Genetic and epigenetic studies of multiple endocrine neoplasia type 1 (MEN1) associated tumours

Katherine A. English

Somerville College

University of Oxford

A Thesis submitted for the degree of

Doctor of Philosophy

Michaelmas Term 2024

Acknowledgements

The work for this thesis was completed in Professor Rajesh Thakker's laboratory at The Academic Unit, Oxford Centre for Endocrinology and Metabolism. I have been most fortunate to work with an excellent team of researchers in this laboratory who have helped me during my thesis journey. During my DPhil, I have had the opportunity to collaborate with two excellent teams led by Dr David Carey at Geisinger and Professor Joe Marsh, MRC Human Genetics Unit, Edinburgh. I would like to thank Dr Lukas Gerasimavicius (MRC Human Genetics Unit), for his help with the *MEN1* missense bioinformatic tools and modelling. This work would also not have been possible without the support from my funder, Cancer Research United Kingdom, who provided me with a Clinical Research Training Fellowship to undertake this work. I would also like to thank my DPhil supervisors Professor Rajesh Thakker, Dr Kate Lines and Professor Eric O'Neill, and my family, Andrew and Harry.

Statement of originality

Unless otherwise stated below or throughout my DPhil, I completed the work undertaken and reported in this thesis.

1) Formalin-fixed paraffin-embedded samples of pancreatic neuroendocrine tumours (PNETs) were provided by Dr Andreas Selberherr, University of Vienna.
2) Dr Kate Lines, Dr Omair Shariq, and I prospectively collected PNETs and normal pancreatic samples from the Oxford Research Biobank.
3) Multiple Endocrine Neoplasia type 1 (*MEN1)* missense variants present in gnomAD and the scores for individual bioinformatic tools (ESCOTT, SPDV, Rosetta and CPT) for all *MEN1* variants were provided by Dr Lukas Gerasimavicius (MRC Human Genetics Unit, Edinburgh).
4) Individual anthropometric and phenotypic data for patients with *MEN1* variants present in the Geisinger cohort, were provided by Dr David Carey's team from Geisinger, United States.
5) Dr Kate Lines converted RNA samples gifted from the Mayo Clinic to complementary DNA (cDNA) for quantitative real-time polymerase chain reaction (qRT-PCR) experiments.
6) Dr Lines and I undertook the experiments for the *ex-vivo* $Men1^{L/L}$*/RIP2 Cre*-ER mouse study.

Table of Contents

Acknowledgements ..... *ii*

Statement of originality ..... *iii*

Table of Contents ..... *iv*

Table of Tables ..... *ix*

Table of Figures ..... *xii*

Abbreviations ..... *xvii*

Abstract *1*

1. Introduction *2*

1.1 Genetics of Pancreatic Neuroendocrine Tumours **2**

1.1.1 Islets of Langerhans (endocrine pancreas) 2

1.1.2 Pancreatic neuroendocrine tumour overview 3

1.1.3 Risk factors for PNETs 4

1.2 PNETs arising in hereditary tumour syndromes **5**

1.3 Multiple Endocrine Neoplasia type 1 ..... **7**

1.3.1 Protein truncating and missense *MEN1* variants 12

1.4. Clinically sporadic PNETs **15**

1.5 Epigenetics: **16**

1.5.1 Epigenetics overview 16

1.5.2 DNA methylation and hydroxymethylation 17

1.5.3 CpG sites 20

1.6 Dysregulated methylome in PNETs **21**

1.6.1 Abnormal hypomethylation in PNETs 21

1.6.2 Abnormal hypermethylation in PNETs 22

1.7 DNA hydroxymethylation in cancer **25**

1.8 DNA 5hmC assessment and use in clinical practice **25**

1.8.1 Prognosis: 25

1.8.2 Screening and/or monitoring of known malignancy 26

1.9 DNA methylation and hydroxymethylation analysis **26**

1.10 Current and future medical treatment for PNETs.....29

1.11 5'hydroxymethylcytosine modulators.....30

1.11.1 Ascorbic acid..... 30

1.11.1.1 cAMP stimulation..... 31

1.11.2 Somatostatin analogue..... 31

2. Aims..... 32

3. Materials and methods..... 33

3.1 Bioinformatic tools to determine the pathogenicity of *MEN1* variants.....33

3.1.1 *MEN1* variants identified in the literature and publicly available databases..... 33

3.1.1.1 PubMed search..... 33

3.1.2 *MEN1* variants identified in a separate cohort (Geisinger)..... 34

3.1.3 Classification of published *MEN1* variants..... 34

3.1.4 Bioinformatic tools used to analyse *MEN1* missense mutations..... 35

3.1.5 Comparing the accuracy of different bioinformatic tools to classify *MEN1* variants in the literature 36

3.1.5.1 Statistical analyses used to compare different bioinformatic tools, and to generate cut-off levels for each tool..... 36

3.1.5.2 Using the Geisinger cohort with matched phenotypic data, to validate the generated cut-off levels for each bioinformatic tool..... 37

3.2 Patient samples for epigenetic studies 40

3.2.1 Vienna Samples..... 40

3.2.2 ORB prospective PNET and normal pancreatic sample..... 44

3.2.3 RNA samples from PNETs from the Mayo clinic..... 45

3.2.4 PNET IHC characterisation..... 46

3.2.4.1 5hmC quantification..... 47

3.2.4.2 Immunohistochemistry antibodies..... 47

3.3 DNA extraction 48

3.3.1 DNA quality assessment..... 48

3.4 Polymerase Chain Reaction (PCR).....49

3.5 Bisulfite and oxidative-bisulfite conversion 52

3.6 Comparison of the methylation CpGs in human PNETs to the normal islet cell.....53

3.7 RNA extraction from FFPE tissue 53

3.7.1 RNA Quality Control..... 53

3.7.2 RNA analysis..... 55

3.8 Tissue culture 56
3.8.1 Drug treatment and cell viability assays 57
3.8.2 DNA extraction 57
3.8.3 RNA extraction 58
3.8.3.1 cDNA synthesis 58
3.8.3.2 qRT-PCR analysis 58
3.9 *SSTR* transfection 59
3.10 Immunocytochemistry 59
3.11 Dot Blot 60
3.12 Statistical Analysis 61
3.12.1 Overview 61
3.12.2 Parametric vs non-parametric testing in *MEN1* SNV bioinformatic tools 62
3.13 DNA methylome analysis using R Studio 63
3.13.1 MethylationEPIC array using R Studio 63
3.13.2 MethylationEPIC array data compared to published data 64
3.14 RNA-seq analysis 64
3.14.1 RNA-seq analysis of BON1 cell lines 64
3.15 Mouse studies 65
3.15.1 Ethics and husbandry 65
3.15.2 Mouse strains 65
3.15.3 MEN1 genotyping 66
3.15.3.1 DNA extraction from mice ear clips 66
3.15.4 *Ex vivo* mouse study 66
4. Results: 70
4.1 Chapter 1: To determine the most accurate bioinformatics tool to classify *MEN1* missense variants 70
4.1.1 Introduction 71
4.1.2 Results 74
4.1.2.1 MEN1 variants reported in the literature 74
4.1.2.2 Protein structure for menin and locations of benign and pathogenic mutations 83
4.1.2.3 Comparison of different bioinformatic tools using published MEN1 missense variants 85
4.1.2.4 Construction of new MEN1 models to improve current bioinformatic tool precision 91

4.1.2.5 Validating the accuracy of bioinformatic tools, Models 1 and 2 in a separate patient cohort from Geisinger 98
4.1.2.6 Bioinformatic tools to assess MEN1 missense mutations 119
4.1.2.7 MEN1 missense variant look-up table with combined VEP pathogenicity score (Model 2) 121
4.1.3 Discussion 123

4.2 Chapter 2: Investigating global DNA 5hmC in normal pancreatic tissue and PNETs by immunohistochemistry 129

4.2.1 Introduction 130
4.2.2 Results 132
4.2.2.1 DNA 5hmC in normal pancreatic tissue 132
4.2.2.2 DNA 5hmC in PNETs 135
4.2.2.3 DNA hydroxymethylation in mice with and without menin expression. 143
4.2.3 Discussion 148

4.3 Chapter 3: Investigating individual CpG DNA hydroxymethylation in normal pancreatic islet tissue and PNETs by methylationEPIC array 152

4.3.1 Introduction 153
4.3.2 Results 155
4.3.2.1 DNA extraction and quality assessment 155
4.3.2.2 Ox-BS conversion 156
4.3.2.3 5hmC in normal pancreatic tissue 157
4.3.2.4 5hmC analysis in PNETs 159
4.3.2.5 Normal islet-containing exocrine pancreatic tissue versus PNETs 162
4.3.2.6 RNA-seq analysis 165
4.3.2.7 RNA-seq analysis validation using a separate PNET cohort 167
4.3.2.8 Comparison of normal islet cells to PNETs 169
4.3.2.9 Free Fatty Acid Receptor 1 gene 179
4.3.2.10 Comparison of FFAR1 and FFAR4 expression in insulinomas compared to normal islet-containing exocrine pancreatic tissue 181
4.3.3 Discussion 184

4.4 Chapter 4: Modulating DNA hydroxymethylation in PNETs 190

4.4.1 Introduction 191
4.4.2 Results 192
4.4.2.1 Effect of somatostatin analogues on DNA 5hmC 192
4.4.2.2 Modulating DNA 5hmC in cell lines 194
4.4.2.3 Modulating DNA 5hmC with ascorbic acid 196

4.4.2.4 Ex-vivo mouse study investigating the effect of Ascorbic acid on DNA5hmC and cell viability. 198
4.4.3 Discussion 201
5. Conclusions *204*
6. Future studies *205*
6.1 *MEN1* missense variants **205**
6.2 DNA hydroxymethylation in normal islets and PNETs **206**
6.2.1 DNA hydroxymethylation in normal islet biology 206
6.2.1.1 DNA 5hmC in each islet subtype 206
6.2.1.2 DNA 5hmC and insulin secretion 206
6.2.2 DNA hydroxymethylation in PNETs 207
6.2.2.1 Correlating genetic and epigenetic changes in PNETs 207
6.2.2.2 AA treatment on gene expression in PNET cell lines 207
6.2.2.3 Assessing DNA 5hmC in a PNET mouse model treated with AA 208
6.2.2.4 Assessment of somatostatin analogues as DNA 5hmC modulators 208
7. References *210*
8. Supplementary Appendix *232*
8.1 Supplementary Tables **232**
8.2 R Code **249**

Table of Tables

Table 1: PNETs associated with hereditary cancer syndromes 6

Table 2: Details from online databases reporting *MEN1* variants 33

Table 3: List of Variant Effect Predictors and associated programs used to categorise *MEN1* SNVs 35

Table 4: ICD codes of interest retrieved for each individual in the Geisinger cohort with a *MEN1* variant 37

Table 5: European Neuroendocrine Tumour Society Staging for PNETs 40

Table 6: Clinical details for insulinoma samples from the University of Vienna 41

Table 7: Clinical details for NF-PNET samples from the University of Vienna 42

Table 8: Clinical details for gastrinoma samples from the University of Vienna 43

Table 9: Clinical details for individuals with MEN1 and for rare functional PNETs from the University of Vienna 43

Table 10: Oxford Research Biobank prospectively collected samples 44

Table 11: Mayo clinic PNET RNA samples 45

Table 12: Primary and secondary antibodies for immunohistochemistry 47

Table 13: DNA PCR primers for *RASSF1* and *MEN1* 50

Table 14: Quality control values for each sample included for DNA 5hmC analysis 51

Table 15: RNA quality control prior to sending for RNA-seq 54

Table 16: List of primers used for qRT-PCR 58

Table 17: All variants classified as LPV or PV by ACMG criteria 76

Table 18: All variants classified as VUS by ACMG criteria 79

Table 19: Descriptive analysis for each bioinformatic tool 87

Table 20: Model 2 coefficients 93

Table 21: ROC analysis results for each bioinformatic tool, Model 1 and Model 2 for the optimal cut-off level and at different sensitivities and specificities 94

Table 22: Numbers of patients in the Geisinger cohort with a *MEN1* protein-truncating variant 99

Table 23: ICD codes enriched in patients with *MEN1* protein-truncating variants in the Geisinger cohort 101

Table 24: ICD codes for each patient with a truncating *MEN1* variant 101

Table 25: *MEN1* variants reported in Geisinger cohort..... 105

Table 26: The reported background and ethnicities of the 693 patients with *MEN1* missense variants in the Geisinger cohort. 112

Table 27: ICD codes enriched in protein truncating *MEN1* variants, found in patients with *MEN1* missense variants reported in the literature to cause MEN1 112

Table 28: An example of a subset of information available for each of the 4028 unique *MEN1* missense variants 122

Table 29: ORB samples collected prospectively with IHC DNA 5hmC islet and exocrine tissue H scores 133

Table 30: Raw H scores for each DNA 5hmC for each PNET 138

Table 31: Descriptive statistics for each distribution 139

Table 32: Calculated F statistic for each comparison, the critical value from the F distribution at α=0.05, and assumption of equal variance 139

Table 33: Univariate and multivariate analyses to determine independent variables associated with tumour H Score..... 142

Table 34: Mouse demographics and raw H scores for DNA 5hmC 144

Table 35: Multivariate model of mice treated with and without tamoxifen 146

Table 36: Summary of DNA 5hmC changes in human and mouse PNETs compared to normal pancreatic islets, exocrine tissue and insulinomas..... 147

Table 37: Differentially hydroxymethylated regions in normal pancreatic tissue using 'Bumphunter' package in R..... 164

Table 38: Descriptive statistics for normal pancreatic cell subtypes..... 177

Table 39: Common genes with ≥1 hypermethylated CpG in both insulinomas and NF-PNETs 179

Table 40: Summary of DNA 5mC and 5hmC changes in human PNETs compared to normal pancreatic islets, exocrine tissue and PNETs. 183

Table 41: *Men1*$^{L/L}$*/RIP2 Cre-ER* mice cohort treated with octreotide..... 193

Table 42: DNA 5hmC in *Men1*$^{L/L}$*/RIP2 Cre*-ER mice and QGP1 and BON1 cells..... 200

Table 43: Quality control data after all samples were uploaded into R 232

Table 44: Significantly hydroxymethylated genes in normal pancreatic tissue using R package "GoMeth' 233

Table 45: Significantly hydroxymethylated genes in insulinomas using R package 'GoMeth' 241

Table 46: DESeq2 gene counts comparing insulinomas to NF-PNETs, shown adjusted $p<0.01$ 244

Table 47: Common hypermethylated genes in insulinomas and NF-PNETs..... 246

Table of Figures

Figure 1: Location of the pancreas within the human body with an example of an Islet of Langerhans..... 2

Figure 2: Three most common tumours in patients with MEN1. 7

Figure 3: Protein Data Bank (PDB) structure for menin 8

Figure 4: Panel A) Chromosome 11 schematic with wild-type/normal *MEN1* located at 11q13 on the maternal chromosome (blue box) and a *MEN1* mutation on the paternal chromosome (orange box). Panel B) An example of a somatic event (epigenetic dysregulation or a *MEN1* mutation) at the *MEN1* locus in the maternal copy. Panel C) Loss of heterozygosity of the full maternal chromosome 11 and Panel D) Homologous translocation of the long arm of chromosome 11 between maternal and paternal alleles. 10

Figure 5: Examples of different mRNA sequences causing protein-truncating mutations. ... 12

Figure 6: Panel A) shows the aligned peptide sequences for a portion of *MEN1* and Panel B) shows the phylogenetic tree. 13

Figure 7: An example of a mRNA sequence illustrating the consequences of a missense mutation with a G substituted with a C, leading to an amino acid change from lysine to asparagine. 14

Figure 8: Heterochromatin characterised by repressive histone marks (H3K27me3) and DNA 5mC, and euchromatin characterised by the active histone mark H3K4me3 and unmodified cytosines and 5hmC. 17

Figure 9: Dynamic DNA methylome and its interaction with the citric acid cycle and chromatin-modifying enzymes..... 19

Figure 10: Example of an actively transcribed gene with DNA 5hmC located at the shores of a CGI. 21

Figure 11: Interactions between KRAS, RASSF1A and menin. 24

Figure 12: A cartoon showing bisulfite-only paired with oxidative-bisulfite conversion of DNA. 28

Figure 13: DNA integrity (DIN) as analysed on the Agilent 2100 Tapestation Bioanalyser for a subset of samples. 49

Figure 14: An example of 4 samples split into two parallel processes so that each sample underwent both paired oxidative bisulfite (oxBS) and bisulfite-only (BS-only) processes. 52

Figure 15: Quality control for RNA-seq from RNA collected from FFPE tissue. 56

Figure 16: Example of DNA Sanger sequencing from a section of DNA. 59

Figure 17: Shows 4 examples of *MEN1* genotyping from 4 $Men1^{L/L}$/RIP2 Cre-ER mice..... 66

Figure 18: Pancreatic neuroendocrine tumours (insulinomas) from β cells in one *$Men1^{L/L}$/RIP2 Cre*-ER mouse, suspended in the top layer of Ficol plaque (yellow)..... 68

Figure 19: Shows and example of insulinomas suspended in media from one 8-month old *$Men1^{L/L}$/RIP2 Cre*-ER mouse. 69

Figure 20: Flow diagram of *MEN1* missense variants reported in the literature..... 75

Figure 21: Protein data bank (PDB) structure (3u85) of the menin protein. Mutations classified as LPV by ACMG are highlighted in blue (core) and pink (surface)..... 84

Figure 22: The proportions of the locations (core (Dark blue) and surface (red)) of the *MEN1* missense mutations in both the benign and pathogenic groups..... 85

Figure 23: Correlation matrix reporting the correlation coefficients between each bioinformatic tool (CADD, REVEL, PremPS, Rosetta, CPT, ESCOTT and SPDV) ***p<0.001. 86

Figure 24: Density histograms for each cohort of *MEN1* variants (BV, VUS and LPV). 88

Figure 25: Comparison of each variant effect predictor, comparing the three *MEN1* cohorts: BVs, VUS and LPVs. 89

Figure 26: Receiver operator curves for each bioinformatic tool to compare LPV in the literature to BV in gnomAD..... 90

Figure 27: ROC curve for Model 1 92

Figure 28: ROC curve for Model 2 92

Figure 29: Percentages of types of *MEN1* mutation reported in Geisinger cohort (Panel A) and gnomAD (Panel B). 98

Figure 30: Protein truncating *MEN1* variants plotted against current age. 100

Figure 31: Missense *MEN1* variants plotted against current age. 115

Figure 32: Reports the menin protein (blue) in complex with MLL (pink) (PDB: 3u85). The locations of the *MEN1* missense mutations in the Geisinger cohort with PHPT are shown in red. 116

Figure 33: Flowchart of *MEN1* missense variants in the Geisinger cohort. 118

Figure 34: 98.7% cut-off levels for each bioinformatic tool and Models 1 and 2 for all *MEN1* SNVs and the validation cohort of *MEN1* SNVs in the Geisinger cohort..... 120

Figure 35: Immunohistochemistry for normal islet 11969 from a 71-year-old female who underwent a non-PNET pancreatic operation. 132

Figure 36: An example of DNA 5hmC intensity in a normal islet vs normal exocrine tissue by IHC.. 133

Figure 37: The differences in DNA 5hmC intensity (H scores calculated in QuPath) between normal islets and exocrine tissue. 134

Figure 38: Immunofluorescence showing that DNA 5hmC higher intensity co-localises with insulin staining (red) in the normal pancreatic islet. 134

Figure 39: IHC for a NF-PNET, insulinoma and a normal adjacent islet. 136

Figure 40: 5hmC staining in an insulinoma and NF-PNET with their normal adjacent islets. 137

Figure 41: Quantification of DNA 5hmC in Qupath. Normal islets compared to insulinomas, NF-PNETs and normal exocrine tissue. 140

Figure 42: Quantification of DNA 5hmC in Qupath. Normal islets and normal exocrine tissue compared between insulinomas and NF-PNETs. 141

Figure 43: Control/wild-type and *Men1*$^{L/L}$ *RIP2-CreER* mice DNA 5hmC by IHC 145

Figure 44: DNA 5hmC in islets and exocrine tissue in control (closed black circles) and Tamoxifen inducible CRE (black stars) in mice aged 6 months. 145

Figure 45: DNA quality was assessed using PCR for *MEN1* and *RASSF1*. 155

Figure 46: Panel A) Density plot for oxBS (O; orange) vs BS samples (B; green) for the 7 paired normal islet-containing exocrine pancreatic tissue samples. Panel B) Multidimensional plot for the oxBS (orange) and BS-only (green) paired samples. 157

Figure 47: Panel A shows an SVD plot for all PNET samples with 4 principal components explaining the variation between samples (PC-1 to PC-4). Panel B reports the Scree plot explaining the percentage of the variation seen within the samples (oxBS vs BS) for all principal components. 158

Figure 48: Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways that are enriched with 5hmC in normal pancreatic tissue. 159

Figure 49: Panel A shows an SVD plot for all PNET samples with 7 principal components explaining the variation between samples (PC-1 to PC-7). Panel B reports the Scree plot explaining the percentage of the variation seen within the samples (oxBS vs BS) for all principal components. 160

Figure 50: Panel A shows an SVD plot for all insulinoma samples with 3 principal components explaining the variation between samples (PC-1 to PC-3). Panel B reports the Scree plot explaining the percentage of the variation seen within the samples (oxBS vs BS) for all principal components. 161

Figure 51: Panel A shows an SVD plot for all NF-PNET samples with 4 principal components explaining the variation between samples (PC-1 to PC-4). Panel B reports the Scree plot explaining the percentage of the variation seen within the samples (oxBS vs BS) for all principal components. 162

Figure 52: Density plot and multidimensional plots for normal islet-containing exocrine pancreatic tissue and PNET subtypes. 162

Figure 53: Heat map for the hydroxymethylated regions in normal islet-containing exocrine pancreatic tissue and for hydroxymethylated genes in insulinomas and NF-PNETs. 163

Figure 54: Volcano plot reporting the top 79 differentially expressed genes (FDR $p<0.01$) comparing insulinomas to NF-PNETs. 166

Figure 55: Differential expression of key genes in insulinomas compared to NF-PNETs. ... 167

Figure 56: Cycle threshold values for insulin and STEAP3 comparing Insulinomas and NF-PNETs. 168

Figure 57: Cycle threshold values for GPR119 and PTPRT comparing NF-PNETs and insulinomas. 168

Figure 58: Shows the beta values for each CpG in nine insulinomas for three genes A) *MEN1*, B) *RASSF1* and C) *DAXX*. 171

Figure 59: GSEA analysis for the top 15 KEGG pathways enriched with aberrantly hypomethylated genes in insulinomas compared to normal β cells. 172

Figure 60: Shows the beta values for each CpG in nine NF-PNETs for three genes A) *MEN1*, B) *RASSF1* and C) *DAXX*. 174

Figure 61: GSEA analysis for the top 15 KEGG pathways enriched with aberrantly hypomethylated genes in NF-PNETs compared to normal α cells. 175

Figure 62: Regions of uninterrupted unmethylated DNA (base pairs) in each pancreatic cell subtype. 177

Figure 63: Venn diagram of the unmethylated regions in normal α and β pancreatic cells. . 178

Figure 64: *FFAR1* with the beta values for the insulinomas (A) and NF-PNETs (B). 180

Figure 65: Volcano plot reporting the top 803 differentially expressed genes (FDR $p<0.01$) comparing insulinomas to normal pancreatic exocrine tissue. 181

Figure 66: The top 24 KEGG enriched pathways in insulinomas vs normal pancreatic exocrine tissue. 182

Figure 67: Schematic diagram of the signalling pathways for activation of FFAR1. 186

Figure 68: IHC samples from two 6-month-old Men1 L/L RIP2 Cre mice treated with either vehicle or high-dose octreotide for 4 weeks. 192

Figure 69: 5hmC by IHC comparing vehicle and octreotide insulinomas in *Men1$^{L/L}$/RIP2 Cre-ER* mice compared to adjacent exocrine tissue. 194

Figure 70: RNA expression for menin in PNET subtypes compared to BON1 cells. 195

Figure 71: RNA expression of SSTR relative to GAPDH (housekeeper) in PNET subtypes INS (insulinomas; red squares) and NF (NF-PNETs; black circles), and in BON-1 cell line (purple stars). 195

Figure 72: Day 7 proliferation assays for both QGP1 (Panel A) and BON1 (Panel B) cells. 196

Figure 73: Panel A) An example of DNA 5hmC dot blot in QGP1 cells treated with 50µM, 100µM and 500µM of AA at day 0. Panel B) DNA 5hmC in QGP1 cells treated with one dose of AA taken over a period of 7 days. 197

Figure 74: DNA 5hmC in BON1 cells treated with one dose of 100µM AA taken over a period of 7 days. 197

Figure 75: Panel A) Overall DNA 5hmC taken over days 1-7 in QGP1 and Panel B) BON1 cells treated with one dose of 100µM of AA. 198

Figure 76: Cell viability assay for mouse islets treated with increasing doses of AA at day 3 (Panel A) and day 5 (Panel B). 199

Figure 77: Immunocytochemistry for the SSTR2 flag antibody showing expression of SSTR2 receptor at the cell membrane. 209

Abbreviations

Abbreviation	Full name
5caC	5-carboxycytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
ΔG	Change in Gibbs Free Energy
ΔΔG	Change in the change in Gibbs Free Energy
α	Alpha
aa	Amino Acid
AA	Ascorbic acid
ACMG	American College of Medical Genetics
AIC	Akaike Information Criterion
αKG	Alpha-ketoglutarate
Ala	Alanine
ANOVA	Analysis of Variance
AR	Androgen receptor
Asn	Asparagine
ATCC	American Type Culture Collection
ATRX	Alpha Thalassemia/Mental Retardation Syndrome X-linked
AUC	Area under the curve
β	Beta
BAZ2A/B	Bromodomain Adjacent To Zinc Finger Domain 2A/B
BER	Base Excision Repair
BET	Bromo- and Extra-Terminal Domain
BMI	Body Mass Index
BRCA1/2	BRCA1/2 DNA Repair Associated
BrDU	Bromodeoxyuridine
BRD4	Bromodomain Containing 4
BSA	Bovine Serum Albumin
BV	Benign Variant
C	Cytosine
CADD	Combined Annotation Dependent Depletion
CAMK2G	Calcium/Calmodulin Dependent Protein Kinase II Gamma
CANX	Calnexin
CEA	Carcinoembryonic acid

cDNA	Complementary DNA
cf-DNA	Cell-free DNA
CDC73	Cell Division Cycle 73
CDKN1B	Cyclin-dependent kinase inhibitor 1B
CDKN2C	Cyclin-dependent kinase inhibitor 2C
CHEK2	Checkpoint Kinase 2
CLPS	Colipase
CPA2	Carboxypeptidase A2
CpG	Cytosine Guanine Dinucleotide
CPT	Cross protein transfer
CREB3L2	CAMP Responsive Element Binding Protein 3 Like 2
CTRC	Chymotrypsin C
Cys	Cysteine
δ	Delta
DAXX	Death Domain-Associated Protein
df	Degrees of Freedom
DHA	Dehydroascorbic acid
DIN	DNA integrity number
dH_2O	Distilled water
DhMP	Differentially hydroxymethylated probe
DhMR	Differentially hydroxymethylated region
DMP	Differentially methylated probes
DNA	Deoxyribonucleic Acid
DNMT	DNA Methyltransferase
ε	Epsilon
EDTA	Ethylenediaminetetraacetic acid
ECL	Electrochemiluminescence
EMT	Epithelial to mesenchymal transition
ENETS	European Neuroendocrine Tumour Society
ESCOTT	Epistatic and Structural model of Mutational Effects
FDR	False Discovery Rate
FFAR1/GPR40	Free Fatty Acid Receptor 1/G-protein-coupled receptor 40
FFPE	Formalin-fixed paraffin-embedded
FORKA1	Forkhead box protein A1
FWER	Family-wise error rate
G1	Grade 1
G2	Grade 2
G3	Grade 3

GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GATA6	GATA binding protein 6
GEP-NET	Gastroenteropancreatic NET
GLP-1	Glucagon-like peptide-1
GLUT1	Glucose transporter 1
Gly	Glysine
GnomAD	Genome Aggregation Database
GPCR	G protein coupled receptor
GPR119	G protein-coupled receptor 119
Gulo	Gulonolactone (L-) oxidase
GULOP	Gulonolactone (L-) Oxidase, Pseudogene
H&E	Haematoxylin and Eosin
HDAC	Histone Deacetylase
H3K4me1	H3K4 mono-methylation
H3K4me2	H3K4 di-methylation
H3K4me3	H3K4 tri-methylation
hMeDIP-seq	5hmC DNA immunoprecipitation sequencing
HOXC6	Homeobox Protein Hox-C6
HOXC8	Homeobox Protein Hox-C8
ICD	International Classification of Disease
IGF-1	Insulin-like growth factor 1
IHC	Immunohistochemistry
IMS	Industrial Methylated Spirits
INS	Insulin
IR	Insulin receptor
IRB	Institutional Review Board
JNK	Jun N-terminal kinase
JunD	JunD proto-oncogene
KDM	Lysine Demethylase
KDM5B	Lysine demethylase 5B
KMT2A/B	Lysine Methyltransferase 2A/B
KEGG	Kyoto Encyclopaedia of Genes
KRAS	KRAS Proto-Oncogene
LB	Likely Benign
LBV	Likely Benign Variant
Leu	Leucine
LINE	Long interspersed nuclear element
LIP	Labile Iron Pool

LOH	Loss of heterozygosity
LOVD	Leiden Open Variation Database
LP	Likely Pathogenic
LPV	Likely Pathogenic Variant
Lys	Lysine
MAF	Minor Allele Frequency
MAX	MYC Associated Factor X
MBD	Methyl binding domain
MDM2	MDM2 Proto-Oncogene
MeCP2	Methyl-CpG Binding Protein 2
MEF	Mouse embryonic fibroblasts
MEN1	Multiple Endocrine Neoplasia type 1
MEN1	Menin 1 (previously Multiple Endocrine Neoplasia type 1)
Met	Methionine
MGMT	O-6-Methylguanine-DNA Methyltransferase
MLL1/2	Mixed-lineage leukaemia gene 1/2
MTA	Materials transfer agreement
MTC	Medullary thyroid cancer
mTOR	Mammalian Target of Rapamycin
MUTYH	MutY DNA Glycosylase
MYC	MYC Proto-Oncogene
NA	Not Available
NET	Neuroendocrine Tumour
NF1	Neurofibromatosis type 1
NPV	Negative predictive value
NR	Not Recordered
OCDEM	Oxford Centre for Diabetes Endocrinology and Metabolism
ORB	Oxford Research Biobank
OUH	Oxford University Hospital
ox-BS	Oxidative-Bisulfite
P	Pathogenic
PA	Pituitary Adenoma
PARP	Poly-ADP Ribose Polymerase
PDB	Protein Data Bank
PC	Principal Component
PDAC	Pancreatic Ductal Adenocarcinoma
PEI	Pancreatic exocrine insufficiency
PFA	Paraformaldehyde

Phe	Phenylalanine
PHPT	Primary Hyperparathyroidism
PNEC	Pancreatic Neuroendocrine Carcinoma
PNEN	Pancreatic Neuroendocrine Neoplasm
PNET	Pancreatic Neuroendocrine Tumour
PPV	Positive predictive value
PRMT5	Protein arginine methyltransferase 5
PRRT	Peptide Receptor Radionuclide Therapy
PTHoma	Parathyroid hormone tumour
PTPRT	Protein tyrosine phosphatase receptor type T
PTV	Protein Truncating Variant
PV	Pathogenic Variant
QC	Quality Control
qRT-PCR	Quantitative Real Time Polymerase Chain Reaction
RASSF1A	Ras-Associated Domain Family 1 Isoform A
RET	Rearranged during transfection
REVEL	Rare Exome Variant Ensemble Learner
RIN	RNA integrity number
RNA	Ribonucleic Acid
ROC	Receiver operator curve
RTK	Receptor tyrosine kinase
rpm	Revolutions per minute
SCSB PDB	Research Collaboratory for Structural Bioinformatics Protein Data Bank
Seq	Sequencing
SETDB1/2	SET Domain Bifurcated Histone Lysine Methyltransferase 1/2
SINE	Short interspersed nuclear element
SLC2A	Sodium Carrier Family 2 Member
SLC23A1	Solute Carrier Family 23 Member 1
SMAD3	Small body size mothers against decapentaplegic homolog 3
SNV	Single Nucleotide Variant
SOX2	SRY-Box Transcription Factor 2
SPDV	Spatial Proximity to Disease Variants
SSA	Somatostatin Analogue
SSC	Saline sodium citrate
SSTR2	Somatostatin 2 receptor
STEAP3	STEAP3 metalloreductase
SVCT1	Sodium vitamin C transporter 1
SVCT2	Sodium vitamin C transporter 2

SVD	Single value decomposition
SYP	Synaptophysin
T2DM	Type 2 diabetes
TBE	Tris-Borate EDTA
TBS-T	Tris-Buffered Saline – 0.1% Tween 20 Detergent
TET	Ten-Eleven-Translocase
TGF-β	Transforming growth factor-β
TSC1/2	Tuberous Sclerosis Complex 1/2
TSG	Tumour Suppressor Gene
TSS	Transcriptional start site
υ	Upsilon
UHRF2	ubiquitin like with PHD and ring finger domains 2
UMD-MEN1	Universal Mutation Database for MEN1
UV	Ultraviolet
VEGFA	Vascular Endothelial Growth Factor A
VEGFR2	Soluble-vascular endothelial growth factor receptor 2
VEP	Variant effect predictor
VIPoma	Vasoactive Intestinal Peptide tumour
VHL	Von Hippel Lindau
VUS	Variant of uncertain significance
WHO	World Health Organisation
WWTR1	WW Domain Containing Transcription Regulator 1
YAP1	Yes1 associated transcriptional regulator
YY1	YY1 transcription factor
YY1	Yin and Yang 1
ZES	Zollinger Ellison syndrome

Abstract

Pancreatic neuroendocrine tumours (PNETs) account for ~10% of all pancreatic tumours, and form part of the presentation of multiple endocrine neoplasia type 1 (MEN1). Genetic MEN1 is caused by abnormal function of the tumour suppressor menin, encoded by multiple endocrine neoplasia type 1/menin 1 (*MEN1*). Most *MEN1* mutations reported to cause MEN1 are protein truncating variants with ~20% caused by *MEN1* missense mutations. However, in large population genetic registries (e.g. gnomAD and ClinVar), >75% are *MEN1* missense variants, with the majority reported as variants of uncertain significance (VUS). Therefore, the aims of my thesis were: 1) to compare the accuracy of bioinformatic tools on a selected population to classify *MEN1* missense mutations; 2) to generate a novel *MEN1* missense mutation classifier and validate it against a separate patient population; 3) to assess a specific DNA methylome mark, DNA hydroxymethylation (5hmC) by immunohistochemistry and by methylationEPIC array in normal human pancreatic tissue and PNET subtypes; and 4) to assess the effect of ascorbic acid (AA) on DNA 5hmC in PNET models.

My studies generated a novel more accurate bioinformatic tool to classify *MEN1* missense variants using a combination of current (e.g. REVEL) and novel (e.g. ESCOTT) scoring systems compared to individual bioinformatic tools alone. The novel *MEN1* model also had improved specificity when validated in a separate cohort of ~175,000 patients from Geisinger. DNA 5hmC was then assessed by immunohistochemistry in a collection of 98 formalin-fixed paraffin embedded (FFPE) PNETs and showed that non-functioning (NF)-PNETs had reduced DNA 5hmC when compared to insulinomas ($p<0.01$) and to normal adjacent islets ($p<0.001$). Thus, 5hmC was ~3 fold lower in NF-PNETs compared to normal adjacent islets ($p<0.001$). Multivariate analysis and methylationEPIC array confirmed this finding. DNA 5hmC was then modulated using AA in two PNET cell lines, QGP1 (*in vitro* model for NF-PNET) and BON1 (*in vitro* model for insulinoma) and showed that AA significantly increased DNA 5hmC in QGP1 cells compared to BON1, and in QGP1 cells, treatment with 100μM of AA decreased cell viability at day 7 ($p<0.05$).

In summary, this thesis highlights the generation of a more accurate novel bioinformatic tool to aid in clinical decision-making in patients with *MEN1* missense variants and that in QGP1 cells (a model for NF-PNETs) DNA 5hmC may be restored with ascorbic acid and is associated with a decrease cell viability.

1. Introduction

1.1 Genetics of Pancreatic Neuroendocrine Tumours

1.1.1 Islets of Langerhans (endocrine pancreas)

The pancreas is located behind the stomach and, in humans, weighs ~90 grams (range: 40 to 180 grams) [1] and is divided into four sections: head, neck, body and tail.[2] Most pancreatic tissue (~90%) forms the exocrine compartment, including acinar cells responsible for synthesising, storing, and secreting digestive enzymes.[2,3] The islets of Langerhans (endocrine pancreas; named after Paul Langerhans, who described these cells using light microscopy in 1869)[4,5] comprise approximately 1-4.5% of the pancreas, with the remaining non-exocrine/non-endocrine tissue consisting of blood vessels, connective tissue and lymphatics **(Figure 1)**.[6]

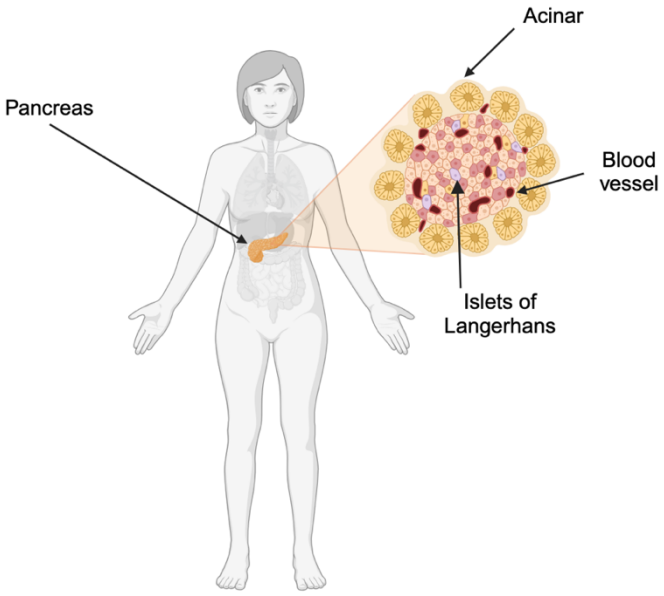


Figure 1: Location of the pancreas within the human body with an example of an Islet of Langerhans. Image generated in BioRender.com.

There are five types of endocrine cells in pancreatic islets, which perform specific functions. Most of the islets are made up of beta (β) cells (54-60%), which are responsible for glucose homeostasis and secrete insulin in response to elevated glucose levels (i.e. hyperglycaemia).

The remaining cell types include alpha (α) cells (~30-35%; secrete glucagon), delta (δ) cells (~8-12%; secrete somatostatin), a small number of upsilon (υ) cells (<5%; secrete pancreatic polypeptide (PP)) and epsilon (ε) cells (<5%; secrete ghrelin; which inhibits insulin secretion from β cells and regulates appetite).[6-10]

1.1.2 Pancreatic neuroendocrine tumour overview

Pancreatic neuroendocrine neoplasms (PNENs) arise from the islets of Langerhans in the pancreas and are the second most common pancreatic cancer, accounting for 1 to 10%.[11-15] PNENs are either well-differentiated tumours (i.e. pancreatic neuroendocrine tumours (PNETs)) or poorly differentiated carcinomas (i.e. pancreatic neuroendocrine carcinomas (PNECs)). PNETs have an incidence of approximately 1-1.5 per 100,000,[13] which appears to be increasing.[11,15-17] According to the World Health Organization (WHO), PNETs are graded from Grades 1-3 (G1-3), with G1 tumours having a proliferative index (Ki67) of <3%, G2 between 3-20% and G3 >20%. Most PNETs are G1 or G2 indolent tumours with a relatively good prognosis compared to PNECs and other malignancies; for example, compared to pancreatic ductal adenocarcinomas (PDAC), the 5-year survival in PNETs is 45-87% compared to <5-10%, for PDACs.[13,16,18,19] However, PNETs do have the ability to metastasise, most commonly to local/regional lymph nodes, the liver or widespread. PNETs arise from one of the endocrine cells in the islets of Langerhans and can be either functional (i.e. hormone secreting) or clinically non-functioning (i.e. non-secreting or PP-secreting, which have no apparent clinical effect). Overall, the most common PNETs are clinically non-functioning (NF-PNETs), and account for 60-80%.[20,21] Of the functional PNETs (20-40%), insulinomas (β cell tumours) are the most common, secreting insulin autonomously and causing symptomatic low blood sugar

levels (hypoglycaemia). Other functional PNETs include glucagonomas (α cell tumours) which secrete excess glucagon and somatostatinomas (δ cells) which secrete excess somatostatin; both may lead to signs of insulin resistance and diabetes, VIPomas which secrete excess VIP causing diarrhoea. Another rare functional PNET includes pancreatic gastrinomas, which are reported to cause 0-15% of Zollinger Ellison Syndrome (ZES; ectopically secreting gastrin causing marked hypersection of gastric acid), with the remaining tumours (60-100%) occurring in the duodenum.[11]

1.1.3 Risk factors for PNETs

PNETs can occur as part of hereditary tumour syndromes, however, most (85-90%) occur as non-hereditary sporadic disease.[22] Risk factors for the development of PNETs have been identified in several retrospective and two prospective case-control studies, and include: recent onset of diabetes (≤12 months), family history of any type of cancer in a first degree relative,[23,24] heavy alcohol consumption, chronic pancreatitis,[23] and smoking.[24] In an Italian prospective cohort: 6.2% (10/162) of patients reported new-onset diabetes within the past 12 months, and overall, 10.4% (17/162) had reported a diagnosis of diabetes (type of diabetes was not recorded);[23] similar findings were also reported in a Han Chinese prospective cohort with 3.9% (15/385) of patients having onset of diabetes within the past 12 months from an overall, 6.2% (24/385) patients with diabetes. Although the authors were unable to definitively ascertain the type of diabetes (e.g. type 1 or type 2), all patients were on oral hypoglycaemic agents and therefore, it was assumed that most had a diagnosis of type 2 diabetes.[24] Although both variables were significant in multivariate analysis, these overall numbers are small and it is unclear if dysglycaemia from diabetes is a cause or consequence of PNETs.[25] Interestingly,

these risk factors are similar to that seen in PDAC patients, 85% of whom were hyperglycaemic on presentation with 50% having T2DM,[26] thereby providing much stronger *a priori* evidence that pathways involved in glucose homeostasis are more critically affected in PDAC compared to PNET; an alternative explanation may be that this is due to a more aggressive disease phenotype of PDACs. Many studies have shown a strong correlation between diabetes and cancer, however it has been unclear if hyperglycaemia itself was a cause, bystander or consequence of PNET and PDAC development.[27,28]

1.2 PNETs arising in hereditary tumour syndromes

Overall, 10-15% of all PNETs arise due to a hereditary tumour syndrome, all of which are due to inactivating mutations in tumour suppressor genes (TSGs), including Multiple Endocrine Neoplasia 1(MEN1) caused by mutations in the Multiple Endocrine Neoplasia type 1 (now termed Menin 1; *MEN1)*), Von Hippel-Lindau disease (Von Hippel-Lindau Tumor Suppressor (*VHL*), Tuberous sclerosis (TSC complex (*TSC*)) and Neurofibromatosis type 1 (Neurofibromin 1 (*NF1*)), and there have been case reports in patients with MEN4 (due to mutations in Cyclin Dependent Kinase Inhibitor 1B (*CDKN1B)*), MEN5 (due to mutations in MYC Associated Factor X (*MAX*); and Hyperparathyroidism jaw tumour syndrome (due to mutations in Cell Division Cycle 73 (*CDC73*)) (**Table 1**). [20,21,29-43]

Table 1: PNETs associated with hereditary cancer syndromes

Syndrome	Gene	Characteristic features	% PNETs
Multiple Endocrine Neoplasia type 1	*MEN1*	PHPT, PA and PNETs (bronchial/thymic/gastric carcinoid tumours, angiofibromas, lipomas, adrenal adenomas).	80%[20]
Von Hippel Lindau disease	*VHL*	Renal cell carcinomas, phaeochromocytomas, PNETs, retinal and central nervous system hemangioblastomas pancreatic, kidney and epididymis cysts.	10-17% >98% NF-PNETs[34-37]
Neurofibromatosis type 1	*NF1*	Café au lait macules (>99%), neurofibromas, skin-fold freckling, iris Lisch nodules and PNETs.	Uncommon [38,39]*
Tuberous sclerosis	*TSC1/2*	Widespread hamartomas (neuro-cutaneous) and PNETs.	0.65-1%[40]
MEN4	*CDKN1B*	PHPT, PA and adrenal, renal and gonad abnormalities.	3 Case reports[29-31]
Hyperparathyroidism jaw tumour syndrome	*CDC73*	PHPT and ossifying fibromas of the jaw.	1 case report[32]
MEN5	*MAX*	Paragangliomas, phaeochromocytomas, PHPT, PA and PNET.	1 case report[33]

*The majority are reported as duodenal somatostatinomas, and PNETs are rare.

Cell Division Cycle 73 (*CDC73*); Cyclin Dependent Kinase Inhibitor 1B (*CDKN1B*); MYC Associated Factor X (*MAX*); Multiple Endocrine Neoplasia type 1 (*MEN1*); Neurofibromatosis type 1 (*NF1*); Pancreatic neuroendocrine tumour (PNETs); Pituitary adenomas (PAs); Primary hyperparathyroidism (PHPT); Tuberous Sclerosis Complex 1/2 (*TSC1/2*); Von Hippel Lindau (*VHL*). Table adapted from Thakker et al., 2012.[44]

PNETs most commonly arise in MEN1, with up to 80% penetrance by the age of 80 years,[20] whereas a PNET phenotype is less common in VHL (10-17%),[34-37] rarely seen in NF1 (<1%)[38-40] and tuberous sclerosis (0.65-1%),[40] and case reports only in MEN4,[29-31] hyperparathyroidism jaw tumour syndrome[32] and MEN5.[33] Therefore, further studies will focus on MEN1.

1.3 Multiple Endocrine Neoplasia type 1

People who inherit germline mutations in *MEN1* are predisposed to developing MEN1, which is characterised by the development of two or more endocrine tumours in one individual and includes primary hyperparathyroidism (PHPT), PNETs and/or anterior pituitary adenomas (PAs; **Figure 2**). MEN1 can also be diagnosed in an individual with only one of these tumours with a first-degree relative with MEN1.[44]

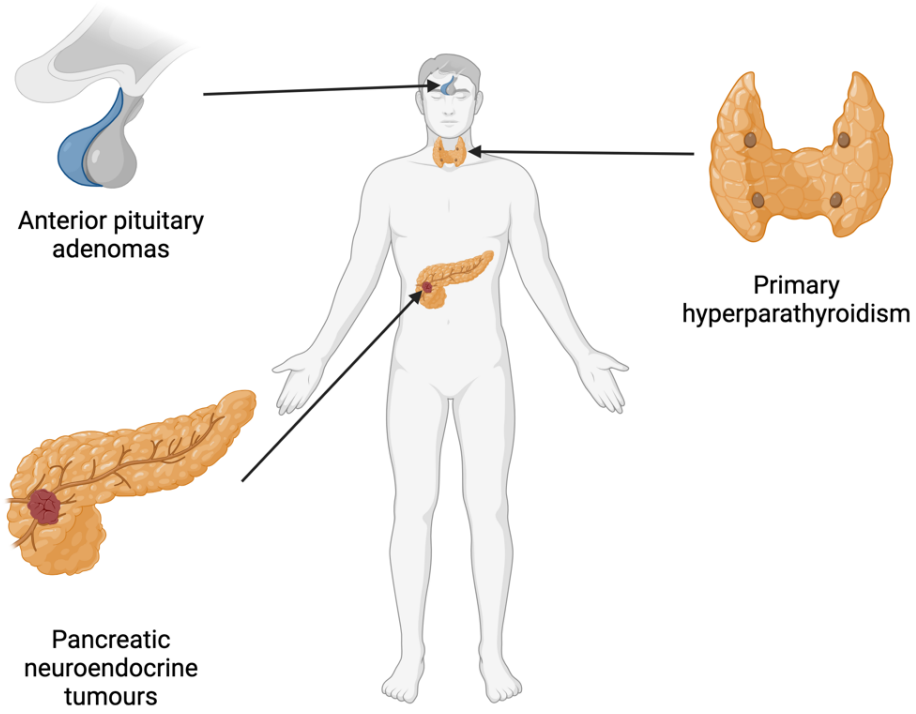


Figure 2: Three most common tumours in patients with MEN1. Pituitary adenomas, primary hyperparathyroidism and pancreatic neuroendocrine tumours. Image generated in BioRender.com.

MEN1 located on chromosome 11q13 was initially identified by positional cloning in 1997,[45,46] encodes for menin, a 610 amino acid (now reported as a 615 amino acid) protein. The crystalline structure of menin was first reported in 2012, with four domains, N-terminal, thumb, palm and fingers, which resembles a curved left hand **(Figure 3)**.[47]

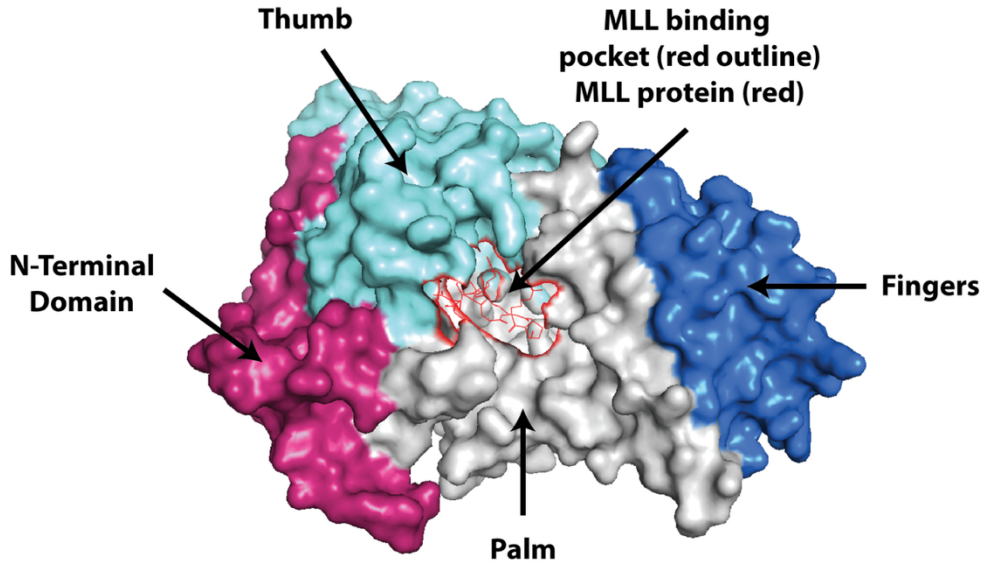


Figure 3: Protein Data Bank (PDB) structure for menin (Magenta N-terminal domain; Cyan – Thumb; Grey- Palm; and blue - fingers) in complex with mixed lineage leukemia (MLL; red),[47] with the image generated in The PyMOL Molecular Graphics System version 2.5.3.

The structure of menin does not show similarities to any other known proteins and has been reported to interact with >40 different binding partners including mixed lineage leukemia (MLL), JunD proto-oncogene (JunD), small body size mothers against decapentaplegic homolog 3 (SMAD3), protein arginine methyltransferase 5 (PRMT5), death domain associated protein (DAXX) and alpha thalassemia/mental retardation syndrome X-Linked (ATRX).[48] One of the most studied interactions is menin binding with MLL, which is encoded by Mixed-lineage leukaemia gene 1/2 (*MLL1/2*) / Lysine Methyltransferase 2A/B (*KMT2A/B*). This histone methyltransferase alters gene expression by increasing H3K4 trimethylation (an active histone mark), thereby increasing DNA accessibility and gene transcription at particular sites e.g.

Homeobox Protein Hox-C6 (*HOXC6*) and Homeobox Protein Hox-C8 (*HOXC8*).[47] JunD, which is a member of the Jun family of transcription factors, also binds to the same menin hydrophobic pocket that MLL binds. Once bound, JunD phosphorylation by Jun N-terminal kinase (JNK) is inhibited, thereby leading to transcriptional repression of JunD targets.[49] Menin is a ubiquitously expressed protein and has cell-specific functions, for example, menin is a tumour suppressor in endocrine cells[50] whereas, menin, when bound to MLL fusion proteins seen in haemopoetic malignancies, forms oncogenic menin/MLL complexes.[51] In endocrine cells, the menin/MLL complex upregulates p27 encoded by *CDKN1B* and p18 encoded by cyclin-dependent kinase inhibitor 2C (*CDKN2C*). In mice, MLL and p27 expression were reported to be ~17 times higher in the pituitary compared to the liver, whereas menin expression was comparable. Additionally, it was reported that in *MLL* knock-out mice, pancreatic islets had significantly decreased levels of both p27 and p18 expression compared to wild-type, whereas no change in these protein levels was seen in the liver.[52] The cell-specific function of MLL on regulation of p27 and p18 may help to explain why patients with MEN1 develop endocrine tumours. Menin has many cellular functions, including regulating gene expression (e.g. apoptotic and cell cycle pathways e.g. p27 and p18) via epigenetic mechanisms (e.g. when bound to MLL), inhibition of proliferative pathways (e.g. decreasing AKT activation and inhibition of JunD mediated transcription), and facilitating anti-proliferative pathways (e.g. SMAD3 and transforming growth factor-β (TGF-β) signaling)[53-57] People with MEN1 develop tumours by Knudson's two-hit hypothesis **(Figure 4)**.[58] For example, if a person has inherited one parental copy of chromosome 11 with a *MEN1* mutation, this would be classified as the first 'hit' in Knudson's two-hit hypothesis. The second 'hit' occurs when the other copy (normal/wild-type) *MEN1* is lost, which can occur due to a somatic mutation, epigenetic

dysregulation (e.g. promoter hypermethylation), loss of heterozygosity due to a full or partial loss of the remaining other normal *MEN1* gene on chromosome 11 (gene deletions) or gene rearrangements e.g. homologous recombination, where during mitosis a daughter cell inherits both copies of the affected *MEN1*.[59,60]

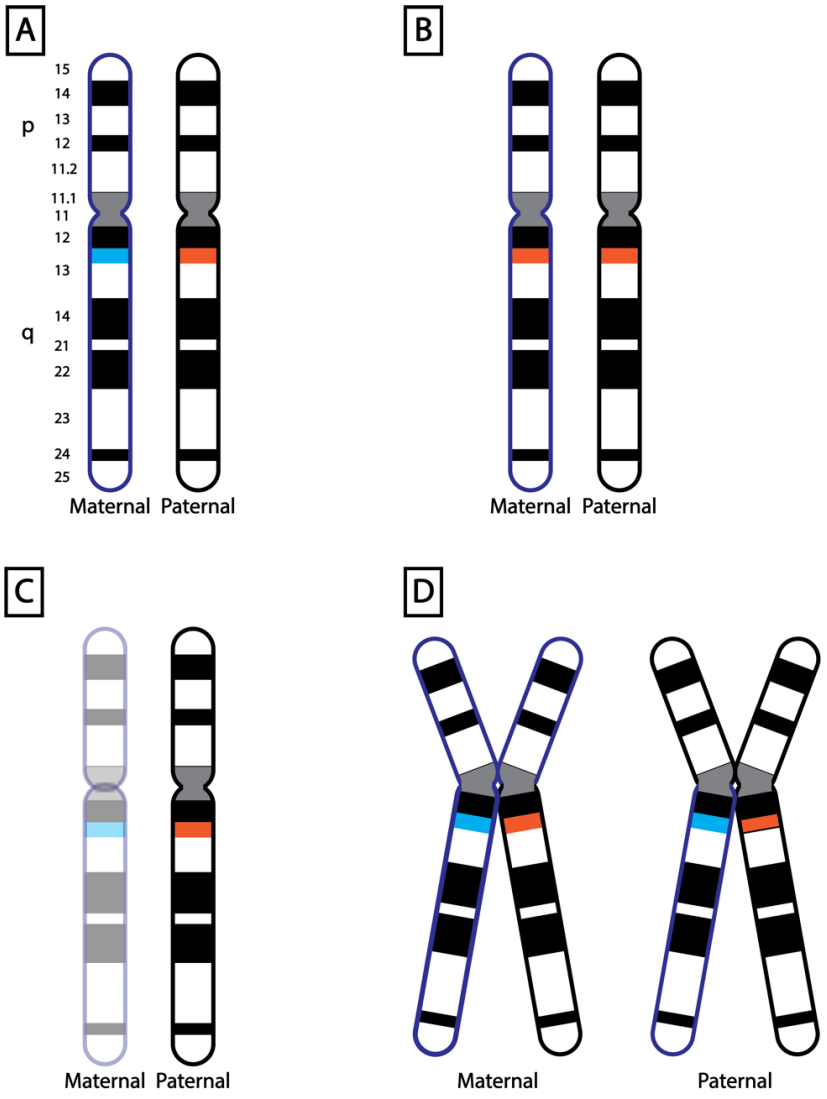


Figure 4: Panel A) Chromosome 11 schematic with wild-type/normal *MEN1* located at 11q13 on the maternal chromosome (blue box) and a *MEN1* mutation on the paternal chromosome (orange box). Panel B) An example of a somatic event (epigenetic dysregulation or a *MEN1* mutation) at the *MEN1* locus in the maternal copy. Panel C) Loss of heterozygosity of the full maternal chromosome 11 and Panel D) Homologous translocation of the long arm of chromosome 11 between maternal and paternal alleles. Image generated in Adobe Illustrator.

In general, most hereditary tumour syndromes are due to biallelic inactivation of TSGs (i.e. both copies are non-functional); however, in some cases, cancer or a phenotype may also develop due to haploinsufficiency, whereby only monoallelic inactivation occurs (germline mutation only, the first 'hit', with no somatic second 'hit') leading to a reduction in the amount of protein. Haploinsufficiency which may cause a phenotype has been reported with germline mutations in Checkpoint Kinase 2 (*CHEK2*; in breast cancer)[61-63] and *MEN1*. Germline *CHEK2* mutations are the second most commonly mutated gene found in hereditary breast cancer after BRCA1/2 DNA Repair Associated (*BRCA1/2*) and it has been reported that some *CHEK2* variants may drive breast cancer tumourigenesis through haploinsufficiency.[62] In heterozygous *MEN1*-knock out mice, endocrine cell proliferation was reported, at double the proliferation rate in affected mouse islets compared to wild-type.[64] In humans, one study reported that in 4 patients with MEN1, islet cell hyperplasia developed without evidence of *MEN1* loss of heterozygosity (LOH), although menin expression by immunohistochemistry was not performed.[65] Whether disease occurs due to *MEN1* haploinsufficiency is particularly important, as with increasing gene sequencing, more *MEN1* mutations have been identified, and at present, most are reported as variants of uncertain significance (VUS). MEN1 due to a *MEN1* mutation are most commonly caused by protein-truncating (PTV) variants (~80%); with ~20% caused by missense mutations.[66] Determining whether *MEN1* mutations are pathogenic is important as current screening guidelines recommend regular blood tests and imaging for tumour surveillance for patients (and their families) with pathogenic *MEN1* mutations.[67] Current screening guidelines in adults recommend both biochemical and radiological screening for tumour surveillance and include: yearly plasma or serum calcium (PHPT), fasting serum gastrin (gastrinoma), prolactin and IGF-1 (pituitary adenoma) levels and radiological screening with abdominal MRI every 2-

3 years (PNET and adrenal adenoma), pituitary MRI every 3-5 years (pituitary adenoma) and CT chest every 3-5 years (thoracic and thymic carcinoid). In asymptomatic children (<19 years of age), imaging and biochemical screening intervals are similar however, the decision of the appropriate time to commence screening is dependent on both the clinician, child and parents.[67]

1.3.1 Protein truncating and missense *MEN1* variants

The two most common types of genetic mutations causing MEN1 are PTVs and missense variants. PTVs are caused by introducing a premature stop codon in the gene sequence. Usually, this results in a shortened protein and depending on the location of the premature stop codon, may affect its function (e.g., loss of function (LOF)). Premature stop codons can occur due to a single base change (e.g. UAU (tyrosine) to UAA (stop)) or an insertion or deletion of base pairs, termed a frameshift mutation **(Figure 5)**.

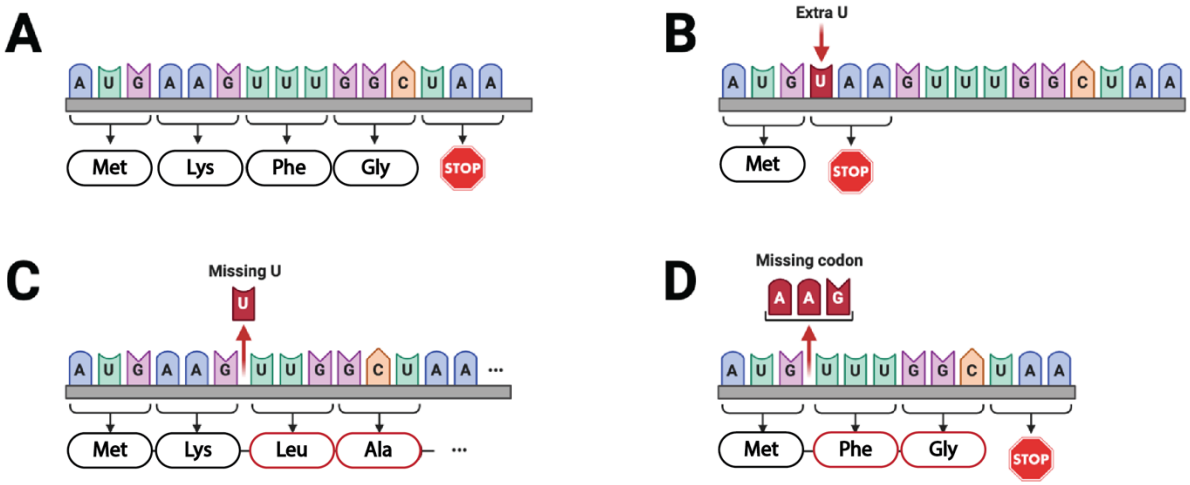


Figure 5: Examples of different mRNA sequences causing protein-truncating mutations. Panel A) represents the normal mRNA sequence; Panel B) 1 base pair insertion of U, results in an immediate stop codon; Panel C) deletion of U, leads to a frameshift in the mRNA sequence with different translated amino acids; and Panel D) in-frame deletion (lysine is removed) resulting in a shortened protein. Met: methionine; Lys: lysine; Phe: phenylalanine; Gly: glysine; Leu: leucine; Ala: alanine. Image generated and modified in BioRender.com and Adobe Illustrator.

