) MR method.* This method assumes that all SNPs are valid IVs. Conventional linear regression of the Wald ratios for each SNP is undertaken, and is then weighted by the inverse variance of the IV estimate. The IVW method constrains the regression such that when β_{ZX} is equal to zero, β_{ZY} is also zero.

(2) *MR-Egger regression.* This method tests for horizontal pleiotropy (whereby genetic variants affect an outcome through pathways other than through the risk factor of interest²⁴⁸), and gives a causal estimate if such pleiotropy is present. As in the IVW method, in MR-Egger, β_{ZY} is again plotted against β_{ZX} . However, unlike in IVW, the intercept is not fixed, so any deviation from the origin gives evidence that there are pleiotropic effects in the corresponding direction. If the intercept is equal to zero, the MR-Egger estimate will be exactly the same as the IVW estimate²⁴⁹. MR-Egger regression relies on the Instrument Strength Independent of Direct Effect (InSIDE) assumption to be valid: the strength of the gene-exposure association does not correlate with the strength of the pleiotropic effects across the IVs²⁶⁸.

(3) *The weighted median method.* In this method, MR estimates are weighted and ordered by the inverse of their variance. Thus, variants with more precise causal estimates contribute more weight to the analysis²⁶⁹. Provided that more than half of the total weight is attributed to SNPs

*** The units of this per-allele difference varies depending on the exposure phenotype. In the case of disease phenotypes, e.g. diabetes, it will be the log-OR of diabetes; conversely, for continuous variables, it will be, e.g. height (in cm), age at menopause (in years), etc.

††† Natural logarithm of the odds ratio.

without any pleiotropic effects, the median MR estimate should remain unbiased to pleiotropy. Hence, the weighted median approach is more robust to violation of the InSIDE assumption than MR-Egger²⁶⁸.

The reason for undertaking three different MR analyses is because causal inference should not be based upon a single method. Rather, it should be corroborated by multiple methods that are predicated on different assumptions²⁴⁹, and some authors consider both MR-Egger and the weighted median to be tools for sensitivity analysis rather than for primary analysis. As MR analyses were undertaken for seven different phenotypes, a Bonferroni-corrected p-value of <0.0071 ($0.05/7$) was defined as the threshold for significance of the MR estimates. Study-specific methods for each of the seven MR analyses are outlined below, and the GWA studies used for these analyses are summarised in **Table 4.2**.

4.2.4.1 Height

A 2014 GIANT consortium meta-analysis of 253,288 individuals of European ancestry identified 697 variants associated with adult height²¹⁵, and these SNPs were used in the height-CTS MR. A total of 601 of these SNPs were present in the CTS GWAS summary statistics and were selected as IVs. Five of these SNPs were genome-wide significant in both the height and CTS GWAS (or were in high LD with such a SNP), so a second MR was performed as a sensitivity analysis with these SNPs excluded.

4.2.4.2 Body mass index (BMI)

Summary statistics from a large meta-analysis of BMI that discovered 97 BMI-associated loci were used for this MR analysis²⁷⁰. SNPs that were genome-wide significant in European

populations (in both sexes) were selected – 96 of these SNPs were present in the CTS GWAS summary statistics.

4.2.4.3 Diabetes

A recent GWAS meta-analysis of type 2 diabetes in individuals of European ancestry was used for this analysis¹³⁷, which identified 139 common variants associated with type 2 diabetes – all SNPs were present in the CTS GWAS summary statistics and were used as IVs. This meta-analysis included c. 63,000 cases and c. 600,000 controls; notably, c. 16,000 of the cases and c. 100,000 of the controls were derived from UK Biobank, meaning that there was a degree of sample overlap between this GWAS meta-analysis and the CTS GWAS, the significance of which will be explored further in the Discussion.

4.2.4.4 Rheumatoid arthritis (RA)

SNPs derived from a GWAS meta-analysis by Okada et al.²⁷¹ were used for this analysis. Forty-two novel genome-wide significant loci were discovered from c. 30,000 RA cases and c. 74,000 controls of European and Asian ancestries, in a three-stage trans-ethnic meta-analysis. One of the key assumptions from two-sample MR is that the two studies represent the same underlying population²⁴⁸; thus, there was a need to negate the effect of heterogeneity in SNP effects arising due to different ancestral populations used in stage 1 (European) and stage 2 (Asian) of the study. A SNP-selection algorithm was devised whereby SNPs were only included as IVs if they were genome-wide significant in stage 1, or significant in the trans-ethnic meta-analysis with no evidence of heterogeneity between stages 1 and 2. Heterogeneity was calculated using Cochran's Q test using the log-OR and SEs of the European vs trans-ethnic estimates. Of the 101 genome-wide significant loci identified in this GWAS, 83 were deemed suitable for use as

IVs on the basis of the SNP selection algorithm, of which 73 were present in the CTS GWAS dataset and were carried through to the MR analysis.

4.2.4.5 Hypothyroidism

Of the 25 index SNPs discovered in a GWAS of hypothyroidism by Pickrell et al.²⁷², 19 were present in the CTS GWAS and were therefore used as IVs in this MR study.

4.2.4.6 Age at menarche

The largest GWAS meta-analysis for age at menarche was selected for this MR study. Day et al. identified 389 independent SNPs significantly associated with age at menarche²⁷³. Of these SNPs, 353 were present in the female-only CTS GWAS summary statistics. Twelve SNPs that were not present in the CTS GWAS had proxy SNPs in high LD ($r^2 > 0.9$), so these SNPs were included as IVs, for a total of 356 SNPs.

4.2.4.7 Age at menopause

A GWAS meta-analysis of menopausal timing published by the same group as above was used for this analysis²⁷⁴. Fifty-four independent signals associated with age at menopause were discovered in this meta-analysis. All were present in the female-only CTS GWAS summary statistics, but one of the SNPs was excluded on account of being in LD ($r^2 = 0.50$) with another SNP at the same locus, leaving a total of 53 SNPs which could be used as IVs.

Table 4.2 Summary of GWA studies used for Mendelian randomisation analyses.

Phenotype	Authors	PMID	N (cases)	N (controls)	Ancestry	Genome-wide significant loci
Height	Wood et al. ²¹⁵	25282103	253,288	N/A	European	697
BMI	Locke et al. ²⁷⁰	25673413	339,224	N/A	European	97
Diabetes	Xue et al. ¹³⁷	30054458	62,892	596,424	European	143
Rheumatoid arthritis	Okada et al. ²⁷¹	24390342	29,880	73,758	European and Asian	101
Hypothyroidism	Pickrell et al. ²⁷²	27182965	17,558	117,083	European	25
Age at menarche	Day et al. ²⁷³	28436984	329,345	N/A	European	389
Age at menopause	Day et al. ²⁷⁴	26414677	69,360	N/A	European	54

4.2.5 Genetic correlations and enrichment between CTS and migraine

Potential genetic correlations between migraine and CTS were examined using two independent methods. Gormley et al.²⁷⁵ performed a meta-analysis of 22 individual GWA studies of c. 375,000 individuals (59,674 cases, 316,078 controls) collected through the International Headache Genetics Consortium (IHGC), discovering 44 independent genome-wide significant SNPs at 38 loci. The summary statistics from the GWAS meta-analysis were not publicly available, so had to be applied for through IHGC. The IHGC were only able to release the summary statistics from a redacted meta-analysis that excluded participants who were recruited through 23andMe, a for-profit genotyping company. Since 23andMe contributed approximately half of the cases and controls to this meta-analysis (30,465 cases and 134,147 controls), the summary statistics obtained from the IHGC for the purposes of this analysis were the product of a significantly under-powered meta-analysis compared to what was published by Gormley et al. Nevertheless, two analyses were undertaken using these summary statistics:

- 1) LD score correlation (LDSC) analysis: In order to compute a figure for the degree of genetic correlation between CTS and migraine, the full summary association statistics from the two GWAS were analysed using LDSC¹⁹⁴ v 1.0.0, using the same methods as described in Chapter 2.
- 2) Cross-phenotype association (CPASSOC): Individual SNP overlaps between migraine and CTS were assessed using the beta-coefficients and standard error values for the SNPs in the two GWAS summary statistics, using the S_{Het} function in CPASSOC²⁷⁶. A p-value $< 5 \times 10^{-8}$ indicated a significant association.

4.3 Results

4.3.1 Nested case-control studies in UK Biobank provide further evidence for CTS-disease associations

The matching algorithm was successful in producing a set of eight well-matched case-control datasets, the demographics of which are summarised in **Table 4.3**.

Table 4.4 shows the odds ratios for association with CTS for the four disease phenotypes in the eight nested case-control studies. The greatest odds ratio for CTS was seen in males with rheumatoid arthritis (OR = 3.75), whereas the only non-statistically significant odds ratios were seen in males with migraine (both matched and unmatched for BMI). With the exception of migraine, matching on BMI had the effect of attenuating the odds ratios for development for CTS.

4.3.2 Polygenic risk score for CTS is raised in females with hypothyroidism

The polygenic risk scores for the eight nested case-control datasets are summarised in **Table 4.5**. With the exception of hypothyroidism, there was no difference in PRS for CTS in any of the case-control pairs. Notably, the greater PRS for CTS seen in hypothyroid cases vs controls (1.573 vs 1.565) was driven principally by female hypothyroid cases vs controls (males: 1.567 vs 1.565; females: 1.574 vs 1.566).

Table 4.3 Demographics of the eight matched nested case-control datasets. Matching of cases to controls was performed at a 1:5 ratio, using the R package MatchIt. The nearest neighbour matching method was selected, and the matching variables were “sex” (exact match) and “year of birth” (nearest match). The “BMI” suffix refers to the dataset produced when matching on body mass index (nearest match) as an additional variable in addition to year of birth and sex. DM = diabetes, RA = rheumatoid arthritis. The total post-QC UK Biobank cohort size from which these nested case-control sets were derived was 401,656.

	DM	DM_BMI	RA	RA_BMI	Hypothyroid	Hypothyroid_BMI	Migraine	Migraine_BMI
CASES								
N cases	24558	24558	6711	6711	23838	23838	14409	14409
Mean BMI	31.6 (5.8)	31.6 (5.8)	28.5 (5.5)	28.5 (5.5)	28.6 (5.5)	28.6 (5.5)	27.1 (5.0)	27.1 (5.0)
Mean Age	70.2 (6.9)	70.2 (6.9)	69.6 (7.1)	69.6 (7.1)	68.9 (7.3)	68.9 (7.3)	65.6 (7.8)	65.6 (7.8)
N males	15309	15309	2175	2175	3994	3994	3368	3368
N females	9249	9249	4536	4536	19844	19844	11041	11041
% females	37.7	37.7	67.6	67.6	83.2	83.2	76.6	76.6
CONTROLS								
N controls	122790	122790	33555	33555	119190	119190	72045	72045
Mean BMI	27.3 (4.3)	30.4 (4.9)	27.3 (4.7)	28.4 (5.3)	27.1 (4.9)	28.5 (5.4)	27.2 (5.0)	27.1 (4.9)
Mean Age	70.2 (6.9)	70.4 (6.8)	69.6 (7.1)	69.7 (7.0)	68.9 (7.3)	69 (7.3)	65.6 (7.8)	65.6 (7.8)
N males	76545	76545	10875	10875	19970	19970	16840	16840
N females	46245	46245	22680	22680	99220	99220	55205	55205
% females	37.7	37.7	67.6	67.6	83.2	83.2	76.6	76.6

Table 4.4 Odds ratios for association with CTS in eight nested case-control datasets. The “BMI” suffix refers to matching on body mass index as an additional matching variable.

		OR FOR CTS (95% CI)	Z-STATISTIC	P-VALUE
DIABETES	Whole dataset	2.31 (2.17-2.46)	26.4	<0.0001
	Males	1.99 (1.81-2.19)	14.3	<0.0001
	Females	2.63 (2.42-2.86)	22.7	<0.0001
DIABETES_BMI	Whole dataset	1.75 (1.65-1.86)	18.4	<0.0001
	Males	1.58 (1.44-1.74)	9.8	<0.0001
	Females	1.92 (1.77-2.08)	15.9	<0.0001
RA	Whole dataset	2.70 (2.44-2.99)	18.9	<0.0001
	Males	3.75 (3.04-4.63)	12.3	<0.0001
	Females	2.46 (2.18-2.77)	14.8	<0.0001
RA_BMI	Whole dataset	1.75 (1.65-1.86)	18.4	<0.0001
	Males	2.97 (2.42-3.64)	10.5	<0.0001
	Females	2.28 (2.03-2.57)	13.7	<0.0001
HYPOTHYROID	Whole dataset	1.47 (1.38-1.57)	11.9	<0.0001
	Males	1.61 (1.32-1.96)	4.7	<0.0001
	Females	1.46 (1.36-1.56)	10.9	<0.0001
HYPOTHYROID_BMI	Whole dataset	1.35 (1.26-1.43)	9.2	<0.0001
	Males	1.42 (1.17-1.72)	3.5	<0.0001
	Females	1.34 (1.25-1.43)	8.5	<0.0001
MIGRAINE	Whole dataset	1.14 (1.04-1.25)	2.76	0.0058
	Males	1.07 (0.82-1.40)	0.51	0.61
	Females	1.15 (1.04-1.28)	2.77	0.0057
MIGRAINE_BMI	Whole dataset	1.15 (1.04-1.26)	2.85	0.0044
	Males	1.01 (0.77-1.31)	0.045	0.96
	Females	1.17 (1.06-1.29)	3.05	0.0023

Table 4.5 Polygenic risk scores for CTS in the eight matched case-control datasets. Statistically significant differences between cases and controls after Bonferroni correction are highlighted in bold italic.

	DM	DM_BMI	RA	RA_BMI	HYPOTHYROID	HYPOTHYROID_BMI	MIGRAINE	MIGRAINE_BMI
WHOLE COHORT								
Mean PRS Cases (S.D.)	1.567 (0.23)	1.567 (0.23)	1.568 (0.23)	1.568 (0.23)	1.573 (0.23)	1.573 (0.23)	1.569 (0.23)	1.569 (0.23)
Mean PRS Controls (S.D.)	1.566 (0.23)	1.569 (0.23)	1.567 (0.23)	1.567 (0.23)	1.565 (0.23)	1.567 (0.23)	1.567 (0.23)	1.567 (0.23)
T-test	0.342	-1.231	0.324	0.228	4.588	3.757	0.601	0.828
P-value	0.733	0.218	0.746	0.82	4.48 × 10 ⁻⁶	0.000172	0.548	0.408
MALES								
Mean PRS Cases (S.D.)	1.566 (0.23)	1.566 (0.23)	1.562 (0.23)	1.562 (0.23)	1.567 (0.23)	1.567 (0.23)	1.567 (0.23)	1.567 (0.23)
Mean PRS Controls (S.D.)	1.567 (0.23)	1.569 (0.23)	1.566 (0.23)	1.567 (0.23)	1.565 (0.23)	1.567 (0.23)	1.568 (0.23)	1.568 (0.24)
T-test	-0.266	-1.2	-0.719	-0.858	0.427	-0.0596	-0.326	-0.251
P-value	0.79	0.23	0.472	0.391	0.669	0.95	0.744	0.802
FEMALES								
Mean PRS Cases (S.D.)	1.568 (0.23)	1.568 (0.23)	1.57 (0.23)	1.57 (0.23)	1.574 (0.23)	1.574 (0.23)	1.569 (0.23)	1.569 (0.23)
Mean PRS Controls (S.D.)	1.566 (0.23)	1.569 (0.23)	1.567 (0.23)	1.567 (0.23)	1.566 (0.23)	1.567 (0.23)	1.567 (0.23)	1.567 (0.23)
T-test	0.901	-0.462	0.886	0.866	4.833	4.141	0.863	1.081
P-value	0.368	0.644	0.376	0.387	1.35 × 10 ⁻⁶	3.47E-05	0.388	0.28

4.3.3 Sex differences in the genetic architecture of CTS

The male-only CTS GWAS did not reveal any genome-wide significant loci (the lowest p-value was 2.0×10^{-7} at the *ADAMTS17* locus). There were nine genome-wide significant loci in the female-only CTS GWAS, seven of which were previously discovered in the full CTS GWAS, and two of which were novel. Summary statistics for the index SNPs at these loci are shown in **Table 4.6**.

The association p-values for the 16 index SNPs from the full CTS GWAS further illustrates these sex differences (**Table 4.7**). In the male-only GWAS, only 2/16 loci SNPs reach suggestive values of significance ($p < 5 \times 10^{-5}$), whereas in the female-only GWAS, all loci are either suggestive (9/16), or genome-wide significant (7/16). This discrepancy can in part be explained by the greater power in the female-only GWAS, with more than double the number of CTS cases. The sex-stratified GWAS with female cases and controls matched to their male counterparts demonstrates that despite eliminating this power effect, females exhibit p-values closer to those in the full CTS GWAS at 14/16 loci, with 7/16 suggestive loci and 1 genome-wide significant locus. However, there was minimal evidence for heterogeneity at the 16 loci between the male-only and unmatched female-only GWAS, with only one locus (chromosome 14) demonstrating nominal evidence of heterogeneity ($I^2 = 0.74$, $p = 0.05$).

Manhattan plots from the three sex-stratified GWAS are shown in **Fig. 4.2**.

Table 4.6 Summary statistics for female-specific CTS risk loci. Shown below are the index SNPs for genome-wide significant loci in the GWAS of all female CTS cases vs all female controls in UK Biobank. EAF = effect allele frequency. INFO = Imputation score. Locus names are designated according to the FUMA-mapped genes, as per Chapter 2, Table 2.2. Novel loci not discovered in the original (non-sex stratified) GWAS are designated as “New”.

SNP	Chromosome	Position	Effect allele	Non-effect allele	EAF	INFO	OR (95% CI)	P-value	Locus
rs77841443	1	219800031	C	T	0.10	0.964	1.17 (1.12-1.24)	1.00×10 ⁻⁹	New
rs59985551	2	56106928	T	C	0.22	0.998	1.12 (1.08-1.16)	1.30×10 ⁻⁹	<i>EFEMP1</i>
rs73032510	2	176891248	G	A	0.79	0.993	1.12 (1.08-1.16)	1.70×10 ⁻⁹	<i>KIAA1715</i>
rs17181956	2	218146080	C	T	0.89	0.978	1.20 (1.14-1.26)	9.20×10 ⁻¹³	<i>DIRC3</i>
rs6977081	7	150542515	G	T	0.67	0.979	1.10 (1.06-1.13)	2.70×10 ⁻⁸	<i>AOC1</i>
rs1866745	15	67034812	A	G	0.35	0.993	1.11 (1.07-1.15)	2.20×10 ⁻¹⁰	<i>SMAD6</i>
rs72755233	15	100692953	A	G	0.11	1.000	1.16 (1.11-1.22)	1.20×10 ⁻⁹	<i>ADAMTS17</i>
rs62621197	19	8670147	T	C	0.04	0.924	1.35 (1.24-1.47)	4.20×10 ⁻¹²	<i>ADAMTS10</i>
rs2691270	19	51595230	A	G	0.56	0.966	1.09 (1.06-1.13)	1.80×10 ⁻⁸	New

Table 4.7 Comparison of CTS GWAS p-values in sex-stratified analyses. The p-values in the original full CTS GWAS are shown under the “Whole cohort” column. “males” = male-only GWAS, “females” = female-only GWAS, “matched females” = female-only GWAS with case and control numbers matched to the male-only GWAS. P-values are colour-coded in the sex-stratified GWAS. Red = genome-wide significant at $p < 5 \times 10^{-8}$; Blue = suggestive of association, at $p < 5 \times 10^{-5}$. I^2 = heterogeneity statistic between male-only GWAS and female-only GWAS (unmatched), given alongside the heterogeneity p-value.

Locus number	SNP	Chromosome	Whole cohort	P (males)	P (females)	P (matched females)	I^2	Heterogeneity P
1	rs12406439	1	1.10×10^{-8}	7.20×10^{-3}	4.40×10^{-7}	1.90×10^{-4}	0	0.60
2	rs12104955	2	3.90×10^{-8}	1.10×10^{-2}	8.60×10^{-7}	9.20×10^{-4}	0	0.50
3	rs3791679	2	2.00×10^{-12}	2.20×10^{-4}	1.60×10^{-9}	1.90×10^{-4}	0	0.80
4	rs1025128	2	2.80×10^{-9}	2.80×10^{-3}	3.00×10^{-7}	9.40×10^{-6}	0	0.70
5	rs847139	2	7.20×10^{-11}	3.80×10^{-3}	2.00×10^{-9}	1.80×10^{-4}	0.06	0.30
6	rs1863190	2	5.40×10^{-13}	1.60×10^{-3}	2.80×10^{-11}	1.50×10^{-7}	0.12	0.29
7	rs4678145	3	4.10×10^{-9}	1.10×10^{-3}	1.00×10^{-6}	1.30×10^{-2}	0	0.97
8	rs6843953	4	5.80×10^{-12}	1.10×10^{-6}	4.10×10^{-7}	4.10×10^{-7}	0.30	0.23
9	rs3828889	6	1.70×10^{-8}	4.60×10^{-4}	8.50×10^{-6}	2.10×10^{-3}	0	0.70
10	rs62422907	6	2.20×10^{-9}	1.70×10^{-3}	5.40×10^{-7}	1.00×10^{-3}	0	0.85
11	rs55841377	7	8.40×10^{-9}	1.30×10^{-2}	1.70×10^{-7}	1.60×10^{-3}	0	0.37
12	rs6977081	7	1.20×10^{-8}	5.20×10^{-2}	2.70×10^{-8}	2.10×10^{-5}	0.51	0.15
13	rs72725608	14	1.10×10^{-8}	3.50×10^{-3}	1.40×10^{-6}	4.80×10^{-6}	0	0.83
14	rs1866745	15	4.20×10^{-10}	6.70×10^{-2}	2.20×10^{-10}	1.20×10^{-5}	0.74	0.05
15	rs72755233	15	2.30×10^{-15}	2.00×10^{-7}	1.20×10^{-9}	2.00×10^{-5}	0	0.33
16	rs62621197	19	7.50×10^{-14}	1.80×10^{-3}	4.20×10^{-12}	2.20×10^{-8}	0.21	0.26

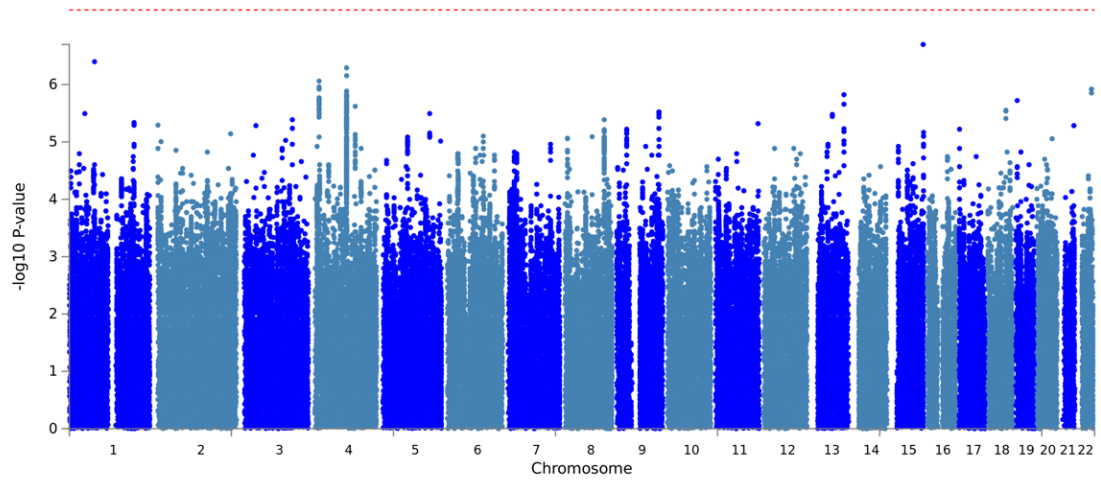
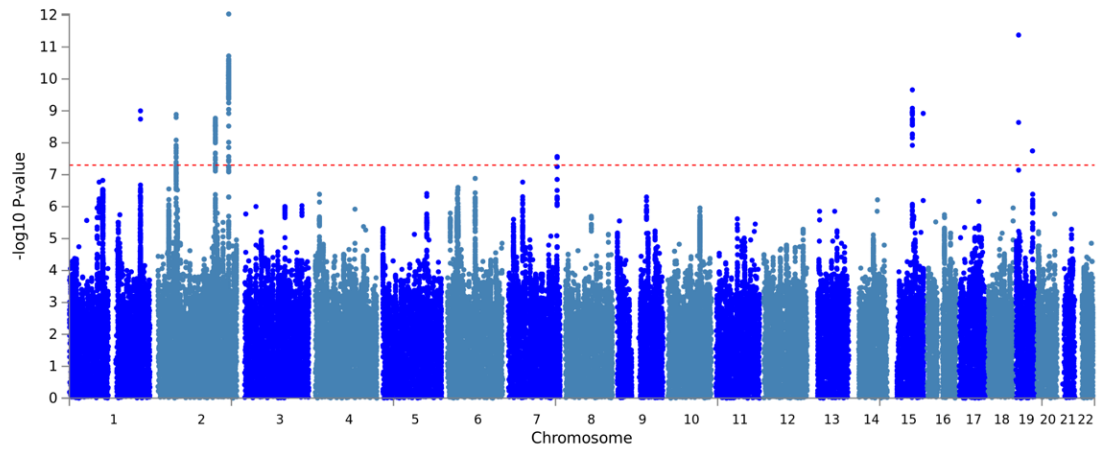
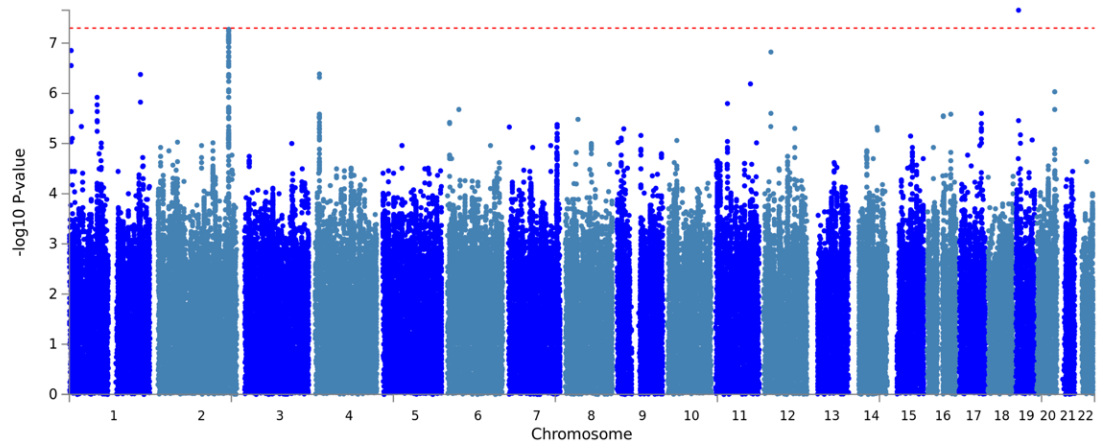
a**b****c**

Fig. 4.2 Sex-stratified GWAS Manhattan plots. (a) Male-only CTS GWAS, (b) Female-only CTS GWAS, (c) Female-only CTS GWAS with case and control numbers matched to those in the male-only GWAS. The dotted red line indicated genome-wide significance at $p=5 \times 10^{-8}$.

4.3.4 Two female-specific risk loci for CTS

The two novel female-specific loci were scrutinised, and the regional plots are shown in **Fig. 4.3**. The locus on 1q41 had two genome-wide significant SNPs (both intergenic), whereas the locus at 19q13.41 was a solitary genome-wide significant SNP (also intergenic). None of these SNPs were eQTLs for any genes in GTEx, and FUMA¹⁸¹ did not map any genes to either of these loci based on annotation and genomic position.

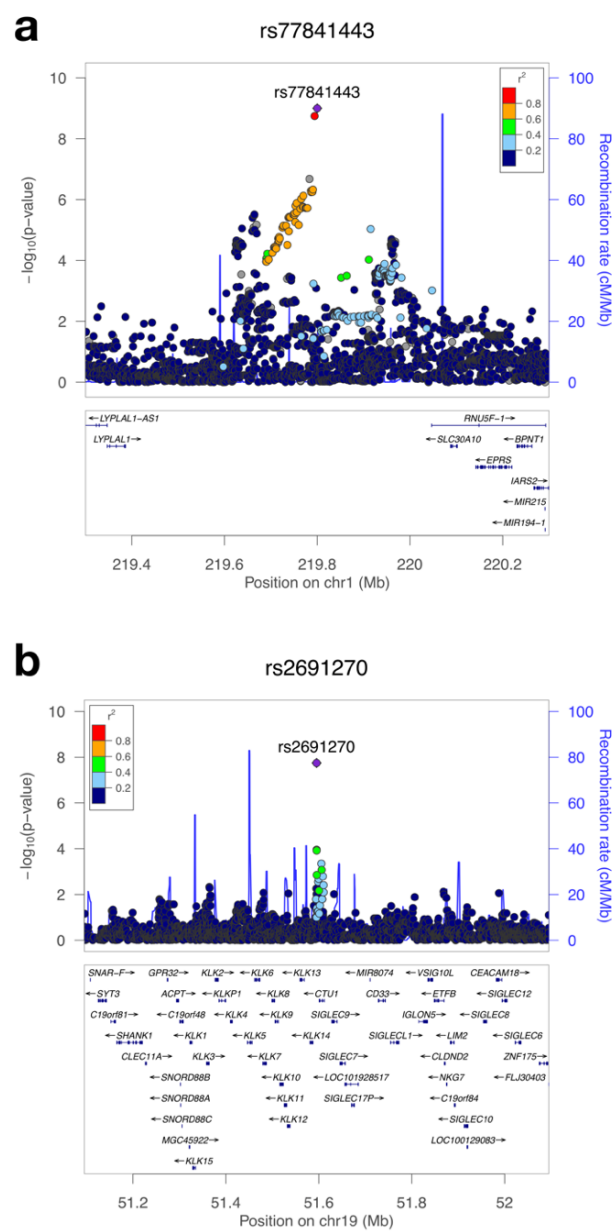


Fig. 4.3 Regional plots of novel loci discovered in the female-only CTS GWAS. (a) rs77841443 on 1q41, (b) rs2691270 on 19q13.41.

4.3.5 Mendelian randomisation analyses suggest a causal role for height, BMI, and age at menarche in CTS aetiology

The results of the various MR analyses are summarised in **Table 4.8**, and MR plots are shown in **Fig. 4.4**. The key findings were as follows:

1. Height is causally associated with CTS

Using the inverse variance-weighted (IVW) MR method, the MR estimate suggests that a 1 S.D. (equivalent to 9.24 cm) increase in genetically instrumented height is associated with an OR^{***} of 0.79 (95% CI: 0.74–0.83, $p = 2.24 \times 10^{-15}$) for CTS development. The weighted median method gave a similar OR of 0.80 (95% CI: 0.75–0.86, $p = 2.33 \times 10^{-9}$), while the MR-Egger analysis gave an OR of 0.72 (95% CI: 0.61–0.84, $p = 2.87 \times 10^{-5}$) and no evidence to support the presence of confounding by unbalanced genetic pleiotropy (intercept = 0.0030; 95% CI: -0.0016, 0.0076, $p = 0.21$). These effects persisted in a sensitivity analysis by removing five SNPs that were genome-wide significant in both the height and the CTS GWAS, which resulted in similar MR estimates. This analysis accords with the fact that CTS patients in UK Biobank of both sexes are on average ≥ 2 cm shorter than controls (Chapter 3, **Table 3.8**).

2. BMI is causally associated with CTS

The IVW method shows that a 1 S.D. (equivalent to 4.8 kg/m²) increase in genetically instrumented BMI is associated with an OR of 1.69 for the development of CTS (95% CI: 1.47–1.95, $p = 4.88 \times 10^{-13}$). The weighted median method yielded an OR of 1.86 (95% CI: 1.54–

^{***} The odds ratios per unit change in the exposure phenotype are derived by exponentiating the MR estimates in Table 4.8.

2.25, $p = 1.56 \times 10^{-10}$), and the MR-Egger analysis gave an OR of 2.44 (95%CI: 1.74–3.43, $p = 2.77 \times 10^{-7}$), with weak evidence to support a modest degree of pleiotropy (intercept = -0.011; 95%CI: -0.020, -0.002, $p = 0.02$). Again, the MR estimates accord with the fact that CTS cases of both sexes in UK Biobank have a BMI that is, on average, $>2.1 \text{ kg/m}^2$ greater than controls.

3. Age at menarche is causally associated with CTS

The IVW method demonstrates that a one-year increment in age at menarche (i.e. delayed onset of menarche) reduces the risk of CTS by an OR of 0.81 (95% CI: 0.75–0.88, $p = 2.39 \times 10^{-7}$), with similar estimates derived from the weighted median method (OR = 0.76; 95% CI: 0.68–0.84, $p = 6.07 \times 10^{-7}$). The pleiotropy-adjusted MR-Egger analysis gave an OR of 0.78 (95% CI: 0.63–0.97, $p = 0.026$) and no evidence to support the presence of confounding by pleiotropy (intercept = 0.0013; 95% CI: -0.0066, 0.0092, $p = 0.75$).

For all other phenotypes tested, there was no evidence to suggest a causal association with risk of CTS, with the exception of IVW in the diabetes MR, which suggested a very marginal effect on CTS risk, with a diagnosis of diabetes increasing CTS risk by an OR of 1.012 (95% CI: 1.008–1.016, $p = 5.68 \times 10^{-8}$).

Table 4.8 Summary of Mendelian randomisation analyses. This table shows the MR estimates for three different MR analyses for seven different phenotypes (plus one sensitivity analysis). MR estimates that are statistically significant at a Bonferroni-corrected p-value threshold are highlighted in bold italic.

<i>Phenotype</i>	<i>Method</i>	<i>MR Estimate</i>	<i>95% CI (lower)</i>	<i>95% CI (upper)</i>	<i>P-value</i>
<i>Height</i>	IVW	-0.241	-0.301	-0.182	2.24×10 ⁻¹⁵
	Weighted median	-0.217	-0.288	-0.146	2.33×10 ⁻⁹
	MR-Egger	-0.335	-0.492	-0.178	2.87×10 ⁻⁵
	MR-Egger intercept	0.003	-0.002	0.008	0.206
<i>Height (sensitivity)</i>	IVW	-0.225	-0.281	-0.168	9.88×10 ⁻¹⁵
	Weighted median	-0.209	-0.280	-0.138	7.86×10 ⁻⁹
	MR-Egger	-0.302	-0.452	-0.152	7.89×10 ⁻⁵
	MR-Egger intercept	0.002	-0.002	0.007	0.275
<i>BMI</i>	IVW	0.524	0.382	0.666	4.88×10 ⁻¹³
	Weighted median	0.622	0.431	0.812	1.56×10 ⁻¹⁰
	MR-Egger	0.892	0.552	1.232	2.77×10 ⁻⁷
	MR-Egger intercept	-0.011	-0.020	-0.002	2.03×10 ⁻²
<i>Diabetes</i>	IVW	0.012	0.008	0.016	5.68×10 ⁻⁸
	Weighted median	0.006	0.001	0.012	2.62×10 ⁻²
	MR-Egger	0.008	-0.002	0.018	0.110
	MR-Egger intercept	0.003	-0.005	0.011	0.412
<i>Rheumatoid arthritis</i>	IVW	2.84×10 ⁻⁴	-0.024	0.024	0.981
	Weighted median	-0.008	-0.036	0.019	0.546
	MR-Egger	-0.019	-0.053	0.015	0.280
	MR-Egger intercept	0.005	-0.001	0.010	0.127
<i>Hypothyroidism</i>	IVW	-0.032	-0.102	0.038	0.363
	Weighted median	0.039	-0.036	0.114	0.306
	MR-Egger	0.026	-0.155	0.207	0.781
	MR-Egger intercept	-0.008	-0.032	0.015	0.494
<i>Age at menarche</i>	IVW	-0.213	-0.293	-0.132	2.39×10 ⁻⁷
	Weighted median	-0.280	-0.390	-0.170	6.07×10 ⁻⁷
	MR-Egger	-0.246	-0.462	-0.029	2.63×10 ⁻²
	MR-Egger intercept	0.001	-0.007	0.009	0.749
<i>Age at menopause</i>	IVW	0.021	-0.018	0.061	0.289
	Weighted median	0.058	0.006	0.110	0.028
	MR-Egger	0.080	-0.012	0.173	0.090
	MR-Egger intercept	-0.012	-0.030	0.005	0.170

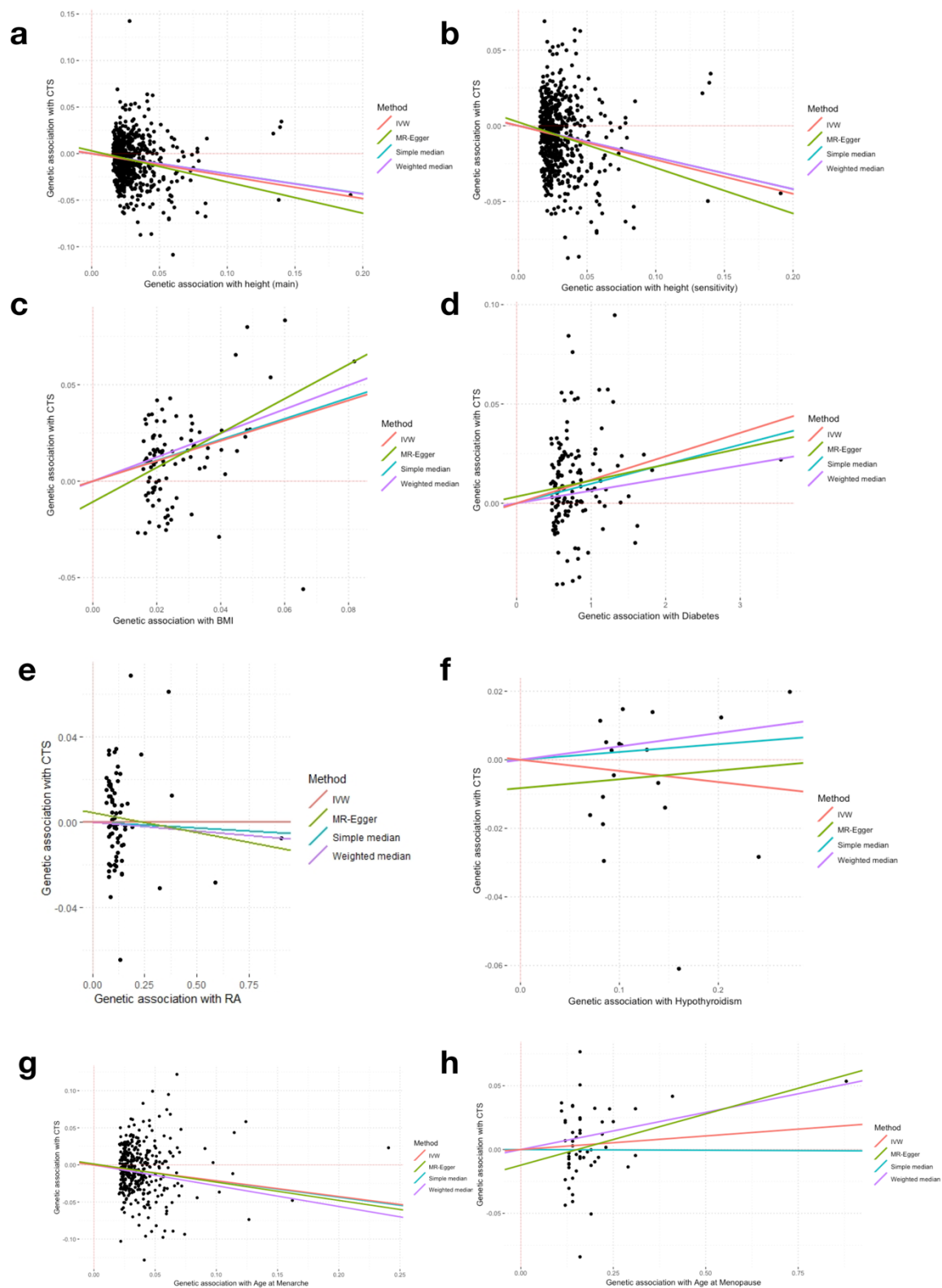


Fig. 4.4 MR plots for eight two-sample MR analyses of CTS. Analyses were for CTS as the outcome and the following phenotypes as the exposure: (a) height, (b) height (sensitivity analysis), (c) BMI, (d) diabetes, (e) rheumatoid arthritis, (f) hypothyroidism, (g) age at menarche, (h) age at menopause.

4.3.6 Weak evidence of genetic correlation between migraine and CTS

The genetic correlation study between the CTS GWAS and the Gormley et al. migraine GWAS yielded a genetic correlation coefficient (r_g) of 0.080 (8%), Z-score 1.62, $p = 0.11$.

Cross-phenotype analysis with CPASSOC identified three SNPs at 9q33.1 that were genome-wide significant when the two GWAS summary statistics were meta-analysed (**Table 4.9**). Regional plots of this associated region in both the CTS and migraine GWAS are shown in **Fig. 4.5**, and the GTEx gene expression of *ASTN2*, a possible causal gene at this locus as identified by Freilinger et al.²⁷⁷, is shown in **Fig. 4.6**.

Table 4.9 Significant SNPs on chromosome 9 in CTS-migraine cross-phenotype analysis. Summary statistics from CPASSOC are shown on the left (EAF = effect allele frequency in UK Biobank). The two boxes on the right show the beta-coefficients and p-values in for these three SNPs in the CTS GWAS and the migraine GWAS.

SNP	Position	P-value (CPASSOC)	Effect allele	EAF	Beta (CTS)	P-value (CTS)	Beta (migraine)	P-value (migraine)
<i>rs1040851</i>	119243634	3.29×10^{-8}	C	0.42	0.0012	1.40×10^{-3}	0.053	5.51×10^{-7}
<i>rs6478241</i>	119252629	1.61×10^{-8}	A	0.36	0.0012	2.10×10^{-3}	0.059	5.95×10^{-8}
<i>rs34979631</i>	119255297	3.73×10^{-8}	T	0.23	0.0016	4.50×10^{-4}	0.072	5.97×10^{-9}

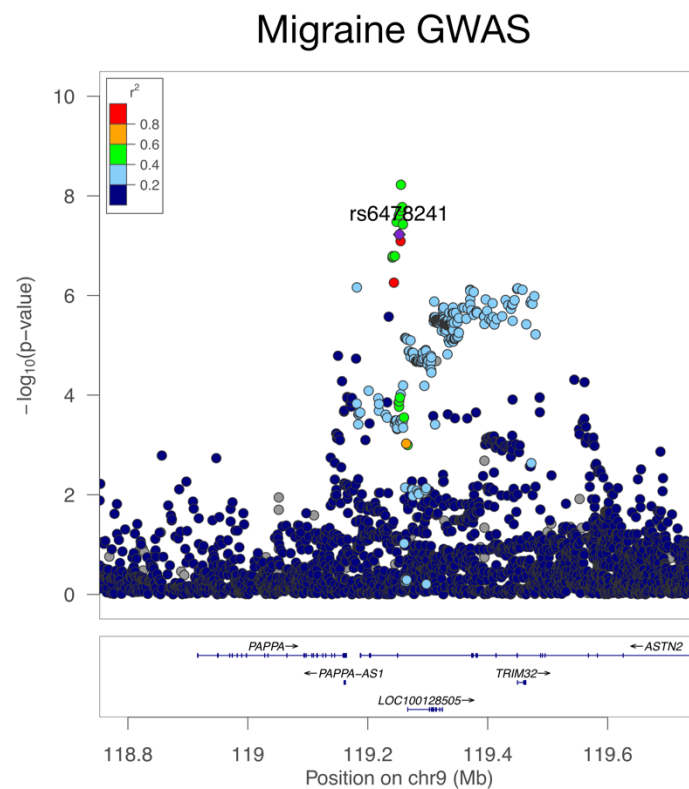
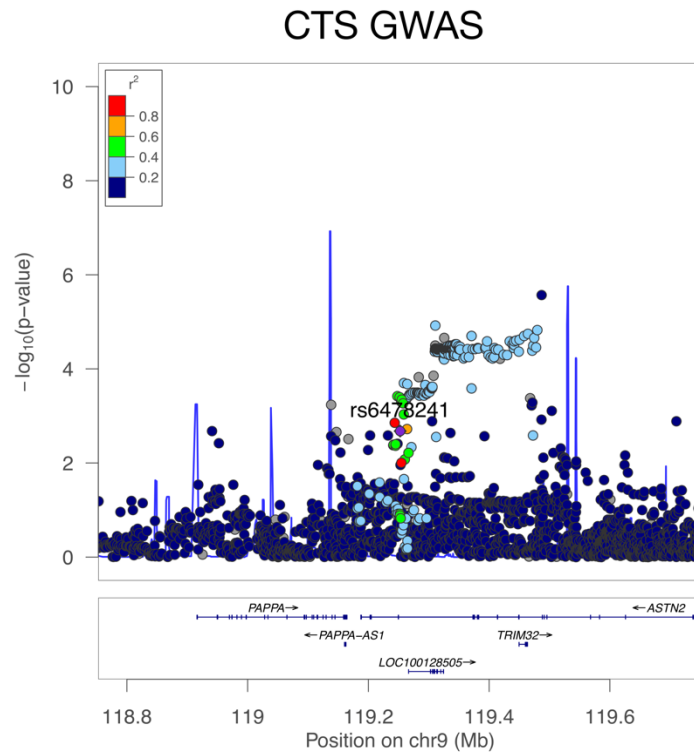


Fig. 4.5 Region of migraine and CTS cross-association. Regional plots for areas of cross-phenotype overlap between the migraine and CTS GWAS are shown, plotted using the summary statistics from the respective GWAS. Both plots are centred on rs6478241 – the SNP that was most significantly associated with both phenotypes in the CPASSOC analysis.

4.4 Discussion

The central purpose of all medical research is to elucidate disease biology and identify opportunities for new treatments and prevention. The size and scale of UK Biobank resource offers researchers a unique and unprecedented opportunity to investigate disease epidemiology and its genetic basis, and the various analyses performed in this chapter have taken advantage of the statistical power conferred by the size of UK Biobank to investigate the relationships between CTS and a multitude of phenotypes.

The first question posed was whether the UK Biobank dataset could replicate previously reported epidemiological associations between CTS and four different disease phenotypes. A series of nested case-control studies were performed in order to answer this question, and a matching algorithm was devised to match individuals of each phenotype of interest to controls at a ratio of 1:5, using two different sets of matching variables that included or excluded BMI.

A nested case-control study is also known as a “case-control in a cohort” design, whereby cases and controls are drawn from the population of a fully enumerated cohort such as UK Biobank. For each case within the cohort, a specified number of matched controls is selected (who have not yet developed the disease by the time of disease onset in the case)²⁷⁸. This method offers several advantages to the traditional case-control design, such as being able to select cases and controls from the same underlying population, the ability to negate potential confounding variables through matching, and the significant economic and logistical benefits of using a population-based cohort such as UK Biobank.

The LD score correlation analyses in Chapter 3 demonstrated that out of all phenotypes tested, the greatest genetic correlation was between CTS and BMI ($r_g=0.346$, $p=5.8\times 10^{-23}$), with eight

out of the top ten correlation observed between CTS and different measures of adiposity. As such, BMI is likely to play a significant role in CTS pathophysiology, hence its inclusion as an additional matching variable in half of the nested case-control datasets. The ≥ 2.1 kg/m² higher BMI in UK Biobank CTS cases vs controls accords with findings from previous (considerably smaller) studies that found that BMI was greater by: 2.0, 2.7, 4.19, and 2.95 kg/m² in CTS cases compared to controls respectively^{279–282}.

In males and females combined, the OR for developing CTS in diabetes, rheumatoid arthritis (RA), and hypothyroidism were 2.31, 2.70, and 1.47, respectively. In all three diseases, matching on BMI as an additional variable in generating the case-control datasets attenuated this effect, with ORs falling to 1.75, 1.75, and 1.35, respectively. This finding is consistent with the notion that raised BMI associated with these three conditions is in part responsible for the increased risk of CTS, and the causal effect of BMI on CTS will be discussed in greater detail below. Whereas diabetes in females conferred a greater risk for developing CTS than in males (OR 2.63 in females vs 1.99 in males), in RA and hypothyroidism, the opposite effect was seen (2.46 in females vs 3.75 in males for RA; 1.46 in females vs 1.61 in males for hypothyroidism). This may in part be an artefact of the relative over-representation of females in the RA and hypothyroidism case-control datasets (67.6% and 83.2%, respectively), in conjunction with the fact that females CTS cases outnumber males by 2:1 in UK Biobank. Within the cohort of males with RA and hypothyroidism, there were only 155 and 134 cases of CTS, respectively (c.f. 431 and 1,144 female CTS cases), suggesting that the OR estimates may have been subject to sparse data bias²⁸³, thus artificially inflating the odds ratios.

As with all epidemiological studies, accurate case ascertainment is key to minimising bias and error, and one weakness of this particular analysis was that the phenotyping of the disease variables was not nearly as stringent as that for CTS (where an independent validation study

was undertaken to ascertain the positive predictive value of an individual designated case being a true case). Moreover, CTS case ascertainment relied heavily on surgical codes (thus capturing the extremes of the phenotype); similar surgical codes are not available for the four medical diagnoses. Another limitation of the nested case-control approach as employed here is that it does not allow for the investigation of temporality of the diagnoses. An alternative method is to perform a cohort analysis with the comorbid diseases as baseline variables and incident CTS diagnosis as the outcome in a Cox proportional hazards regression, using data on the date of assignment of secondary care and self-reported diagnostic codes within the UK Biobank dataset. UK Biobank has proven to be a fertile resource for undertaking such studies^{284,285}.

Nevertheless, this analysis, which to the best of my knowledge is the largest series of case-control studies examining the effects of these disease phenotypes on CTS, have provided strong evidence to confirm the associations between CTS and diabetes, RA and hypothyroidism. The relationship between migraine and CTS will be discussed separately at the end of this section.

Given these robust observed associations between CTS and the three disease phenotypes, it may at first seem intriguing that the polygenic risk score (PRS) analyses did not demonstrate a greater polygenic risk for CTS in diabetics compared to non-diabetics, nor in RA cases compared to non-RA controls, suggesting that sufferers of these diseases do not have a greater inherent genetic liability for CTS. The only notable discrepancy in PRS observed was in hypothyroid cases vs non-hypothyroid controls, which showed a very modest separation in PRS of 1.573 vs 1.565, principally in females. There are several ways in which one could interpret these findings. A ‘mechanical’ interpretation might be that the lack of a raised genetic risk for CTS in these diseases provides evidence that the increased risk is mediated by non-genetic factors. For example, in the case of RA, it may be the proliferation of tenosynovium around the flexor tendons within the carpal tunnel that raises CTS risk through a purely mechanical effect.

A similar explanation has been posited for hypothyroidism, whereby myxoedema-associated swelling of the sub-synovial connective tissues (SSCT) raises pressure within the carpal tunnel¹⁰⁴. In the case of diabetes, it may be the glycosylation of collagen fibres that stiffen and alter the compliance of the SSCT²⁸⁶, or perhaps within the median nerve itself. While these speculations may have some merit, and there is a temptation to reconcile these pre-existing hypotheses with the negative findings in this PRS analysis, the ‘inherently genetic’ vs ‘non-genetic/mechanical’ explanations of CTS aetiology present somewhat of a false dichotomy, and the true answer is likely to be more nuanced. We may recall from the previous chapter that rheumatoid arthritis and type 2 diabetes showed moderate and nominally significant genetic correlations with CTS (RA: $r_g = 0.113$, $p = 9.8 \times 10^{-3}$; diabetes $r_g = 0.141$, $p = 0.011$).

A more parsimonious explanation might be found in the conservative methodology employed in constructing the risk score. The weighted PRS was derived from the 16 index SNPs that were genome-wide significant in the CTS GWAS, and therefore all sub-threshold association signals were ignored, many of which are likely to be true positive associations and thus contribute to the genetic architecture of CTS¹⁶⁸. Other less stringent methods for constructing polygenic risk scores exist, whereby the p-value thresholds are set at lower values, such as 1×10^{-5} , or a range of thresholds are used²⁸⁷. Given that the associated SNPs for CTS were derived from a single-stage GWAS without replication, it was felt prudent to recruit only the significantly associated SNPs to build the PRS, at a cost of losing power when it is used as an analytic tool. The PRS analysis in Chapter 3 demonstrated that the separation in PRS between CTS cases and controls is relatively modest (1.620 vs 1.566), so the absence of a raised PRS in individuals of CTS-associated disease phenotypes might have been expected.

The sex-stratified GWAS demonstrated evidence of sex differences in terms of the magnitude of the genetic associations at the 16 loci, albeit with minimal evidence of heterogeneity

between the sexes (**Table 4.7**). It seems reasonable to conclude that females exhibit somewhat stronger evidence of genetic architecture for CTS than males. This was suggested in the comparison of the male-only GWAS with the size-matched female-only GWAS (**Fig. 4.2a** vs **Fig. 4.2c**), where despite the power discrepancy owing to the greater number of female CTS cases having been abrogated, the p-values in the matched female-only GWAS were lower than in males (**Table 4.7**). There is a stochastic element to this analysis of males vs matched females, so some caution is required in its interpretation. However, the interpretation that females are, on the whole, more genetically liable to develop CTS than males accords neatly with the near-universal observation that CTS is more common in females.

It is worth noting here that the only study of heritability in CTS was performed in a cohort of female twin-pairs¹¹⁶. The heritability estimate of 0.46 is therefore specific to females, and there is no estimate in males. Several highly powered studies of sex-dependent liability have found that heritability estimates for most human phenotypes are largely similar between males and females²⁸⁸, although a large phenome-wide heritability study in UK Biobank found that in phenotypes that have sex differences in heritability, females consistently exhibit higher heritability estimates than males²⁸⁹.

The two novel loci discovered in the full female-only GWAS are difficult to interpret. In the case of the 1q41 locus, there are two genome-wide significant SNPs in an intergenic region, with a cluster of suggestively associated SNPs in moderate LD with them just downstream (**Fig. 4.3a**). Although no eQTLs could be found within this group of SNPs, of potential relevance is the fact that this cluster of SNPs is situated 3-400 kb away from *LYPLAL1* (lysophospholipase-like 1) and its antisense counterpart, *LYPLAL1-AS1*. *LYPLAL1* is a triglyceride lipase believed to be up-regulated in the subcutaneous adipose tissue of obese individuals²⁹⁰, and has been implicated in a GWAS meta-analysis of waist-hip ratio²⁹¹. The 19q13.41 locus, on the other

hand, comprises a single genome-wide significant SNP, with a small peak of nominally associated SNPs in low LD underlying it (**Fig. 4.3b**). The region spanning ± 500 kb of these SNPs is highly gene-dense, and resolving SNP-gene interactions in this region is difficult in the absence of promising functional candidate SNPs. The possibility that this lone SNP might represent a false positive association must be considered.

Chapter 1 gave a summary of how female reproductive phenotypes (pregnancy, menopause and female sex hormone levels) might influence CTS risk, with particularly compelling evidence for the role of hypooestrogenaemia in CTS aetiology. This theory was garnered from four different observations: oral contraceptive use, bilateral oophorectomy, and aromatase inhibitor use all increase CTS incidence, whereas the opposite effect is seen with hormone replacement therapy. Part of the reason for undertaking the sex-stratified GWAS was to enable the interrogation of the potential causal role of two female reproductive phenotypes on CTS risk: age at menarche and age at menopause.

The two-sample MR analysis of age at menarche and CTS found strong evidence to suggest that increasing age at menarche is protective in CTS. All three MR analyses gave robust estimates, with a one-year delay in genetically instrumented age at menarche being associated with an odds ratio of between 0.76 (weighted median) to 0.81 (IVW) for development of CTS. The MR-Egger intercept did not suggest confounding by pleiotropy. Increasing age at menarche is associated with a shorter lifetime reproductive period: a retrospective cohort study of over 300,000 women found that females with age at menarche ≤ 9 years had reproductive period that was 9 years longer compared to females whose age at menarche was ≥ 17 years²⁹². Is a shorter reproductive period (with less cumulative exposure to endogenous oestrogen) therefore associated with reduced risk of CTS? This would run counter to the

hypooestrogenaemia-CTS hypothesis posited above, which implies a protective effect of oestrogen on CTS risk.

The MR analysis of age at menopause had fewer than one-sixth the number of genetic instruments than the menarche MR, which may explain why the results were inconclusive. The weighted median MR was nominally significant, and suggested that a one-year delay in the onset of menopause is associated with an OR for CTS development of 1.06 (95% CI: 1.01-1.12; $p=0.028$). Although the IVW and MR-Egger gave disparate (and non-significant) MR estimates, they were directionally concordant with the weighted median – i.e. later menopause is associated with increased CTS risk. If we accept this result, it is again counterintuitive in the context of the hypooestrogenaemia-CTS hypothesis.

Clearly the interplay between female sex hormones, reproductive lifespan and CTS is highly complex, and will not be resolved by these two MR analyses alone. Moreover, it is prudent to consider other variables through which the effects of early menarche on CTS risk might be mediated. Earlier onset of menarche has consistently been associated with adverse health outcomes, including obesity, type 2 diabetes and cardiovascular disease^{293,294}, as well as shorter height in adulthood²⁹⁵. A pan-European case-control study of nearly 30,000 participants found that each year's delay in the onset of menarche was associated with a 0.32 kg/m² lower BMI in adulthood²⁹⁴. Furthermore, there is a body of evidence linking high BMI with a later onset of menopause²⁹⁶. Therefore, both of the causal effects suggested in the female reproductive MR analyses (i.e. CTS risk being associated with early menarche and late menopause) may be mediated by BMI, height, or by both. BMI, along with height, were the two other MR analyses in which strong causal effects on CTS were demonstrated, and we will consider each of these in turn.

All three MR analyses for BMI gave robust MR estimates, with each S.D. increase in BMI corresponding to ORs for development of CTS ranging from 1.69 (IVW) to 2.44 (MR-Egger). While these estimates are significant and strongly suggest a causal role of BMI in CTS pathogenesis, the mechanism by which this occurs remains to be elucidated. Increased adiposity has been posited to increase the volume of fat in the carpal tunnel, although evidence for this is weak²⁸⁰. Another hypothesis centres around obesity-related metabolic syndrome, which can lead to glycosylation of extracellular proteins and cause fat deposition in nerves, leading ultimately to nerve dysfunction²⁴⁵.

A further possibility is that increased BMI mediates its effect on CTS through diabetes. The estimates from the diabetes-CTS MR analysis were modest (IVW: OR of 1.012, $p = 5.68 \times 10^{-8}$), and thus considerably lower than the observational OR for diabetes in UK Biobank. The diabetes-CTS MR analysis suffered from a methodological shortcoming in the form of sample overlap: the diabetes GWAS meta-analysis included approximately 16,000 cases and 100,000 controls derived from UK Biobank, meaning that there was a sample overlap of approximately 17% between the two GWAS cohorts. However, the significance of any bias resulting from this overlap is questionable – it has been shown that such sample overlap tends to bias the MR estimate towards the observational association rather than towards the null²⁹⁷ (leading to an inflation of the estimate in this particular case). Given the near-null OR derived from the IVW analysis, the interpretation of the results would remain unchanged even if there had been some inflation of the MR estimate due to sample overlap. One method to mitigate this bias induced by sample overlap is to create a set of GWAS summary statistics for CTS by performing an association study in a subset of participants that excludes all diabetics; the revised IV-outcome effect estimates will reduce the type 1 error rate in a Mendelian randomisation analysis of diabetes on CTS.²⁹⁷

Further studies that may prove helpful in untangling the complex relationship between CTS and diabetes/BMI would include MR analyses of other measures of adiposity (such as waist-to-hip ratio), metabolic variables (such as circulating lipids and inflammatory markers), in addition to a multivariable MR analysis in females with BMI, diabetes and age at menarche as exposures.

The height-CTS MR provided strong evidence for an inverse causal effect of height on CTS status, with estimates of the OR for developing CTS per 1 S.D. increase in height ranging from 0.72 (MR-Egger) to 0.80 (weighted median). These findings persisted in the sensitivity analysis, providing further evidence for their validity. It must be noted that it may not be height *per se* that is inversely causal in CTS; rather, it is more likely that height is a proxy for altered skeletal growth, which manifests in the form of altered hand and wrist dimensions that predispose to CTS that were discussed in Chapter 1. Whether there is a direct correlation between height and wrist index remains to be investigated. The notion of altered skeletal growth contributing to CTS predisposition is corroborated by the genetic correlation analyses in the previous chapter, with strong correlations observed between CTS and several height phenotypes, as well as bone mineral density. Furthermore, the finding that osteoblasts were the top enrichment in the partitioned heritability analysis across multiple tissue and cell types lends further credence to this hypothesis.

The final analyses in this chapter investigated the epidemiological associations between migraine and CTS. Within UK Biobank, the overall odds ratio for association with CTS in migraine patients was 1.14 (95% CI: 1.04-1.25). In the sex-stratified analyses, the odds ratio was greater in females (1.15; 95% CI 1.04-1.288) than in males (1.07; 95%CI 0.82-1.40). By matching migraine cases to controls on BMI in addition to age and sex, the odds ratio became

marginally more significant in females (1.17; 95%CI 1.06-1.29). Thus, there is reasonably strong evidence to suggest a modest association between these two phenotypes in females only.

The one study that has reported an association between migraine and CTS was in fact reporting an association between “migraine *or* severe headache” and CTS, a limitation that was acknowledged by the authors²⁶². Out of 25,880 respondents in the study by Law et al., 3.7% (952) had a self-reported diagnosis of CTS and 4,212 (16.3%) had migraine (according to their broad case definition). Within the UK Biobank post-QC cohort of 401,656 from which the nested case-control datasets were derived, 3.1% (12,312) had a CTS diagnosis and 3.6% (14,409) had a migraine diagnosis. There is a significant discrepancy in the migraine prevalence between the two studies. Based on a systematic review of 302 studies involving over 6 million study subjects worldwide, the global prevalence of migraine has been estimated at approximately 10%²⁹⁸, with estimates varying from 2.6% to 21.7%²⁹⁹. Thus, while both migraine prevalence estimates are within the range of previously reported studies, the UK Biobank prevalence may be an underestimate, while that of Law et al. may be an overestimate, in part due to the case ascertainment method. This discrepancy probably explains to a large degree the very different odds ratios for CTS found in the two studies (1.14 vs 2.67), and further investigations are warranted to better establish this epidemiological link.