In-frame, non-synonymous missense mutations are due to a single nucleotide base pair change, which results in a different translated amino acid at that position. Each protein tolerates missense mutations differently, and generally, amino acids that are evolutionary conserved across species tend to tolerate missense mutations poorly at these locations. Comparing *MEN1* across species (Drosophila melanogaster (Fruit fly), Anas zonorhyncha (Eastern spot-billed duck), Xenopus tropicalis (Western clawed frog; Silurana tropicalis), Pseudonaia textilis (Eastern brown snake), Vombatus ursinus (Common wombat), Mus musculus (Mouse), Delphinapterus leucas (Beluga whale), Homo sapiens (Human) and Gorilla gorilla gorilla (Western lowland gorilla) shows that the aligned DNA sequence is highly conserved **(Figure 6)**.

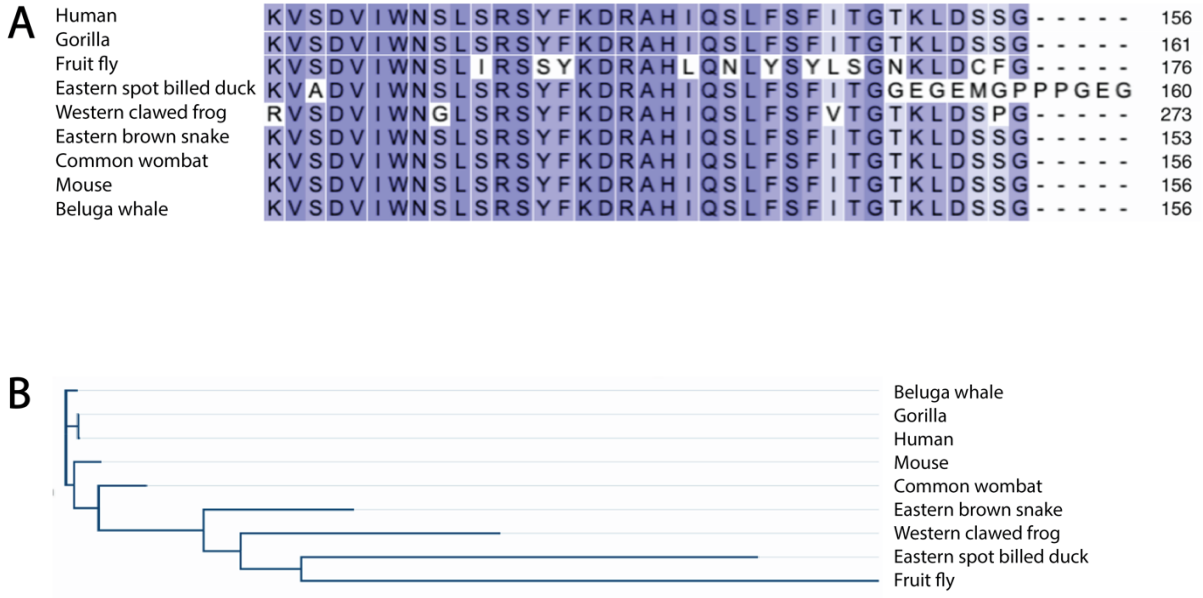


Figure 6: Panel A) shows the aligned peptide sequences for a portion of *MEN1* and Panel B) shows the phylogenetic tree. The image was adapted from the Basic Local Alignment Search Tool (BLAST) of the National Center for Biotechnology Information (NCBI) in Adobe Illustrator.

The American College of Medical Genetics (ACMG) guidelines classify missense mutations as either pathogenic (P), likely pathogenic (LP), uncertain significance (VUS), likely benign (LB) and benign (B). An example of a missense mutation where a cytosine (C) has been substituted for a guanine (G), which results in lysine (Lys; AAG) now mutated to asparagine (Asn; AAC; **Figure 7**).

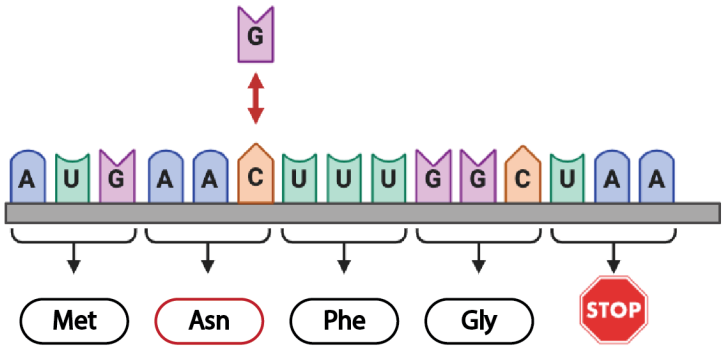


Figure 7: An example of a mRNA sequence illustrating the consequences of a missense mutation with a G substituted with a C, leading to an amino acid change from lysine to asparagine. Met: methionine; Asn: asparagine; Phe: phenylalanine; Gly: glysine. Image generated and modified in BioRender.com and Adobe Illustrator.

Less commonly, hereditary tumour syndromes may be due to activating mutations (as opposed to inactivation of TSGs). For example, activating mutations in the proto-oncogene rearranged during transfection (*RET)* which encodes for the tyrosine kinase receptor, ret, causes multiple endocrine neoplasia type 2 (MEN2; formerly MEN2A; characterised by medullary thyroid cancer (MTC), phaeochromocytoma and PA) or MTC-only, and multiple endocrine neoplasia type 3 (MEN3; formerly MEN2B) characterised by MTC, phaeochromocytomas, mucosal neuroma, marfanoid habitus, medullated corneal nerve fibers and megacolon.[68] These activating *RET* mutations are due to single amino acid substitutions (missense variants) at specific codons, which can cause extracellular changes e.g. ret homodimerisation or

intracellular domain mutations leading to constitutive activation.[69] However, there are no reported cases of hereditary tumour syndromes due to activating proto-oncogenes causing PNETs.

1.4. Clinically sporadic PNETs

Germline mutations of patients with clinically sporadic PNETs (i.e. not associated with a hereditary tumour syndrome) are found in >15%. These have been reported to occur in *MEN1, VHL, TSC, NF1* and in MutY DNA Glycosylase (*MUTYH)*, *BRCA2*, *CDKN1B* and *CHEK2* genes. [22,41,70,71] However, the penetrance for a subset of these gene mutations is low; therefore, not all patients will develop a PNET phenotype. *MEN1* is the most mutated gene in both hereditary and sporadic forms of PNETs, with 35-40% of sporadic PNETs harbouring mutations in *MEN1*.[55,72] Additionally, *ATRX* or *DAXX* are mutated in 11-17% and 22-25% of sporadic PNETs, respectively, and are involved in alternative lengthening of telomeres.[48] ATRX and DAXX also form complexes with menin and are involved" in chromosomal re-modelling.[55,73] Other somatic mutations in genes involved in chromosomal re-modelling and reported in sporadic PNETs include SET Domain Containing 2, Histone Lysine Methyltransferase (*SETD2*), Lysine Methyltransferase 2C (*KMT2C*) and AT-Rich Interaction Domaine 1A (*ARID1A*).[41] Down regulation of the tumour suppressors Phosphatase and Tensin homolog (PTEN) and Tuberous Sclerosis Complex 2 (TSC2; reported in up to 75% of sporadic PNETs),[74] and Plekstrin homology-like domain family A member 3 (*PHLDA3*; reported in ~72% of sporadic PNETs; either by loss of heterozygosity or promoter methylation)[75,76] lead to over activity of the proliferative member pathway. In sporadic insulinomas, it has been reported that up to 30% harbour a somatic mutation at Thr372Arg in the YY1 transcription factor (*YY1*) gene, which encodes Yin Yang 1 (YY1), a ubiquitous protein involved in both mTOR pathway

regulation[77,78] and in the recruitment of epigenetic modulators e.g. histone deacetylases and histone acetyltransferases to either repress (facilitate a closed chromatin state) or activate (facilitate an open chromatin state) target genes.[79,80]

1.5 Epigenetics:

1.5.1 Epigenetics overview

Epigenetics refers to processes (e.g. DNA methylation, acetylation, phosphorylation, ubiquitination and sumoylation) that alter gene activity without changing the underlying DNA code. Epigenetic modifications regulate cell-specific transcription by changing the DNA structure, thereby controlling DNA accessibility to other proteins. In humans, DNA has a total length of >2m and is wrapped around histone proteins to form nucleosomes which then form chromatin, with a diameter of ~10-20μm.[81] Chromatin can either be unwound to form euchromatin or 'open chromatin' for transcription to occur, or may be wound tightly around histone proteins to form heterochromatin or 'closed chromatin', which inhibits transcription **(Figure 8)**. DNA methylation and histone modifications are two common and dynamic epigenetic modifications determining chromatin state. For transcription to occur, DNA must be unwound to form open or euchromatin, characterised by active histone marks eg. H3K4me3 which adds a methyl group to lysine 4 of histone protein H3, forming the active histone mark H3K4me3. MLL1/2 / KMT2A/B are 'writers' as they add histone modifications, whereas 'erasers' do the opposite and remove these marks. Histone 'readers' are transcription factors that determine which marks are present on a particular histone (i.e. repressive or active). There are > 75 different 'writers', 'erasers' and 'readers' involved in methylation maintenance of histone H3.[82,83]

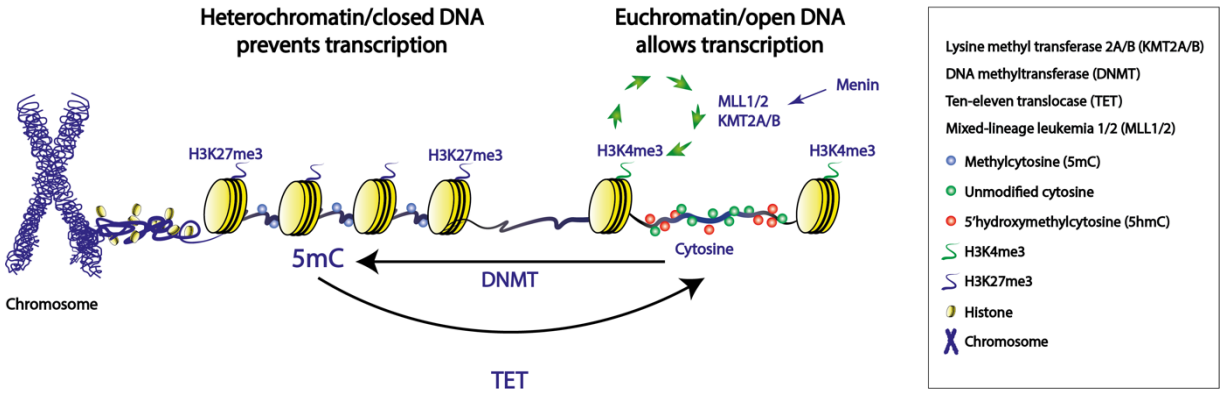


Figure 8: Heterochromatin characterised by repressive histone marks (H3K27me3) and DNA 5mC, and euchromatin characterised by the active histone mark H3K4me3 and unmodified cytosines and 5hmC. Figure adapted from English et al., 2023.[84]

1.5.2 DNA methylation and hydroxymethylation

DNA modifications are distinct but connected to histone modifications, with DNA methylation being the most well-studied DNA mark. DNA methylation can lead to long-term gene silencing through the modified base 5-methylcytosine (5mC), which recruits transcriptional repressors to DNA promoters and/or prevents transcription factors from binding, thereby silencing gene expression.[85,86] Cytosine modifications are recognised by various transcription factors, each with a preference for specific modifications.[87] Several 'reader' proteins, mainly those with a methyl binding domain (MBD) domain, interact with 5mC to mediate transcriptional repression. These ‘reader’ proteins include MBD1-6, Methyl-CpG Binding Protein 2 (MeCP2), SET Domain Bifurcated Histone Lysine Methyltransferase 1/2 (SETDB1/2), and Bromodomain Adjacent To Zinc Finger Domain 2A/B (BAZ2A/B), which either recruit other repressive proteins or by directly interact with histone modifications (e.g., MeCP2 interacts with histone 'writers' and 'erasers' involved in histone deacetylation and methylation), leading to heterochromatin/closed chromatin.[88] In humans, 5-methylcytosine (5mC) is formed when DNA

Methyltransferase (DNMT3A, DNMT3B and DNMT1) enzymes transfer a methyl group from S-adenosyl-methionine (SAM) to the 5th carbon position of cytosine. [86] DNMT3A and DNMT3B are responsible for *de novo* methylation, and DNMT1 for methylation maintenance during DNA replication. This methyl group can then be removed (actively or passively) through consecutive reactions catalysed by the Ten-Eleven Translocation (TET) enzyme family, TET1, TET2, and TET3, to either an intermediate DNA modification (e.g. 5hmC) or to an unmodified cytosine. [85,89,90] The DNA methylome is dynamic and is closely intertwined with the citric acid cycle (TCA; Kreb's cycle), with TET and a subset of Lysine Demethylases (KDM's) dependent on α-ketoglutarate. α-ketoglutarate is part of the TCA cycle and functions as a cofactor for chromatin-modifying enzymes including lysine demethylase 5B (KDM5B) which demethylates H3K4me3, H3K4 di-methylation (H3K4me2) and H3K4 mono-methylation (H3K4me1) histone marks **(Figure 9)** and TET enzymes. TET enzymes require active iron (Fe^{2+}), which can be recycled from Fe^{4+} (inactive) by ascorbic acid (AA; vitamin C). It has been reported that increased cytoplasmic cyclic AMP (cAMP) levels leads to the acidification of endosomes and the subsequent release of labile free iron (Fe^{2+}) into the cytoplasm, thereby increasing DNA 5hmC.[91] When menin expression is lost, H3K4me3 levels decrease, and DNMT enzymes are no longer inhibited, causing increased DNA 5mC, which has been reported in both human and mouse parathyroid adenomas and PNETs.[75,92,93]

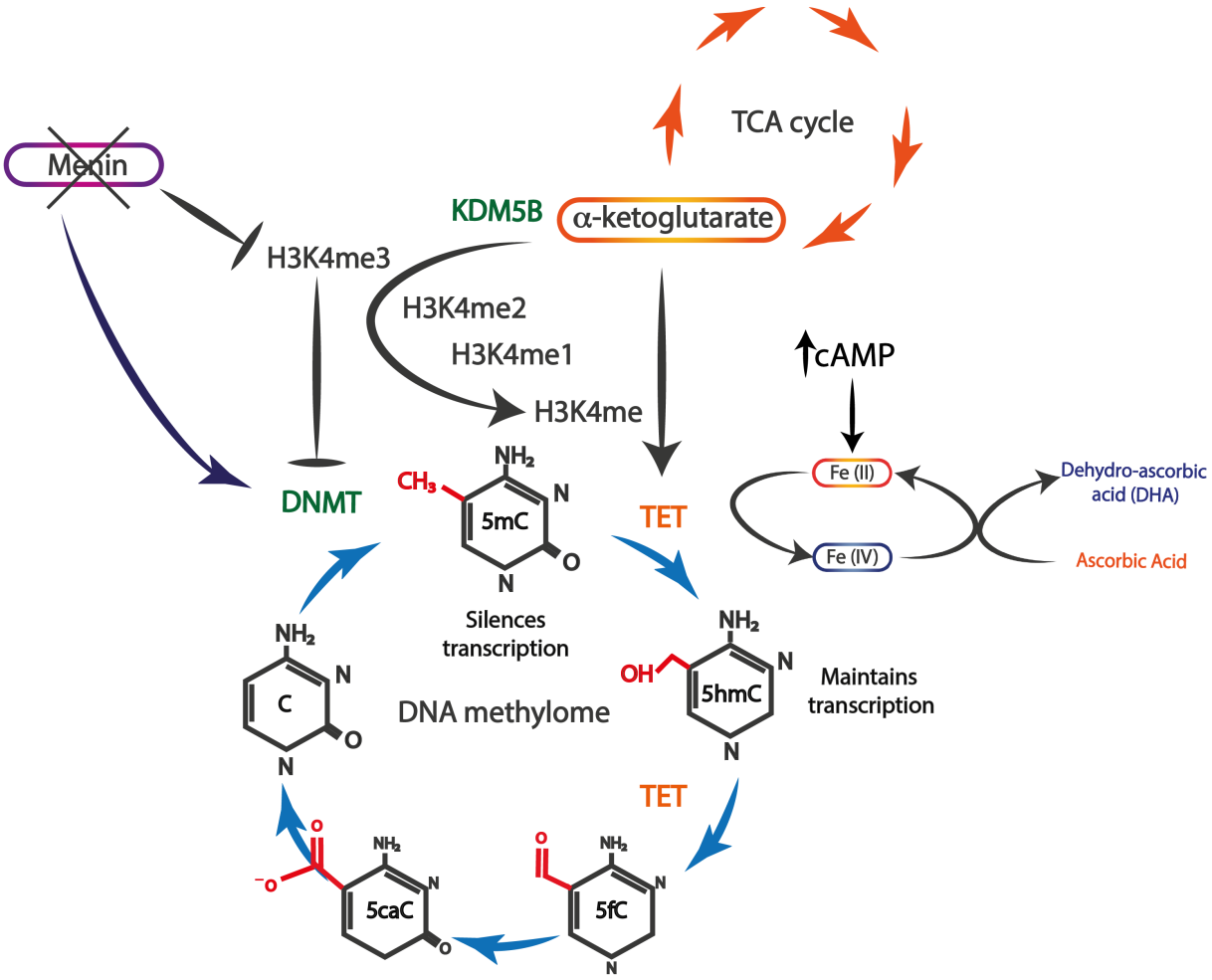


Figure 9: Dynamic DNA methylome and its interaction with the citric acid cycle and chromatin-modifying enzymes. 5'methylcytosine (5mC); 5'hydroxymethylcytosine (5hmC); 5'formylcytosine (5fC); 5'carboxycytosine (5caC); Base Excision Repair (BER); The citric acid cycle (TCA); Fe^{2+} (Fe (II)); Fe^{4+}(Fe (IV)); Dehydro-ascorbic acid (DHA); Cyclic AMP (cAMP) and Lysine demethylase 5B (KDM5B). Figure reproduced and adapted from Kohli et al., 2013[85] and English et al., 2023.[84]

During DNA demethylation by TET enzymes, 5mC is converted to 5hmC. Although less abundant than 5mC, 5hmC is a stable epigenetic mark, promoting gene transcription and protecting against the formation of the repressive mark, 5mC. [85,94] 5mC and 5hmC occur almost

exclusively at sites where cytosine is followed by a guanine (CpG) on cis-DNA (i.e. the same strand of DNA).

1.5.3 CpG sites

Throughout the genome, CpGs are scattered at lower than expected frequency, likely due to 5mC undergoing spontaneous deamination to thiamine.[95] Clusters of CpGs can also be found in sections of DNA 0.5-2kB in length, termed CpG islands (CGIs). There are ~28 million CpG sites that make up ~30,000 CGIs throughout the human genome; these CGIs are usually found in promoters, specifically of housekeeper genes.[95,96] CGIs are flanked on either side by regions of DNA termed shores (within 2kB), shelves (between 2-4kB) and DNA beyond these regions are termed open sea, i.e. >4kB away from the CGI. Genes located within open chromatin, i.e., those with the potential for gene transcription, contain CpG sites within their respective CGI that are hypomethylated and flanked on either side by 5hmC at CGI shores. Reported mechanisms by which 5hmC marks protect against aberrant methylation include inhibition of MBDs,[97] interference with DNMT1 enzymes responsible for DNA 5mC maintenance during cell division[98,99] and by binding of specific enzymes/transcription factors to DNA 5hmC, for example ubiquitin like with PHD and ring finger domains 2 (UHRF2).[99-101] DNA 5hmC enrichment positively correlates with histone markers associated with gene expression e.g. H3K4me3 **(Figure 10)**.[102]

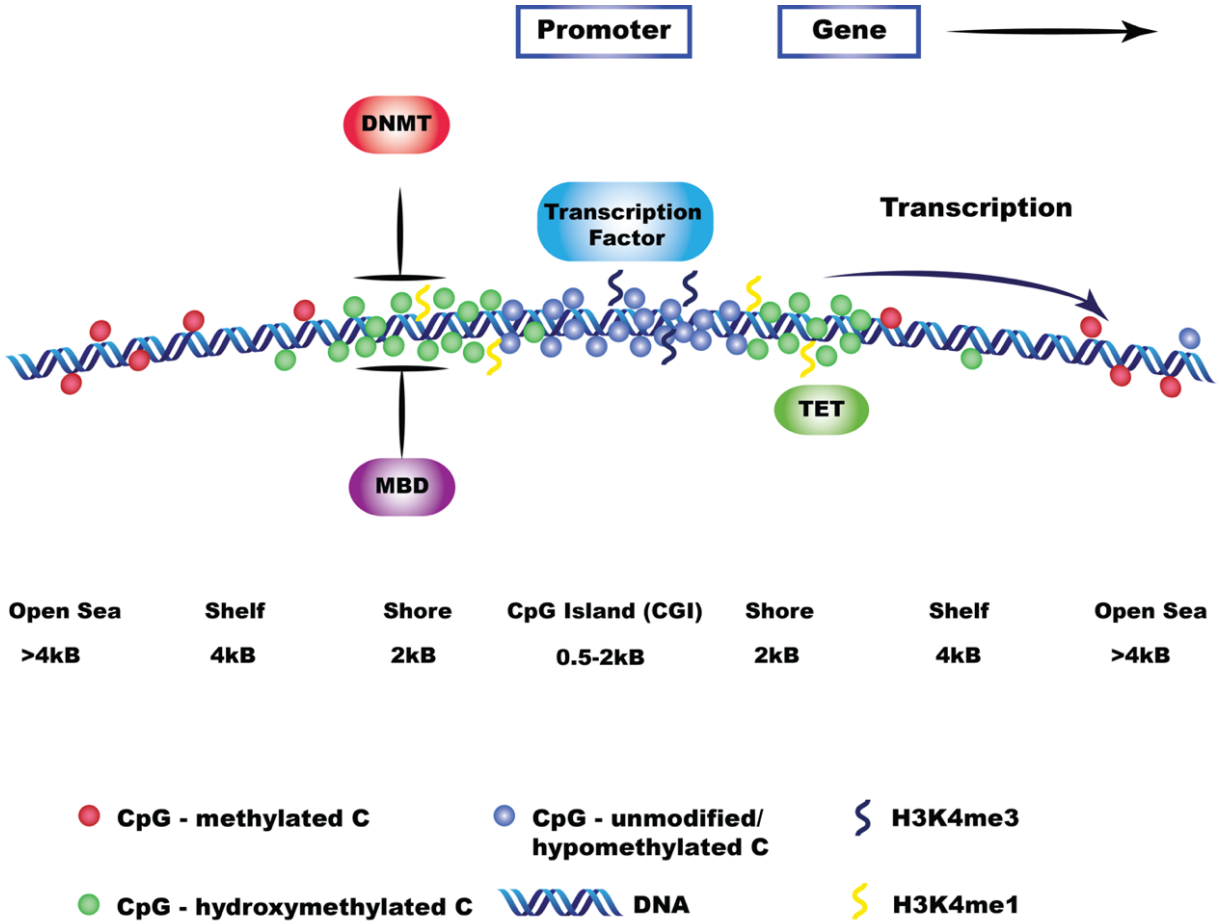


Figure 10: Example of an actively transcribed gene with DNA 5hmC located at the shores of a CGI. DNA methyltransferase (DNMT); Methyl Binding Domain (MBD); Ten eleven translocase (TET); CpG Island (CGI). Figure reproduced and adapted from English et al., 2023.[84]

1.6 Dysregulated methylome in PNETs

1.6.1 Abnormal hypomethylation in PNETs

Carcinogenesis is associated with global DNA hypomethylation (including regions of short and long interspersed nuclear elements (SINEs and LINEs)) and hypermethylation at certain regions e.g. TSGs. SINEs and LINEs comprise ~50% of the human genome and are usually hypermethylated to prevent transcription, maintaining chromosome stability.[103,104] LINE-1 (6,000kb GC-poor region of DNA) is the most abundant LINE and encodes for two proteins,

which can 'copy and paste' itself into other parts of the genome; LINE-1 is termed a retro-transposon, and its methylation status has been used as a marker of global DNA methylation in different cancers (e.g. breast, colon, liver and prostate), and in five studies investigating PNET hypomethylation.[105-111] *Alu* repeats (a type of SINE; ~300bp) are present in ~10% of the human genome,[112] initially thought of as 'junk' DNA, these SINEs have been shown to play an important role in evolution by generating genetic diversity and disease (e.g. cancer) through deleterious insertional mutagenesis.[113] The methylation status of *Alu* repeats have also been used as a marker of global DNA methylation levels and has been reported in one study of 11 PNETs with matched controls.[108] This study reported an inverse correlation between O-6-Methylguanine-DNA Methyltransferase (*MGMT*) promoter methylation and global *Alu* methylation.[108]

1.6.2 Abnormal hypermethylation in PNETs

Hypermethylation at certain regions (e.g. TSGs) occurs during carcinogenesis. The PNET methylome has been investigated in 32 studies,[75,76,107-110,114-138] in a gene-specific manner,[76,107-109,114-133] by array-based technologies[75,122,134-138] and globally.[110] A detailed review of the current studies investigating the PNET methylome has recently been reported.[84] The most commonly reported hypermethylated TSG in PNETs is transcript A of the Ras Association Domain Family Member 1 (*RASSF1)*, termed RASSF1A.[114,117] In pancreatic islets, KRAS Proto-Oncogene (*KRAS*) activates RASSF1A, which controls cellular stability and homeostasis.[139] In pancreatic islets, overexpression of KRAS does not lead to cellular proliferation (as seen in >95% of PDACs), as the RASSF1A pathway likely dominates over the competing proliferative MAPK pathway (latter inhibited by menin; **Figure 11**).[54] Conversely,

menin cannot inhibit PDAC proliferation due to activating *KRAS* mutations. RASSF1A also interacts with DAXX, which subsequently stabilises the tumour suppressor p53, via inhibiting MDM2 Proto-Oncogene (MDM2), thereby enhancing MDM2-DAXX interaction.[140] MDM2, a ubiquitin ligase, promotes p53 ubiquitination and degradation. Thereby, its activation modulates the p53-cell cycle arrest/apoptotic pathway.[140] RASSF1A also acts via the Hippo pathway, ultimately leading to yes-associated protein 1 (YAP1) and WW Domain Containing Transcription Regulator 1 (TAZ/WWTR1) phosphorylation (encoded by Yes1 Associated Transcriptional Regular (*YAP1*) and WW Domain Containing Transcription Regulator 1 (*WWTR1*), respectively), thereby inhibiting proliferation/inducing apoptosis.[139] Free fatty acid receptor 1 (FFAR1; GPR40), activation by Docosahexaenoic acid (DHA) has been reported to activate the Hippo pathway and induce YAP phosphorylation thereby sequestering YAP in the cytoplasm.[141]

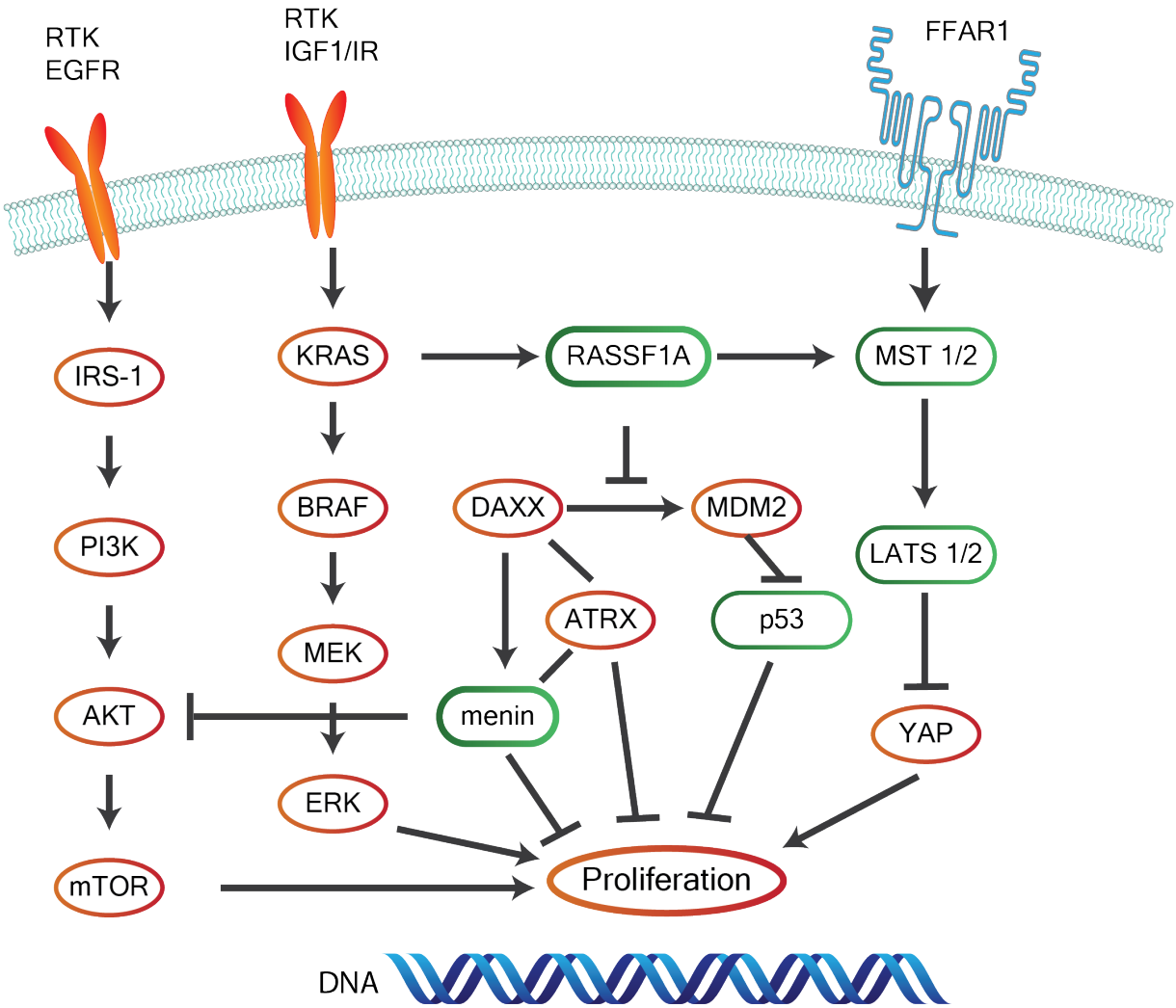


Figure 11: Interactions between KRAS, RASSF1A and menin. Pathways shown: mTOR, MEK/ERK and Hippo pathways. Receptor Tyrosine Kinase (RTK); Epidermal Growth Factor Receptor (EGFR); ATRX Chromatin Remodeler (ATRX); B-Raf proto-oncogene (BRAF); Death domain associated protein (DAXX); Extracellular signal-Regulated Kinase (ERK); Free fatty acid receptor 1 (FFAR1); GTPase Kras (KRAS); Insulin-like growth factor 1 receptor/insulin receptor (IGF-1/IR); Insulin receptor substrate 1 (IRS-1); Large tumor suppressor kinase 1/2 (LATS ½) mechanistic Target of Rapamycin (mTOR); Mitogen-activated protein kinase kinase (MEK); Macrophage stimulating 1/2 (MST1/2); Phosphoinositide 3-kinase (PI3K); Protein kinase B (AKT); Ras association domain family 1 isoform A (RASSF1A); yes-associated protein 1 (YAP); Red circled transcription pathways (proliferative) and Green circled are tumour suppressor pathways and TSGs. Figure reproduced and adapted from Stevenson et al., 2018.[48]

1.7 DNA hydroxymethylation in cancer

DNA 5hmC is associated with gene transcription[142] and is found at promoters/transcriptional start sites, gene bodies and enhancer regions.[143-146] DNA 5hmC protects against 5mC;[147] however, when 5hmC is lost, aberrant DNA methylation can occur, leading to inappropriate gene silencing. DNA 5hmC has been reported in many cancers, including intrahepatic cholangiocarcinoma,[148] endometrial adenocarcinoma,[149] gastric cancer,[150] PNETs[151] and multiple myeloma.[152] Similar to the changes seen in DNA 5mC in malignancy (i.e. aberrant methylation of TSGs), DNA 5hmC has been reported in some malignancies to have aberrant hydroxymethylation of oncogenes, for example, in prostate cancer, there is increased expression of the androgen receptor (*AR*) forkhead box protein A1 (*FORKA1*) which was associated with DNA 5hmC of enhancers upstream from the transcriptional start site (TSS).[153] In PDAC, DNA 5hmC gain was associated with oncogenic pathways, including KRAS, MYC Proto-Oncogene (*MYC*), Vascular Endothelial Growth Factor A (*VEGFA*), and Bromodomain Containing 4 (*BRD4*).[154]

1.8 DNA 5hmC assessment and use in clinical practice

1.8.1 Prognosis:

Loss of DNA 5hmC in cancers compared to normal tissue has been associated with poorer overall survival and has been reported in different malignancies, including intrahepatic cholangiocarcinoma,[148] endometrial adenocarcinoma[149] and gastric cancer.[150] Therefore, DNA 5hmC assessment in tumours may provide prognostic information.

1.8.2 Screening and/or monitoring of known malignancy

Different cell types and malignancies have specific epigenetic signatures, some of which can be measured by cell-free DNA (cf-DNA).[155] Cf-DNA is a broad term used to describe DNA that circulates in the blood and is found in low concentrations in healthy individuals (0-100ng/mL).[156] In cancer patients, the concentration of cf-DNA (termed circulating tumour DNA (ct-DNA)) increases dramatically to >1000ng/mL.[157,158] The PAZONET phase II study reported that ct-DNA polymorphisms in the soluble-vascular endothelial growth factor receptor 2 (*VEGFR2*) and *VEGFR3* confer a potentially favourable response to NETs treated with pazopanib, a receptor tyrosine kinase (RTK) inhibitor.[159] However, given that PNETs have a low and non-specific genetic mutational burden, the clinical utility of using mutations in ct-DNA to monitor and/or diagnose NETs has been reported as not a helpful test in a Delphi Consensus statement.[160] The 'NETest' (measures circulating messenger RNA signatures) has been used as a biomarker for PNETs (and in other NETs), and has shown superior efficacy compared to chromogranin A.[161] Cf-DNA 5hmC has been measured in a few studies and has reported superior efficacy to conventional biomarkers; for example, cf-DNA was superior to carcinoembryonic acid (CEA) levels in colorectal cancer.[162] However, no study has investigated ct-DNA 5hmC signatures in PNETs.

1.9 DNA methylation and hydroxymethylation analysis

The most common method to investigate DNA 5mC is based on bisulfite conversion, which converts unmodified cytosine, 5fC and 5caC to uracil; however, it cannot deaminate 5mC or 5hmC. Therefore, both 5mC and 5hmC are not converted to uracil (and subsequently thiamine) and are indistinguishable when either sequenced or analysed using array-based methods. Only one study assessed the global level of 5hmC by immunohistochemistry in 60 PNETs (subtype

unavailable) and reported that loss of 5hmC, presence of nodal metastasis and lymphovascular invasion independently predicted distant metastasis.[151] No studies have been reported looking at specific CpG 5hmC in PNETs using array-based technologies or gene sequencing.

Given that 5hmC and 5mC have opposing functions, this may explain the reported inconsistencies between protein expression and 'methylated' promoters in PNETs (e.g. *MEN1* and *MGMT*) and in PNECs (SRY-Box Transcription Factor 2 (*SOX2)*).[116,127,138,163] Current techniques for measuring specific CpG DNA 5hmC include oxBS-seq,[164,165] TAB-seq[166] and DNA affinity-based methods, for example, 5hmC DNA immunoprecipitation (hMeDIP-seq).[167] Other methods (i.e. not sequencing) include array-based techniques developed to reduce cost, whilst maintaining coverage. The most recent array-based platform to measure DNA 5mC is Illumina's methylationEPIC array, which interrogates >850,000 CpG sites. These CpG sites cover various genes, including CGI, promoters and enhancers.[168] However, the genomic coverage is not uniform, and some genes have more CpG coverage, which need to be accounted for during analysis.[169] Studies have reported that CpG's close to each other have the same methylation profile.[170] Therefore, the array is intended to give an overview of the whole DNA methylome. Studies that have used these array-based technologies in PNETs have performed bisulfite-only conversion, which cannot distinguish between 5mC and 5hmC. Therefore, CpGs protected from methylation (i.e. that have the 5hmC DNA mark) are classified inappropriately as being methylated/are transcriptionally repressed. A newer technique, whereby the sample to be analysed for 5hmC is split into two, with half undergoing an additional oxidative step to convert 5hmC to 5fC prior to bisulfite conversion results in only the 5mC mark being correctly identified as being methylated. By undertaking these paired experiments, CpGs with the 5hmC mark can be distinguished from those with the 5mC mark **(Figure 12)**.

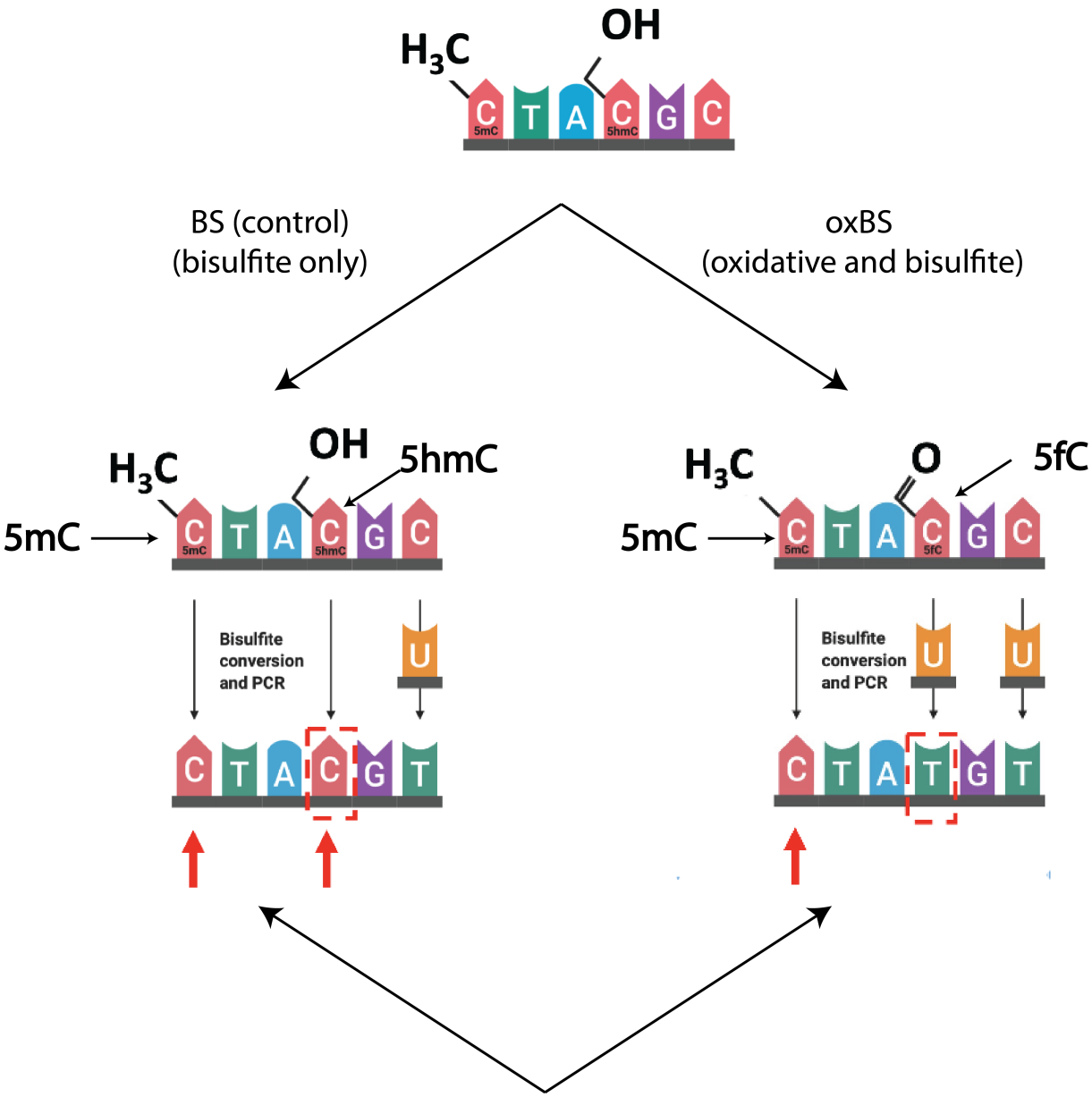


Figure 12: A cartoon showing bisulfite-only paired with oxidative-bisulfite conversion of DNA. Methylcytosine (5mC); Hydroxymethylcytosine (5hmC); Formylcytosine (5fC); Cytosine (C); Guanine (G); Polymerase Chain Reaction (PCR); Thiamine (T) and Uracil (U). Figure adapted from the TrueMethyl oxBS-Seq Module.

1.10 Current and future medical treatment for PNETs

Currently, there are limited treatments for PNETs, and treatment decisions are based on clinical grounds and guidelines.[171] Somatostatin analogues (SSA) possess anti-proliferative and anti-secretory effects on PNETs and improve progression-free survival.[172] Peptide Receptor Radionuclide Therapy (PRRT), which involves radio-isotopes conjugated to a SSA (e.g. targeted radiotherapy), has shown promising results, with improvement in progression-free survival (NETTER-1 trial) in mid-gut NETs.[173] PRRT is not without risks, with reports of radiation nephropathy and progressive renal impairment appearing months after treatment.[174] Additionally, patients with renal impairment are at increased risk of myelosuppression given that most [177Lu]Lu-Dotatate used in PRRT is excreted renally.[175] RTKs (sunitinib)[176] and mTOR inhibitors (everolimus)[177] are effective; however, in-depth molecular characterisation of sporadic PNETs before treatment initiation is not currently feasible, leading to variable treatment responses. Newer single and/or combination therapies have shown promise in pre-clinical studies (e.g. Bromodomain and Extra-Terminal Domain (BET), poly (ADP-ribose) polymerase (PARP) or histone deacetylase (HDAC) inhibitors) however therapy is limited by off-target effects.[178-181] SSAs are the initial medical treatment in patients with PNETs with evidence of progression, however, studies have not addressed the effects of these drugs on the PNET epigenome.

1.11 5'hydroxymethylcytosine modulators

1.11.1 Ascorbic acid

Ascorbic acid (AA; L-ascorbate, or vitamin C) is a water-soluble vitamin found in fresh fruit and vegetables. AA was named from the Latin word for 'scorbutus', meaning 'scurvy', which is the physical manifestation of AA deficiency. Unlike most vertebrates (including mice), humans cannot generate AA from glucose as they lack a functional enzyme, L-gulono-gamma-lactone oxidase, encoded by Gulonolactone (L-) Oxidase, Pseudogene (*GULOP)*, responsible for the final step in AA synthesis. Therefore, AA must be obtained from the diet and is absorbed primarily in the terminal ileum via the sodium vitamin C transporter 1 (SVCT1; encoded by Solute Carrier Family 23 Member 1 (*SLC23A1*)). AA is then transported in the blood (usual maximum serum concentration of 50-85μM) and can be transported into cells by the sodium vitamin C transporter 2 (SVCT2) transporter or via the Sodium Carrier Family 2 Member proteins (*SLC2A*), for example *SLC2A1*, *SLC2A3* and *SLC2A4* which encode for glucose transporter 1,3 and 4 (GLUT1, 3 and 4), respectively). GLUT1 transporters have been reported to be upregulated in different malignancies, including prostate adenocarcinoma, thyroid, squamous cell carcinomas of the head and neck and uterine cervix and PNETs.[182] In humans, AA is an essential cofactor for many enzymes, including iron-dependent/alpha-ketoglutarate (αKG) dioxygenases, for example, JmjC-domain-containing histone demethylases and the TET family of enzymes responsible for DNA demethylation. AA is converted/oxidised to dehydroascorbic acid (DHA), which yields 2 electrons, that reduces Fe^{4+} to Fe^{2+} (active). Fe^{2+} is the active form that can interact with αKG and oxygen to demethylate DNA 5mC to 5hmC. During this reaction, iron is converted back to its 'inactive' Fe^{4+} state and requires AA to cycle

iron back into its active Fe^{2+} state.[183] Therefore, disorders with AA metabolism or sub-physiological intake may lead to a dysregulated epigenome.

1.11.1.1 cAMP stimulation

There has been only one study which has investigated the upstream regulators of DNA 5hmC.[91] Authors reported that compounds which increase cAMP expression (or inhibit its degradation) increase DNA 5hmC by increasing the intracellular labile iron pool (LIP or Fe^{2+}). DNA 5hmC was also assessed after treatment with 50μM of AA, which reported no change in the LIP. Additionally, cAMP did not increase TET enzymes and TET2 and TET3 levels decreased in cell lines HEK-293, mouse embryonic fibroblasts (MEF) and SH-SY5Y. In this study, cell lines were also treated with a G protein coupled receptors (GPCR) agonist, isoproterenol (which binds β-adrenergic receptor) and calcitonin gene-related peptide (CGRP), which binds to the GPCR receptors, both of which act via Gs to stimulate adenylate cyclase and increase cAMP production.[91]

1.11.2 Somatostatin analogue

SSAs are the most common medical treatment for PNETs expressing the somatostatin 2 receptor (SSTR2). SSAs are overexpressed in most well-differentiated PNETs, and treatment with SSAs has been reported to increase progression-free survival.[184,185] There are five SSTRs, and most neuroendocrine tumours express SSTR2 and, to a lesser extent, SSTR5. The two most common long-acting SSAs are lanreotide and octreotide LAR, which predominantly act via SSTR2. SSAs stimulate GPCRs, which decrease cAMP via Gi, and are first-line medical therapy for well-differentiated PNETs; the effects of SSAs on DNA 5hmC has not been investigated.

2. Aims

The aims of my thesis are:

1. A) To determine the prevalence of MEN1 syndrome and *MEN1* mutations using bioinformatic tools and clinical phenotypes from publications and patients in the Geisinger cohort with *MEN1* variants.

   B) Develop a model to predict the pathogenicity of previously unreported *MEN1* missense mutations.

2. A) To investigate global DNA hydroxymethylation in normal pancreatic tissue and PNET subtypes by immunohistochemistry.

   B) To determine predictors for DNA hydroxymethylation.

3. A) To investigate individual CpG DNA hydroxymethylation in normal pancreatic tissue and PNETs by methylationEPIC array

   B) To determine pathways responsible for any differences seen in DNA 5hmC in PNET subtypes.

4. To investigate compounds which can modulate DNA hydroxymethylation

3. Materials and methods

3.1 Bioinformatic tools to determine the pathogenicity of *MEN1* variants

3.1.1 *MEN1* variants identified in the literature and publicly available databases

A manually curated dataset was collated with all reported *MEN1* missense mutations in the literature. First, *MEN1* missense mutations were identified from two published studies examining all reported *MEN1* mutations.[66,186] All *MEN1* missense mutations were then reviewed in four freely available online databases: ClinVar, The Universal Mutation Database – MEN1 (UMD-MEN1), the Leiden Open Variation Database (LOVD) and the genome aggregation database (gnomAD; **Table 2**).

Table 2: Details from online databases reporting *MEN1* variants

Database	Information
ClinVar (USA)	Public archive of DNA mutations with corresponding pathogenicity annotations including references to published papers. https://www.ncbi.nlm.nih.gov/clinvar/
UMD-MEN1 (France)	Public archive of *MEN1* mutations from four diagnostic laboratories in France. Detailed information is not publicly available. Each variant has been given a pathogenicity score. http://umd.be/MEN1/
LOVD (Netherlands)	Database reporting DNA mutations, including ClinVar pathogenicity classification and references. https://www.lovd.nl/
gnomAD (USA)	Database reporting DNA mutations in the population from a variety of large-scale sequencing projects. https://gnomad.broadinstitute.org/

Universal Mutation Database – Multiple Endocrine Neoplasia type 1 (UMD-MEN1); Leiden Open Variation Database (LOVD); Genome Aggregation Database (gnomAD) and United States of America (USA).

3.1.1.1 PubMed search

A PubMed search was undertaken, with Mesh terms: (("Multiple Endocrine Neoplasia Type 1"[Mesh]) AND ("Multiple Endocrine Neoplasia Type 1/classification"[Mesh] OR "Multiple Endocrine Neoplasia Type 1/diagnosis"[Mesh] OR "Multiple Endocrine

Neoplasia Type 1/genetics"[Mesh])) AND ("Mutation, Missense/genetics"[Mesh]) and ("Multiple Endocrine Neoplasia Type 1"[Mesh]) OR "Multiple Endocrine Neoplasia Type 1/genetics"[Mesh] – filter from 2008. Other PubMed search terms included: "MEN1", "Multiple Endocrine Neoplasia Type 1", "Missense mutation", "insulinoma", "PNET".

3.1.2 *MEN1* variants identified in a separate cohort (Geisinger)

A separate cohort of individuals who carry a *MEN1* variant were identified from the Geisinger database. This large database of ~175,000 individuals in the United States of America, includes genome sequencing information (i.e. DNA mutations) which is linked to individual's health record, and contains patient diagnoses which have been entered by health care providers in the form on International Classification of Diseases (ICD) codes. Each individual's health record contains their sex, age, linked family codes, ethnicity, body mass index (BMI), current medications and previous biochemistry (e.g. calcium and parathyroid hormone values).[187,188]

3.1.3 Classification of published *MEN1* variants

The American College for Medical Genetics (ACMG) classification for each published *MEN1* variant reported to cause MEN1 was undertaken, and all variants marked as likely pathogenic variant (LPV) or pathogenic variant (PV) were listed. All *MEN1* single nucleotide variants (SNVs), regardless of their classification in these databases were removed from the LPV/PV dataset if there was no associated clinical information. To generate the benign variant (BV) dataset, *MEN1* missense variants were reviewed in gnomAD, and all variants which were reported to be associated with any neuroendocrine tumour (germline or sporadic) were removed.

3.1.4 Bioinformatic tools used to analyse *MEN1* missense mutations

Bioinformatic scoring systems **(Table 3)**: Combined Annotation Dependent Depletion (CADD),[189] Rare Exome Variant Ensemble Learner (REVEL),[190] and Predicting the impact of missense mutations of protein stability (PremPS)[191] scores were generated and/or annotated for all possible *MEN1* variants.

Table 3: List of Variant Effect Predictors and associated programs used to categorise *MEN1* SNVs

VEP	Information
REVEL[190]	Metapredictor: combines 13 individual in silico tools (MutPred2,[192] fathmm v2.3,[193] VEST 3.0,[194] PolyPhen-2,[195] SIFT,[196] PROVEAN,[197] MutationAssessor,[198] MutationTaster,[199] LRT,[200] GERP++,[201] SiPhy,[202] phyloP,[203] and phaseCons)[204] to give each missense variant a score between 0 (benign) and 1 (pathogenic).[190]
CADD[189]	Machine learning based scoring system, including variables from phyloP,[203] and phaseCons)[204], GERP++,[201], PolyPhen-2,[195] SIFT,[196], to give a raw and scaled CADD score (PHRED).
ESCOTT [205]	Model including evolutionary conservation and structural features (e.g. location of the amino acids, surface vs core).
SPDV	Spatial Proximity to Disease Variants (SPDV) scores, generated and provided by Dr Lukas Gerasimavicius, which utilises known disease variant locations within AlphaFold 2 protein models to derive Euclidean proximity scores for variants of interest. The methodology is currently unpublished.
CPT[206]	Cross protein transfer (CPT) scores, provided by Dr Lukas Gerasimavicius.
Rosetta[207,208]	Generates thermostability scores (developed in Dr David Baker's laboratory, University of Washington; Dr Baker jointly awarded the Nobel Prize in Chemistry in 2024)
PremPS[191]	Generates thermostability scores. The *MEN1* protein structure generated by Alphafold was used to generate a subset of PremPS scores.
Alphafold[209]	Predicts the 3D structure of proteins, which can be input into PremPS software to provide thermostability scores (missing from the 3u85 PDB structure).
RCSB PBD	Repository for storing 3D structures for proteins that can be freely downloaded and viewed (eg. PymOL) and used to generate thermostability scores (PremPS)
PyMOL	Platform for viewing and mutating amino acids from known protein structures (e.g downloaded from RCSB PBD or Alphafold).

REVEL: Rare Exome Variant Ensemble Learner; CADD: Combined Annotation Dependent Depletion; CPT: Cross protein transfer; ESCOTT: Epistatic and Structural model of Mutational Effects; SPDV: Spatial Proximity to Disease Variants; 3D: 3-Dimensional; SCSB PDB: Research Collaboratory for Structural Bioinformatics Protein Bank Database; PyMOL: PyMOL Molecular Graphics System; VEP: Variant effector predictor.

CADD scores were downloaded from (https://cadd.gs.washington.edu/download) on the 13th of July 2022, REVEL scores were downloaded from (https://sites.google.com/site/revelgenomics/about) on the 3rd of May 2021, and PremPS scores[191] were generated using freely available software (https://lilab.jysw.suda.edu.cn/research/PremPS/ on the 24th of October 2023. For thermostability scores, PremPS was used with the protein data bank (PDB) structure identification number: 3u85 (menin in complex with MLL)[47] downloaded from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB-PBD) (https://www.rcsb.org/) and for the remaining amino acids not captured in the 3u85 structure, the menin protein was downloaded from Alphafold (https://alphafold.ebi.ac.uk/entry/E7EN32) on the 27th September 2023.[209] Other bioinformatic scores (Rosetta,[207,208] ESCOTT,[205] CPT and SPDV[unpublished]; **Table 3**) and *MEN1* gnomAD variants were provided by Dr Lukas Gerasimavicius in Professor Joe Marsh's Medical Research Council (MRC) Human Genetics Unit in Edinburgh. All bioinformatic scores were linked to each *MEN1* variant in R Studio (version 4.2.0 "Vigorous Calisthenics"). MEN1 protein structural analysis was viewed using PyMOL Molecular Graphics System, version 2.5.3.

3.1.5 Comparing the accuracy of different bioinformatic tools to classify *MEN1* variants in the literature

3.1.5.1 Statistical analyses used to compare different bioinformatic tools, and to generate cut-off levels for each tool

Receiver Operator Curves (ROC) were performed with ACMG LPV/PV dataset curated from the literature and controls as neutral/benign *MEN1* SNVs reported in gnomAD. ROC analyses determined the cut-off levels at 100%, 95%, 90% and 85% sensitivity and specificity and for the optimal cut-off level (Youden's index) for each of the *MEN1*

predictors (Rosetta, CPT, ESCOTT, REVEL, CADD, SPDV and PremPS). The Youden index is calculated by maximising the sensitivity and specificity:

$$\text{Youden Index: } Maximum\ (Specificity + Sensitivity - 100)$$

3.1.5.2 Using the Geisinger cohort with matched phenotypic data, to validate the generated cut-off levels for each bioinformatic tool

Individuals with a *MEN1* variant in the Geisinger cohort were retrieved and a subset of ICD codes for each patient were reviewed (**Table 4**).

Table 4: ICD codes of interest retrieved for each individual in the Geisinger cohort with a *MEN1* variant

Base Code	ICD Description
C11	Malignant neoplasm of nasopharynx
C75	Malignant neoplasm of other endocrine glands and related structures
C95	Leukemia of unspecified cell type
C26	Malignant neoplasm of other and ill-defined digestive organs
C69	Malignant neoplasm of eye and adnexa
L99	Other disorders of skin and subcutaneous tissue in diseases classified
C23	Malignant neoplasm of gallbladder
C30	Malignant neoplasm of nasal cavity and middle ear
C17	Malignant neoplasm of small intestine
D30	Benign neoplasm of urinary organs
K38	Other diseases of appendix
D40	Neoplasm of uncertain behavior of male genital organs
C07	Malignant neoplasm of parotid gland
C48	Malignant neoplasm of retroperitoneum and peritoneum
C10	Malignant neoplasm of oropharynx
C21	Malignant neoplasm of anus and anal canal
C41	Malignant neoplasm of bone and articular cartilage of other and
C51	Malignant neoplasm of vulva
D15	Benign neoplasm of other and unspecified intrathoracic organs
C82	Follicular lymphoma
C32	Malignant neoplasm of larynx
C83	Non-follicular lymphoma
C19	Malignant neoplasm of rectosigmoid junction
C71	Malignant neoplasm of brain
C56	Malignant neoplasm of ovary
C93	Monocytic leukemia
E32	Diseases of thymus
C7B	Secondary neuroendocrine tumors
D38	Neoplasm of uncertain behavior of middle ear and respiratory and
D07	Carcinoma in situ of other and unspecified genital organs
C68	Malignant neoplasm of other and unspecified urinary organs
C09	Malignant neoplasm of tonsil
C62	Malignant neoplasm of testis
D28	Benign neoplasm of other and unspecified female genital organs
C92	Myeloid leukemia
C22	Malignant neoplasm of liver and intrahepatic bile ducts
N46	Male infertility

D34	Benign neoplasm of thyroid gland
D44	Neoplasm of uncertain behavior of endocrine glands
C06	Malignant neoplasm of other and unspecified parts of mouth
E24	Cushing's syndrome
D39	Neoplasm of uncertain behavior of female genital organs
C20	Malignant neoplasm of rectum
C55	Malignant neoplasm of uterus part unspecified
N53	Other male sexual dysfunction
E20	Hypoparathyroidism
D09	Carcinoma in situ of other and unspecified sites
D11	Benign neoplasm of major salivary glands
A40	Streptococcal sepsis
Q93	Monosomies and deletions from the autosomes not elsewhere classified
D41	Neoplasm of uncertain behavior of urinary organs
C7A	Malignant neuroendocrine tumors
D13	Benign neoplasm of other and ill-defined parts of digestive system
D10	Benign neoplasm of mouth and pharynx
C53	Malignant neoplasm of cervix uteri
C77	Secondary and unspecified malignant neoplasm of lymph nodes
E85	Amyloidosis
D16	Benign neoplasm of bone and articular cartilage
Q99	Other chromosome abnormalities not elsewhere classified
C90	Multiple myeloma and malignant plasma cell neoplasms
C73	Malignant neoplasm of thyroid gland
D33	Benign neoplasm of brain and other parts of central nervous system
D03	Melanoma in situ
D80	Immunodeficiency with predominantly antibody defects
K28	Gastrojejunal ulcer
I74	Arterial embolism and thrombosis
C64	Malignant neoplasm of kidney except renal pelvis
D46	Myelodysplastic syndromes
C80	Malignant neoplasm without specification of site
I43	Cardiomyopathy in diseases classified elsewhere
D06	Carcinoma in situ of cervix uteri
D32	Benign neoplasm of meninges
D31	Benign neoplasm of eye and adnexa
C84	Mature T/NK-cell lymphomas
D27	Benign neoplasm of ovary
C91	Lymphoid leukemia
D61	Other aplastic anemias and other bone marrow failure syndromes
E29	Testicular dysfunction
E13	Other specified diabetes mellitus
E34	Other endocrine disorders
C67	Malignant neoplasm of bladder
C54	Malignant neoplasm of corpus uteri
C85	Other specified and unspecified types of non-Hodgkin lymphoma
E22	Hyperfunction of pituitary gland
K26	Duodenal ulcer
D3A	Benign neuroendocrine tumors
S12	Fracture of cervical vertebra and other parts of neck
R57	Shock not elsewhere classified
D05	Carcinoma in situ of breast
C78	Secondary malignant neoplasm of respiratory and digestive organs
C25	Malignant neoplasm of pancreas
D89	Other disorders involving the immune mechanism not elsewhere classified
L74	Eccrine sweat disorders
R77	Other abnormalities of plasma proteins
D36	Benign neoplasm of other and unspecified sites

D37	Neoplasm of uncertain behavior of oral cavity and digestive organs
D21	Other benign neoplasms of connective and other soft tissue
D24	Benign neoplasm of breast
D47	Other neoplasms of uncertain behavior of lymphoid hematopoietic
C43	Malignant melanoma of skin
D04	Carcinoma in situ of skin
K85	Acute pancreatitis
K25	Gastric ulcer
E07	Other disorders of thyroid
E08	Diabetes mellitus due to underlying condition
E27	Other disorders of adrenal gland
M80	Osteoporosis with current pathological fracture
C61	Malignant neoplasm of prostate
C34	Malignant neoplasm of bronchus and lung
E05	Thyrotoxicosis [hyperthyroidism]
C18	Malignant neoplasm of colon
E31	Polyglandular dysfunction
E06	Thyroiditis
E23	Hypofunction and other disorders of the pituitary gland
C79	Secondary malignant neoplasm of other and unspecified sites
S72	Fracture of femur
M86	Osteomyelitis
R12	Heartburn
E28	Ovarian dysfunction
D35	Benign neoplasm of other and unspecified endocrine glands
N97	Female infertility
B36	Other superficial mycoses
K86	Other diseases of pancreas
R76	Other abnormal immunological findings in serum
S42	Fracture of shoulder and upper arm
D68	Other coagulation defects
I80	Phlebitis and thrombophlebitis
S32	Fracture of lumbar spine and pelvis
R97	Abnormal tumor markers
S92	Fracture of foot and toe except ankle
E10	Type 1 diabetes mellitus
I63	Cerebral infarction
S62	Fracture at wrist and hand level
E16	Other disorders of pancreatic internal secretion
S82	Fracture of lower leg including ankle
I42	Cardiomyopathy
I11	Hypertensive heart disease
E21	Hyperparathyroidism and other disorders of parathyroid gland
S22	Fracture of rib(s) sternum and thoracic spine
D17	Benign lipomatous neoplasm
D72	Other disorders of white blood cells
D12	Benign neoplasm of colon rectum anus and anal canal
S52	Fracture of forearm
C50	Malignant neoplasm of breast
I82	Other venous embolism and thrombosis
N64	Other disorders of breast
N52	Male erectile dysfunction
E83	Disorders of mineral metabolism
N20	Calculus of kidney and ureter
D49	Neoplasms of unspecified behavior
L02	Cutaneous abscess furuncle and carbuncle
R22	Localized swelling mass and lump of skin and subcutaneous tissue
M81	Osteoporosis without current pathological fracture

D48	Neoplasm of uncertain behavior of other and unspecified sites
D23	Other benign neoplasms of skin
E87	Other disorders of fluid electrolyte and acid-base balance
E55	Vitamin D deficiency
E03	Other hypothyroidism
I50	Heart failure
E11	Type 2 diabetes mellitus
M79	Other and unspecified soft tissue disorders not elsewhere classified

3.2 Patient samples for epigenetic studies

Patient samples in this thesis included historic formalin-fixed paraffin-embedded (FFPE) tissue gifted to the University of Oxford by Dr Andreas Selbeherr, University of Vienna and prospectively collected samples through the Oxford Research Biobank (ORB). Separate RNA samples were provided by the Mayo Clinic in Rochester, Minnesota, USA.

3.2.1 Vienna Samples

FFPE gastroenteropancreatic NET (GEP-NET) samples were available from 95 patients (under a material transfer agreement (MTA)), and for each sample, information was available for the age and sex of the individual, follow-up duration, status (i.e. dead or alive) at last follow-up, PNET grade (World Health Organisation (WHO)), Ki67 (histological marker of cellular proliferation), European Neuroendocrine Tumour Society (ENETs) stage (T1-4; **Table 5**) and tumour size. The WHO classification system to grade PNETs uses both tumour differentiation (well vs poorly differentiated) and how proliferative the tumour is, using Ki67 (G1 <3%, G2 3-20% and G3 >20%), with G3 tumours either classified as a G3 PNET (well-differentiated), or G3 PNEC (poorly-differentiated).[210] The ENETs tumour stage is used as a marker of tumour size and spread (**Table 5**).

Table 5: European Neuroendocrine Tumour Society Staging for PNETs

ENETs stage	Pancreatic neuroendocrine tumour characteristics
T1	Tumour <2cm
T2	Tumour 2-4cm
T3	Tumour >4cm, or invading duodenum or common bile duct
T4	Tumour has spread outside the pancreas, and invaded adjacent structures or arterial vessels

European Neuroendocrine Tumour Society (ENETs) and tumour stage (T).

The breakdown for the 95 PNET samples included 30 insulinomas **(Table 6)**, 40 NF-PNETs **(Table 7)**, 17 gastrinomas **(Table 8)**, 7 rare functioning PNETs (3 VIPomas, 2 somatostatinomas, 1 PTHoma and 1 glucagonoma) and 7 patients with MEN1 (**Table 9)**.

Table 6: Clinical details for insulinoma samples from the University of Vienna

ID	Age[a]	Sex	F/up	Status[b]	NET Grade	Ki67	Tumour size (mm)	ENETs
INS1[c]	73	F	170	1	1	1	15	1
INS2[d,e]	53	NA	55	1	1	1	30	3
INS3[c]	12	M	138	0	1	1	15	1
INS4[c]	33	M	236	0	1	1	13	1
INS5	30	NA	302	0	1	1	45	1
INS6	47	M	244	0	1	1	20	1
INS7	55	NA	122	1	1	1	20	1
INS8[c]	48	F	192	0	1	1	20	1
INS9[c]	62	M	51	0	1	0.8	5	1
INS10	74	F	107	0	1	1	15	1
INS11[c]	80	M	33	0	1	2	8	1
INS12[c]	94	F	43	1	1	1.6	17	1
INS13[c]	46	M	94	0	2	3	13	1
INS14[d]	41	F	154	0	2	3	50	3
INS15[c]	43	F	153	0	1	1	20	1
INS16	61	F	86	0	1	1	15	1
INS17	42	M	129	0	1	1	10	1
INS18[c]	44	M	289	1	1	1	10	1
INS19[c]	68	M	25	1	1	1	25	2
INS20	72	F	1	1	NA	NA	15	1
INS21[c]	71	F	88	1	1	1	18	1
INS22[c]	68	F	1	0	2	2.4	18	1
INS23[c]	58	F	1	0	1	2.2	13	1
INS24[c]	31	M	43	0	2	3.4	10	1
INS25[c]	80	F	1	1	1	0.8	19	1
INS26	29	F	30	0	1	1	30	2
INS27[c]	40	M	26	0	1	1	18	1
INS28[c]	28	M	95	0	1	1.5	15	1

[a]Age at diagnosis, [b]dead = 1, alive=0 at last follow-up, [c]normal adjacent islets on the same histology slide as the PNET, [d]Liver-directed therapy, [e]Chemotherapy, no patients had peptide receptor radiotherapy. European Neuroendocrine Tumour Society Staging; Follow-up (F/up; months); Millimetres (mm); Not available (NA).

Table 7: Clinical details for NF-PNET samples from the University of Vienna

ID	Age[a]	Sex	F/up	Status[b]	NET Grade	Ki67	Tumour size (mm)	ENETs
NF1	69	NA	29	0	1	0.5	45	3
NF2[c]	66	M	5	0	1	2	27	2
NF3	71	M	180	1	1	1	60	3
NF4	35	F	59	0	2	5	50	3
NF5[c]	50	F	116	0	1	1	48	3
NF6	52	F	13	1	2	3	38	3
NF7	55	M	212	1	1	1	100	3
NF8[c]	28	F	267	0	2	4	35	2
NF9	41	NA	124	1	1	1	43	3
NF10[d]	44	M	10	1	2	9	50	3
NF11	61	F	87	1	1	1	50	3
NF12[d]	39	M	8	1	1	1	65	3
NF13[c]	49	F	150	0	1	1	17	1
NF14[e]	45	F	71	1	1	1	70	3
NF15[c]	69	F	13	1	1	1	80	2
NF16[c]	60	M	0	1	1	1	45	3
NF17[f]	53	F	114	1	1	1	100	3
NF18[c]	59	M	6	1	3	70	21	2
NF19	26	F	184	0	1	1	53	3
NF20[c]	53	F	154	0	1	1	40	2
NF21	59	F	1	0	2	7	30	4
NF22[c,d]	77	M	22	0	2	3	40	3
NF23[d]	58	M	52	0	2	5	27	2
NF24[f]	40	F	21	1	3	30	180	3
NF25	64	F	256	0	1	1	22	2
NF26	51	F	65	1	1	2	NA	NA
NF27[d]	37	F	166	0	1	2	80	3
NF28	62	M	100	0	1	1.3	16	2
NF29[f]	68	F	21	1	1	1	80	3
NF30	55	M	32	0	3	25	40	3
NF31[c]	68	M	5	0	1	1	17	1
NF32[c]	63	F	0	0	1	1	15	1
NF33[c]	34	M	87	0	2	8	30	2
NF34[c]	44	M	84	0	2	3.2	22	2
NF35[c]	77	M	2	0	1	1	10	1
NF36[c]	55	F	58	0	2	5	22	2
NF37[d,e]	52	F	80	0	2	10	NA	0
NF38[c,f]	48	M	24	1	2	4	50	3

[a]Age at diagnosis, [b]dead = 1, alive=0 at last follow-up, [c]normal adjacent islets on the same histology slide as the PNET, [d]Liver-directed therapy, [e]Somatostatin analogue therapy, [f]Chemotherapy, no patients had peptide receptor radiotherapy. European Neuroendocrine Tumour Society Staging; Follow-up (F/up; months); Millimetres (mm); Not available (NA).

Table 8: Clinical details for gastrinoma samples from the University of Vienna

ID	Age[a]	Sex	F/up	Status[b]	NET Grade	Ki67	Tumour size (mm)	ENETs
GS1	22	M	235	0	1	1	60	3
GS2	39	M	161	0	1	1	NA	NA
GS3	74	NA	1	0	1	1.6	30	2
GS4	26	F	60	0	2	6	15	1
GS5[c]	38	F	46	1	1	1	200	4
GS6[c]	55	M	36	1	1	1	70	4
GS7	36	M	269	0	1	1	30	2
GS8	77	M	29	0	1	1	10	1
GS9	40	M	255	0	1	1.5	30	2
GS10	67	M	52	0	1	1	11	2
GS11	62	F	54	1	1	1	10	1
GS12	60	F	0	1	1	1	NA	3
GS13	42	M	0	1	1	1	40	3
GS14[d,e]	61	F	10	0	3	25	42	3
GS15	22	M	69	0	2	3	11	2

[a]Age at diagnosis, [b]dead = 1, alive=0 at last follow-up, [c]chemotherapy, [d]Liver-directed therapy, [e]Somatostatin analogue therapy, no patients had peptide receptor radiotherapy. European Neuroendocrine Tumour Society Staging; Follow-up (F/up; months); Millimetres (mm); Not available (NA)

Table 9: Clinical details for individuals with MEN1 and for rare functional PNETs from the University of Vienna

ID	Age[a]	Sex	F/up	Status[b]	NET Grade	Ki67	Tumour size (mm)	ENETs
MEN1								
Gastrinoma	47	M	3	0	1	1	17	1
VIPoma	29	M	177	0	1	2	80	3
Insulinoma	15	M	128	0	2	3	20	1
Non-functioning	60	F	127	0	2	5	30	3
Gastrinoma	43	M	190	0	1	1	40	2
Non-functioning	60	NA	96	0	1	1	25	2
Insulinoma	33	NA	286	0	1	1.2	25	2
Rare PNETs								
PTHoma	63	M	73	0	1	1	70	3
VIPoma	37	NA	148	0	1	2	15	1
Somatostatinoma	50	M	128	0	1	1	12	1
Glucagonoma	51	NA	62	0	1	1	110	3
VIPoma	38	NA	0	0	1	1	40	2
VIPoma	29	F	73	0	2	5	30	2
Somatostatinoma	73	F	57	1	1	1	60	3

[a]Age at diagnosis, [b]dead = 1, alive=0 at last follow-up. European Neuroendocrine Tumour Society Staging; Follow-up (F/up; months); Millimetres (mm); Not available (NA).

3.2.2 ORB prospective PNET and normal pancreatic sample

Prospective PNETs (n=3) and normal pancreatic tissue (n=21) collection commenced in October 2019 from the ORB (Research Ethics Committee reference: 19/A131; 09/H0606/5+5)(**Table 10**). Tissue and blood samples (PNET patients only) were collected on the day of surgery. Samples were divided into media (1mm^3 pieces) and stored at -80°C or fixed with 4% paraformaldehyde (PFA), for embedding. After 24 hours, 4% PFA fixed samples were transferred to 70% ethanol and embedded in paraffin.

Table 10: Oxford Research Biobank prospectively collected samples

ID	Age (years)	Sex (M/F)	Primary reason for pancreatic resection
Samples analysed			
11882	69	M	Low grade IPMN
11893[a]	50	M	PDAC
11898	77	F	PDAC
11982	70	M	PDAC
11969[a]	71	F	Grade 2 NF-PNET
11974[b]	52	F	Duodenal adenocarcinoma
11981	75	M	PDAC
11995	69	M	PDAC
11999[b]	55	M	No malignancy
12002[a]	68	M	IPMN
12073[b]	41	M	Grade 2 NF-PNET
12081	61	M	Grade 1 NF-PNET
12109	73	M	PDAC
Samples without adjacent normal islets			
NA	75	Female	PDAC
NA	63	Female	PDAC
NA	60	Male	PDAC
NA	58	Male	PDAC
NA	62	Male	PDAC
NA	73	Female	PDAC
NA	78	Female	PDAC
NA	79	Female	PDAC
NA	54	Male	Pancreatic cyst
NA	69	M	PDAC
NA	54	M	Benign cyst

Male (M); Female (F); Non-functioning PNET (NF-PNET); Intraductal papillary mucinous neoplasm (IPMN); Pancreatic ductal adenocarcinoma (PDAC). [a]Samples sent for methylationEPIC array 5hmC analysis; [b]Samples sent for methylationEPIC array bisulfite-only analysis.

3.2.3 RNA samples from PNETs from the Mayo clinic

RNA samples from 39 PNETs (14 insulinomas and 25 NF-PNETs) were sent from the Mayo clinic on dry ice **(Table 11).** RNA was provided as a gift from Professor Gloria Petersen, Professor of Epidemiology, through an MTA under Mayo Clinic Institutional Review Board (IRB) number #354-06.

Table 11: Mayo clinic PNET RNA samples

ID	Age (years)	Sex (F/M)	NET Grade	Tumour Size (mm)	ENETs Stage
Insulinoma					
1	60	F	1	NR	NR
2	63	M	1	NR	NR
3	54	F	1	NR	NR
4	73	M	1	NR	NR
5	57	M	1	NR	NR
6	54	F	1	NR	NR
7	50	F	1	NR	NR
8	59	M	1	NR	NR
9	39	F	1	NR	NR
10	72	M	2	NR	NR
11	37	F	2	NR	NR
12	53	M	2	NR	NR
13	73	M	1	NR	NR
14	69	F	1	NR	NR
NF-PNET					
1	57	F	1	NR	NR
2	74	F	1	NR	NR
3	70	M	1	NR	NR
4	78	M	1	NR	NR
5	72	F	1	NR	NR
6	61	F	2	NR	NR
7	63	F	2	NR	NR
8	20	F	2	NR	NR
9	64	F	2	NR	NR
10	56	F	2	NR	NR
11	76	M	2	NR	NR
12	57	F	2	NR	NR
13	61	F	2	NR	NR
14	73	M	3	NR	NR
15	72	M	3	NR	NR
16	50	F	3	NR	NR
17	64	F	1	NR	NR
18	66	F	1	NR	NR
19	39	F	1	NR	NR
20	33	M	2	NR	NR
21	61	M	1	NR	NR
22	66	M	2	NR	NR
23	68	F	2	NR	NR
24	51	F	1	NR	NR
25	57	M	2	NR	NR

Female (F); Identifier (ID); Male (M); Millimetres (mm) and not recorded (NR).

3.2.4 PNET IHC characterisation

Vienna FFPE samples and ORB samples were sectioned using a microtome (Leica RM2255 rotary microtome; Leica Biosystems) at 0.4µm thick sections. Sections were floated in distilled water at 47°C, mounted on positively charged glass slides (Superfrost^{TM} Plus glass slides; Thermo Fisher Scientific), incubated at 37°C overnight, catalogued and stored. Each primary antibody was optimised for use with historic FFPE tissue. Sections were deparaffinised in Histoclear x 3 (National Diagnostics), rehydrated in graduated industrial methylated spirits (IMS) concentrations (100% x 2, 95% x 2, 80% x 1; 5 minutes each), washed in running tap water and distilled water (dH_2O) prior to antigen retrieval using Citrate Buffer pH 6 (Cytoplasmic antigens: no retrieval (staining for glucagon) or pressure cooker with a 10 minute cycle at 121°C (staining for insulin, synaptophysin, chromogranin A); nuclear antigens with an autoclave (Osprey 40L benchtop autoclave; LTE Scientific) for 20 minutes at 121°C (DNA 5hmc and menin). Sections were blocked with 10% hydrogen peroxide (30% w/w in dH_2O, Sigma, STBH5484) for 10 minutes, washed with TBS-T (10 minutes x 2), blocked for 30 minutes with 10% serum from the host that the secondary antibody was raised prior to primary antibody application. Primary and secondary antibody concentrations are shown in **Table 12**. Slides were incubated overnight (at least 16 hours) at 4°C. Sections were washed in TBS-T (10 minutes x 2), secondary antibody applied for 1.5 hours, washed in TBS-T (10 minutes x 2) and stained with diaminobenzidine (DAB) peroxidase (Vector Laboratories, SK-4100) for 10 minutes prior to counterstaining with haematoxylin QS (Vector Laboratories, H-3404). Slides were then washed, dehydrated for 5 minutes each in increasing IMS concentrations (80%, 90% x 2, 100% x 2, Histoclear x 2 followed by a final xylene wash and mounted with Vectamount (Vector Laboratories, H-5000-60) with a glass covering (SLS Coverslips No 1. 22 x 64mm, MIC3228). For haematoxylin and eosin (H&E) staining, slides were rehydrated (same protocol), washed

with dH_2O, haematoxylin QS stained for 45 seconds, washed in running tap water for 10 minutes, rinsed in dH_2O, dipped in 95% IMS, stained with 0.5% Eosin Y (Sigma Aldrich, 318906) with 0.25% glacial acetic acid (Scientific Laboratory Supplies, CHE1018) for 7 minutes, dipped in 70% IMS, rehydrated and mounted (same protocol).

3.2.4.1 5hmC quantification

Digital images were taken using the Eclipse E400 microscope (Nikon) and captured at 20X magnification (DXM1200 digital camera; Nikon). 5hmC quantification was undertaken using QuPath software (version 0.3.2) to obtain H scores for each cell compartment. Cell detection and H scores were analysed using the following parameters: Detection image (Optical density (OD) sum), with Intensity Threshold Parameters: Nucleus DAB OD mean, threshold 1+ 0.1, threshold 2+ 0.2, threshold 3+ 0.3. Paired raw H scores between cell compartments were analysed using a Paired t-test or One-way ANOVA with p-values adjusted for multiple comparisons using Prism software (version 9.4.0). Fold changes between cell compartments were analysed using adjusted H scores (normal islets adjusted to an H score of 150) to account for different staining intensities between slides.

3.2.4.2 Immunohistochemistry antibodies

Primary and secondary antibodies used in immunohistochemistry are shown in **Table 12**.

Table 12: Primary and secondary antibodies for immunohistochemistry

Antibody	Company	Catalogue no.	IHC dilution
Primary			
Insulin	Abcam	Ab 7842	1:50
Glucagon	Abcam	Ab 92517	1:1000
Chromogranin A	Abcam	Ab 15160	1:200
Synaptophysin	Abcam	Ab 32127	1:500
5hmC	Active Motif	39769	1:250
Ki67	Abcam	Ab 15580	1:500
Menin	Abcam	Ab 2605	1:500
Secondary			
Goat anti-rabbit	Jackson ImmunoResearch	111-035-006/ 148674	1:200
Rabbit anti-guinea pig	Dako	P0141 / 00074093	1:100

3.3 DNA extraction

DNA extraction was performed using the DNA FFPE Tissue Kit, Qiagen, 56404. 1cm^2 of tissue x 10µm-thick sections were obtained using the microtome (Leica RM2255 rotary microtome; Leica Biosystems) and excess paraffin removed with a scalpel and placed in a 1.5mL Eppendorf tubes. The QIAamp protocol was followed as per the manufacturer's instructions apart from 1) deparaffination solution (500µL; Qiagen 19093) was used, and samples were digested overnight with proteinase K (30µL) at 56°C. The following morning all samples were mixed gently with a pipette and if not fully digested, further proteinase K (20µL) was added and digestion continued until the sample was completely lysed. RNAse A (2µL; Qiagen 19101) was added to each sample. DNA was eluted in nuclease free water (35-50µL) and stored at -80°C.

3.3.1 DNA quality assessment

DNA was extracted from all possible tumour samples, and where possible, DNA from adjacent normal pancreatic tissue containing islets. DNA quantity was assessed with the NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific), and Agilent Tapestation 2100. The average DNA concentration on Agilent tapestation for all samples was 68.38 ± 43.81ng/µL compared to 246ng/µL ± 193.6ng/µL using the Nanodrop. DNA purity was assessed using the 260/280 and 260/230 ratios on the NanoDrop. Single nucleotides, RNA and DNA all absorb at a wavelength of 260nm, and to assess for sample purity the 260/280 ratio can be used, with ratio ~>1.8 indicating pure DNA. Lower ratios can be seen with contamination (e.g. proteins and some reagents used for DNA extraction). The 260/230 ratio also assesses DNA purity, with an expected value ~2-2.2, this assesses for the contaminants to absorb at a 230nm wavelength (e.g. ethylenediaminetetraacetic acid (EDTA) and phenol). The average DNA purity on NanoDrop for 260/280 and 260/230 was

1.896 ± 0.040 and 1.917 ± 0.226, respectively. The average DNA integrity number (DIN) was 3.27 ± 1.292 for all samples. An example of the DNA integrity as analysed by Agilent Tapestation is shown in **Figure 13**. DNA integrity numbers (DIN) values range from 1 to 10, where 1 is a degraded sample and 10 is intact.

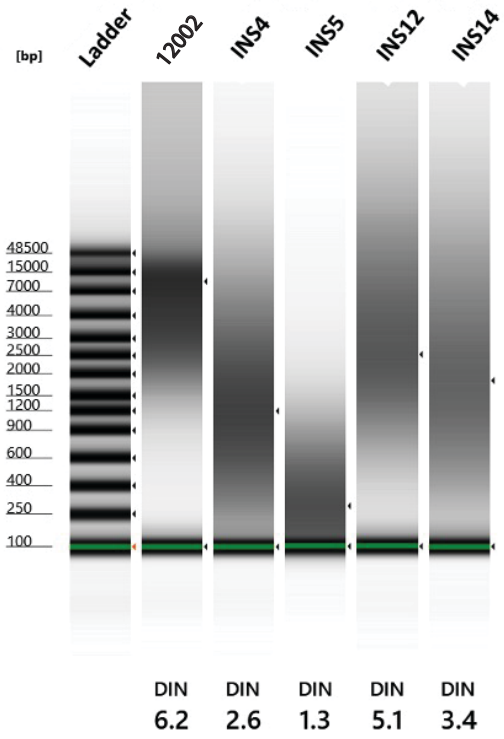


Figure 13: DNA integrity (DIN) as analysed on the Agilent 2100 Tapestation Bioanalyser for a subset of samples.

3.4 Polymerase Chain Reaction (PCR)

DNA quality for use in downstream applications was assessed via polymerase chain reaction (PCR) using generated primers (Primer-BLAST) for *RASSF1* 244, 516 and 775 base pairs and *MEN1* of 250, 500 and 750 base pairs (Sigma Aldrich). PCR was performed on a thermal cycler (Pendragon Scientific LTD, MJ Research, Tetrad PTC 225) with a cycle length of 35 cycles using the primer sequences and annealing temperatures (**Table 13)**. Tetrad program was as follows:

TETRAD program was as follows:

Step 1 95°C 3 minutes

Step 2 95°C 30 seconds

Step 3 (Specific optimized temp per primer set) 40 seconds

Step 4 72°C 30 seconds

Step 5 Go to Step 2, 30-35 times

Step 6 72°C 5 minutes

Step 7 4°C 5 minutes

Step 8 END

DNA ladder (Hyperladder 100bp, Bioline, HR-719101A) was electrophoresed with samples on a 1.25% agarose gel (UltraPure Agarose, Invitrogen, 15510-027) diluted in Tris-Borate-EDTA (TBE; Sigma T4415-4L) with 1μL of ethidium bromide (Sigma, 1001891876) per 50mL of agarose gel. The electrophoresis of the gel was performed in TBE solution at 100 volts for 2 hours prior to imaging with a ChemiDoc (Bio-Rad).

Table 13: DNA PCR primers for *RASSF1* and *MEN1*

Gene	Generated Primer Sequence	Temperature (° Celsius)
RASSF1A (244) F	5' – CCGCAAGTTTGCACTCTTTGAGC	67
RASSF1A (244) R	5' – CTTCAGGACAAAGCTCAGGGC	
RASSF1A (516) F	5' - ACACCTGACCTTTCTCAAGC	68
RASSF1A (516) R	5' – AAGAACTTTCGCAGCAGGGC	
RASSF1A (775) F	5' – AGACACCTGACCTTTCTCAAGC	65
RASSF1A (775) R	5' – CTTCAGGACAAGCTCAGGGC	
MEN1 (250) F	5' – GGCAGAAGGTGCGCATAGTG	67
MEN1 (250) R	5' – TGAGCCGTGCTGCCAACCTTC	
MEN1 (500) F	5' – GGCAGAAGGTGCGCATAGTG	62
MEN1 (500) R	5' – CACAAGCGGTCCGAAGTCC	
MEN1 (750) F	5' – GCAGAAGGTGCGCATAGTG	64
MEN1 (750) R	5' – GACTGGTAATTAGGACCCAGCG	

Forward (F); Reverse (R).

DNA concentration and quality control values were analysed on two machines NanoDrop (Thermo Fisher Scientific) and Agilent Tapestation 2100 Bioanalyser **(Table 14)**.

Table 14: Quality control values for each sample included for DNA 5hmC analysis

ID	Type	NanoDrop[a]			TapeStation[b]					
		DNA (conc ng/µL)	260/280	260/230	Total DNA (ng)	DNA conc (ng/µl)	DIN	Total DNA (ng)	RASSF1 250bp	RASSF1 500bp
NF35 (T)	NF-PNET	237.3	1.91	2.04	8305.5	58.6	2.7	2051	1	1
NF35 (N)	Adj normal	184.9	1.88	1.94	6471.5	44.8	2.4	1568	1	0
NF36 (T)	NF-PNET	590.4	1.9	2.14	70848	89.8	3.2	10776	1	1
NF36 (N)	Adj normal	165.4	1.89	1.64	6616	76.7	5.6	3068	1	1
11947 (T)	NF-PNET	126.8	1.96	1.93	4438	43.6	5.8	1526[c]	1	1
11969 (N)	Adj normal	153.5	1.96	1.9	5372.5	69.8	5.5	2443	1	1
NF24	NF-PNET	253.3	1.92	2.01	17731	59.8	5.1	4186	1	1
NF34	NF-PNET	367.2	1.92	2.16	12852	73.4	4	2569	1	1
INS9 (T)	Insulinoma	176.8	1.89	1.96	7072	84.7	4.6	3388	1	1
INS9 (N)	Adj normal	171.7	1.9	1.93	6868	74.4	3.8	2976	1	1
INS13T	NF-PNET	218.9	1.9	1.85	15323	61.7	2.2	4319	1	1
INS13 (N)	NF-PNET	209.8	1.9	2.04	10490	67.9	3	3395	1	1
INS23	Insulinoma	282.4	1.91	2.11	19768	94.1	4	6587	1	1
INS27	Insulinoma	543.9	1.91	2.16	43512	110	4.1	8800	1	1
11893 (N)	Normal pancreas	415.2	1.94	2.15	49824	66.3	2.4	7956	1	0
12002 (N)	Normal pancreas	NA	NA	NA	2395	47.9	6.2	2395	NA	NA

[a]DNA concentration and quality were assessed by NanoDrop Microvolume Sphectrophotometer (Thermo Fisher Scientific)
[b]DNA concentration and quality were assessed by the Agilent Tapestation 2100 Bioanalyser.
[c]Sample 11947 was combined with a second DNA extraction to make up to 2ng of total DNA.
Adjacent normal pancreatic tissue (Adj normal); DNA integrity number (DIN); Identification (ID); Normal pancreatic tissue (N); Not available (NA); Tumour (T).

3.5 Bisulfite and oxidative-bisulfite conversion

Samples from 11 people with sporadic PNETs **(Methods 3.2, Tables 6, 7, 10 and 14)** underwent oxidate-bisulfite (oxBS) conversion using the TrueMethyl oxBS Module (Part No. 0414-32). DNA (2ng) was obtained for each sample and split so that 1ng of DNA could undergo paired oxBS conversion and bisulfite conversion only. The protocol was followed as per previously reported[211] apart from during the bisulfite conversion, where instead of 40µL of solution undergoing bisulfite conversion in one 200µL PCR tube, samples were split into two 20µL volumes in two 200µL PCR tubes. **Figure 14** shows an example of the oxBS-paired samples. Importantly, following the oxidative conversion, the oxBS samples were orange, which indicates a successful oxidation, and the bisulfite-only samples remain clear. The oxidant was stored on ice and transferred to each PCR tube in a darkened room, away from possible airborne contaminants (particularly ethanol) for the oxidative process. The protocol was then continued as per manufacturer's instructions and as previously reported.[211]

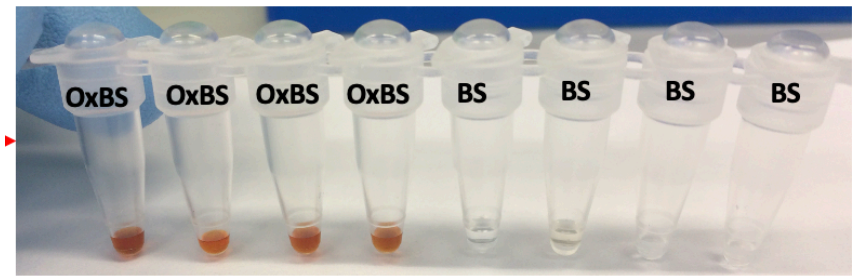


Figure 14: An example of 4 samples split into two parallel processes so that each sample underwent both paired oxidative bisulfite (oxBS) and bisulfite-only (BS-only) processes.

A further 12 samples underwent bisulfite conversion only (Insulinomas (n=5), NF-PNETs (n=5) and BON1 (n=1) and QGP1 (n=1) cell lines **(Methods 3.2, Tables 6,7 and 10)**. Samples were sent for methylationEPIC array analysis at Cambridge epigenetics, and data

was received back via red and green channel jpg files and analysed using R (ChAMP) pipeline.[212]

3.6 Comparison of the methylation CpGs in human PNETs to the normal islet cell.

Methylation values for CpGs were available for 18 PNETs (insulinomas (n=9) and NF-PNETs (n=9). Loyfer et al., reported unmethylated DNA regions in the normal β and α cell, which required at least 4 consecutive unmethylated CpGs before that region was deemed unmethylated.[213] The remaining CpGs not present within these unmethylated regions in the normal β cell may have been methylated or in groups of <4 consecutive unmethylated CpGs. Therefore, for this analysis 'methylation' is defined as a beta level >0.3, as previously reported.[214] Unmethylated regions of normal β and α cells were downloaded from Loyfer et al., to be used to compare the methylated samples for FFPE fixed PNETs.[213]

3.7 RNA extraction from FFPE tissue

3.7.1 RNA Quality Control

RNA was extracted from FFPE tissue (Qiagen RNeasy FFPE ID: 73504) and analysed on the Tapestation Agilent 2000 using (Agilent RNA Screentape analysis ID: 5067-5576). All patient samples that had been sent for methylationEPIC array were sent for consideration of RNA-seq. RNA samples, volumes and Quality Control (QC) data are shown in **Table 15**.

Table 15: RNA quality control prior to sending for RNA-seq

Patient ID	NanoDrop conc (ng/µL)[a]	260/280	260/230	Tapestation conc (ng/µL)[b]	RIN	DV200	DV100
Insulinoma							
INS14	87.8	1.9	1.7	24.4	2.3	17	35.65
INS1	113.7	1.78	1.13	15.8	2.1	13.22	NA
INS12	342.3	2.03	2.01	60.2	2.1	7.69	27.46
INS13(N)	449.1	2.02	2.06	145	1.4	29.4	59.42
INS9(N)	277.8	2.01	1.94	115	1.6	23.07	51.25
INS9(T)	641.5	2	2.06	124	1.9	13.74	37.71
INS23	170.7	2.02	2.02	40.8	1.8	13.15	39.27
INS22	214.1	1.97	1.69	62.8	1.6	9.11	30.66
INS25	293	1.91	1.47	80.1	1.6	21.78	49.95
INS27	212.2	2.06	1.91	61	2.1	13.9	39.95
INS26	646.1	1.98	1.53	249	2.4	61.62	80.13
Non-functioning							
NF10	129.6	1.66	0.88	17.2	2.3	17.78	NA
NF14	103.1	1.89	1.99	26	1.4	25.59	50.45
NF24	51.5	1.97	1.55	22.1	2.3	13.55	34.89
NF34	124.5	2.02	1.83	68.1	1.8	27.79	56.31
NF35(T)	80.3	1.88	1.37	53.7	2	49.37	71.6
NF35(N)	287.9	1.98	1.83	199	2.5	63.52	81.89
NF36(T)	156.6	1.95	1.66	47.8	1.8	26.32	52.09
NF36(N)	512.4	2	1.69	109	1.8	21.8	49.68
ORB 12066	10.4	1.16	1.54	3.26	NA	25.73	40.91
ORB 11947	103.8	1.87	1.28	71.1	2.2	77.76	88.53
ORB 12064	28.3	1.79	0.94	13.1	NA	66.6	79.58
Normal pancreas							
ORB 11893(N)	782.4	2.05	2.1	387	1.7	65.9	83.12
ORB 11999(N)	100.6	1.79	1.18	40.9	3.1	61.79	75.89
ORB 11969(N)	181.9	1.96	1.55	104	3.1	85.5	NA
ORB 11974(N)	22.1	1.73	0.93	7.93	NA	4.26	NA
ORB 12073(N)	16.5	1.91	1.52	13.9	8.8	29.14	50.18
ORB 12002(N)	116.1	2.04	2.06	57	1.8	36.2	63.28

Adjacent (adj); Normal pancrease (N); Not available (NA); Non-function PNET (NF); Oxford Research Biobank (ORB); RNA integrity number (RIN). Percentage of RNA fragments >100bp (DV100; Percentage of RNA fragments >200bp (DV200).

[a]DNA concentration and quality were assessed by NanoDrop Microvolume Sphectrophotometer (Thermo Fisher Scientific)

[b]DNA concentration and quality were assessed by the Agilent Tapestation 2000.

3.7.2 RNA analysis

RNA was sent to the Oxford Genomics Centre and samples were run on the NEBNext Ultra I Directional RNA Library Kit for Illumina using the ribodepletion protocol for FFPE samples.[215] Results were available for download (using terminal and checked with md5sum) for FASTQ (stores nucleotide base sequences and quality scores)[216] BAM files (compressed file used for alignment to a reference genome) from paired-end sequencing reads. Samples had been run through the Oxford Genomics pipeline and further analysis was undertaken using the provided BAM files. All BAM files were uploaded to the Galaxy server, an open source, web-based platform where sequencing data can be uploaded and analysed. RNA-seq files were annotated using the GTF file (stores information about gene structure) homo_sapiens.GRCh37.75.gtf. Matched BAM files from two sequencing runs were joined using Samtools, and then Feature counts undertaken. DEseq2 analysis was then performed, which reports differentially expressed genes. Quality control metrics for RNA-seq samples are shown in **Figure 15.** Sample groups compared included: insulinomas vs controls; NF-PNETs vs controls; all tumours vs controls; and insulinomas vs NF-PNETs. Differentially expressed genes were represented as Volcano plots and generated using the Galaxy server. DE2seq2 output from Galaxy server was subsequently uploaded into R Studio (version 4.2.0 "Vigorous Calisthenics") and GOseq analysis was completed[217] (**Supplementary Appendix 8.2, RNA-seq code**).

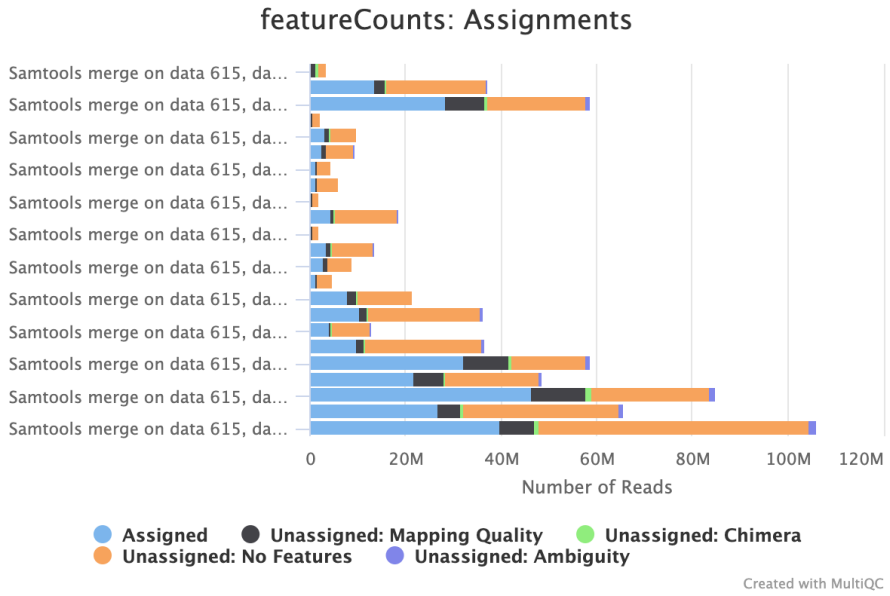


Figure 15: Quality control for RNA-seq from RNA collected from FFPE tissue.

3.8 Tissue culture

Two human PNET cell lines, BON1 (derived from the lymph node from a 28-year-old patient with metastatic insulinoma)[218] and QGP1 (derived from a metastatic somatostatinoma),[219] were used for drug treatment and cell viability assays. The BON1 cell line was used as a model for insulinomas and the QGP1 cell line was used as a model for a NF-PNETs. BON1 cells were provided by Dr Simona Grozinsky-Glasberg (Hadassah-Hebrew University Hospital, Israel) as a gift to the University of Oxford and were grown in Dulbecco's Modified Eagle Medium/F12 (DMEM/F12; Gibco) with 10% fetal bovine serum (FBS; Sigma-Aldrich). QGP1 cells were purchased from the Japanese Collection of Research Bioresources Cell Bank (Osaka) and were grown in Roswell Park Memorial Institute (RPMI)-1640 media (Gibco) with 10% FBS. Cell lines were authenticated using Short Tandem Repeats (STR) profiling, organised by Dr Kate Lines.[220] Cells were grown in 75T flasks (Corning) and incubated at 37°C with 5% CO_2. Cells were grown to 75% confluence and either passaged using 0.05% Trypsin-EDTA (Thermo Fisher Scientific) or used in experiments.

3.8.1 Drug treatment and cell viability assays

Cells were trypsinised for 5 minutes in 75T flasks at 37°C, 5mL of appropriate media was added to stop trypsinisation. Cells were removed and centrifuged at 1000g for 3 minutes, supernatant was removed, and fresh media applied, cells were then counted using a haemocytometer (Hirschmann Em-Techolor). Cells were resuspended in the appropriate volume of media for further experiments. For cell viability assays, cells were seeded (50µL) into 96-well black plates and a further 50µL of drug and media added. Ascorbic acid (AA) (Abcam; ab142830) was added to specific wells to make concentrations of 50µM, 100µM and 500µM, of AA respectively. Cell viability was performed at baseline and at day 7, using CellTiter-Blue cell viability assay, and day 7 readings normalised to baseline. DNA was collected at baseline and day 7 for dot blot analysis. Cells were grown 12 well plates to ensure enough DNA was able to be extracted for dot blot analysis.

3.8.2 DNA extraction

DNA from cells was extracted, as previously described.[211] Briefly, plates were washed in phosphate buffered saline (PBS) and trypsin (200µL) was added to each well and incubated at 37°C for 3 minutes. Media (200µL) were added to each well, and samples transferred to 1.5mL Eppendorf tubes. Samples were centrifuged at maximum speed for 1 minute and the supernatant was carefully poured out. Samples were pipetted and vortexed to resuspend the cells in solution prior to adding cell lysis buffer (600µL) and re-vortexed at maximum intensity for 20 seconds (DNA extraction kit). Samples were incubated at 37°C to ensure complete cell lysis. RNAse (2µL) was added. Protein precipitate (600µL) was added and vortexed for 10 seconds. Isopranolol (600µL) was added which allowed DNA to clump, samples were then centrifuged at maximum speed for 1 minute, and isopranolol supernatant was removed. 70% ETOH (600µL) was added and tubes allowed to air dry. Samples were re-suspended in nuclease free water (50µL) and incubated for 1 hour at 65°C. DNA

concentration was analysed using the NanoDrop (Thermo Scientific Fisher) and samples were adjusted with nuclease free water to a concentration of 100ng/µL.

3.8.3 RNA extraction

Plates containing cells treated with and without AA for 7 days were washed in PBS and RTL buffer (200µL) was added to each well and stored at -80°C. RNA was extracted from cell lines using the RNeasy Kits for RNA Purification (Cat. No. 74104).

3.8.3.1 cDNA synthesis

1µg of RNA was converted to cDNA using the QuantiTect Reverse Transcription Kit (Qiagen). Genomic DNA was removed using gDNA Wipeout Buffer for 2 minutes at 42°C. All samples were put on ice, and mixed with Reverse Transcriptase, RT buffer and RT primer mix as per manufacturer's instructions. The samples were then heated for 15 minutes at 42°C, followed by 3 minutes at 95°C. cDNA was diluted 1:4 in nuclease free water and used for subsequent Quantitative reverse transcription polymerase chain reaction (qRT-PCR).

3.8.3.2 qRT-PCR analysis

qRT-PCR was performed using the QuantiTect SYBR Green PCR Kit (Qiagen). Gene specific primers are shown in **Table 16**. Briefly, 4µL of cDNA template was added to reaction mixture (36µL; QuantiTect SYBR Green Master mix, including 12µL of nuclease free water). The total solution (40µL) was added to each PCR well (9.5µL), and samples were analysed in quadruplicate. qRT-PCR reaction was performed on a Rotor-Gene Q cycler (Qiagen).

Table 16: List of primers used for qRT-PCR

Gene symbol	Gene name	Qiagen catalogue number
INS	*Insulin*	QT01531040
PTPRT	*Protein tyrosine phosphatase receptor type T*	QT00097923
GPR119	*G protein-coupled receptor 119*	QT00231609
STEAP3	*STEAP3 Metalloreductase*	QT00201033
CANX	*Calnexin*	QT00092995
GAPDH	*Glyceraldehyde-3-phosphate dehydrogenase*	QT00079247

3.9 *SSTR* transfection

BON1 cells were transfected with *SSTR2* receptor using the pCMV6-Entry Mammalian Vector with C-terminal Myc-DDY Tag. *SSTR2* sequence was checked by Sanger sequencing **(Figure 16).**

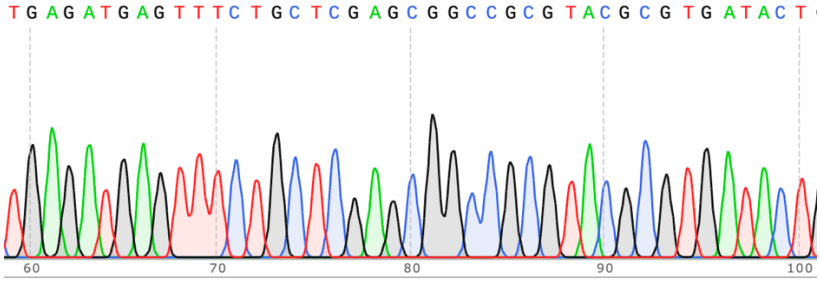


Figure 16: Example of DNA Sanger sequencing from a section of DNA.

BON1 and QGP1 cells were transfected with *SSTR2* (Stock plasmid *SSTR2* (100ng/µL)). Cells were transfected in a 24 well plate with two replicates of untreated and transfected cells for each cell type. Solution A) plasmid *SSTR2* solution (4µL) was added to an Eppendorf made up to 100µL of serum free media, and Solution B) lipofectamine (4µL) was added to a second Eppendorf made up to 100µL with serum free media. Solutions A and B were incubated together for 5 minutes. Cells were left for 3 days and then treated with geneticin (400µL of 50mg/mL per 50mL of media). Cells were monitored until all untransfected cells died. Cells in the transfected wells were then put into 25mL flasks and grown under selection media (geneticin). SSTR2 expression was confirmed by immunocytochemistry.

3.10 Immunocytochemistry

Immunocytochemistry was performed as previously described (BestProtocols: ICC Methanol Fixed Cells-Direct Method; Thermo Fisher Scientific;

https://www.thermofisher.com/ie/en/home/references/protocols/cell-and-tissue-analysis/protocols/icc-methanol-fixed-cells-direct-method.html). Briefly, circular coverslips were placed into 6 well plates and then washed with 100% ethanol for 10 minutes. Ethanol was removed and allowed to airdry, the coverslips were then washed for 5 minutes in PBS x 3. Poly-D-lysine hydrobromide (300µL; Sigma-Aldrich P6407-5MG) was added to the centre of the coverslip and incubated at room temperature for 20 minutes. Coverslips were washed for 5 minutes in PBS x 3 and allowed to air dry. 500µL of 5 x 10^3 cells were added to each coverslip and left for 24 – 48 hours to allow cells to adhere to the coverslip. Cells were washed for 5 minutes in PBS x 3, and 100% methanol (cooled to -20°C) was added to each well for 5 minutes, followed by PBS x 3 washes (5 minutes each). Coverslips were blocked in 1% BSA for 30 minutes (0.5mg in 50mL PBS), followed by a further 3 x 5-minute washes in PBS. The primary antibody to the DYK Tag (DYKDDDDK Tag Rabbit AbCell signalling 2368S) was added to each coverslip at 1:1000 concentration and placed on the rocker for 1.5 hours, with one sample as a negative control. Samples were washed in 3 x PBS for 5 minutes prior to the addition of the secondary antibody (GAR fluoro 568) at 1:2000 concentration, under an aluminium foil cover for 1 hour on the rocker. Samples were washed in 3 x 5-minute washes in PBS. Coverslips were reverse mounted on slides with DAPI (4',6-diamidino-20-phenylindole) gold and imaged using the Eclipse E400 microscope (Nikon) and captured at 20X magnification (DXM1200 digital camera; Nikon).

3.11 Dot Blot

Dot blot analysis of DNA was undertaken as previously described.[211] Briefly, DNA was denatured in 200µM NaOH with 20µM EDTA and 2 x saline sodium citrate (SSC) at 1:1:2 (DNA:NaOH:2 x SSC), and left at room temperature for 15 minutes. DNA was blotted on N+ charged bond paper in duplicates of 2µL (50ng DNA) and 1µL (25ng DNA). After

drying, DNA was crosslinked at 1200 Joules (ultraviolet (UV) Stratlinker) and blocked in 5% milk for 30 minutes. The primary antibody (5hmc) was incubated on the dot blot overnight at 4°C. The blot was then washed in TBS-T x 3 for 5 minutes on a rocker. The secondary antibody was applied for 30 minutes and re-washed in TBS-T x 3 for 5 minutes on a rocker. Electrochemiluminescence (ECL) solution was added to the blot and left for 3 - 5 minutes before analysis on the ChemiDoc (Bio-Rad). Dot blot images were transferred to the computer and the intensities were analysed using ImageJ. To ensure similar quantities of DNA were blotted for each sample, a separate dot blot was completed. After UV crosslinking, the dot blot was washed in 100% ethanol for 5 minutes on a rocker, prior to multiple washes in tap water. Methylene blue 0.2% in 0.4M sodium acetate and 0.4M acetic acid was added to the blotting paper and incubated for 20 minutes prior to multiple washes in tap water to remove excess stain. The paper was then read as a colorimetric image using a ChemiDoc (Bio-Rad) and images analysed in ImageJ.

3.12 Statistical Analysis

3.12.1 Overview

Kaplan Meier and Cox proportional hazards models were constructed and model assumptions verified, as described.[221,222] To assesses a difference in proportions, Chi Squared test was performed. All statistical analysis was undertaken using Stata/IC 16.1, MedCalc Statistical Software version 19.4.1 (MedCalc Software Ltd, Ostend, Belgium) and R Studio.

Ordinary one-way ANOVA[223] was used to compare the differences in bioinformatic scores for each group: benign, possibly pathogenic and definitely pathogenic using GraphPad Prism 10 for macOS (version 10.2.2, March 19 2024). ROC curves were generated using

GraphPad Prism 10 and the Youden index calculated in Microsoft Excel for Mac (version 16.85). Comparisons between ROC curves, positive predictive value (PPV), negative predictive value (NPV), sensitivity and specificity were analysed using the "DTComPair" package (version 1.2.4) in R Studio (version 4.2.0 "Vigorous Calisthenics"). For all samples, ANOVA with a post hoc Bonferroni correction was performed to control for multiple testing. Samples that were not normally distributed were analysed with non-parametric tests.

3.12.2 Parametric vs non-parametric testing in *MEN1* SNV bioinformatic tools

For *MEN1* SNV bioinformatic tools, the distributions did not meet the Kolmogorov-Smirnov (K-S)[224,225] statistical test for the normal distribution assumption ($p<0.0001$), however, at larger sample sizes (>300), this test is not an accurate test for normal distributions and will inappropriately reject the null hypothesis (i.e., that the sample is normally distributed; type 1 error).[226] Other tests for normal distributions include an 'eye-ball' assessment of the samples' distribution with a histogram and formally assessing the distribution's skewness and kurtosis, with an absolute skewness ≤ 2 and an absolute excess kurtosis ≤ 4 suggesting that the data is normally distributed in larger sample sizes (>300).[226] The skewness of the data refers to whether there is a long tail on one side of the histogram compared to the normal distribution (i.e., is there a lack of symmetry in the distribution). Negative skewness indicates a long tail on the left-hand side.[227] Kurtosis is a marker of the length (or heaviness) of the tails compared to the normal distribution; if the tails of the distribution are less than that of the normal distribution, kurtosis will be negative, whereas if the tails are longer (i.e. more outliers), kurtosis will be positive.[228] Excess Kurtosis is defined as the distribution's kurtosis -3 (the kurtosis value for the normal distribution).[226]

3.13 DNA methylome analysis using R Studio

3.13.1 MethylationEPIC array using R Studio

Methylation data were loaded into R Studio (version 4.2.0 "Vigorous Calisthenics") as 'beta' values and normalised using Beta Mixture Quantile dilation (BMIQ) method, using the champ.load function in 'ChAMP' pipeline as previously described.[212,229] Data were analysed twice, initially as beta values (more biologically significant, although less statistically robust) and then subsequently underwent a log base 2 logarithmic transformation to give 'M' values, as differential methylation analysis performed using 'M' values. (more statistically valid, although results may not be as biologically important due to small effect sizes). The dataset was assessed for confounding (batch and array effects, block age, patient sex and age) using Singular Value Decomposition Analysis (SVD) analysis. 5hmC CpGs sites were determined using differentially hydroxymethylated probes (DhMP) in the R ChAMP pipeline.[212] Gene set enrichment analysis (GSEA) was performed using the 'GOmeth'[230] and 'GOregion'[231] pipelines, to account for the 'multi-gene' bias (1 CpG site annotated to >1 gene), for the different number of CpGs interrogating each gene and for gene length.[230,231] The false discovery rate was set at 0.05, beta cutoff value of 0.05 and M value of 0.1 unless otherwise stated. The quality control values for each sample uploaded into R is shown in **Supplementary Appendix 8.1, Table S1.**

3.13.2 MethylationEPIC array data compared to published data

Tumour suppressor genes reported in humans to cause disease were downloaded from the TSGene (Tumour Suppresor Gene Database) on the 9th October 2024.[232] Methylation data for normal β and α cells were downloaded from the supplementary material freely available by Loyfer et al.[213] The normal islet subtype methylation data was uploaded into R, and merged with the methylationEPIC array data based on hg19 genome co-ordinates and gene ID.

3.14 RNA-seq analysis

3.14.1 RNA-seq analysis of BON1 cell lines

FASTQ files were available for RNA-seq analysis of BON1 cell line (3 replicates).

WTCHG_195872_228_1.fastq.gz

WTCHG_195872_228_2.fastq.gz

WTCHG_195872_228_1.fastq.gz

WTCHG_195872_229_2.fastq.gz

WTCHG_195872_230_1.fastq.gz

WTCHG_195872_230_2.fastq.gz

FASTQ files were uploaded to Galaxy Server (https://usegalaxy.org/) on the 26th of April 2023. I performed all RNA-seq analysis using a combination of bioinformatic tools available on the galaxy platform (FastQC, MultiQC, Cutadapt, HISAT2, feature counts and column join). The generated feature count table was downloaded from Galaxy and uploaded into R Studio. Gene ID names were obtained from g:Profiler (https://biit.cs.ut.ee/gprofiler/convert) on the 23rd February 2024, downloaded and uploaded into R Studio. BON1 feature count tables and Gene ID names, were merged in R Studio. Differential gene expression was analysed using DESeq2 package in R Studio.

3.15 Mouse studies

3.15.1 Ethics and husbandry

Animals were kept in the accordance with UK Home Office welfare guidelines. Regulated procedures were carried out under Procedure Project Licence (PPL) number 30/2383 and my personal licence (PIL) number I84117966. Mice were housed at the Biomedical Services Building, John Radcliffe Hospital (Oxford), in a pathogen-free facility, had free access to water and a standard diet. Breeding cages were set up and included two females and one male in each cage. Mice were monitored and cared for by Animal Technicians daily, who in addition weaned the mice and housed them into separate cages. Earclipping for genotyping and mouse identification from the age of 8 weeks was undertaken by Dr Kate Lines, Dr Omair Shariq and I. As per protocol, aged mice previously involved with breeding were not used in any experimental procedures and were humanely culled by Schedule 1 method. Additionally, any mouse deemed to be suffering was culled humanely (Schedule 1).

3.15.2 Mouse strains

The mice used in our experiments were the *Men1*$^{L/L}$*/RIP2 Cre*-ER mice. Briefly, a LoxP site was present on either side of the mouse *Men1* gene (flanking exons 3-8), additionally these mice had Cre-recombinase under the control of the rat insulin promoter. Therefore, these genetically altered mice lost menin expression specifically in the pancreatic β cell, where Cre-recombinase was expressed. *Men1*$^{L/L}$*/RIP2 Cre*-ER developed insulinomas by the age of 6 months.

3.15.3 MEN1 genotyping

3.15.3.1 DNA extraction from mice ear clips

Men1$^{L/L}$*/RIP2 Cre*-ER mice were earclipped for identification and genotyping **(Figure 17)**, and the earclips stored at -20°C. Earclips were digested DNA extraction buffer for 3h at 56°C, and then at 95°C for 10 minutes. DNA extraction buffer comprised of 10mL of 1M Tis HCL pH 8.5, 0.4mL 0.5M EDTA pH 8, 10mL 10% Tween and made up to 200mL with distilled water. Additionally, 2µL of 25mg/mL of proteinase K was added per 150µL.

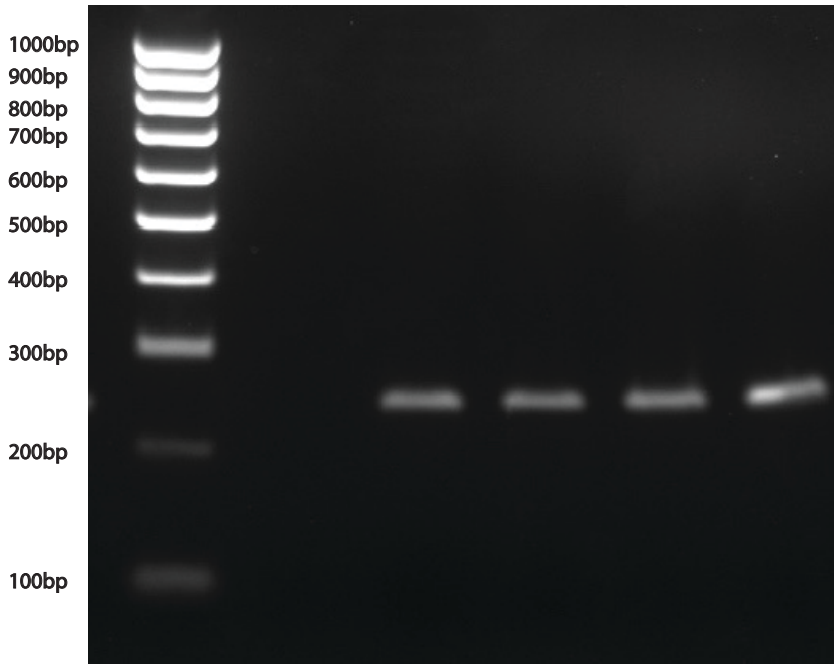


Figure 17: Shows 4 examples of *MEN1* genotyping from 4 Men1$^{L/L}$/RIP2 Cre-ER mice. One band seen at ~236bp is consistent with the mutant copy of *MEN1*. Therefore, these 4 mice are homozygous for the mutant copy of *MEN1*.

3.15.4 *Ex vivo* mouse study

Men1$^{L/L}$*/RIP2 Cre*-ER mice were transferred from the Biomedical Service Building to the pod at the Oxford Centre for Diabetes Endocrinology and Metabolism (OCDEM). Mice were killed by Schedule 1 method by Dr Kate Lines. Mice were pinned to a cork board in a supine position and cleaned from sternum to pelvis with IMS. An incision from sternum

to pelvis was made using scissors and the sternum clamped and retracted superiorly. The bowel was identified and moved to the side to allow adequate visualisation of the pancreas *in situ*. Luciferase (1.5mL) was injected into the main pancreatic duct *in-situ* using a 30-gauge (yellow) needle. if issues arose cannulating the main pancreatic duct, then luciferase was injected throughout the pancreas. Correct cannulation of the pancreatic duct caused the pancreas to visibly inflate. The pancreas was then removed and suspended in 2mL of 0.2% bovine serum albumin (BSA) Hanks Balanced Salt Solution (1g BSA per every 500mL of Hanks solution) and placed on ice. Up to four mice were included in each *ex vivo* study. All tubes were transferred to a water bath set at 37°C for 13 minutes with intermittent shaking of each tube every 2 minutes. Each tube was gently shaken 65 times via inversion. 0.2% BSA Hanks solution (5mL) was then added to each tube and the top layer was removed x 3 (this contains exocrine tissue, whereas the islets/PNETs are sitting at the bottom of the tube). The 0.2% BSA Hanks solution was then removed, and 10mL of Ficol plaque was added to the islet clump at the bottom of the tube and shaken to fully re-suspend the tissue. Each tube was then centrifuged for 20 minutes at 1200rpm. After centrifuging, the pancreatic islets/PNETs were seen in the supernatant **(Figure 18)**.



Figure 18: Pancreatic neuroendocrine tumours (insulinomas) from β cells in one *Men1*$^{L/L}$*/RIP2 Cre*-ER mouse, suspended in the top layer of Ficol plaque (yellow).

The supernatant containing PNETs was removed and pipetted into 25mL of 0.2% BSA Hanks buffer, and the new 50mL tube was centrifuged at 1500rpm for 4 minutes. The supernatant was removed and discarded, with the PNETs in a pellet at the bottom of the tube. The pellet was washed again with 10mL of 0.2% BSA Hanks Solution and centrifuged at 1200rpm for 3 minutes. The supernatant was removed, and the islets were re-suspended in the appropriate media (~10mL; RPMI media with 10% FBS, with 0.5mL of Pen-Strep (10,000 Units penicillin and 10mg/mL of streptomycin; Signa-Aldrich) and 90mg of glucose to the 50mL tube of media (10mM glucose concentration). The tube containing PNETs suspended in media was then placed in a clear plastic dish and placed under a light microscope **(Figure 19)**.

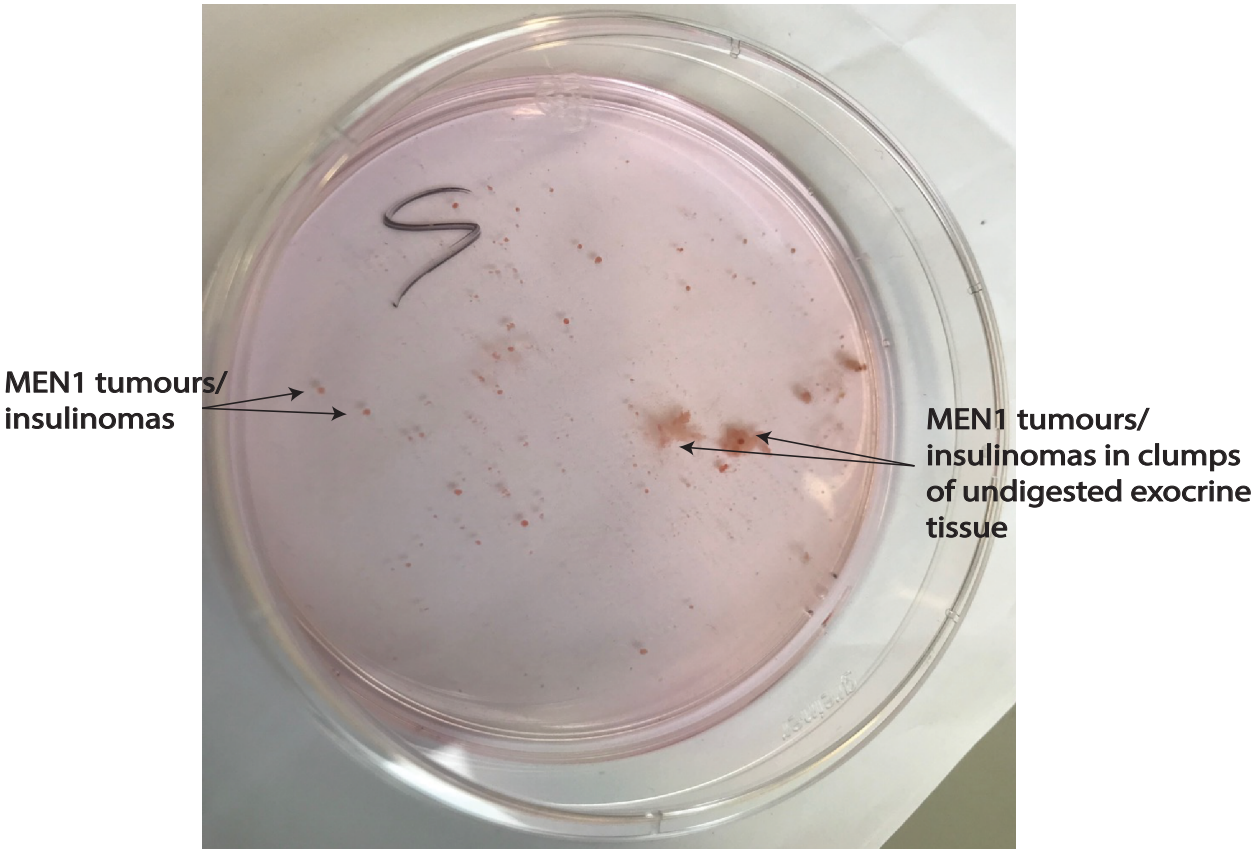


Figure 19: Shows and example of insulinomas suspended in media from one 8-month old *Men1*$^{L/L}$*/RIP2 Cre*-ER mouse. Most islet cell tumours (insulinomas) have had the surrounded exocrine tissue digested, however some islets are still trapped within exocrine tissue, as shown above.

Islet cell tumours were then extracted using a P200 micropipette under an inverted light microscope, and seeded into 96 well plates, with the aim to have the same number and size of islets in each well for each replicate. Each well contained 50µL of media and islets, and a further 50µL of media containing AA was added to each well. Each replicate contained a control, vehicle, 100µM AA and 500µM AA. Cell viability assays were performed using CellTiter-Blue as previously described at days 0, 3 and 5. Additionally, DNA extraction was attempted from the *ex vivo* islets, however, there was insufficient samples to obtain enough DNA for dot blot examination of DNA 5hmC.