Evidence for a genetic correlation between migraine and CTS was weak ($r_g = 0.080$, $p = 0.11$). As mentioned, it was not possible to obtain the full set of summary association statistics from the Gormley et al. migraine GWAS meta-analysis, so the summary statistics were derived from a cohort with half of the cases and controls redacted, which resulted in a significant loss of power. Thus, it is possible that a greater genetic correlation estimate would be obtained if this study were to be repeated on the full set of migraine GWAS summary statistics. To put the observed genetic correlation of 8% in context, rheumatoid arthritis and ulcerative colitis are

two diseases that have previously been linked in several epidemiological and experimental observations, and the two diseases share many common biological pathways involved in autoimmune disease³⁰⁰. Bulik-Sullivan et al. found a genetic correlation of 9% between the two diseases, which did not reach statistical significance¹⁹⁴. Nevertheless, there are significant therapeutic implications to this modest degree of shared genetic architecture, in that TNF- α blocking agents are now used to treat both RA and ulcerative colitis³⁰¹. Thus, the magnitude of the genetic correlation is not necessarily a reflection of the importance of shared biology, especially in a therapeutic context.

The cross-phenotype association study between CTS and migraine found one region of genome-wide significance across the two studies, at 9q33.1 (**Table 4.9**; **Fig. 4.5**). This locus was genome-wide significant in the Gormley et al. GWAS, and its association with migraine was originally discovered in an earlier GWAS that identified *ASTN2* as the likely causal gene at this locus²⁷⁷. *ASTN2* is a member of the astrotactin gene family, which are membrane proteins believed to play a role in neuronal migration and neurodevelopment, specifically in the development of the laminar architecture of the cortex^{302,303}. Transcriptomic studies across multiple tissue types in GTEx show that this gene is expressed most highly in the brain, but it is also expressed highly in tibial nerve, suggesting that it may also play a role in the peripheral nervous system (**Fig. 4.6**). The three SNPs identified in the CPASSOC analysis are all eQTLs for *ASTN2* (albeit in testis and not in brain or nerve).

The potential contribution of central nervous system dysfunction to CTS pathophysiology was discussed in Chapter 1, including the evidence suggesting a role played by central sensitisation in CTS pain¹⁰⁶. The headache phase of migraine is believed to be triggered by the activation of nociceptors in the trigeminal nerve that innervate blood vessels in the meninges (the trigeminovascular system)²⁵⁰, and repeated activation of this system eventually leads to central

sensitisation, which is the hallmark of chronic migraine³⁰⁴. In fact, injection of Botulinum toxin A to multiple sites in the head and neck^{§§§} for treatment of chronic migraine is widely practiced, and has been endorsed by the National Institute for Clinical Excellence in the UK since 2012³⁰⁵. The toxin is believed to attenuate neuropeptide and neurotransmitter exocytosis in peripheral sensory neurons in the head and neck, thereby reducing central sensitisation^{306,307}. Migraine has a similar heritability estimate to CTS, at 42%³⁰⁸, and it is interesting to postulate that the areas of overlap identified in the genetic correlation studies represent a shared genetic propensity towards central sensitisation, possibly through peripheral nerve perturbations. However, both diseases are known to have significant pain and psychological comorbidities^{109,251}; thus, the possibility that the genetic overlap corresponds to a propensity towards a pain phenotype in both CTS and migraine must also be considered.

In summary, this chapter has shed considerable light on the complex relationship between CTS and its associated phenotypes by examining them through the lens of genetic epidemiology. The following chapter will stray from the “big data” approach that has been central to the last three chapters, and instead will bring the lens into sharper focus to allow the genetics of CTS to be examined at a much higher resolution.

§§§ Peripheral nerve decompression surgery in migraine targets the occipital nerves and branches of the trigeminal nerve in these very areas, and are purported to offer a more permanent resolution to migraine symptoms.

Chapter 5

From genotype to gene expression and phenotype

5.1 Introduction

The judicious use of big data lies at the very core of modern genetic association studies, and the previous chapters have investigated the genetics of CTS by leveraging the very large number of CTS cases and controls present in UK Biobank. The summary association statistics generated in Chapter 2 were the basis for identifying candidate genes and biological pathways relevant to CTS in Chapter 3, and for performing an array of genetic epidemiological investigations in Chapter 4.

The focus will now shift from a large population-based cohort to a much smaller, albeit deeply phenotyped, cohort. In this final results chapter, I will use combined genotypic and phenotypic data from two cohorts of CTS patients in Oxford to investigate potential correlations between the genetic liability for CTS with more granular phenotypes that are relevant to CTS pathology. I will also attempt to identify which of the previously identified CTS candidate genes are under- or over-expressed as a result of variations in genotype at the genome-wide significant SNPs.

5.1.1 From genotype to phenotype

In defining the central dogma of molecular biology, Francis Crick wrote, “*the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible*”³⁰⁹. That is to say that the blueprint for all life is written in DNA, which is replicated with every cell division, transcribed into messenger RNA, and translated into proteins. The reverse of this sequence – proteins being used as a template for the synthesis of RNA or DNA – is not known to occur.

Biological phenotypes are therefore in theory the ultimate manifestation of this sequence of DNA replication-transcription-translation, but this only tells part of the story. Phenotype is principally determined by the interaction between two variables – the genotype carried by each individual, and the environment to which they are exposed. To add to the complexity, most phenotypes are not determined by single alleles, but by the additive effects of multiple alleles at multiple loci³¹⁰. Furthermore, all biological systems are at the mercy of stochastic factors at the molecular level³¹¹. If we add epigenetics to the mix, it becomes readily apparent that predicting phenotype from genotype, one of the central aspirations of personalised medicine, is an enormously challenging task. In spite of some recent promising results¹⁴⁵, the utility of polygenic risk estimation is currently limited¹⁴³. The analyses in this chapter will seek to explore the degree to which the polygenic risk score (PRS) for CTS developed in Chapter 3 is relevant to an array of measures that characterise the CTS phenotype.

5.1.2 Focus on the sub-synovial connective tissues (SSCT)

The second half of this chapter will focus on gene expression. In Chapter 1, I summarised the considerable body of evidence to suggest that molecular aberrations in the SSCT are

contributors to CTS pathophysiology, and the bioinformatic analyses in Chapter 3 demonstrated that the CTS GWAS summary association statistics were strongly enriched for extracellular matrix (ECM)-associated genes. The SSCT are dense in ECM components such as collagen, elastin and proteoglycans³¹²; it therefore follows logically that the SSCT should be the natural starting point for studying gene expression in CTS^{****}. An added benefit is that this tissue, in the form of tenosynovium surrounding the flexor tendons within the carpal tunnel, is relatively easy to harvest during carpal tunnel surgery. The presence of oestrogen and progesterone receptors has been immunohistochemically demonstrated in the tenosynovium of CTS patients^{313,314}; thus, a study of gene expression in the SSCT also presents an opportunity to verify this finding and to revisit the female hormonal hypothesis of CTS.

5.1.3 RNA-Sequencing

Over the last decade or so, the analysis of the transcriptome^{†††} has gradually shifted from the use of hybridisation-based microarrays to RNA-Sequencing (RNA-Seq). This has been made possible by advances in high-throughput next-generation sequencing technologies and bioinformatics, in conjunction with a massive reduction in the cost of such sequencing³¹⁵. One of the predominant advantages of RNA-Seq over microarrays is that microarrays are reliant on pre-designed detection probes to be present on the array, so can only detect transcripts for which the sequence is already known. In contrast, RNA-Seq can detect novel, unannotated genes, meaning that RNA-Seq studies can be updated as new information about the transcriptome becomes available. Insofar as RNA-Seq is free of any presuppositions, one could argue that it shares some philosophical similarities with GWAS. Other advantages of RNA-Seq

**** The median nerve is probably the most informative tissue in which to study gene expression in CTS, but ethical considerations preclude even resecting a small portion of the nerve for the purpose of such studies.

††† A transcriptome is defined as the full range of mRNA molecules expressed by an organism or a cell.

include the ability to detect a higher proportion of differentially expressed genes³¹⁶ and improved detection of alternatively spliced transcripts³¹⁷.

Although there are several methods for performing an RNA-Seq experiment, the basic principles are as follows: RNA is isolated and (usually) filtered using poly(A) tails to keep only the mRNA. This RNA is fragmented to the appropriate length for the sequencing machine, and reverse transcribed to cDNA to create a cDNA library. Short oligonucleotide adapter sequences are ligated to the cDNA, which is then amplified by PCR. This DNA can then be sequenced using a high-throughput sequencer to produce hundreds of millions of reads, which are mapped to a reference genome. Prior to commencing data analysis, quality control is as important in RNA-Seq as it is in GWAS.

No transcriptomic study of CTS tissues has hitherto been undertaken, so this chapter provides an excellent opportunity to study the intermediate step in the central dogma between DNA and protein, in carpal tunnel tenosynovium.

5.1.4 Aims

The specific aims of this chapter are as follows:

- 1) Perform whole-genome genotyping and imputation in the Oxford CTS cohort.
- 2) Compare PRS between CTS cases and healthy controls in the Oxford cohort by constructing a polygenic risk score (PRS).
- 3) Correlate PRS with various phenotypic measures relevant to CTS in the Oxford cohort.
- 4) Perform an RNA-Seq experiment using RNA extracted from surgically resected tenosynovium from CTS patients in the Oxford cohort.

- 5) Investigate whether candidate genes identified in Chapter 3 are expressed in carpal tunnel tenosynovium, and quantify their expression levels.
- 6) Investigate whether oestrogen and progesterone receptor genes are expressed in carpal tunnel tenosynovium.
- 7) Investigate if genotypes at the index SNPs identified in the GWAS are correlated with the expression levels of genes mapped to those loci.

The fundamental hypothesis underlying these aims is that the genome-wide significant SNPs discovered in UK Biobank are still of relevance in a considerably smaller, independent cohort. Thus, the PRS is expected to be greater in the Oxford cases compared to the controls, and PRS is expected to correlate with certain phenotypic variables relevant to CTS. The top candidate genes identified in Chapter 3 (*EFEMP1*, *ADAMTS17* and *ADAMTS10*) are expected to be highly expressed in carpal tunnel tenosynovium, and female sex hormone receptor genes are expected to be expressed in tenosynovium. Genotypes at the index SNPs are expected to correlate with expression levels of at least one of the genes mapped to each locus, with the exception of *ADAMTS17* and *ADAMTS10*, in which the identified GWAS variants confer coding changes and are therefore unlikely to manifest as a change in gene expression levels.

5.2 Methods

5.2.1 Description of Oxford cohorts

The gene expression and genotype-phenotype correlation studies in this chapter were performed in participants recruited to two ethically approved and related studies based in Oxford: PiNS (10/H07056/35) and MGCTS (16/LO/1920). The cohorts are described below.

5.2.1.1 PiNS

The Pain in Neuropathy Study (PiNS) is a cross-sectional phenotyping/genotyping study concerned with patients suffering from a range of peripheral neuropathies. The broad aim of the study is to elucidate correlates between pain symptomatology with intra-epidermal nerve fibre density, nerve conduction studies, psychological co-morbidities, measures of quality of life and functionality, and genotype/gene expression³¹⁸. A CTS sub-study of PiNS³¹⁹ has thus far recruited approximately 120 CTS patients with CTS, and approximately 30 healthy age- and sex-matched controls. The CTS cases had clinically¹¹ and electrodiagnostically³²⁰ confirmed CTS and were recruited from the surgical waiting lists at Oxford University Hospitals NHS Foundation Trust. Exclusion criteria included electrodiagnostic evidence of abnormalities other than CTS, a history of upper limb or neck trauma, cervical radiculopathy, pregnancy or diabetes.

The majority of patients underwent carpal tunnel decompression surgery, and most had the following tests and investigations, both pre- and post-operatively:

- Quantitative sensory testing in the hand

- Electrodiagnostic tests
- Completion of multiple symptom and function questionnaires
- Collection of blood
- Skin biopsy from the median-nerve innervated territory of the index finger (the radio-volar aspect overlying the proximal phalanx)

A sub-cohort within the CTS group have also undergone collection of tissues at the time of carpal tunnel decompression surgery (**Fig 5.1**). The tissues collected are:

1. A strip of palmar skin
2. A piece of the transverse carpal ligament
3. Tenosynovium surrounding the flexor tendons within the carpal tunnel
4. Connective tissues surrounding the median nerve

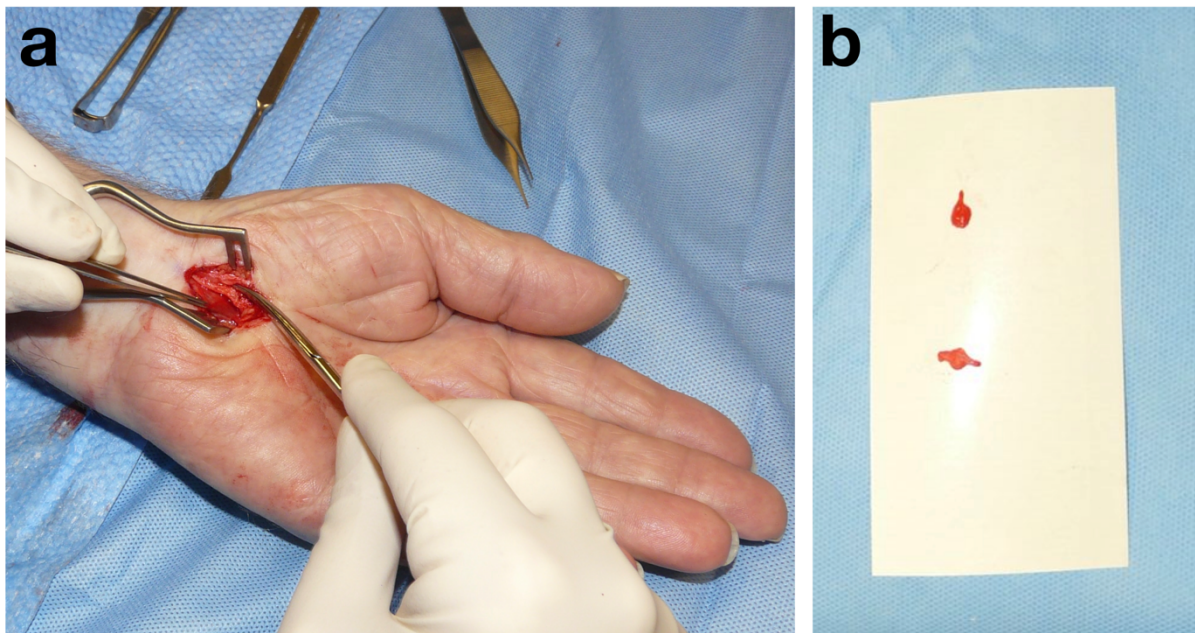


Fig. 5.1 Collection of surgical specimens during carpal tunnel decompression surgery. (a) A typical carpal incision, showing the divided transverse carpal ligament. (b) Resected tenosynovial specimens. Photograph provided by Professor Dominic Furniss.

5.2.1.2 MGCTS

The Molecular Genetics of Carpal Tunnel Syndrome (MGCTS) is a cohort study closely aligned to PiNS, designed to supplement the number of surgical CTS cases. The study was conceived with the specific aim of collecting more surgical tissues to confer greater statistical power to the genotype-specific gene expression studies described in this chapter. Recruitment is ongoing, and at the time of writing, c. 70 participants have been recruited. MGCTS is an entirely surgical CTS cohort, and participants are recruited through hand surgery clinics in Oxford at the time when they are listed to undergo carpal tunnel decompression surgery.

Unlike PiNS, not all of the participants have undergone electrophysiological testing, and the CTS status of MGCTS participants is instead based on a clinical diagnosis of CTS made by one of the consultant hand surgeons at the Nuffield Orthopaedic Centre. Patients recruited to the study complete an abbreviated set of demographic, symptom and function questionnaires that are administered to PiNS participants. Blood and surgical tissues (but no index finger biopsies) are collected as for PiNS.

5.2.2 PiNS phenotyping

Phenotyping of the entire PiNS cohort was performed by a single examiner (Dr Annina Schmid), and details of the methodology for the CTS PiNS cohort have been described previously³¹⁹. The methods relevant to the phenotypes analysed in this chapter are briefly outlined below.

5.2.2.1 Questionnaires

In addition to the collection of basic demographic, medical, and family history data, patients completed the visual analogue scale (VAS) to quantify the level of pain experienced in the past 24 hours on a 0-10 scale (0 indicating no pain and 10 indicating the worst pain ever). They also completed the Boston Carpal Tunnel Questionnaire³²¹, which assesses both the symptomatic and functional severity of CTS, with scores ranging from zero (no symptoms or disability) to 5 (severe symptoms or disability). For surgical patients, the same questionnaires were administered post-operatively at 3 months (N=5), or 6 months (N=66).

5.2.2.2 Electrodiagnostic tests

Electrodiagnostic testing (EDT) was undertaken using the ADVANCE system (Waltham, MA, USA), with hand temperature standardised to $>31^{\circ}\text{C}$. Parameters tested include sensory nerve action potentials (SNAPs) and motor studies of the median nerve. For my analyses, I used median nerve conduction velocity as the sensory outcome and abductor pollicis brevis (APB) latency as the motor outcome. Both of these measures contribute to the electrodiagnostic test severity of CTS, which was graded according to the seven-point Canterbury scale, where 0 denotes no neurophysiological abnormality and 6 denotes extremely severe CTS³²⁰. Electrodiagnostic tests reflect large nerve fibre function. The presence of neuropathies other than CTS was excluded by performing sensory and motor nerve conduction studies of the ulnar and radial nerves.

5.2.2.3 Quantitative sensory testing (QST)

QST was performed according to a standardised protocol³²². For the purposes of these analyses, I specifically used the thermal thresholds, which are reflective of small fibre function and cannot be assessed with standard electrodiagnostic tests. Cold and warm detection thresholds in median nerve-innervated skin (on the palmar side of the index finger) were measured with a Thermotester. Median nerve thermal sensory limen (TSL) – the difference threshold for alternating cool and warm stimuli – was used in these analyses as a composite measure of both cold (A δ fibres) and warm (C-fibres) detection. All QST data were log-transformed to achieve a normal distribution and expressed as Z-scores.

5.2.2.4 Skin biopsy and confocal imaging

The majority of PiNS CTS and cases and controls underwent skin biopsy in the index finger, using a 3mm diameter punch biopsy. Biopsy specimens were fixed in 2% periodate-lysine-paraformaldehyde for 30 minutes, washed in 0.1 M phosphate buffer, stored for 2-3 days in 15% sucrose in 0.1 M phosphate buffer, embedded in OCT and stored at -80°C. Specimens were cut on the cryostat in 50 μ m sections, and double-stained with myelin basic protein (MBP) and protein gene product 9.5 (PGP). Intra-epidermal nerve fibre density (IENFD) was determined on microscopy, using three sections per participant, and expressed as fibres per mm of epidermis. Additionally, nodes of Ranvier were identified in the sub-epidermal plexus (≥ 15 per patient), and their length measured. Nodes were considered elongated if they exceeded 6.1 μ m in length³²³, and the percentage of elongated nodes was calculated for each individual. In the analyses described below, individuals were dichotomised into a “no node elongation” group, and a “>0% elongated nodes” group.

5.2.3 DNA extraction

DNA extraction was undertaken for a total of 166 participants: (1) PiNS CTS participants with a full or near-full complement of electrophysiological, QST, skin biopsy and questionnaire data, regardless of whether they were in the surgical CTS cohort (N=86), (2) all PiNS control participants (N=27), and (3) all MGCTS participants who had been recruited by that stage (N=53).

DNA was extracted from whole blood using the PureLink Genomic DNA Kit (Invitrogen, Carlsbad, CA, USA). Frozen blood specimens were defrosted on ice, and 200 µl transferred to a sterile microcentrifuge tube. 20 µl of Proteinase K and 20 µl of RNase A were added to each sample, vortexed, and incubated at room temperature for 2 minutes. 200 µl of Lysis/Binding Buffer was then added to each sample, vortexed, and incubated at 55°C for 10 minutes to promote protein digestion. Following incubation, 200 µl of 100% molecular grade ethanol was added to the sample and vortexed for 5 seconds to yield a homogenous solution. 640 µl of the resulting lysate was loaded onto spin columns and centrifuged at 10,000 g for 1 minute at room temperature, and the collection tubes discarded. Columns were then washed with wash buffer 1 and centrifuged at 10,000 g for 1 minute, followed by wash buffer 2 and centrifuged at 17,000 g for 3 minutes. DNA was eluted in 50 µl of genomic elution buffer. After 1 minute of incubation at room temperature, the columns were centrifuged at 17,000 g for 1 minute, and the purified DNA was collected in DNase/RNase-free Eppendorf tubes. After purification, the yield of purified DNA was estimated by UV absorbance at 260 nm on a NanoDrop spectrophotometer (Thermo Fisher Scientific, Paisley, UK), and samples stored at -80°C until further use.

5.2.4 Whole-genome genotyping

Whole-genome genotyping of 166 genomic DNA samples was performed by the Oxford Genomics Centre at the Wellcome Trust Centre for Human Genetics. The genotyping array used was the Infinium Global Screening Array-24 v2.0 (Illumina, San Diego, CA, USA), which carries approximately 700,000 markers. One sample failed QC due to having a call rate of 96.3%, which is below the 98% call-rate cut-off imposed by the array. The mean call rate for the remaining 165 samples was 98.8%.

5.2.5 Quality control of genotyped data

Quality control (QC) of the remaining 165 samples followed almost exactly the same pipeline that I used for QC of the UK Biobank dataset, described in Chapter 2. QC was limited to 709,053 SNPs on the 22 autosomes. Briefly, QC began by removing all 4,997 SNPs with call rate <90%. This was followed by sample QC, where six samples with heterozygosity >3 S.D. from the mean were flagged for exclusion. No samples were flagged for exclusion based on <98% call rate, nor on sex discrepancy, as determined using the PLINK --check-sex command on the X chromosomes.

After excluding six samples which exhibited high heterozygosity, the remaining 149 samples were merged with publicly available data from the 1000 Genomes Project, and principal component analysis (PCA) was performed using flashpca¹⁵⁴. The first principal component was plotted against the second, which demonstrated that the majority of the cohort overlap with the European ancestral groups from the 1000 Genomes Project, particularly GBR (British in England and Scotland). There were four notable ethnic outliers with principal component 1 >

-0.1 (**Fig. 5.2**), and these individuals were flagged for exclusion prior to commencing PRS-phenotype (but not genotype-gene expression) analyses.

PCA was followed by SNP QC. 11,957 SNPs with call rate <98% were flagged for exclusion, as were 1,178 SNPs with HWE p-value < 0.0001. A total of 691,354 SNPs and 159 samples remained following the QC protocol.

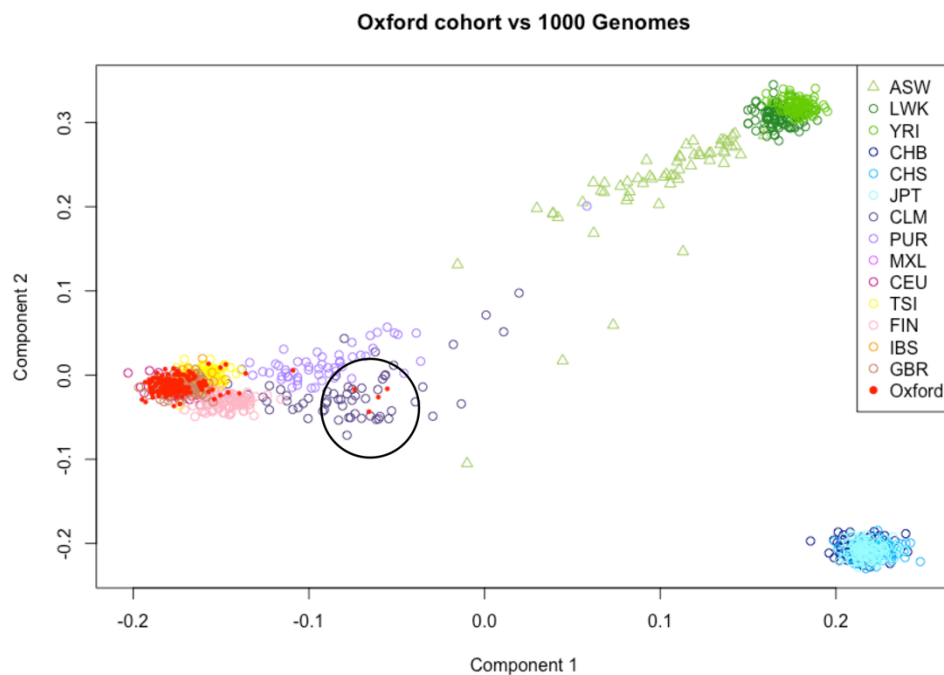


Fig. 5.2 Principal component analysis for identification of ethnic outliers. Each red dot indicates an individual in the Oxford cohort. Most of the samples overlap with European ancestral groups (GBR, TSI, FIN). The four circled individuals were identified as ethnic outliers.

5.2.6 Imputation

Prior to pre-phasing and imputation, all monomorphic SNPs (MAF = 0), multiallelic SNPs, indels and duplicated SNPs were removed. SNP IDs, reference alleles and strand orientations were aligned with the Haplotype Reference Consortium (HRC) reference panel v1.1, using a script provided by Dr William Rayner (Wellcome Trust Centre for Human Genetics, Oxford).

The PLINK files were converted to Variant Call Format (VCF) files and uploaded to the Sanger Imputation Service (Wellcome Trust, Hinxton, UK). Pre-phasing and imputation were performed using EAGLE2¹⁶¹ and the PBWT algorithm³²⁴.

5.2.7 Polygenic risk score in the Oxford cohorts

The previously identified four ethnic outliers were excluded from the dataset following imputation, leaving a total of 155 samples (81 PiNS cases, 50 MGCTS cases, and 24 PiNS controls).

A polygenic risk score (PRS) was calculated for all 155 individuals using the weighted genetic risk score method outlined in Chapter 3. Out of the 16 CTS-associated loci, there was a low density of genotyped SNPs at locus 2 (rs12104955 on chromosome 2) and locus 5 (rs847139 on chromosome 2). The index SNPs discovered in the association study were not available in the imputed dataset, and there were no nearby SNPs in high LD that could act as proxies. Thus, the PRS was constructed based on 14 out of 16 index SNPs for which genotype data were available.

Mean PRS was calculated for all cases (PiNS and MGCTS), all controls, and for the PiNS and MGCTS cases separately. T-tests were performed to compare the differences in mean PRS between the three case groups and the controls. Furthermore, the mean allelic dosage of the risk allele for each index SNP was compared between all cases and controls.

5.2.8 PRS-phenotype analyses

Twelve independent tests of PRS-phenotype association/correlation were performed for CTS cases, spanning a variety of measurable phenotypes pertinent to CTS:

1. Whether individuals had a family history of CTS
2. Whether individuals demonstrated any evidence of elongated nodes of Ranvier ($>6.1\text{ }\mu\text{m}$) in the small sensory fibres of the sub-epidermal plexus
3. Whether the VAS pain score was zero or greater than zero
4. Electrodiagnostic (EDT) severity grade
5. Boston symptom score
6. Difference in Boston symptom score before and after surgery (surgical cases only)
7. Boston function score
8. Difference in Boston function score before and after surgery (surgical cases only)
9. Sensory nerve conduction as measured by nerve conduction velocity in the median nerve
10. Motor nerve conduction as measured by abductor pollicis brevis latency
11. Median nerve thermal sensory limen (TSL)
12. Intra-epidermal nerve fibre density (IENFD) (fibres/mm of epidermis)

The variables were grouped according to type of data (binary, ordinal, or continuous), and all statistical analyses performed in R. Differences in PRS for binary variables were assessed with a t-test, whereas ANOVA was performed to look for differences in mean PRS across the six grades of EDT severity. Correlation tests between PRS and the various parameters were performed for continuous variables, and a Shapiro-Wilks test of normality undertaken to assess

whether the data were normally distributed. Pearson's r was performed to assess correlation in normally distributed variables, and Spearman's ρ for the non-normally distributed variables.

To account for multiple testing across 12 phenotypes, an adjusted p-value was derived for each test, using the Benjamini-Hochberg method. Given that some of these variables are likely to be correlated (e.g. thermal thresholds and intra-epidermal nerve fibre density), the Benjamini-Hochberg method to control the false discovery rate was deemed to be more appropriate than the more conservative Bonferroni correction.

5.2.9 RNA extraction from carpal tunnel tenosynovium

Surgically resected specimens were immediately immersed in RNALater stabilisation solution (Thermo Fisher Scientific, Paisley, UK) and stored at -80°C until further use. RNA was extracted using the High Pure RNA Isolation Kit (Roche, Penzberg, Germany). After defrosting over ice, samples were homogenised in TRIzol (Invitrogen, Carlsbad, CA, USA), mixed with chloroform, and incubated for 10 minutes at room temperature. The samples were centrifuged at $12,000\text{ g}$ at 4°C for 15 minutes, after which $200\text{ }\mu\text{l}$ of the supernatant was added to an equal volume of 70% ethanol and transferred to filter columns. The samples were repeatedly centrifuged at $13,000\text{ g}$ for 30 seconds until there was no more flow-through. The samples were then incubated with $100\text{ }\mu\text{l}$ of DNase for 15 minutes at room temperature. This was followed by two wash-centrifugation cycles for each of two wash buffers, followed by a final centrifugation step to dryness. RNA was eluted in $30\text{ }\mu\text{l}$ of elution buffer, and the purified RNA collected in RNase-free Eppendorf tubes. After purification, the yield of purified RNA was estimated by UV absorbance at 260 nm on a NanoDrop spectrophotometer, and samples stored at -80°C until further use.

5.2.10 RNA-Seq of tenosynovial RNA

RNA samples were sent to the Oxford Genomics Centre at the Wellcome Trust Centre for Human Genetics for sequencing. RNA-Seq was performed in two batches (batch 1: PiNS N = 41 specimens, July 2017; batch 2, MGCTS, N = 43 specimens, March 2019). The same sequencing protocol was applied to both batches.

Library preparation was poly(A)-enriched and directional, and performed using the Illumina TruSeq Stranded mRNA Library Prep Kit and standard universal Illumina multiplexing adaptors. Conversion of the poly(A)-selected RNA to cDNA was performed using the strand-specific dUTP strand-marking protocol³²⁵, and amplification used unique dual indexing. Sequencing was performed using the Illumina HiSeq4000 system with 75 bp read length and paired-end reads. The Phred score was used to encode quality metrics in the resulting FastQ sequencing files, following the Sanger standard³²⁶. All sequencing lanes gave a high yield, high-quality base content, and consistent GC content. Individuals were multiplexed in lanes.

Raw sequencing data were mapped to the hg19 (GRCh37) human genome assembly using STAR aligner¹⁹¹ with default ENCODE options. The genome was indexed using `-sjbdOverhang` (Read mate length 1), and counts were assigned using HTSeq¹⁹² against the GRCh37.87 ENSEMBL gene annotation. Data processing was performed in R, and normalised for effective library size using DESeq2 transformation¹⁹³. Read counts were further log2-transformed using the regularised log transformation¹⁹³, which also moderates log-fold changes by estimating a parameter that is proportional to the expected true abundance of each gene. In this way, data are scaled and centred so as to be directly comparable across samples, log2-transformed, and effect sizes are shrunk towards zero to reduce overestimation (mostly for lowly expressed genes).