4. Results:

4.1 Chapter 1: To determine the most accurate bioinformatics tool to classify *MEN1* missense variants

4.1.1 Introduction

The most common somatic genetic mutation found in sporadic PNETs is in the TSG Multiple Endocrine Neoplasia type 1, now referred to as Menin 1 *(MEN1)*, and is found in ~40% of cases.[55,72] It has been reported that loss of menin in endocrine tumours is associated with increased DNA 5mC.[75,92,93] Therefore, menin plays a role in modulating the DNA methylome. Patients with the hereditary tumour syndrome MEN1, develop multifocal PNETs in up to 80%.[20] Therefore, this gene is important in PNET pathogenesis. *MEN1* is highly conserved with only two common non-synonymous protein-coding missense variants reported in gnomAD with a MAF >1% (Thr541Ala (minor allele frequency (MAF) of 97%) and Arg171Gln (MAF of 1.76%)). Importantly, only one of these variants, Thr541Ala, meets the criteria for a standalone benign variant (>5%) using ACMG criteria.[233] There are other rare *MEN1* missense variants that may be pathogenic, and out of 1570 *MEN1* mutations reported in Clinvar, ~73% are due to missense variants, with the majority (~80%) classified as a variant of uncertain significance (VUS) and only ~14% as pathogenic/likely pathogenic (PV/LPV). Most *MEN1* mutations reported to cause MEN1 have a case finding/ascertainment bias, as most patients initially present with disease suggestive of MEN1 (or have disease present in a family member) and proceed to genetic testing; therefore, the prevalence of 'asymptomatic' *MEN1* missense mutations in the general population is unknown. With the increasing availability of gene sequencing, more robust techniques are needed to determine the pathogenicity of the *MEN1* missense variants, as this will determine appropriate tumour surveillance and genetic counselling for individuals and their families. There are many variant effect predictors (VEPs) that have been generated to try to improve missense variant classification **(Methods 3.1, Table 2)**. One of the most widely used VEPs is the metapredictor, REVEL (Rare Exome Variant Ensemble Learner), which combines 13 *in silico* (e.g. computational/bioinformatic) tools (MutPred2,[192] fathmm v2.3,[193] VEST 3.0,[194] PolyPhen-2,[195] SIFT,[196] PROVEAN,[197] MutationAssessor,[198]

MutationTaster,[199] LRT,[200] GERP++,[201] SiPhy,[202] phyloP,[203] and phaseCons)[204] to give each missense variant a score between 0 (benign) and 1 (pathogenic).[190] The REVEL score cut-off to determine variant pathogenicity depends on the gene; a gene-specific REVEL score cut-off for *MEN1* has not been reported.[234] In general, VEPs use diverse computational methods, to determine what effect a SNV may have on protein function and stability. These may include analysing evolutionary sequence conservation across species, phylogenetic relationships, protein-protein interactions and thermostability changes.[235] Additionally, it has been shown that combining scores improves the overall efficacy of VEPs.[236] Thermostability is an important marker of protein stability and function, and multiple computational methods have been reported to classify SNVs based on changes in this parameter, including FoldX,[237] Rosetta,[207,208] and PremPS.[191] Thermostability is defined as the change in Gibbs free energy from the denatured (unfolded) state to its native (folded) state and is reported as the delta G (ΔG). In general, proteins like to be in a state of low energy (i.e. folded/native state); therefore, ΔG will be negative, and the lower this number, the more stable the protein.

ΔG = Gibbs free energy of the folded protein – Gibbs free energy of the unfolded protein

Missense mutations in proteins will lead to a different conformational state of the folded protein and, subsequently, a different ΔG. The differences in these two ΔG's are defined as the ΔΔG.

$$\Delta\Delta G = \Delta G \text{ (mutant protein)} - \Delta G \text{ (wild-type protein)}$$

If a missense mutation results in abnormal (or unstable) protein folding, the ΔG of the mutant protein will be higher than that of the normal protein, and therefore, the ΔΔG will be positive.[191] Whereas for stabilising missense mutations, the ΔΔG will be negative. Protein thermostability has also been reported to discriminate between benign and pathogenic *MEN1* missense mutations using FoldX.[238] Other VEPs include ESCOTT,[205] CPT and SPDV. Importantly, a benign variant (BV) dataset is needed to compare the accuracy of VEPs to distinguish between benign and pathogenic variants, and studies have used variants reported in gnomAD (and not known to cause disease) as a 'putatively benign' dataset.[236] Therefore, the same methodology has been used to study VEP accuracy in *MEN1* variants, with variants classified as either BV, VUS or LPV.

Most *MEN1* SNVs are reported as VUS, which makes it difficult for clinicians to determine whether patients should enter intensive tumour surveillance screening programs. Therefore, this chapter aims first to: 1) compare the accuracy of different bioinformatic tools to score *MEN1* SNVs from affected MEN1 patients from the literature, 2) try and improve upon current *MEN1* VEPs by generating a new *MEN1* model, 3) review the phenotypes of patients with *MEN1* SNVs using the Geisinger cohort (**Methods 3.1.2**), which is a database of ~175,000 patients sequenced genomes with matched phenotypic information and 4) validate *MEN1* VEPs using this cohort.

4.1.2 Results

4.1.2.1 MEN1 variants reported in the literature

MEN1 variants reported to cause MEN1 were collected from publicly available online databases, ClinVar (archive of disease-causing genetic variants), GnomAD (genetic variants reported in humans), UMD-MEN1 (Universal Mutation Database for MEN1) and LOVD (Leiden Open Variation Database). All papers reported in these databases were reviewed. Variants were also reviewed from two review papers looking at *MEN1* SNVs causing MEN1.[66,186] PubMed was searched using MeSH terms "MEN1", "Multiple Endocrine Neoplasia Type 1", "Missense mutation", "insulinoma", "PNET" for new papers reported after these publications **(Methods 3.1.1.1)**. Reference lists from all published papers were reviewed. *MEN1* SNVs were only included if there was clear phenotypic information reported in the paper to confirm that this variant was associated with MEN1. *MEN1* SNVs reported in online databases to be pathogenic with no phenotypic information were excluded.

4.1.2.1.1 *MEN1* variants reported to cause neuroendocrine tumours

There were 319 unique *MEN1* SNVs reported in the literature that were associated with NETs. One hundred and forty-seven variants were excluded, 37 variants were reported in somatic NETs (i.e. DNA from tumours; NF-PNETs, functional PNETs, NF and functioning pituitary adenomas, PHPT, lung carcinoid and adrenal adenoma), 83 were reported in the UMD-MEN1 database (with no attached detailed clinical information) and 7 variants were associated with a germline *MEN1* variants who displayed a non-MEN1 clinical phenotype. One hundred and seventy-two unique *MEN1* SNVs were reported to cause MEN1 (≥2 PHPT, PNET or pituitary adenoma in an index patient) **(Figure 20)** and were classified according to the ACMG guidelines, with 74 reported as PV/LPV **(Table 17)** and 98 as VUS **(Table 18)**. If the same *MEN1* variant was reported in a patient with a somatic *MEN1*

mutation in a sporadic NET and also in a patient with a germline *MEN1* variant causing MEN1, this *MEN1* variant was included for further analysis.

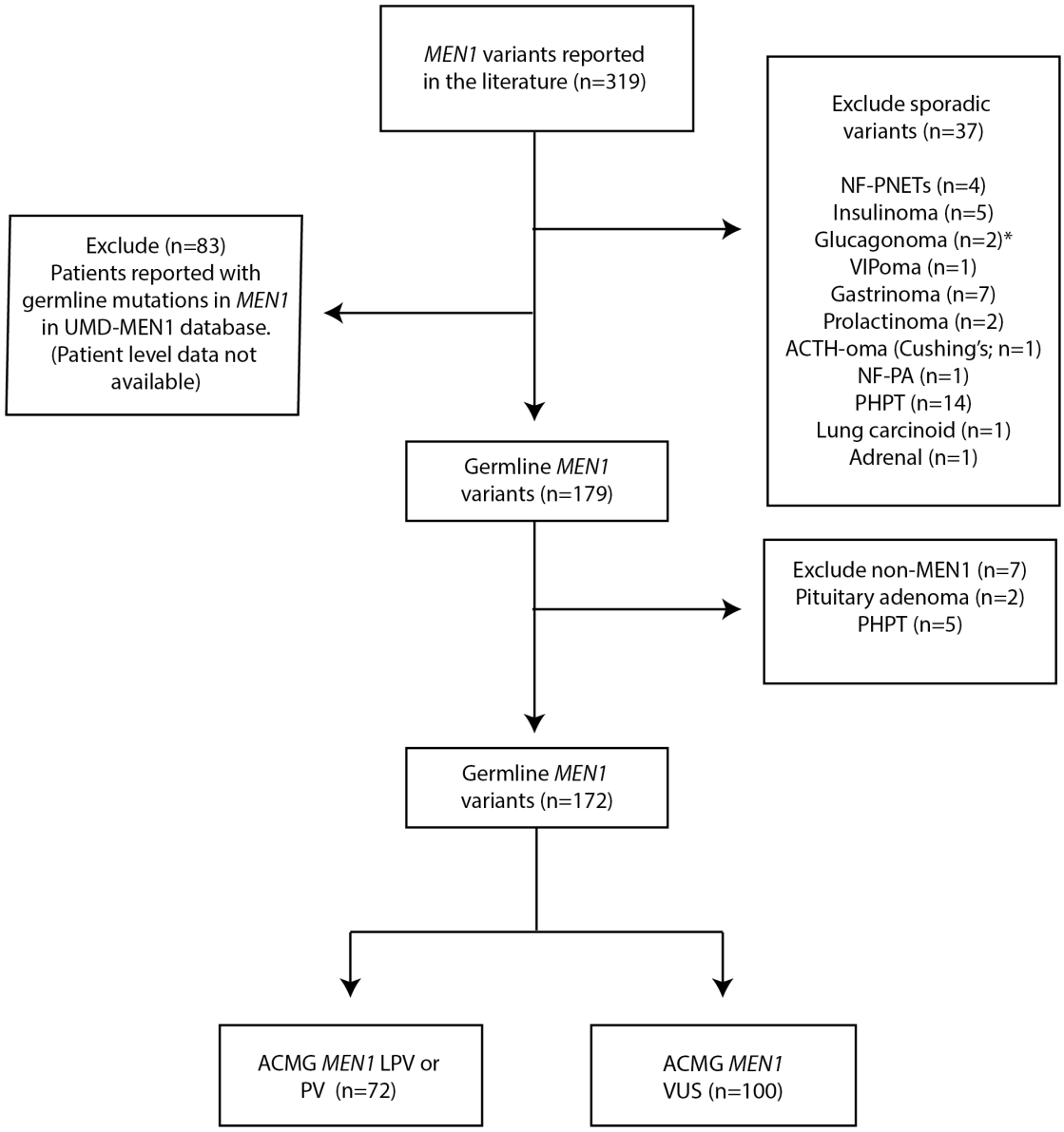


Figure 20: Flow diagram of *MEN1* missense variants reported in the literature.

ACMG: American College of Medical Genetics; ACTH: adrenocorticotrophin releasing hormone; LPV: Likely Pathogenic Variant; NF-PA: non-functioning pituitary adenoma; NF-PNETs: non-functioning pancreatic neuroendocrine tumours; PV: Pathogenic Variant; PHPT: primary hyperparathyroidism; VIPoma: Vasoactive intestinal polypeptide tumour and VUS: Variant of uncertain significance.

*** The same *MEN1* SNV was reported in a somatic glucagonoma and PHPT, and in a somatic glucagonoma and NF-PNET.**

Table 17: All variants classified as LPV or PV by ACMG criteria

Variant 615AA	GRCh38	ClinVar	Domains	Features	PHRED	REVEL	PremPS	SPDV	Rosetta	CPT	ESCOTT	PP1	PP3	PP5	PM1	PM2	PM (PS4)	PM5	Number of people	Number of papers
L22R	64810045	PV/LPV	NTD	NA	32	0.914	2.48	5.9601146	6.12857143	0.43691105	-4.3425646		1	1		1	1		3	3
E26K	64810034	PV/LPV	NTD	NA	23.9	0.826	0.67	5.9601146	4.87494331	0.41917643	-3.8140824	1		1		1	1		9	7
G42R	64809986	LPV	NTD	NA	33	0.954	0.82	4.51793708	8.06609977	0.65282569	-5.456203		1	1		1	1	1	7	2
G42S	64809986	Conflicting	NTD	NA	33	0.935	0.81	4.51793708	5.51122449	0.51813118	-5.456203		1			1	1	1	3	2
G42V	64809985	LPV	NTD	NA	33	0.937	-0.13	4.51793708	6.04727891	0.59962932	-5.456203		1	1		1	1	1	2	2
G42A	64809985	PV	NTD	NA	32	0.92	0.71	4.51793708	3.30816327	0.4759802	-5.456203		1	1		1	0	1	1	1
G42D	64809985	NA	NTD	NA	NA	0.911	1.24	4.51793708	6.61836735	0.63609	-5.456203	1	1			1	1		8	2
E45K	64809977	PV	NTD	NA	28.8	0.895	0.67	4.7067392	2.46020408	0.42876773	-5.2068036		1	1		1	1		10	6
E45Q	64809977	PV	NTD	NA	26.8	0.878	0.88	4.7067392	-0.7316327	0.41524707	-3.1394315		1	1		1	1		8	2
E45G	64809976	PV	NTD	NA	32	0.938	0.43	4.7067392	2.68163265	0.45527673	-5.2068036		1	1		1	1		6	6
H139Y	64809695	PV	Thumb	MLL	28	0.953	0.02	5.47781737	0.40918367	0.33421645	-3.0180106		1	1	1	1	0	1	1	1
H139N	64809695	NA	Thumb	MLL	28.2	0.867	0.84	5.47781737	1.51054422	0.44301208	-5.0135086		1		1	1	1		2	1
H139D	64809695	PV	Thumb	MLL	29.3	0.936	1.33	5.47781737	1.98503401	0.46025343	-5.0135086		1	1	1	1	1		4	4
H139P	64809694	LPV	Thumb	MLL	28.7	0.862	0.98	5.47781737	8.42006803	0.53839882	-5.0135086		1	1	1	1	0	1	1	1
H139R	64809694	PV	Thumb	MLL	27.2	0.855	0.95	5.47781737	1.71938776	0.47912518	-4.7984944	1	1	1	1	1	0	1	1	1
F144V	64809680	Conflicting	Thumb	NA	25.8	0.973	1.26	4.51793708	2.55748299	0.32024525	-3.3572126	1	1			1	1		6	3
F144C	64809679	LPV	Thumb	NA	32	0.96	1.36	4.51793708	3.4329932	0.25884259	-5.3333867		1	1		1	0	1	2	1
D158E	64808086	NA	Thumb	NA	24.5	0.94	0.53	4.68384722	4.17653061	0.51898056	-3.2199897		1			1	1		9	2
G161R	64808079	PV/LPV	Thumb	MLL	25.2	0.932	0.34	3.70166599	9.63537415	0.76777949	-5.7338871		1	1	1	1	0	1	3	1
G161S	64808079	Conflicting	Thumb	MLL	24.9	0.972	0.46	3.70166599	1.89138322	0.53177539	-4.9131853		1		1	1	0	1	4	1
G161V	64808078	NA	Thumb	MLL	24.6	0.926	0.13	3.70166599	6.13537415	0.586873	-5.7338871		1		1	1	0	1	2	2
G161D	64808078	PV/LPV	Thumb	MLL	24.5	0.941	0.82	3.70166599	5.69421769	0.68123301	-4.9897296	1	1	1	1	1	1		4	4
A165T	64808067	NA	Thumb	NA	25.6	0.936	0.46	5.68356931	0.82131519	0.41277246	-1.4804535		1			1	1	1	2	2
A165P	64808067	PV	Thumb	NA	25.6	0.954	0.68	5.68356931	12.8020408	0.75422513	-5.4816706	1	1	1		1	1		3	3
A169P	64808055	NA	Thumb	NA	25.9	0.925	0.73	3.72504422	14.1891157	0.77561671	-4.7621297	1	1			1	0	1	2	1

Variant 615AA	GRCh38	ClinVar	Domains	Features	PHRED	REVEL	PremPS	SPDV	Rosetta	CPT	ESCOTT	PP1	PP3	PP5	PM1	PM2	PM (PS4)	PM5	Number of people	Number of papers
A169D	64808054	LPV	Thumb	NA	27.2	0.971	1.22	3.72504422	6.24784581	0.85075122	-4.7621297	1	1	1		1	1		9	4
C170R	64808052	NA	Thumb	NA	27.9	0.959	1.54	3.72504422	9.20725624	0.83698547	-5.1395869		1			1	1		6	4
C170Y	64808051	LPV	Thumb	NA	25.9	0.953	0.41	3.72504422	5.06428571	0.75587946	-3.7685118		1	1		1	1		2	1
L173P	64808042	PV/LPV	Thumb	NA	28.3	0.956	2.14	5.3136143	7.5537415	0.78188124	-4.4746687	1	1	1		1	1		23	6
D177Y	64808031	PV/LPV	Thumb	NA	32	0.943	-0.11	8.04135844	1.6670068	0.60032504	-4.0699141		1	1		1	1		7	7
L180P	64808021	NA	Thumb	MLL	28.8	0.965	2.87	7.01319915	6.71292517	0.80161835	-5.8079019	0	1		1	1	1		2	1
A181P	64808019	PV/LPV	Thumb	MLL	26.5	0.942	1.25	3.78519903	5.38072562	0.73961009	-5.433204		1	1	1	1	0		1	1
E184K	64808010	LPV	Thumb	MLL	31	0.937	0.84	5.4745069	4.78945578	0.576586	-5.3616764		1	1	1	1	1		2	2
E184Q	64808010	NA	Thumb	MLL	27.8	0.924	0.81	5.4745069	0.88718821	0.51783239	-3.5881441		1		1	1	0	1	1	1
E184D	64808008	VUS	Thumb	MLL	23.2	0.893	1.15	5.4745069	1.10079365	0.52148912	-3.5881441		1		1	1	1		2	2
H186R	64808003	PV/LPV	Thumb	NA	26.3	0.93	1.02	5.55664144	1.51609977	0.4568298	-5.4799392		1	1		1	1		7	3
W188R	64807998	NA	Thumb	MLL	32	0.929	2.06	5.55664144	6.04591837	0.81976482	-6.0271532		1		1	1	1		2	2
W188S	64807997	NA	Thumb	MLL	32	0.903	2.08	5.55664144	6.10147392	0.78565288	-6.0271532	1	1		1	1	1		6	4
W188C	64807996	PV	Thumb	MLL	34	0.923	1.64	5.55664144	5.238322	0.66071504	-6.0271532		1	1	1	1	1		5	5
V189E	64807994	PV	Thumb	NA	32	0.933	2.09	5.33733099	5.62301587	0.85521191	-3.5743764		1	1		1	1		2	1
W225R	64807677	LPV	Thumb	NA	36	0.921	2.19	6.09373603	5.14013605	0.67111529	-3.5231895		1	1		1	1		6	3
L228P	64807667	PV	Thumb	NA	33	0.941	2.46	5.45967483	4.47369615	0.6674272	-4.803716		1	1		1	1		10	5
G230R	64807662	LPV	Thumb	NA	32	0.942	0.85	5.45967483	2.17414966	0.60391305	-3.3205246	0	1	1		1	1		3	3
E240K	64807632	PV	Palm	NA	27.3	0.948	0.79	10.4282441	5.0255102	0.52362095	-5.2645295		1	1		1	1		2	2
C246R	64807614	LPV	Palm	MLL	25.1	0.881	1.04	9.2355282	8.71893424	0.4772558	-5.2438904		1	1	1	1	1		6	4
C246Y	64807613	PV/LPV	Palm	MLL	26	0.918	0.38	9.2355282	6.04671202	0.42651957	-5.0184235	0	1	1	1	1	1		4	3
C246F	64807613	NA	Palm	MLL	26.1	0.92	0.1	9.2355282	4.41439909	0.4745394	-5.0184235		1		1	1	1		10	3
A247V*	64807610	PV/LPV	Palm	MLL	28.7	0.96	0.51	4.7082239	2.81870748	0.37288409	-5.0928131		1	1	1	1	1		2	2
L261F	64807569	PV/LPV	Palm	NA	25.5	0.898	0.67	6.40868157	7.12823129	0.42023246	-4.4870996		1	1		1	1		8	2
R280K*	64807179	PV	Palm	NA	NA	0.802	-0.11	8.00568221	0.78412698	0.11549395	-0.1966469		1	1		1	1		12	2
G286R	64807082	LPV	Palm	MLL	28.4	0.967	1.05	4.54971061	14.1765306	0.64263586	-5.2735503		1	1	1	1	1		7	3

Variant 615AA	GRCh38	ClinVar	Domains	Features	PHRED	REVEL	PremPS	SPDV	Rosetta	CPT	ESCOTT	PP1	PP3	PP5	PM1	PM2	PM (PS4)	PM5	Number of people	Number of papers
G286E	64807081	NA	Palm	MLL	27	0.965	1.41	4.54971061	4.46632653	0.69473071	-4.5886654		1		1	1	0	1	NA	1
A289E	64807072	LPV	Palm	MLL	28	0.967	1.54	4.54971061	14.4	0.59111492	-4.3053844	1	1	1	1	1	1		4	2
H322R	64806331	LPV	Palm	NA	26.3	0.947	1.36	6.22065227	1.47380952	0.44491401	-4.6657926		1	1		1	1		3	2
H322Y	64806332	LPV	Palm	NA	26	0.956	0	6.22065227	-1.1929705	0.37776724	-1.8581627		1	1		1	1	1	2	2
P325R	64806322	PV	Palm	MLL	28.7	0.958	1.34	7.66962856	5.49671202	0.65904358	-5.7860273		1	1	1	1	0	1	NA	1
P325L	64806322	PV/LPV	Palm	MLL	29.5	0.959	0.3	7.66962856	-1.5363946	0.54688173	-5.7860273		1	1	1	1	1		NA	2
A342P	64806272	LPV	Palm	NA	26.3	0.952	1.23	4.75037845	12.1153061	0.81646114	-5.4894553		1	1		1	0	1	1	1
A342D	64806271	PV	Palm	NA	26.1	0.948	1.69	4.75037845	5.49852608	0.83680702	-5.4894553	1	1	1		1	1		8	7
W346R	64806260	NA	Palm	NA	26.6	0.935	2.11	6.67701209	3.40181406	0.79657141	-5.9274057	1	1			1	1		8	3
T349R	64806250	PV	Palm	NA	26.9	0.812	1.21	6.67701209	1.71088435	0.52759538	-5.2075828		1	1		1	1		12	4
Y356H	64805769	NA	Palm	NA	32	0.962	1.51	6.73212199	1.81972789	0.45226375	-2.8093119	1	1			1	1		8	2
I365F	64805742	VUS	Palm	NA	26.8	0.883	1.24	5.61352148	3.80997732	0.45666102	-5.3406062	1	1			1	1		3	2
A373D	64805717	PV	Palm	NA	28.6	0.961	1.58	6.19814652	5.14988662	0.66814376	-3.0971523	1	1	1		1	1		2	2
P378S	64805703	LPV	Palm	NA	27.7	0.972	1.33	6.28395246	-1.6880952	0.42266878	-3.2951904		1	1		1	1		NA	2
D423N	64805132	PV	Fingers	NA	27.3	0.891	0.71	4.66157106	1.9590703	0.42407245	-2.9864879	1	1	1		1	1		33	13
D423H	64805132	PV	Fingers	NA	27.2	0.908	0.43	4.66157106	10.4510204	0.44895368	-4.96409		1	1		1	1	1	5	3
W428R	64805117	LPV	Fingers	NA	29.1	0.945	2.12	9.29405331	5.32891157	0.56401339	-5.7081596	0	1	1		1	1		9	5
W428S	64805116	NA	Fingers	NA	28	0.925	2.15	9.29405331	6.25011338	0.56412049	-5.7081596		1			1	1	1	2	2
W441R	64805078	PV	Fingers	NA	28.6	0.934	1.86	11.2311386	5.09818594	0.48201751	-5.6533836		1	1		1	1		4	3
S448Y	64805056	LPV	Fingers	NA	28.4	0.938	0.09	4.46278573	4.19115646	0.33266274	-3.0398893	1	1	1		1	1		6	3
F452S	64805044	LPV	Fingers	NA	28.2	0.927	2.65	6.02517859	5.06961451	0.40519209	-5.4458103		1	1		1	1		3	3
S560N	64804503	PV/LPV	Fingers	NLSa	25.2	0.747	0.84	4.66157106	3.0292517	0.45408772	-4.3788854		1	1		1	1		4	3
S560R	64804502	NA	Fingers	NLSa	23.2	0.759	1.1	4.66157106	5.619161	0.37692086	-4.0814691		1			1	1	1	2	2

*All variants located on the surface of the protein, the remaining are located within the core.

Amino acid (AA); Genome reference consortium human build 38 (GRCh38); Likely pathogenic variant (LPV); Not reported (NA); Mixed lineage leukemia (MLL); Nuclear localisation sequence a (NLSa); Pathogenic variant (PV); Supporting evidence of pathogenicity (PP1-5); Moderate evidence of pathogenicity (PM1,2 and 5); Strong evidence of pathogenicity (PS4); Variant of uncertain significance (VUS).

Table 18: All variants classified as VUS by ACMG criteria

Variant 615AA	GRCh38	ClinVar	Domains	Features	PHRED	REVEL	PremPS	SPDV	Rosetta	CPT	ESCOTT	PP1	PP3	PP5	PM1	PM2	PM (PS4)	PM5	BP4	Number of people	Number of papers
P12L*	64810075	LP	NTD	NA	32	0.892	0.13	10.6932194	2.07721088	0.33300697	-4.0174589	0	1	1		1	0			1	1
P32R	64810015	Con	NTD	NA	32	0.798	1.21	4.99700847	6.87913832	0.38305886	-4.4115226		1			1	1			2	2
S38P	64809998	LP	NTD	NA	32	0.951	1.22	4.19862814	5.71462585	0.55152783	-5.6554146		1	1		1	0			2	1
L39W	64809994	NA	NTD	NA	31	0.973	1.87	5.33118338	7.98129252	0.48942031	-4.9537032	0	1			1	1			4	4
E45A	64809976	NA	NTD	NA	31	0.94	0.44	4.7067392	0.55272109	0.46245214	-5.2068036		1			1	0	1		1	1
H46P*	64809973	VUS	NTD	NA	28.1	0.894	-0.44	6.14069043	14.3238095	0.45174544	-4.1970902	1	1			1	0			2	1
A49V	64809964	VUS	NTD	NA	29.2	0.836	-0.21	5.79056315	1.78639456	0.28704052	-2.8982688		1			1	0			1	1
R52G*	64809956	VUS	NTD	NA	28.3	0.865	0.31	10.5173708	1.97653061	0.2539486	-2.0447978	0	1			1	0			1	1
R98P*	64809817	NA	NTD	NA	28.9	0.825	0.65	5.94450084	9.47619048	0.45673781	-3.324357		1			1	0			1	1
R98L*	64809817	NA	NTD	NA	24.8	0.817	0.05	5.94450084	1.29489796	0.25224144	-0.8903745					1	0			2	1
G110E*	64809781	Con	Thumb	NA	22.8	0.671	0.71	9.39106736	1.04761905	0.12513826	-0.2009503					1	0		1	NA	1
E116G*	64809763	NA	Thumb	NA	33	0.942	0.38	13.0096501	1.13503401	0.30259146	-1.6433241		1			1	0			2	1
R137W*	64809701	VUS	Thumb	NA	31	0.846	0.43	6.26316854	-0.414966	0.48163193	-3.6874797		1			1	0			2	1
L143H	64809682	NA	Thumb	NA	29.8	0.971	2.7	5.62824167	3.13061225	0.5127503	-4.6802546	1	1			1	0			3	1
F144S	64809679	NA	Thumb	NA	32	0.964	1.77	4.51793708	4.21530612	0.26776065	-5.3333867		1			1	0	1		1	1
F146S	64809673	NA	Thumb	NA	32	0.899	2.43	5.4469511	4.86734694	0.28046239	-4.43849		1			1	0			2	1
D158V*	64808087	P	Thumb	NA	29.5	0.953	-0.21	4.68384722	3.13367347	0.45108874	-5.3431302		1	1		1	0			2	1
S159I*	64808084	VUS	Thumb	NA	25.3	0.848	-0.04	3.86163607	-2.465873	0.61021534	-4.7375982		1			1	1			3	2
S160F*	64808081	VUS	Thumb	NA	21.4	0.687	-1.72	3.70166599	5.26496599	0.17159808	-0.9804931					1	1		1	7	3
V167F	64808061	NA	Thumb	NA	25.4	0.957	1.03	5.08107362	6.48979592	0.67165693	-4.4528761		1			1	1			3	2
V167A	64808060	VUS	Thumb	NA	26.2	0.966	2.08	5.08107362	1.80283447	0.46840133	-2.6310961		1			1	0			1	1
Q171P	64808048	VUS	Thumb	NA	26.6	0.97	0.99	5.90635645	9.4755102	0.77402208	-5.0843137		1			1	0			NA	1
G174S*	64808040	VUS	Thumb	NA	25.9	0.944	0.73	4.66309515	1.07789116	0.55305777	-2.1644364		1			1	0			1	1
R176W*	64808034	on	Thumb	NA	24.8	0.657	0.08	6.89491521	0.29693878	0.22610698	-0.4514957					1	0		0	1	1
V201G	64807958	NA	Thumb	NA	28.6	0.959	2.58	7.84935502	4.36814059	0.74365841	-5.2672349		1			1	0			7	1

Variant 615AA	GRCh38	ClinVar	Domains	Features	PHRED	REVEL	PremPS	SPDV	Rosetta	CPT	ESCOTT	PP1	PP3	PP5	PM1	PM2	PM (PS4)	PM5	BP4	Number of people	Number of papers
H204D	64807950	NA	Thumb	NA	26.5	0.967	1.23	5.75778939	2.90385488	0.65517407	-4.629422					1	0			1	1
V220M*	64807902	P/LP	Thumb	NA	25	0.799	0.67	5.89185416	1.77301587	0.32664339	-0.8952475			1		1	1			7	5
W225G	64807677	NA	Thumb	NA	32	0.937	2.15	6.09373603	6.48129252	0.67692191	-6.0832151		1			1	0	1		2	1
W225S	64807676	NA	Thumb	NA	33	0.901	2.22	6.09373603	5.73537415	0.66176114	-5.8220095		1			1	0	1		NA	1
L228R	64807667	NA	Thumb	NA	31	0.968	2.43	5.45967483	5.65816327	0.58096695	-4.803716		1			1	0	1		1	1
K229T*	64807664	NA	Thumb	NA	33	0.758	0.5	5.13649113	2.44319728	0.31855006	-2.575486					1	0			1	1
S231P	64807659	NA	Thumb	NA	24.6	0.777	0.88	3.53449619	5.18401361	0.35338464	-3.6618771	1				1	0			3	1
R234L*	64807649	NA	Thumb	NA	22.8	0.598	-0.59	7.5863935	0.92664399	0.25164481	-0.8359238					1	1		1	2	2
R234H*	64807649	Con	Thumb	NA	22.9	0.622	0.59	7.5863935	2.57052154	0.30723742	-1.917684					1	0		1	1	1
C235R	64807647	LP	Thumb	NA	28.5	0.959	1.96	8.8650714	7.04376417	0.55274209	-5.3154182		1	1		1	0			2	1
L254P*	64807589	NA	Palm	NA	24.6	0.768	0.31	12.6139385	1.78741497	0.23160665	-1.323824					1	0			5	1
S258P	64807578	NA	Palm	NA	26.5	0.872	1.16	8.61180087	5.37210884	0.43265039	-4.4159906	0	1			1	0			NA	1
S258W	64807577	P	Palm	NA	26.7	0.951	0.68	8.61180087	6.0914966	0.53085199	-4.1151466		1	1		1	0			NA	1
L259R*	64807574	NA	Palm	NA	24.4	0.769	1.53	12.5873916	0.89455782	0.20263958	-2.3446363					1	0			NA	1
E260K	64807572	VUS	Palm	NA	26.1	0.937	0.5	10.0921188	4.43027211	0.49105885	-3.9537539		1			1	0			NA	1
L261P	64807568	NA	Palm	NA	29.4	0.955	2.51	6.40868157	10.2545352	0.56818959	-4.9078122		1			1	0	1		1	1
L264P	64807559	Con	Palm	NA	29.3	0.962	2.29	9.82442608	16.1982993	0.65375665	-4.9528922		1			1	0			NA	1
Q265P	64807556	NA	Palm	NA	28.4	0.962	0.96	6.36634292	7.11485261	0.64107177	-5.3429312	0	1			1	1			1	1
L269P	64807212	P	Palm	NA	31	0.931	2.76	7.21447994	14.2345805	0.7664003	-5.3782385	1	1	1		1	0			NA	1
L278P	64807185	LP	Palm	NA	30	0.96	2.36	8.13561595	8.97528345	0.57437766	-3.2219972		1	1		1	0			6	1
L278R	64807185	LP	Palm	NA	31	0.955	2.32	8.13561595	10.4515873	0.53758274	-2.3573861		1	1		1	0			6	1
E279A*	64807182	NA	Palm	NA	23.6	0.687	-0.12	4.55658503	1.04727891	0.16633823	-0.7982556					1	0		1	NA	1
A284D	64807087	NA	Palm	NA	28.6	0.986	1.61	4.98036847	3.78786848	0.76570112	-5.4208632		1			1	0			1	1
A284V	64807087	NA	Palm	NA	33	0.958	0.66	4.98036847	2.0531746	0.47482317	-4.9529057		1			1	1			3	1
L291P	64807066	NA	Palm	NA	29.6	0.966	2.45	5.5894943	9.97346939	0.73030174	-4.9581684		1			1	0			NA	1
L306R	64807021	NA	Palm	NA	27.7	0.93	2.75	5.39173387	5.88072562	0.5052269	-4.5266713		1			1	0			9	1

Variant 615AA	GRCh38	ClinVar	Domains	Features	PHRED	REVEL	PremPS	SPDV	Rosetta	CPT	ESCOTT	PP1	PP3	PP5	PM1	PM2	PM (PS4)	PM5	BP4	Number of people	Number of papers
G310R	64806368	NA	Palm	NA	33	0.934	0.78	7.60303839	11.7564626	0.43002004	-4.1317013		1			1	0			2	1
A314P	64806356	NA	Palm	NA	27.4	0.924	1.2	5.63827539	9.18208617	0.69214162	-3.8695664		1			1	0			NA	1
T316P*	64806350	NA	Palm	NA	23.4	0.72	0.65	8.00568221	7.82029479	0.31301347	-2.3469221					1	1		0	3	3
R319P*	64806340	NA	Palm	NA	22.4	0.682	0.37	7.68172861	1.24387755	0.26155368	-2.4358727	1				1	1		1	2	2
D320Y*	64806338	Con	Palm	NA	28.7	0.875	0.22	8.68081726	2.26927438	0.29826979	-3.9229724		1			1	1			5	2
A330P	64806308	NA	Palm	NA	27.1	0.91	1.22	5.16474967	9.31043084	0.76982587	-4.8059851	1	1			1	0			NA	1
L343P	64806268	LP	Palm	NA	26.4	0.937	2.3	5.40684002	10.8054422	0.75555715	-4.8979505		1	1		1	0			3	1
A347P	64806257	LP	Palm	NA	24.8	0.955	1.14	5.33964109	11.5943311	0.74274591	-5.0259323		1	1		1	0			NA	1
A350P	64806248	NA	Palm	NA	25.2	0.909	1.2	4.14108026	8.58707483	0.67655256	-4.8201941		1			1	0			1	1
I353N	64806238	NA	Palm	NA	28.1	0.867	2.68	4.7399457	3.57460318	0.7239028	-4.8447305	0	1			1	0			1	1
Y358D	64805763	NA	Palm	NA	29	0.942	1.8	5.70186286	4.14829932	0.6303364	-4.7649061		1			1	1			NA	2
C359F	64805759	VUS	Palm	NA	24	0.572	0.1	8.53592552	2.15827664	0.13211192	-0.866045					1	0		0	4	1
R360W*	64805757	VUS	Palm	NA	33	0.937	0.17	13.129591	0.19637188	0.38700364	-3.9120563		1			1	0			NA	1
D362H	64805751	NA	Palm	NA	28.8	0.908	0.12	5.20206186	8.89444444	0.58666338	-5.5833125		1			1	0			2	1
F369C	64805729	NA	Palm	NA	29.7	0.966	2.39	7.09155643	4.49501134	0.44543396	-5.023548		1			1	0			1	1
F369S	64805729	NA	Palm	NA	30	0.945	2.64	7.09155643	5.67800454	0.60385165	-5.023548	1	1			1	0			2	1
E371D*	64805722	Con	Palm	NA	14.94	0.611	0.03	10.2463327	0.36496599	0.0547711	-0.2235004					1	0		1	3	1
V372I	64805721	Con	Palm	NA	18.54	0.422	-0.76	5.84069485	-0.4159864	0.07678903	-0.3120171					1	0		1	3	1
I377M	64805704	VUS	Palm	NA	23.8	0.837	1.03	7.02760362	1.21133787	0.36071631	-1.475974		1			1	1			2	2
P378A	64805703	Con	Palm	NA	26.4	0.959	1.13	6.28395246	-2.8111111	0.42084361	-3.9549511		1			1	0	1		2	1
L380P	64805696	NA	Palm	NA	28.7	0.91	2.18	9.91587604	11.7547619	0.51992877	-2.4915811		1			1	0			NA	1
L381P	64805693	LP	Palm	NA	29.3	0.94	2.43	6.11925852	11.7901361	0.53541732	-4.4695295		1	1		1	0			1	1
L381R	64805693	VUS	Palm	NA	28.7	0.941	2.49	6.11925852	6.33390023	0.44532886	-4.4695295		1			1	0			NA	1
A390V*	64805666	Con	Palm	NA	24.1	0.608	0.26	17.1774515	1.45748299	0.1698457	-1.2250193					1	1		1	3	2
L418P	64805146	NA	Fingers	NA	29.1	0.944	2.6	7.07100706	9.96893424	0.62836012	-5.0057579		1			1	1			2	2
L419R	64805143	VUS	Fingers	NA	28.7	0.923	1.92	7.0416052	4.37426304	0.56638554	-5.0823991		1			1	0			1	1

Variant 615AA	GRCh38	ClinVar	Domains	Features	PHRED	REVEL	PremPS	SPDV	Rosetta	CPT	ESCOTT	PP1	PP3	PP5	PM1	PM2	PM (PS4)	PM5	BP4	Number of people	Number of papers
L419Q	64805143	NA	Fingers	NA	28.5	0.934	2.71	7.0416052	2.62267574	0.57491016	-5.0823991		1			1	1			3	1
R420P*	64805140	Con	Fingers	NA	28.3	0.912	1.09	6.8689947	6.57120181	0.40067585	-3.7376409		1			1	1			NA	2
G424V	64805128	LP	Fingers	NA	25.8	0.903	0.63	4.14200028	1.51417234	0.48945365	-4.9300057		1	1		1	0			1	1
G424D	64805128	NA	Fingers	NA	25.7	0.912	1.47	4.14200028	2.06519274	0.52879715	-4.2555456		1			1	1			2	2
C426Y	64805122	P/LP	Fingers	NA	26.1	0.952	0.54	5.20456849	4.56723356	0.48048786	-4.84383		1	1		1	0			NA	1
S432R	64805103	LP	Fingers	NA	24.4	0.867	1.02	7.12722545	5.68945578	0.40857984	-4.021278		1	1		1	1			7	2
W441C	64805076	P/LP	Fingers	NA	27.8	0.932	1.43	11.2311386	5.37494331	0.39848588	-5.0895653		1	1		1	0	1		NA	1
L449P	64805053	NA	Fingers	NA	29.2	0.967	2.59	4.46278573	11.6212018	0.52356055	-4.2406786		1			1	1			16	4
F452L	64805043	VUS	Fingers	NA	14.77	0.767	1.47	6.02517859	2.22063492	0.29947154	-4.7400272		1			1	1	1		2	1
G536A	64804575	NA	Fingers	NA	0.246	0.322	-0.29	39.4529519	0.13095238	0.07110102	-0.1240016					1	0		1	NA	1
G537C	64804573	Con	Fingers	NA	14.59	0.67	0.27	40.2522763	4.84693878	0.20562011	-0.529555					1	1		0	4	3
P545S	64804549	Con	Fingers	NA	21.4	0.69	0.33	30.5667239	0.39251701	0.10112655	-0.1383942	0				1	1		0	7	1
P549S	64804537	VUS	Fingers	NA	22.1	0.696	0.46	24.4219585	0.92823129	0.08939171	-0.333587					1	0		0	2	1
K562E	64804498	LP	Fingers	NLSa	28.2	0.879	1.2	5.74678387	4.83095238	0.31460319	-4.1951481		1	1		1	0			NA	1
M563K	64804494	NA	Fingers	NLSa	27.7	0.945	2	5.25313283	5.01439909	0.38462988	-4.2519997		1			1	0			NA	1
M566R	64804485	LP	Fingers	NLSa	28.4	0.92	2.3	8.96284257	3.63117914	0.28193631	-4.1552941		1	1		1	0			1	1
L570Q	64804473	NA	Fingers	NLSa	28.2	0.914	2.44	6.57469227	3.28764172	0.42785791	-4.4036557	1	1			1	0			3	1
I580N	64804443	NA	Fingers	NA	28.8	0.921	2.53	8.7108629	3.37494331	0.41836713	-4.067605		1			1	1			6	2
L584P	64804431	NA	Fingers	NA	28.7	0.979	2.52	7.82693237	11.1911565	0.42455929	-4.2605814		1			1	0			NA	1
S588P	64804420	NA	Fingers	NA	25.1	0.963	0.59	18.4919818	0.95578231	0.29164825	-4.264663	0	1			1	0			1	1
S588L	64804419	NA	Fingers	NA	25.7	0.879	0.61	18.4919818	0.02981859	0.22702643	-4.264663		1			1	0			1	1

*All variants located on the surface of the protein, the remaining are located within the core.

Amino acid (AA); Supporting evidence of benign impact (BP4); Conflicting evidence of pathogenicity (Con); Genome reference consortium human build 38 (GRCh38); Likely pathogenic variant (LPV); Not reported (NA); Mixed lineage leukemia (MLL); Nuclear localisation sequence a (NLSa); Pathogenic variant (PV); Supporting evidence of pathogenicity (PP1-5); Moderate evidence of pathogenicity (PM1,2 and 5); Strong evidence of pathogenicity (PS4); Variant of uncertain significance (VUS).

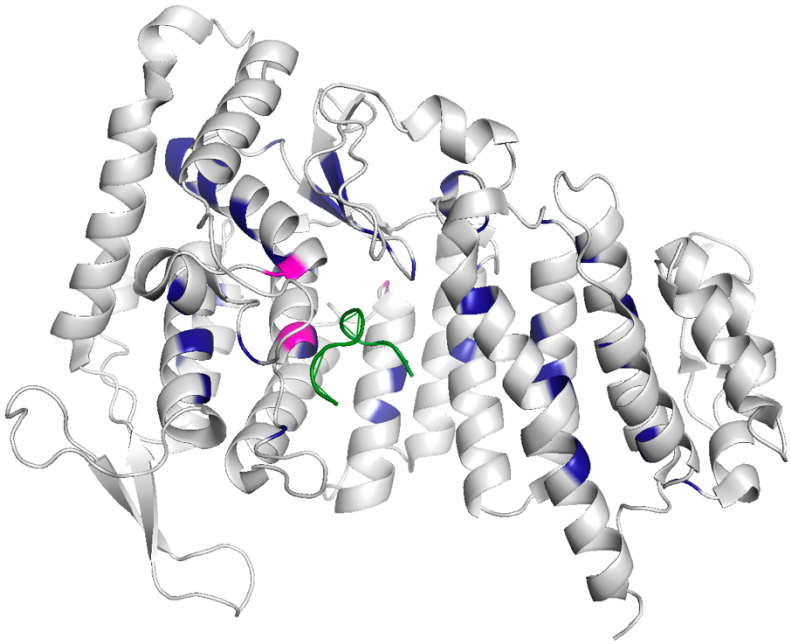
4.1.2.1.2 Benign variant dataset from gnomAD

For the BV dataset, 463 unique *MEN1* SNVs were reported in gnomAD, which did not include any *MEN1* missense variants in the literature associated with NET (germline or sporadic). Out of the 463 variants, ~55.7% (258/463) have been reported in ClinVar, and are classified as follows: benign/likely benign (B/LB; 1.5% (7/463)), VUS (47.5% (220/463)) and conflicting reports of pathogenicity (6.7% (31/463)).

4.1.2.2 Protein structure for menin and locations of benign and pathogenic mutations

Menin is a 610 amino acid (aa; now reported as 615aa) protein. The protein data bank (PDB) structure of menin is reported in its 610aa transcript and does not include a portion of the C terminus from 386-401aa, 460-519aa, 528-547aa and 582-610aa. For consistency, all *MEN1* mutations are reported in the 615aa format. The 74 *MEN1* missense variants classified as LPV/PV, are located throughout the protein, do not cluster around a specific mutational hotspot and are buried within the protein core **(Figure 21; Panel A)**. Whereas, most benign *MEN1* variants, are found on the surface of the protein **(Figure 21; Panel B)**.

Panel A



Panel B

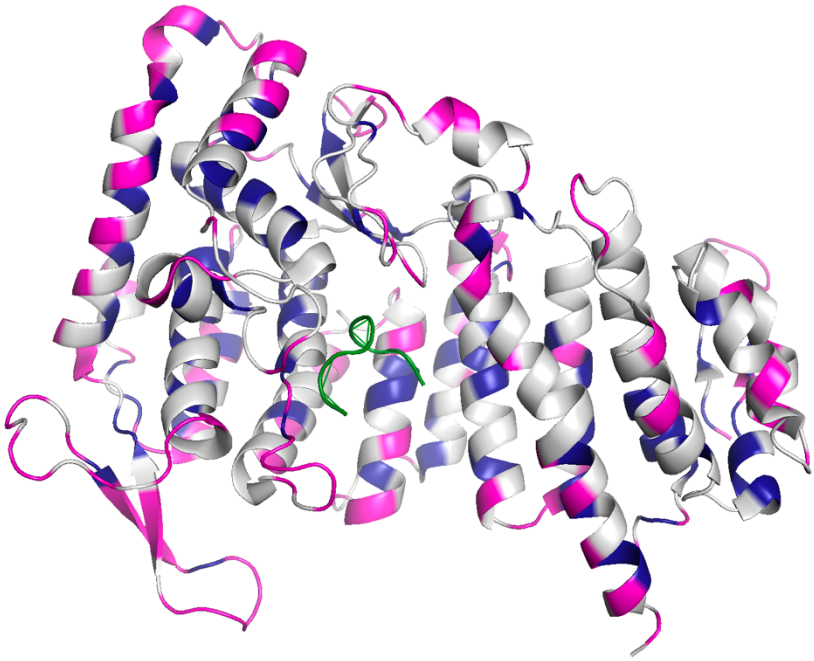


Figure 21: Protein data bank (PDB) structure (3u85) of the menin protein. Mutations classified as LPV by ACMG are highlighted in blue (core) and pink (surface). MLL protein is present in the pocket of the menin protein (green) (Panel A). Benign *MEN1* variants reported in gnomAD. Surface *MEN1* variants (pink). Core *MEN1* variants (blue) and not reported (white/grey).

Pathogenic *MEN1* mutations are significantly enriched within the core of the protein, 95.9% (71/74), compared to the benign/neutral missense mutations, 30% (139/463) p<0.001 **(Figure 22).**

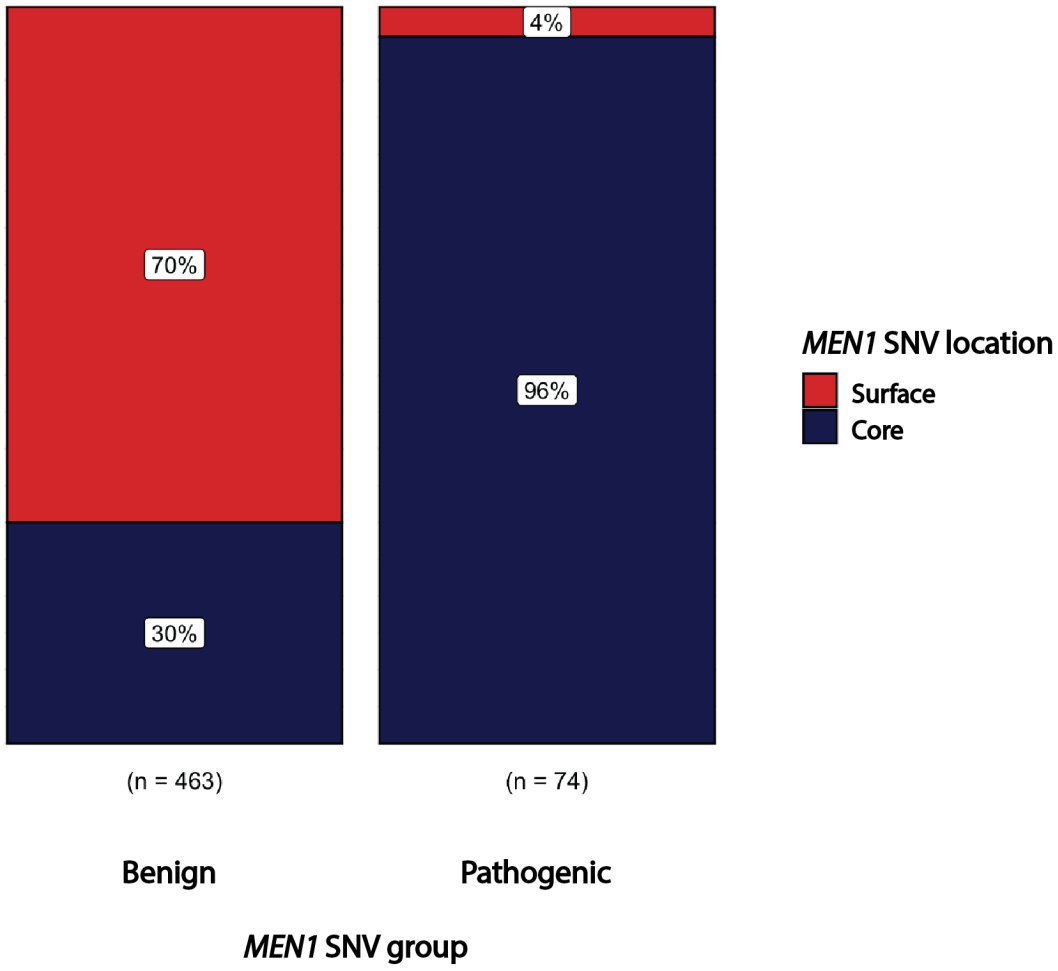


Figure 22: The proportions of the locations (core (Dark blue) and surface (red)) of the *MEN1* missense mutations in both the benign and pathogenic groups.

4.1.2.3 Comparison of different bioinformatic tools using published MEN1 missense variants

Bioinformatic tools were first compared using all *MEN1* missense variants published in gnomAD and *MEN1* variants published in the literature to cause MEN1. All bioinformatic tools were significantly correlated with each other p<0.001 (Correlation matrix; **Figure 23**). There is a strong positive correlation between CADD and REVEL scores (0.82) and a strong

negative correlation between CPT and ESCOTT (-0.86). CADD and REVEL scores use similar bioinformatic tools to construct their overall pathogenicity score, and therefore, are significantly highly correlated.

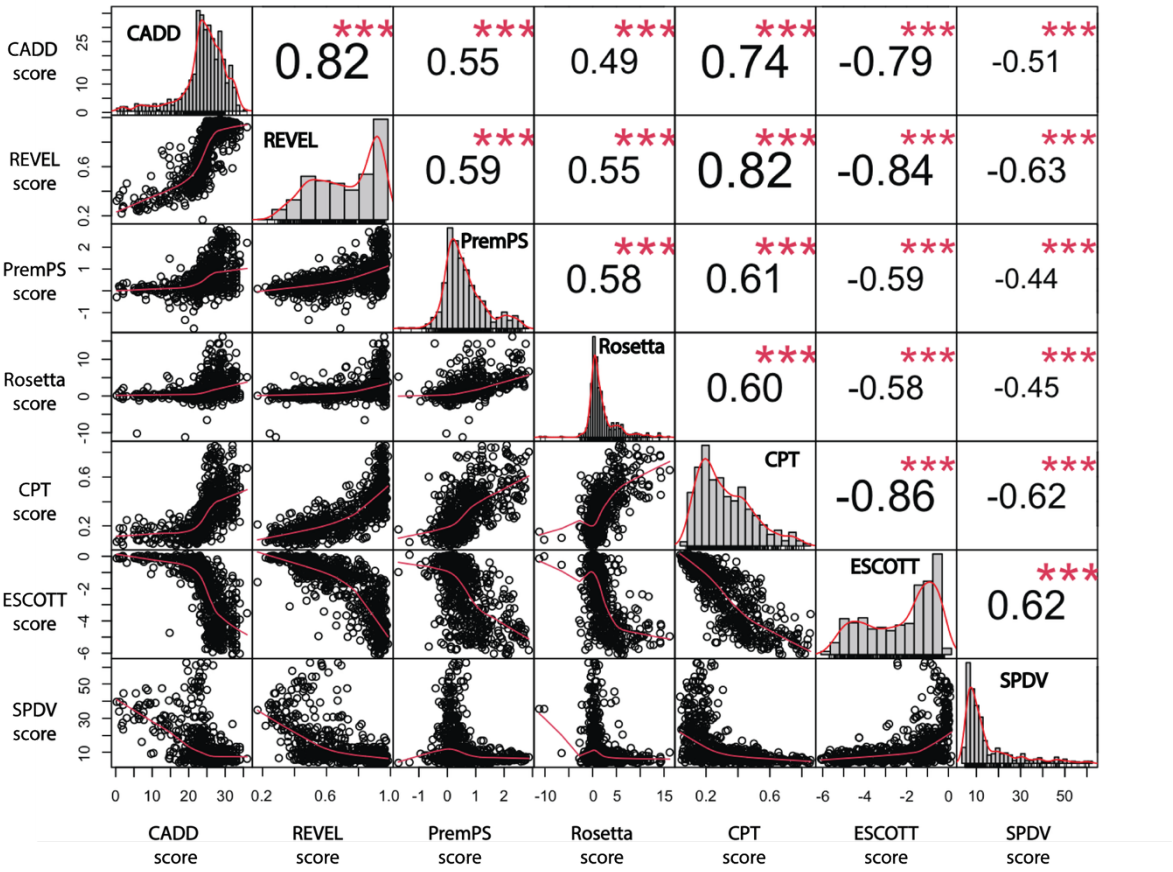


Figure 23: Correlation matrix reporting the correlation coefficients between each bioinformatic tool (CADD, REVEL, PremPS, Rosetta, CPT, ESCOTT and SPDV) ***p<0.001. Each scoring system is scored differently, therefore, the axes values (X and Y) are specific and have different ranges for each bioinformatic tool.

The descriptive statistics for each bioinformatic tool and assessment for normal distribution were undertaken **(Table 19)**. Rosetta and CADD did not meet the assumption for normal distribution (Kurtosis 13.54 (≥4) for Rosetta BV and the CADD VUS group (Skew -2.440 and Kurtosis 10.97) and therefore were analysed by non-parametric methods (Kruskal Wallis).

Table 19: Descriptive analysis for each bioinformatic tool

	Rosetta			CPT			ESCOTT			REVEL			CADD			PremPS		
MEN1 predictors	**BV**	**VUS**	**LPV**	**BV**	**VUS**	**LPV**	**BV**	**VUS**	**LPV**	**BV**	**VUS**	**LPV**	**BV**	**VUS**	**LPV**	**BV**	**VUS**	**LPV**
Minimum	-11.25	-2.811	-1.688	0.0375	0.0548	0.1155	-5.604	-6.083	-6.027	0.1670	0.3220	0.7470	0.9240	0.2460	23.20	-1.330	-1.720	-0.1300
25%[a]	0.0527	1.582	1.966	0.1456	0.2986	0.4435	-2.322	-4.947	-5.456	0.4680	0.8278	0.9118	21.10	24.88	26.10	0.0900	0.3925	0.6700
Median	0.5793	4.402	4.950	0.2116	0.4483	0.5256	-1.112	-4.254	-5.014	0.6190	0.9220	0.9360	23.40	28.10	27.85	0.3200	1.115	0.9650
75%[a]	1.478	7.003	6.122	0.3277	0.5728	0.6704	-0.4095	-2.450	-3.780	0.8420	0.9545	0.9538	26.20	29.38	30.63	0.7100	2.210	1.485
Maximum	9.967	16.20	14.40	0.7686	0.7740	0.8552	0.1391	-0.1240	-0.1966	0.9740	0.9860	0.9730	34.00	33.00	36.00	2.610	2.760	2.870
Range	21.21	19.01	16.09	0.7310	0.7193	0.7397	5.743	5.959	5.831	0.8070	0.6640	0.2260	33.08	32.75	12.80	3.940	4.480	3.000
Mean	0.8103	4.803	4.825	0.2444	0.4408	0.5585	-1.510	-3.699	-4.612	0.6404	0.8670	0.9237	22.81	26.96	28.30	0.4448	1.187	1.078
Std. Dev.	1.665	3.968	3.501	0.1280	0.1854	0.1622	1.356	1.637	1.199	0.2083	0.1277	0.0448	6.165	4.625	2.855	0.5570	1.007	0.6950
SEM	0.0774	0.3968	0.4126	0.006	0.0185	0.0191	0.0630	0.1637	0.1413	0.0097	0.01277	0.0053	0.2865	0.4625	0.3365	0.0259	0.1007	0.0819
Skewness	-0.4372	0.6130	0.8598	0.8024	-0.1054	0.0827	-1.028	0.9003	1.485	-0.0180	-1.846	-1.834	-1.279	-2.440	0.5907	0.9711	-0.1071	0.5047
Kurtosis	13.54	-0.1291	0.9598	0.2199	-0.7290	-0.4358	0.1776	-0.4239	2.499	-1.167	3.642	3.741	2.155	10.97	-0.4636	1.698	-0.7484	-0.1692

[a]percentile

REVEL, CPT, ESCOTT and PremPS tools had an approximate normal distribution, however unequal variance and were analysed by a parametric test, one-way ANOVA (Analysis of Variance) with a Browne-Forsythe and Welch statistic and a Dunnett's C post-hoc test. To illustrate these different distributions and the need for different statistical tests, an example of density histograms for Rosetta and CPT are shown in **Figure 24**. Rosetta has a non-normal distribution for the BV, as shown by the long left-sided tail (Kurtosis) and density line, which is taller and narrower than the normal distribution. Whereas all CPT cohorts are normally distributed (similar normal distribution line (red) and density

line (blue)), however, there is a significant difference between group variances, as shown by the different distribution widths and calculated mathematically using the ratio of variances (F value) and comparing it to the F-distribution (look-up table; (http://www.socr.ucla.edu/Applets.dir/F_Table.html)).

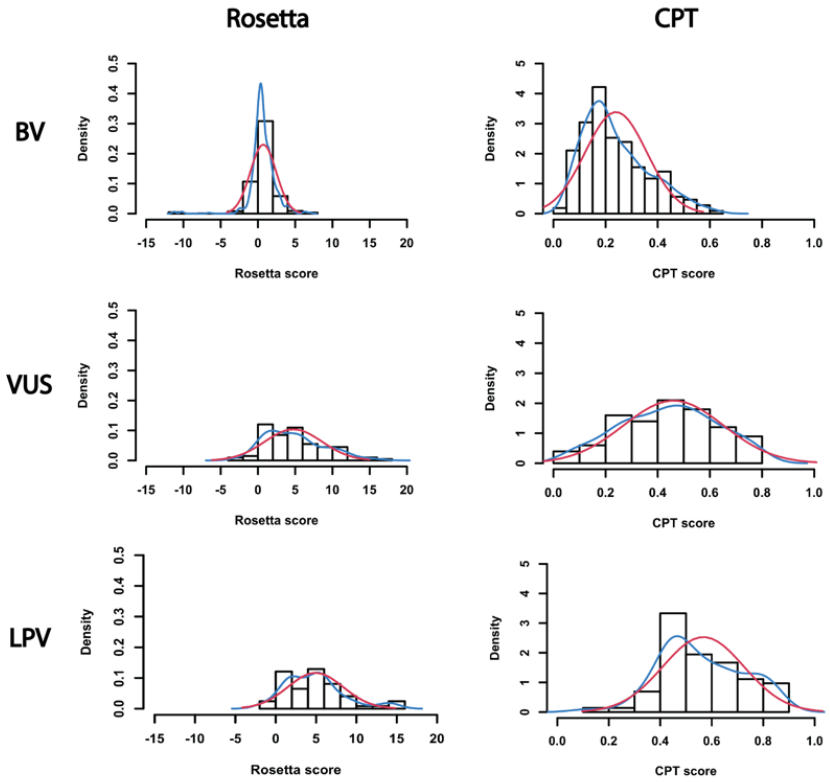


Figure 24: Density histograms for each cohort of *MEN1* variants (BV, VUS and LPV). The red lines on the density plots indicate the normal distribution for the dataset, and the blue indicates the density line. Benign variant (BV); Variant of uncertain significance (VUS) and likely pathogenic variant (LPV).

Bioinformatic scores for each VEP were compared between the three cohorts (BV, VUS, LPV; **Figure 25**). There were 463 BVs, 98 VUS and 74 variants in the LPV cohort. All bioinformatic tools reported a significant difference between the BV and VUS cohorts, and between the BV and LPV cohorts, $p<0.0001$. There was a significant difference in bioinformatic scores between *MEN1* VUS and LPV for CPT and ESCOTT ($p<0.0001$), REVEL ($p<0.001$) and SPDV ($p<0.001$).