As the tenosynovial RNA-seq data were produced in two library batches, the batch effect was removed by fitting a linear model and removing the component attributed only to the batch effect, using the limma package³²⁷ in R. PCA was used to confirm removal of the batch effect and to identify any outliers. Three outliers (one from PiNS; two from MGCTS) were excluded on this basis (Fig. 5.3).

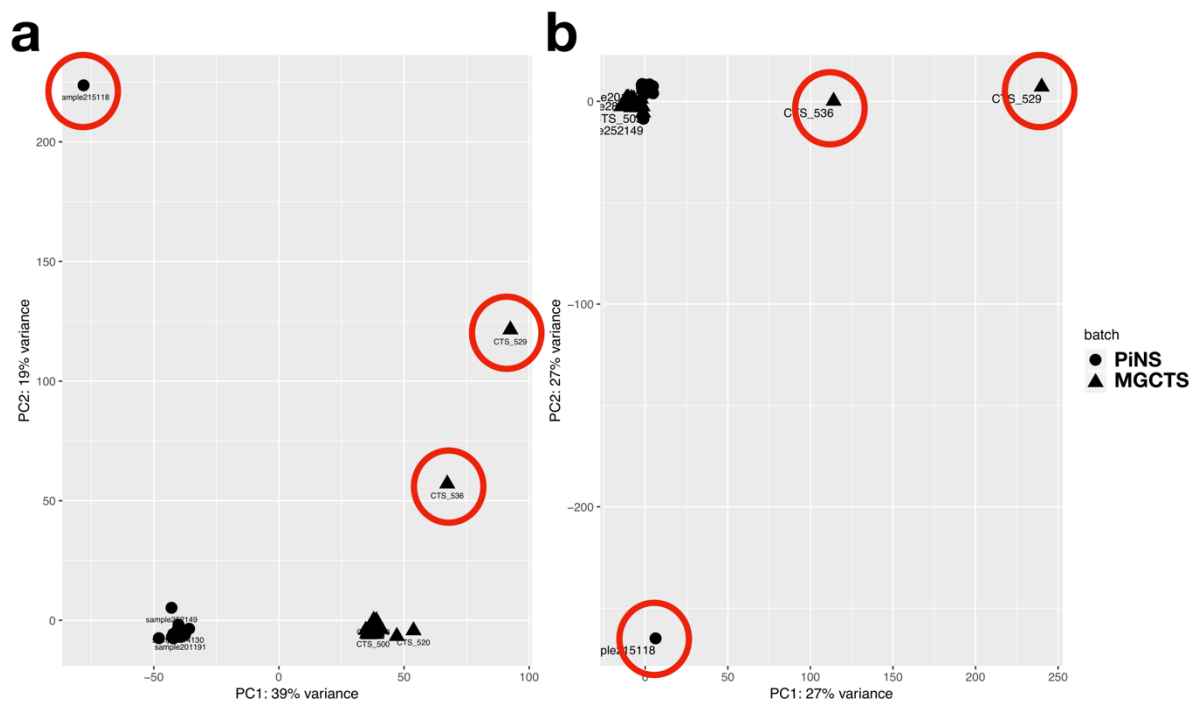


Fig. 5.3 Principal component analysis of gene expression in the Oxford cohorts. (a) Before removal of batch effect, and (b) after removal of batch effect. The three outliers are circled in red. These three samples persisted as outliers even after removal of the batch effect, so were excluded from all analyses.

Of the four ethnic outliers excluded from the PRS-phenotype analyses, two donated tenosynovial specimens from which RNA was extracted. These two individuals were included in the genotype-specific expression analyses for two reasons: (1) they did not appear as outliers in the PCA of gene expression and clustered with the other white British participants, and (2) the specific hypothesis being tested in the genotype-specific expression analyses was whether

genotype at a particular index SNP correlates with the expression of a particular gene, and this is independent of the allele frequencies of that SNP in different ancestral populations. In contrast, ethnic outliers were excluded for the PRS analyses because the PRS was built using additive effects of the genotypes at 14 independent SNPs discovered in a white British population, so can therefore be distorted by discrepancies in allele frequencies and different patterns of LD.

5.2.11 Genotype-specific gene expression in tenosynovium

Firstly, log2-transformed expression levels of the 25 FUMA-mapped genes in tenosynovium were compared against the median, top quartile, and top decile expression levels for all genes.

The genotype-specific analyses were performed on a total of 78 individuals for whom both RNA-Seq data and genotyping data were available (40 PiNS and 38 MGCTS). Genotype at the index SNPs at the GWAS loci was used to ascertain genotype-specific expression for the genes of interest. The genes selected for these analyses were 22 genes mapped by FUMA (Chapter 2, **Table 2.2**) to 14 index SNPs for which genotype data were available.

For each index SNP, the 78 individuals with both RNA-Seq and genotype data were grouped according to genotype; for imputed SNPs, individuals were grouped according to the inferred genotype on imputation. Four out of 14 index SNPs in the Oxford cohort were genotyped, and the 10 imputed SNPs were of high quality, with a mean INFO score of 0.94 (S.D. 0.07). All genotype-specific gene expression analyses were performed in R, and the methodology is best illustrated by example:

The index SNP at the first locus on chromosome 1 was rs12406439, and FUMA mapped a single gene to this locus – *COL11A1*. T is the effect allele at this SNP, and C the non-effect allele. The individuals with genotypes CC, CT, and TT were coded as 0, 1, and 2 in terms of effect allele dosage, and there were 13, 29, and 34 individuals with these genotypes, respectively. The mean and S.D. of the log2-transformed expression levels were calculated for each of the three groups, and a box plot produced. One-way ANOVA was performed to test for differences between the group means, and a post-hoc pairwise t-test was performed between each pair of groups (i.e. 0 vs 1, 0 vs 2, and 1 vs 2). A Benjamini-Hochberg correction was applied to the t-test p-values to account for three tests being performed.

In cases where the index SNPs had >1 gene mapped to that locus by FUMA, e.g. (locus 11, rs55841377: four genes mapped), the same tests as above were performed between SNP dosage at the index SNP and gene expression for all genes. The pairwise t-test p-values were Benjamini-Hochberg corrected, as above, but there was no further correction to account for the fact that four different genes were examined at this locus, given that SNPs that reside in regulatory regions frequently modulate the expression of several genes.

In cases where there were no individuals with a particular genetic dosage (e.g. locus 7, rs4678145: zero individuals with an effect allele dosage of 0), a 2-tailed t-test rather than ANOVA was performed between the gene expression in the two groups available, and the t-test p-values were left unadjusted. In cases where there was only one individual with a particular dosage (e.g. locus 13, rs72725608: one individual with an effect allele dosage of 2), an ANOVA test could be performed, but a t-test could only be performed between the two dosage groups with > 1 individual, and the p-values were therefore unadjusted.

5.2.12 Comparison of gene expression in tenosynovium vs index finger skin

An analysis was performed to quantify the relative gene expression of three candidate genes in tenosynovium versus a control tissue. The three genes chosen for this analysis were those that had the strongest evidence for functionality in the *in silico* analyses described in Chapter 3: *EFEMP1*, *ADAMTS17* and *ADAMTS10*.

Ideally, this control tissue should be the carpal tunnel tenosynovium of individuals without CTS. However, in the absence of such an ideal control tissue, index finger skin collected from punch biopsies of six healthy (i.e. non-CTS) PiNS participants was used as a surrogate “control” tissue. Skin and tenosynovium, though histologically distinct, both contain a significant population of fibroblasts^{314,328} and are rich in connective tissue components such as collagen and elastin^{58,329}. The index finger skin biopsy specimens were snap frozen in liquid nitrogen immediately after collection, and RNA extraction performed following the same protocol as described above for tenosynovium. RNA-Seq and downstream analyses were performed at the Oxford Genomics Centre using the same protocol (but in different library batches) as for the tenosynovial RNA specimens.

Differential expression analysis between the index skin of six control individuals and the tenosynovium of 41 PiNS CTS cases was performed in DESeq with a zero-mean normal prior and log2 fold-change shrinkage. Counts were normalised and transformed using the regularised log2 transformation $\pm$ the standard error of the mean. P-values were computed using the Wald test and were FDR-adjusted.

5.2.13 Analysis of female sex hormone receptors

Expression levels for four female sex hormone receptor genes were also analysed in the 78 PiNS and MGCTS tenosynovial samples (their protein products are given in parentheses): *ESR1* (Oestrogen receptor α), *ESR2* (Oestrogen receptor β), *GPER* (G protein-coupled oestrogen receptor 1), and *PGR* (progesterone receptor).

Sex-stratified analyses were performed to examine whether these genes are expressed in greater quantities in females than males. Two-tailed t-tests were performed for each gene, and p-values adjusted for multiple testing across four genes using the Benjamini-Hochberg method.

5.3 Results

5.3.1 Polygenic risk score analyses

Demographic details of the genotyped cohort are shown in **Table 5.1**, broken down into whether individuals were PiNS cases, PiNS controls, or MGCTS cases.

Table 5.1 Demographics of the Oxford cohorts. All individuals for whom genotype data were available are shown.

	PiNS cases	PiNS controls	MGCTS cases
N	81	24	50
N females (%)	52 (64%)	15 (63%)	32 (64%)
Mean age (S.D.)	61.0 (12.7)	55.0 (18.9)	60 (14.6)
N surgery (%)	71 (88%)	0 (0%)	50 (100%)

The results of the PRS analyses are shown in **Table 5.2** and **Fig 5.4**. **Fig. 5.4a** demonstrates that the PRS were normally distributed across all cases and controls (range 0.85-1.86). CTS cases had a significantly higher PRS than the controls (1.38 vs 1.21, $p = 0.003$). Sub-group analysis of the cases found that both the PiNS (PRS = 1.37) and MGCTS (PRS = 1.39) cohorts independently demonstrated a greater PRS than the controls ($p = 0.008$ for PINS; $p = 0.005$ for MGCTS).

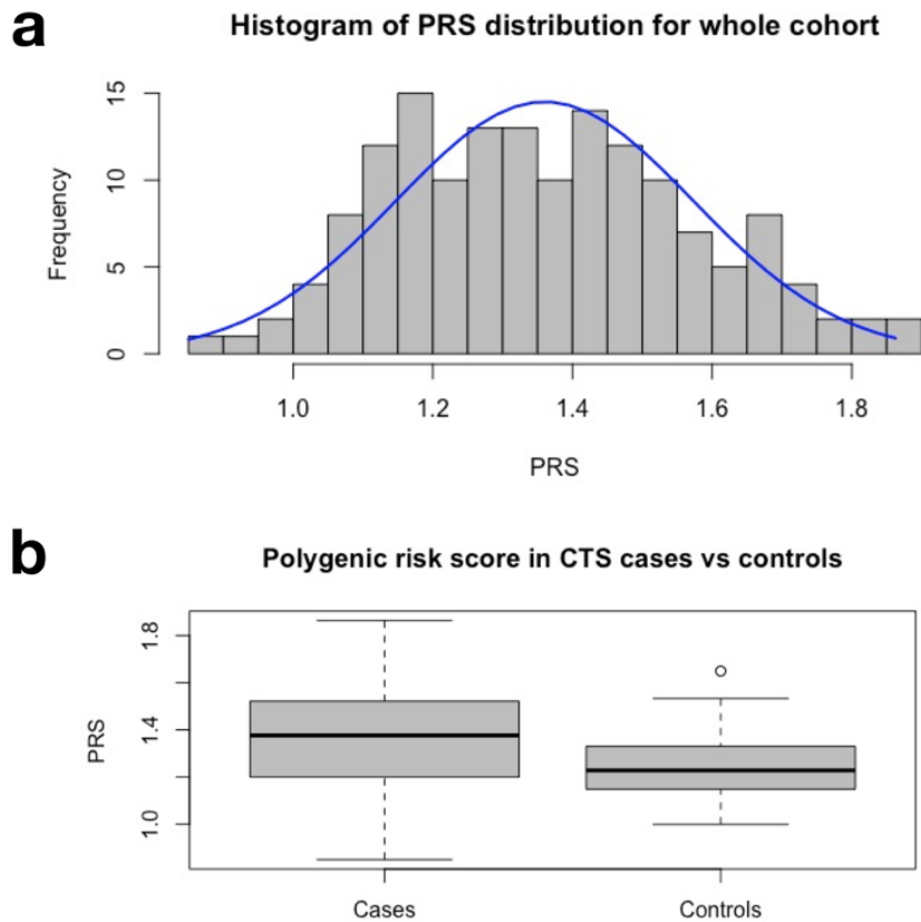


Fig. 5.4 Polygenic risk score analyses in the Oxford cohort. (a) Histogram showing the distribution of PRS in both CTS cases and controls. (b) Box plots showing mean PRS after removal of batch effect. Central line = median, box = upper quartile to lower quartile, whiskers = maximum/minimum values excluding outliers (circled).

The mean allelic dosages of the fourteen risk alleles incorporated into the PRS were computed for cases and controls. The allelic dosages were greater in the cases at 12/14 loci (**Table 5.2b**).

Table 5.2 Polygenic risk score (PRS) analyses (a) Mean PRS in all CTS cases (and broken down by cohort), and in all controls. (b) Comparison of allelic dosage of the risk allele between CTS cases and controls at each of the 14 index SNPs. Values where the dosage was greater in controls than in cases are highlighted in ***bold italic***.

(a)

Group	N	PRS	S.D.	P (t-test vs all controls)
All cases	131	1.379	1.212	0.003
PiNS cases	81	1.372	0.213	0.008
MGCTS cases	50	1.391	0.221	0.005
All controls	24	1.256	0.167	-

(b)

	Mean dosage of risk allele	
	All cases	Controls
Locus 1	1.2595	1.1250
Locus 3	0.4198	0.3333
Locus 4	1.2374	1.0417
Locus 6	1.5767	1.5771
Locus 7	1.7744	1.7083
Locus 8	0.2679	0.2479
Locus 9	1.4649	1.4167
Locus 10	1.7863	1.7083
Locus 11	1.5676	1.3333
Locus 12	1.4046	1.5000
Locus 13	0.0916	0.0000
Locus 14	0.7542	0.6667
Locus 15	0.2336	0.1750
Locus 16	0.1092	0.0417

5.3.2 PRS-phenotype association and correlation analyses

The results of these analyses are summarised in **Table 5.3** and **Fig 5.5**. The binary variables (family history, elongation of Nodes of Ranvier, pain symptoms) demonstrated no statistically significant differences in PRS between the two outcome groups, even before p-value adjustment. Likewise, ANOVA of EDT grade did not show any statistically significant differences in PRS between groups, and there was no trend of either increasing or decreasing PRS with increasing electrodiagnostic disease severity (**Fig 5.5d**). None of the eight correlation tests were statistically significant, except between intraepidermal nerve fibre density and PRS (**Fig 5.4l**; Spearman's $\rho = 0.25$, $p = 0.03$), with the finding that greater PRS correlated with increased intra-epidermal nerve fibre density. However, this parameter was not statistically significant when p-values were corrected for multiple testing.

Table 5.3 PRS-phenotype analyses. Summary of the twelve genotype-phenotype analyses performed, broken down by nature of the data being examined. P_{adj} = adjusted p-value following Benjamini-Hochberg correction.

Binary variables	Population	N (Yes / No)	"Yes" - Mean PRS (S.D.)	"No" - Mean (S.D.)	Test	P	P _{adj}
Family history of CTS (Yes / No)	PiNS + MGCTS cases	123 (42 / 81)	1.391 (0.205)	1.377 (0.225)	t-test	0.73	0.83
% Elongated nodes > 0 (Yes / No)	PiNS cases	69 (22 / 47)	1.324 (0.219)	1.380 (0.217)	t-test	0.33	0.80
VAS pain > 0 (Yes / No)	PiNS cases	80 (52 / 28)	1.383 (0.232)	1.367 (0.207)	t-test	0.76	0.83
Ordinal variables	Population	N total	N by EDT grade	Mean PRS (S.D.)	Test	P	P _{adj}
Electrodiagnostic severity	PiNS cases	81			ANOVA	0.08	0.48
1			5	1.57 (0.17)			
2			14	1.31 (0.22)			
3			27	1.42 (0.19)			
4			17	1.29 (0.22)			
5			15	1.36 (1.36)			
6			3	1.36 (0.09)			
Continuous variables (parametric)	Population	N	Mean (S.D.)	Correlation test	Coefficient	P	P _{adj}
Boston Symptom Score	PiNS cases	81	2.69 (0.67)	Pearson	0.081	0.47	0.80
Boston Symptom post-op minus pre-op	PiNS cases	71	1.37 (0.21)	Pearson	0.14	0.25	0.80
Sensory (median NCV)	PiNS cases	45	21.5 (12.7)	Pearson	0.12	0.43	0.80
Motor (APB latency)	PiNS cases	78	28.8 (12.5)	Pearson	0.072	0.53	0.80
Continuous variables (non-parametric)	Population	N	Median (IQR)	Correlation test	Coefficient	P	P _{adj}
Boston Function Score	PiNS cases	81	2.13 (1.25)	Spearman	0.035	0.76	0.83
Boston Function post-op minus pre-op	PiNS cases	71	0.63 (0.81)	Spearman	0.005	0.97	0.97
Warm vs cold detection (TSL)	PiNS cases	81	7.23 (5.57)	Spearman	-0.097	0.39	0.80
IENFD	PiNS cases	81	4.19 (4.26)	Spearman	0.25	0.03	0.32

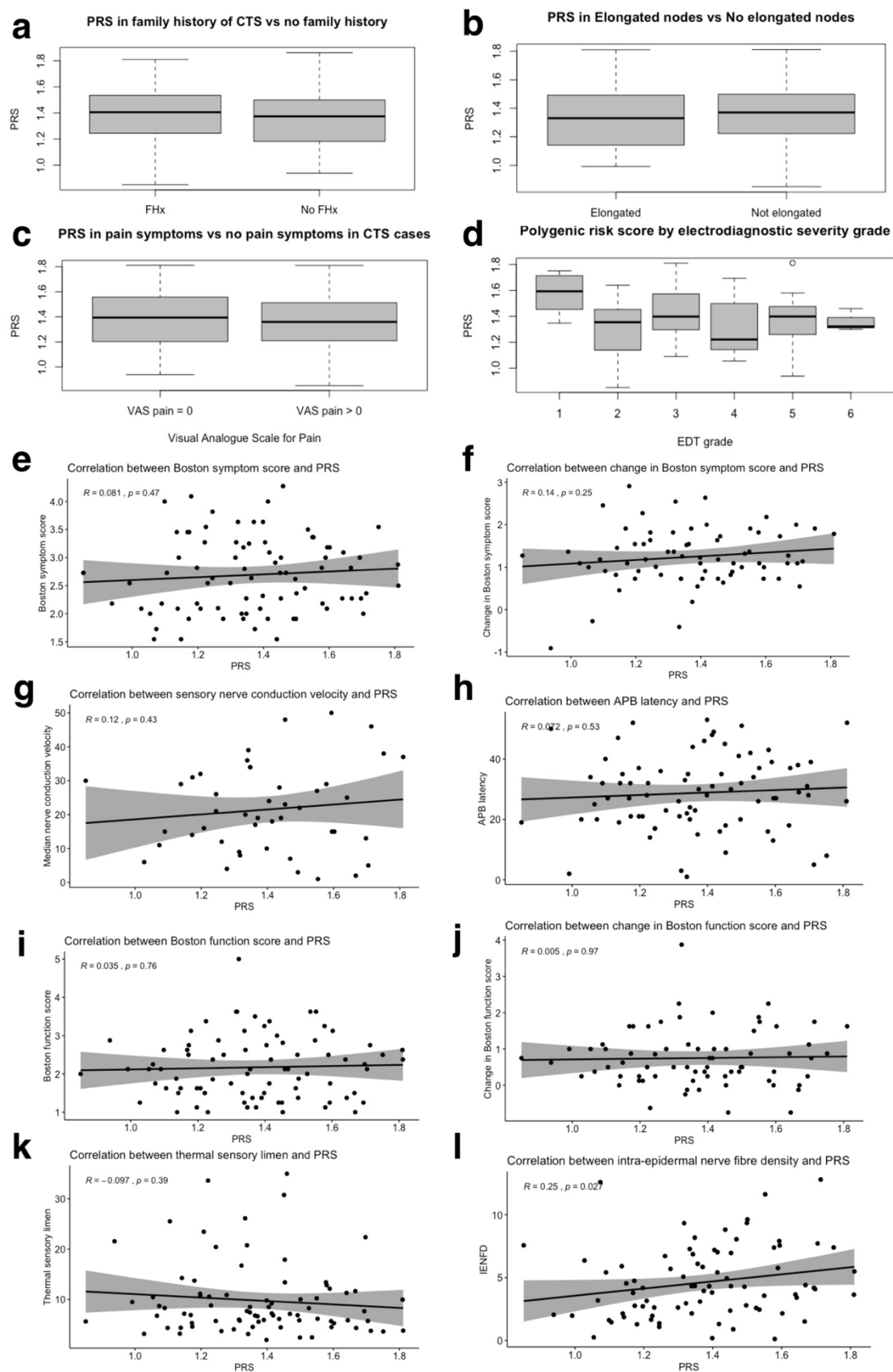


Fig. 5.5 PRS-phenotype correlation analyses. (a)-(l) Results of all analyses. Each figure has a title describing which parameters were being analysed.

5.3.3 Candidate genes are highly expressed in tenosynovium

Of the 25 FUMA-mapped genes, all were expressed in tenosynovium at levels above the median expression for all genes, including *EFEMP1*, *ADAMTS17* and *ADAMTS10*. Seventeen genes had expression levels in the top quartile, and ten in the top decile (Table 5.4; Fig. 5.6).

Table 5.4 Mean expression of 25 FUMA-mapped genes. Expression values are expressed as log2-transformed and normalised counts. 3rd / 4th quartile cut-off = 6.51. 9th / 10th decile cut-off = 10.10.

FUMA-mapped genes	Mean expression	S.D.	Top Quartile?	Top Decile?
<i>COL11A1</i>	5.07	1.04	No	No
<i>LTBP1</i>	12.32	0.32	Yes	Yes
<i>EFEMP1</i>	14.00	0.52	Yes	Yes
<i>KIAA1715</i>	10.53	0.32	Yes	Yes
<i>EVX2</i>	1.20	0.43	No	No
<i>DIRC3</i>	6.06	0.78	No	No
<i>ITGB5</i>	14.05	0.26	Yes	Yes
<i>UMPS</i>	9.13	0.16	Yes	No
<i>KALRN</i>	8.48	0.63	Yes	No
<i>MUC13</i>	2.73	0.37	No	No
<i>AEBP1</i>	13.15	0.56	Yes	Yes
<i>PLD2</i>	10.86	0.20	Yes	Yes
<i>MYL7</i>	1.56	0.28	No	No
<i>GCK</i>	4.69	0.44	No	No
<i>AOC1</i>	5.19	1.36	No	No
<i>TMEM176A</i>	9.96	0.67	Yes	No
<i>TMEM176B</i>	10.81	0.72	Yes	Yes
<i>KCNH2</i>	2.12	0.73	No	No
<i>IFT43</i>	8.79	0.25	Yes	No
<i>TTLL5</i>	9.08	0.25	Yes	No
<i>TGFB3</i>	10.65	0.47	Yes	Yes
<i>SMAD6</i>	8.61	0.40	Yes	No
<i>ADAMTS17</i>	8.66	0.45	Yes	No
<i>ADAMTS10</i>	11.50	0.40	Yes	Yes
<i>MYO1F</i>	10.40	0.41	Yes	Yes
Hormonal genes	Mean expression	S.D.	Top Quartile?	Top Decile?
<i>ESR1</i>	9.19	0.40	Yes	No
<i>ESR2</i>	1.45	0.40	No	No
<i>GPER</i>	9.36	0.35	Yes	No
<i>PGR</i>	7.28	0.44	Yes	No

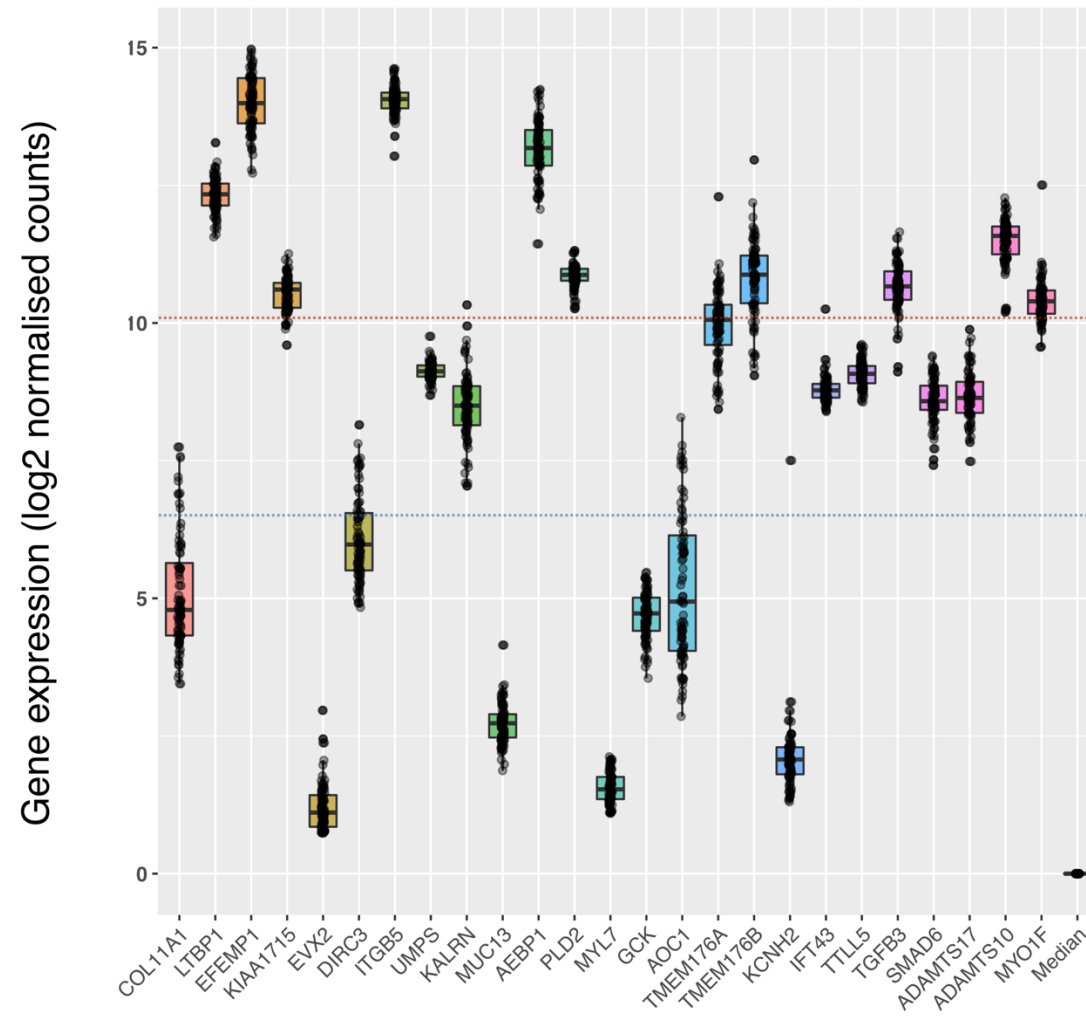


Fig. 5.6 Expression of 25 FUMA-mapped genes in CTS tenosynovium. The blue dotted line indicates the 3rd/4th quartile threshold for mean expression across all genes, and the red line indicates the 9th/10th decile threshold.

Comparing CTS tenosynovium and index finger skin from healthy controls, *ADAMTS10* and *EFEMP1* were significantly upregulated in CTS tenosynovium compared with healthy skin (*ADAMTS10*: log2 fold change (lfc) = 0.65; $p = 2.6 \times 10^{-3}$; *EFEMP1*: lfc = 2.29 = 1.9×10^{-14}), although *ADAMTS17* was not (Fig. 5.7).

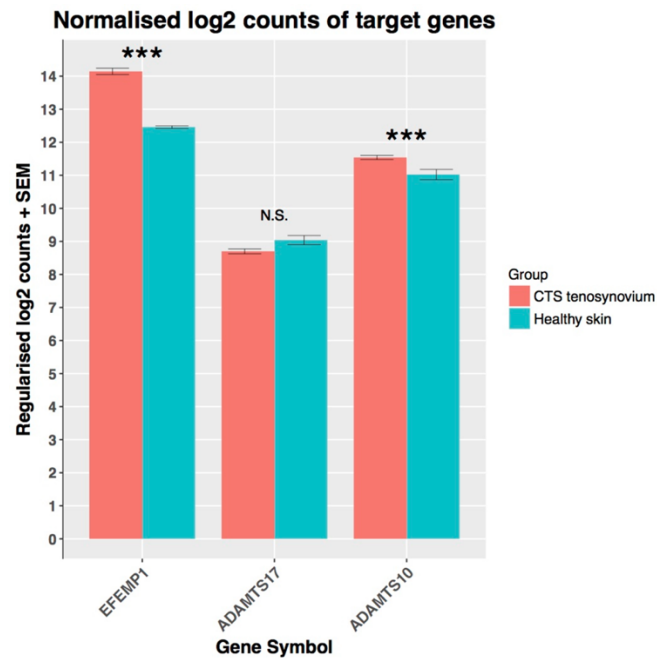


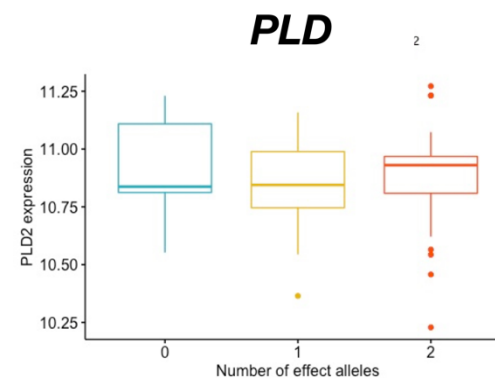
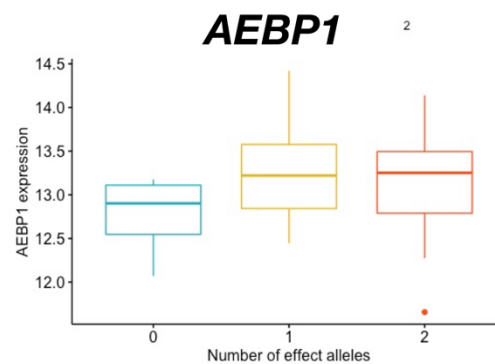
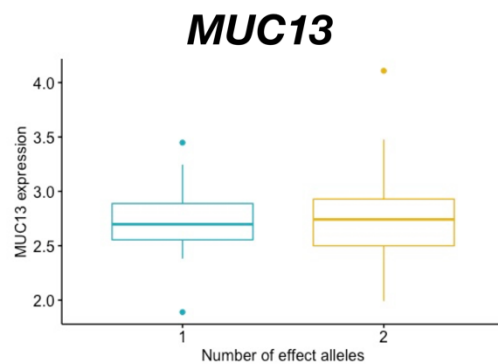
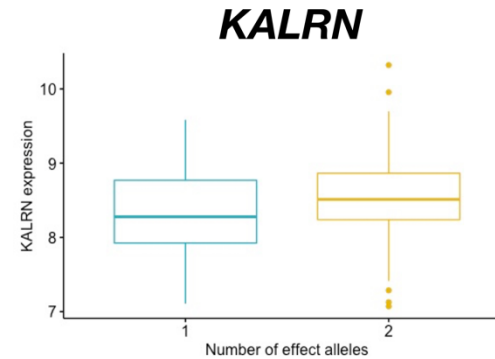
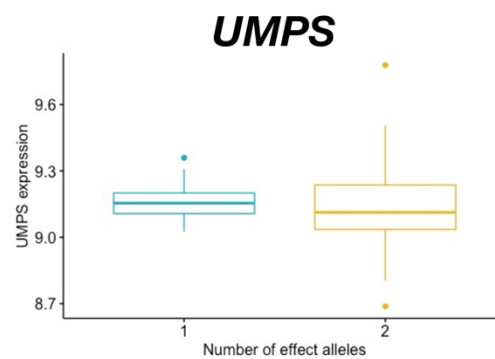
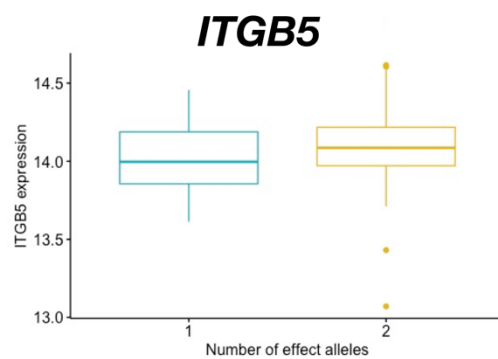
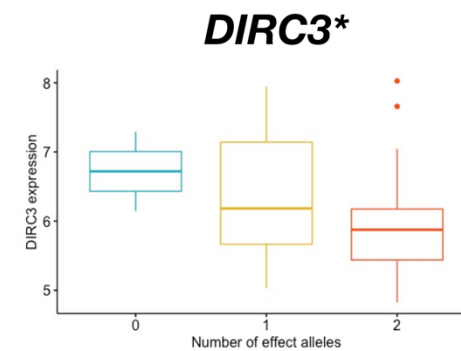
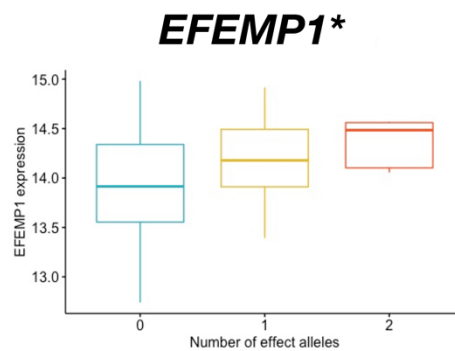
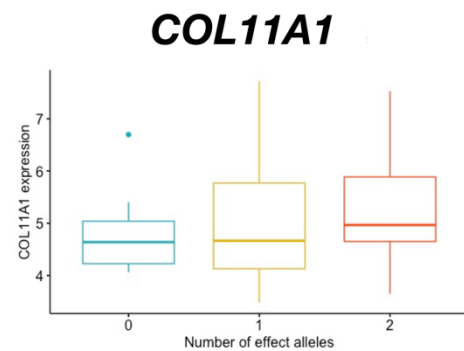
Fig. 5.7 Expression of three candidate genes in CTS tenosynovium vs index finger skin. Error bars = standard error of the mean of the regularised log2 counts. *** = $p < 0.01$; N.S. = non-significant.

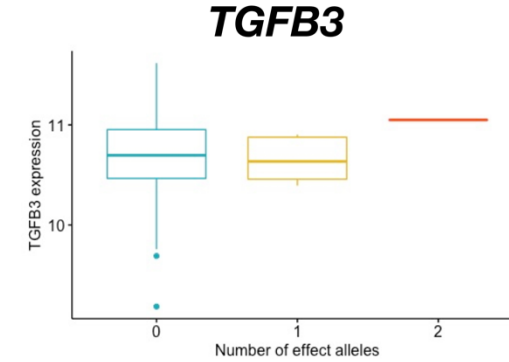
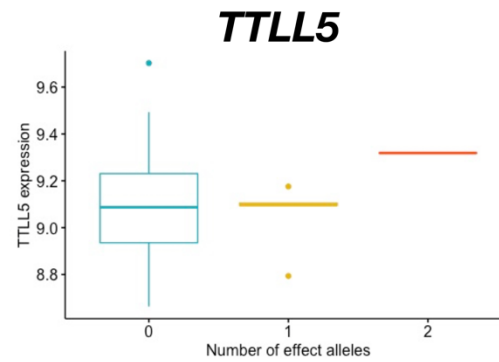
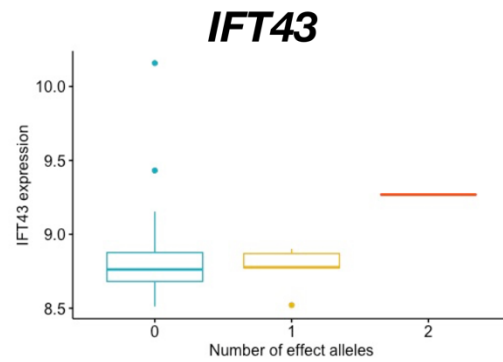
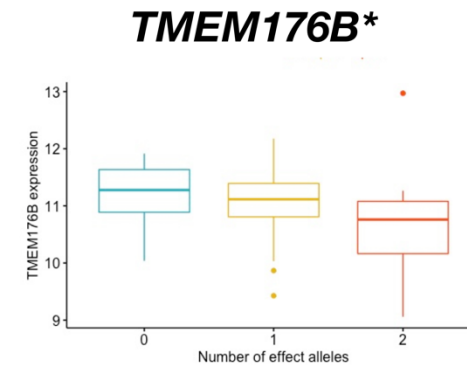
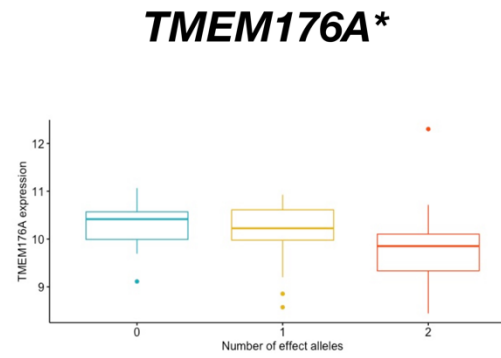
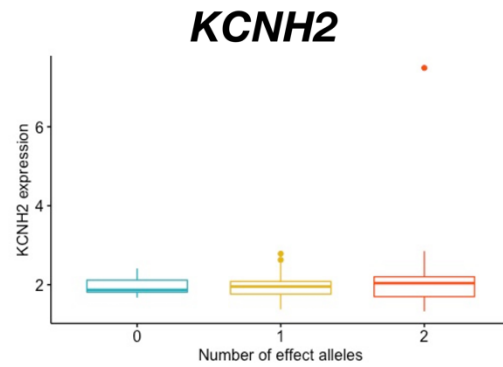
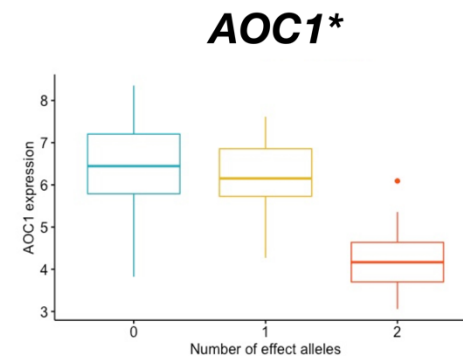
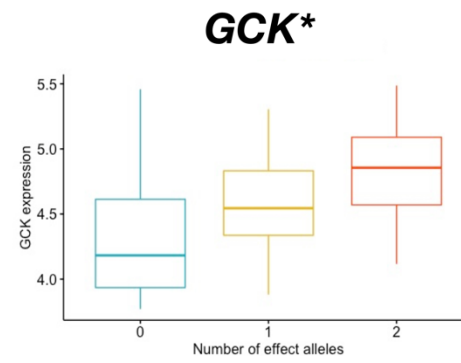
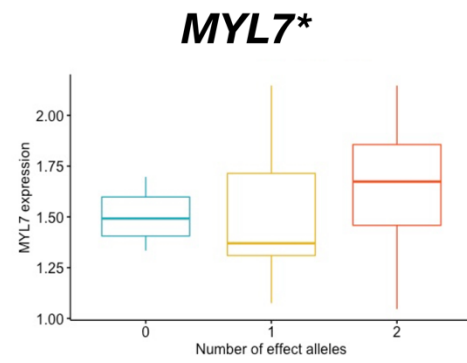
5.3.4 Genotype-specific expression analyses

Genotype-specific expression analyses found a statistically significant difference in mean expressions levels between individuals with differing effect allele dosages for seven out of 22 genes: *EFEMP1*, *DIRC3*, *MYL7*, *GCK*, *AOC1*, *TMEM176A* and *TMEM176B*. None of the genes that mapped to SNPs where there were only 0 or 1 individuals with a particular genotype dosage demonstrated statistical significance in the t-tests. Results are summarised in **Table 5.5** and **Figure 5.8**.

Table 5.5 Genotype-specific expression analyses. T-tests, when performed between all three allele dosage groups, are Benjamini-Hochberg adjusted. Other p-values are unadjusted. EAF = effect allele frequency. ***Bold italic*** = significant.

Locus	Index SNP	Gene	EAF	Individuals per effect allele dosage			Gene expression per allele dosage (S.D.)			t-test p-value			
				N (0)	N (1)	N (2)	Count (0)	Count (1)	Count (2)	ANOVA p	(0 vs 1)	(0 vs 2)	(1 vs 2)
1	rs12406439	<i>COL11A1</i>	0.58	13	29	34	4.78 (0.71)	5.03 (1.25)	5.26 (0.89)	0.34	0.41	0.2	0.41
3	rs3791679	<i>EFEMP1</i>	0.24	49	24	5	13.9 (0.53)	14.2 (0.43)	14.4 (0.25)	<i>0.035</i>	<i>0.039</i>	<i>0.039</i>	0.29
6	rs1863190	<i>DIRC3</i>	0.69	3	24	51	6.72 (0.58)	6.35 (0.88)	5.91 (0.65)	<i>0.017</i>	0.4	0.19	0.10
7	rs4678145	<i>ITGB5</i>	0.88	0	16	62	–	14.0 (0.25)	14.1 (0.26)	–	–	–	0.35
		<i>UMPS</i>		0	16	62	–	9.16 (0.09)	9.13 (0.18)	–	–	–	0.36
		<i>KALRN</i>		0	16	62	–	8.33 (0.60)	8.53 (0.63)	–	–	–	0.26
		<i>MUC13</i>		0	16	62	–	2.73 (0.38)	2.75 (0.36)	–	–	–	0.82
11	rs55841377	<i>AEBP1</i>	0.77	5	25	48	12.8 (0.46)	13.3 (0.55)	13.2 (0.50)	0.13	0.16	0.16	0.49
		<i>PLD2</i>		5	25	48	10.9 (0.27)	10.8 (0.19)	10.9 (0.19)	0.69	0.83	0.83	0.83
		<i>MYL7</i>		5	25	48	1.51 (0.15)	1.48 (0.29)	1.65 (0.29)	<i>0.043</i>	0.79	0.14	0.06
		<i>GCK</i>		5	25	48	4.39 (0.68)	4.57 (0.38)	4.84 (0.37)	<i>0.005</i>	0.59	0.32	<i>0.02</i>
12	rs6977081	<i>AOC1</i>	0.64	10	31	37	6.25 (1.49)	6.17 (0.98)	4.24 (0.66)	<i>3.87×10⁻¹³</i>	0.88	<i>0.003</i>	<i>4.0×10⁻¹²</i>
		<i>KCNH2</i>		10	31	37	1.96 (0.23)	1.96 (0.35)	2.12 (0.97)	0.62	0.98	0.54	0.54
		<i>TMEM176A</i>		10	31	37	10.3 (0.57)	10.2 (0.58)	9.75 (0.71)	<i>0.013</i>	0.56	<i>0.04</i>	<i>0.04</i>
		<i>TMEM176B</i>		10	31	37	11.2 (0.57)	11.0 (0.58)	10.6 (0.77)	<i>0.004</i>	0.38	<i>0.01</i>	<i>0.01</i>
13	rs72725608	<i>IFT43</i>	0.03	72	5	1	8.80 (0.23)	8.77 (0.15)	9.27	0.13	0.66	–	–
		<i>TTLL5</i>		72	5	1	9.09 (0.22)	9.05 (0.15)	9.32	0.52	0.62	–	–
		<i>TGFB3</i>		72	5	1	10.7 (0.45)	10.7 (0.23)	11	0.70	0.79	–	–
14	rs1866745	<i>SMAD6</i>	0.36	32	35	11	8.65 (0.39)	8.55 (0.35)	8.73 (0.36)	0.33	0.47	0.53	0.47
15	rs72755233	<i>ADAMTS17</i>	0.09	58	19	1	8.71 (0.44)	8.65 (0.40)	8.07	0.32	0.58	–	–
16	rs62621197	<i>ADAMTS10</i>	0.04	70	8	0	11.5 (0.39)	11.6 (0.28)	–	–	0.18	–	–
		<i>MYO1F</i>		70	8	0	10.4 (0.41)	10.4 (0.24)	–	–	0.90	–	–





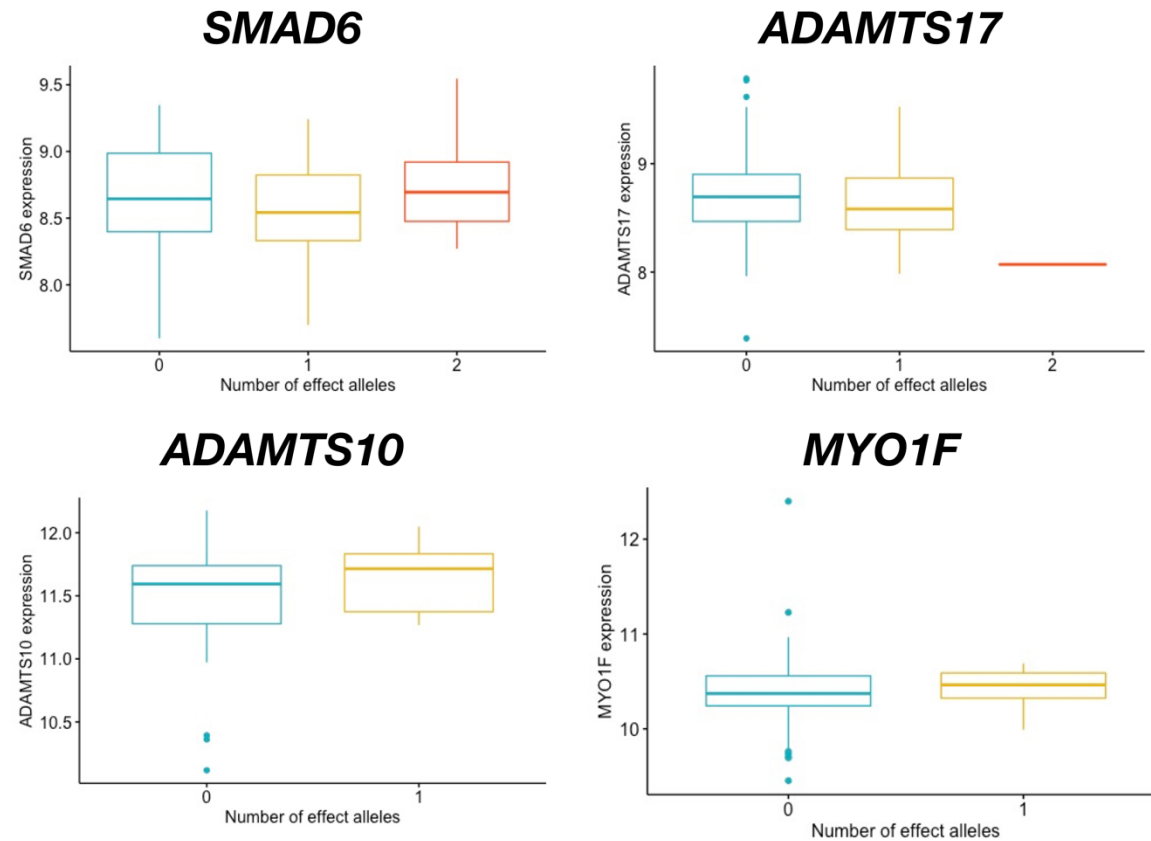


Fig. 5.8 Genotype-specific gene expression analysis. Relative expression levels (log2-transformed and normalised counts) of the 22 FUMA-mapped genes are shown for each genotype at the index SNP, indicated by the dosage of the effect allele (0, 1 or 2). Gene names followed by an asterisk indicates that ANOVA testing demonstrated a statistically significant difference in mean gene expression between genotypes.

5.3.5 Oestrogen and progesterone receptors are expressed in tenosynovium

Like the 25 FUMA-mapped genes, the four female sex hormone receptor genes were all expressed relatively highly in tenosynovium (all above the median; 3/4 in the top quartile) (Fig. 5.9). *ESR1* and *GPER* showed greater expression in females than in males. However, none of the sex differences in expression of the four genes reached statistical significance that withstood correction for multiple testing (Table 5.6).

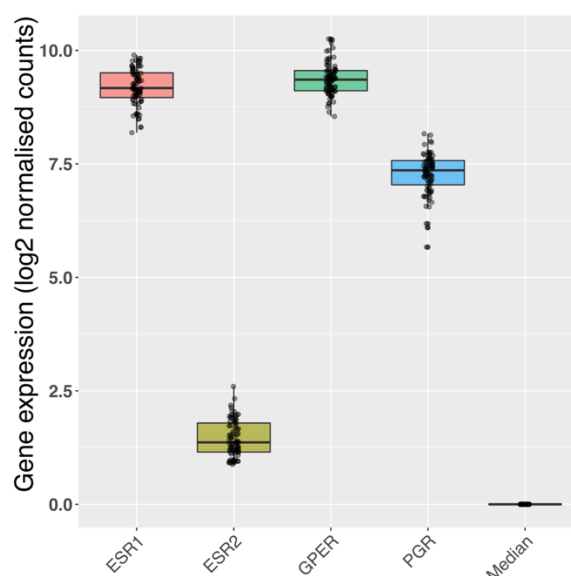


Fig. 5.9 Expression of four female sex hormone receptor genes in CTS tenosynovium.

Table 5.6 Expression of female sex hormones in carpal tunnel tenosynovium. Mean expression for each gene is given as log2-transformed normalised counts, with its S.D. Both unadjusted and Benjamini-Hochberg adjusted p-values are given.

	Females	Males	P	P _{adj}
N	49	29		
<i>ESR1</i>	9.26 (0.39)	9.14 (0.41)	0.22	0.43
<i>ESR2</i>	1.17 (0.30)	1.22 (0.32)	0.55	0.63
<i>GPER</i>	9.45 (0.32)	9.27 (0.35)	0.02	0.1
<i>PGR</i>	7.29 (0.46)	7.33 (0.37)	0.63	0.63

5.4 Discussion

The analyses in this chapter were undertaken with the aim of establishing whether the CTS-associated loci discovered in the UK Biobank GWAS could be informative in a much smaller, independent, and deeply phenotyped cohort of CTS patients. The inverse of this question was also asked, i.e. whether gene expression studies in this smaller cohort could provide evidence for potential functionality of the GWAS loci. Broadly, the various analyses succeeded in achieving these aims, in spite of the absence of correlation between PRS and the various granular CTS phenotypes.

Validation of the PRS (and thus the GWAS loci) was achieved in the Oxford cohort in spite of the relatively small sample size and the attenuated PRS, based on 14 loci and not 16. The robustness of the PRS reflects the careful phenotyping of both the UK Biobank and Oxford CTS cohorts, with particular attention paid to ensuring a homogenous ancestral population in both datasets. The PiNS cohort was considerably more deeply phenotyped than the MGCTS cohort, with all CTS cases having an electrophysiological diagnosis of CTS. Nevertheless, the PRS amongst the MGCTS cases was slightly higher than in the PiNS CTS cases (**Table 5.2a**). This suggests that selecting only surgical cases in the absence of any other phenotypic data is a crude but effective form of CTS phenotyping, as (1) hand surgeons are less likely to have misdiagnosed individuals they deem to require surgery, and (2) these individuals are on the phenotypically more severe end of the spectrum, and therefore likely to have a greater genetic liability for CTS.

When effect allele frequencies were compared between the Oxford cases and controls at the 14 individual loci, 12/14 loci showed concordance in terms of cases having a greater allelic dosage of the risk allele than controls (**Table 5.2b**). At one of the discordant loci, the difference was

negligible (locus 6: 1.5767 in cases vs 1.5771 in controls), whereas in the other, the difference was slightly larger but not statistically significant (1.40 in cases vs 1.50 in controls, t-test $p = 0.48$). This provides further evidence in support of the fact that phenotyping of CTS cases in UK Biobank and in PiNS/MGCTS had succeeded in ascertaining the same phenotype. In contrast, the Estonian replication cohort demonstrated directional concordance at 13/16 loci despite being >300 times larger.

The PRS-phenotype correlation analyses aimed to distil the full spectrum of phenotypic manifestations of CTS down to as few variables as possible. Phenotypes selected for analysis therefore ranged from electrodiagnostic measures to pain and symptom severity to target-organ innervation in the form of intra-epidermal nerve fibre density (IENFD) in median-innervated index skin. The selection of such a diverse range of measures was motivated by the fact that CTS manifests differently in different individuals, and symptoms do not necessarily correlate with each other – for instance, neurophysiology correlates poorly with patient-reported symptoms and level of disability³³⁰. The rationale underlying the decision to correlate PRS with these twelve phenotypes rather than genotype at individual loci were twofold. Firstly, most phenotypes are not determined by single alleles, but by joint effects of several alleles at multiple loci³¹⁰. Secondly, *post hoc* corrections for multiple testing ($14 \text{ loci} \times 12 \text{ phenotypes}$) would result in a substantial loss of statistical power.

None of the twelve phenotypes correlated with phenotype with the exception of IENFD, which counterintuitively suggested a greater innervation density with higher PRS. Nevertheless, this correlation did not withstand correction for multiple testing, which underscores the importance of such statistical conservatism when performing multiple pairwise tests, where the chance of a making a false positive discovery is high. There are several potential important reasons why

none of these analyses demonstrated an association or correlation between PRS and CTS phenotype.

Firstly, there is the ever-present matter of statistical power, which is relevant at both the genetic and sample levels. For reasons explained in Chapter 4, choosing to build a PRS based only on genome-wide significant SNPs (and not including suggestive SNPs) was a deliberately conservative strategy that was chosen given that this was a single-stage GWAS. If we factor in the attenuated PRS in the form of two missing loci and the fact that the CTS GWAS explains just over 2% of the SNP-based heritability of CTS, it becomes apparent that the PRS in its current form is only capturing a small fraction of the genetic architecture of CTS, and is consequently a weak instrument.

If any phenotype was going to show an association with PRS, it was most likely to be whether an individual has a family history of CTS. This analysis demonstrated that CTS cases with a family history did indeed have a higher PRS than those without (1.391 vs 1.377), but the difference was non-significant. This was in spite of the fact that this particular analysis was the best-powered of the twelve, with 123 individuals from both PiNS and MGCTS. Thus, statistical power was also lacking at the sample level, especially in the remaining 11 analyses that only used the deeply phenotyped PiNS cases, with a maximum of 81 individuals with genotype data. Deeply phenotyping even this number has been an inordinately labour-intensive endeavour for the single examiner, and it is unlikely that this cohort will grow significantly beyond its current size.

It must also be borne in mind that phenotypic variance is a product of both genes and environment. Within the Oxford cohort, none of the non-genetic variables were controlled for, including symptom duration – perhaps the most important parameter of all. Participants were

recruited at various stages of their disease, and the symptom duration recorded at recruitment ranged from months to decades. The natural history of CTS has not been extensively reported in the literature, but most studies seem to concur that most individuals exhibit temporal fluctuations in symptoms, and the natural history varies widely between individuals, with a progressive deterioration in most, but self-limited disease progression in some^{331,332}. There is likely to be a genetic factor that influences disease progression in individuals, and this would arguably have been the most interesting phenotype to study in relation to polygenic risk. However, this is a very difficult phenotype to quantify. Consideration was given to normalising all of the phenotypic measures by factoring in symptom duration, but patients' recall of symptom duration is notoriously prone to recall bias³³³, thus it was felt imprudent to introduce such a variable to phenotypic measures that are already subject to several layers of uncertainty and complexity.

Perhaps most important, however, is the fact that the polygenic risk score is a construct composed of all the different facets of the genetic architecture of CTS. For instance, pain in CTS is known not to correlate with the severity of nerve injury³³⁴; therefore, the pain phenotype in CTS is likely to have its own underlying genetic architecture. The same goes for the other facets of genetic predisposition to CTS that have been uncovered thus far in the thesis, such as altered skeletal growth and altered gene expression within the sub-synovial connective tissues – each of these reflects an aberration in a different biological pathway. Individuals will display varying degrees of genetic liability in these distinct domains of genetic predisposition. Consequently, the PRS, which is a cumulative tally of the magnitude of innate predisposition in these different domains, is too blunt an instrument to resolve the granular phenotypes within CTS.

In summary, therefore, there were several limitations inherent in the PRS-phenotype analyses, and it was perhaps optimistic to expect any of the measures to show a correlation with CTS.

The gene expression studies, in contrast, were considerably more informative. Firstly, all of the 25 FUMA-mapped genes at the associated loci were expressed in carpal tunnel tenosynovium, all of them above the median expression for all genes, with more than two-thirds in the top quartile and two-fifths in the top decile (**Fig. 5.6; Table 5.5**). While this fact alone does not prove that these genes are responsible for the perturbations in the SSCT seen in CTS, it does lend credibility to the notion that some of these identified genes are biologically plausible modulators of the SSCT.

All three of the top candidate genes identified in the bioinformatic analyses in Chapter 3 (*EFEMP1*, *ADAMTS17*, *ADAMTS10*) were amongst those relatively highly expressed genes. *EFEMP1* and *ADAMTS10* were significantly more highly expressed in tenosynovium compared to index finger skin in non-CTS participants, and *ADAMTS17* expression was lower (but the difference was non-significant) (**Fig. 5.7**). It is acknowledged that skin is a poor surrogate control tissue for the ideal control tissue, which is carpal tunnel tenosynovium collected from non-CTS controls. The harvesting of such tissue in healthy individuals is unethical, and while it could theoretically be collected in the course of hand trauma surgery or tendon transfer surgery, gene expression in such cases is likely to be perturbed as a result of the insult of trauma or the disease process necessitating tendon transfer surgery, respectively. Thus, while only a limited amount can be inferred from this analysis, it does show that these three genes are expressed at greater or comparable levels when measured against another tissue with dense connective tissue architecture.

In the genotype-specific gene expression analyses, seven genes exhibited a significant difference in mean expression levels between individuals of differing genotype: *EFEMP1*, *DIRC3*, *MYL7*, *GCK*, *AOC1*, *TMEM176A* and *TMEM176B* (**Fig. 5.8, Table 5.5**). Of these genes, only *EFEMP1* had emerged as a strong candidate gene in CTS pathophysiology from my previous bioinformatic analyses, so the other genes deserve some scrutiny.

GCK encodes glucokinase, an enzyme that phosphorylates glucose to glucose-6-phosphate and is a critical component of hepatic glucose metabolism. It serves as an important glucose sensor in pancreatic β -cells, and activators of this enzyme have been shown to potentiate the action of insulin in the liver, leading to trials of glucokinase activators as a therapeutic target for type 2 diabetes³³⁵. Polymorphisms at the *GCK* locus have been associated with an array of metabolic phenotypes in previous GWA studies, including fasting glucose³³⁶, offspring birth weight³³⁷, and type 2 diabetes³³⁸. Thus, in the context of my previous analyses of diabetes/adiposity and CTS, *GCK* is a gene that has considerable biological plausibility as a modulator of CTS risk.

TMEM176A and *TMEM176B* both encode transmembrane proteins that physically interact with each other. *TMEM176A* has been linked to glioblastoma³³⁹, oesophageal³⁴⁰ and hepatocellular carcinoma³⁴¹, while *TMEM176B* has been associated with small cell lung carcinoma³⁴² and regulation of immune cell activation and tolerance³⁴³. Although their mechanisms of action are as yet unclear, it is noteworthy that *TMEM176B* was shown to be upregulated in tumour vessels of human renal cell carcinoma patients, suggesting that it has a role in tumour angiogenesis³⁴⁴. It is conceivable, therefore, that it could play a similar angiogenic role in the SSCT in CTS.

DIRC3 (disrupted in renal cell carcinoma 3), a non-coding RNA gene, was originally discovered in a familial study of renal cell carcinoma³⁴⁵, but aberrations in this gene have been linked to other cancers, including melanoma. Loss of function in melanoma cell lines has recently been

shown to lead to increased anchorage-independent growth, a strong predictor of metastatic potential; target genes of *DIRC3* are thus hypothesized to be associated with processes such as cell adhesion, migration and proliferation³⁴⁶.

The potential relevance of *MYL7* (Myosin Light Chain 7) and *AOC1* (amine oxidase copper-containing 1) to CTS are less clear. *MYL7* is expressed almost exclusively in the atrium of the heart¹⁸⁰, and is overexpressed in hypertrophic cardiomyopathy³⁴⁷. In the genotype-specific expression studies, *AOC1* demonstrated the strongest evidence of allele-dependent gene expression. This gene encodes a secreted glycoprotein that catalyses the breakdown of putrescine and histamine, and has been implicated in kidney organogenesis as a downstream target gene of the Wilms Tumour Protein (WT1)³⁴⁸.

It is worth noting that the majority of genes that failed to demonstrate convincing evidence of genotype-specific expression were mapped to index SNPs with low minor allele frequencies (locus 7: MAF = 0.12; locus 13: MAF = 0.03; locus 15: MAF = 0.09, locus 16: MAF = 0.04). Consistent with the Hardy-Weinberg principle, the number of individuals who were homozygous for the minor alleles at these loci was very low - either 0 or 1. The 78 samples included in this analysis therefore failed to confer sufficient statistical power for examining gene expression in the individuals homozygous at the minor alleles at these loci. It was this limitation in statistical power that motivated the decision to examine only the genotype-gene expression relationships for the index SNPs at each locus with the FUMA-mapped genes at that locus, rather than performing a formal eQTL analysis.

ADAMTS17 and *ADAMTS10* are exceptions to this reasoning. My *a priori* hypothesis for these two genes was that the causal variants confer CTS risk through missense variants that alter protein function rather than gene expression levels. For these two genes, therefore, the failure

to find a significant difference between the two testable genotypes serves to strengthen the argument that the GWAS had indeed identified the true causal variants.

The sex-stratified GWAS in the previous chapter strongly suggested differences in the genetic architecture of CTS between the sexes. Toesca et al.³¹³ used immunohistochemical methods to demonstrate expression of oestrogen receptor α (a product of the *ESR1* gene) in fibroblasts and synovial lining cells in the carpal tunnel tenosynovium of 30 CTS cases, and expression was found to be higher than in four non-CTS controls. They reported progesterone receptor expression (*PGR* gene) in the transverse carpal ligament, but not in tenosynovium. Kim et al.³¹⁴ found that both oestrogen receptor α and oestrogen receptor β (*ESR2* gene) were expressed more highly in the tenosynovium of 14 postmenopausal female CTS cases compared to six controls.