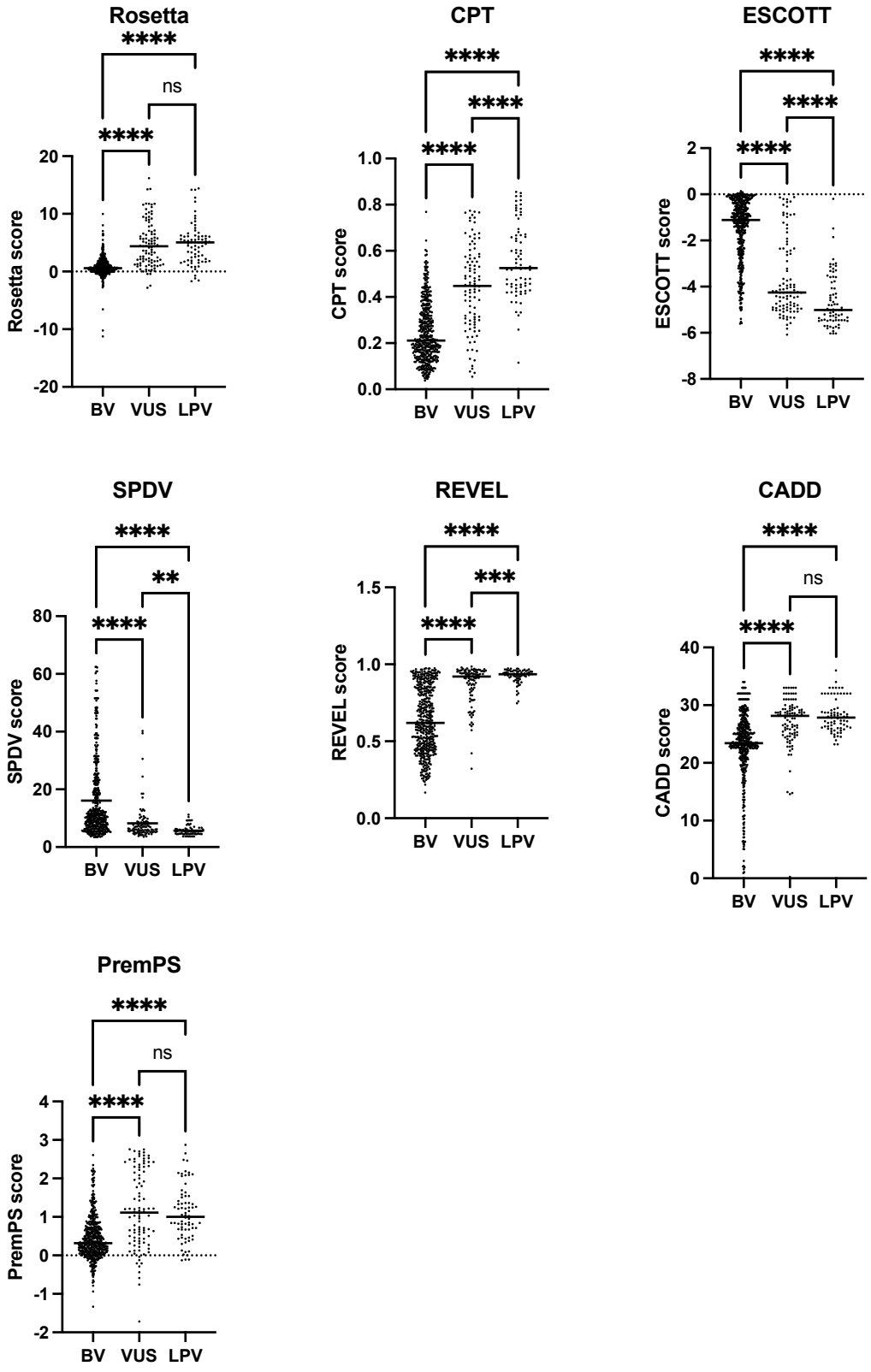


Figure 25: Comparison of each variant effect predictor, comparing the three *MEN1* cohorts: BVs, VUS and LPVs. Benign variant (BV); Variant of uncertain significance (VUS) and likely pathogenic variant (LPV).

ROC analysis was undertaken using all variants in the LPV group (n=74) and BV group (n=463) with Area Under the Curve (AUC) values of 0.8750, 0.9332, 0.9395, 0.9001, 0.8166, 0.7696 and 0.8617, p<0.0001 for Rosetta, CPT, ESCOTT, REVEL, CADD, PremPS and SPDV, respectively **(Figure 26)**. Each VEP significantly distinguished BV from LPVs, with optimal cut-off values having different sensitivities and specificities.

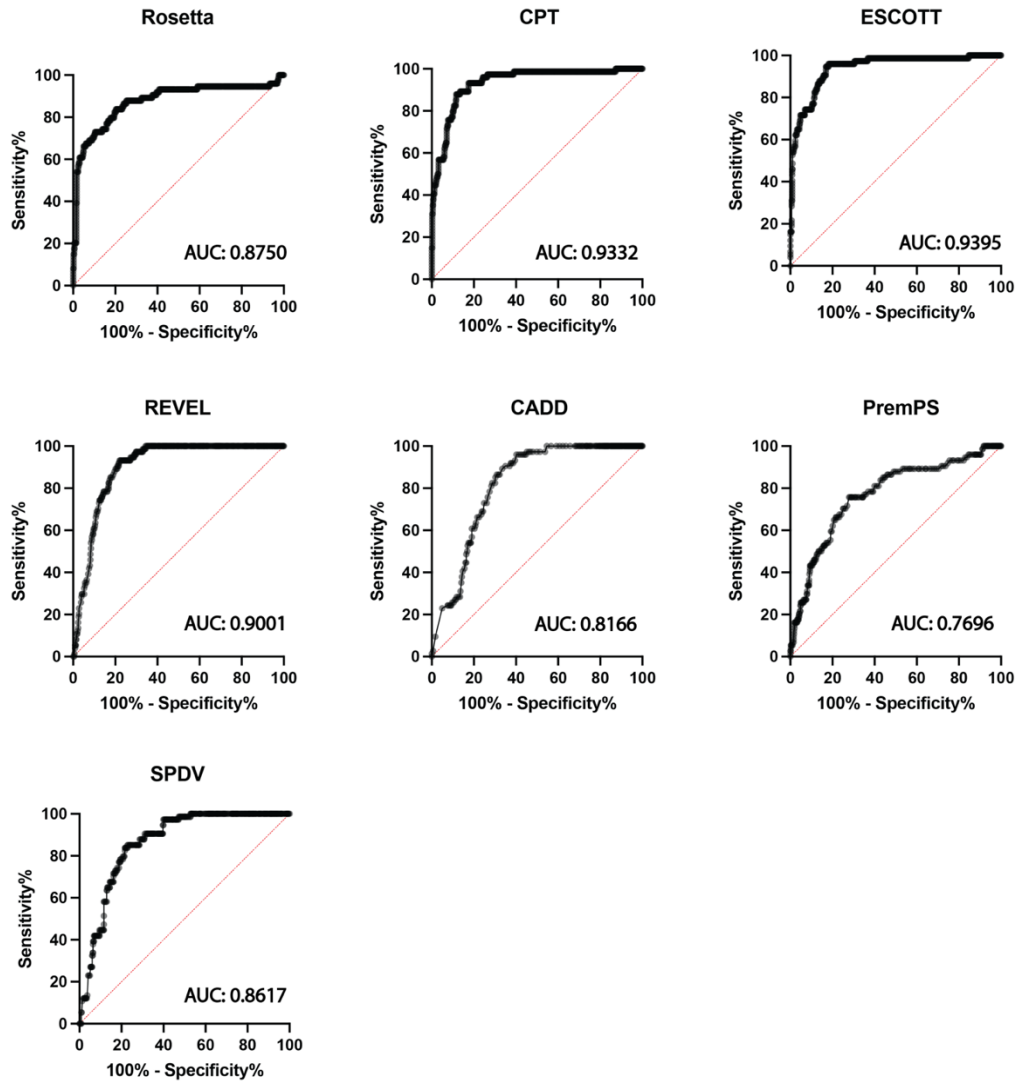


Figure 26: Receiver operator curves for each bioinformatic tool to compare LPV in the literature to BV in gnomAD.

The sensitivity of a specific cut-off level for each VEP refers the proportion of correctly identified LPVs, for example, a long horizontal line at the top of the graph (e.g. REVEL and SPDV), is consistent with a greater sensitivity than the other VEPs (e.g. CADD and PremPS) at the optimal cut-off level. This means that most LPVs will be correctly classified, however

this is at the expense of a lower specificity (i.e. a higher proportion of BVs will be inappropriately classified as being LPV). Whereas, the initial sharp vertical rise of the graphs (e.g. Rosetta, ESCOTT and CPT) indicates that these optimal cut-off levels favour a higher specificity (i.e. less BVs will be inappropriately classified as pathogenic, however at the expense of misclassifying LPVs as BVs). Therefore, given that each individual VEP has strengths with increased sensitivity or specificity (not both), if certain VEPs were combined a more accurate VEP may be generated to classify *MEN1* variants.

4.1.2.4 Construction of new MEN1 models to improve current bioinformatic tool precision

Multiple logistic regression was undertaken to determine a new scoring system for *MEN1* missense variants. This generated two models using the current bioinformatic tools. The first model was generated in GraphPad Prism, and the second in R Studio (due to R Studio's greater functionality compared to Prism and the ability to generate a weighted model).

4.1.2.4.1 Model 1

The initial model was generated in GraphPad Prism and included, all bioinformatic tools (Rosetta, CPT, ESCOTT, REVEL, CADD, PremPS and SPDV), location of amino acid (dichotomous variable: Surface vs Core) and if a particular amino acid mutation was already present in the LPV group (dichotomous variable: yes/no). Model reduction was undertaken, and all variables (apart from CADD) were included in the final model. The AUC of the final model was 0.9786, with superior accuracy compared to all other tools **(Figure 27)**. The Akaike Information Criterion (AIC) value for Model 1 was 143.58. The AIC is useful for determining which model (full versus reduced) best explains the data and is used during model reduction, with a lower AIC favoured for a particular model.

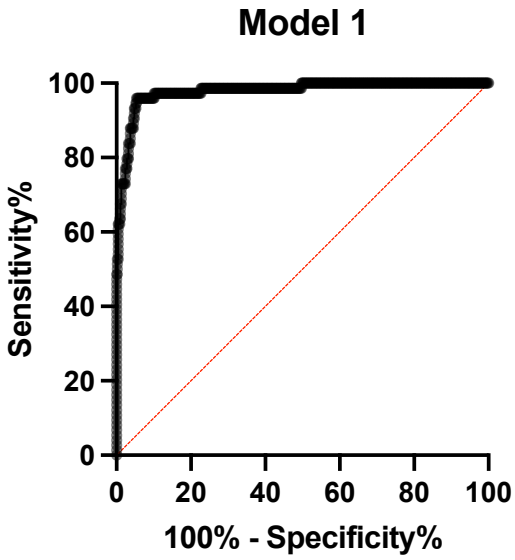


Figure 27: ROC curve for Model 1

4.1.2.4.2 Model 2

Given the difference in the number of variants in the LPV vs the BV group, a weighted logistic regression model was constructed in R Studio. The model initially included all variables and was subsequently reduced to overcome likely overfitting with Model 1. The final reduced Model 2 included only REVEL, SPDV and ESCOTT scores, with LPV variants weighted at 0.84 and BV at 0.16. The AUC for Model 2 was 0.9618, $p<0.0001$, and the ROC curve is shown in **Figure 28**. The AIC for model 2 was 41.766.

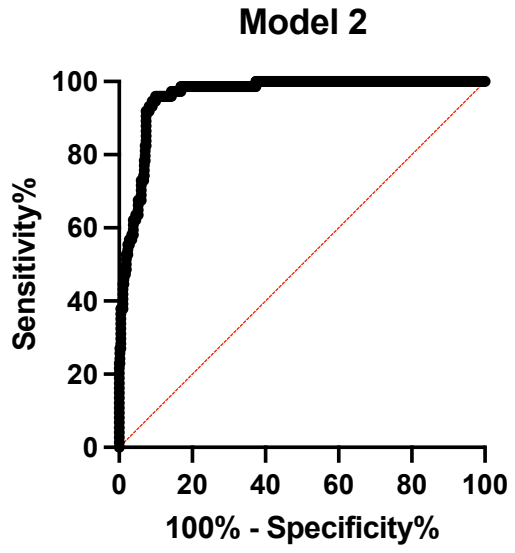


Figure 28: ROC curve for Model 2

Models 1 and 2 have comparable accuracy, however Model 2 has a significantly lower AIC and, therefore, this new generated model with improved diagnostic accuracy of *MEN1* missense variant classification was used. The coefficients for each variable in Model 2 are shown in **Table 20.**

Table 20: Model 2 coefficients

Coefficients	Estimate	Std. Error	P value
Intercept	-8.3387	4.0927	0.0416*
ESCOTT	-0.9708	0.2602	0.0002***
REVEL	9.3162	4.5662	0.0413*
SPDV	-0.4488	0.1555	0.0039**

Standard (Std.)
* p<0.05; **p<0.01 and ***p<0.001

On reviewing the ROC curves for Model 2 compared to each individual VEP, the AUC is higher at 0.9618 as compared to the highest individual VEP, ESCOTT (AUC 0.9395). To further investigate the accuracy of each VEP and model, the cut-off levels for 100%, 98.7%, 97.3%, 95%, 90% and 85% sensitivity and specificity, and the optimal cut-off level (Youden's index) were generated **(Table 21)**. The accuracy for the optimal cut-off levels for Models 1 and 2 are greater than the individual models.

Table 21: ROC analysis results for each bioinformatic tool, Model 1 and Model 2 for the optimal cut-off level and at different sensitivities and specificities

MEN1 SNV predictors	REVEL	CADD	ESCOTT	Rosetta	CPT	SPDV	PremPS	Model 1	Model 2
AUC	0.9001	0.8166	0.9395	0.8750	0.9332	0.8617	0.7696	0.9786	0.9618
Std. Error	0.01374	0.01972	0.01465	0.02756	0.01517	0.01784	0.03058	0.007919	0.008266
95% CI	0.8732 to 0.9270	0.7779 to 0.8552	0.9108 to 0.9682	0.8209 to 0.9290	0.9035 to 0.9630	0.8268 to 0.8967	0.7097 to 0.8296	0.9631 to 0.9942	0.9456 to 0.9780
Optimal cut-off level	> 0.8545	> 24.45	< -2.788	> 1.665	> 0.4125	< 7.020	> 0.6650	> 0.2173	>0.5775
Youden's index	71.21	55.78	77.81	63.48	76.18	62.25	47.82	90.55	86.01
Sensitivity	93.24	95.95	95.95	83.78	87.84	85.14	75.68	95.95	95.95
Specificity	77.97	59.83	81.86	79.7	88.34	77.11	72.14	94.6	90.06
LR	4.233	2.388	5.288	4.127	7.531	3.719	2.716	17.77	9.657
PPV	43.45	27.63	45.10	39.74	54.62	37.28	30.27	73.96	60.68
NPV	97.19	98.93	98.70	96.85	97.85	97.01	94.89	99.32	99.29
Accuracy	82.68	64.80	83.43	80.26	88.27	78.21	72.63	94.79	90.88
100% Sensitivity									
Cut-off level	> 0.7455	> 23.15	< -0.1964	> -1.702	> 0.1136	<11.26	> -0.1400	> 0.00153	> 0.01374
Specificity	65.66	45.36	15.33	2.376	12.74	47.08	8.207	50.32	62.63
LR	2.912	1.83	1.181	1.024	1.146	1.89	1.089	2.013	2.676
PPV	31.76	22.63	15.88	14.07	15.48	23.20	14.83	24.34	29.96
NPV	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Accuracy	70.39	52.89	27.00	15.83	24.77	54.38	20.86	57.17	67.78
98.7% Sensitivity									
Cut-off level	> 0.7585	> 23.15	<-1.479	>-1.605	>0.2587	<10.44	>-0.1150	>0.01799	> 0.3224
Specificity	66.95	45.36	63.28	2.592	61.12	52.7	8.855	77.11	83.15

MEN1 SNV predictors	REVEL	CADD	ESCOTT	Rosetta	CPT	SPDV	PremPS	Model 1	Model 2
LR	2.985	1.83	2.687	1.013	2.537	2.086	1.082	4.309	5.856
PPV	32.30	22.63	30.04	13.93	28.85	25.00	14.75	41.01	48.34
NPV	99.68	100.00	99.66	92.31	99.65	99.59	97.62	99.72	99.74
Accuracy	71.32	52.89	68.16	15.83	66.29	59.03	21.23	80.26	85.29
97.3% Sensitivity									
Cut-off level	>0.8010	>23.85	<-1.856	>-1.212	>0.3200	<9.343	>-0.1150	> 0.08529	> 0.3960
Specificity	70.19	54.86	69.55	3.024	73.65	60.04	8.855	89.85	85.75
LR	3.264	2.155	3.195	1.003	3.693	2.435	1.082	9.585	6.826
PPV	34.29	25.62	33.80	13.82	37.11	28.02	14.75	60.50	52.17
NPV	99.39	99.22	99.38	87.50	99.42	99.29	97.62	99.52	99.50
Accuracy	73.93	60.71	73.37	16.01	76.91	65.18	21.23	90.88	87.34
95% Sensitivity									
Cut-off level	> 0.8110	> 24.45	< -2.788	> -0.7353	> 0.3319	<9.343	> -0.00500	> 0.2173	> 0.5775
Specificity	71.06	59.83	81.86	6.695	75.81	60.04	15.33	94.6	90.06
LR	3.315	2.388	5.288	1.028	3.966	2.325	1.133	17.77	9.657
PPV	34.63	27.63	45.81	14.12	38.80	28.02	15.33	73.96	60.68
NPV	99.10	98.93	99.21	91.18	99.15	99.29	95.95	99.32	99.29
Accuracy	74.49	64.80	83.80	18.99	78.58	65.18	26.44	94.79	90.88
90% Sensitivity									
Cut-off level	> 0.8660	> 24.85	< -3.093	> 0.8839	> 0.3773	<8.070	> 0.1250	> 0.2503	> 0.6841
Specificity	79.05	62.42	84.23	> 0.8839	82.51	68.9	28.29	95.46	92.66
LR	4.322	2.445	5.743	2.409	5.175	2.911	28.29	19.96	12.33
PPV	40.85	28.10	47.86	27.80	45.27	31.75	16.79	76.14	66.34

MEN1 SNV predictors	REVEL	CADD	ESCOTT	Rosetta	CPT	SPDV	PremPS	Model 1	Model 2
NPV	98.12	97.97	98.24	97.64	98.20	97.85	94.93	98.44	98.39
Accuracy	80.63	66.48	85.10	66.29	83.61	71.88	36.87	94.79	92.36
85% Sensitivity									
Cut-off level	> 0.8895	> 25.55	< -3.290	> 1.515	> 0.4188	<0.2595	28.29	> 0.3919	> 0.6927
Specificity	82.29	69.76	87.04	76.67	88.55	92.66	28.29	96.33	92.66
LR	4.807	2.816	6.645	3.65	7.437	4.625	1.932	23.19	11.59
PPV	43.45	31.03	51.61	36.84	54.31	13.78	23.60	78.75	64.95
NPV	97.19	96.71	97.58	96.99	97.39	NA	95.93	97.59	97.50
Accuracy	82.68	71.88	86.96	77.84	88.08	13.78	59.96	94.79	91.62
100% specificity									
Cut-off level	> 0.9735	> 35.00	< -5.629	> 10.21	> 0.7721	<3.440	> 2.630	> 0.8923	> 0.9702
Sensitivity	0	1.351	16.22	8.108	14.86	0	2.703	48.65	22.97
LR	0	NA	NA	NA	NA	NA	NA	NA	NA
PPV	0.00	100.00	100.00	100.00	100.00	0.00	100.00	100.00	100.00
NPV	86.17	86.38	88.19	87.19	88.02	86.19	86.54	92.42	89.04
Accuracy	85.85	86.41	88.45	87.34	88.27	86.03	86.59	92.92	89.39
95% specificity									
Cut-off level	> 0.9500	> 31.50	< -4.296	> 3.288	> 0.4904	<4.548	>1.505	>0.2360	> 0.8282
Sensitivity	29.73	22.97	71.62	66.22	56.76	22.97	25.68	93.24	63.51
LR	6.882	4.625	14.42	13.33	11.43	4.625	5.169	18.77	12.79
PPV	52.38	42.50	69.74	68.06	64.62	42.50	45.24	75.00	67.14
NPV	89.49	88.53	95.44	94.62	93.22	88.53	88.89	98.88	94.22
Accuracy	86.59	85.10	91.81	91.06	89.76	85.10	85.47	94.79	90.69

MEN1 SNV predictors	REVEL	CADD	ESCOTT	Rosetta	CPT	SPDV	PremPS	Model 1	Model 2
90% specificity									
Cut-off level	> 0.9285	> 29.45	< -3.832	> 2.527	> 0.4259	<4.910	>1.170	>0.08894	> 0.5775
Sensitivity	60.81	25.68	74.32	71.62	79.73	41.89	43.25	95.95	95.95
LR	6.257	2.642	7.481	7.209	8.025	5.878	4.353	9.657	9.657
PPV	50.00	29.69	54.46	53.54	56.19	48.44	41.03	60.68	60.68
NPV	93.51	88.37	95.64	95.21	96.53	90.91	90.85	99.29	99.29
Accuracy	86.22	81.38	87.90	87.52	88.64	85.85	83.61	90.88	90.88
85% specificity									
Cut-off level	> 0.9065	> 28.35	< -3.171	> 2.053	> 0.3940	<5.813	>0.9700	>0.04847	> 0.3624
Sensitivity	78.38	40.54	87.84	74.32	89.19	67.57	51.35	97.3	97.3
LR	5.337	2.76	5.894	7.209	5.985	4.534	3.446	6.529	6.529
PPV	46.03	30.61	48.51	44.35	48.89	42.02	35.51	51.06	51.06
NPV	96.11	89.98	97.77	95.40	98.01	94.26	91.63	99.49	99.49
Accuracy	84.36	79.14	85.47	83.61	85.66	82.68	80.45	86.78	86.78

AUC: Area under the curve; CI: Confidence Interval; LR: likelihood ratio; NPV: negative predictive value; PPV: positive predictive value; Std: Standard; ROC: Receiver Operator Curves; SNV: Single nucleotide variant.

4.1.2.5 Validating the accuracy of bioinformatic tools, Models 1 and 2 in a separate patient cohort from Geisinger

Out of ~175,000 patients from the Geisinger cohort, 1075 patients had 215 *MEN1* variants; the breakdown of variants is shown in **Figure 29 Panel A**. Most mutations were missense variants (78%; 168/215), with a distribution similar to what has been reported in gnomAD **(Figure 29 Panel B)**. The average age of the Geisinger population is 57.2 years, with 730 people ≥50 years.

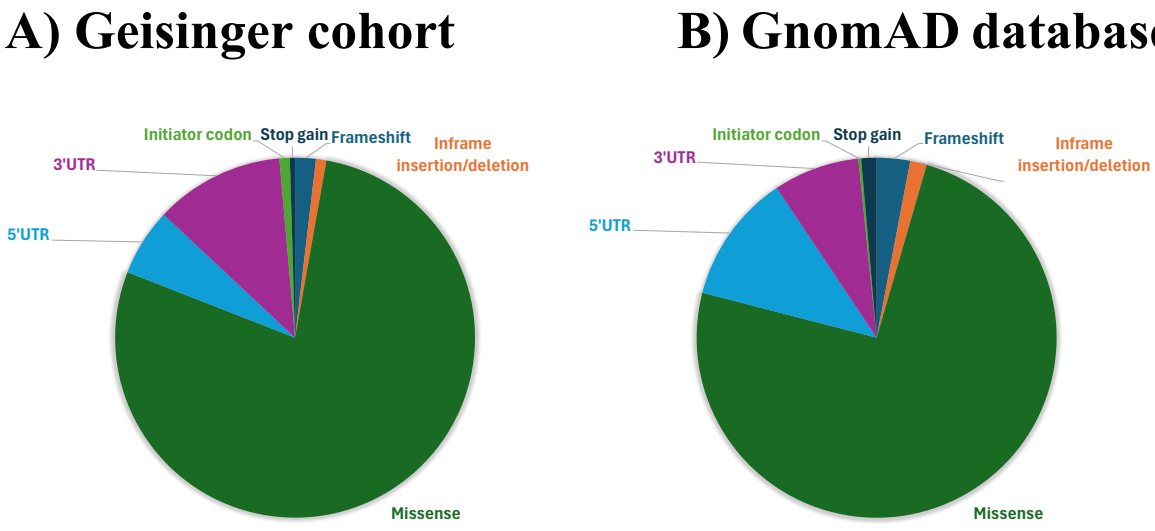


Figure 29: Percentages of types of *MEN1* mutation reported in Geisinger cohort (Panel A) and gnomAD (Panel B).

Patients with PTVs are more likely to develop a MEN1 phenotype than patients with missense variants from single nucleotide variants (SNVs); therefore, the ICD codes and serum calcium levels from patients with PTVs were reviewed. PHPT (the most common manifestation of MEN1) usually occurs before the age of 50 years. Therefore, this ICD, E21 (Hyperparathyroidism and other disorders of the parathyroid gland), was reviewed first.

4.1.2.5.1 Truncating variants

In the Geisinger cohort, 7 *MEN1* PTVs (4 frameshift, 1 termination, and 2 initiator codon), which have been reported in the literature to cause MEN1, were found in 22 people from 14 families **(Table 22)**. All patients were reported to be of white background.

Table 22: Numbers of patients in the Geisinger cohort with a *MEN1* protein-truncating variant

Truncating	Patients affected
p.Ile85Serfs*33	2
p.Leu103Cysfs*16	15
p.Arg108Ter	1
p.Met1Thr	1
p.Trp270Alafs*4	1
p.Ala469Argfs*98	1
p.Met1Ile	1

PHPT was diagnosed in 77% (17/22) **(Figure 30)**. Out of the 5 people remaining without a diagnosis of PHPT: two people with *MEN1* variants M1I and L103Cfs*16, <50 years had normal serum calcium levels, one person >70y with *MEN1* variant W270Afs*4 had no serum calcium level, one person >50 years old with *MEN1* variant M1T, had a normal serum calcium and one person with *MEN1* variant R108Ter, had no ICD code for PHPT had a maximum elevated serum calcium ≥11.2mg/dL (i.e. >2.8mmol/L). Initiator codon variants may not always manifest a phenotype, and people <50 years old may not have had time to develop a calcium phenotype. When these individuals (<50 years old) and all patients with no serum calcium estimations were excluded, then all the remaining patients (18/18) had either a diagnosis of PHPT or were hypercalcaemic ≥11.2mg/dL (i.e. >2.8mmol/L). The average age of all people with protein-truncating *MEN1* variants in the Geisinger population was 55.7 years.

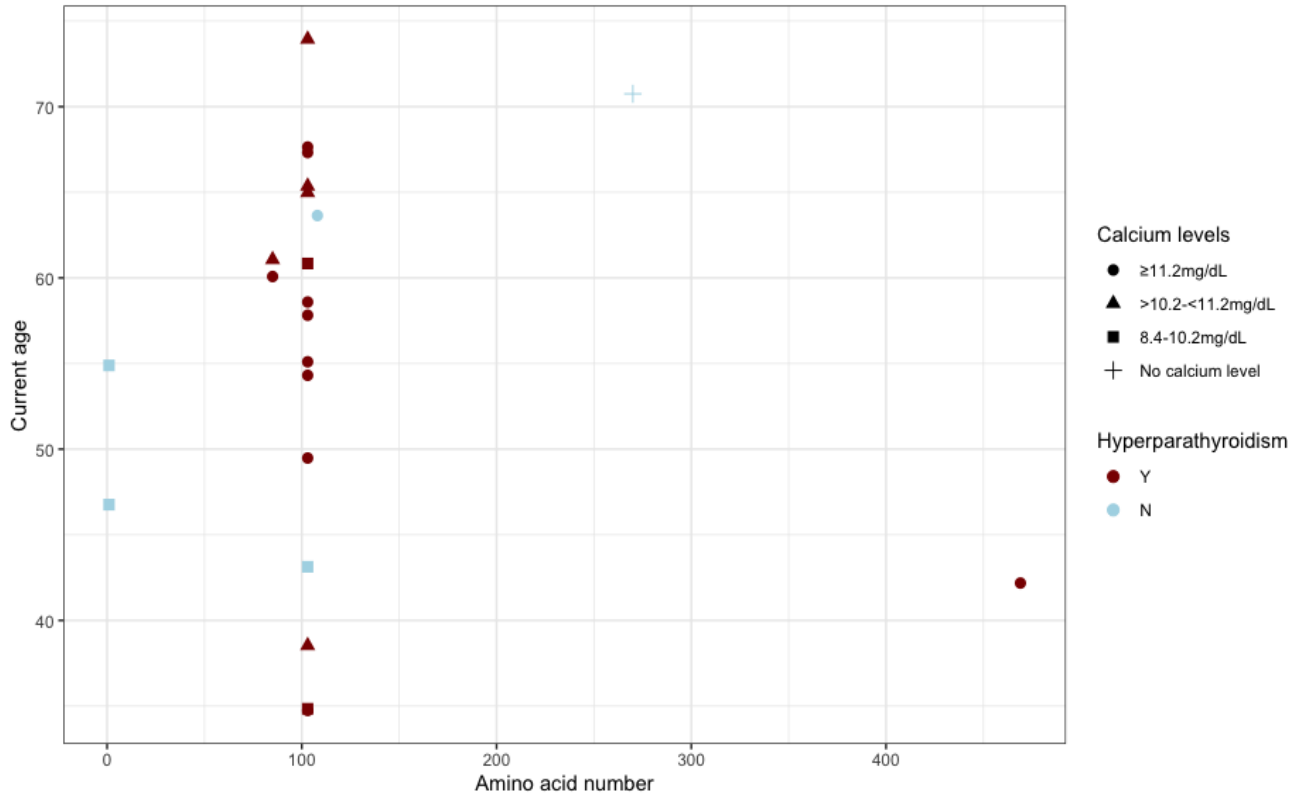


Figure 30: Protein truncating *MEN1* variants plotted against current age. Red shapes indicate the patient has an ICD code for PHPT, and blue indicates they do not.

All ICD codes of interest enriched in patients with *MEN1* PTVs (adj.p value <0.01) were extracted **(Table 23)**. These enriched ICD codes include common tumours associated with MEN1 (e.g. E21: Hyperparathyroidism and other disorders of parathyroid gland, E23: Hypofunction and other disorders of the pituitary gland, C25: Malignant neoplasm of pancreas, D3A: Benign neuroendocrine tumours, and C25: Malignant neuroendocrine tumours) or sequelae from either the tumour itself (e.g. K26: Duodenal ulcer, E83 Disorders of mineral metabolism and N20: Calculus of kidney and ureter) or the consequence of its treatment (E20: Hypoparathyroidism, E13 Other specified diabetes mellitus). The average number of ICD codes enriched in people with a truncating *MEN1* variant is 4.8. Both patients with initiator codon variants had no ICD enriched in patients with *MEN1* PTVs.

Table 23: ICD codes enriched in patients with *MEN1* protein-truncating variants in the Geisinger cohort

Base Code	ICD Description	Count	adj. p value
E21	Hyperparathyroidism and other disorders of parathyroid gland	17	0
E31	Polyglandular dysfunction	15	0
D3A	Benign neuroendocrine tumors	6	4.68E-13
D35	Benign neoplasm of other and unspecified endocrine glands	10	2.18E-12
D37	Neoplasm of uncertain behavior of oral cavity and digestive organs	6	4.88E-10
E20	Hypoparathyroidism	4	1.16E-08
E23	Hypofunction and other disorders of the pituitary gland	7	1.20E-08
C25	Malignant neoplasm of pancreas	5	4.89E-08
E83	Disorders of mineral metabolism	11	2.94E-06
K86	Other diseases of pancreas	7	4.34E-06
C75	Malignant neoplasm of other endocrine glands and related structure	1	4.59E-06
C7A	Malignant neuroendocrine tumors	2	0.00031
D13	Benign neoplasm of other and ill-defined parts of digestive system	2	0.000761
C7B	Secondary neuroendocrine tumors	1	0.001593
N20	Calculus of kidney and ureter	9	0.002763
K26	Duodenal ulcer	2	0.003950
D44	Neoplasm of uncertain behavior of endocrine glands	1	0.006347
E13	Other specified diabetes mellitus	2	0.009474

Adjusted p value (adj. p value); International Classification of Diseases (ICD).

The ICD codes for each patient with a *MEN1* PTV that are enriched in the whole cohort of *MEN1* PTV are shown in **Table 24**.

Table 24: ICD codes for each patient with a truncating *MEN1* variant

Variant	Age (years)	ICD codes
p.Ile85Serfs*33 **(carrier count 2)**		
PT1008884[a]	60	Benign neoplasm of other and unspecified endocrine glands Neoplasm of uncertain behavior of oral cavity and digestive organs Hyperparathyroidism and other disorders of parathyroid gland Polyglandular dysfunction
PT1164093[a]	61	Benign neoplasm of other and unspecified endocrine glands Benign neuroendocrine tumors Disorders of mineral metabolism Duodenal ulcer Hyperparathyroidism and other disorders of parathyroid gland Hypofunction and other disorders of the pituitary gland Hypoparathyroidism Malignant neoplasm of pancreas Neoplasm of uncertain behavior of oral cavity and digestive orga

		Other diseases of pancreas
		Polyglandular dysfunction
p.Leu103Cysfs*16 **(carrier count 16)**		
PT312753[b]	61	Benign neoplasm of other and unspecified endocrine glands
		Hyperparathyroidism and other disorders of parathyroid gland
		Hypofunction and other disorders of the pituitary gland
		Disorders of mineral metabolism
PT462116[b]	74	Hyperparathyroidism and other disorders of parathyroid gland
		Calculus of kidney and ureter
PT1221533[c]	65	Other specified diabetes mellitus
		Hyperparathyroidism and other disorders of parathyroid gland
PT670588[c]	54	Malignant neuroendocrine tumors
		Benign neoplasm of other and unspecified endocrine glands
		Neoplasm of uncertain behavior of oral cavity and digestive organs
		Hyperparathyroidism and other disorders of parathyroid gland
		Polyglandular dysfunction
		Disorders of mineral metabolism
PT1053799	59	Malignant neoplasm of pancreas
		Benign neoplasm of other and unspecified endocrine glands
		Hyperparathyroidism and other disorders of parathyroid gland
		Polyglandular dysfunction
		Other diseases of pancreas
PT1054084	67	Hyperparathyroidism and other disorders of parathyroid gland
		Polyglandular dysfunction
		Disorders of mineral metabolism
PT1205590	34	Benign neoplasm of other and unspecified endocrine glands
		Benign neuroendocrine tumors
		Hypoparathyroidism
		Hyperparathyroidism and other disorders of parathyroid gland
		Hypofunction and other disorders of the pituitary gland
		Polyglandular dysfunction
		Disorders of mineral metabolism
		Other diseases of pancreas
		Calculus of kidney and ureter
PT1207875	55	Neoplasm of uncertain behavior of oral cavity and digestive organs
		Benign neuroendocrine tumors
		Hyperparathyroidism and other disorders of parathyroid gland
		Polyglandular dysfunction
		Disorders of mineral metabolism
		Calculus of kidney and ureter
PT1273481	43	Polyglandular dysfunction
PT1421142	57	Benign neoplasm of other and ill-defined parts of digestive system

		Benign neoplasm of other and unspecified endocrine glands Neoplasm of uncertain behavior of endocrine glands Hyperparathyroidism and other disorders of parathyroid gland Hypofunction and other disorders of the pituitary gland Polyglandular dysfunction Disorders of mineral metabolism Other diseases of pancreas
PT264066	49	Malignant neuroendocrine tumors Secondary neuroendocrine tumors Benign neoplasm of other and unspecified endocrine glands Benign neuroendocrine tumors Hyperparathyroidism and other disorders of parathyroid gland Polyglandular dysfunction Disorders of mineral metabolism Calculus of kidney and ureter
PT1031133	67	Hyperparathyroidism and other disorders of parathyroid gland Hypofunction and other disorders of the pituitary gland Polyglandular dysfunction Disorders of mineral metabolism Calculus of kidney and ureter
PT321017	34	Benign neoplasm of other and ill-defined parts of digestive system Benign neuroendocrine tumors Hyperparathyroidism and other disorders of parathyroid gland Hypofunction and other disorders of the pituitary gland Polyglandular dysfunction Other diseases of pancreas Calculus of kidney and ureter
PT660913	65	Malignant neoplasm of pancreas Benign neoplasm of other and unspecified endocrine glands Benign neuroendocrine tumors Other specified diabetes mellitus Hypoparathyroidism Hyperparathyroidism and other disorders of parathyroid gland Polyglandular dysfunction Calculus of kidney and ureter
PT795026	39	Malignant neoplasm of pancreas Neoplasm of uncertain behavior of oral cavity and digestive organs Hyperparathyroidism and other disorders of parathyroid gland Hypofunction and other disorders of the pituitary gland Polyglandular dysfunction Disorders of mineral metabolism Other diseases of pancreas Calculus of kidney and ureter
p.Ala469Argfs*98 **(carrier count 1)**		
PT966834	42	Malignant neoplasm of pancreas

		Malignant neoplasm of other endocrine glands and related structures Benign neoplasm of other and unspecified endocrine glands Neoplasm of uncertain behavior of oral cavity and digestive organs Hypoparathyroidism Hyperparathyroidism and other disorders of parathyroid gland Polyglandular dysfunction Disorders of mineral metabolism Duodenal ulcer Other diseases of pancreas Calculus of kidney and ureter
p.Arg108Ter **(carrier count 1)**		
PT1235014	64	Polyglandular dysfunction Secondary malignant neoplasm of other and unspecified sites*
p.Trp270Alafs*4 **(carrier count 1)**		
PT853222	71	Other benign neoplasms of skin Neoplasm of uncertain behavior of other and unspecified sites Other and unspecified soft tissue disorders not elsewhere classified Osteoporosis without current pathological fracture Fracture of forearm
p.Met1Ile **(carrier count 1)**		
PT72989	47	Type 2 diabetes mellitus* Vitamin D deficiency* Other disorders of fluid electrolyte and acid-base balance*
p.Met1Th **(carrier count 1)**		
PT720261	55	Other venous embolism and thrombosis* Other and unspecified soft tissue disorders not elsewhere classified* Fracture of lower leg including ankle*

[a]Family 1; [b]Family 2 and [c]Family 3.*Other ICD codes reported in patient records, not enriched in *MEN1* PTV group.

4.1.2.5.2 Missense variants

There were 168 unique *MEN1* missense variants from 693 patients reported in the Geisinger cohort, with 11% (18/163) of these variants reported in the literature to be associated with MEN1 (LPV n=5 (G161S, L173P, C246Y, A342P and D423N); and VUS n=13 (P12L, V167A, G174S, R176W, R234H, L254P, C359F, E371D, V372I, A390V, G537C, P454S and P549S)**(Table 25)**.

Table 25: *MEN1* variants reported in Geisinger cohort

Variant	GRCh38	ClinVar	Domains	No. of people	No. of papers	PHRED	REVEL	PremPS	Rosetta	CPT	ESCOTT	SPDV	gnomAD (1=yes)
A6S	64810094	VUS	NTD	NA	NA	21.2	0.486	0.18	0.34251701	0.17249676	-0.2424402	8.73374153	1
P12L	64810075	LP	NTD	1	1	32	0.892	0.13	2.077210884	0.33300697	-4.0174589	10.6932194	0
R14C	64810070	VUS	NTD	1*	NA	32	0.829	0.28	1.102040816	0.19342611	-0.8355662	13.0536891	1
S15F	64810066	VUS	NTD	NA	NA	29	0.584	-0.07	1.36428571	0.38032528	-3.4204212	10.6948379	1
D18Y	64810058	VUS	NTD	NA	NA	29.4	0.744	-0.09	0.43367347	0.33718542	-3.0617066	7.61425676	1
R21S	64810049	Con	NTD	1*	NA	32	0.773	0.17	0.900226757	0.1356565	-0.6710692	6.36829136	1
R21C	64810049	VUS	NTD	1*	NA	23.4	0.577	0.45	1.207029478	0.21722927	-2.068456	6.36829136	0
R29Q	64810024	NA	NTD	NA	NA	23.1	0.434	0.6	-0.2772109	0.19056051	-0.3400738	4.64953032	0
E30G	64810021	Con	NTD	NA	NA	24.1	0.523	0.16	0.90839002	0.13095169	-0.5823538	9.7760902	0
P32S	64810016	Con	NTD	NA	NA	30	0.848	1.18	3.06893424	0.36222702	-3.8979327	4.99700847	1
V35L	64810007	VUS	NTD	NA	NA	22.8	0.517	0.88	0.76485261	0.19929804	-1.2569224	5.83254411	1
V40L	64809992	NA	NTD	NA	NA	22.7	0.688	0.49	2.01020408	0.30377268	-1.1936935	4.96937549	0
T56I	64809943	VUS	NTD	NA	NA	23.1	0.483	-0.05	0.40986395	0.09686283	-0.354456	20.0023524	0
P59S	64809935	VUS	NTD	NA	NA	22.6	0.715	0.06	0.74319728	0.11629907	-0.6420559	18.1552052	1
F63L	64809923	NA	NTD	NA	NA	23.4	0.669	0.87	3.26802721	0.21817272	-1.3835606	14.2616152	1
A68T	64809908	VUS	NTD	NA	NA	19.38	0.5	0.57	0.03639456	0.13981765	-0.3589889	19.6436479	0
A68V	64809907	VUS	NTD	NA	NA	22.6	0.513	-0.35	0.24965986	0.26975415	-0.3620479	19.6436479	1
P69S	64809905	VUS	NTD	NA	NA	18.37	0.585	0.31	0.57210884	0.22279162	-0.4533064	19.7779585	0
D70N	64809902	VUS	NTD	NA	NA	23.9	0.638	0.01	0.18673469	0.19337041	-0.6165751	24.0069787	1
D70H	64809902	NA	NTD	NA	NA	22.8	0.383	-0.08	0.84285714	0.16462506	-1.1964678	24.0069787	0
D70E	64809900	VUS	NTD	NA	NA	18.81	0.479	0.13	-0.0068027	0.05599703	-0.1372857	24.0069787	1
G73D	64809892	NA	NTD	NA	NA	21.6	0.662	0.81	1.22585034	0.32368797	-0.0800728	22.3890046	1
T76P	64809884	NA	NTD	NA	NA	22.9	0.535	0.08	2.02482993	0.16125668	-0.6922572	15.4589823	0
Y77H	64809881	VUS	NTD	NA	NA	25.5	0.573	0.49	-1.0040816	0.19587306	-2.4891291	16.1340725	0
A81T	64809869	VUS	NTD	NA	NA	27.6	0.486	0.25	2.33526077	0.23680372	-2.3910583	10.1004861	0

Variant	GRCh38	ClinVar	Domains	No. of people	No. of papers	PHRED	REVEL	PremPS	Rosetta	CPT	ESCOTT	SPDV	gnomAD (1=yes)
D82E	64809864	VUS	NTD	NA	NA	16.56	0.504	-0.28	0.65600907	0.0595247	-0.1879758	12.1872226	1
Y90C	64809841	VUS	NTD	NA	NA	29.1	0.835	1.43	2.59591837	0.23366377	-1.9358979	7.53440308	1
R92H	64809835	VUS	NTD	NA	NA	32	0.856	0.51	0.51099773	0.35118859	-3.847295	10.0474783	1
Y106C	64809793	VUS	Thumb	1*	NA	29.5	0.921	1.48	3.456802721	0.26080729	-2.8537636	10.4784661	0
G110A	64809781	VUS	Thumb	NA	NA	22.2	0.457	0.32	1.26292517	0.08654471	-0.2359578	9.39106736	1
G111D	64809778	VUS	Thumb	NA	NA	23.6	0.53	0.69	0.17732426	0.38503996	-0.8849273	6.79371294	1
S113F	64809772	VUS	Thumb	NA	NA	31	0.92	-0.25	6.40578231	0.40209085	-4.0076158	5.08695658	1
R131C	64809719	VUS	Thumb	NA	NA	32	0.977	0.63	1.0085034	0.30785198	-3.0361582	16.6806062	0
R137P	64809700	NA	Thumb	NA	NA	29.6	0.941	0.82	2.94965986	0.53378809	-4.2707989	6.26316854	0
P152Q	64809655	NA	Thumb	NA	NA	10.15	0.043	0.38	1.6377551	0.17557602	-1.1735224	15.4816257	0
K156R	64808093	VUS	Thumb	NA	NA	24.2	0.769	0.53	0.36394558	0.18751174	-0.9661406	8.66022467	1
G161S	64808079	Con	Thumb[a]	4	1	24.9	0.972	0.46	1.89138322	0.53177539	-4.9131853	3.70166599	1
V167A	64808060	VUS	Thumb	1	1	26.2	0.966	2.08	1.802834467	0.46840133	-2.6310961	5.08107362	1
C170W	64808050	NA	Thumb	NA	NA	26.5	0.917	0.61	7.53197279	0.80533103	-4.216749	3.72504422	0
L173P	64808042	P/LPV	Thumb	23	6	28.3	0.956	2.14	7.553741497	0.78188124	-4.4746687	5.3136143	1
G174S	64808040	VUS	Thumb	1	1	25.9	0.944	0.73	1.077891156	0.55305777	-2.1644364	4.66309515	1
R176W	64808034	Con	Thumb	1	1	24.8	0.657	0.08	0.296938776	0.22610698	-0.4514957	6.89491521	1
A181S	64808019	Con	Thumb[a]	NA	NA	25.7	0.943	0.57	1.75464853	0.42036633	-1.3277196	3.78519903	1
A181T	64808019	VUS	Thumb[a]	NA	NA	24.9	0.868	0.44	1.09297052	0.54972597	-2.4340789	3.78519903	1
L182P	64808015	VUS	Thumb	NA	NA	27.1	0.941	2.2	4.28696145	0.7685558	-5.5237791	6.25970506	1
F191S	64807988	NA	Thumb	NA	NA	28.4	0.957	2.84	4.73684807	0.65224899	-4.8197584	6.0707208	0
P193T	64807983	NA	Thumb	NA	NA	22.9	0.468	0.51	1.60204082	0.18735081	-2.0889994	7.8786442	1
P193L	64807982	Con	Thumb	1*	NA	23.7	0.678	0.34	1.470748299	0.19982208	-1.1354771	7.8786442	1
N194S	64807979	Con	Thumb	NA	NA	18.59	0.419	0	0.7037415	0.0897527	-0.4775963	9.43360551	1
G195R	64807977	VUS	Thumb	NA	NA	22.6	0.805	1.02	0.90646259	0.29347768	-1.4226273	7.4966281	1
T215P	64807917	NA	Thumb	NA	NA	18.03	0.568	-0.57	-0.8450113	0.20632554	-0.3377248	9.22196641	0

Variant	GRCh38	ClinVar	Domains	No. of people	No. of papers	PHRED	REVEL	PremPS	Rosetta	CPT	ESCOTT	SPDV	gnomAD (1=yes)
N217D	64807911	VUS	Thumb	NA	NA	17.89	0.436	0.23	-0.14161	0.13288977	-0.2185552	5.53211141	1
N217S	64807910	VUS	Thumb	NA	NA	16.35	0.397	-0.36	0.07460318	0.04789096	-0.2930365	5.53211141	1
A218D	64807907	VUS	Thumb	NA	NA	22	0.784	0.78	-0.7725624	0.37219174	-1.1523211	5.34216446	0
G219S	64807905	VUS	Thumb	3	§	23.3	0.879	0.88	0.471768707	0.40962365	-1.6027188	3.64952614	1
R223Q	64807892	VUS	Thumb	NA	NA	22.3	0.557	0.59	-0.1673469	0.2669693	-1.2687554	7.14504585	1
S224G	64807680	VUS	Thumb	NA	NA	23.9	0.85	0.8	1.40442177	0.33273621	-2.3689622	6.29877058	0
M233I	64807651	VUS	Thumb	NA	NA	21.9	0.577	-0.71	0.52687075	0.1394627	-1.1575278	8.09860516	1
R234C	64807650	Con	Thumb	NA	NA	25.2	0.624	0.29	2.1085034	0.200467	-0.837476	7.5863935	1
R234H	64807649	Con	Thumb	1	1	22.9	0.622	0.59	2.570521542	0.30723742	-1.917684	7.5863935	1
R237C	64807641	Con	Palm	NA	NA	29.8	0.932	0.06	1.45272109	0.34537896	-3.8059798	9.59127718	1
R237H	64807640	VUS	Palm	NA	NA	27.7	0.944	0.85	0.6223356	0.41847334	-4.7456668	9.59127718	1
M244T	64807619	VUS	Palm	NA	NA	25.4	0.928	1.58	2.30226757	0.34761824	-3.9862752	7.19063987	1
C246Y	64807613	P/LPV	Palm[a]	4	3	26	0.918	0.38	6.046712018	0.42651957	-5.0184235	9.2355282	0
I252V	64807596	VUS	Palm	NA	NA	23.3	0.528	0.57	0.26632653	0.17296442	-0.4094951	8.73223107	1
I252T	64807595	VUS	Palm	1*	NA	27.3	0.876	1.38	1.139795918	0.37829714	-2.9038729	8.73223107	1
L254P	64807589	NA	Palm	5	1	24.6	0.768	0.31	1.787414966	0.23160665	-1.323824	12.6139385	0
T256A	64807584	NA	Palm	1	§	22.7	0.825	0.64	0.079251701	0.19850422	-0.7936091	12.6981979	0
T256S	64807583	NA	Palm	NA	NA	21.3	0.653	0.29	0.05680272	0.1386615	-0.19287	12.6981979	0
D257N	64807581	VUS	Palm	NA	NA	26.3	0.942	0.48	-0.0329932	0.29567924	-1.3263308	11.1239832	1
D257H	64807581	VUS	Palm	NA	NA	26.4	0.806	0.43	-0.0302721	0.37190486	-4.2803857	11.1239832	1
L261V	64807569	NA	Palm	NA	NA	18.23	0.674	0.58	1.70396825	0.26579277	-1.0724447	6.40868157	0
Q263E	64807563	VUS	Palm	NA	NA	19.12	0.576	0.16	1.46666667	0.18153461	-0.9489084	13.3314637	1
Q263H	64807561	Con	Palm	NA	NA	21	0.815	0.65	0.75079365	0.19057487	-3.1984535	13.3314637	1
K267R	64807218	VUS	Palm	NA	NA	22.9	0.539	-0.01	0.32721088	0.07150167	-0.1955931	12.6554034	0
L269F	64807213	NA	Palm	NA	NA	28.6	0.916	0.78	1.40657596	0.49406844	-3.0183954	7.21447994	0
Y273C	64807200	VUS	Palm	2*	NA	29.7	0.884	1.29	2.769727891	0.39137684	-4.4965937	11.3900105	1

Variant	GRCh38	ClinVar	Domains	No. of people	No. of papers	PHRED	REVEL	PremPS	Rosetta	CPT	ESCOTT	SPDV	gnomAD (1=yes)
L275P	64807194	VUS	Palm	NA	NA	25.2	0.725	0.81	8.0329932	0.52393656	-3.6644985	12.3994626	1
H277Y	64807189	VUS	Palm	NA	NA	25	0.769	-0.21	-1.2942177	0.2124343	-0.991315	8.74470235	0
R280G	64807180	VUS	Palm	NA	NA	33	0.906	0.62	2.49671202	0.3944151	-2.2433788	8.00568221	1
E295A	64807054	VUS	Palm	NA	NA	22.4	0.517	0.11	0.33027211	0.07589871	-0.543604	10.9576879	1
T297I	64807048	VUS	Palm	NA	NA	24.1	0.651	0.59	-0.8517007	0.09060615	-0.6289153	10.5756057	1
R300W	64807040	VUS	Palm	NA	NA	32	0.778	0.67	2.97845805	0.38619728	-3.5868874	10.9600348	1
P303S	64807031	NA	Palm	NA	NA	22.1	0.706	0.8	1.53673469	0.2606129	-1.0298725	4.61324679	1
P303Q	64807030	NA	Palm	NA	NA	25.3	0.843	1.31	1.17517007	0.48839069	-3.7705038	4.61324679	0
T305I	64807024	NA	NA	NA	NA	NA	0.505	0.21	-0.3414966	0.29289499	-1.0567071	8.71269766	0
R319Q	64806340	Con	Palm	NA	NA	20.4	0.37	0.24	-0.0465986	0.08940648	0.13906291	7.68172861	1
D320H	64806338	VUS	Palm	NA	NA	26.1	0.819	0.47	2.32596372	0.21135118	-3.9229724	8.68081726	0
Y326C	64806319	NA	Palm	NA	NA	28.2	0.794	2.19	4.69410431	0.36363679	-4.7009543	5.15227075	0
M327K	64806316	VUS	Palm	NA	NA	24.2	0.773	1.47	1.63673469	0.32901548	-4.9319265	5.89151394	0
Y332C	64806301	VUS	Palm	NA	NA	25.7	0.853	1	2.70839002	0.38713397	-2.5065146	8.3005039	1
R335C	64806293	VUS	Palm	NA	NA	32	0.959	0.1	2.76995465	0.44832101	-4.9047354	10.4535607	1
R335H	64806292	VUS	Palm	NA	NA	28.4	0.923	0.65	2.31746032	0.4820963	-4.9047354	10.4535607	1
R337C	64806287	VUS	Palm	NA	NA	25.8	0.863	0.26	1.64387755	0.22624004	-1.4240364	12.7312004	0
R340W	64806278	VUS	Palm	NA	NA	31	0.868	0.42	0.7537415	0.29360296	-3.0657384	11.0992851	1
A342P	64806272	LP	Palm	1	1	26.3	0.952	1.23	12.11530612	0.81646114	-5.4894553	4.75037845	0
Q354E	64806236	NA	Palm	NA	NA	23.9	0.769	0.95	1.98072562	0.2803958	-2.6506345	7.17408764	0
C359F	64805759	VUS	Palm	4	1	24	0.572	0.1	2.158276644	0.13211192	-0.866045	8.53592552	1
C359Y	64805759	VUS	Palm	NA	NA	24.1	0.628	0.35	3.85351474	0.28605515	-0.9274702	8.53592552	1
E363K	64805748	VUS	Palm[a]	NA	NA	31	0.854	0.29	-0.0591837	0.38793081	-4.7206226	10.2058873	1
E371D	64805722	Con	Palm	3	1	14.28	0.611	0.03	0.364965986	0.0547711	-0.2235004	10.2463327	1
V372I	64805721	Con	Palm	3	1	18.54	0.422	-0.76	-0.4159864	0.07678903	-0.3120171	5.84069485	1
V376I	64805709	VUS	Palm	NA	NA	22.5	0.573	-0.47	-0.0783447	0.15815699	-0.4647274	7.92777774	1

Variant	GRCh38	ClinVar	Domains	No. of people	No. of papers	PHRED	REVEL	PremPS	Rosetta	CPT	ESCOTT	SPDV	gnomAD (1=yes)
N379S	64805699	Con	Palm	NA	NA	21.9	0.468	0.32	-0.1229025	0.08471298	-0.7175739	10.1072199	1
A384T	64805685	VUS	Palm	NA	NA	23.2	0.463	0.46	2.82052154	0.10857907	-0.2449759	9.92310028	0
A390V	64805666	Con	Palm	3	2	24.1	0.608	0.26	1.457482993	0.1698457	-1.2250193	17.1774515	1
G391S	64805664	VUS	Palm	NA	NA	18.51	0.366	0.1	0.68095238	0.16543832	-0.0774329	17.4442419	0
R394W	64805655	Con	Fingers	NA	NA	26.1	0.532	0.11	0.48639456	0.21565851	-2.4926392	18.6618636	1
R394Q	64805654	Con	Fingers	NA	NA	24	0.534	0.31	0.82312925	0.18141872	-0.5178133	18.6618636	1
R394P	64805654	VUS	Fingers	NA	NA	23.5	0.443	0.45	0.89455782	0.13783096	-0.5910245	18.6618636	1
P395L	64805651	Con	Fingers	NA	NA	20.9	0.429	-0.05	0.90034014	0.10162868	-0.9243914	23.994405	1
G406V	64805182	VUS	Fingers	NA	NA	23	0.583	-0.03	-0.6146259	0.14561487	-0.067626	12.4953358	1
A408T	64805177	VUS	Fingers	1*	NA	25.5	0.729	0.11	-0.60759637	0.27681718	-2.2396713	7.39551412	1
P412L	64805164	NA	Fingers	NA	NA	28	0.883	0.39	2.01167801	0.34157436	-3.0704541	8.26982419	0
C414Y	64805158	VUS	Fingers	NA	NA	25.7	0.917	0.7	6.54603175	0.42977604	-3.347858	7.52029923	1
A416S	64805153	VUS	Fingers	NA	NA	25.9	0.824	0.54	1.46904762	0.26192586	-0.9730477	9.57152386	0
D423N	64805132	P	Fingers	33	13	27.3	0.891	0.71	1.959070295	0.42407245	-2.9864879	4.66157106	0
P433R	64805101	NA	Fingers	NA	NA	23.5	0.801	0.56	0.3829932	0.20510604	-2.8717908	9.98347304	0
V446M	64805063	NA	Fingers	NA	NA	25.8	0.83	0.42	1.59172336	0.37424919	-1.1866451	5.25039062	0
R457Q	64804812	Con	Fingers	NA	NA	28.3	0.949	1.14	1.02131519	0.3208329	-4.4953452	6.08647398	1
R457P	64804812	Con	Fingers	NA	NA	29	0.932	1.45	9.96689342	0.44042779	-4.4953452	6.08647398	1
S464N	64804791	VUS	Fingers	NA	NA	19.69	0.496	0.34	0.09557823	0.21528424	-1.2313234	10.6919137	1
R465P	64804788	VUS	Fingers	NA	NA	23.4	0.623	0.15	-0.4781179	0.14009341	-1.0801453	11.6660804	0
E468Q	64804780	VUS	Fingers	NA	NA	22.9	0.634	0.1	-0.2784581	0.20904346	-0.7597576	16.8806332	1
E471V	64804770	VUS	Fingers	NA	NA	23.2	0.484	0.23	0.47687075	0.14532357	-1.2509547	21.6442716	1
P475L	64804758	Con	Fingers	1*	NA	19.17	0.426	0.16	0.802040816	0.08537682	-0.1505785	30.7473644	1
G477D	64804752	NA	Fingers	NA	NA	19.73	0.468	0.49	0.04659864	0.22999202	-0.7671059	37.0587948	0
E478K	64804750	NA	Fingers	NA	NA	22.7	0.56	-0.03	0.88061225	0.21370869	-0.4515907	39.0949962	1
E482D	64804736	VUS	Fingers	NA	NA	22.7	0.438	0.02	0.47857143	0.15153922	-0.1949686	51.6074817	1

Variant	GRCh38	ClinVar	Domains	No. of people	No. of papers	PHRED	REVEL	PremPS	Rosetta	CPT	ESCOTT	SPDV	gnomAD (1=yes)
R486W	64804726	VUS	Fingers[b]	NA	NA	27.2	0.749	0.42	0.11904762	0.242534	-2.4515399	62.2944025	0
R490Q	64804713	VUS	Fingers[b]	NA	NA	24.1	0.622	0.26	-0.1666667	0.2400709	-1.0164045	64.583455	0
P497L	64804692	VUS	Fingers[b]	NA	NA	22.8	0.447	0.06	-0.372449	0.16578017	-0.5939184	54.208404	1
P498S	64804690	VUS	Fingers[b]	NA	NA	20.8	0.453	0.49	0.69285714	0.11566336	-0.1136014	51.546115	1
P498Q	64804689	NA	Fingers[b]	NA	NA	23.4	0.418	0.57	0.75895692	0.13400566	-0.1467769	51.546115	1
P499L	64804686	VUS	Fingers[b]	NA	NA	21.7	0.402	0.05	1.86167801	0.07909462	-0.1005189	50.7582031	1
K507N	64804661	Con	Fingers	NA	NA	18.86	0.306	0.38	0.43129252	0.18013156	-0.3670303	47.9440183	1
G508D	64804659	LB	Fingers	NA	NA	12.68	0.73	0.06	-0.5098639	0.22467004	-0.3149773	47.375448	1
T511P	64804651	VUS	Fingers	NA	NA	2.132	0.272	-0.01	2.17959184	0.11865596	-0.0374065	44.4790504	1
Q513K	64804645	Con	Fingers	2*	NA	14.86	0.554	0.05	0.332312925	0.14161653	-0.0868667	40.7171874	1
S517L	64804632	Con	Fingers	2*	NA	17.23	0.36	0	1.107482993	0.16303057	-0.0754262	33.4044424	1
P519T	64804627	Con	Fingers	NA	NA	12.22	0.321	0.01	0.41530612	0.12865396	-0.0705319	29.5886783	1
P519R	64804626	VUS	Fingers	NA	NA	15.95	0.276	0.31	1.11802721	0.14119793	-0.0677452	29.5886783	1
P519L	64804626	VUS	Fingers	NA	NA	17.96	0.356	-0.01	0.46734694	0.07125752	-0.077247	29.5886783	1
P520S	64804624	VUS	Fingers	NA	NA	10.78	0.27	-0.02	0.4037415	0.08965823	0.0345337	28.1743172	1
R521G	64804621	Con	Fingers	NA	NA	23.5	0.697	0.18	0.81666667	0.13083438	-0.2105469	31.5599568	1
R521W	64804621	VUS	Fingers	NA	NA	13.13	0.698	0.38	0.10952381	0.19638968	-0.9942481	31.5599568	1
R521Q	64804620	VUS	Fingers	1*	NA	17.34	0.384	0.08	0.263605442	0.18860859	-0.3492075	31.5599568	1
P524L	64804611	NA	Fingers	NA	NA	17.37	0.344	0.03	-0.3506803	0.09424986	-0.0006193	27.2227197	1
G525R	64804609	Con	Fingers	NA	NA	14.09	0.4	0.07	5.47029479	0.11162142	-0.1834548	26.3863516	1
T526S	64804605	VUS	Fingers	NA	NA	1.456	0.227	0.19	1.02142857	0.14251035	-0.1447207	25.4630464	0
V527L	64804603	VUS	Fingers	NA	NA	5.085	0.29	0.13	0.07857143	0.1281906	-0.1128183	28.8762598	1
A528P	64804600	Con	Fingers	1*	NA	4.847	0.282	-0.05	-0.95544218	0.00438925	-0.0350654	29.4007841	1
A528T	64804600	Con	Fingers	NA	NA	6.374	0.258	0.08	0.27346939	0.17055345	-0.1852145	29.4007841	1
G537C	64804573	Con	Fingers	4	3	15.64	0.67	0.27	4.846938776	0.20562011	-0.529555	40.2522763	1
S538G	64804570	VUS	Fingers	NA	NA	9.289	0.287	0.21	0.12108844	0.0635452	-0.0259321	37.8796499	1

Variant	GRCh38	ClinVar	Domains	No. of people	No. of papers	PHRED	REVEL	PremPS	Rosetta	CPT	ESCOTT	SPDV	gnomAD (1=yes)
Q541R	64804560	VUS	Fingers	NA	NA	9.524	0.398	0.13	-0.0986395	0.12625455	-0.1230309	38.7038544	1
V542M	64804558	Con	Fingers	NA	NA	8.412	0.352	0	0.23503401	0.20285933	-0.199644	33.8068526	0
P545S	64804549	Con	Fingers	7	1	21.4	0.69	0.33	0.392517007	0.10112655	-0.1383942	30.5667239	1
P549S	64804537	VUS	Fingers	2	1	22.1	0.696	0.46	0.928231293	0.08939171	-0.333587	24.4219585	1
E552K	64804528	VUS	Fingers[c]	NA	NA	16.87	0.488	-0.51	0.6329932	0.15805529	-0.7238734	17.3569187	1
E552G	64804527	LB	Fingers[c]	NA	NA	21.1	0.47	0.04	0.84115646	0.07491244	-0.4522911	17.3569187	1
V555L	64804519	Con	Fingers[c]	1*	NA	19.97	0.709	0.17	1.930272109	0.16010049	-0.3535194	13.6234764	1
V555A	64804518	VUS	Fingers[c]	NA	NA	22.4	0.599	0.72	1.61712018	0.24015179	-0.9701167	13.6234764	0
T573S	64804465	NA	Fingers[c]	NA	NA	20.8	0.392	-0.08	-0.4205215	0.13289579	-0.642025	11.017349	1
S588A	64804420	VUS	Fingers	NA	NA	24.4	0.846	0.27	0.12131519	0.16904766	-1.734277	18.4919818	0
Q595R	64804398	NA	Fingers[d]	NA	NA	20.9	0.43	-0.4	1.09727891	0.08267881	-0.8777121	30.4329499	0
S601G	64804381	NA	Fingers[d]	NA	NA	22	0.465	0.36	0.97199547	0.0740705	-0.2336951	20.3713393	0
R612C	64804348	VUS	Fingers[d]	NA	NA	32	0.82	0.25	0.37789116	0.19415138	-2.83827	26.4504203	1

[a]MLL binding, [b]NLS1, [c]NLSa, [d]NLS2

*At least one person reported; § Reported in UMD-MEN1 with no detailed clinical information.