This transcriptomic study has therefore confirmed the observation that *ESR1* and *ESR2* are expressed in carpal tunnel tenosynovium and has also provided the first ever demonstration of progesterone receptor expression in this tissue. Expression of *GPER* (G protein-coupled oestrogen receptor 1, which binds oestradiol) in tenosynovium has not been previously reported either. Both *PGR* and *GPER* were highly expressed relative to all genes examined in the RNA-Seq (**Fig. 5.9**).

Progesterone has been shown to play an important role in myelin formation^{349,350}, and Schwann cells are known to both synthesise progesterone³⁵¹ and express its receptor³⁵². Progesterone is therefore a physiological regulator of peripheral nerve function, including nerve recovery following injury³⁵³, as demonstrated by impaired remyelination following inhibition of progesterone in a mouse model of sciatic nerve injury³⁵⁴. Based on this preclinical data, two small randomised control trials of progesterone vs corticosteroid injection for CTS have been

undertaken by the same group^{355,356}. Although these trials were both small and of questionable methodological quality, the second study suggested that post-injection improvement in Boston function score at six months was modestly greater in the progesterone group compared to the corticosteroid group³⁵⁶. A more seemingly robust trial of corticosteroid vs progesterone has been registered by a different group³⁵⁷, and it will be interesting to see if their results are more noteworthy. All trials involve the injection of progesterone into the carpal tunnel, so any observed effects may be mediated by the effects of progesterone on its receptors in the SSCT rather than in the median nerve.

The observation that there were no significant differences between the expression levels of the four sex hormone receptor genes between males and females accords with what Toesca et al. reported for oestrogen receptor α and progesterone receptor³¹³. These receptors, all encoded in autosomal genes, are nuclear receptors for steroid sex hormones. Therefore, the serum concentration of the circulating hormone (a variable with significant sexual dimorphism), and not the receptor density, is likely to be the principal determinant of any downstream effects. It has been posited that oestrogen receptors, once activated by oestrogen, bind to nuclear factor (NF)- $\kappa\beta$ to repress the transcription of pro-inflammatory genes such as IL-6, thus attenuating the cascade of events that lead to angiogenesis, oedema and fibrosis within the SSCT in CTS^{314,358}. The observation that the sex hormone receptors are expressed equally between the sexes is therefore not at odds with the postulate that female sex hormones play a role in CTS pathogenesis through their actions in the SSCT.

In conclusion, the analyses in this chapter have achieved the aims of validating the PRS for CTS in a local cohort and providing evidence for functionality at several of the GWAS loci. The functional plausibility of *EFEMP1*, *ADAMTS17* and *ADAMTS10* as CTS risk genes in the SSCT was cemented, while other credible candidate genes were identified. The relative success of

the genotype-specific expression analyses compared to the genotype-phenotype analyses serves as a useful illustration of a biological inevitability: there is a dramatic attenuation in the signal-to-noise ratio as the measured variable gets further removed from the level of the transcriptional machinery in the nucleus.

Chapter 6

Discussion

6.1 Summary of key findings

The overarching aim of this thesis was to gain an understanding of the genetic architecture of CTS, and the principal findings from the various studies presented herein can be summarised as follows:

- A GWAS of CTS revealed 16 genome-wide significant loci associated with CTS and implicated a number of biologically plausible candidate genes that are likely to be involved in CTS pathogenesis, including *ADAMTS17*, *ADAMTS10*, *EFEMP1*, *COL11A1*, *TGFB3*, *LTBP1*, and *SMAD6*. Strong functional candidate SNPs were identified for the first three, whereas the latter three belong to the TGF- β /SMAD signalling pathway.
- A disproportionate number of genes implicated in the GWAS are important components and modulators of the extracellular matrix (ECM).
- There was strong evidence of genetic correlation between CTS and anthropometric variables – namely shorter height and raised BMI.
- A polygenic risk score (PRS) for CTS demonstrated that within UK Biobank, individuals in the top quintile for PRS had a relative risk for developing CTS of 1.88 compared to individuals in the bottom quintile. Also, CTS cases who had undergone carpal tunnel

surgery had a greater PRS than CTS cases who had not undergone surgery, suggesting that PRS associates with disease severity.

- Nested case-control studies within the UK Biobank cohort confirmed epidemiological associations between CTS and type 2 diabetes, rheumatoid arthritis, hypothyroidism, and to a lesser extent, migraine.
- Mendelian randomisation found a causal role for shorter stature, raised BMI, and earlier age at menarche in the development of CTS.
- There were clear sex differences in the genetic architecture of CTS, with stronger associations in females. Two genome-wide significant loci specific to females were identified.
- There was moderately strong evidence of shared genetic architecture between CTS and migraine, with a demonstration of genetic overlap in the vicinity of *ASTN2* on chromosome 9.
- A cohort of deeply phenotyped CTS cases in Oxford had a significantly higher PRS for CTS compared to controls, providing external validation for the GWAS in an independent cohort. However, PRS within this cohort did not demonstrate correlation with granular phenotypes relevant to CTS.
- RNA-Seq of carpal tunnel tenosynovium found that there was high expression of genes mapped to the GWAS loci, in addition to oestrogen and progesterone receptor genes.
- Several of the genes that mapped to the GWAS loci demonstrated evidence of genotype-specific expression, suggesting that variants at the GWAS loci are functional.

6.2 Four postulates on the genetic aetiology of CTS

How can these findings be integrated into what is already known about CTS? In the interests of conceptual clarity, I will attempt to answer this question by distilling the key findings down to four fundamental postulates about how genetic variants may contribute to CTS liability. These postulates are based on overlapping strands of evidence accrued from different analytical methods employed throughout the thesis, and they all accord with our current understanding of CTS.

6.2.1 Postulate 1: Genetic variants linked to skeletal growth predispose to CTS

The Mendelian randomisation analysis of height demonstrated that shorter height was causal in CTS aetiology, and this phenomenon was independent of any pleiotropic effects. There was an abundance of evidence linking the genetic architecture of CTS to skeletal growth. Several of the GWAS loci were previously reported for their association with height, and there was a strong negative genetic correlation between height and CTS. Observationally, UK Biobank CTS cases were ≥ 2 cm shorter than controls. It is doubtful that shorter height *per se* is the mediator for increased CTS risk; rather, the inference is that the genetic propensity towards altered skeletal growth predisposes to the squarer wrist shape that has been reported in CTS patients in several studies. This, in turn, creates an anatomical configuration that is likely to result in greater carpal tunnel pressures.

6.2.2 Postulate 2: Genetic variants linked to adiposity predispose to CTS

As with height, the Mendelian randomisation study using strong genetic instruments for BMI found a causal role for raised BMI in CTS risk; again, this accorded with the observation in UK

Biobank that there is a ≥ 2.1 kg/m² discrepancy in BMI between CTS cases and controls. The association between obesity and CTS is well-established in the literature, and various measures of adiposity showed the strongest genetic correlations with CTS out of all phenotypes examined. It appears that the causal role of BMI on CTS is not mediated entirely through increased risk of type 2 diabetes, as seen in the Mendelian randomisation analysis of diabetes. Rather, I believe there is a role played by the altered metabolic profile associated with obesity, and in support of this, metabolic syndrome and dyslipidaemia are associated with peripheral neuropathy independent of diabetes³⁵⁹. Furthermore, the genotype-specific expression analyses in the previous chapter provided robust evidence for variants at one of the associated loci on chromosome 7 altering *GCK* (glucokinase) expression. This metabolic theory of adiposity and CTS seems more credible than the previously posited theory of increased fat volume within the carpal tunnel²⁸⁰.

6.2.3 Postulate 3: Genetic variants acting in the SSCT predispose to CTS

The set of GWAS-mapped genes showed strong enrichment for ECM components (e.g. *EFEMP1* and *COL11A1*) and modulators of the ECM (e.g. the three genes in the TGF- β /SMAD signalling pathway and probably the *ADAMTS* genes). Several of the implicated genes play a role in cell adhesion and proliferation, which is perhaps why so many of them have been found to be perturbed in various cancers.

The hypothesis examined in the final chapter was whether these genes exert their effects in CTS through their expression in the SSCT, and there was strong evidence in favour of this. There is an abundance of literature focusing on changes within the SSCT in CTS that goes back several decades, and my findings are certainly compatible with these observations. What remains to be examined is the hypothesis that the median nerve is rendered more vulnerable

through the over- or under-expression of these genes within the ECM of the nerve itself. This is a considerably more difficult hypothesis to prosecute due to the obstacles inherent in obtaining median nerve specimens. As discussed in *Chapter 3*, *ADAMTS17* is highly expressed in Schwann cells, and was shown to be upregulated in the distal nerve stump in a mouse nerve injury model²⁴⁰. The role of this gene in nerve recovery following crush injury is an intriguing and testable functional hypothesis.

6.2.4 Postulate 4: Genetic variants that predispose to CTS exert greater effects in females

It is almost universally accepted that CTS is more common in females than it is in males. A credible explanation for this phenomenon was borne out in my sex-stratified GWAS, which demonstrated stronger association signals in females than males, even when the two cohorts were size-matched. There are various studies in the literature that paint a compelling (though not entirely coherent) picture of the role played by female sex hormones in CTS aetiology. I believe that the demonstration of progesterone receptor expression in CTS tenosynovium and the Mendelian randomisation analysis demonstrating a strong causal role for younger age at menarche in CTS predisposition both add to this melee of interesting hypotheses, if not exactly providing greater clarity. Further experimental, observational, and genetic epidemiological analyses are required to disentangle the complex web of female sex hormones, age at menarche, parity, metabolism and adiposity.

6.2.5 Integration

At the outset of Chapter 2, I gave two *a priori* hypotheses for how genetic variants might feasibly contribute to CTS predisposition: ‘nerve structure/environment’ vs ‘nerve/pain

vulnerability’. The intention was to create two broad conceptual frameworks under which the findings in this thesis could be classified, with the qualification that the two hypotheses were not mutually exclusive.

The analyses in this thesis are presented in chronological order. After completing the *in silico* analyses in Chapter 3, there was reason to believe that the nerve environment hypothesis might predominate over the nerve vulnerability hypothesis. However, the subsequent analyses have highlighted the likely role of metabolic and hormonal variables in CTS aetiology, and the picture is now considerably less clear; postulates 2, 3, and 4 are certainly all compatible with the nerve vulnerability hypothesis.

The genetic architecture of a complex trait cannot be neatly compartmentalised into discrete domains or groups of genes that are independent of one another. Biology is a science of complex systems and lacks the determinism of classical physics. While it is frequently necessary to dissect these systems into their constituent parts to enable them to be studied, there are grave limitations to this sort of reductionism. The emergent properties of biological systems cannot be explained by studying the individual components in isolation³⁶⁰, as illustrated by the interdependency of the genome and by pleiotropy.

This is certainly consistent with the finding that the more data accrued in this thesis, the greater the resistance to any attempts at *post-hoc* rationalisation. The postulates should therefore not be considered as four irreconcilable propositions, but rather thought of as four overlapping Venn diagrams.

6.3 Broader implications

6.3.1 CTS as a genetic disease

The work encapsulated in this thesis has brought us closer to understanding CTS as a complex disease, and I believe a strong case has been made for there being an underlying genetic architecture.

Within the public sphere, there is a strong perception that CTS is caused by hand overuse, and curiously, this perception remains well-entrenched even among members of the medical profession. Any patient performing an internet search in an attempt to understand why they have developed CTS is unlikely to stumble across any mention of their innate genetic propensity towards developing the disease. They will, however, find countless sources of information purporting that their CTS is a result of computer overuse or poor ergonomics.

CTS has been listed as an occupational disease in the European Union's register of occupational diseases since 2003³⁶¹, and there is a vast amount of literature on workplace-related CTS, much of it emanating from the USA. According to the US Bureau of Labor Statistics, CTS is classified as an injury associated with repetitive trauma, and there were 46,000 cases of workplace-related CTS in 1986, with an increase to 281,800 by 1992. This was accompanied by a corresponding rise in workers' compensation claims. As of 2003, CTS is the leading cause of lost workdays due to "workplace injury" ³⁶².

One of the loudest dissenting voices to the occupational school of thought of CTS aetiology has come from David Ring, whose group wrote the following in the *Journal of Hand Surgery*¹¹³:

“According to our analysis of the available literature, the etiology of CTS is largely structural, genetic, and biological, with environmental and occupational factors such as repetitive hand use playing a minor and more debatable role. There is insufficient evidence to implicate hand use of any type, or typing in particular, as an important and direct cause of CTS, and to do so with confidence might be considered scientifically irresponsible. Given the impact of labeling a disease as work-related, and the sobering history of such negative speculative illness concepts, such causal links should not be adopted without strong scientific support. Current scientific evidence is inadequate to implicate environmental or occupational factors in carpal tunnel syndrome”.

Based on the work presented in this thesis, I am inclined to add my voice to theirs. A disease with strong anthropometric and metabolic correlates, a heritability of 0.46, and robust evidence of an underlying genetic architecture should probably not be labelled as “occupational”. The narrative certainly needs to be more nuanced, as it is likely that gene-environment interactions are central to the ‘occupational theory’ of CTS aetiology: individuals with an innate genetic liability may be more likely to develop CTS if they aggravate an already vulnerable median nerve through repetitive hand movements or tool use.

The matter of workers’ compensation in the not-too-distant era of polygenic risk scores will undoubtedly prove difficult to navigate. However, it is hoped that the work presented in this thesis will help to steer the discourse towards a more nuanced understanding of the role played by non-genetic factors in CTS pathogenesis.

6.3.2 Other entrapment neuropathies

Although CTS is by far the most common entrapment neuropathy, it is not the only one. The next most common entrapment neuropathy is cubital tunnel syndrome³⁶³, a compression of the

ulnar nerve under Osborne's ligament, which forms the roof over the bony concavity formed by the olecranon process and medial epicondyle. Moreover, there are several other sites of median, ulnar and radial nerve compression in the upper limb. Moving from the upper limb to the lower limb, there is also meralgia paraesthetica (compression of the lateral femoral cutaneous nerve), and tarsal tunnel syndrome (compression of the tibial nerve in the lower limb equivalent of the carpal tunnel). What these diseases and CTS share in common is the impingement of peripheral nerves by anatomical structures.

The implications of this thesis for the understanding of these other entrapment neuropathies are unclear. If a degree of innate nerve vulnerability is one of the consequences of having a high polygenic risk for CTS (as appears might be the case), it seems a perfectly credible proposition that the other entrapment neuropathies should share some of this genetic architecture. Given that CTS is by far the commonest entrapment neuropathy, adequately powering a study to examine these genetic correlations will be a considerable challenge. Within the UK Biobank showcase of c. 500,000 participants, fewer than 800 have an operative code for cubital tunnel release. Thus, while cross-phenotype association studies are not currently a realistic prospect, CTS is likely to be a useful model system for understanding certain genetic factors that manifest in these other entrapment neuropathies.

6.3.3 BMI – a modifiable risk factor?

Mendelian randomisation analysis was originally conceived by epidemiologists who wished to establish the causal role played by *modifiable* risk factors in disease²⁴⁸. The science underlying causal inference is valid whether or not the risk factors are modifiable, but there are fewer tangible public health benefits from the study of non-modifiable risk factors.

Of the three Mendelian randomisation analyses in CTS that showed a strong causal effect, BMI is the only modifiable risk factor. A strong case can therefore be made for weight loss therapy to be instituted as one of the first-line interventions for overweight patients who present to primary care with mild symptoms of CTS. The expectation is that patients who successfully lose weight might find that their CTS symptoms improve, or even resolve. If we factor in the litany of health benefits that result from weight loss in the obese, it would be difficult to argue against the trial of such an intervention. However, a pessimist (or perhaps, geneticist) may point to the high heritability estimates for BMI (ranging from anywhere between 31% to 90%)^{364,365} and argue that such interventions are inherently futile. Such a genetically deterministic outlook is not entirely baseless – many weight loss interventions have been trialled, but any effects tend to be modest, and weight regain is common³⁶⁶.

6.4 Future directions

There are a great number of directions in which I could proceed on the basis of the findings presented in this thesis, and I will describe some of these potential avenues.

6.4.1 X-chromosome analysis

One of the most significant findings in the course of this work was the sex differences in the genetic architecture of CTS. In the face of such genetic and epidemiological evidence, analysis of the sex chromosomes was conspicuously lacking from my GWAS. The sex chromosomes contribute to the variance seen in sexually differentiated phenotypes, but have thus far been excluded from the vast majority of GWAS²⁸⁸. The principal reason for this neglect has been the paucity of statistically robust methods to account for the uniqueness of the sex chromosomes. For instance, there is reduced power as a result of a smaller allelic sample size (X-inactivation in females and males only carrying one allele). However, analytic methods for X-wide association studies (XWAS) have been developed and are being increasingly refined³⁶⁷, and novel trait- and disease-associated loci on the X-chromosome have been discovered^{368,369}. It would be prudent to attempt such an analysis in CTS, as it would strongly complement the sex-stratified analyses.

6.4.2 Functional studies

There has been a deluge of GWAS data in the past decade. As discussed in Chapter 1, biological understanding of the association signals lags considerably behind the discovery of novel loci. Marjoram et al.³⁷⁰ eloquently argue for the need for ‘more biology’ over ‘more data’:

“..we face a fork in the road in which one path is labeled ‘more data’, whereas the other is labeled ‘more biology’. Although both paths are clearly useful, we believe there is much to be gained by considering the latter fork and exploiting the ability of biological knowledge to inform statistical analysis as we attempt to move from association to function”

I concur with this sentiment. The first steps that I would take in the laboratory would be to perform immunohistochemical and immunocytochemical studies to identify sub-cellular localisation of the GWAS candidate genes within tenosynovium and also transverse carpal ligament. I would also seek to determine the expression of these genes in nerve tissue collected from healthy controls. Transverse carpal ligament has been collected along with tenosynovium in both MGCTS and PiNS, and collection of control nerve from healthy volunteers is ongoing as part of these studies. These investigations can be expanded to a range of cell and tissue types *in vitro*, and I would be particularly interested to assess the expression of these genes in human induced pluripotent stem cell-derived sensory neurons co-cultured with rat Schwann cells³⁷¹. Such a model would allow the study of the effect of knock-down or overexpression of these genes on myelination.

ADAMTS17 is one of the most interesting genes that emerged from this study, not only because it has not yet been fully characterised, but because the GWAS strongly implicated a missense variation (with likely deleterious effect) as the causal variant. While *ADAMTS17* is one of the GWAS genes that is known to affect skeletal growth, it is also highly expressed in Schwann cells, and I have previously discussed the strands of evidence for its potential role in the peripheral nerve. Firstly, the substrate of *ADAMTS17* must be identified, and mass spectrometry-based proteomic techniques can be employed to this end³⁷². Furthermore, an *ADAMTS17*-knockout mouse will be generated. This will allow, among other things, the study of nerve recovery in the mutant vs wild type mouse using a chronic nerve compression model,

potentially yielding valuable insights into the function of this gene in the pathophysiology of nerve compression.

6.4.3 Migraine and CTS

Within this thesis, I had the opportunity to study the intriguing link between migraine and CTS. There was a strong suggestion of overlap in genetic architecture near the *ASTN2* (Astrotactin 2) on chromosome 9, which is expressed highly in both the central and peripheral nervous systems.

Many of the similar techniques used to phenotype the CTS participants in PiNS could potentially be applied to migraine patients. One of the “trigger sites” targeted in migraine surgery is in the glabellar area, where neurolysis of the supratrochlear and supraorbital nerves is performed. A punch biopsy taken from the forehead skin innervated by these nerves before and three months after surgery would allow quantification of intra-epidermal nerve fibre density (IENFD). The same tissue could also be used for transcriptomic studies to assess which genes are differentially expressed before and after surgery, and *ASTN2* would be a strong *a priori* candidate for showing such differential expression. Evidence for a change in IENFD or gene expression that correlated with symptomatic improvement post-op would add considerable weight to the idea that the reported success of migraine surgery is not down to placebo alone.

The CTS-migraine analyses highlighted the possibility that altered central pain processing may be a common denominator in both diseases, and this is worthy of further exploration. In particular, the genetic relationship between pain phenotypes and CTS is an area that has hitherto not been investigated. A GWAS of multisite chronic pain has recently been

published³⁷³, as has a protocol for a proposed GWAS of baseline pain sensitivity and nociceptive sensitisation in healthy volunteers³⁷⁴. In future, I plan to investigate genetic correlations between these pain phenotypes and CTS, including Mendelian randomisation analyses to elucidate whether these pain phenotypes lie on the causal pathway for CTS. This work could potentially be extended to psychiatric comorbidities of CTS, such as anxiety and depression.

6.5 Concluding remarks

This year marks 66 years since Watson and Crick's discovery of the structure of the double helix in 1953. By way of comparison, 66 years after the Wright brothers' first demonstration of powered flight in 1903, Neil Armstrong walked on the moon. The term "genomic revolution" has been used liberally since early this century to describe the state of progress in human genetics, but do the tangible fruits of genetics justify the use of such a term?

Yes and no. The sequencing of the first human genome nearly twenty years ago brought with it immense promise for the future of medicine³⁷⁵. There have been a considerable number of well-publicised success stories, but the overwhelming majority of these concern gene therapy for rare, monogenic disorders³⁷⁶. When it comes to complex diseases, success has been more muted. While "personalised medicine" from the use of polygenic risk scores promises improved diagnosis and precise therapeutic targeting, the bioinformatic methods developed for disease association in large population cohorts do not currently confer the precision necessary to resolve disease risk at an individual level^{143,377}. As was touched upon in the previous chapter, the personalised utility of polygenic risk scores is hampered by the immense complexity of the genotype-phenotype relationship. As things currently stand, polygenic risk scores are not the genetic fortune tellers that they are sometimes purported to be, and we need to temper our expectations.

That being said, we can be guardedly optimistic that we are on the right trajectory. UK Biobank has recently released a tranche of whole exome sequencing (WES) data for c. 50,000 participants³⁷⁸, greatly expanding its utility as a genetic resource. NHS England launched the 100,000 genomes project in 2012, with the aim of sequencing the whole genomes (including cancer genomes) of 70,000 NHS patients with cancer, rare disorders and infectious diseases³⁷⁹.

Fuelled by the exponential fall in the cost of sequencing and bioinformatic advances, whole-genome sequencing on a population level is a credible proposition within the next two decades. Sixty-six years on from the double helix discovery, I believe that genetics is still awaiting its moon landing, but we are probably not very far off.

Amid such optimism for the future of genomic medicine, there is cause for pessimism for the future of CTS. NHS England, which proclaims on its website that it stands at the forefront of the genomic revolution³⁸⁰ also recently published a recommendation to restrict funding for 17 surgical procedures³⁸¹, one of which is carpal tunnel surgery. This is regrettable as much as it is misguided³⁸². CTS incidence is increasing³⁷ in the midst of the obesity epidemic in the West. If surgical treatment is significantly rationed, the likelihood of patients presenting to a hand surgeon after irreversible nerve damage has occurred is considerably greater.

CTS will never receive the same publicity or funding as cancer or cardiovascular disease. However, as clinicians who treat the disease, we see how the seemingly trivial pathology of a “pinched nerve” can profoundly affect the lives of our patients. It is incontrovertibly the case that there is an unmet clinical need for better treatments; gaining a more sophisticated understanding of the biology of the disease is the necessary first step to achieve this. The research presented herein has opened the door to the feasibility of using computational genetics in the pursuit of this endeavour. With the advent of population-wide whole genome sequencing, genomics has the potential to assist in the diagnosis and prognostication of CTS, such as the identification of individuals at risk of rapid disease progression, or those for whom repeated surgery is likely to be futile. There is also a strong case to be made for research into the development of novel non-surgical treatments that are less transient than steroid injections.

The work presented in this thesis represents the first crude sketch of the genetic landscape of CTS on a canvas that was previously blank. It is sincerely hoped that this first exposition of the genetic architecture of CTS will motivate future investigations in the field.

Scientific outputs

Principal publication

Wiberg A, Ng M, Schmid AB, Smillie RW, Baskozos G, Holmes MV, Bennett DL, Furniss D. A genome-wide association analysis identifies 16 novel susceptibility loci for Carpal Tunnel Syndrome. *Nature Communications*. 2019 Mar 4;10(1):1030.

Published abstracts

Wiberg A, Ng M, Schmid AB, Baskozos G, Bennett DL, Furniss D. A genetic risk score predicts the need for surgery in Carpal Tunnel Syndrome. *J Hand Surg (European Volume)* (2018):43(Supplement 3) S14. doi: 10.1177/1753193418800428

Wiberg A, Ng M, Schmid AB, Baskozos G, Bennett DL, Furniss D. Abstract: A Genetic Risk Score for Carpal Tunnel Syndrome. *PRS Global Open* (2018):6 Issue 8S 64. PSTM 2018 Abstract Supplement. doi: 10.1097/01.GOX.0000546801.74134.68

Wiberg A, Ng M, Schmid AB, Baskozos G, Bennett DL, Furniss D. A genome-wide association study of carpal tunnel syndrome. *J Hand Surg (European Volume)* (2018):43(Supplement 2) S142-S143. doi: 10.1177/1753193418769834

Presentations at scientific meetings

Sandy-Hindmarch O, Furniss D, **Wiberg A**, Baskozos G, Blat D, Bennett DL, Schmid AB. “Investigating Interactions between the Immune System and Peripheral Nervous System in the Initiation and Maintenance of Neuropathic Pain”. (Poster presentation at the 7th International Congress on Neuropathic Pain, London, May 2019).

Smillie RW, Ng M, Schmid AB, Baskozos G, Holmes MV, Bennett DL, Furniss D, **Wiberg A**. “Genetic evidence for a causal role of body mass index in the aetiology of carpal tunnel syndrome”. (Oral presentation at the BSSH Spring Scientific Meeting, Swansea, April 2019).

Wiberg A, Ng M, Schmid AB, Baskozos G, Bennett DL, Furniss D. “A Genetic Risk Score for Carpal Tunnel Syndrome”. (Oral presentations at the British Association of Plastic, Reconstructive and Aesthetic Surgeons (BAPRAS) Winter Scientific Meeting, London, November 2018).

Wiberg A, Ng M, Schmid AB, Baskozos G, Bennett DL, Furniss D. “A genetic risk score predicts the need for surgery in Carpal Tunnel Syndrome”. (Oral presentation at the British Society for Surgery of the Hand (BSSH) Autumn Scientific Meeting, London, October 2018).

Wiberg A, Ng M, Schmid AB, Baskozos G, Bennett DL, Furniss D. “Identification of 13 Novel Susceptibility Loci in a Genome-wide Association Study of Carpal Tunnel Syndrome”. (Poster presentation by W. Morita on behalf of the authors, at the 33rd Annual Research Meeting of the Japanese Orthopaedic Association, Nara, Japan, October 2018).

Wiberg A, Ng M, Schmid AB, Baskozos G, Bennett DL, Furniss D. “A Genetic Risk Score for Carpal Tunnel Syndrome”. (Oral presentation at the Plastic Surgery: The Meeting (American Society of Plastic Surgeons annual congress), Chicago, Illinois, September 2018).

Wiberg A, Ng M, Schmid AB, Smillie RW, Baskozos G, Holmes MV, Bennett DL, Furniss D. “A Genome-wide Association Study of Carpal Tunnel Syndrome Reveals 16 Novel Susceptibility Loci”. (Poster presentation at the Genetics of Neuropathic Pain and its Comorbidities Meeting, The Jackson Laboratory, Bar Harbor, Maine, September 2018).

Wiberg A, Furniss D. “Insights into the Molecular Pathology of Carpal Tunnel Syndrome” and “The Genetic Basis of Human Handedness”. (Oral presentation at the National Plastic Surgery Research Forum, Royal Free Hospital, London, June 2018).

Wiberg A, Ng M, Schmid AB, Baskozos G, Bennett DL, Furniss D. “A Genome-wide Association Study of Carpal Tunnel Syndrome”. (Oral presentation at the Federation of European Societies for Surgery of the Hand (FESSH) Congress, Copenhagen, Denmark, June 2018).

Wiberg A, Ng M, Schmid AB, Baskozos G, Bennett DL, Furniss D. “A Genome-wide Association Study of Carpal Tunnel Syndrome reveals 13 novel susceptibility loci”. (Oral presentation at the British Society for Surgery of the Hand (BSSH) Spring Scientific Meeting, Cardiff, May 2018).

Bibliography

1. Alpenfels, E. The anthropology and social significance of the human hand. *Artif Limbs* **2**, 4–21 (1995).
2. Timpson, N. J., Greenwood, C. M. T., Soranzo, N., Lawson, D. J. & Richards, J. B. Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nat. Rev. Genet.* **19**, 110–124 (2017).
3. Young, R. W. Evolution of the human hand: The role of throwing and clubbing. *J. Anat.* **202**, 165–174 (2003).
4. Duncan, S. F. M., Bhate, O. & Mustaly, H. Pathophysiology of carpal tunnel syndrome. in *Carpal Tunnel Syndrome and Related Median Neuropathies* (Springer International Publishing AG, 2017).
5. American Academy of Orthopaedic Surgeons Work Group Panel. *Clinical practice guideline on the diagnosis of carpal tunnel syndrome*. (2009).
6. England, J. D. Entrapment neuropathies. *Curr. Opin. Neurol.* **12**, 597–602 (1999).
7. Padua, L. *et al.* Carpal tunnel syndrome: clinical features, diagnosis, and management. *Lancet Neurol.* **15**, 1273–1284 (2016).
8. Atroshi, I. *et al.* Prevalence of carpal tunnel syndrome in a general population. *JAMA* **282**, 153–158 (1999).
9. Ferry, S., Pritchard, T., Keenan, J., Croft, P. & Silman, A. J. Estimating the Prevalence of Delayed Median Nerve Conduction in the General Population. *Br. J. Rheumatol.* **37**, 630–635 (1998).
10. de Krom, M. C. *et al.* Carpal tunnel syndrome: prevalence in the general population. *J. Clin. Epidemiol.* **45**, 373–376 (1992).
11. Practice parameter for carpal tunnel syndrome (Summary statement). *Neurology* **43**, 2406–2406 (1993).
12. Bland, J. D. P. Clinical surveillance of carpal tunnel syndrome in two areas of the United Kingdom. *J Neurol Neurosurg Psychiatry* **74**, 1674–1679 (2003).
13. Stevens, J. C., Sun, S., Beard, C. M., O’Fallon, W. M. & Kurland, L. T. Carpal tunnel syndrome in Rochester, Minnesota, 1961 to 1980. *Neurology* **38**, 134–8 (1988).