Conflicting classifications of pathogenicity (Con); Likely benign (LB); Likely pathogenic (LP); Not available (NA); Number (No.); Pathogenic (P); Variant of uncertain significance (VUS).

Most patients (95%; 656/693) reported to be of White background; 3% (20/693) reported a Black or African American background; and <1% each reported a background of Asian, Native Hawaiian or Other Pacific Islander, American Indian or Alaska Native or were either not available or declined to report **(Table 26)**.

Table 26: The reported background and ethnicities of the 693 patients with *MEN1* missense variants in the Geisinger cohort.

Background	Hispanic or Latino ethnicity	Not Hispanic or Latino ethnicity	Hispanic or Latino ethnicity NR	Total
White	24	621	11	656
Black or African American	0	20	0	20
Asian	0	6	0	6
Native Hawaiian or Other Pacific Islander	2	1	0	3
American Indian or Alaska Native	0	2	0	2
Not reported or declined to report	2	0	4	6
Total	28	650	15	693

Not reported (NR).

At least one ICD code that was enriched in patients with *MEN1* PTVs was found in 6.3% (44/693) patients with missense *MEN1* variants **(Table 27)**.

Table 27: ICD codes enriched in protein truncating *MEN1* variants, found in patients with *MEN1* missense variants reported in the literature to cause MEN1

Variant (n=26)	Age/ Sex (n=44)	Carrier count	Max calcium[#]	Max PTH[§]	ICD codes
ACMG LPV					
C246Y	25F	1	11	78	Hyperparathyroidism and other disorders of parathyroid gland Benign neuroendocrine tumors Polyglandular dysfunction
A342P	42F	1	11.4	71	Polyglandular dysfunction Secondary neuroendocrine tumors Malignant neoplasm of pancreas Hyperparathyroidism and other disorders of parathyroid gland
D423N	40M[a]	3	10.3	NA	Malignant neoplasm of pancreas Polyglandular dysfunction Malignant neuroendocrine tumors Benign neuroendocrine tumors
	40F[a+]		12.6	120	Polyglandular dysfunction

	47F[a]		10.7	147	Polyglandular dysfunction Hyperparathyroidism and other disorders of parathyroid gland
ACMG VUS					
P12L	58F	1	9.9	NA	Hyperparathyroidism and other disorders of parathyroid gland
C359F	64F	2	10.1	85	Other specified diabetes mellitus
G537C	58F	15	10.1	23	Other specified diabetes mellitus
P545S	83F	17	9.8	135	Neoplasm of uncertain behavior of endocrine glands Other specified diabetes mellitus
UMD-MEN1					
P193L	58F	32	10.1	39	Hyperparathyroidism and other disorders of parathyroid gland
	67M		9.6	NA	Duodenal ulcer
Q513K	46F	28	9.9	36	Polyglandular dysfunction
	66M		10.3	120	Hyperparathyroidism and other disorders of parathyroid gland
V555L	85F	2	11.2	154	Hyperparathyroidism and other disorders of parathyroid gland
gnomAD					
S15F	75F	19	11.3	83	Neoplasm of uncertain behavior of endocrine glands Hyperparathyroidism and other disorders of parathyroid gland
	30F		9.4	35	Benign neuroendocrine tumors
	68F		8	295	Hyperparathyroidism and other disorders of parathyroid gland
L275P	61F	28	10.4	59	Other specified diabetes mellitus
	85M		9.7	NA	Other specified diabetes mellitus
E295A	74F	16	10.4	59	Duodenal ulcer
R335H	83F	2	10.2	NA	Duodenal ulcer
P499L	51F	25	NA	196	Hyperparathyroidism and other disorders of parathyroid gland
	66M		9.9	NA	Benign neuroendocrine tumors
G508D	64F	16	10.8	13	Hyperparathyroidism and other disorders of parathyroid gland
P519R	74F	6	10.2	45	Hyperparathyroidism and other disorders of parathyroid gland
Q541R	55M	3	10.4	211	Duodenal ulcer Hyperparathyroidism and other disorders of parathyroid gland
Unreported					

R137P	43M	9	9.8	26	Other specified diabetes mellitus
P193L	58F	32	10.1	39	Hyperparathyroidism and other disorders of parathyroid gland
	67M		9.6	NA	Duodenal ulcer
G219S	81M	7	10.5	71	Duodenal ulcer Hyperparathyroidism and other disorders of parathyroid gland
A416S	74F	3	11.1	48	Hyperparathyroidism and other disorders of parathyroid gland
V446M	63F	2	11.1	101	Hyperparathyroidism and other disorders of parathyroid gland
R465P	42F	81	10	0	Cushing's syndrome
	60F		9.8	75	Hyperparathyroidism and other disorders of parathyroid gland
	61M		9.9	37	Other specified diabetes mellitus
	65F		11.6	142	Hyperparathyroidism and other disorders of parathyroid gland
	73F		11	110	Duodenal ulcer
	80M		9.8	41	Malignant neoplasm of liver and intrahepatic bile ducts
	81F		11.3	103	Hyperparathyroidism and other disorders of parathyroid gland
	85M		10.4	94	Duodenal ulcer
	>90F		10.3	114	Hyperparathyroidism and other disorders of parathyroid gland
Q513K	46F	28	9.9	36	Polyglandular dysfunction
	66M		10.3	120	Hyperparathyroidism and other disorders of parathyroid gland
V555A	51M	5	10.2	79	Hyperparathyroidism and other disorders of parathyroid gland

[a]From one family; [+]Maximum serum calcium >11.2mg/dL; [§] Calcium reference range: (8.4-10.2mg/dL); [#]PT reference range: (15-65 pg/mL). American college of medical genetics (ACMG); Female (F); International Classification of Disease (ICD); Likely pathogenic variant (LPV); Male (M); Not available (NA); The Universal Mutation Database (UMD); Variant of uncertain significance (VUS).

The individual ICD code for PHPT and serum calcium levels were plotted against patient age **(Figure 31)**. At least one serum calcium measurement was available in 94% (650/693) of patients, with 24% (154/650) having at least one measurement above the upper limit of normal (>10.2mg/dL (i.e. >2.55mmol/L)).

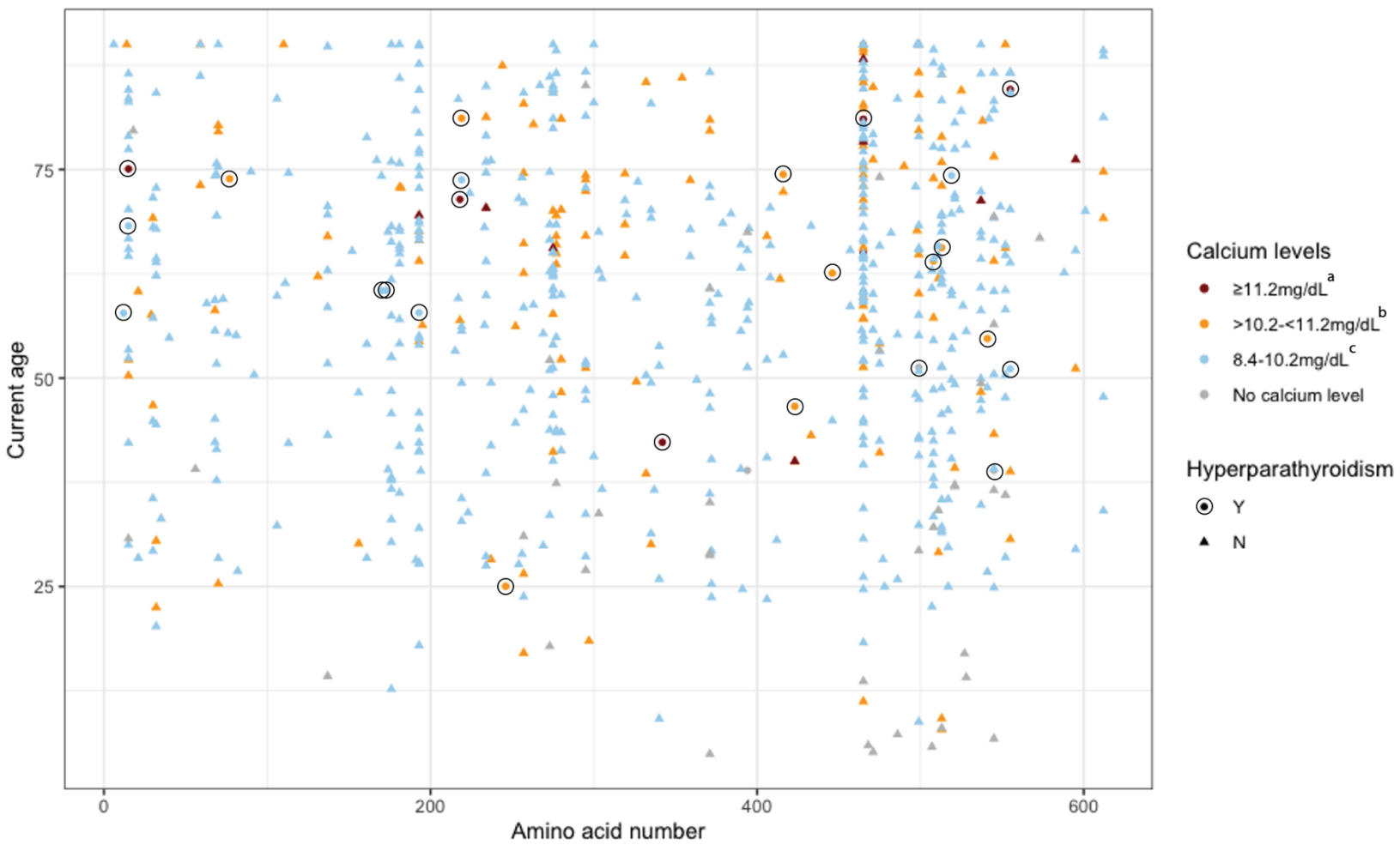


Figure 31: Missense *MEN1* variants plotted against current age. Target symbol indicates an International Classification of Disease (ICD) code for primary hyperpararthyroidism (PHPT) and was present in 3% (24/693) of patients. A normal calcium level (blue) and no ICD code for PHPT was seen in 71% (489/693) patients.

[a](≥2.8mmol/L); [b](>2.55-<2.8mmol/L); [c](2.1-2.55mmol/L).

An ICD code for PHPT was diagnosed in 3% (24/693). Fourteen patients with *MEN1* missense variants had a maximum serum calcium level >11.2mg/dL (including 6 patients with already diagnosed PHPT by ICD code) **(Figure 31)**. Therefore, a calcium phenotype may have been present in an additional 8 patients (7/8 of these patients had non-suppressed parathyroid hormone (PTH) levels and one patient had no PTH level taken). Most patients with a *MEN1* missense variant did not exhibit a calcium phenotype. PHPT and hypercalcaemia were seen in patients with *MEN1* missense mutations scattered throughout *MEN1* and did not cluster at a particular location **(Figures 31 and 32).** The missense variants with a diagnosis of PHPT included P12L, S15F (x2), Y77H, P193L, A218D, G219S (x2), C246Y, A342P, D423N, V446M, R465P (x4), P499L, G508D, Q513K, P519R, Q541R, V555A and V555L.

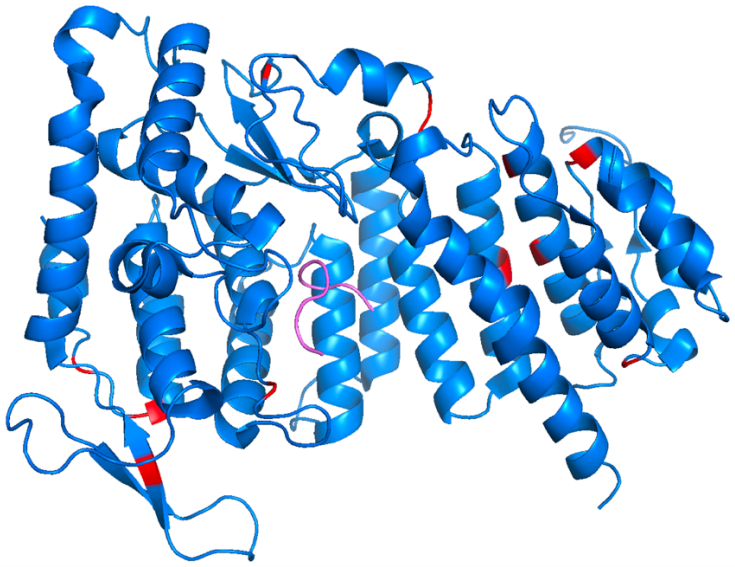


Figure 32: Reports the menin protein (blue) in complex with MLL (pink) (PDB: 3u85). The locations of the *MEN1* missense mutations in the Geisinger cohort with PHPT are shown in red.

4.1.2.5.3 Enriched ICD codes in missense variants

Out of the ICD codes that were enriched in patients with *MEN1* PTVs, ≥1 of these ICD codes were present in 6% (44/693) of patients with *MEN1* missense variants (4 ICD codes (n=2), 3

ICD codes (n=1), 2 ICD codes (n=5) and 1 ICD code (n=36)). There were 168 unique *MEN1* missense variants reported in the Geisinger cohort **(Figure 33)**. Out of these, only three variants (C246Y, 342P and D423N) were associated with clinical MEN1 based on their ICD codes and defined as ≥2 MEN1 phenotypes in one individual or a first-degree relative with MEN1 and the individual ≥1 MEN1 phenotype. All three variants are reported in the literature as LPV (ACMG criteria). Out of the 165 unique *MEN1* missense variants reported in Geisinger with no clinical diagnosis of MEN1 based on ICD codes, 88 of these variants have been reported in gnomAD, and two as LPV based on ACMG criteria.

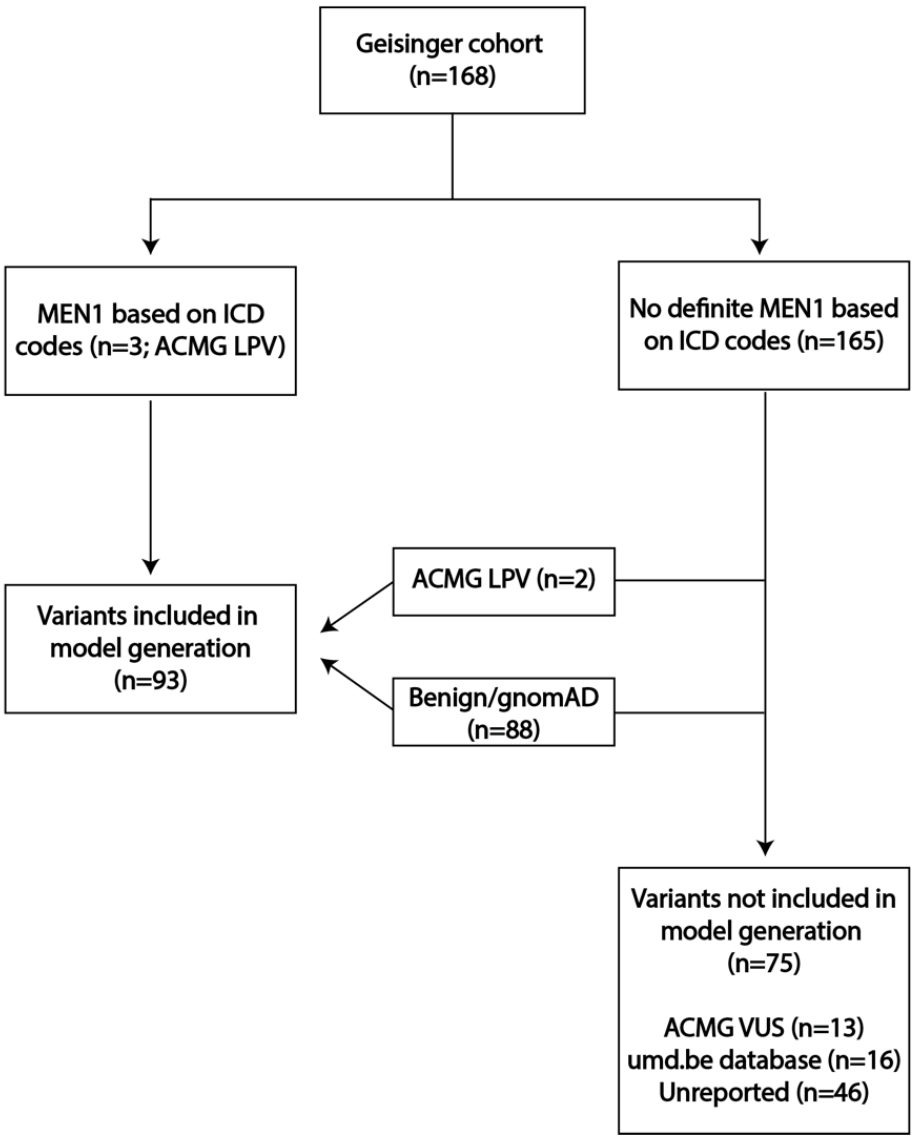


Figure 33: Flowchart of *MEN1* missense variants in the Geisinger cohort. International Classification of Diseases (ICD); American college of medical genetics (ACMG); Likely pathogenic variant (LPV); Variant of uncertain significance (VUS).

These previously reported *MEN1* SNVs had been included in the BV and LPV cohorts to generate the cut-off levels for each bioinformatic VEP. In total, there were 75 *MEN1* missense variants in the Geisinger cohort that were not used to generate cut-off levels for each VEP and were subsequently used in model validation.

4.1.2.5.4 Prevalence of MEN1 in the Geisinger population

A total of 23 patients had a genetic diagnosis of MEN1, thereby yielding a prevalence of MEN1 in the Geisinger population of ~0.01% (23/175,000). People with MEN1 were from 17 families of white background and self-reported of non-Hispanic or Latino ethnicity. This included 18 people with *MEN1* PTVs and either PHPT (17/18) or maximum serum calcium ≥11.2mg/dL. The prevalence of MEN1 in patients with a missense *MEN1* mutation in the Geisinger cohort was 0.7-1% (5/693 – 5/476), with 3 out of these five people from one family with the same *MEN1* mutation (D423N). The average age of people with MEN1 due to a *MEN1* missense variant in the Geisinger population is 38.8 years, compared to 59 years in the remaining cohort with *MEN1* missense mutations.

4.1.2.6 Bioinformatic tools to assess MEN1 missense mutations

4.1.2.6.1 Validation set

In the Geisinger cohort there were 168 unique *MEN1* variants, with 3 *MEN1* missense variants (D423N, C246Y, and A342P from 5 people) as the positive cases and 165 *MEN1* missense variants with no phenotype as the negative cases (control group). In the Geisinger cohort, there were 75 *MEN1* SNVs from patients in the control group who were not used to determine the cut-off levels for each tool. Using a sensitivity cut-off of 98.7%, Models 1 and 2 were the most specific to rule out pathogenic SNVs at 88% and 85.33%, respectively **(Figure 34)**.

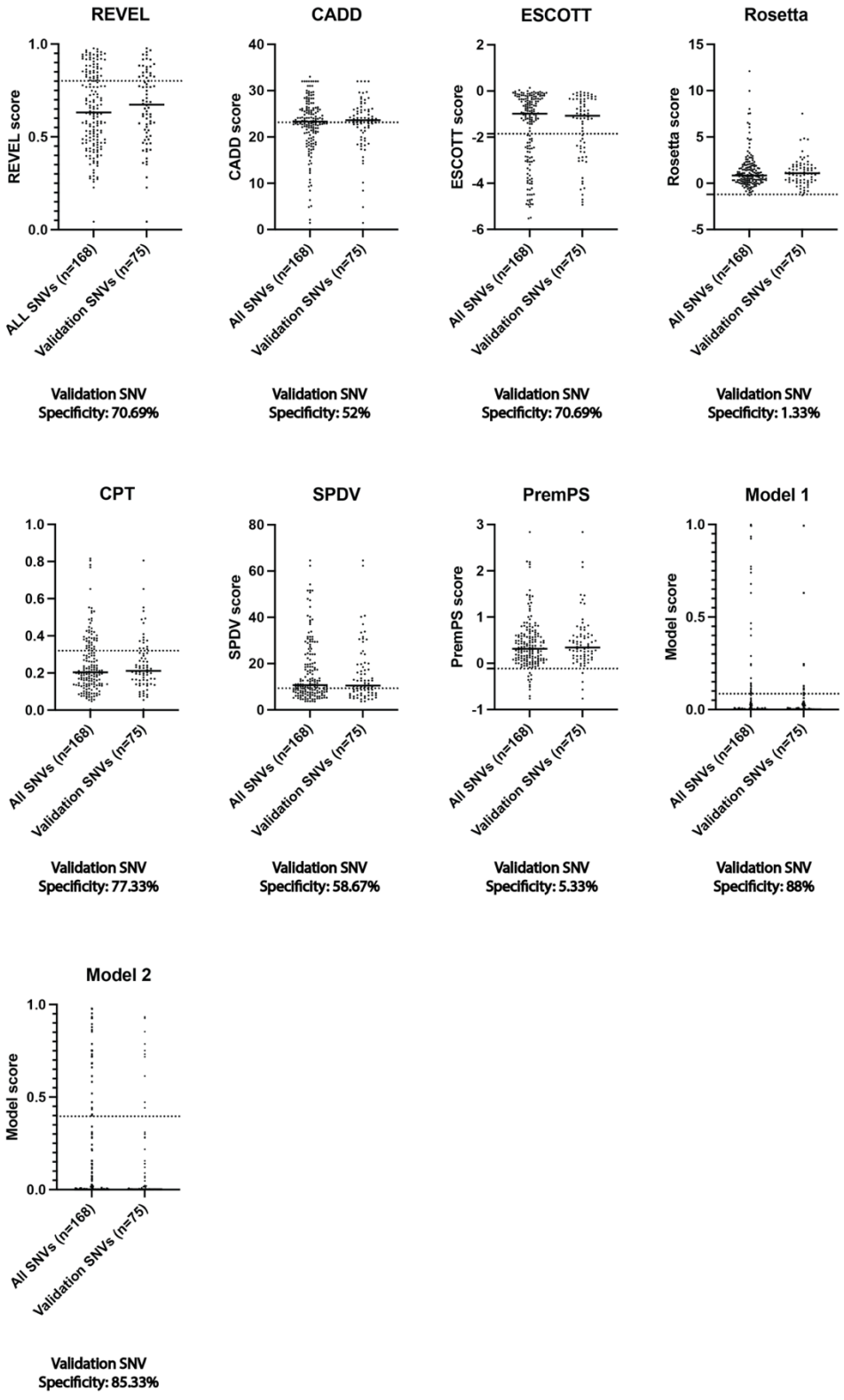


Figure 34: 98.7% cut-off levels for each bioinformatic tool and Models 1 and 2 for all *MEN1* SNVs and the validation cohort of *MEN1* SNVs in the Geisinger cohort.

All models correctly classified the 3 pathogenic variants in the Geisinger cohort, which is not unsurprising as these *MEN1* variants were part of generating the original cut-off levels. Each VEP had variable specificity ranging from relatively high for CPT (77.33%), ESCOTT (70.69%) and REVEL (70.69%) to very low (e.g. Rosetta (1.33%) and PremPS (5.33%)). Both MEN1 models outperformed the other VEPs using the benign validation Geisinger cohort; this is seen with a high specificity of 88% and 85.33% for Models 1 and 2, respectively. Therefore, combining individual VEPs can improve the accuracy of classifying *MEN1* variants.

4.1.2.7 MEN1 missense variant look-up table with combined VEP pathogenicity score (Model 2)

All *MEN1* missense variants that could occur due to an SNV were initially downloaded with associated REVEL scores (n=4028; (https://sites.google.com/site/revelgenomics/about) on the 3rd of May 2021). Each *MEN1* SNV was then annotated with information from ClinVar, individual VEP and Model scores, protein structural features and information regarding published cases of MEN1. An example of the information available for each of the *MEN1* missense variable is shown in **Table 28.** This resource will provide clinicians with a *MEN1* missense look-up table to aid in variant classification. .

Table 28: An example of a subset of information available for each of the 4028 unique *MEN1* missense variants

Variant	GRCh38	number of people	Number of papers	PHRED	REVEL	PremPS	Location	Rosetta	CPT	ESCOTT	SPDV	Clinical significance (Last reviewed)	Domains	Model 2 score
G2R	64810106	0	0	29.3	0.878	0.68	SUR	-0.3605442	0.30401896	-2.9816395	11.8268954	NA	NTD	0.07093522
G2W	64810106	0	0	31	0.894	0.41	SUR	0.12244898	0.26198847	-2.9816395	11.8268954	NA	NTD	0.08140922
G2A	64810105	0	0	23	0.76	0.3	SUR	-0.864966	0.10939101	-1.1409517	11.8268954	NA	NTD	0.00424118
G2E	64810105	0	0	28	0.901	0.68	SUR	-0.75	0.23635209	-2.9816395	11.8268954	Uncertain significance (Last reviewed: Aug 30, 2021)	NTD	0.08642102
G2V	64810105	0	0	27.6	0.937	0.19	SUR	-0.4857143	0.22203877	-2.9816395	11.8268954	NA	NTD	0.11683435
L3M	64810103	0	0	26.4	0.565	0.76	COR	0.43639456	0.19731299	-0.8888181	10.5714239	NA	NTD	0.00095139
L3V	64810103	0	0	24	0.547	0.7	COR	1.6776644	0.21768789	-2.2890959	10.5714239	NA	NTD	0.00312578
L3P	64810102	0	0	32	0.83	1.63	COR	2.97891157	0.26258427	-3.673324	10.5714239	NA	NTD	0.14373038
L3Q	64810102	0	0	31	0.739	1.62	COR	2.82006803	0.27644168	-3.673324	10.5714239	NA	NTD	0.06708082
….	….	….	….	….	….	….	….	….	….	….	….	….	….	….
G42C	64809986	1	1	33	0.948	0.22	COR	4.39489796	0.49761727	-5.456203	4.51793708	NA	NTD	0.97729691
G42R	64809986	7	2	33	0.954	0.82	COR	8.06609977	0.65282569	-5.456203	4.51793708	Likely pathogenic (Last reviewed: Nov 24, 2021)	NTD	0.9785046
G42S	64809986	3	2	33	0.935	0.81	COR	5.51122449	0.51813118	-5.456203	4.51793708	Conflicting interpretations of pathogenicity (Last reviewed: Sep 1, 2021)	NTD	0.97444858
G42V	64809985	2	2	33	0.937	-0.13	COR	6.04727891	0.59962932	-5.456203	4.51793708	Likely pathogenic (Last reviewed: Jan 23, 2018)	NTD	0.97490842
G42A	64809985	1	1	32	0.92	0.71	COR	3.30816327	0.4759802	-5.456203	4.51793708	Pathogenic (Last reviewed: Sep 20, 2022)	NTD	0.97072859
G42D	64809985	8	2	33	0.911	1.24	COR	6.61836735	0.63609	-5.456203	4.51793708	NA	NTD	0.96824976

N-terminal domain (NTD).

4.1.3 Discussion

The results of this chapter have shown that by combining VEPs, a novel, more accurate *MEN1* missense mutation variant predictor can be generated. Additionally, this new model can provide VEP scores for novel *MEN1* missense variants, and clinicians can choose which cut-off level they deem acceptable based on the sensitivity and/or specificity. This is vitally important, as with the recent rise in the availability of high-throughput DNA sequencing, clinicians are faced with increasing diagnostic uncertainty regarding *MEN1* missense variants. At present, to determine if a *MEN1* missense variant is pathogenic, clinicians rely on published case reports of patients with MEN1 with a particular *MEN1* missense variant, bioinformatic scores (e.g. REVEL; however, a *MEN1* gene-specific cut-off has not been reported), and annotated databases e.g. ClinVar. ClinVar is a publicly available database with variants submitted for ~331,000 from 5 continents,[239] and there have been ongoing improvements in the available phenotypic information and classification of variants; however, the database does include misclassified variants.[240] However, for *MEN1*, most of these missense variants in ClinVar are classified as VUS,[239] which does not help with clinical decision-making e.g. to enrol a patient in an intensive endocrine tumour screening program. To compare and then improve upon current *MEN1* VEPs, first a solid dataset must be generated, with both pathogenic and benign variants. To generate a robust pathogenic dataset, the literature was reviewed, and this identified a total of 319 unique missense *MEN1* SNVs, and out of these, 74 were LPVs as defined by ACMG criteria. For comparison, variants submitted to gnomAD (and not present in any publication of a patient with documented MEN1 or any NET (germline or sporadic)) were used as the benign cohort. Although there will be some variants in the 'putatively benign' dataset that will be misclassified, this is likely to be negligible and the use of variants published

in gnomAD as a comparison/benign dataset has previously been reported.[236] Importantly, these cut-off levels for each individual VEP and the new combined model (Model 2) that has been generated were based on LPV/LP *MEN1* variants, with all *MEN1* VUS removed, therefore clinical judgement must still be utilised. For example, if a patient presents with a *MEN1* SNV that strongly co-segregates with MEN1 in the family, regardless of bioinformatic VEPs, this patient will be enrolled into a MEN1 screening program. Genetic modifiers may also play a role in patients with *MEN1* missense variants of predicted low pathogenicity.[241]

All bioinformatic tools assessed significantly distinguished *MEN1* BVs reported in gnomAD from MEN1-causing variants reported in the literature (LPV); however, only REVEL, ESCOTT, SPDV and CPT were able to distinguish between possible and definite-causing *MEN1* missense variants, suggesting that these are more precise variant classifiers. ESCOTT includes information regarding the location of the mutated amino acid (e.g surface versus core) in its pathogenicity score, and independent of the ESCOTT score, it has also been shown that pathogenic *MEN1* mutations are localised within the core of the MEN1 protein when compared to all reported amino acid mutations in gnomAD. It has been reported that SNVs within the core are likely to destabilise the protein and thereby increase its degradation/decrease its function.[242] Whereas SNVs located in amino acids on the surface (or solvent-accessible regions) are more likely to disrupt ligand binding. The SPDV scores were generated by Dr Lukas Gerasimavicius, through a collaboration with the University of Edinburgh, using AlphaFold 2 protein models to derive Euclidean proximity scores for variants of interest (unpublished methodology). AlphaFold 2 protein models were also used to derive thermostability scores using the PremPS method.[191] The amino acid location in menin's 3D structure is not included in the commonly used REVEL score. The REVEL score has previously

been shown to be superior to other bioinformatic tools,[234] however, newer methods to classify missense variants are emerging, for example cross-protein transfer (CPT).[206] CPT is a newer method to score variants based on data trained on functional assay data from a subset of proteins and extrapolated to other proteins and has been reported to be superior to REVEL in predicting SNVs.[206] Therefore, these newer VEPs were tested against the more commonly used REVEL and CADD scores.

In this study, using *MEN1* variants reported in gnomAD as the BV dataset (n=463) we report that a REVEL score of ≤0.7455 rules out MEN1 with 100% sensitivity, and a REVEL cut-off score of >0.8545 optimally classifies pathogenic vs benign mutations (accuracy of 82.68%). Cut-off levels for the remaining bioinformatic tools were undertaken using this method, which reported that CPT was the most accurate at 88.27% and had the highest PPV in distinguishing benign vs pathogenic *MEN1* mutations. Different combinations were modelled to try and incorporate both the sensitivity of REVEL and the specificity of the other newer VEPs. Two Models were constructed: Model 1 showed apparent superior accuracy over Model 2; however, this was likely due to model overfitting and correlation between variant tools. Additionally, the AIC was significantly lower for Model 2 compared to Model 1, due to the lower number of variables within the model and will likely be more generalisable when applied to other *MEN1* variants. Therefore, Model 2 is most likely to have a greater accuracy in classifying unknown *MEN1* variants. To validate Model 2 and newer bioinformatic tools, a separate cohort from Geisinger, a biobank in the United States, was reviewed.

There are >20 major biobanks globally, which collect and 'bank' information on people (30,000 up to >500,000) and may include access to biospecimens (e.g. blood, urine, saliva), genetic information and phenotypic information.[243] The Geisinger cohort is one of these major biobanks

and comprises ~175,000 individuals who have attended a medical facility in the United States. Each patient has had their genome sequenced, demographics collected and access to biochemistry and phenotypic information is available in the form of ICD codes. Compared to most studies looking at *MEN1* mutations causing MEN1, which are confounded by ascertainment bias, the Geisinger cohort provides a relatively unbiased population selection. The reported estimated prevalence of MEN1 in the population is 1 in 5,000 up to 1 in 40,000.[244,245] Our study reports that the population prevalence of genetic MEN1 in the Geisinger cohort as ~0.01% or 1 in 10,000, with ~1/5 of these attributable to a missense mutation, which is similar to that reported in the literature.[66] Approximately 0.4% of patients in the Geisinger cohort had a non-synonymous missense mutation in the protein-coding region of *MEN1,* which was one of 168 unique *MEN1* mutations. Only three of these mutations (C246Y, 342P and D423N; present in 5 individuals) had ICD codes consistent with a diagnosis of MEN1. This result was unexpected, as *MEN1* is a highly conserved gene, it was hypothesised that most patients with a *MEN1* mutation would have a phenotype. Out of the 693 people in the Geisinger cohort with a *MEN1* missense variant, most (even those who carry a mutation previously identified in patients with MEN1) had either normal serum calcium levels and/or no ICD consistent with MEN1. Therefore, *MEN1* missense mutations are either less penetrant than previously thought; they may be influenced by genetic modifiers; people in the Geisinger cohort may have subclinical disease, or a higher proportion of *MEN1* missense mutations may be benign. This provides important information for clinicians to aid in counselling patients who are asymptomatic, with no family history of MEN1, and who carry a *MEN1* missense mutation, as clearly, not all these patients require enrolment into an intensive endocrine tumour screening program. However, what should be considered is the cohort ethnicity studied, as >90% of the

Geisinger cohort is of reported white background.[187] Additionally, most LPV/PV *MEN1* missense variants identified in large registries were from European countries and therefore, this novel VEP, was predominantly trained on and validated on patients of white background and may not be as applicable to other patient ethnicities.

Intuitively, it would be expected that initiator codon variants would cause a phenotype, similar to other PTVs, which was not seen in our cohort.[246] Most commonly, translation starts at the AUG start codon, which usually resides within the Kozak sequence (GCCACC**A**UGG; the underlined codons -3 (A) and +4 (G) positions have the strongest effect on translation, with the start of the AUG marked as +1 (bold)).[246] In the Geisinger cohort, no patients had -3 (A) or +4 (G) missense mutations. However, two patients had SNVs at the start codon (i.e. initiator codon variants), at +2 position (T>C) and +3 (G>A), altering the start amino acid Methionine (Met) to Threonine (Thr) and Isoleucine (Ile), respectively. Both patients with initiator codon variants had normal serum calcium levels and no ICD codes that were enriched in patients with *MEN1* PTVs, which may be due to age-related penetrance of MEN1 and this may be consistent with reports of initiator codon variants being associated with a protracted disease phenotype.[247] Of note, there are no close in frame ATG sequences downstream of the initiator codon in *MEN1*, with the nearest ATG DNA sequence at codon 228 (of the 615 amino acid protein).

One of the most striking features in the *MEN1* missense mutation cohort was that the prevalence of PHPT was low, at only 3% compared to 100% with *MEN1* PTVs, suggesting that most of these *MEN1* missense variants were either benign or non-penetrant. This finding was unexpected, as *MEN1* is a highly conserved tumour suppressor in endocrine tissue, and it was hypothesised that missense mutations in this gene would be causative of disease. On review of

the literature, one study compared *MEN1* somatic mutations in different tumours to germline mutations in ClinVar, and reported that most *MEN1* germline missense mutations were less likely to be pathogenic/drivers of neoplasia as they were not seen frequently in somatic tumours.[248] However, this can be explained by the tissue-specific function of menin. Menin acts as a TSG in endocrine tumours; however, it can indirectly be a driver of malignancy in specific types of leukaemia, for example, in MLL fusion leukaemia, where *MLL* has been translocated to a different part of the genome, excess translated MLL is able to bind with menin, and lead to a deleterious increase in the active histone mark H3K4me3 and subsequent increase in *HOX* gene expression.[249] Additionally, it is not clear if *MEN1* SNVs in somatic tumours of different organs are in fact, drivers of tumourigenesis or passengers. It has been reported that ~half of the genetic mutations driving cancer are undiscovered.[250] Our study provides evidence that most *MEN1* missense mutations do not carry a phenotype typical of MEN1.

Conclusion: Overall, the key findings from this chapter include: 1) In the absence of clinical features suggestive of MEN1, an incidental finding of a *MEN1* missense mutation (as opposed to a *MEN1* PTVs) does not necessarily mean that all of these patients should be enrolled in an intensive MEN1 screening program, as almost all of these mutations have either low penetrance or are associated with a milder clinical phenotype and 2) A more accurate novel *MEN1* pathogenicity model using multiple VEPs has been generated to classify *MEN1* missense variants, including those not currently reported and will aid in clinical decision-making.

4.2 Chapter 2: Investigating global DNA 5hmC in normal pancreatic tissue and PNETs by immunohistochemistry

4.2.1 Introduction

Genetic testing may be useful for patients who present with an MEN1 phenotype or in a subset of patients who have PNETs but do not have other MEN1-associated tumours. In addition to genetic testing, analysis of the tumour itself may be useful to confirm the diagnosis and provide guidance regarding prognosis and treatment. It is known that menin plays a role in epigenetic regulation, including DNA methylation (5mC) in endocrine tissue.[75,92,93] Loss of menin leads to increased DNA 5mC compared to tumours with a functional menin.[75,92,93] In PNETs, there have been many studies which have either investigated DNA methylation at specific genes[75,76,107-109,114-138] or globally.[110] However, only one study has investigated DNA hydyroxymethylation (5hmC)[151] and it has been reported that DNA 5mC and 5hmC have opposing functions. It has been reported that normal pancreatic islets stain strongly for 5mC on IHC, and although not explicitly stated in the paper, published IHC images of two normal pancreatic tissue samples show comparable 5mC in normal pancreatic exocrine tissue.[110] Additionally, global DNA 5hmC has been reported to increase with beta cell maturation with a concomitant decrease in DNA 5mC.[251] DNA 5hmC can be measured using a polyclonal antibody from Active Motif antibody which is reported to be specific for DNA 5hmC (not DNA 5mC). This 5hmC antibody has been used to investigate the effects of vitamin C on DNA 5hmC levels in bladder cancer cell lines,[252] and reported that DNA 5hmC significantly increased and DNA 5mC levels slightly decreased with vitamin C treatment.[252] This antibody has also been used to study DNA 5hmC in renal cell cancer compared to normal adjacent kidney tissue and reported loss of DNA 5hmC in certain tumours compared to adjacent normal kidney tissue, however, no change in DNA 5mC.[253] This suggests that the DNA 5hmC antibody from Active Motif is specific for DNA 5hmC. Current bisulphite-only techniques are unable to distinguish

between the closed DNA mark 5mC and the open DNA mark 5hmC. Thus, in PNETs it may be that certain genes previously reported as being methylated, are in fact, hydroxymethylated. Therefore, this chapter aims to assess DNA 5hmC in normal pancreatic tissue and in PNETs, and to explore predictors of 5hmC.

4.2.2 Results

4.2.2.1 DNA 5hmC in normal pancreatic tissue

Normal pancreatic tissue from 20 patients was collected prospectively from patients undergoing non-PNET-related pancreatic surgery at The Oxford University Hospitals (OUH) NHS Trust, with samples collected by the Oxford Research Biobank (ORB). Samples were characterised by IHC to determine 1) pancreatic origin of the sample and 2) islet content with synaptophysin and H&E staining. An example of a characterised islet is shown in **Figure 35,** which confirms PNET origin with H&E, synaptophysin and insulin staining. Thirteen samples had normal islets and exocrine tissue on the same tissue slide and were used for subsequent analysis.

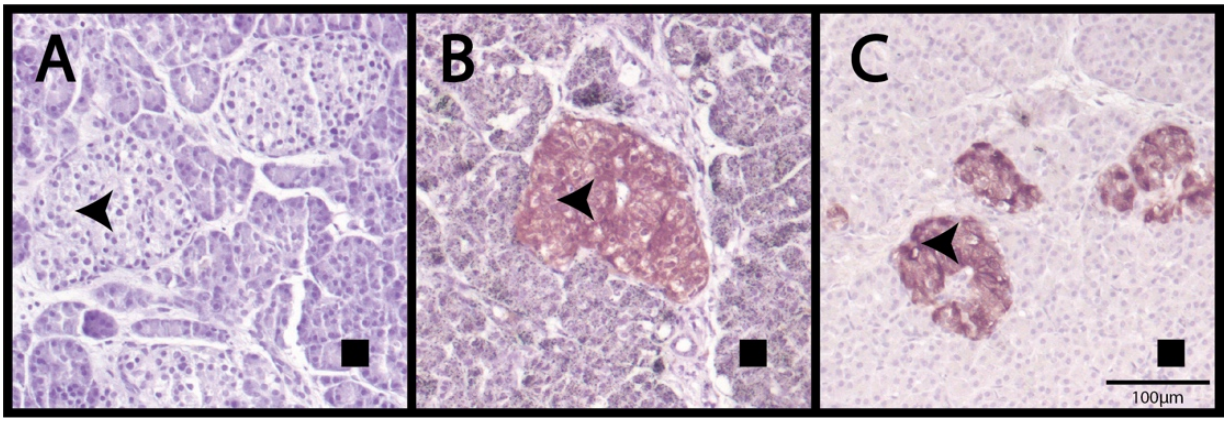


Figure 35: Immunohistochemistry for normal islet 11969 from a 71-year-old female who underwent a non-PNET pancreatic operation. Panel A) H+&E stain, Panel B) synaptophysin staining and Panel C) insulin staining. Normal islets are shown as arrow heads, and exocrine tissue as black squares. All images were taken at 20x.

Staining intensity for DNA 5hmC was determined using H scores which are derived from IHC images analysed using QuPath software (version 0.3.2). H scores range from 0-300, where 0 is no staining identified and 300 is maximal staining. 5hmC staining in normal pancreatic tissue **(Figures 36 and 37)**, revealed that in patients without PNETs, normal islets had an ~2 fold higher 5hmC staining intensity compared to pancreatic exocrine tissue (51 islets assessed in total, range 1-9 islets/sample from 34 fields of 305,866um^2 studied; $p<0.001$).

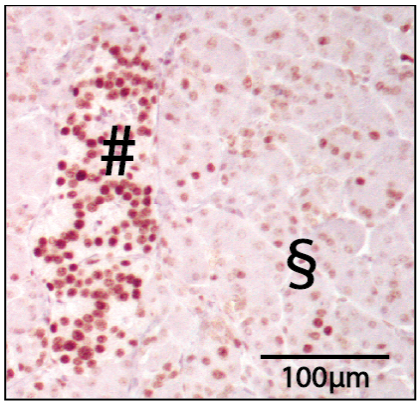


Figure 36: An example of DNA 5hmC intensity in a normal islet vs normal exocrine tissue by IHC. Normal pancreatic islet (#) and normal exocrine tissue (§).

Patient demographics and raw H scores for 13 normal pancreatic samples from the ORB are shown in **Table 29.**

Table 29: ORB samples collected prospectively with IHC DNA 5hmC islet and exocrine tissue H scores

Sample	Age (years)	Sex (M/F)	Islet H score	Exocrine H score
11882	79	M	139.071333	70.1688
11893	50	M	239.155467	134.0913
11898	77	F	248.2016	242.266433
11982	70	M	208.757933	130.890567
11969	71	F	263.863367	147.050867
11974	52	F	215.6776	121.336067
11981	75	M	103.0303	90.5325
11995	69	M	171.6228	120.417
11999	55	M	271.9677	169.300625
12002	68	M	262.78235	187.6665
12073	41	M	280.9934	247.0975
12081	61	M	226.7647	170.4724
12109	73	M	261.8182	233.6377

Male (M); Female (F); Non-functioning PNET (NF-PNET)

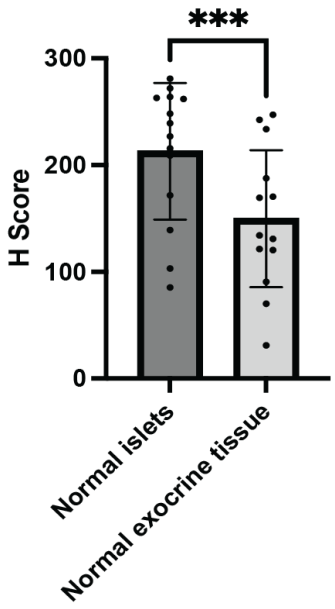


Figure 37: The differences in DNA 5hmC intensity (H scores calculated in QuPath) between normal islets and exocrine tissue. * denotes p <0.001.**

Immunofluorescence was also undertaken to show that the higher intensity 5hmC co-localises with insulin staining **(Figure 38)**.

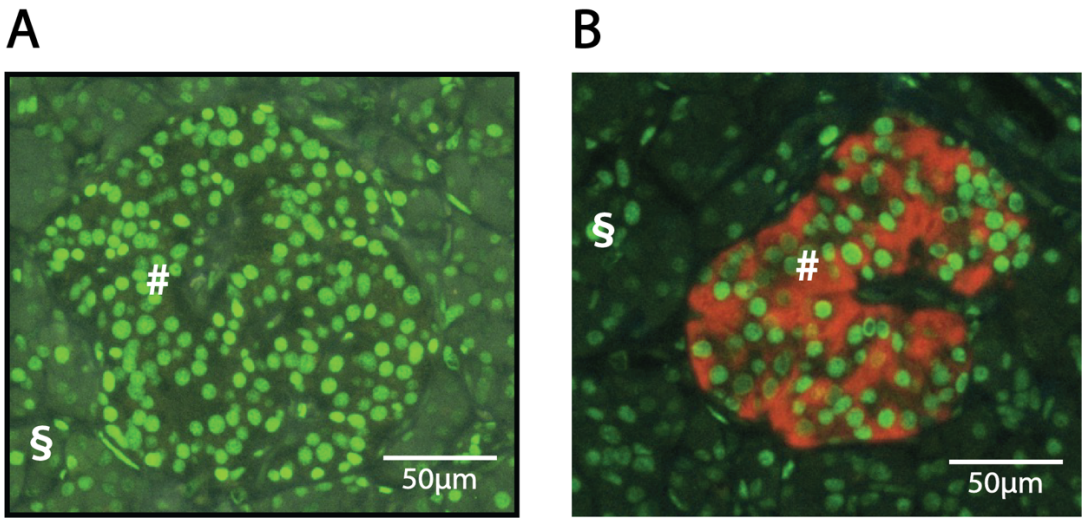


Figure 38: Immunofluorescence showing that DNA 5hmC higher intensity co-localises with insulin staining (red) in the normal pancreatic islet. Normal islets (#) and normal pancreatic exocrine tissue (§).

4.2.2.2 DNA 5hmC in PNETs

4.2.2.2.1 PNET samples from Vienna

In total, 122 pancreatic samples (PNET n=98) and normal pancreas (n=24) were obtained either through collaboration with the Medical University of Vienna (n=95 PNETs) or the Oxford Radcliffe Biobank (ORB; PNETs (n=3) and normal pancreata (n=24)). Samples were either already fixed and embedded in wax blocks (Vienna samples), or prospectively fixed and embedded (ORB samples). PNET subtypes were as follows: Gastrinoma (n=15), insulinoma (n=28), NF-PNET (n=38), rare tumours (n=7) and patients with MEN1 (n=7).

4.2.2.2.2 IHC characterisation of FFPE samples

Pancreatic origin was confirmed with IHC for all PNETs: tumours were stained for synaptophysin, chromogranin, insulin, glucagon, menin and H&E. All tumours were reviewed and assessed for the presence of normal adjacent islets, sufficient pancreatic material, and IHC markers of neuroendocrine origin. An example of H&E, menin and synaptophysin staining for a NF-PNET, an insulinoma and an adjacent normal islet are shown in **Figure 39**. Tumours and normal pancreatic islets stain for synaptophysin **(Figure 39 (D-F))**, a neuroendocrine tumour origin marker. These tumours and normal islets also express the tumour suppressor menin **(Figure 39 (G-I))**.

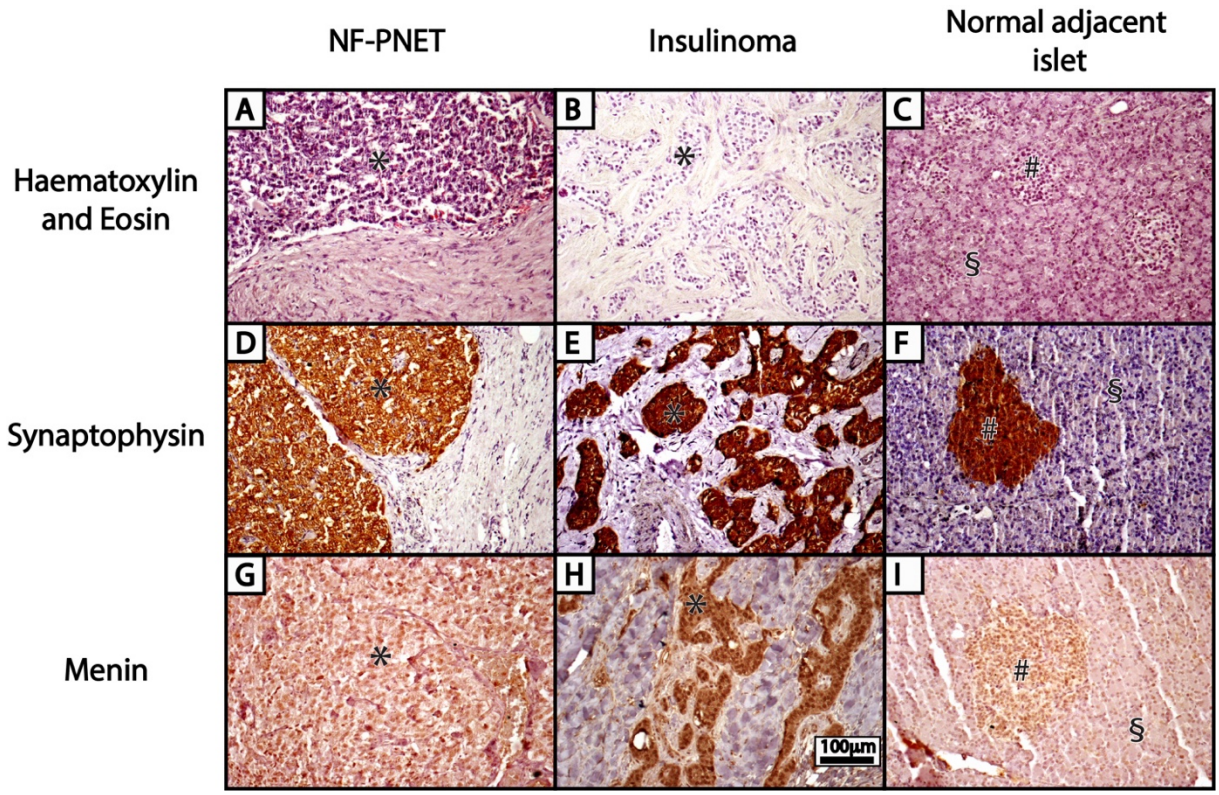


Figure 39: IHC for a NF-PNET, insulinoma and a normal adjacent islet. H&E staining (A-C), synaptophysin staining (D-F) and menin staining (G-I). Normal pancreatic islet (#); normal exocrine tissue (§) and NF-PNET (*).

For subsequent analysis, insulinomas and NF-PNETs in patients without a history of MEN1 were studied, as these were the two most common PNET subtypes.

4.2.2.2.3 DNA 5hmC differences between normal islets and PNET subtypes

5hmC staining was initially undertaken in all confirmed PNETs. Subgroup analysis was subsequently undertaken between the NF-PNETs (n=31, mean age 54 ± 13 years (y), 14 males (M), WHO tumour grade (G) G1=15, G2=13, G3=3) and insulinomas (n=21, mean age 50 ± 21y, 9M, G1=17 and G2=4). These patients did not have an underlying genetic syndrome associated with PNETs or been treated with somatostatin analogue therapy (SSA). Given the historic nature of most PNET samples, this information was obtained through clinical records,

as opposed to genetic testing to exclude hereditary tumour syndromes. An example of DNA 5hmC staining in an insulinoma and NF-PNET with their respective normal adjacent islet is shown in **Figure 40.**

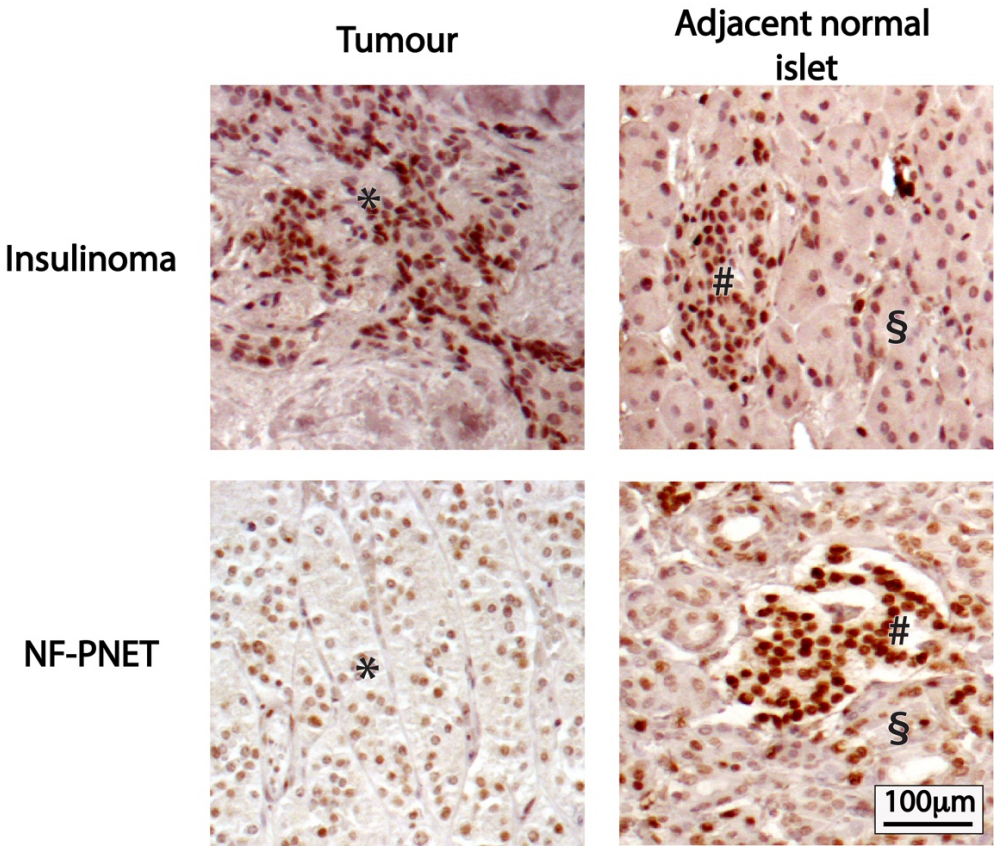


Figure 40: 5hmC staining in an insulinoma and NF-PNET with their normal adjacent islets. Normal pancreatic islet (#); normal exocrine tissue (§) and NF-PNET (*).

Out of these 52 PNETs, 33 had normal adjacent islets present on the same IHC slide (NF-PNETs (n=15) and insulinomas (n=17). The raw H scores for each PNET are shown in **Table 30**.

Table 30: Raw H scores for each DNA 5hmC for each PNET

PNET subtype	Normal islets	Normal exocrine tissue	PNET
Insulinoma			
INS1	38.17465	10.7581	42.0717
INS3	147.333575	109.25805	182.949175
INS4	167.7	67.14495	59.714225
INS8	83.7928	24.61465	88.8322
INS9	247.956025	78.3713	112.503775
INS12	24.851525	4.921375	25.168475
INS13	4.4988	3.2686	47.077
INS15	274.900233	208.827067	148.556075
INS18	225.748175	124.7397	163.642875
INS19	205.55555	108.66415	209.89925
INS21	217.2414	63.3028	126.146475
INS22	186.56355	51.6664	46.0171
INS23	17.0829	1.08295	2.74955
INS24	276.1598	133.845083	168.0281
INS25	102.204533	6.7294	21.0481
INS27	18.5815	6.00876	0.9332
INS28	250.014175	97.81785	107.57795
Average	146.3741	64.76595	91.34795
NF-PNET			
NF2	274.769567	135.844667	106.1172
NF5	87.1541667	34.1698	1.207225
NF8	57.6156333	31.1732333	18.27375
NF13	278.103133	180.002067	102.1039
NF15	172.8855	53.2499667	56.1589
NF16	43.4763667	20.3804333	14.6030333
NF18	13.1634	1.1302	0
NF20	33.3188	0	0.4793
NF22	54.2896	4.4866	8.153
NF31	35.1987	9.35753333	37.1319
NF32	28.0507333	12.6238667	17.635725
NF33	256.2471	106.411067	97.3217
NF34	252.384167	148.830833	140.048733
NF35	98.4025667	15.52	0.27486667
NF36	14.85845	0.9077	2.052725
Average	113.327859	50.2725312	40.1041305

Descriptive statistics were initially reviewed for the H scores of each group (normal islets, insulinomas, NF-PNETs and normal exocrine tissue). Both skewness and kurtosis suggested that the sample values had a normal distribution **(Table 31)**.

Table 31: Descriptive statistics for each distribution

Tissue subtype	Degrees of freedom	Kurtosis	Skewness	Mean	Standard deviation	Variance
Normal islets	32	-1.605	0.2832	128.7	99.84	9968
Exocrine tissue	31	-0.2021	0.9053	57.97	59.91	3589
Insulinoma	16	-1.246	0.2614	91.35	66.87	4471
NF-PNET	15	-0.2251	1.106	37.85	47.13	2221

Non-functioning pancreatic neuroendocrine tumour (NF-PNET)

However, on the assessment of equal variance between each group (which is required for a standard ANOVA), the variances were unequal **(Table 32)**.

Table 32: Calculated F statistic for each comparison, the critical value from the F distribution at α=0.05, and assumption of equal variance

Comparisons	Calculated F statistic	Critical value from F distribution α=0.05	Assume equal variance
Normal islets (df 32)			
Insulinoma	2.22947886	2.1938	No
Exocrine tissue	2.77737531	1.8409	No
NF-PNET	4.48806844	2.2468	No
Insulinoma (df 16)			
Exocrine tissue	1.24575091	2.0148	Yes
NF-PNET	2.01305718	2.4034	Yes
Exocrine tissue (df 31)			
NF-PNET	1.61593877	2.2468	Yes

Degrees of freedom (df)

For the insulinoma, NF-PNET and exocrine tissue groups, there is insufficient evidence that they are not from the same distribution (i.e. F statistics is > critical value from the F distribution),

however, the variance for normal islet H scores is significantly different to all other groups and the assumption of equal variance is violated. Therefore, an ANOVA using a Brown-Forsythe and Welch statistics was used.

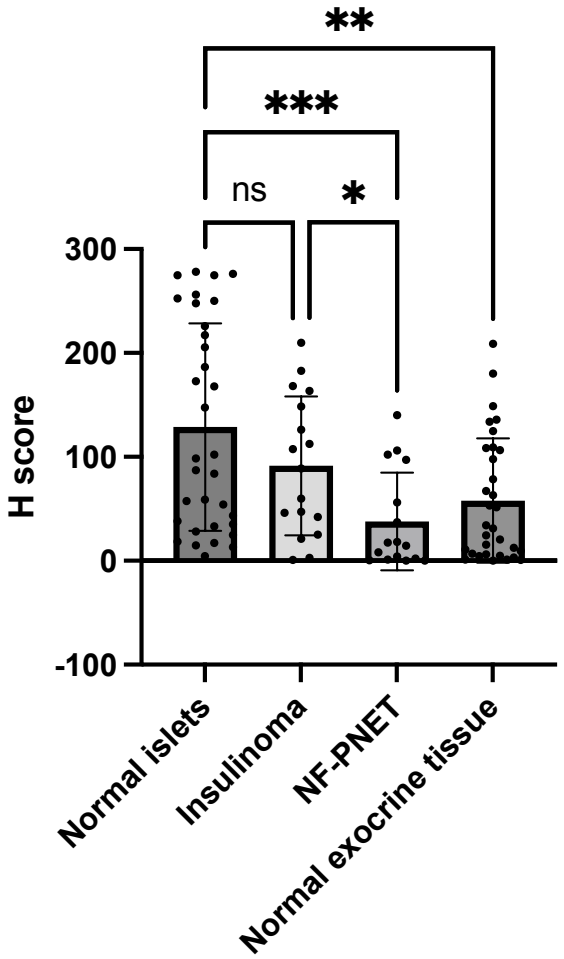


Figure 41: Quantification of DNA 5hmC in Qupath. Normal islets compared to insulinomas, NF-PNETs and normal exocrine tissue. *** p <0.001; ** p <0.01 and * p <0.05.

NF-PNETs had reduced DNA 5hmC when compared to insulinomas ($p<0.05$) and to normal adjacent islets ($p<0.001$) **(Figure 41)**. Thus, DNA 5hmC was ~3 fold lower in NF-PNETs

compared to normal adjacent islets (n=16; p<0.001), whilst 5hmC between insulinomas and normal adjacent islets (n=17) was not significantly different. Subtype group analyses were also undertaken, which reported no significant difference between the H scores of normal islets and exocrine tissue between insulinomas and NF-PNETs **(Figure 42)**.