14. Mondelli, M., Giannini, F. & Giacchi, M. Carpal tunnel syndrome incidence in a general population. *Neurology* **58**, 289–294 (2002).
15. Szabo, R. & Madison, M. Carpal Tunnel Syndrome. *Orthop Clin North Am* **23**, 103–9 (1992).
16. Bland, J. D. P. Carpal tunnel syndrome. *BMJ* **335**, 343–346 (2007).
17. Bröske, J., Bednarski, M., Grzelec, H. & Zyluk, A. The usefulness of the Phalen test and the Hoffmann-Tinel sign in the diagnosis of carpal tunnel syndrome. *Acta Orthop. Belg.* **68**, 141–5 (2002).
18. AAEM Quality Assurance Committee *et al.* Literature review of the usefulness of nerve conduction studies and electromyography for the evaluation of patients with carpal tunnel syndrome. *Muscle Nerve* **16**, 1392–414 (1993).
19. Bland, J. D. P. Do nerve conduction studies predict the outcome of carpal tunnel decompression? *Muscle and Nerve* **24**, 935–40 (2001).
20. Fowler, J. R., Gaughan, J. P. & Ilyas, A. M. The sensitivity and specificity of ultrasound for the diagnosis of carpal tunnel syndrome: A meta-analysis. *Clin. Orthop. Relat. Res.* **469**, 1089–94 (2011).
21. Brienza, M. *et al.* 3T diffusion tensor imaging and electroneurography of peripheral nerve: A morphofunctional analysis in carpal tunnel syndrome. *J. Neuroradiol.* **41**, 124–30 (2014).
22. Gerritsen, A. A. M. *et al.* Splinting vs surgery in the treatment of Carpal tunnel syndrome: A randomized controlled trial. *JAMA* **288**, 1245–51 (2002).
23. Atroshi, I., Flondell, M., Hofer, M. & Ranstam, J. Methylprednisolone injections for the carpal tunnel syndrome: A randomized, placebo-controlled trial. *Ann. Intern. Med.* **159**, 309–17 (2013).
24. Bickel, K. Carpal tunnel syndrome. *J Hand Surg Am* **35**, 147–152 (2010).
25. Shi, Q. & MacDermid, J. C. Is surgical intervention more effective than non-surgical treatment for carpal tunnel syndrome? A systematic review. *J. Orthop. Surg. Res.* **6**, 17 (2011).
26. Soltani, A. M., Allan, B. J., Best, M. J., Mir, H. S. & Panthaki, Z. J. Revision decompression and collagen nerve wrap for recurrent and persistent compression neuropathies of the upper extremity. *Ann. Plast. Surg.* **72**, 572–8 (2014).
27. Benson, L. *et al.* Complications of endoscopic and open carpal tunnel release. *Arthroscopy* **22**, 919–924 (2006).
28. Louie, D., Earp, B. & Blazar, P. Long-term outcomes of carpal tunnel release: A critical review of the literature. *Hand* **7**, 242–246 (2012).

29. Bland, J. Carpal tunnel syndrome. *Curr. Opin. Neurol.* **18**, 581–585 (2005).
30. Stütz, N. M., Gohritz, A., van Schoonhoven, J. & Lanz, U. Revision surgery after carpal tunnel release - Analysis of the pathology in 200 cases during a 2 year period. *J. Hand Surg. Am.* **31**, 68–71 (2006).
31. Tung, T. & Mackinnon, S. Secondary carpal tunnel surgery. *Plast. Reconstr. Surg.* **107**, 1830–1843 (2001).
32. Hutting, K. H. & van Uchelen, J. H. Treatment of Recurrent and Persistent Carpal Tunnel Syndrome: The Fasciocutaneous Dorsal Ulnar Artery Flap. *J. Hand Surg. Am.* **43**, 686.e1–686.e5 (2018).
33. Zieske, L., Ebersole, G. C., Davidge, K., Fox, I. & Mackinnon, S. E. Revision carpal tunnel surgery: A 10-year review of intraoperative findings and outcomes. *J. Hand Surg. Am.* **38**, 1530–1539 (2013).
34. Palmer, D. H. & Hanrahan, L. P. Social and economic costs of carpal tunnel surgery. *Instr. Course Lect.* **44**, 167–72 (1995).
35. Burton, C. L., Chen, Y., Chesterton, L. S. & Van Der Windt, D. A. Trends in the prevalence, incidence and surgical management of carpal tunnel syndrome between 1993 and 2013: An observational analysis of UK primary care records. *BMJ Open* **8**, 1–11 (2018).
36. Foley, M.; Silverstein, B.; Polissar, N. The Economic Burdens of Carpal Tunnel Syndrome: Long-Term Earnings of CTS Claimants in Washington State. *Am. J. Ind. Med.* **50**, 155–172 (2007).
37. Bebbington, E. & Furniss, D. Linear regression analysis of Hospital Episode Statistics predicts a large increase in demand for elective hand surgery in England. *J. Plast. Reconstr. Aesthetic Surg.* **68**, 243–251 (2015).
38. Stecco, C. & Aldegheri, R. Historical review of carpal tunnel syndrome. *Chir. Organi Mov.* **92**, 7–10 (2008).
39. Putnam, J. A series of cases of paresthesias, mainly of the hand, or periodic recurrence, and possibly of vaso-motor origin. *Arch Med* **4**, 147–162 (1880).
40. Hunt, J. The neural atrophy of the muscle of the hand, without sensory disturbances. *Rev Neurol Psych* **12**, 137–148 (1914).
41. Marie, P. & Foix, C. Atrophie isolée de l'éminence thenar d'origine névritique: rôle du ligament annulaire antérieur du carpe dans la pathogénie de la lésion. *Rev Neurol* **26**, 647–649 (1913).
42. Cannon, B. & Love, J. G. Tardy median palsy; median neuritis; median thenar neuritis amenable to surgery. *Surgery* **20**, 210–216 (1946).

43. Gelberman, R. H., Hergenroeder, P. T., Hargens, A. R., Lundborg, G. N. & Akeson, W. H. The carpal tunnel syndrome. A study of carpal canal pressures. *J. Bone Joint Surg. Am.* **63**, 380–3 (1981).
44. Seradge, H., Jia, Y. C. & Owens, W. In vivo measurement of carpal tunnel pressure in the functioning hand. *J. Hand Surg. Am.* **20**, 855–859 (1995).
45. Coppieters, M. W., Schmid, A. B., Kubler, P. A. & Hodges, P. W. Description, reliability and validity of a novel method to measure carpal tunnel pressure in patients with carpal tunnel syndrome. *Man. Ther.* **17**, 589–592 (2012).
46. Rydevik, B., Lundborg, G. & Bagge, U. Effects of graded compression on intraneural blood flow: An in vivo study on rabbit tibial nerve. *J. Hand Surg. Am.* **6**, 3–12 (1981).
47. Lundborg, G., Gelberman, R. H., Minteer-Convery, M., Lee, Y. F. & Hargens, A. R. Median nerve compression in the carpal tunnel—Functional response to experimentally induced controlled pressure. *J. Hand Surg. Am.* **7**, 252–259 (1982).
48. Rydevik, B. & Lundborg, G. Permeability of intraneural microvessels and perineurium following acute, graded experimental nerve compression. *Scand. J. Plast. Reconstr. Surg.* **11**, 179–87 (1977).
49. Mackinnon, S. E. Pathophysiology of nerve compression. *Hand Clin.* **18**, 231–241 (2002).
50. Millesi, H., Zöch, G. & Rath, T. The gliding apparatus of peripheral nerve and its clinical significance. *Ann. Chir. la Main* **9**, 87–97 (1990).
51. Byl, C., Puttlitz, C., Byl, N., Lotz, J. & Topp, K. Strain in the median and ulnar nerves during upper-extremity positioning. *J. Hand Surg. Am.* **27**, 1032–1040 (2002).
52. Ogata, K. & Naito, M. Blood flow of peripheral nerve effects of dissection, stretching and compression. *J. Hand Surg. Br.* **11**, 10–14 (1986).
53. Dahlin, L. B., Sjöstrand, J. & McLean, W. G. Graded inhibition of retrograde axonal transport by compression of rabbit vagus nerve. *J. Neurol. Sci.* **76**, 221–230 (1986).
54. Wall, E. J. *et al.* Experimental stretch neuropathy. Changes in nerve conduction under tension. *J. Bone Joint Surg. Br.* **74**, 126–129 (1992).
55. Kiernan, M., Mogyoros, I. & Burke, D. Conduction block in carpal tunnel syndrome. *Brain* **122**, 933–941 (1999).
56. Thomas, P. K. & Fullerton, P. M. Nerve Fibre Size in the Carpal Tunnel Syndrome. *J. Neurol. Neurosurg. Psychiatry* **26**, 520–527 (1963).
57. Schmid, A. B., Bland, J. D. P., Bhat, M. A. & Bennett, D. L. H. The relationship of nerve fibre pathology to sensory function in entrapment neuropathy. *Brain* **137**, 3186–3199 (2014).

58. Oh, J. *et al.* Immunolocalization of collagen types in the subsynovial connective tissue within the carpal tunnel in humans. *J. Orthop. Res.* **23**, 1226–1231 (2005).
59. Festen-Schrier, V. J. M. M. & Amadio, P. C. The biomechanics of subsynovial connective tissue in health and its role in carpal tunnel syndrome. *J. Electromyogr. Kinesiol.* **38**, 232–239 (2018).
60. Jinrok, O. *et al.* Vascular pathologic changes in the flexor tenosynovium (subsynovial connective tissue) in idiopathic carpal tunnel syndrome. *J. Orthop. Res.* **22**, 1310–1315 (2004).
61. Fomby, P. & Cherlin, A. J. Sonographic Measurements of Subsynovial Connective Tissue Thickness in Patients With Carpal Tunnel Syndrome. *J Ultrasound Med* **72**, 181–204 (2011).
62. Fuchs, P. C., Nathan, P. A. & Myers, L. D. Synovial histology in carpal tunnel syndrome. *J. Hand Surg. Am.* **16**, 753–758 (1991).
63. Kerr, C. D., Sybert, D. R. & Albarracin, N. S. An analysis of the flexor synovium in idiopathic carpal tunnel syndrome: Report of 625 cases. *J. Hand Surg. Am.* **17**, 1028–1030 (1992).
64. Nakamichi, K. I. & Tachibana, S. Histology of the transverse carpal ligament and flexor tenosynovium in idiopathic carpal tunnel syndrome. *J. Hand Surg. Am.* **23**, 1015–1024 (1998).
65. Schuind, F., Ventura, M. & Pasteels, J. L. Idiopathic carpal tunnel syndrome: Histologic study of flexor tendon synovium. *J. Hand Surg. Am.* **15**, 497–503 (1990).
66. Chikenji, T. *et al.* Transforming growth factor- β (TGF- β) expression is increased in the subsynovial connective tissues of patients with idiopathic carpal tunnel syndrome. *J. Orthop. Res.* **32**, 116–122 (2014).
67. Ettema, A. M., Amadio, P. C., Zhao, C., Wold, L. E. & An, K.-N. A histological and immunohistochemical study of the subsynovial connective tissue in idiopathic carpal tunnel syndrome. *J. Bone Joint Surg. Am.* **86-A**, 1458–66 (2004).
68. Kim, J. K., Koh, Y. Do, Kim, J. S., Hann, H. J. & Kim, M. J. Oxidative stress in subsynovial connective tissue of idiopathic carpal tunnel syndrome. *J. Orthop. Res.* **28**, 1463–1468 (2010).
69. Hirata, H. *et al.* The relationship of VEGF and PGE2 expression to extracellular matrix remodelling of the tenosynovium in the carpal tunnel syndrome. *J. Pathol.* **204**, 605–612 (2004).
70. Hirata, H. *et al.* MMP-2 expression is associated with rapidly proliferative arteriosclerosis in the flexor tenosynovium and pain severity in carpal tunnel syndrome. *J. Pathol.* **205**,

- 443–450 (2005).
71. Chammas, M. *et al.* Carpal tunnel syndrome - Part I (anatomy, physiology, etiology and diagnosis). *Rev. Bras. Ortop.* **49**, 429–436 (2014).
 72. Ibrahim, I., Khan, W. ., Goddard, N. & Smitham, P. Carpal Tunnel Syndrome: A Review of the Recent Literature. *Open Orthop. J.* **6**, 69–76 (2012).
 73. Moghtaderi, A., Izadi, S. & Sharafadinzadeh, N. An evaluation of gender, body mass index, wrist circumference and wrist ratio as independent risk factors for carpal tunnel syndrome. *Acta Neurol. Scand.* **112**, 375–379 (2005).
 74. Johnson, E. W., Gatens, T., Poindexter, D. & Bowers, D. Wrist Dimensions: correlation with median sensory latencies. *Arch. Phys. Med. Rehabil.* **64**, 556–557 (1983).
 75. Kouyoumdjian, J. A., Morita, M. P. A., Rocha, P. R. F., Miranda, R. C. & Gouveia, G. M. Wrist and palm indexes in carpal tunnel syndrome. *Arq. Neuropsiquiatr.* **58**, 625–629 (2000).
 76. Mondelli, M. *et al.* Anthropometric Measurements as a Screening Test for Carpal Tunnel Syndrome: Receiver Operating Characteristic Curves and Accuracy. *Arthritis Care Res. (Hoboken)*. **67**, 691–700 (2015).
 77. Violante, F. S. *et al.* Carpal tunnel syndrome and manual work: A longitudinal study. *J. Occup. Environ. Med.* **11**, 1189–96 (2007).
 78. Mattioli, S. *et al.* Risk factors for operated carpal tunnel syndrome: A multicenter population-based case-control study. *BMC Public Health* **9**, 343 (2009).
 79. Shiri, R., Pourmemari, M. H., Falah-Hassani, K. & Viikari-Juntura, E. The effect of excess body mass on the risk of carpal tunnel syndrome: A meta-analysis of 58 studies. *Obes. Rev.* **16**, 1094–1104 (2015).
 80. Kurt, S. *et al.* Obesity and carpal tunnel syndrome: is there a causal relationship? *Eur Neurol* **59**, 253–257 (2008).
 81. Kozak, A. *et al.* Association between work-related biomechanical risk factors and the occurrence of carpal tunnel syndrome: An overview of systematic reviews and a meta-analysis of current research. *BMC Musculoskelet. Disord.* **16**, (2015).
 82. Frederick, M., Musolin, K., Beckman, J., Harrison, R. & Jackson, R. Rates of Carpal Tunnel Syndrome in a State Workers' Compensation Information System, by Industry and Occupation — California, 2007–2014. *Morb. Mortal. Wkly. Rep.* **67**, 1094–1097 (2018).
 83. Luckhaupt, S. E. *et al.* Prevalence and work-relatedness of carpal tunnel syndrome in the working population, United States, 2010 national health interview survey. *Am. J. Ind. Med.* **56**, 615–624 (2013).

84. Padua, L. *et al.* Systematic review of pregnancy-related carpal tunnel syndrome. *Muscle and Nerve* **42**, 697–702 (2010).
85. Padua, L. *et al.* Symptoms and neurophysiological picture of carpal tunnel syndrome in pregnancy. *Clin. Neurophysiol.* **112**, 1946–1951 (2001).
86. Kaplan, Y., Kurt, S. G. & Karaer, H. Carpal tunnel syndrome in postmenopausal women. *J. Neurol. Sci.* **270**, 77–81 (2008).
87. Vessey, M. P., Villard-Mackintosh, L. & Yeates, D. Epidemiology of carpal tunnel syndrome in women of childbearing age. Findings in a large cohort study. *Int. J. Epidemiol.* **19**, 655–9 (1990).
88. Pascual, E. *et al.* Higher incidence of carpal tunnel syndrome in oophorectomized women. *Rheumatology* **30**, 60–2 (1991).
89. Manson, J. E. *et al.* Menopausal hormone therapy and the incidence of carpal tunnel syndrome in postmenopausal women: Findings from the Women’s Health Initiative. *PLoS One* **13**, e0207509 (2018).
90. Spagnolo, F., Sestak, I., Howell, A., Forbes, J. F. & Cuzick, J. Anastrozole-induced carpal tunnel syndrome: Results from the international breast cancer intervention study II prevention trial. *J. Clin. Oncol.* **34**, 139–143 (2016).
91. Khosrawi, S. & Maghrouri, R. The prevalence and severity of carpal tunnel syndrome during pregnancy. *Adv. Biomed. Res.* **1**, 43 (2012).
92. Tseng, C. H. *et al.* Medical and non-medical correlates of carpal tunnel syndrome in a Taiwan cohort of one million. *Eur. J. Neurol.* **19**, 91–97 (2012).
93. Pourmemari, M. H. & Shiri, R. Diabetes as a risk factor for carpal tunnel syndrome: A systematic review and meta-analysis. *Diabet. Med.* **33**, 10–16 (2016).
94. Singh, R., Gamble, G. & Cundy, T. Lifetime risk of symptomatic carpal tunnel syndrome in Type 1 diabetes. *Diabet. Med.* **22**, 625–630 (2005).
95. Dahlin, L. B. *et al.* Low myelinated nerve-fibre density may lead to symptoms associated with nerve entrapment in vibration-induced neuropathy. *J. Occup. Med. Toxicol.* **9**, 1–6 (2014).
96. Goldin, A., Beckman, J. A., Schmidt, A. M. & Creager, M. A. Advanced glycation end products: Sparking the development of diabetic vascular injury. *Circulation* **114**, 597–605 (2006).
97. Esposito, K. *et al.* Inflammatory cytokine concentrations are acutely increased by hyperglycemia in humans: Role of oxidative stress. *Circulation* **106**, 2067–72 (2002).
98. Ali, R. *et al.* Molecular and pathological studies in the posterior interosseous nerve of diabetic and non-diabetic patients with carpal tunnel syndrome. *Diabetologia* **57**, 1711–

- 1719 (2014).
99. Geoghegan, J. M., Clark, D. I., Bainbridge, L. C., Smith, C. & Hubbard, R. Risk factors in carpal tunnel syndrome. *J. Hand Surg. Am.* **29**, 315–320 (2004).
 100. Karadag, O. *et al.* Sonographic assessment of carpal tunnel syndrome in rheumatoid arthritis: Prevalence and correlation with disease activity. *Rheumatol. Int.* **32**, 2313–9 (2012).
 101. Oktayoglu, P. *et al.* Assessment of the presence of carpal tunnel syndrome in patients with diabetes mellitus, hypothyroidism and acromegaly. *J. Clin. Diagnostic Res.* **9**, OC14–OC18 (2015).
 102. Eslamian, F. *et al.* Electrophysiologic changes in patients with untreated primary hypothyroidism. *J. Clin. Neurophysiol.* **28**, 323–8 (2011).
 103. Bottacchi, E. *et al.* Polyneuropathy in hypothyroidism: clinical, electrophysiological and morphological findings in four cases. *J. Neurol. Neurosurg. Psychiatry* **50**, 1454–1460 (2008).
 104. Karne, S. S. & Bhalerao, N. S. Carpal tunnel syndrome in hypothyroidism. *J. Clin. Diagnostic Res.* **10**, OC36–OC38 (2016).
 105. Kasem, A. A. Carpal Tunnel Syndrome in Hypothyroid Patients: The Effect of Hormone Replacement Therapy. *Am. J. Intern. Med.* **2**, 54 (2014).
 106. Soon, B., Vicenzino, B., Schmid, A. B. & Coppieters, M. W. Facilitatory and inhibitory pain mechanisms are altered in patients with carpal tunnel syndrome. *PLoS One* **12**, 1–11 (2017).
 107. Fernandez-De-Las-Peas, C. *et al.* Bilateral widespread mechanical pain sensitivity in carpal tunnel syndrome: Evidence of central processing in unilateral neuropathy. *Brain* **132**, 1472–1479 (2009).
 108. Zanette, G., Cacciatori, C. & Tamburin, S. Central sensitization in carpal tunnel syndrome with extraterritorial spread of sensory symptoms. *Pain* **148**, 227–36 (2010).
 109. McCallum, L. M., Damms, N. A., Sarrigiannis, P. G. & Zis, P. Anxiety and depression in patients with suspected carpal tunnel syndrome - A case controlled study. *Brain Behav.* **9**, e01342 (2019).
 110. Beleckas, C. M. *et al.* Relative Prevalence of Anxiety and Depression in Patients With Upper Extremity Conditions. *J. Hand Surg. Am.* **43**, 571.e1-571.e8 (2018).
 111. Nunez, F., Vranceanu, A. M. & Ring, D. Determinants of pain in patients with carpal tunnel syndrome. *Clin. Orthop. Relat. Res.* **468**, 3328–32 (2010).
 112. Shin, Y. H., Yoon, J. O., Kim, Y. K. & Kim, J. K. Psychological Status Is Associated With Symptom Severity in Patients With Carpal Tunnel Syndrome. *J. Hand Surg. Am.* **43**,

- 484.e1-484.e8 (2018).
113. Lozano-Calderón, S., Anthony, S. & Ring, D. The Quality and Strength of Evidence for Etiology: Example of Carpal Tunnel Syndrome. *J. Hand Surg. Am.* **33**, 525–538 (2008).
 114. Radecki, P. The familial occurrence of carpal tunnel syndrome. *Muscle Nerve* **17**, 325–330 (1994).
 115. Hemminki, K., Li, X. & Sundquist, K. Familial risks for nerve, nerve root and plexus disorders in siblings based on hospitalisations in Sweden. *J. Epidemiol. Community Health* **61**, 80–4 (2007).
 116. Hakim, A. J., Cherkas, L., El Zayat, S., MacGregor, A. J. & Spector, T. D. The genetic contribution to carpal tunnel syndrome in women: a twin study. *Arthritis Rheum.* **47**, 275–279 (2002).
 117. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era - Concepts and misconceptions. *Nat. Rev. Genet.* **9**, 255–266 (2008).
 118. Mucci, L. A. *et al.* Familial risk and heritability of cancer among twins in nordic countries. *JAMA* **315**, 68–76 (2016).
 119. Wirdefeldt, K., Gatz, M., Reynolds, C. A., Prescott, C. A. & Pedersen, N. L. Heritability of Parkinson disease in Swedish twins: A longitudinal study. *Neurobiol. Aging* **32**, 1923.e1-1923.e8 (2011).
 120. Paul, N. L. M. *et al.* Genetic Heritability of Ischemic Stroke and the Contribution of Previously Reported Candidate Gene and Genomewide Associations. *Stroke* **43**, 3161–3167 (2012).
 121. Elstner, M. *et al.* Familial carpal tunnel syndrome: further evidence for a genetic contribution. *Clin. Genet.* **69**, 179–82 (2006).
 122. Gossett, J. G. & Chance, P. F. Is there a familial carpal tunnel syndrome? An evaluation and literature review. *Muscle Nerve* **21**, 1533–1536 (1998).
 123. Danta, G. Familial carpal tunnel syndrome with onset in childhood. *J. Neurol. Neurosurg. Psychiatry* **38**, 350–355 (1975).
 124. Michaud, L. J., Hays, R. M., Dudgeon, B. J. & Kropp, R. J. Congenital carpal tunnel syndrome: case report of autosomal dominant inheritance and review of the literature. *Arch. Phys. Med. Rehabil.* **71**, 430–2 (1990).
 125. Braddom, R. Familial carpal tunnel syndrome in three generations of a black family. *Am J Phys Med* **64**, 227–34 (1985).
 126. De Smet, L. & Fabry, G. Carpal tunnel syndrome: Familial occurrence presenting in childhood. *J. Pediatr. Orthop. Part B* **8**, 127–8 (1999).
 127. Alford, J., Weiss, A. & Akelman, E. The familial incidence of carpal tunnel syndrome in

- patients with unilateral and bilateral disease. *Am J Orthop (Belle Mead NJ)* **33**, 397–400 (2040).
128. Chance, P. F. *et al.* DNA deletion associated with hereditary neuropathy with liability to pressure palsies. *Cell* **72**, 143–151 (1993).
 129. Lupski, J. R. *et al.* Whole-Genome Sequencing in a Patient with Charcot–Marie–Tooth Neuropathy. *N. Engl. J. Med.* **362**, 1181–1191 (2010).
 130. Hund, E., Linke, R. P., Willig, F. & Grau, A. Transthyretin-associated neuropathic amyloidosis: Pathogenesis and treatment. *Neurology* **56**, 431–435 (2001).
 131. Valentino, P. *et al.* Genetic screening for familial amyloid polyneuropathy in patients with idiopathic carpal tunnel syndrome. *J. Peripher. Nerv. Syst.* **13**, 151–2 (2008).
 132. Burger, M. C., de Wet, H. & Collins, M. Interleukin and growth factor gene variants and risk of carpal tunnel syndrome. *Gene* **564**, 67–72 (2015).
 133. Burger, M., de Wet, H. & Collins, M. The COL5A1 gene is associated with increased risk of carpal tunnel syndrome. *Clin. Rheumatol.* **34**, 767–774 (2015).
 134. Dada, S., Burger, M. C., Massij, F., de Wet, H. & Collins, M. Carpal tunnel syndrome: The role of collagen gene variants. *Gene* **587**, 53–58 (2016).
 135. Burger, M. C., De Wet, H. & Collins, M. The BGN and ACAN genes and carpal tunnel syndrome. *Gene* **551**, 160–166 (2014).
 136. McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* **9**, 356–369 (2008).
 137. Xue, A. *et al.* Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat. Commun.* **9**, 2941 (2018).
 138. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
 139. Wang, W. Y. S., Barratt, B. J., Clayton, D. G. & Todd, J. A. Genome-wide association studies: Theoretical and practical concerns. *Nat. Rev. Genet.* **6**, 109–118 (2005).
 140. Price, A. L., Spencer, C. C. A. & Donnelly, P. Progress and promise in understanding the genetic basis of common diseases. *Proc. R. Soc. B Biol. Sci.* **282**, (2015).
 141. Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371–5 (2014).
 142. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
 143. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* **19**, 581–590 (2018).
 144. Chasioti, D., Yan, J., Nho, K. & Saykin, A. J. Progress in Polygenic Composite Scores in

- Alzheimer's and Other Complex Diseases. *Trends Genet.* **35**, 371–382 (2019).
145. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
 146. Gratten, J., Wray, N. R., Keller, M. C. & Visscher, P. M. Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nat. Neurosci.* **17**, 782–790 (2014).
 147. Kraft, P., Zeggini, E. & Ioannidis, J. P. A. Replication in Genome-Wide Association Studies. *Stat. Sci.* **24**, 561–573 (2009).
 148. Collins, R. What makes UK Biobank special? *The Lancet* **379**, 1173–1174 (2012).
 149. Delgado-Rodríguez, M. & Llorca, J. Bias. *J. Epidemiol. Community Health* **58**, 635–666 (2004).
 150. Fry, A. *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with Those of the General Population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).
 151. Data providers and dates of data availability. Available at: https://biobank.ctsu.ox.ac.uk/crystal/exinfo.cgi?src=Data_providers_and_dates. (Accessed: 29th July 2019)
 152. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
 153. Hellwege, J. N. *et al.* Population Stratification in Genetic Association Studies. *Current protocols in human genetics* (2017). doi:10.1002/cphg.48
 154. Abraham, G. & Inouye, M. Fast principal component analysis of large-scale genome-wide data. *PLoS One* **9**, e93766 (2014).
 155. Lane, J. *et al.* Can we use routinely collected hospital and GP data for epidemiological study of common hand conditions? A UK Biobank based validation project. *bioRxiv* (2018). doi:<http://dx.doi.org/10.1101/274167>
 156. Loh, P. R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
 157. Loh, P. R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nature Genetics* **50**, 906–908 (2018).
 158. Day, F. R., Loh, P. R., Scott, R. A., Ong, K. K. & Perry, J. R. B. A Robust Example of Collider Bias in a Genetic Association Study. *Am. J. Hum. Genet.* **98**, 392–393 (2016).
 159. Pruim, R. J. *et al.* LocusZoom: Regional visualization of genome-wide association scan results. in *Bioinformatics* **26**, 2336–2337 (2010).
 160. Leitsalu, L. *et al.* Cohort profile: Estonian Biobank of the Estonian Genome Center,

- University of Tartu. *Int. J. Epidemiol.* **44**, 1137–1147 (2015).
161. Loh, P. R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
 162. Mitt, M. *et al.* Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur. J. Hum. Genet.* **25**, 869–876 (2017).
 163. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
 164. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
 165. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
 166. Benonisdottir, S. *et al.* Epigenetic and genetic components of height regulation. *Nat. Commun.* **7**, 13490 (2016).
 167. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
 168. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–53 (2009).
 169. Maher, B. Personal genomes: The case of the missing heritability. *Nature* **456**, 18–21 (2008).
 170. Wainschtein, P., Jain, D. P., Yengo, L. & Zheng, Z. Recovery of trait heritability from whole genome sequence data. *bioRxiv* doi: <http://dx.doi.org/10.1101/588020> (2019).
 171. Skol, A. D., Scott, L. J., Abecasis, G. R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* **38**, 209–213 (2006).
 172. Poirier, J. G. *et al.* Resampling to Address the Winner’s Curse in Genetic Association Analysis of Time to Event. *Genet. Epidemiol.* **39**, 518–528 (2015).
 173. Colquhoun, D. An investigation of the false discovery rate and the misinterpretation of p-values Subject Areas : Author for correspondence : *R. Soc. Open Sci.* **1**, 140216 (2014).
 174. Göring, H. H. H., Terwilliger, J. D. & Blangero, J. Large Upward Bias in Estimation of Locus-Specific Effects from Genomewide Scans. *Am. J. Hum. Genet.* **69**, 1357–1369 (2001).
 175. Gallagher, M. D. & Chen-Plotkin, A. S. The Post-GWAS Era: From Association to Function. *Am. J. Hum. Genet.* **102**, 717–730 (2018).

176. Fang, H., Knezevic, B., Burnham, K. L. & Knight, J. C. XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits. *Genome Med.* **8**, 129 (2016).
177. Nicolae, D. L. *et al.* Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
178. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
179. Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
180. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
181. Watanabe, K., Taskesen, E., Van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
182. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
183. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
184. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
185. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
186. Wu, Y. *et al.* Identification of the primate-specific gene *BTN3A2* as an additional schizophrenia risk gene in the MHC loci. *EBioMedicine* **44**, 530–541 (2019).
187. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
188. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
189. Fehrmann, R. S. N. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47**, 115–125 (2015).
190. Weiss T, Taschner-Mandl S, Bileck A, Slany A, Kromp F, Rifatbegovic F, Frech C, Windhager R, Kitzinger H, Tzou CH, Ambros PF, Gerner C, A. I. Proteomics and transcriptomics of peripheral nerve tissue and cells unravel new aspects of the human Schwann cell repair phenotype. *Glia* **64**, 2133–2153 (2016).
191. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

192. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
193. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
194. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
195. Zheng, J. *et al.* LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
196. De Jager, P. L. *et al.* Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *Lancet Neurol.* **8**, 1111–1119 (2009).
197. Amendola, L. M. *et al.* Actionable exomic incidental findings in 6503 participants: Challenges of variant classification. *Genome Res.* **25**, 305–315 (2015).
198. Wen, W. *et al.* Genome-wide association studies in East Asians identify new loci for waist-hip ratio and waist circumference. *Sci. Rep.* **6**, 17958 (2016).
199. He, L. *et al.* α -Mannosidase 2C1 attenuates PTEN function in prostate cancer cells. *Nat. Commun.* **2**, 307 (2011).
200. Yue, W. *et al.* Suppression of 6A8 α -mannosidase gene expression reduced the potentiality of growth and metastasis of human nasopharyngeal carcinoma. *Int. J. Cancer* **108**, 189–95 (2004).
201. Hill, W. D. *et al.* Molecular genetic aetiology of general cognitive function is enriched in evolutionarily conserved regions. *Transl. Psychiatry* **6**, e980 (2016).
202. Kelwick, R., Desanlis, I., Wheeler, G. N. & Edwards, D. R. The ADAMTS (A Disintegrin and Metalloproteinase with Thrombospondin motifs) family. *Genome Biol.* **16**, 113 (2015).
203. Hubmacher, D. *et al.* Unusual life cycle and impact on microfibril assembly of ADAMTS17, a secreted metalloprotease mutated in genetic eye disease. *Sci. Rep.* **7**, 41871 (2017).
204. Brocker, C. N., Vasiliou, V. & Nebert, D. W. Evolutionary divergence and functions of the ADAM and ADAMTS gene families. *Hum. Genomics* **4**, 43–55 (2009).
205. Tachmazidou, I. *et al.* Whole-Genome Sequencing Coupled to Imputation Discovers Genetic Signals for Anthropometric Traits. *Am. J. Hum. Genet.* **100**, 865–884 (2017).
206. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1082 (2009).

207. Izidoro, M. A. *et al.* A study of human furin specificity using synthetic peptides derived from natural substrates, and effects of potassium ions. *Arch. Biochem. Biophys.* **487**, 105–114 (2009).
208. Kamolz, L.-P. *et al.* Carpal tunnel syndrome: a question of hand and wrist configurations? *J. Hand Surg. Br.* **29**, 321–4 (2004).
209. Li, Z. M., Gabra, J. N., Marquardt, T. L. & Kim, D. H. Narrowing carpal arch width to increase cross-sectional area of carpal tunnel - a cadaveric study. *Clin. Biomech.* **28**, 402–407 (2013).
210. Dagoneau, N. *et al.* ADAMTS10 mutations in autosomal recessive Weill-Marchesani syndrome. *Am. J. Hum. Genet.* **75**, 801–6 (2004).
211. Shah, M. H., Bhat, V., Shetty, J. S. & Kumar, A. Whole exome sequencing identifies a novel splice-site mutation in ADAMTS17 in an Indian family with Weill-Marchesani syndrome. *Mol. Vis.* **20**, 790–6 (2014).
212. Gorlin, R., L'Heureux, R. & Shapiro, I. Weill-Marchesani syndrome in two generations: genetic heterogeneity or pseudodominance? *J. Pediatr. Ophthalmol. Strabismus* **11**, 139–144 (1974).
213. Dellon, A. L., Trojak, J. E. & Rochman, G. M. Median nerve compression in Weill Marchesani syndrome. *Plast. Reconstr. Surg.* **74**, 127–130 (1984).
214. Lin, Y. J. *et al.* Association of human height-related genetic variants with familial short stature in Han Chinese in Taiwan. *Sci. Rep.* **7**, 6372 (2017).
215. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
216. Lluch, A. Thickening of the synovium of the digital flexor tendons: cause or consequence of the carpal tunnel syndrome? *J. Hand Surg. Br.* **17**, 209–12 (1992).
217. Armstrong, T. J., Castelli, W. a, Evans, F. G. & Diaz-Perez, R. Some histological changes in carpal tunnel contents and their biomechanical implications. *J. Occup. Med.* **26**, 197–201 (1984).
218. Ellis, R., Blyth, R., Arnold, N. & Miner-Williams, W. Is there a relationship between impaired median nerve excursion and carpal tunnel syndrome? A systematic review. *J. Hand Ther.* **30**, 3–12 (2017).
219. Gupta, R., Rowshan, K., Chao, T., Mozaffar, T. & Steward, O. Chronic nerve compression induces local demyelination and remyelination in a rat model of carpal tunnel syndrome. *Exp. Neurol.* **187**, 500–508 (2004).
220. Zhang, Y. & Marmorstein, L. Y. Focus on molecules: Fibulin-3 (EFEMP1). *Experimental Eye Research* **90**, 374–375 (2010).

221. Zhou, Y.-H. *et al.* Weaponizing human EGF-containing fibulin-like extracellular matrix protein 1 (EFEMP1) for 21st century cancer therapeutics. *Oncoscience* **3**, (2016).
222. Jorgenson, E. *et al.* A genome-wide association study identifies four novel susceptibility loci underlying inguinal hernia. *Nat. Commun.* **6**, 10130 (2015).
223. Marmorstein, L. Y. *et al.* Aberrant accumulation of EFEMP1 underlies drusen formation in Malattia Leventinese and age-related macular degeneration. *Proc. Natl. Acad. Sci.* **99**, 13067–72 (2002).
224. McLaughlin, P. J. *et al.* Lack of fibulin-3 causes early aging and herniation, but not macular degeneration in mice. *Hum. Mol. Genet.* **16**, 3059–3070 (2007).
225. Lei, D. *et al.* MiR-338-5p suppresses proliferation, migration, invasion, and promote apoptosis of glioblastoma cells by directly targeting EFEMP1. *Biomed. Pharmacother.* **89**, 957–965 (2017).
226. Fang, M., Jacob, R., McDougal, O. & Oxford, J. T. Minor fibrillar collagens, variable regions alternative splicing, intrinsic disorder, and tyrosine sulfation. *Protein and Cell* **3**, 419–433 (2012).
227. Liu, W. *et al.* A genetic variant in COL11A1 is functionally associated with lumbar disc herniation in Chinese population. *J. Genet.* **96**, 867–872 (2017).
228. Hay, M. *et al.* Association of type XI collagen genes with chronic Achilles tendinopathy in independent populations from South Africa and Australia. *Br. J. Sports Med.* **47**, 569–574 (2013).
229. Dada, S., Burger, M. C., Massij, F., de Wet, H. & Collins, M. Carpal tunnel syndrome: The role of collagen gene variants. *Gene* **587**, 53–58 (2016).
230. Verrecchia, F. & Mauviel, A. Transforming growth factor- β signaling through the Smad pathway: Role in extracellular matrix gene expression and regulation. *Journal of Investigative Dermatology* **118**, 211–215 (2002).
231. Yamanaka, Y. *et al.* Blocking fibrotic signaling in fibroblasts from patients with carpal tunnel syndrome. *J. Cell. Physiol.* **233**, 2067–2074 (2017).
232. Dallas, S. L. *et al.* Fibronectin regulates latent transforming growth factor- β (TGF β) by controlling matrix assembly of latent TGF β -binding protein-1. *J. Biol. Chem.* **280**, 18871–18880 (2005).
233. Marshall, S., Tardif, G. & Ashworth, N. Local corticosteroid injection for carpal tunnel syndrome. *Cochrane database Syst. Rev.* CD001554 (2007).
234. Tritschler, I. *et al.* Modulation of TGF- β activity by latent TGF- β -binding protein 1 in human malignant glioma cells. *Int. J. Cancer* **125**, 530–540 (2009).
235. Eble, J. & Niland, S. The Extracellular Matrix of Blood Vessels. *Curr. Pharm. Des.* **15**,

- 1385–1400 (2009).
236. Rhodes, J. M. & Simons, M. The extracellular matrix and blood vessel formation: not just a scaffold. *J. Cell. Mol. Med.* **11**, 176–205 (2007).
 237. Barros, C. S., Franco, S. J. & Müller, U. Extracellular Matrix: Functions in the nervous system. *Cold Spring Harb. Perspect. Biol.* **3**, 1–24 (2011).
 238. Eldridge, C. F., Bunge, M. B. & Bunge, R. P. Differentiation of axon-related Schwann cells in vitro: II. Control of myelin formation by basal lamina. *J. Neurosci.* **9**, 625–638 (1989).
 239. Bunge, M. B., Bunge, R. P., Kleitman, N. & Dean, A. C. Role of peripheral nerve extracellular matrix in schwann cell function and in neurite regeneration. *Dev. Neurosci.* **11**, 348–360 (1989).
 240. Fricker, F. R. *et al.* Axonal neuregulin 1 is a rate limiting but not essential factor for nerve remyelination. *Brain* **136**, 2279–2297 (2013).
 241. Chance, P. F. Overview of hereditary neuropathy with liability to pressure palsies. *Ann. N. Y. Acad. Sci.* **883**, 14–21 (1999).
 242. Bai, Y. *et al.* Conduction Block in PMP22 Deficiency. *J. Neurosci.* **30**, 600–608 (2010).
 243. Panosyan, F. B. *et al.* Carpal tunnel syndrome in inherited neuropathies: A retrospective survey. *Muscle and Nerve* **57**, 388–394 (2018).
 244. Hendriks, S. H. *et al.* Type 2 diabetes seems not to be a risk factor for the carpal tunnel syndrome: A case control study. *BMC Musculoskelet. Disord.* **15**, 1–5 (2014).
 245. Shiri, R., Pourmemari, M. H., Falah-Hassani, K. & Viikari-Juntura, E. The effect of excess body mass on the risk of carpal tunnel syndrome: A meta-analysis of 58 studies. *Obes. Rev.* **16**, 1094–1104 (2015).
 246. Fischer, R. A. Statistical methods in genetics. *Int. J. Epidemiol.* **39**, 329–335 (2010).
 247. Smith, G. D. *et al.* Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med.* **4**, e352 (2007).
 248. Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. *BMJ* **362**, k601 (2018).
 249. Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* **32**, 377–389 (2017).
 250. Nosedá R & Burstein R. Migraine pathophysiology: anatomy of the trigeminovascular pathway and associated neurological symptoms, CSD, sensitization and modulation of pain. *Pain* **154**, Suppl 1: doi:10.1016/j.pain.2013.07.021 (2013).
 251. Bergman-Bock, S. Associations Between Migraine and the Most Common Psychiatric Co-Morbidities. *Headache* **58**, 346–353 (2018).

252. Guyuron, B. Is Migraine Surgery Ready for Prime Time? The Surgical Team's View. *Headache* **55**, 1464–1473 (2015).
253. Guyuron, B., Varghai, A., Michelow, B. J., Thomas, T. & Davis, J. Corrugator supercilii muscle resection and migraine headaches. *Plast. Reconstr. Surg.* **106**, 429–434 (2000).
254. Gfrerer, L., Raposio, E., Ortiz, R. & Austen, W. G. Surgical Treatment of Migraine Headache. *Plast. Reconstr. Surg.* **142**, 1036–1045 (2018).
255. Janis, J. E. *et al.* A review of current evidence in the surgical treatment of migraine headaches. *Plast. Reconstr. Surg.* **134**, 131S–41S (2014).
256. Guyuron, B., Kriegler, J. S., Davis, J. & Amini, S. B. Comprehensive surgical treatment of migraine headaches. *Plast. Reconstr. Surg.* **115**, 1–9 (2005).
257. Guyuron, B., Kriegler, J. S., Davis, J. & Amini, S. B. Five-year outcome of surgical treatment of migraine headaches. *Plast. Reconstr. Surg.* **127**, 603–8 (2011).
258. Guyuron, B. *et al.* A placebo-controlled surgical trial of the treatment of migraine headaches. *Plast. Reconstr. Surg.* **124**, 461–468 (2009).
259. Gaul, C., Sandor, P. S. & Diener, H.-C. Questions on Surgical Treatment of Migraine. *Plast. Reconstr. Surg.* **126**, 669 (2010).
260. McGeeney, B. E. Migraine trigger site surgery is all placebo. *Headache* **55**, 779–81 (2015).
261. Mathew, P. G. A critical evaluation of migraine trigger site deactivation surgery. *Headache* **54**, 142–152 (2014).
262. Law, H.-Z., Amirlak, B., Cheng, J. & Sammer, D. M. An Association between Carpal Tunnel Syndrome and Migraine Headaches—National Health Interview Survey, 2010. *Plast. Reconstr. Surg. - Glob. Open* **3**, e333 (2015).
263. Ho, D. E., Imai, K., King, G. & Stuart, E. A. MatchIt : Nonparametric Preprocessing for Parametric Causal Inference. *J. Stat. Softw.* **42**, doi: 10.18637/jss.v042.i08 (2011).
264. Biondi, B. Thyroid and obesity: An intriguing relationship. *J. Clin. Endocrinol. Metab.* **95**, 3614–7 (2010).
265. Feng, J. *et al.* Body mass index and risk of rheumatoid arthritis: A meta-analysis of observational studies. *Medicine (Baltimore)*. **95**, e2859 (2016).
266. Martorell-Marugan, J., Toro-Dominguez, D., Alarcon-Riquelme, M. E. & Carmona-Saez, P. MetaGenyo: A web tool for meta-analysis of genetic association studies. *BMC Bioinformatics* **18**, 563 (2017).
267. Yavorska, O. O. & Burgess, S. MendelianRandomization: An R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46**, 1734–1739 (2017).

268. Bowden, J., Smith, G. D. & Burgess, S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–25 (2015).
269. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet. Epidemiol.* **40**, 304–14 (2016).
270. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
271. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–81 (2014).
272. Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709–17 (2016).
273. Day, F. R. *et al.* Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat. Genet.* **49**, 834–841 (2017).
274. Day, F. R. *et al.* Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat. Genet.* **47**, 1294–1303 (2015).
275. Gormley, P. *et al.* Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat. Genet.* **48**, 856–66 (2016).
276. Zhu, X. *et al.* Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am. J. Hum. Genet.* **96**, 21–36 (2015).
277. Freilinger, T. *et al.* Genome-wide association analysis identifies susceptibility loci for migraine without aura. *Nat. Genet.* **44**, 777–82 (2012).
278. Ernster, V. L. Nested Case-Control Studies. *Prev. Med. (Baltim.)* **23**, 587–90 (1994).
279. Dieck, G. S. & Kelsey, J. L. An epidemiologic study of the carpal tunnel syndrome in an adult female population. *Prev Med* **14**, 63–69 (1985).
280. Werner, R. A., Albers, J. W., Franzblau, A. & Armstrong, T. J. The relationship between body mass index and the diagnosis of carpal tunnel syndrome. *Muscle Nerve* **17**, 632–6 (1994).
281. Stallings, S. P., Kasdan, M. L., Soergel, T. M. & Corwin, H. M. A case-control study of obesity as a risk factor for carpal tunnel syndrome in a population of 600 patients presenting for independent medical examination. *J. Hand Surg. Am.* **22**, 211–5 (1997).
282. Kouyoumdjian, J. A., Penha Ananias Morita, M. D. A., Rocha, P. R. F., Miranda, R. C. & Gouveia, G. M. Body mass index and carpal tunnel syndrome. *Arq. Neuropsiquiatr.* **58**,

- 252–6 (2000).
283. Greenland, S., Mansournia, M. A. & Altman, D. G. Sparse data bias: A problem hiding in plain sight. *BMJ* **353**, 1–6 (2016).
 284. Celis-Morales, C. A. *et al.* Associations between diabetes and both cardiovascular disease and all-cause mortality are modified by grip strength: Evidence from UK Biobank, a prospective population-based cohort study. *Diabetes Care* **40**, 1710–1718 (2017).
 285. Weng, S. F., Vaz, L., Qureshi, N. & Kai, J. Prediction of premature all-cause mortality: A prospective general population cohort study comparing machine-learning and standard epidemiological approaches. *PLoS One* **14**, e0214365 (2019).
 286. Rosenbloom, A. L. & Silverstein, J. H. Connective tissue and joint disease in diabetes mellitus. *Endocrinol. Metab. Clin. North Am.* **25**, 473–83 (1996).
 287. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466–1468 (2015).
 288. Khramtsova, E. A., Davis, L. K. & Stranger, B. E. The role of sex in the genomics of human complex traits. *Nat. Rev. Genet.* **20**, 173–190 (2018).
 289. Ge, T., Chen, C. Y., Neale, B. M., Sabuncu, M. R. & Smoller, J. W. Phenome-wide heritability analysis of the UK Biobank. *PLoS Genet.* **13**, e1006711 (2017).
 290. Steinberg, G. R., Kemp, B. E. & Watt, M. J. Adipocyte triglyceride lipase expression in human obesity. *Am. J. Physiol. Metab.* **293**, E958–E64 (2007).
 291. Lindgren, C. M. *et al.* Genome-wide association scan meta-analysis identifies three loci influencing adiposity and fat distribution. *PLoS Genet.* **5**, e1000508 (2009).
 292. Bjelland, E. K., Hofvind, S., Byberg, L. & Eskild, A. The relation of age at menarche with age at natural menopause: a population study of 336 788 women in Norway. *Hum. Reprod.* **33**, 1149–1157 (2018).
 293. Walvoord, E. C. The timing of puberty: Is it changing? Does it matter? *J. Adolesc. Heal.* **47**, 433–439 (2010).
 294. Elks, C. E. *et al.* Age at menarche and type 2 diabetes risk: The EPIC-InterAct study. *Diabetes Care* **36**, 3526–3534 (2013).
 295. Onland-Moret, N. C. *et al.* Age at Menarche in Relation to Adult Height. *Am. J. Epidemiol.* **162**, 623–632 (2005).
 296. Zhu, D. *et al.* Body mass index and age at natural menopause: an international pooled analysis of 11 prospective studies. *Eur. J. Epidemiol.* **33**, 699–710 (2018).
 297. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet. Epidemiol.* **40**, 597–608 (2016).
 298. Woldeamanuel, Y. W. & Cowan, R. P. Migraine affects 1 in 10 people worldwide

- featuring recent rise: A systematic review and meta-analysis of community-based studies involving 6 million participants. *J. Neurol. Sci.* **372**, 307–315 (2017).
299. Yeh, W. Z., Blizzard, L. & Taylor, B. V. What is the actual prevalence of migraine? *Brain Behav.* **8**, 6–11 (2018).
 300. Rodríguez-Reyna, T. S., Martínez-Reyes, C. & Yamamoto-Furusho, J. K. Rheumatic manifestations of inflammatory bowel disease. *World J. Gastroenterol.* **15**, 5517–24 (2009).
 301. Park, S. C. & Jeen, Y. T. Current and emerging biologics for ulcerative colitis. *Gut Liver* **9**, 18–27 (2015).
 302. Wilson, P. M., Fryer, R. H., Fang, Y. & Hatten, M. E. Astn2, A Novel Member of the Astrotactin Gene Family, Regulates the Trafficking of ASTN1 during Glial-Guided Neuronal Migration. *J. Neurosci.* **30**, 8529–40 (2010).
 303. Ni, T., Harlos, K. & Gilbert, R. Structure of astrotactin-2 : a conserved membrane protein involved in neuronal development. *Open Biol.* **6**, 160053 (2016).
 304. Boyer, N., Dallel, R., Artola, A. & Monconduit, L. General trigeminospinal central sensitization and impaired descending pain inhibitory controls contribute to migraine progression. *Pain* **155**, 1196–205 (2014).
 305. Bendtsen, L. *et al.* Guideline on the use of onabotulinumtoxinA in chronic migraine: a consensus statement from the European Headache Federation. *J. Headache Pain* **19**, 91 (2018).
 306. Sprenger, T., Viana, M. & Tassorelli, C. Current Prophylactic Medications for Migraine and Their Potential Mechanisms of Action. *Neurotherapeutics* **15**, 313–323 (2018).
 307. Aoki, K. R. Review of a proposed mechanism for the antinociceptive action of botulinum toxin type A. *Neurotoxicology* **26**, 785–93 (2005).
 308. Polderman, T. J. C. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).
 309. Crick, F. On protein synthesis. in *Symposia of the Society for Experimental Biology, Number XII: The Biological Replication of Macromolecules* 138–163 (1958).
 310. Mitteroecker, P., Cheverud, J. M. & Pavlicev, M. Multivariate analysis of genotype–phenotype association. *Genetics* **202**, 1345–1363 (2016).
 311. Burga, A. & Lehner, B. Beyond genotype to phenotype: Why the phenotype of an individual cannot always be predicted from their genome sequence and the environment that they experience. *FEBS J.* **279**, 3765–3775 (2012).
 312. Oh, J. *et al.* Morphological changes of collagen fibrils in the subsynovial connective tissue in carpal tunnel syndrome. *J. Bone Jt. Surg. Am.* **88**, 824–831 (2006).

313. Toesca, A., Pagnotta, A., Zumbo, A. & Sadun, R. Estrogen and progesterone receptors in carpal tunnel syndrome. *Cell Biol. Int.* **32**, 75–79 (2008).
314. Kim, J. K., Hann, H. J., Kim, M. J. & Kim, J. S. The expression of estrogen receptors in the tenosynovium of postmenopausal women with idiopathic carpal tunnel syndrome. *J. Orthop. Res.* **28**, 1469–1474 (2010).
315. Ozsolak, F. & Milos, P. M. RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* **12**, 87–98 (2011).
316. Wang, C. *et al.* The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat. Biotechnol.* **32**, 926–932 (2014).
317. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
318. Themistocleous, A. C. *et al.* The Pain in Neuropathy Study (PiNS): A cross-sectional observational study determining the somatosensory phenotype of painful and painless diabetic neuropathy. *Pain* **157**, 1132–1145 (2016).
319. Schmid, A. B., Bland, J. D. P., Bhat, M. A. & Bennett, D. L. H. The relationship of nerve fibre pathology to sensory function in entrapment neuropathy. *Brain* **137**, 3186–3199 (2014).
320. Bland, J. D. P. A neurophysiological grading scale for carpal tunnel syndrome. *Muscle and Nerve* **23**, 1280–1283 (2000).
321. Levine, D. *et al.* A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. *J Bone Jt. Surg Am* **75**, 1585–92 (1993).
322. Rolke, R. *et al.* Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): Standardized protocol and reference values. *Pain* **123**, 231–243 (2006).
323. Doppler, K., Werner, C. & Sommer, C. Disruption of nodal architecture in skin biopsies of patients with demyelinating neuropathies. *J. Peripher. Nerv. Syst.* **18**, 168–176 (2013).
324. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
325. Borodina, T., Adjaye, J. & Sultan, M. A strand-specific library preparation protocol for RNA sequencing. *Methods Enzymol.* **500**, 79–98 (2011).
326. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
327. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing

- and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
328. Sorrell, J. M. & Caplan, A. I. Fibroblast heterogeneity: more than skin deep. *J. Cell Sci.* **117**, 667–75 (2004).
 329. de Araujo, R., Lobo, M., Trindade, K., Silva, D. & Pereira, N. Fibroblast Growth Factors: A Controlling Mechanism of Skin Aging. *Skin Pharmacol. Physiol.* 1–8 (2019).
 330. Mondelli, M., Reale, F., Sicurelli, F. & Padua, L. Relationship between the self-administered Boston questionnaire and electrophysiological findings in follow-up of surgically-treated carpal tunnel syndrome. *J. Hand Surg. Br.* **25**, 128–34 (2000).
 331. Padua, L., Padua, R., Aprile, I., Pasqualetti, P. & Tonali, P. Multiperspective follow-up of untreated carpal tunnel syndrome: A multicenter study. *Neurology* **56**, 1459–66 (2001).
 332. Padua, L. *et al.* Natural history of carpal tunnel syndrome according to the neurophysiological classification. *Ital. J. Neurol. Sci.* **19**, 357–361 (2006).
 333. Schmier, J. & Halpern, M. T. Patient recall and recall bias of health state and health status. *Expert Rev. Pharmacoeconomics Outcomes Res.* **4**, 159–163 (2004).
 334. Gürsoy, A. E. *et al.* Relationship between electrodiagnostic severity and neuropathic pain assessed by the LANSS pain scale in carpal tunnel syndrome. *Neuropsychiatr. Dis. Treat.* **9**, 65–71 (2012).
 335. Matschinsky, F. M. Assessing the potential of glucokinase activators in diabetes therapy. *Nat. Rev. Drug Discov.* **8**, 399–416 (2009).
 336. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105–116 (2010).
 337. Beaumont, R. N. *et al.* Genome-wide association study of offspring birth weight in 86 577 women identifies five novel loci and highlights maternal genetic effects that are independent of fetal genetics. *Hum. Mol. Genet.* **27**, 742–756 (2018).
 338. Imamura, M. *et al.* Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes. *Nat. Commun.* **7**, 10531 (2016).
 339. Liu, Z. *et al.* Potential targets of TMEM176A in the growth of glioblastoma cells. *Oncotargets Ther* **11**, 7763–7775 (2018).
 340. Wang, Y., Zhang, Y., Herman, J. G., Linghu, E. & Guo, M. Epigenetic silencing of TMEM176A promotes esophageal squamous cell cancer development. *Oncotarget* **8**, 70035–70048 (2017).
 341. Wang, Y. *et al.* Large scale identification of human hepatocellular carcinoma-associated antigens by autoantibodies. *J. Immunol.* **169**, 1102–9 (2002).
 342. Gottschling, S. *et al.* Establishment and comparative characterization of novel squamous cell non-small cell lung cancer cell lines and their corresponding tumor tissue. *Lung*

- Cancer* **75**, 45–57 (2012).
343. Condamine, T. *et al.* Tmem176B and Tmem176A are associated with the immature state of dendritic cells. *J. Leukoc. Biol.* **88**, 507–15 (2010).
 344. Otsubo, T. *et al.* Identification of novel targets for antiangiogenic therapy by comparing the gene expressions of tumor and normal endothelial cells. *Cancer Sci.* **105**, 560–567 (2014).
 345. Bodmer, D., Schepens, M., Eleveld, M. J., Schoenmakers, E. F. P. M. & Van Kessel, A. G. Disruption of a novel gene, DIRC3, and expression of DIRC3-HSPBAPI fusion transcripts in a case of familial renal cell cancer and t(2;3)(q35;q21). *Genes Chromosom. Cancer* **38**, 107–116 (2003).
 346. Coe, E. *et al.* The MITF-SOX10 regulated long non-coding RNA DIRC3 is a melanoma tumour suppressor. *bioRxiv* (2019). doi:<http://dx.doi.org/10.1101/591065>
 347. Lim, D. S., Roberts, R. & Marian, A. J. Expression profiling of cardiac genes in human hypertrophic cardiomyopathy: Insight into the pathogenesis of phenotypes. *J. Am. Coll. Cardiol.* **38**, 1175–1180 (2001).
 348. Kirschner, K. M. *et al.* Amine Oxidase Copper-containing 1 (AOC1) is a downstream target gene of the Wilms tumor protein, WT1, during kidney development. *J. Biol. Chem.* **289**, 24452–24462 (2014).
 349. Melcangi, R. C. *et al.* Peripheral nerves: A target for the action of neuroactive steroids. *Brain Res. Rev.* **48**, 328–38 (2005).
 350. Schumacher, M. *et al.* Progesterone synthesis and myelin formation in peripheral nerves. *Brain Res. Rev.* **37**, 343–359 (2001).
 351. Baulieu, E. E. Neurosteroids: A novel function of the brain. *Psychoneuroendocrinology* **23**, 963–87 (1998).
 352. Jung-Testas, I., Schumacher, M., Robel, P. & Baulieu, E. E. Demonstration of progesterone receptors in rat Schwann cells. *J. Steroid Biochem. Mol. Biol.* **68**, 77–82 (1996).
 353. Melcangi, R. C. *et al.* Role of neuroactive steroids in the peripheral nervous system. *Front. Endocrinol. (Lausanne)*. **2**, 1–8 (2011).
 354. Koenig, H. L. *et al.* Progesterone synthesis and myelin formation by Schwann cells. *Obstet. Gynecol. Surv.* **50**, 792–793 (1995).
 355. Bahrami, M. H., Shahraeeni, S. & Raeissadat, S. A. Comparison between the effects of progesterone versus corticosteroid local injections in mild and moderate carpal tunnel syndrome: A randomized clinical trial. *BMC Musculoskelet. Disord.* **16**, 1–6 (2015).
 356. Raeissadat, S. A., Shahraeeni, S., Sedighipour, L. & Vahdatpour, B. Randomized

- controlled trial of local progesterone vs corticosteroid injection for carpal tunnel syndrome. *Acta Neurol. Scand.* **136**, 365–371 (2017).
357. Milani, P., Mondelli, M., Ginanneschi, F., Mazzocchio, R. & Rossi, A. Progesterone - new therapy in mild carpal tunnel syndrome? Study design of a randomized clinical trial for local therapy. *J. Brachial Plex. Peripher. Nerve Inj.* **5**, 1–7 (2010).
 358. Kawasaki, T., Ushiyama, T., Inoue, K. & Hukuda, S. Effects of estrogen on interleukin-6 production in rheumatoid fibroblast-like synoviocytes. *Clin. Exp. Rheumatol.* **18**, 743–5 (2000).
 359. Hanewinkel, R. *et al.* Metabolic syndrome is related to polyneuropathy and impaired peripheral nerve function: A prospective population-based cohort study. *J. Neurol. Neurosurg. Psychiatry* **87**, 1336–1342 (2016).
 360. Regenmortel, M. H. V. Van. Reductionism and complexity in molecular biology. *EMBO Rep.* **5**, 1016–1020 (2004).
 361. Giersiepen, K. & Spallek, M. Carpal Tunnel Syndrome as an Occupational Disease. *Dtsch Arztebl Int.* **108**, 238–242 (2011).
 362. Kao, S. Y. Carpal Tunnel Syndrome As an Occupational Disease. *J. Am. Board Fam. Med.* **16**, 533–542 (2009).
 363. K., A., A., R., A., P. & N., E. Cubital tunnel syndrome: Anatomy, clinical presentation, and management. *J. Orthop.* **15**, 832–836 (2018).
 364. Silventoinen, K. *et al.* Genetic and environmental effects on body mass index from infancy to the onset of adulthood: An individual-based pooled analysis of 45 twin cohorts participating in the COllaborative project of Development of Anthropometrical measures in Twins (CODATwins). *Am. J. Clin. Nutr.* **104**, 371–9 (2016).
 365. Feng, R. How much do we know about the heritability of BMI? *Am. J. Clin. Nutr.* **104**, 243–4 (2016).
 366. Loveman, E. *et al.* The clinical effectiveness and cost-effectiveness of long-term weight management schemes for adults: A systematic review. *Health Technol. Assess. (Rockv).* **15**, 1–182 (2011).
 367. Chang, D. *et al.* Accounting for eXentricities: Analysis of the X Chromosome in GWAS reveals X-linked genes implicated in autoimmune diseases. *PLoS One* **9**, 1–31 (2014).
 368. Tukiainen, T. *et al.* Chromosome X-Wide Association Study Identifies Loci for Fasting Insulin and Height and Evidence for Incomplete Dosage Compensation. *PLoS Genet.* **10**, (2014).
 369. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121

- (2018).
370. Marjoram, P., Zubair, A. & Nuzhdin, S. V. Post-GWAS: where next? More samples, more SNPs or more biology? *Heredity (Edinb)*. **112**, 79–88 (2014).
 371. Clark, A. J. *et al.* Co-cultures with stem cell-derived human sensory neurons reveal regulators of peripheral myelination. *Brain* **140**, 898–913 (2017).
 372. Müller, S. A., Scilabra, S. D. & Lichtenthaler, S. F. Proteomic Substrate Identification for Membrane Proteases in the Brain. *Front. Mol. Neurosci.* **9**, 1–16 (2016).
 373. Johnston, K. J. A. *et al.* Genome-wide association study of multisite chronic pain in UK Biobank. *PLOS Genet.* **15**, e1008164 (2019).
 374. Schmid, A. B. *et al.* Genetic components of human pain sensitivity: a protocol for a genome-wide association study of experimental pain in healthy volunteers. *BMJ Open* **9**, e025530 (2019).
 375. Collins, F. S. & McKusick, V. A. Implications of the human genome project for medical science. *JAMA* **285**, 540 (2001).
 376. Naldini, L. Gene therapy returns to centre stage. *Nature* **526**, 351–60 (2015).
 377. Bardakjian, T. & Gonzalez-Alegre, P. Towards precision medicine. *Handb. Clin. Neurol.* **147**, 93–102 (2018).
 378. Hout, C. V. Van *et al.* Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. *bioRxiv* (2019). doi:10.1101/572347
 379. Turnbull, C. *et al.* The 100 000 Genomes Project: Bringing whole genome sequencing to the NHS. *BMJ* **361**, k1687 (2018).
 380. Hill, S. The genomic revolution – its future. (2018). Available at: <https://www.england.nhs.uk/blog/genomic-revolution/>. (Accessed: 17th May 2019)
 381. Iacobucci, G. NHS proposes to stop funding 17 ‘unnecessary’ procedures. *BMJ* **362**, k2903 (2018).
 382. Shewring, D., Hobby, J. & Warwick, D. NHS plan to restrict funding is not based on high quality data. *BMJ* **362**, k3743 (2018).