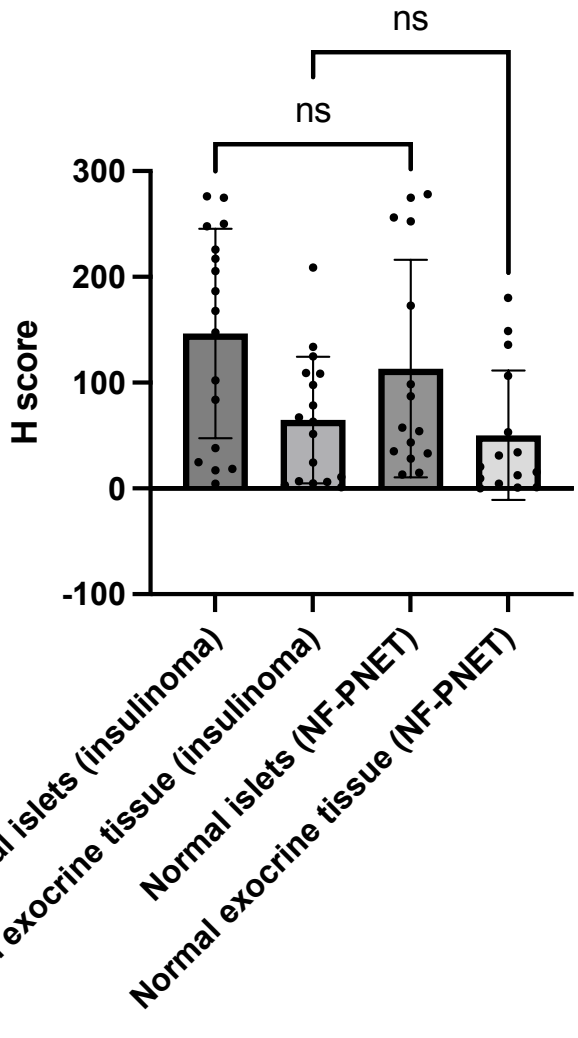


Figure 42: Quantification of DNA 5hmC in Qupath. Normal islets and normal exocrine tissue compared between insulinomas and NF-PNETs. Non-significant (ns).

4.2.2.2.4 Univariate and multivariate analysis

To determine which variables may be associated with the lower DNA 5hmC staining in NF-PNET compared to insulinomas, univariate and multivariate analyses were undertaken using tumour and patient characteristics, with the results shown in **Table 33**.

Table 33: Univariate and multivariate analyses to determine independent variables associated with tumour H Score

Variable	ß estimate	Standard Error	P value
Univariate analysis			
PNET subtype (NF-PNET)	-78.01	28.69	0.0106
Age	-1.654	0.8512	0.0611
Sex (F)	-35.41	31.41	0.2703
Tumour Grade	-34.89	29.34	0.2434
Tumour size	-1.676	1.017	0.1093
Menin status	96.72	40.97	0.0247
Normal adjacent islet H score	0.8065	0.1228	<0.0001
Multivariate analysis (reduced model)		R^2 0.7371	
Normal adjacent islet H score	0.9505	0.3683	0.0156
PNET subtype (NF-PNET)	-61.56	17.80	0.0018
Age	0.1769	1.345	0.8963
Sex (F)	-13.21	18.58	0.4831
Age:Normal adjacent islet H score*	-0.004141	0.006434	0.5253

Female (F); non-functioning pancreatic neuroendocrine tumour (NF-PNET).
*Interactive term

On univariate analysis, PNET subtype, menin status and normal islet H score (5hmC staining intensity) were significantly associated with tumour 5hmC intensity. There was a trend towards increasing patient age associated with tumour 5hmC intensity (p=0.0611). Tumour grade and size were not significantly associated with tumour 5hmC intensity. All significant variables and variables known to be associated with 5hmC levels were entered into the initial multivariate

model before model reduction. Multivariate analysis confirmed that NF-PNETs were associated with lower 5hmC when controlled for patient age, sex, normal islet H score and an age:normal islet H score interacting term ($p<0.002$). Menin status was initially entered into the multivariate model, but because all PNETs associated with loss of menin were NF-PNETs, including menin in the final model did not improve its overall fit and lost significance when controlled for using other variables. Therefore, to evaluate further whether menin had an effect on islet 5hmC, DNA 5hmC was evaluated in an *Men* $^{L/L}$*/Rip2-Cre ER* mouse model.[254]

4.2.2.3 DNA hydroxymethylation in mice with and without menin expression.

Men1$^{L/L}$ mice (*Men1* gene floxed with LoxP sites) were crossbred with mice expressing inducible Cre recombinase under the control of the rat insulin promoter (*RIP2-CreER*). These mice where phenotypically normal until given tamoxifen which would selectively delete the *Men1* gene in insulin-producing β cells of the pancreas, and mice would subsequently develop insulinomas. Previously sectioned unstained immunohistochemistry pancreata from 17 mice aged 6 months of age: 10 control mice (either without CRE recombinase (i.e. wild type) or *MEN*$^{L/L}$*/RIP-2 CreER*, not given tamoxifen) and 7 mice carrying CRE recombinase given tamoxifen at 3 months of age, were stained for DNA 5hmC by IHC. These mice have been fully characterised (including proliferation data) and published by Lines et al., 2016.[254] Sections taken from these mice were stained for 5hmC, the raw H scores and mouse characteristics are shown in **Table 34.**

Table 34: Mouse demographics and raw H scores for DNA 5hmC

Mouse ID	Sex	Genotype	Islet H score	Exocrine H score	Proliferation (Normalised)
Control					
38_1	F	Hom/ERT-Cre	19.5899667	17.2554667	NA
38_7	F	Hom/ERT-Cre	289.87725	143.99305	NA
39_7	M	Hom/ERT-Cre	200.5862	99.6041	NA
44_8*	M	Het	271.579833	122.887733	0.39789461
45_2*	F	Hom	277.1086	115.336367	0.46461469
45_3*	F	wt/ERT-Cre	299.912867	188.118267	1.36658276
48_1*	M	Hom	104.5591	21.9377	0.49402383
48_2*	M	Het	207.684733	19.0616667	0.64493607
48_3*	M	Het	291.6441	138.0921	0.37109609
51_5*	F	Het	293.4729	103.659567	0.5626021
PNET					
67_3*	M	Hom/ERT-Cre	268.082567	154.431433	1.57841103
66_4*	F	Hom/ERT-Cre	289.4617	124.519167	1.53162376
51_1*	M	Hom/ERT-Cre	250.3942	81.9697667	2.19359608
44_4*	M	Hom/ERT-Cre	298.557	116.363667	1.35707462
45_1*	F	Hom/ERT-Cre	192.195233	57.1311333	0.71623987
45_5*	F	Hom/ERT-Cre	280.044067	126.645433	0.85098225
45_6*	F	Hom/ERT-Cre	295.477433	174.342333	1.04089194

* Mice given Tamoxifen. Cre recombinase and a mutated estrogen receptor (ERT-Cre); Female (F); Homozygous (Hom); Heterozygous (Het); Identification (ID); Male (M); Not available (NA).

There was no difference seen in DNA 5hmC between control mice and mice who had developed insulinomas (**Figure 43;** immunohistochemistry and **Figure 44**; Quantification of H scores using QuPath).

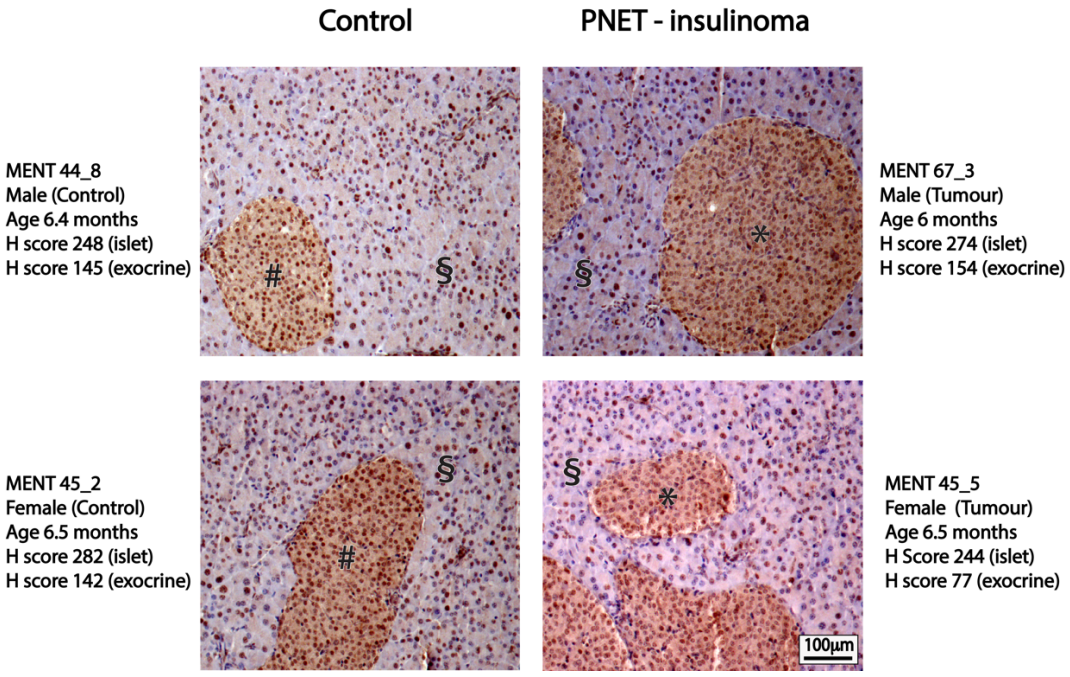


Figure 43: Control/wild-type and *Men1*$^{L/L}$ *RIP2-CreER* mice DNA 5hmC by immunohistochemistry. Normal pancreatic islet (#); normal exocrine tissue (§) and NF-PNET (*).

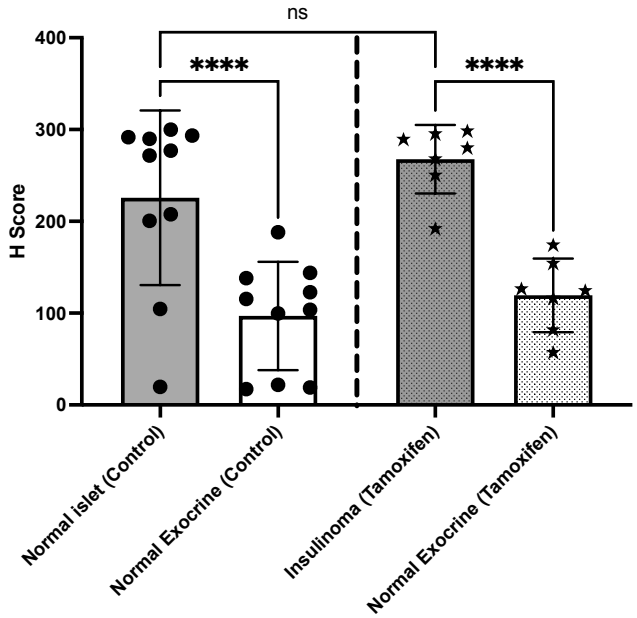


Figure 44: DNA 5hmC in islets and exocrine tissue in control (closed black circles) and Tamoxifen inducible CRE (black stars) in mice aged 6 months. **** p<0.0001 and not significant (ns).

To determine if mouse sex or islet cell proliferation was associated with DNA 5hmC, linear regression analysis was undertaken, which revealed neither sex nor proliferation to be associated with DNA 5hmC, as shown in the multivariate model **(Table 35)**. The only significant predictor for each sample's islet 5hmC score was the adjacent exocrine tissue's 5hmC H score, p=0.0045. It is not unsurprising that adjacent exocrine tissue (on the same slide) is an independent predictor of the islet H score, as samples may have been processed differently at each stage (e.g. different timings for tumour fixation in paraformaldehyde (PFA), alcohol, embedding, sample age or quality). Therefore, the H score of the adjacent exocrine tissue acts to normalise the islet H scores on the same slide to enable a more accurate comparison between the islets of mice with tumours vs controls, as menin expression is not lost in exocrine tissue in *Men1*$^{L/L}$ *RIP2-CreER* mice and therefore it's DNA 5hmC should be same between each group. After controlling for differences between each sample, tumour or normal islet and proliferation indices were not associated with islet 5hmC intensity.

Table 35: Multivariate model of mice treated with and without tamoxifen

Variable	Estimate	Standard error	95% Confidence interval	P value
Intercept	161.0	32.69	87.03 to 234.9	0.0008
Exocrine	0.8678	0.2313	0.3445 to 1.391	0.0045
Tumour[Y]	-2.124	27.83	-65.07 to 60.82	0.9408
Sex[M]	-5.177	22.54	-56.17 to 45.81	0.8235
Profileration	5.704	26.81	-54.95 to 66.36	0.8363

Male (M) and Yes (Y).

Therefore, menin expression does not change DNA 5hmC.

In summary, DNA 5hmC is higher in normal islet cells compared to NF-PNETs and pancreatic exocrine tissue, and this difference seen in DNA 5hmC does not appear to be a function of menin **(Table 36)**.

Table 36: Summary of DNA 5hmC changes in human and mouse PNETs compared to normal pancreatic islets, exocrine tissue and insulinomas.

Human	Normal islet	Insulinoma	NF-PNET	Pancreatic exocrine tissue
Insulinoma	↔	**NA**	↓	↔
NF-PNET	↑	↑	**NA**	↔
Pancreatic exocrine tissue	↑	↔	↔	**NA**

Mouse	Normal islet (menin present)	Insulinoma (no menin)	Pancreatic exocrine tissue
Insulinoma (no menin)	↔	**NA**	↓
Pancreatic exocrine tissue	↑	↑	**NA**

Not available; Non-functioning pancreatic neuroendocrine tumour (PNET).

4.2.3 Discussion

Out of 22 normal pancreatic tissue samples collected prospectively from the ORB (with Dr Kate Lines and Dr Omair Shariq), DNA 5hmC was assessed in 14 samples after IHC characterisation (H&E, synaptophysin and chromogranin A). The most common reason for excluding samples was the absence of islets found during sectioning despite testing multiple samples through the block. This was not an unexpected finding, as for some samples, there was only minimal tissue available. Normal human islets consistently show a significant increase in DNA 5hmC compared to exocrine tissue, suggesting that 5hmC is important for normal islet cell function. One confounder with using these samples was that normal pancreatic tissue was obtained from patients who were undergoing a non-PNET surgery, most commonly for pancreatic ductal adenocarcinoma (PDAC). Therefore, these tissue specimens were the best 'normal' pancreatic samples we could obtain. Co-immunofluorescence of the normal pancreatic tissue with 5hmC and insulin, showed that DNA 5hmC co-localises with normal islet cells.

FFPE blocks are the most common long-term storage of tumour tissue[255] and given the rare nature of PNETs, protocols which have been adapted or specifically designed to analyse these historic samples are needed. Consistent with a previous report by Grillo et al.,[255] examining the extent of antigen loss (cytoplasmic, nuclear and membranous) using old historic FFPE samples from the 1960's compared to newer samples from 2010, in our cohort cytoplasmic antigens (synaptophysin and chromogranin A) did not lose antigenicity over time; however, the nuclear antigen 5hmC required prolonged heat treatment for antigen retrieval for membrane permeabilisation before antigen detection.[255] In our cohort, IHC staining for synaptophysin and chromogranin A (specific markers for neuroendocrine tumours[256]) were present in over

95% of tumours and were concordant with the clinical details that these tumours are of neuroendocrine origin. Insulin staining was undertaken using two separate insulin antibodies, however, both failed to detect insulin in insulinomas (although did detect insulin in adjacent normal islets). This does not refute the diagnosis of an insulinoma, and current guidelines report that insulin staining is not always positive and the diagnosis of an insulinoma is based on clinical features i.e. Whipple's triad.[257] It was unclear from sectioning and IHC staining if the gastrinomas were of pancreatic origin and therefore were not carried through for further analysis. Additionally, there were rare secreting pancreatic islet tumours including VIPomas, glucagonomas and ACTHomas, which were not analysed in detail due to low numbers. PNET samples were available for seven MEN1 patients (two with insulinomas and two with NF-PNETs) and DNA 5hmC was assessed. However, as all MEN1 patients had multifocal tumours, it was unclear which PNET was a functioning insulinoma vs a NF-PNET. Therefore, a cohort of non-MEN1 insulinomas and NF-PNETs obtained from the ORB and from Vienna were fully characterised (e.g. stained by IHC for insulin, glucagon, chromogranin and synaptophysin) and for DNA 5hmC. Given that most PNETs are of low grade, and with the known difficulties with detecting some nuclear antigens, Ki67 was not repeated and these values for each tumour sample were taken from the clinical details provided with the specimens. Although there was a total of 28 insulinomas and 41 NF-PNETs, only 17 and 16 were analysed for DNA 5hmC, respectively, as these samples had adjacent normal islets on the same slide. Comparing protein and/or epigenetic expression of specific markers in tumours to normal islets on the same slide is important as variables such as block age, embedding techniques and IHC staining protocols (for example, the length of time slides are incubated with diaminobenzidine (DAB; a chromagen which stains nuclei brown)) will change the staining intensity, and therefore, it is not accurate

to compare the expression of a specific marker only between tumours on different slides. There was no significant difference in normal pancreatic islet and exocrine cell DNC 5hmC between insulinomas and NF-PNETs. NF-PNETs had an approximate 3-fold low DNA 5hmC when compared to normal islets, whereas there was no significant difference between normal islet DNA 5hmC and insulinomas. Given that other factors can determine DNA 5hmC, univariate and then multivariate analysis was undertaken to control for these variables including patient sex and age, and DNA 5hmC of normal adjacent islet. This also confirmed that DNA 5hmC is significantly associated with the type of tumour, with NF-PNETs having ~60 lower H score (H score 0 – 300; 0 (no staining), 300 (maximum staining)).

To determine if menin expression was associated with changes in DNA 5hmC, 6-month-old control/wildtype mice and 6-month-old *MEN*$^{L/L}$*/RIP-2 CreER* mice (who had been given tamoxifen at 3 months to knockout menin expression in pancreatic islets) were compared. There were no differences in DNA 5hmC in islet cells between mice with menin expression (control) and mice without menin expression (who developed PNETs), even when controlled for sex and islet cell proliferation. This suggests that menin and tumour proliferation is not a determinant of DNA 5hmC in mice. It has been reported that DNA hydroxymethylation may be a function of proliferation, with rapidly dividing cells losing hydroxymethylation. The mouse data does not support this. Explanations for this may include, a low proliferation index, that is generally found in PNETs, or it may be that tumours cells are able to regenerate the 5hmC mark.

Conclusion: Overall PNET cells lose DNA 5hmC compared with normal islet cells and that this loss in nuclear staining is more pronounced in NF-PNET cells compared to insulinoma cells. Normal pancreatic islet cells are heavily hydroxymethylated compared with exocrine cells suggesting this DNA mark may have an important role in normal pancreatic islet cells. Therefore, in the next chapter, I will investigate this DNA 5hmC at the CpG level and determine possible pathways enriched in this DNA mark and its role.

4.3 Chapter 3: Investigating individual CpG DNA hydroxymethylation in normal pancreatic islet tissue and PNETs by methylationEPIC array

4.3.1 Introduction

PNET subtypes (insulinomas and NF-PNETs) have different global DNA 5hmC, and both have lower DNA 5hmC when compared to normal pancreatic islets **(Chapter 2, 4.2)**. To further investigate for epigenetic alterations in these PNETs, the methylationEPIC array was used to determine CpG-specific methylation values, as it interrogates >850,000 CpG sites across the human genome. Pipelines have been designed to analyse raw data generated from the methylationEPIC and 450K arrays in R.[212] Although it was not initially designed to investigate the hydroxymethylome, the methylationEPIC array can be adapted.[211] The normal pancreas contains ~1-4.5% islets, with the majority being pancreatic exocrine tissue.[6,258] Human pancreatic islets are grouped in the head of the pancreas, and gradually disperse towards the pancreatic tail.[258] As normal pancreatic islets are heavily hydroxymethylated, despite most normal pancreas comprising exocrine tissue, a comparison between hydroxymethylated PNET CpGs and normal pancreatic tissue may be possible. Insulinomas are derived from normal β cells. However, the cell of origin for most NF-PNETs is unclear, with a proportion similar to normal α cells or a β-like cell.[72,135,259] It has been reported that RASSF1A is the most commonly methylated tumour suppressor in sporadic PNETs.[114-117,123,128] Loss of RASSF1A leads to decreased phosphorylation of yes-associated protein 1 (YAP1), thereby allowing YAP1 to interact with nuclear transcription factors and promote TEA domain transcription (TEAD) transcription **(Introduction 1.6.2, Figure 11)**.[260] Free Fatty Acid Receptor 1/G-protein-coupled receptor 40 (FFAR1/GPR40) agonists have recently been reported to activate the Hippo signalling pathway in a prostate cancer cell line, by increasing phosphorylation of YAP, which sequesters YAP in the cytoplasm and thereby acting as a tumour suppressor.[141] FFAR1 is highly expressed in human pancreatic islet cells, with ~20 fold increased expression compared to

pancreatic exocrine tissue,[261] and its endogenous ligands include medium to long chain fatty acids.[262] Current interest in this receptor is mainly due to its expression in human pancreatic β cells, where activation leads to glucose-mediated insulin secretion through the PLC/IP3/Ca^{2+} pathway; therefore, it is a target for type 2 diabetes therapeutics.[263,264] In rodents, FFAR1 is reported to be predominantly expressed in pancreatic β cells and not in α cells,[262] whereas, it is unclear if FFAR1 is expressed in human pancreatic α cells. FFAR1 and FFAR4 receptors play a role in tumourigenesis, and both receptor activation or inhibition have opposing effects in different cancer cell types.[265] This chapter aims to explore DNA 5mC and 5hmC in normal pancreatic exocrine tissue (containing groups of pancreatic islets) and PNET subtypes and compare this with RNA expression, and to review the methylation of key and novel genes involved in PNET pathogenesis and Hippo signalling.

4.3.2 Results

4.3.2.1 DNA extraction and quality assessment

To determine which specific CpG sites were hydroxymethylated, DNA was extracted from FFPE tissue, oxidative and bisulfite converted and then sent for methylationEPIC array analysis.[168] DNA was extracted from 59 FFPE PNET samples (>70% tumour content), 14 normal pancreatic exocrine tissue samples containing groups of normal pancreatic islets (matched to PNETs (n=12) and Oxford Research Biobank (ORB) (n=2)). The islet content of the normal pancreatic exocrine tissue was verified by H&E staining before wax scrolls were sectioned off the FFPE block for DNA extraction. DNA fragmentation and quality were assessed by PCR (forward and reverse primers generated using primer-BLAST for *MEN1* (9 PNET samples) and *RASSF1* (35 PNET samples). Primer efficacy and optimisation were initially evaluated with control DNA, with all primers amplifying DNA as expected. *MEN1* PCR products were amplified in PNET samples at lengths of: 250bp 7/8 (88%), 500bp 0/8 (0%) and 750bp 0/8 (0%) and in *RASSF1* samples: 244bp 34/35 (97%), 516bp 21/35 (60%) and 775bp 2/8 (25%). An example of amplified DNA extracted from an insulinoma is shown in **Figure 45**, which reports *MEN1* can be amplified to 250bp and *RASSF1* up to 750bp.

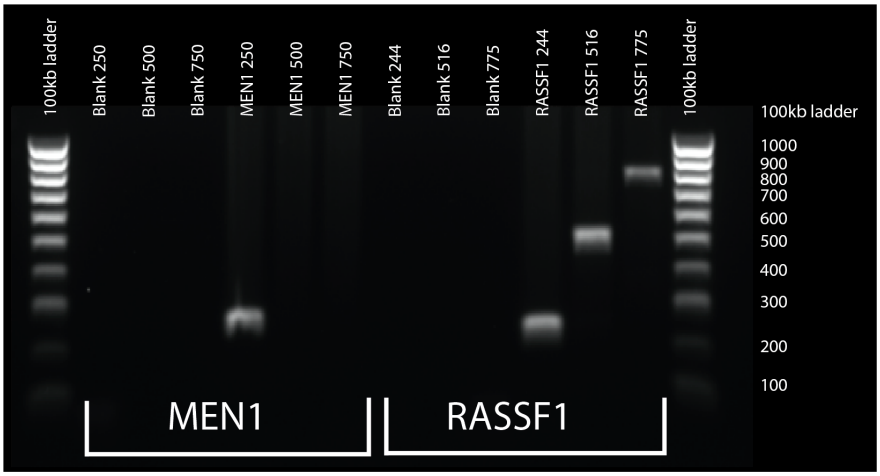


Figure 45: DNA quality was assessed using PCR for *MEN1* and *RASSF1*.

Samples were ranked by quality using the (260/230) and (260/280) ratios from NanoDrop readings, by DIN (DNA integrity number), ability to PCR *RASSF1*, total amount of DNA extracted and sample type. Preference was given to tumour samples with adequate amounts of adjacent normal exocrine pancreatic tissue containing normal pancreatic islets, to improve matching.

4.3.2.2 Ox-BS conversion

Samples from 11 people with sporadic PNETs (n=9) and normal islet-containing exocrine pancreatic tissue (n=2; from patients who underwent pancreatic surgery for PDAC). Samples were taken from 7 normal islet-containing exocrine pancreatic tissue (5 matched to tumour tissue), 4 insulinomas and 5 NF-PNETs **(Methods 3.2, Tables 6,7 and 10)**. Additionally, 10 PNETs (5 insulinomas and 5 NF-PNETs **(Methods 3.2, Tables 6,7 and 10)** and 2 cell lines (BON1 and QGP1) were bisulfite converted only. BON1 cells which derived from the lymph node from a 28-year-old patient with metastatic insulinoma[218] were provided by Dr Simona Grozinsky-Glasberg (Hadassah-Hebrew University Hospital, Israel) as a gift to the University of Oxford, and QGP1 cells which derived from a metastatic somatostatinoma,[219] were purchased from the Japanese Collection of Research Bioresources Cell Bank (Osaka). The BON1 cell line was used as a model for insulinomas and the QGP1 cell line was used as a model for a NF-PNETs. Samples were then sent for methylationEPIC array analysis at Cambridge Epigenetics. Raw data results were received and analysed using 'ChAMP' package in R Studio (version 4.2.0 "Vigorous Calisthenics").

4.3.2.3 5hmC in normal pancreatic tissue

Seven pairs of normal islet-containing exocrine pancreatic tissue underwent both oxBS and BS conversion and underwent methylationEPIC array. Samples were analysed in R Studio.

4.3.2.3.1 Quality control for the loaded samples

In total, 727,322 probes were available for analysis from seven paired (i.e. 14) samples. Quality control data for all uploaded samples are shown in **Supplementary Appendix 8.1, Table 43.** Samples that underwent oxidation before bisulfite conversion show a greater proportion of unmethylated CpG sites (larger peak at beta<0.3) and a lower proportion of methylated CpG sites as evidenced by a smaller peak at beta >0.7 **(Figure 46 Panel A)** and the multidimensional (MDS) plot for the normal islet-containing exocrine pancreatic tissue samples **(Figure 46 Panel B)**, shows that oxBS and the BS samples cluster together. Singular Value Decomposition (SVD) analysis assessed any significant effects in sample variation that may be attributed to other variables (technical effects, patient anthropometric data and tumour characteristics).

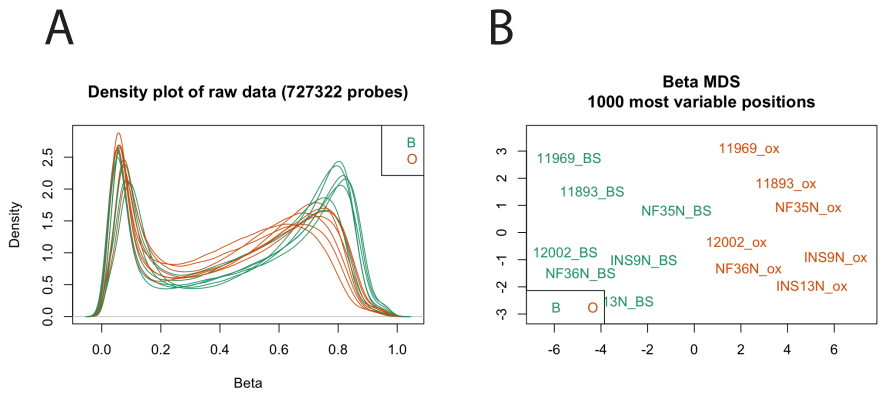


Figure 46: Panel A) Density plot for oxBS (O; orange) vs BS samples (B; green) for the 7 paired normal islet-containing exocrine pancreatic tissue samples. Panel B) Multidimensional plot for the oxBS (orange) and BS-only (green) paired samples.

SVD analysis reported that the main driver for the variation seen within the samples (~73% variation) was due to the samples being assessed (i.e. oxBS vs BS; p<0.01; Principal components (PC)1 and 3) and the array position (p<0.05; PC-1) **(Figure 47)**. Block age (~18% variation; PC-2) and sex (~8% variation; PC-4) were other significantly associated variables between the oxBS and BS samples, p<0.01 and <0.05, respectively **(Figure 47)**.

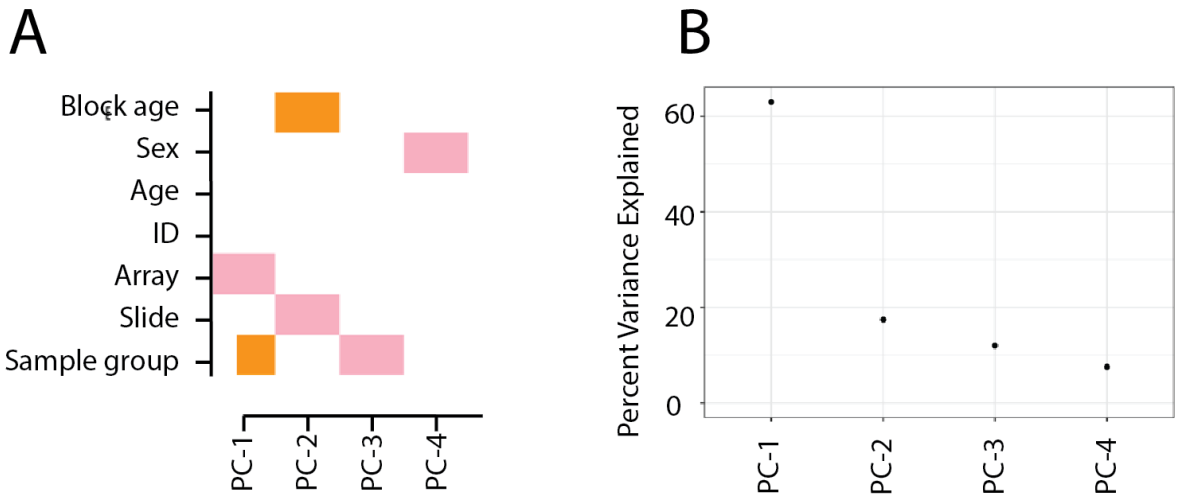


Figure 47: Panel A shows an SVD plot for all PNET samples with 4 principal components explaining the variation between samples (PC-1 to PC-4). p>0.05 (white boxes), p<0.05 (pink boxes) and p<0.01 (orange boxes). Panel B reports the Scree plot explaining the percentage of the variation seen within the samples (oxBS vs BS) for all principal components.

4.3.2.3.2 Gene Set Enrichment Analysis

Using a false discovery rate (FDR) of 0.05 and a beta cut-off of 0.05, 64,478 differentially methylated probes (DMPs) i.e. 5hmC CpGs, were identified (using R package "GoMeth") with an adjusted p<0.01. In normal pancreatic tissue 5hmC was enriched in 36 Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways. The first 9 pathways are shown in **Figure 48**.

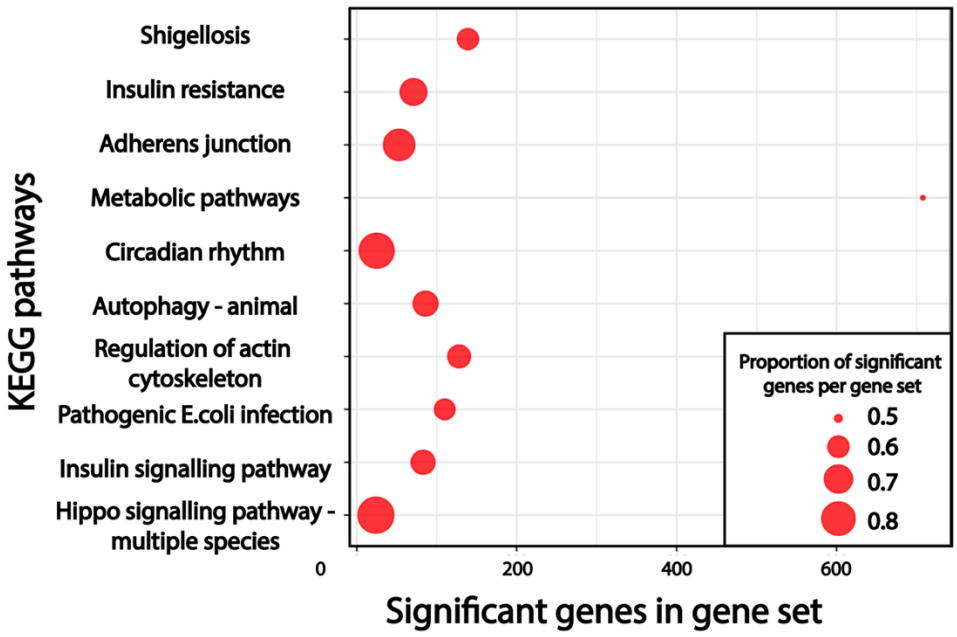


Figure 48: Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways that are enriched with 5hmC in normal pancreatic tissue.

4.3.2.4 5hmC analysis in PNETs

The methodology for analysis of PNET 5hmC CpGs was repeated for normal islet-containing exocrine pancreatic tissue samples. The final number of CpG probes for the 9 pairs of PNETs (4 insulinomas and 5 NF-PNETs) was 723,389. Analysis was performed using both 'GoMeth' and 'GoREGION'. There was a total of 57 DMPs with a $p<0.01$, using an FDR of 0.05 and a beta cut-off of 0.05, and no genes were enriched in GSEA pathways. To determine why there were so few significant 5hmC CpGs in PNETs, SVD analysis was undertaken, which reported that tumour type (insulinoma vs PNET), grade and size, patient age and sex were significant patient variables contributing to the variation in 5hmC between samples **(Figure 49)**. PC-1 (which explains ~33% sample variation) was significantly associated with patient age ($p<0.01$), tumour size ($p<0.05$) and tumour grade ($p<0.05$). Sample group (oxBS vs BS; PC-7 contributed

to <5% of the variation seen in all PNET samples, p<0.05). Given that there may be confounding between these univariate variables and that insulinomas have more 5hmC than NF-PNETs on 5hmC IHC staining, PNET analysis was repeated separately for insulinomas and NF-PNETs.

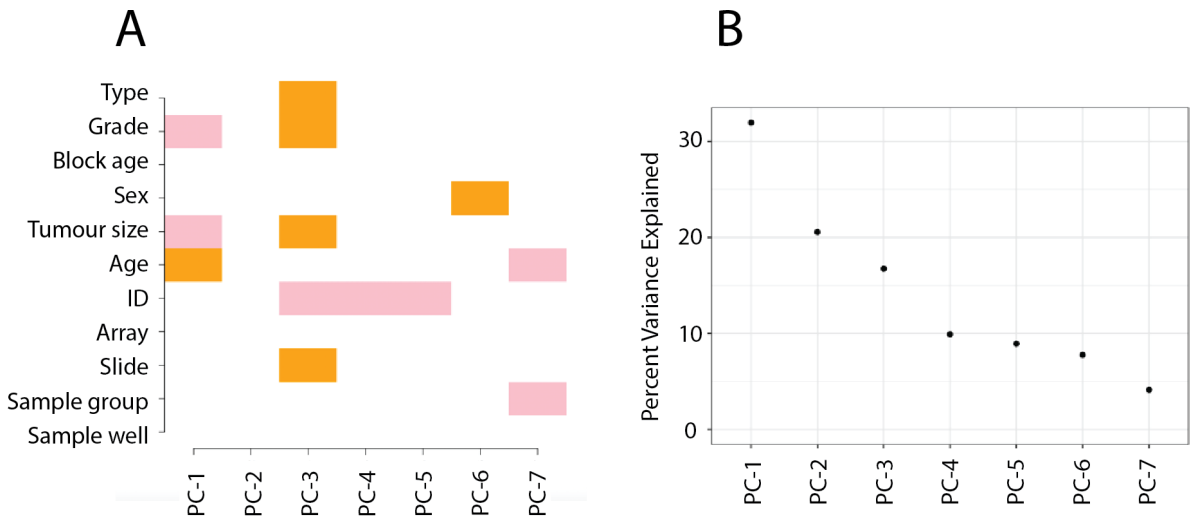


Figure 49: Panel A shows an SVD plot for all PNET samples with 7 principal components explaining the variation between samples (PC-1 to PC-7). p>0.05 (white boxes), p<0.05 (pink boxes) and p<0.01 (orange boxes). Panel B reports the Scree plot explaining the percentage of the variation seen within the samples (oxBS vs BS) for all principal components.

4.3.2.4.1 Insulinomas

There were eight paired insulinoma samples analysed (from 4 patients). A total of 734,439 probes were analysed. SVD analysis revealed that block age (~62% variation, PC-1, p<0.01), tumour size (~20% variation, PC-2, p<0.05) and sample group (20% variation, PC-3, oxBS vs BS; p<0.05) were significantly associated with the variation seen within the samples **(Figure 50)**. Using the same criteria used in normal pancreatic tissue, 0.03% (2306/734,439) 5hmC CpGs (adjusted p value cut-off of <0.08).

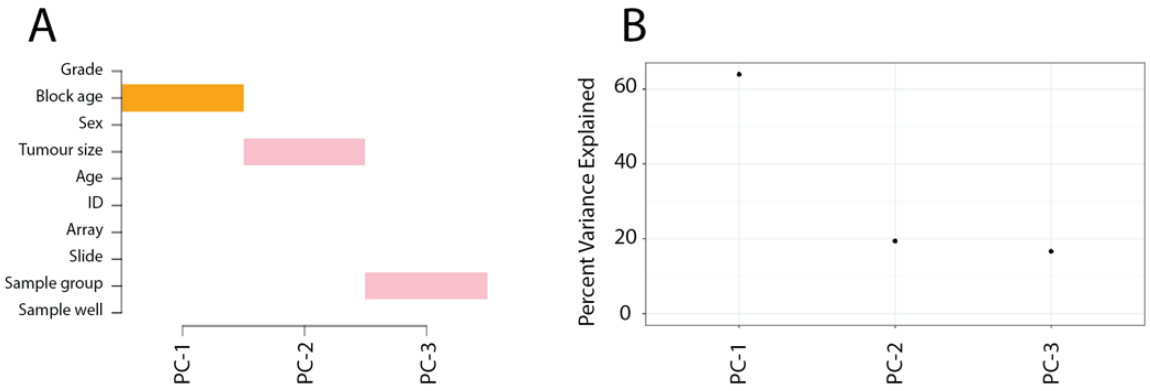


Figure 50: Panel A shows an SVD plot for all insulinoma samples with 3 principal components explaining the variation between samples (PC-1 to PC-3). p>0.05 (white boxes), p<0.05 (pink boxes) and p<0.01 (orange boxes). Panel B reports the Scree plot explaining the percentage of the variation seen within the samples (oxBS vs BS) for all principal components.

4.3.2.4.2 NF-PNETs

There were 10 NF-PNET samples analysed (from 5 patients). A total of 730,927 probes were analysed. SVD analysis revealed that age ($p<0.05$), tumour size ($p<0.01$) and grade ($p<0.05$) accounted for ~35% variation (PC-1), sex ($p<0.05$, PC-2, ~32% variation), and block age ($p<0.01$), sex ($p<0.05$ and slide ($p<0.05$), PC-4, accounted for ~15% variation. were associated with sample variation **(Figure 51)**. There were no 5hmC CpG sites, even with an adj p value of 0.5.

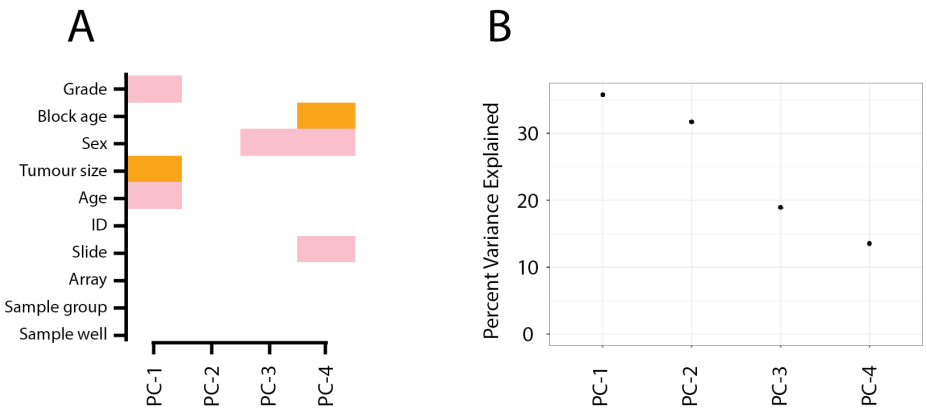


Figure 51: Panel A shows an SVD plot for all NF-PNET samples with 4 principal components explaining the variation between samples (PC-1 to PC-4). p>0.05 (white boxes), p<0.05 (pink boxes) and p<0.01 (orange boxes). Panel B reports the Scree plot explaining the percentage of the variation seen within the samples (oxBS vs BS) for all principal components.

4.3.2.5 Normal islet-containing exocrine pancreatic tissue versus PNETs

4.3.2.5.1 Hydroxymethylated CpGs in each group

The density plot and multidimensional plots for normal islet-containing exocrine pancreatic tissue and PNET subtypes are shown in **Figure 52**. Insulinomas and NF-PNETs form two separate groups, and insulinomas are tightly clustered compared to the more heterogeneous NF-PNETs.

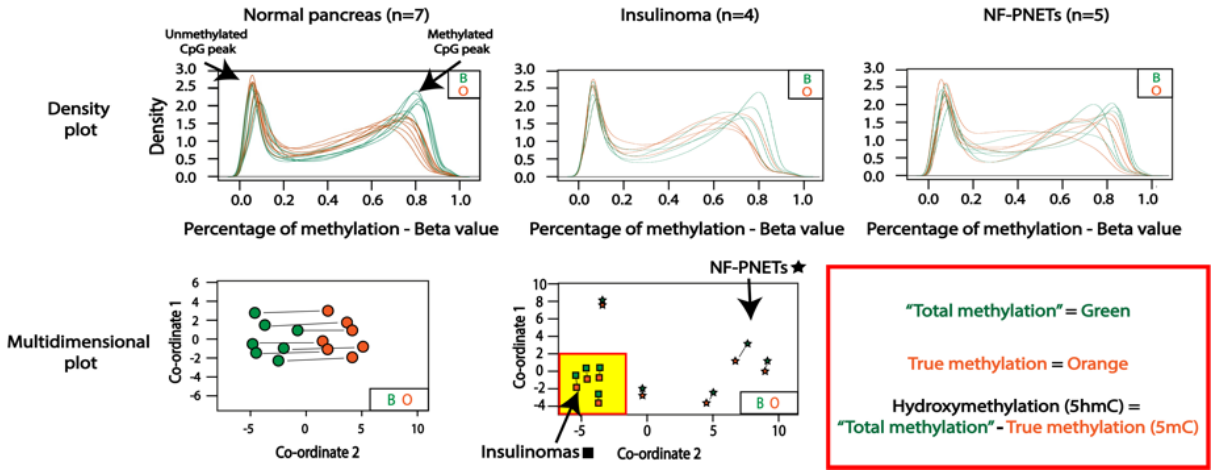


Figure 52: Density plot and multidimensional plots for normal islet-containing exocrine pancreatic tissue and PNET subtypes.

The number of hydroxymethylated CpGs across the genome was assessed for each group (normal islet-containing exocrine pancreatic tissue, insulinomas and NF-PNETs) using the R package 'GoMETH'. Normal islet-containing exocrine pancreatic tissue had the highest number of hydroxymethylated CpGs at 8.9% (64,478,727,322) using an adjusted P value of <0.01). Insulinomas had 0.03% (2306/734,439) hydroxymethylated CpGs using an adjusted P value cut off of <0.08 (there were no significantly hydroxymethylated CpGs using an adjusted p value of <0.01), and for NF-PNETs, there were no significantly hydroxymethylated CpGs up to an adjusted P Value of <0.5). Two insulinomas and 3 NF-PNETs had paired normal islet-containing exocrine pancreatic tissue analysed. Heatmaps are shown in **Figure 53**, which clearly shows that NF-PNETs have no common hydroxymethylated genes.

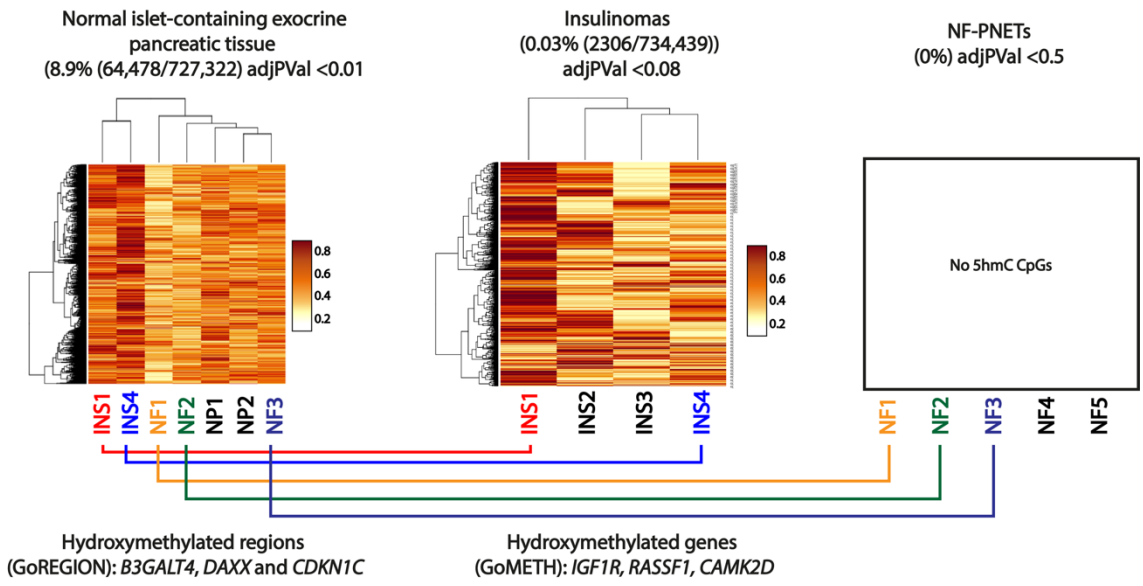


Figure 53: Heat map for the hydroxymethylated regions in normal islet-containing exocrine pancreatic tissue and for hydroxymethylated genes in insulinomas and NF-PNETs.

4.3.2.5.2 Significantly hydroxymethylated genes in each group

To obtain a broader picture of which genes or DNA regions (DMRs) were hydroxymethylated in each group (normal islet-containing exocrine pancreas, insulinoma and NF-PNETs), 'GoMeth' (returns 5hmC genes) and 'GoREGION' and 'Bumphunter' (returns 5hmC regions) packages were run in R. Normal islet-containing exocrine pancreatic tissue had 3338 significantly hydroxymethylated genes **(Supplementary Appendix 8.1, Table 44)**, and 15 differentially hydroxymethylated regions (DhMRs) with a family-wise error rate (FWER) or <0.2. The first 10 DhMRs are shown in **Table 37**. Insulinomas had 1480 significantly 5hmC genes **(Supplementary Appendix 8.1, Table 45)**, including *RASSF1A* and *IGF-1*, but not *MEN1* or *DAXX*. There were no significantly hydroxymethylated genes for NF-PNETs, and no hydroxymethylated regions for insulinomas or NF-PNETs. There were no significantly hydroxymethylated CpGs in the *MEN1* or *DAXX* in insulinomas, however, *RASSF1A* and *IGF-1* had significantly hydroxymethylated CpGs.

Table 37: Differentially hydroxymethylated regions in normal pancreatic tissue using 'Bumphunter' package in R

DhMR	Chr	Start	End	Length	Number of CpGs	Gene	p.value	DhMR.fwer
1	chr6	33244752	33246488	1736	46	*B3GALT4*[a]	0	0
2	chr6	31598379	31601226	2847	38	*PRRC2A*	7.4e-06	0.004
3	chr15	44956098	44956258	160	5	*SPG11*[a]	2.2e-05	0.012
4	chr5	132201234	132201652	418	6	*GDF9*[a]	9.6e-05	0.04
5	chr11	2922051	2922852	801	12	*SLC22A18*	0.000140	0.052
6	chr11	2481206	2482929	1723	22	*KCNQ1*	0.000118	0.064
7	chr3	23851254	23851909	655	9	*UBE2E1*[a]	0.000163	0.068
8	chr5	177659926	177660509	583	8	*PHYKPL*[a]	0.0001698	0.068
9	chr4	1166528	1167714	1186	16	*SPON2*[a]	0.0002290	0.092
10	chr11	44972672	44973218	546	12	*TP53I11*	0.0002586	0.104

[a]H3K27Acetylation mark. Beta-1,3-Galactosyltransferase (*B3GALT4*); Chromosome (Chr); Differentially hydroxymethylated regions (DhMR); family-wise error rate (FWER); Growth Differentiation Factor 9 *(GDF9)*; Potassium Voltage-Gated Channel Subfamily Q Member 1 (*KCNQ1);* 5-Phosphohydroxyl-L-Lysine Phospho-Lyase *(PHYKPL);* Proline rich coiled-coil 2A *(PRRC2A);* Solute Carrier Family 22 Member 18 *(SLC22A18);* SPG11 Vesicle Trafficking Associated, Spatacsin *(SPG11);* Spondin 1 *(SPON2);* Tumor Protein P53 Inducible Protein 11 *(TP53I11);* Ubiquitin Conjugating Enzyme E2E1 *(UBE2E1)*

In normal islet-containing exocrine pancreatic tissue, there were 9 DhMRs with a FWER of <0.1 (i.e. false positives could occur ~10%). DhMRs were associated with the histone mark, H3K27Acetylation (Acetylation of lyinse 27 on histone 3) in 66.7% (6/9) of regions. H3K27Acetylation can be found at active enhancer regions, is associated with an open chromatin state and is a critical regulator of gene expression.[266]

4.3.2.6 RNA-seq analysis

To determine possible mechanisms underlying these differences in tumour subtype 5hmC, RNA-Seq analysis using NF-PNETs (n=7) and insulinomas (n=6) was performed. This revealed 170 genes to be differentially expressed (FDR $p<0.05$), of which 91 were upregulated and 78 downregulated, and 79 genes to be differentially expressed (FDR $p<0.01$), of which 33 were upregulated (including Calcium/Calmodulin Dependent Protein Kinase 2 Gamma (*CAMK2G*)); and 46 were down-regulated (including Glucagon Like peptide 1 receptor (*GLP1R*)) in NF-PNETs **(Figure 54; Supplementary Appendix 8.1, Table 46)**.

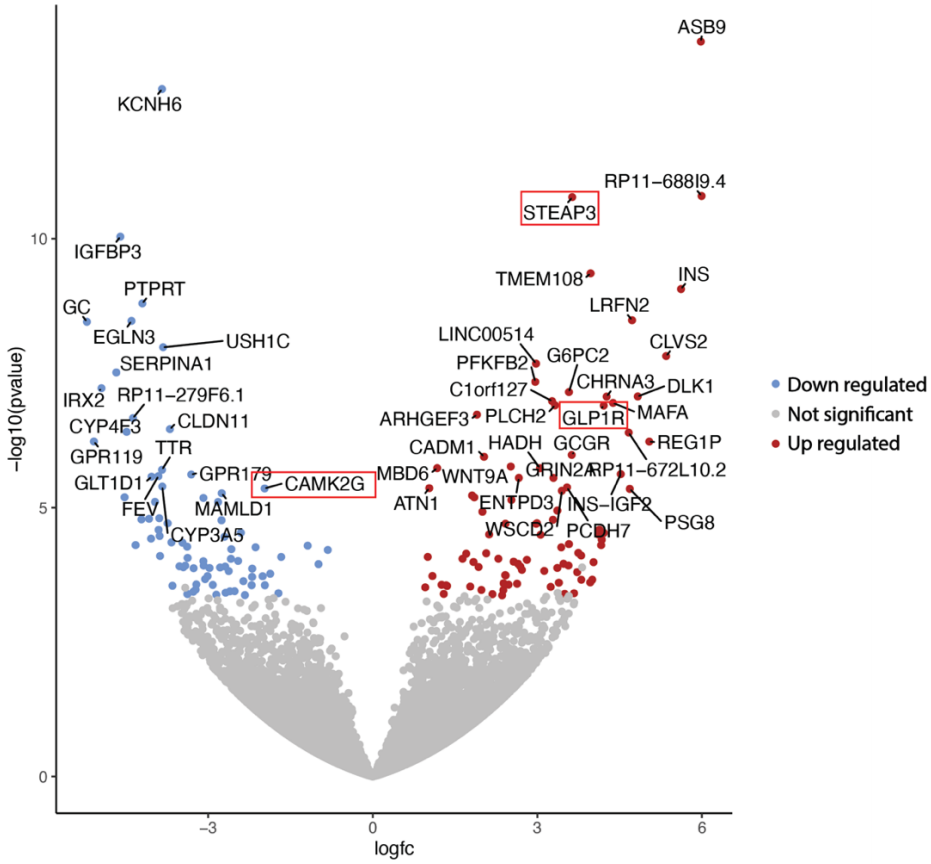


Figure 54: Volcano plot reporting the top 79 differentially expressed genes (FDR p<0.01) comparing insulinomas to NF-PNETs. The red circles are genes upregulated, and blue circles are genes downregulated in insulinomas compared to NF-PNETs. The top 50 genes (out of 79) are annotated with the gene name. Genes possibly involved in DNA 5hmC are in red boxes.

On review of these genes, STEAP3 metalloreductase (*STEAP3)* and glucagon-like peptide-1 (*GLP1)* (FDR p<0.01) and CAMP Responsive Element Binding Protein 3 Like 2 (*CREB3L2)* (FDR p<0.05) showed significantly increased expression in insulinomas compared to NF-PNETs. Additionally, Calcium/Calmodulin Dependent Protein Kinase II Gamma *(CAMK2G)* (FDR p<0.01) was increased in NF-PNETs compared to insulinomas. These genes may be involved in DNA 5hmC regulation **(Figure 55)**.

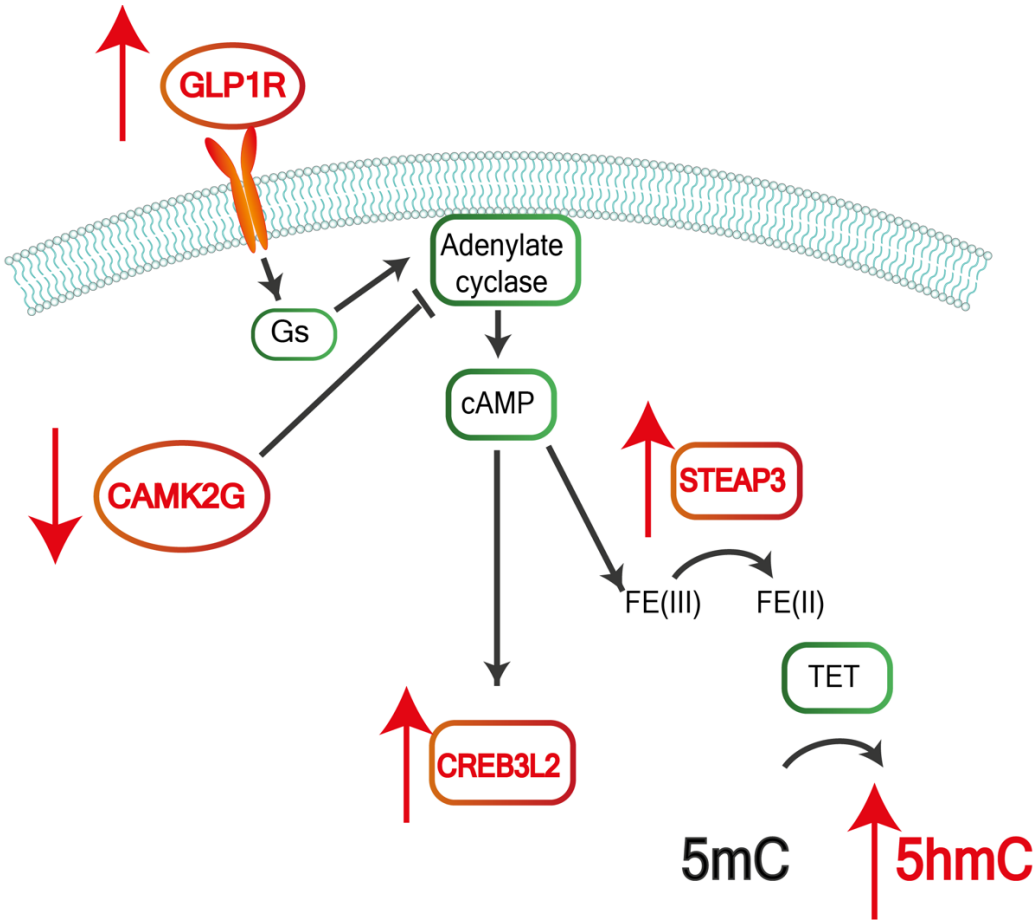


Figure 55: Differential expression of key genes in insulinomas compared to NF-PNETs. CAMP Responsive Element Binding Protein 3 Like 2 (CREB3L2); Calcium/Calmodulin Dependent Protein Kinase II Gamma (CAMK2G); Glucagon-like peptide-1 (GLP-1); STEAP3 metalloreductase (STEAP3); Ten-eleven translocase (TET); DNA methylation (5mC); DNA hydroxymethylation (5hmC).

4.3.2.7 RNA-seq analysis validation using a separate PNET cohort

PNET RNA samples from the Mayo cohort **(Methods 3.2.3 Table 11)** were normalised and converted to cDNA by Dr Kate Lines. Quantitative real time-PCR (qRT-PCR) was subsequently performed on all samples (insulinomas (n=14), NF-PNETs (n=25) and normal islet-containing exocrine pancreatic tissue (n=6)) by Dr Kate Lines and me. Samples were run between 40-45 cycles. Three housekeeper genes were run (*GAPDH*, *CANX* and *B2M*) and RNA-seq data was validated with key genes of interest identified in the RNA-Seq analysis (*INS*,

CAMK2G, *PTPRT*, *GPR119*, *GLP-1* and *STEAP3*). **Figure 56** confirms that in a separate cohort of PNETs, insulin and STEAP3 are more highly expressed in insulinomas than NF-PNETs.

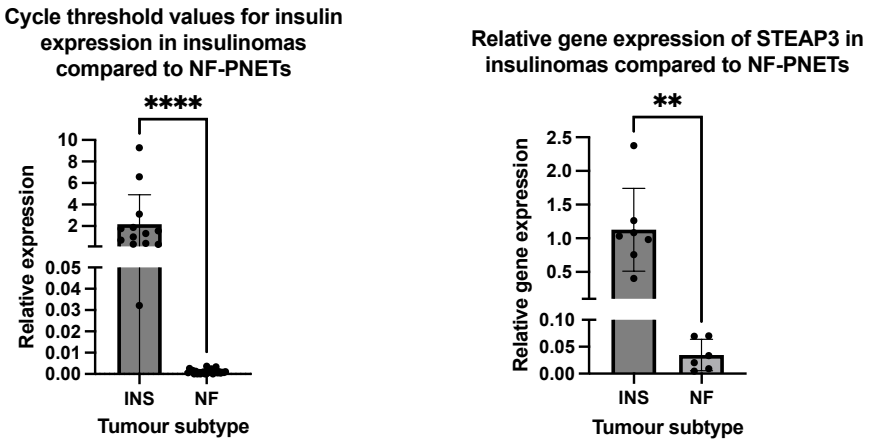


Figure 56: Cycle threshold values for insulin and STEAP3 comparing Insulinomas and NF-PNETs. Non-functioning PNET (NF) and Insulinoma (INS).

For the two genes tested for which NF-PNETs had higher expression than in insulinomas, no PCR product was available for insulinomas despite increasing the amplification cycle up to 45 cycles **(Figure 57)**.

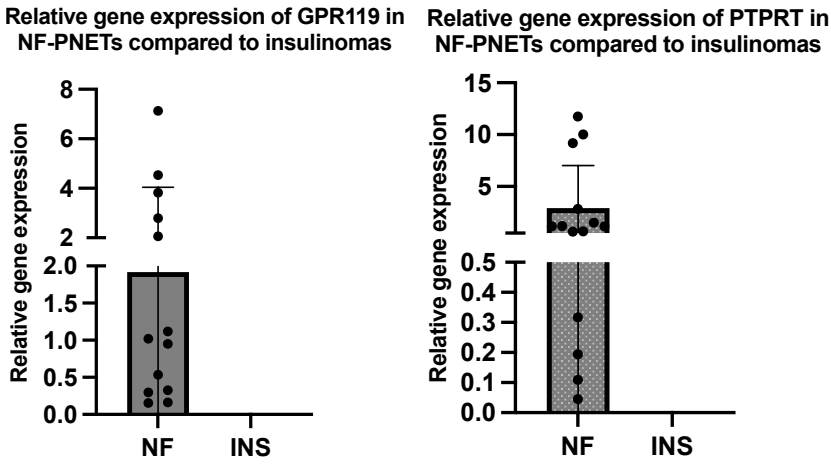


Figure 57: Cycle threshold values for GPR119 and PTPRT comparing NF-PNETs and insulinomas. Non-functioning PNET (NF) and Insulinoma (INS).

Overall, the Mayo cohort PNET RNA qRT-PCR confirmed the RNA-seq results from the PNET samples the University of Vienna, supports the possible mechanism by which insulinomas have higher DNA 5hmC than NF-PNETs.

4.3.2.8 Comparison of normal islet cells to PNETs

4.3.2.8.1 Comparison of human insulinoma cells to normal β cells

The bisulfite-only CpG methylation values for 9 insulinomas were first analysed and compared to that of the normal β cell. The bisulfite-only (as opposed to ox-BS) data for insulinomas were used as the normal pancreatic islet cell subtype DNA had also only been bisulfite converted. Additionally, no CpGs on *FFAR1* were significantly hydroxymethylated in insulinomas. Raw data files were first uploaded into R and analysed using the 'ChAMP' pipeline. The genomic co-ordinates (hg 19) for unmethylated DNA regions in the normal β cell were reported, including the gene ID and histone modifications (H3K4me3, H3K27me3). Using the genomic co-ordinates and gene ID for matching, methylation data from CpGs (beta values) from insulinomas were annotated with the regions of unmethylated DNA in normal β cells. There were 22.3% (161,538/725,289) CpG sites covered in the methylationEPIC array that were unmethylated in normal β cells. Out of these 0.9% (1,434/161,538) were inappropriately methylated in insulinomas (i.e average beta value >0.3). No significant KEGG pathways were enriched with inappropriately methylated genes with an FDR <0.05. Of the remaining CpGs in insulinomas that did not meet the criteria for being unmethylated in normal pancreatic β cells (i.e. were methylated), 15.5% (89,969/579,011) were hypomethylated. The methylation concordance between normal β cells and insulinomas was 89.5% (649,146/725,289). In regions

of unmethylated DNA in normal β cells, *MEN1* and *RASSF1* showed no aberrant hypermethylation; however, *DAXX* had 1% (3/315) CpGs that had a beta value >0.3 **(Figure 58).** This suggests that in insulinomas, *DAXX* may have aberrantly methylated CpG sites at the CGI of the *DAXX* promoter **(Figure 58 Panel C)**. Although *ATRX* is covered in the methylationEPIC array, all CpGs within this region did not return a beta value and were removed during the quality control process. Compared to both *MEN1* and *DAXX*, *RASSF1* had significant levels of DNA 5hmC across the whole gene, $p<0.01$ and $p<0.0001$, respectively. There was no difference in the levels of DNA 5hmC between *MEN1* and *DAXX* **(Figure 58 Panel D)**.

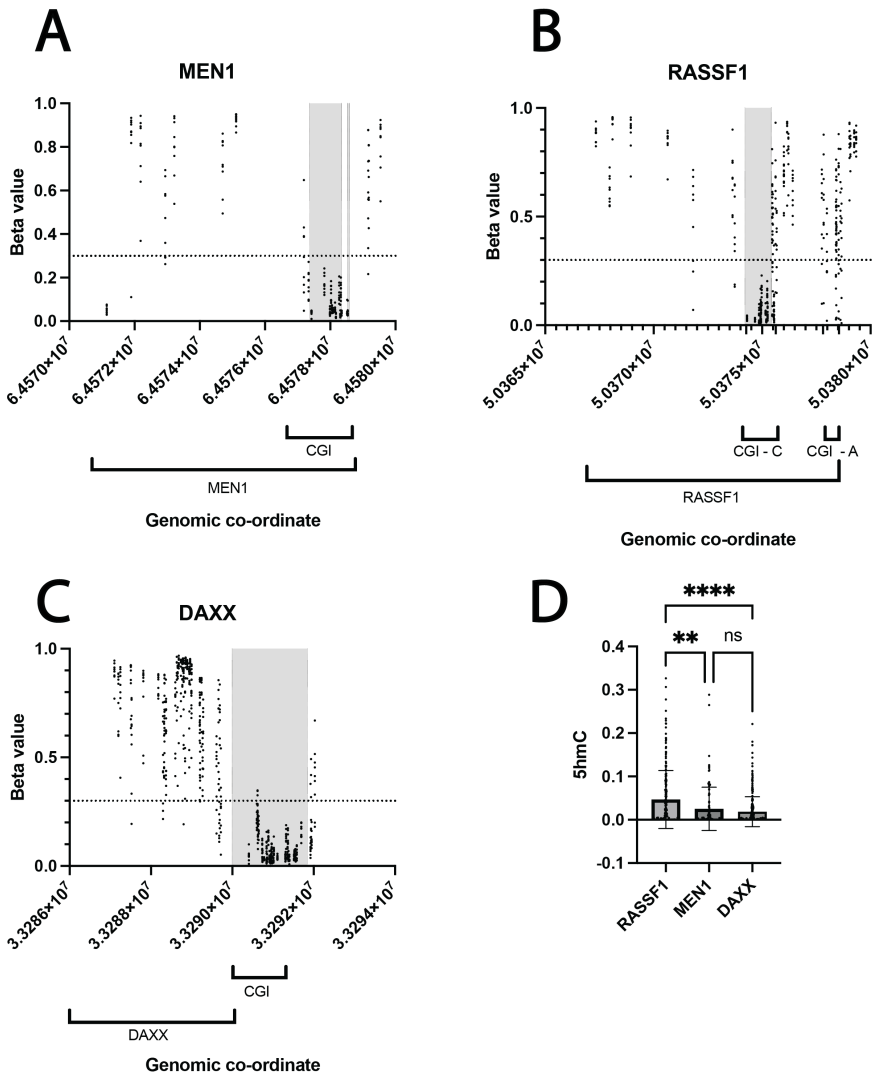


Figure 58: Shows the beta values for each CpG in nine insulinomas for three genes A) *MEN1*, B) *RASSF1* and C) *DAXX*. The horizontal dotted line indicates a beta value threshold of 0.3, and beta values below 0.3 correspond to unmethylated CpG sites. The shaded grey areas indicate the regions of unmethylated DNA in normal β cells, and if there is no aberrant methylation in insulinomas, then all points that lie within the shaded region should be below 0.3 (i.e. concordant with being unmethylated in the normal β cell) and any points above 0.3, suggests that the CpG site in the insulinoma is inappropriately methylated. For *MEN1* (Panel A) and *RASSF1* (Panel B), there is methylation concordance, however, for *DAXX* (Panel C) 3 CpG points within the shaded area are above 0.3 and suggest that these CpG sites are inappropriately methylated in insulinomas. Panel D) shows DNA 5hmC for each CpG covered in *RASSF1*, *MEN1* and *DAXX*. *RASSF1* has significantly higher DNA 5hmC than both *MEN1* and *DAXX*. CpG island (CGI). Not significant (ns).* p<0.05, ** p<0.01, ***p<0.001, ****p<0.0001.

Insulinomas had significantly higher aberrant hypomethylation than hypermethylation, $p<0.0001$, compared to normal β cells. GSEA analysis reported 62 KEGG pathways enriched with hypomethylated genes in insulinomas compared to normal β cells FDR <0.05; the top 15 pathways are shown in **Figure 59**. Insulin secretion pathway is significantly hypomethylated, as are Pathways in cancer and Proteoglycans in cancer. Overall, hypomethylated genes may be associated with increased gene transcription, and these results suggest that inappropriate overactivity of these pathways may contribute to insulinoma pathogenesis.

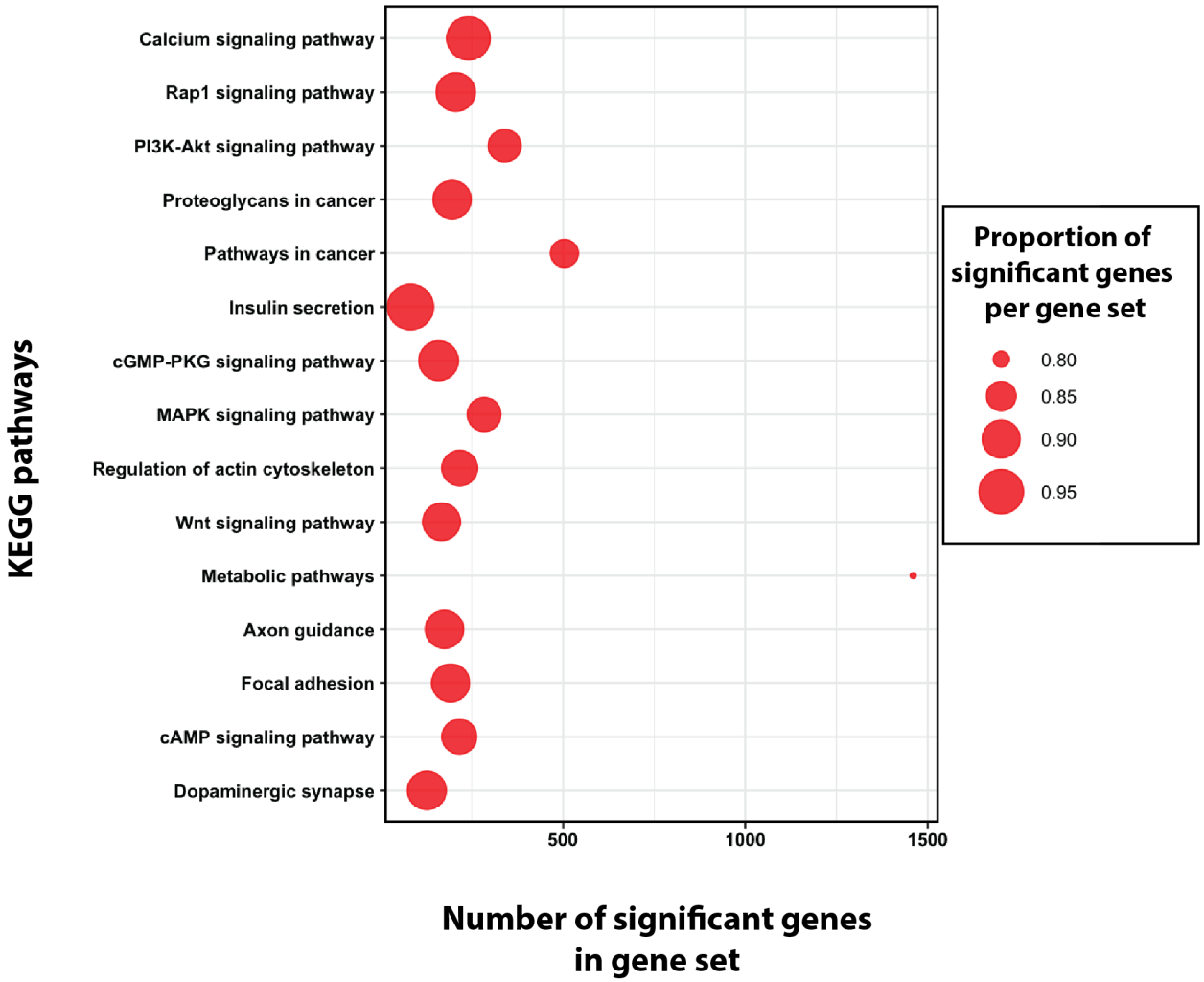


Figure 59: GSEA analysis for the top 15 KEGG pathways enriched with aberrantly hypomethylated genes in insulinomas compared to normal β cells.

4.3.2.8.2 Comparison of human NF-PNET cells to the normal α cell

There were 23.3% (166,788/715,522) CpG sites from 9 bisulfite-only converted NF-PNETs covered in the methylationEPIC array that were unmethylated in normal α cells. Out of these, 1.5% (2,572/166,788) were inappropriately methylated in NF-PNETs (i.e beta value >0.3). No significant KEGG pathways were enriched with inappropriately methylated genes with an FDR <0.05. Of the remaining CpGs that did not meet the criteria for being unmethylated in normal pancreatic α cells, 13.5% (77,241/574,141) CpGs in NF-PNETs were hypomethylated. The methylation concordance between normal α cells and NF-PNETs was 92.4% (661,116/715,522). In regions of unmethylated DNA in normal α cells, *MEN1* and *RASSF1* showed no aberrant hypermethylation; however, *DAXX* had 2% (7/315) CpGs that had a beta value >0.3 **(Figure 60).** Compared to normal α cells, NF-PNETs do not have aberrant hypermethylation of the TSGs *RASSF1*, which has been reported to be one of the most hypermethylated TSGs in PNETs. Whereas, in *DAXX* there are hypermethylated CpGs at the location of the CGI in NF-PNETs (an area of DNA which is important for gene transcription), which suggests a potential different cause of DAXX inactivation (as opposed to LOF genetic mutations seen in PNETs). A comparison between DNA 5hmC in *MEN1*, *RASSF1* and *DAXX* was unable to be undertaken, as there were no significantly hydroxmethyated CpGs **(Figure 53)**.

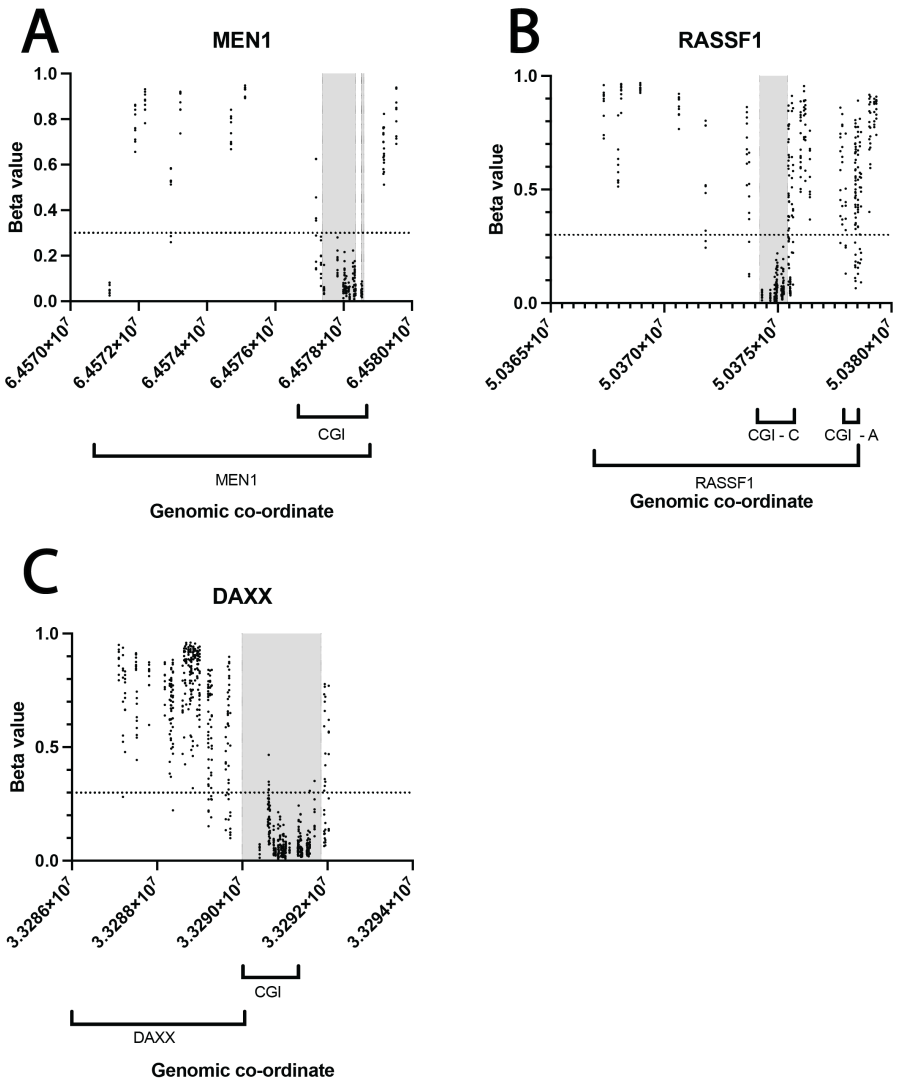


Figure 60: Shows the beta values for each CpG in nine NF-PNETs for three genes A) *MEN1*, B) *RASSF1* and C) *DAXX*. The horizontal dotted line indicates a beta value threshold of 0.3 and beta values below 0.3 correspond to unmethylated CpG sites. The shaded grey areas indicate the regions of unmethylated DNA in normal α cells, and if there is no aberrant methylation in NF-PNETs, then all points that lie within the shaded region should be below 0.3 (i.e. concordant with being unmethylated in the normal α cell) and any points above 0.3, suggests that the CpG site in the NF-PNET is inappropriately methylated. For *MEN1* (Panel A) and *RASSF1* (Panel B), there is methylation concordance, however, for *DAXX* (Panel C) 7 CpG sites within the shaded area are above 0.3 and suggest that these CpG sites are inappropriately methylated in NF-PNETs. The shaded grey areas indicate the regions of DNA unmethylated in normal β cells. CpG island (CGI).

NF-PNETs had significantly higher aberrant hypomethylation than hypermethylation, p<0.0001, which is consistent with aberrant DNA methylome changes seen in cancer. GSEA analysis reported 65 KEGG pathways enriched with hypomethylated genes in NF-PNETs compared to normal α cells FDR <0.05; the top 15 pathways are shown in **Figure 61**.

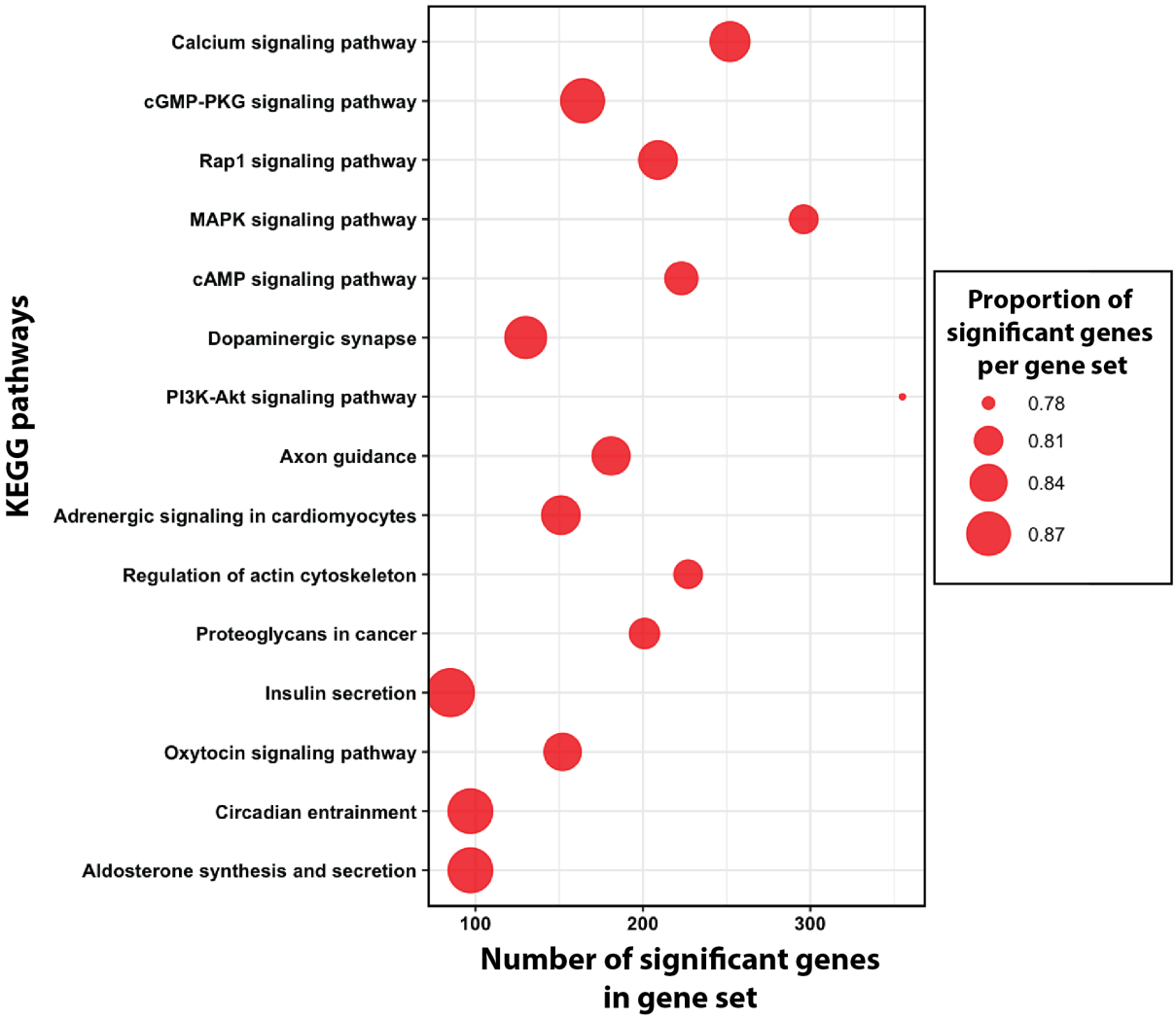


Figure 61: GSEA analysis for the top 15 KEGG pathways enriched with aberrantly hypomethylated genes in NF-PNETs compared to normal α cells.

Hypomethylated gene pathways included Proteoglycans in cancer, consistent with genes upregulated in carcinogenesis. Of note, the Insulin secretion pathway is hypomethylated, which suggests that a subset of NF-PNETs may have a β cell-like cell of origin. In both NF-PNETs and insulinomas, the Calcium signalling pathway was the most significantly hypomethylated, and suggests that upregulation of genes involved in calcium signalling may contribute to PNET pathogenesis.

4.3.2.8.3 Comparison of published DNA methylation (bisulphite only) for normal pancreatic tissue subtypes

Regions of unmethylated DNA (data generated using bisulphite only converted DNA) from published normal pancreatic cell subtypes (β, α, delta (δ), acinar and ductal cells) were extracted and compared. Pancreatic exocrine cell subtypes (acinar and ductal) had significantly shorter lengths of unmethylated DNA compared to all endocrine cell subtypes (β, α, δ), however, were more abundant **(Table 38 and Figure 62)**. Global DNA hypomethylation of CpGs interrogated by the EPIC methylation array, were significantly lower in endocrine cell subtypes compared to exocrine cell subtypes $p<0.0001$, however, the difference was small; global unmethylated DNA in acinar cell subtypes versus β cell subtypes were 20.61% (178,428/865,918) versus 18.66% (161,538/865,918), respectively, with a difference of 1.95% $p<0.0001$. Given that islet cells are significantly hydroxymethylated compared to exocrine cells, it is unclear if normal islet cells have comparable, lower or higher DNA methylation levels than pancreatic exocrine cell subtypes.

Table 38: Descriptive statistics for normal pancreatic cell subtypes

	Beta (β)	Alpha (α)	Delta (δ)	Acinar	Ductal
Median	702	576	785	317	413
25% percentile	322	263	389	153	199
75% percentile	1,206	1,127	1,254	847	986
Mean	843.1	762.3	901.6	566.4	652.4
Std. Dev.	633.4	628.3	624.7	586.9	606.2
SEM	3.89	3.59	33.99	2.71	3.08
Number of unmethylated DNA regions	26,523	30,640	24,542	46,788	38,749
Total DNA (base pairs)	22,360,378	23,357,108	22,127,718	26,499,517	25,27,276
Percentage of unmethylated CpGs in the EPIC array	18.66% (161,538/ 865,918)	19.26% (166,788/ 865,918)	18.52% (160,369/ 865,918)	20.61% (178,428/ 865,918)	20.14% (174,353/ 865,918)

Standard error of the mean (SEM); Standard deviation (Std. Dev).

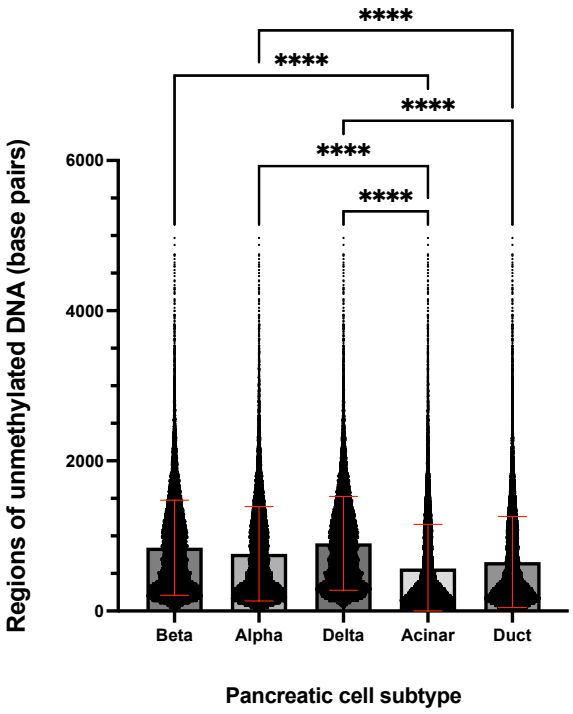


Figure 62: Regions of uninterrupted unmethylated DNA (base pairs) in each pancreatic cell subtype. Delta cells have the longest regions of unmethylated DNA, followed by Beta, Alpha, Ductal and Acinar cell subtypes.

4.3.2.8.4 Comparison of DNA methylation for normal α and β cells

When normal α and β cells were compared, the unmethylated regions covered by the methylationEPIC array were similar. There were 159,035 common unmethylated CpGs between normal β cells (159,035 /161,538; 98.5%) and normal α cells (159,035 /166,788; 95.4%) **(Figure 63)**.

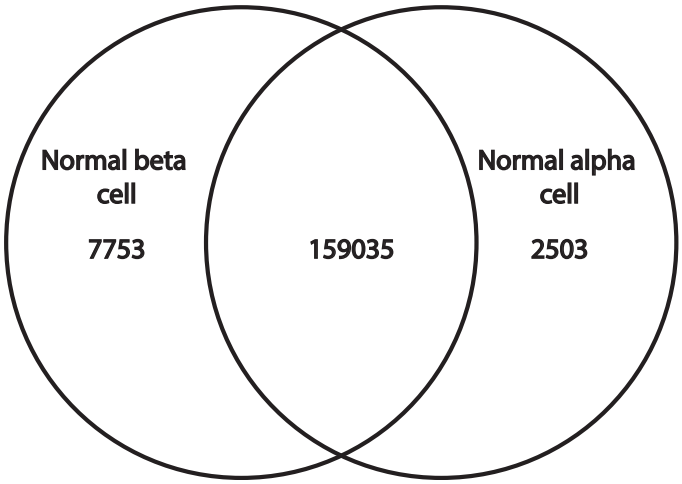


Figure 63: Venn diagram of the unmethylated regions in normal α and β pancreatic cells.

4.3.2.8.5 Comparison of DNA methylation between PNET subtypes

Using the averages of the beta values for each CpG site, there were 562,536 common hypomethylated CpGs (average beta value <0.3) in insulinomas (562,536/579,011; 97.2%) and NF-PNETs (562,536/574,141; 98.0%), and 1,190 common hypermethylated CpGs (average beta value >0.7) in insulinomas (1,190/1,434; 83.0%) and NF-PNETs (1,190/2,572; 46.3%). Out of the common hypermethylated genes, there were 1113 unique genes with at least 1 hypermethylated CpG site, of which 39 had been reported to be a human tumour suppressor gene **(Table 39).**

Table 39: Common genes with ≥1 hypermethylated CpG in both insulinomas and NF-PNETs

Genes								
AHCYL1	CDK2AP1	EIF3F	FOXP1	MAD1L1	PHLDA3	PTPN12	SOCS3	TRIM3
APC	CDKN1B	EPB41	GATA4	NDRG4	PPP2R1B	RASSF5	SPTBN1	TSC2
CADM1	CHST10	EXT2	HIC1	NEDD4L	PPP2R5C	RPS6KA2	ST5	WWOX
CCAR2	DLG1	EYA4	IRF4	NKX2-8	PRDM11	RUNX3	STARD13	
CDH4	EBF3	FHIT	LTF	PER2	PTCH2	SLC9A3R1	SYK	

Out of the 1113 uniquely hypermethylated genes **(Supplementary Appendix 8.1, Table 47)**, *FFAR1* is significantly hypermethylated across multiple CpG sites (n=4; cg00878516, cg03782727, cg06090864 and cg18410168), and has been associated with Hippo pathway modulation, therefore this gene was analysed.

4.3.2.9 Free Fatty Acid Receptor 1 gene

4.3.2.9.1 CpG methylation values of *FFAR1* and *FFAR4* compared to normal pancreatic β cells

The methylationEPIC array interrogates 12 CpGs across *FFAR1*, with 9 CpG sites available for analysis for both insulinomas and NF-PNETs. **Figure 64** shows the individual methylation values for the 9 insulinomas across *FFAR1* and the three regions of unmethylated CpGs in normal β cells. In all insulinomas (n=9), 55.6% (45/81) CpGs across the *FFAR1* are in regions of unmethylated DNA in normal β cells, and 64.4% (29/45) were inappropriately methylated in insulinomas. These 29 inappropriately methylated CpGs, were from 7 of the 9 patients with insulinomas. The remaining 2 patients with insulinomas had methylation concordance (i.e. unmethylated CpGs in tumours that were also unmethylated in the normal β cell). Of note, there were no significantly hydroxymethylated CpGs across *FFAR1*. In NF-PNETs, there were no methhylated CpGs in the region of unmethylated DNA in normal α cells **(Figure 64)**. There

were 3 CpGs sites covered across *FFAR4* and these were not in regions of unmethylated DNA in either α or β cells.

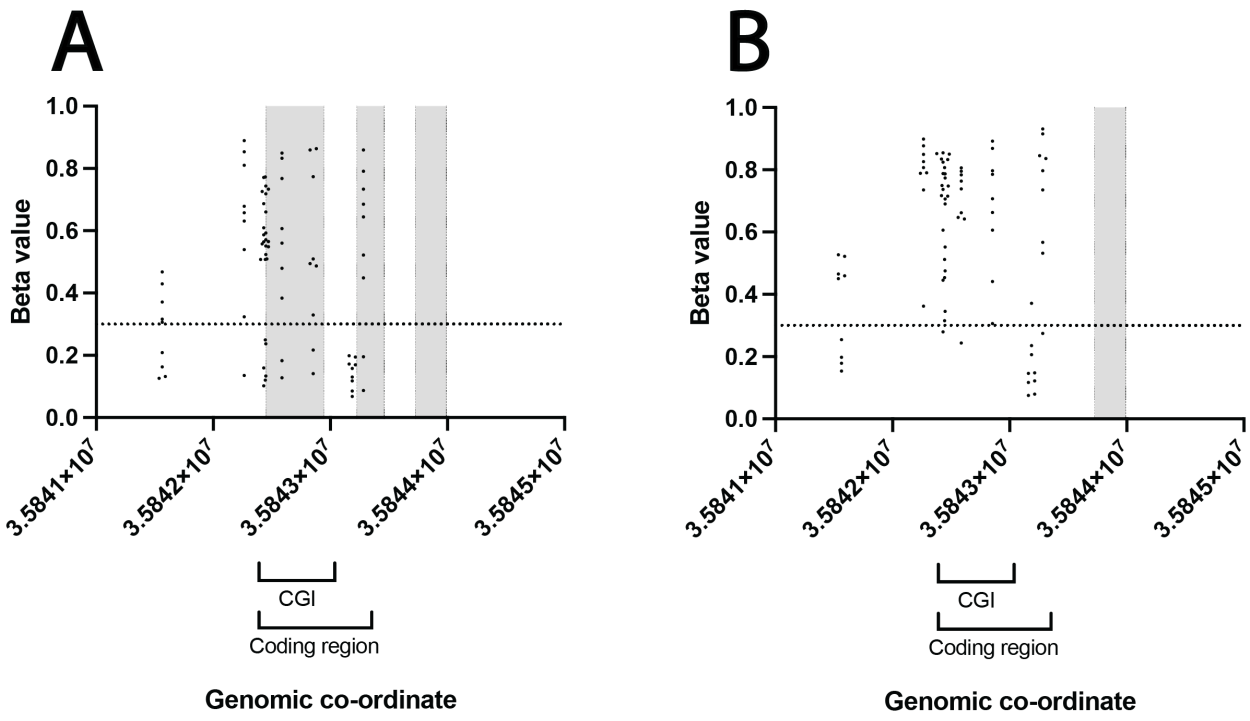


Figure 64: *FFAR1* with the beta values for the insulinomas (A) and NF-PNETs (B). Grey shaded areas indicate regions of unmethylated CpGs in normal β and α cells, respectively. The horizontal dotted line at 0.3 indicates the cut-off level to define aberrant hypermethylation, and beta values below 0.3 correspond to unmethylated CpG sites. The shaded grey areas indicate the regions of unmethylated DNA in normal β and α cells, respectively, and if there is no aberrant methylation in PNETs, then all points that lie within the shaded regions should be below 0.3 (i.e. concordant with being unmethylated in the normal β or α cell) and any points above 0.3, suggests that the CpG site in the PNET is inappropriately methylated. For *FFAR1* in insulinomas (Panel A) 29 CpG sites within the shaded areas are above 0.3 and suggest that these CpG sites are inappropriately methylated in insulinomas. Whereas, for *FFAR1* in NF-PNETs (Panel B), there are no inappropriately methylated CpGs. CpG island (CGI).

4.3.2.10 Comparison of FFAR1 and FFAR4 expression in insulinomas compared to normal islet-containing exocrine pancreatic tissue

RNA-seq data was compared between normal islet-containing exocrine pancreatic tissue (n=10) and insulinomas (tumour only; n=6). This revealed 803 differentially expressed genes (FDR p<0.01), of which 610 were upregulated and 193 downregulated in insulinomas compared to normal pancreatic exocrine tissue **(Figure 65)**. As expected, mRNA expression in insulinomas was enriched for genes including *INS-IGR-2*, synaptophysin (*SYP*) and pancreatic and duodenal homeobox 1 (*PDX1*) and in normal pancreatic tissue, it was enriched for Chymotrypsin C (*CTRC*), Carboxypeptidase A2 (*CPA2*) and Colipase (*CLPS*) genes.

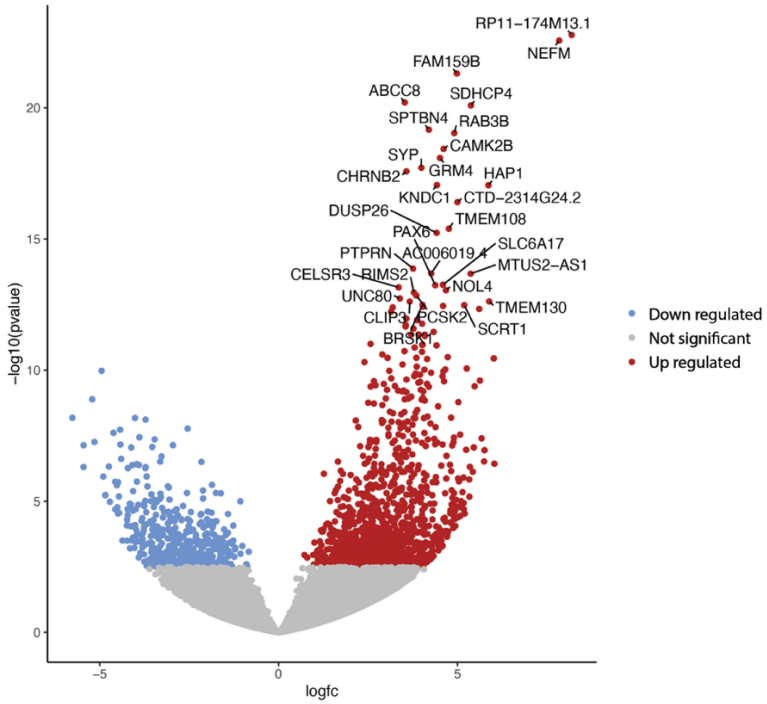


Figure 65: Volcano plot reporting the top 803 differentially expressed genes (FDR p<0.01) comparing insulinomas to normal pancreatic exocrine tissue. The red circles are genes upregulated, and blue circles are genes downregulated in insulinomas compared to normal pancreatic exocrine tissue. The top 30 genes (out of 803) are annotated with the gene name.

GSEA analysis shows the KEGG-enriched pathways between insulinomas and normal islet-containing exocrine pancreatic tissue **(Figure 66)**. The two most enriched pathways are in

pancreatic secretion (exocrine tissue) and maturity onset diabetes of the young (insulinomas). This provides further evidence of the validity of using RNA from FFPE samples, as GSEA pathways clearly distinguish between insulinomas and normal islet-containing exocrine pancreatic tissue. As expected, most mRNA extracted was of pancreatic exocrine origin, and this was confirmed by mRNA sequencing results.

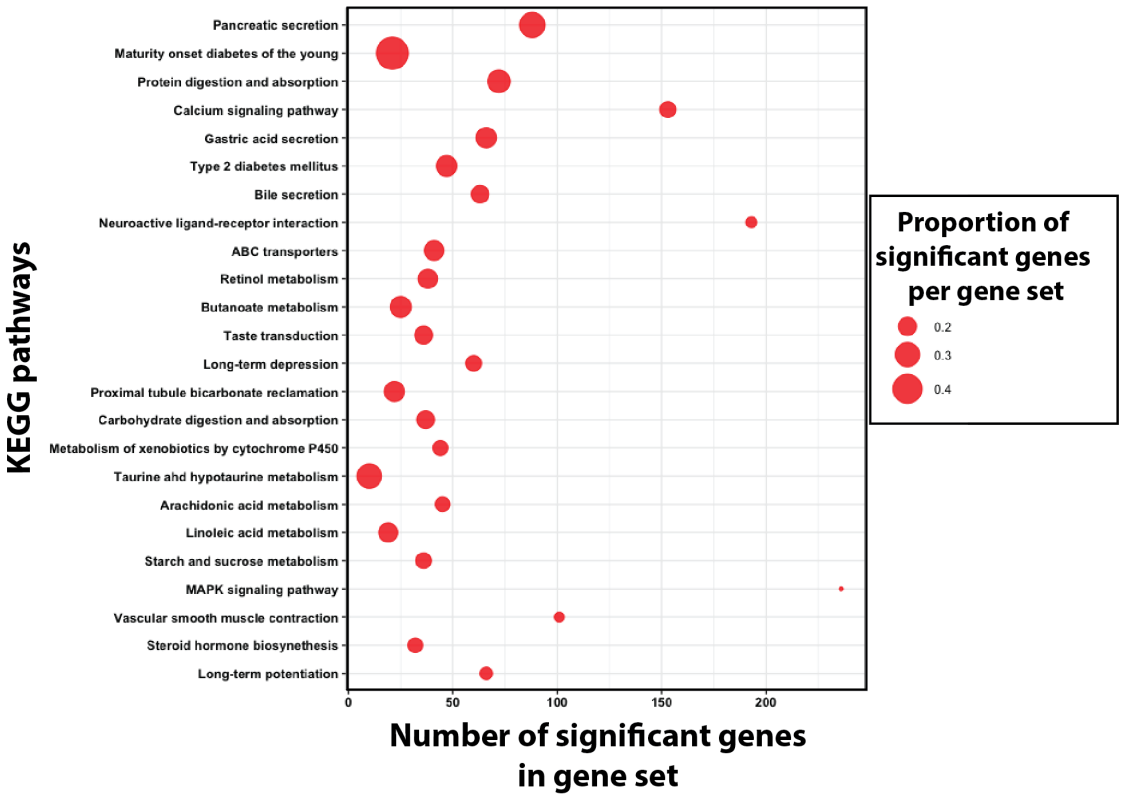


Figure 66: The top 24 KEGG enriched pathways in insulinomas vs normal pancreatic exocrine tissue. The two most significant KEGG pathways were pancreatic secretion (pancreatic exocrine tissue), and maturity onset diabetes of the young (insulinomas).

There was no difference in mRNA expression levels between insulinomas and normal islet-containing exocrine pancreatic tissue for *FFAR1*, *FFAR4*, *YAP1*, *DAXX*, *MEN1* or *RASSF1*. Given that *FFAR1* is highly expressed in normal β cells compared to normal islet-containing exocrine pancreatic tissue, as there is no difference between *FFAR1* mRNA expression between

insulinomas and normal islet-containing exocrine pancreatic tissue, this suggests that *FFAR1* translation is lost in insulinomas, which may be due to inappropriate methylation.

In summary, DNA 5hmC is increased in pancreatic islet cell subtypes compared to pancreatic exocrine cell subtypes; and in PNETs, DNA 5hmC is lower in insulinomas and significantly lower in NF-PNETs, compared to normal islet cells **(Table 40)**. DNA 5mC (bisulfite only) in pancreatic endocrine cell subtypes is significantly higher than pancreatic exocrine cell subtypes, although the difference is small which could be attributable to DNA 5hmC inappropriately categorised as DNA 5mC. Therefore, future studies investigating the DNA methylome in human islet cells, must account for DNA 5hmC.

Table 40: Summary of DNA 5mC and 5hmC changes in human PNETs compared to normal pancreatic islets, exocrine tissue and PNETs.

DNA 5mC (bisulfite only)	Normal islet	Insulinoma	NF-PNET	Pancreatic exocrine tissue
Insulinoma	↑	NA	↔	NA
NF-PNET	↑	↔	NA	NA
Normal islet	NA	↓	↓	↔/↓
Pancreatic exocrine tissue	↔/↑	NA	NA	NA
DNA 5hmC	**Normal islet**	**Insulinoma**	**NF-PNET**	**Pancreatic exocrine tissue**
Insulinoma	↔	NA	↓	↔
NF-PNET	↑	↑	NA	↔
Normal islet	NA	↔	↓	↓
Pancreatic exocrine tissue	↑	↔	NA	NA

4.3.3 Discussion

The results from this Chapter confirm that DNA 5hmC is higher in insulinomas than NF-PNETs, and that both PNET subtypes have lower 5hmC than normal islet-containing exocrine pancreatic tissue. Given that normal pancreatic islet cells are heavily hydroxymethylated compared to normal pancreatic exocrine tissue, GSEA pathways enriched in normal islet-containing exocrine pancreatic tissue may reflect those of normal islet cells. Two of the most hydroxymethylated pathways in normal islet-containing exocrine pancreatic tissue included the Insulin signalling and the Hippo signalling pathways. The insulin signalling pathway is essential for pancreatic β cells;[267] therefore, it is expected that many genes within this pathway are hydroxymethylated. The Hippo signalling pathway is important for overall pancreatic cell development, including β cell survival, proliferation and regeneration.[268,269] Although present early on in pancreatic cell development, it has been reported that YAP expression, which is part of the Hippo signalling pathway, is lost in mature endocrine islet cells,[270] however is maintained in normal pancreatic exocrine cells, and is required to maintain acinar cell differentiation.[271] YAP can either be phosphorylated (inactive) and sequestered in the cytoplasm, or unphosphorylated (active) where it translocates to the nucleus leading to cellular proliferation by increasing the expression of TEAD transcription factors **(Introduction 1.6.2, Figure 11)**.[260] However, insulinomas and NF-PNETs aberrantly express YAP mRNA at levels that are not significantly different from normal islet-containing exocrine pancreatic tissue. *YAP* is not significantly hydroxymethylated in PNETs, and in normal pancreatic β and α cells, most DNA is methylated, which is consistent with YAP not being expressed in normal pancreatic islets.[270] In PNETs, there is a higher number of hypomethylated CpG sites in *YAP*, therefore, aberrant hypomethylation at *YAP* may be a driver and/or contribute to PNET tumourigenesis. Recently,

it was reported that FFAR1/FFAR4 agonists caused YAP phosphorylation and its sequestration in the cytoplasm, inhibiting its proliferative action.[141] FFAR1 is a Class A GPCR and is reported to be highly expressed in normal β cells and is encoded by *FFAR1* located on 19q13.[272] FFAR1 expression is modulated by PDX1 and the STAT family of proteins binding to the evolutionary conserved HR2 domain located 5' to the TSS of *FFAR1* and decreased expression can be seen in gluco-lipotoxic conditions.[273] It has been reported that GPCRs can signal through multiple Gα proteins,[274] and FFAR1 has been reported to signal through both G_q, G_s and partially G_i.[272,275] FFAR1 has been reported to activate the PLC/IP3/Ca^{2+} pathway through G_q and potentiates glucose stimulated insulin release, and to increase cAMP production through G_q activation of and adenylate cyclase 2 (ADCY2),[275] and increased levels of cAMP has been reported to increase DNA 5hmC.[91] FFAR1 activation of G_s proteins has been reported to activate the Hippo signalling pathway,[141] and therefore loss of FFAR1 may contribute to tumourigenesis **(Figure 67)**.

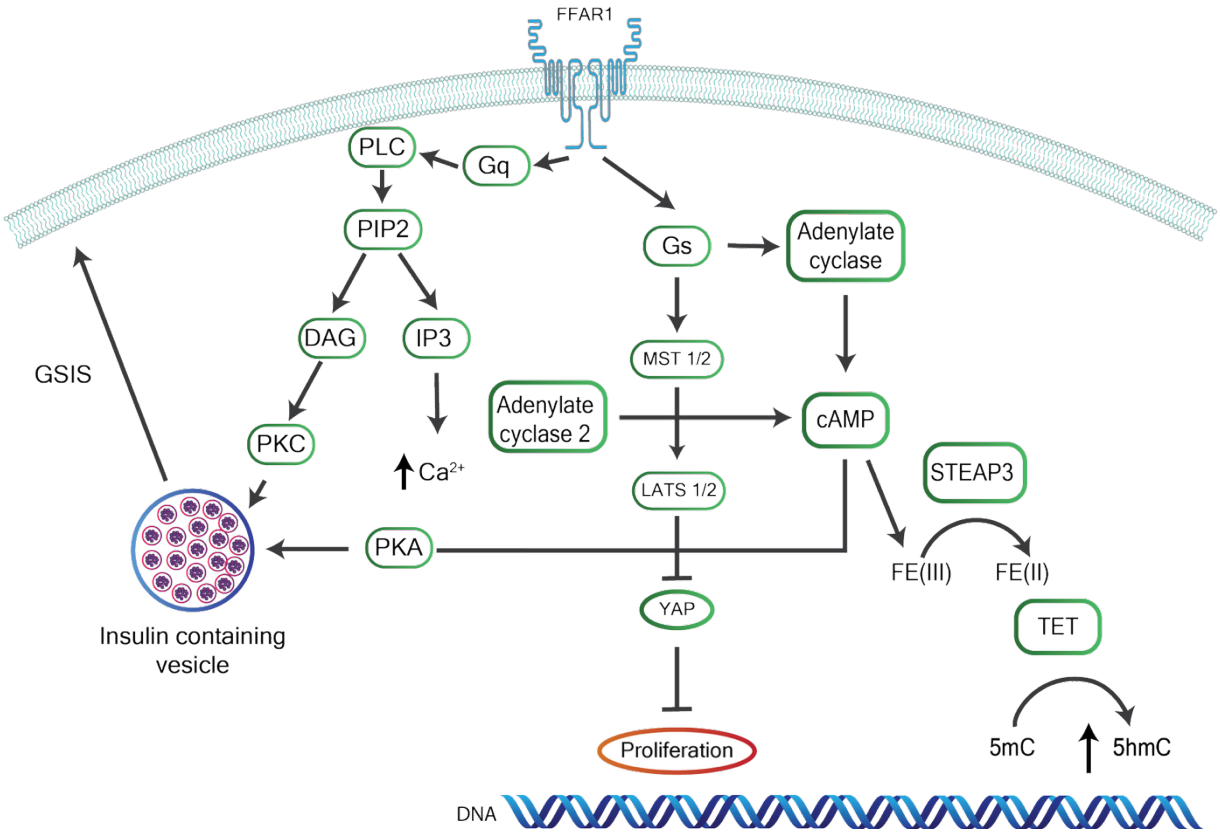


Figure 67: Schematic diagram of the signalling pathways for activation of FFAR1. FFAR1 activation of Gs activates the anti-proliferative Hippo signalling pathway and also increases cAMP. It has been reported that cAMP increases DNA 5hmC. Both Gs and Gq pathways increase PKA and PKC, respectively to stimulate glucose stimulated insulin secretion. In insulinomas, FFAR1 is decreased, and YAP is increased which may decrease DNA 5hmC and increase cell proliferation. Cyclic AMP (cAMP); Diacylglycerol (DAG); Free fatty acid receptor 1 (FFAR1); G protein q (Gq); G protein s (Gs); Glucose stimulated insulin secretion (GSIS); Inositol 1,4,5-triphosphate (IP3); Large tumor suppressor kinase 1/2 (LATS 1/2); Macrophage stimulating 1/2 (MST1/2); Phospholipase A (PKA); Phospholipase C (PLC); Phosphatidylinositol 4,5-bisphosphonate (PIP2); STEAP3 Metalloreductase (STEAP3); Ten eleven translocase (TET); yes-associated protein 1 (YAP).

In our study, up to two thirds of CpGs in *FFAR1* which are unmethylated in normal β cells, were methylated in insulinomas. FFAR1 may be an important modulator of the Hippo signalling pathway in β cells, and its loss due to aberrant DNA methylation may contribute to PNET pathogenesis **(Figure 67)**. It has been reported that *FFAR1* mRNA is highly expressed

in both normal islet cells (n=2) and in insulinomas (n=2), but not in a glucagonoma (n=1) or gastrinoma (n=1), and that FFAR1 mRNA expression is ~20 fold higher in normal pancreatic islets compared to whole pancreatic tissue (endocrine and exocrine pancreas).[261] This is partly in keeping with the results from this study, which showed no significant difference in *FFAR1* mRNA expression in islet-containing pancreatic exocrine tissue and NF-PNETs, however, *FFAR1* expression was also low in insulinomas (not high), which was concordant with the increased methylation of *FFAR1* seen in this study. As it is unclear if DNA 5hmC is higher in normal pancreatic β cells compared to α cells, and whether FFAR1 is expressed on normal pancreatic α cells, it isn't possible to determine if aberrant expression of FFAR1 in PNET subtypes contribute to the differences seen in DNA 5hmC in insulinomas compared to NF-PNETs.

RASSF1A has been reported to be the most aberrantly methylated tumour suppressor in PNETs.[114-117,123,128]. However, in 18 PNETs (9 insulinomas and 9 NF-PNETs), there were no methylated CpGs in regions of DNA reported to be unmethylated in normal α or β cells, which suggests that aberrant methylation of *RASSF1* is not a driver in our PNET cohort. *RASSF1* has two promoters (CGI A and B), and 7 different transcripts (A-G),[260] and it has been reported that transcript A (RASSF1A) is a common TSG methylated in cancer. However, *RASSF1* was found to be significantly hydroxymethylated in insulinomas (but not in NF-PNETs). Therefore, discrepancies reported between *RASSF1* tissue expression and DNA methylation may be due to the presence of DNA 5hmC at *RASSF1*. In both insulinomas and NF-PNETs, inappropriate methylation was seen at CpG sites located within the CpG island at the *DAXX* promoter. Similar results have been reported in one study of 48 PNETs, with almost all PNETs having *DAXX*

hypermethylation[138] which suggests that aberrant methylation of *DAXX* may be an important mechanism for DAXX inactivation. There was no aberrant DNA methylation seen in *MEN1* in either PNET subtype.

Overall, PNETs were hypomethylated compared to normal α and β cells. The most hypomethylated pathway in PNET subtypes was the calcium signalling pathway, which has not been reported in PNETs. Increased calcium signalling has been reported to be crucial to tumour progression in cancers (e.g. neuroendocrine differentiation of prostate cancer[276] and phaeochromocytoma cell lines),[277] and in epithelial to mesenchymal transition (EMT).[278] Additionally, pharmacological agents have been trialled to target increased calcium signalling through the inhibition of voltage-gated calcium channels.[276]

It is unclear why normal pancreatic islet cells are heavily hydroxymethylated compared to other tissue. DNA 5hmC is higher in insulinomas vs NF-PNETs, which may be due to: 1) Insulinomas are a defined subset of PNETs that retain their ability to secrete insulin and compared to NF-PNETs, are more likely to be epigenetically similar. 2) Insulinomas retain the ability to secrete the correct peptide from their parent cell i.e. the β cell. They are therefore more likely more differentiated than the heterogenous clinically non-secretory NF-PNETs, of which the cell of origin is unclear, although a proportion are derived from the α cell. DNA 5hmC is also important for β cell differentiation and maturation[279] and 3) looking at differentially expressed genes between insulinomas and NF-PNETs, three key genes were identified which may contribute to 5hmC differences in PNET subtypes including *STEAP3*, *GLP1* and *CAMK2G* **(Figure 54)**. It has been reported that cAMP activation, increases DNA 5hmC, via increasing

the labile iron pool.[91] Free iron and α-ketoglutarate are essential for TET enzymes function and are responsible for maintaining the DNA hydroxymethylome **(Introduction, section 1.5.2, Figure 9)**.[85] It has been reported that GLP-1 is over-expressed in insulinomas, and expression decreases/is lost in metastatic disease.[280] The GLP-1 receptor is a key target for T2DM drugs, and GLP-1 agonists are important for glucose-mediated insulin secretion.[281] STEAP3, a metalloprotease, is an important enzyme responsible for converting inactive iron (Fe^{3+}) to active iron (Fe^{2+}), and it is Fe^{2+} which is needed as a cofactor for TETs action. Therefore, increased STEAP3 as seen in insulinomas (or loss of STEAP3 in NF-PNETs), may help to maintain β cell DNA 5hmC. CAMK2G activation suppresses adenylate cyclase, thereby decreasing cAMP levels.

Conclusion: Overall, the results from this Chapter strengthen the case that PNETs lose DNA 5hmC compared to normal pancreatic islets, and that aberrant loss of expression of key β cell genes (eg. *FFAR1*) or *DAXX* may contribute to PNET pathogenesis. Improving the DNA hydroxymethylome with pharmacological agents may help to protect against aberrant methylation and may improve tumour phenotype, given that DNA 5hmC is associated with cell differentiation.

4.4 Chapter 4: Modulating DNA hydroxymethylation in PNETs

4.4.1 Introduction

As shown in Chapter 3, human insulinomas have more DNA 5hmC when compared to NF-PNETs. This difference in DNA 5hmC could be due to increased activation of cAMP and STEAP3, thereby increasing DNA 5hmC in insulinomas compared to NF-PNETs. Two compounds, somatostatin analogues (SSAs) and ascorbic acid (AA) could be used to modulate DNA 5hmC in PNETs, by targeting different pathways. SSAs are the most commonly used medical therapy to treat both functional consequences of PNETs and have an anti-proliferative action.[282] SSAs can modulate cAMP, and it has been shown that SSAs decrease cAMP.[283] The effect of SSAs on DNA 5hmC has not been reported in PNETs. Another compound, which has a similar function to STEAP3 is AA. AA has been shown to re-express genes previously repressed in PDAC (i.e. GATA binding protein 6 (*GATA6)*) leading to an improved cancer phenotype.[211] It has been reported that AA could be used as a cancer therapeutic, and parental AA has mainly been used to achieve supraphysiological levels of AA, as these high doses have been shown to have a cytotoxic effect.[284] Therefore, the effects of AA on DNA 5hmC in a mouse model for MEN1 and PNET cell lines were assessed.

Two human PNET cell lines were used in this study, BON1 (derived from the lymph node from a 28-year-old patient with metastatic insulinoma)[218] and QGP1 cells (derived from a metastatic somatostatinoma).[219] QGP1 cells show no evidence of hormone secretion *in vitro*,[219] and therefore, are used as a model for a NF-PNET.

4.4.2 Results

4.4.2.1 Effect of somatostatin analogues on DNA 5hmC

4.4.2.1.1 *Men1$^{L/L}$/RIP2 Cre-ER* mice treated with octreotide

Men1$^{L/L}$/RIP2 Cre-ER mice, a model for MEN1 that develops PNETs, were treated with intraperitoneal injections 50mg/kg of Octreotide twice weekly for 1 month. Cut sections already mounted on IHC slides were available for analysis from 11 mice treated with vehicle, and 4 mice treated with high dose octreotide. **Figure 68** shows an example of DNA 5hmC staining in vehicle vs mice treated with high-dose octreotide.

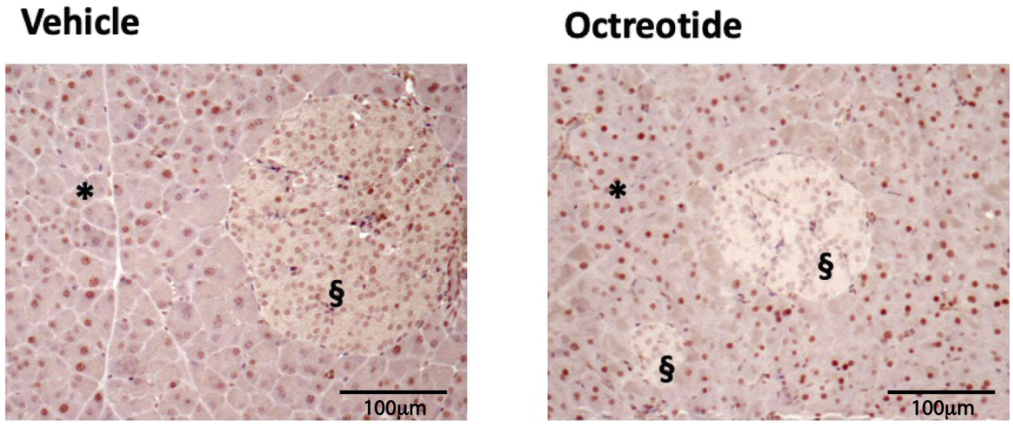


Figure 68: IHC samples from two 6-month-old Men1 L/L RIP2 Cre mice treated with either vehicle or high-dose octreotide for 4 weeks. * Exocrine tissue; § Islet cell tumours.

The mouse identification, sex and raw H scores for tumours and adjacent exocrine tissue are shown in **Table 41**.

Table 41: *Men1*$^{L/L}$*/RIP2 Cre-ER* mice cohort treated with octreotide

Treatment group	Identification	Sex	Tumour (H score)	Exocrine Tissue (H score)
Vehicle				
	MENC 24.2d	M	102.225	84.385
	MENC 36.4b	M	139.586667	103.2957
	MENC 36.4g	F	106.356667	114.216667
	MENC 38.1k	F	17.525	88.015
	MENC 53.2a	M	97.35405	121.877875
	MENC 54.1b	M	94.2190333	100.415067
	MENC 55.1d	F	102.692333	98.5076
	MENC 27.1a	M	89.0541	83.7206333
	MENC 30.2i	F	81.7678667	46.5217333
	MENC 36.4a	F	92.8515667	50.7763667
	MENC 55.1c	F	97.9066	102.647333
High-dose Octreotide 30mg/kg				
	MENC 36.4d	M	47.36	102.095
	MENC 36.4c	M	38.1269667	142.7196
	MENC 36.4h	F	22.0933333	102.968367
	MENC 36.4i	F	52.6366667	147.178467

High dose octreotide significantly decreased DNA 5hmC in mouse insulinomas compared to adjacent exocrine tissue, with an ~3 fold decrease in 5hmC **(Figure 69)**. There was no significant difference in 5hmC in insulinomas compared to exocrine tissue in mice treated with vehicle.

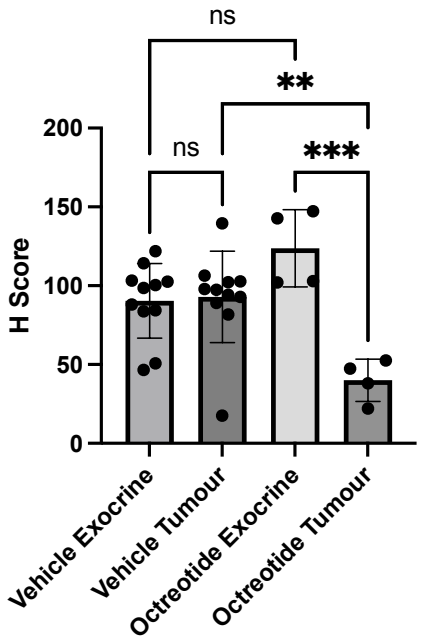


Figure 69: 5hmC by IHC comparing vehicle and octreotide insulinomas in *Men1$^{L/L}$/RIP2 Cre-ER* mice compared to adjacent exocrine tissue. **p <0.01, ***p<0.001.

4.4.2.2 Modulating DNA 5hmC in cell lines

SSTR subtype and menin expression were determined in the BON1 cell line. Raw FASTQ files were available from three replicates from the BON1 cell line and RNA-seq analysis was undertaken on the online Galaxy server and in R Studio.

4.4.2.2.1 RNA for SSTR and menin expression

Given that menin is important for PNET pathogenesis in all MEN1 patients and ~40% of non-MEN1 patients, first, its expression was compared across 13 human PNETs (insulinomas n=6) and (NF-PNETs =7) and the BON1 cell line. There was no significant difference in menin expression in insulinomas compared to NF-PNETs. However, menin expression in BON1 cells was significantly lower (almost zero) than in both PNET subtypes (~370-fold lower; **Figure 70**).

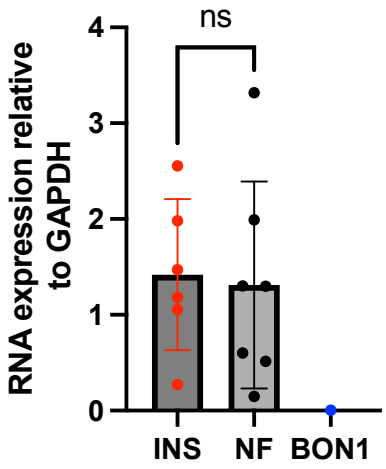


Figure 70: RNA expression for menin in PNET subtypes compared to BON1 cells. INS: Insulinomas (red circles); NF: NF-PNETs (black circles) and BON1 cell line (blue circles).

Next, SSTR expression was reviewed in BON1 cell lines and human insulinomas and NF-PNETs. There was a lower expression of SSTR1, SSTR2 and SSTR5 in PNETs compared to the BON1 cell line **(Figure 71)**. Therefore, assessing the effect of SSTR2 activation on DNA 5hmC requires SSTR transfection into the BON1 cell line.

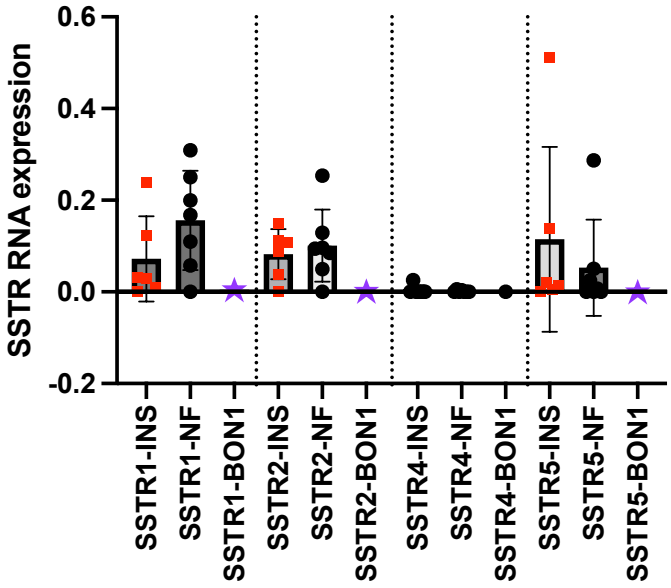


Figure 71: RNA expression of SSTR relative to GAPDH (housekeeper) in PNET subtypes INS (insulinomas; red squares) and NF (NF-PNETs; black circles), and in BON-1 cell line (purple stars).

4.4.2.3 Modulating DNA 5hmC with ascorbic acid

QGP1 and BON1 cells were treated with increasing doses of ascorbic acid (AA) only, and cell viability was assessed on day 7. AA had a greater effect on cell viability in QGP1 cells compared to BON1 cells **(Figure 72)**. There was a significant decrease in cell viability with 500µM of AA in both cell lines and at 100µM for QGP1 cells only.

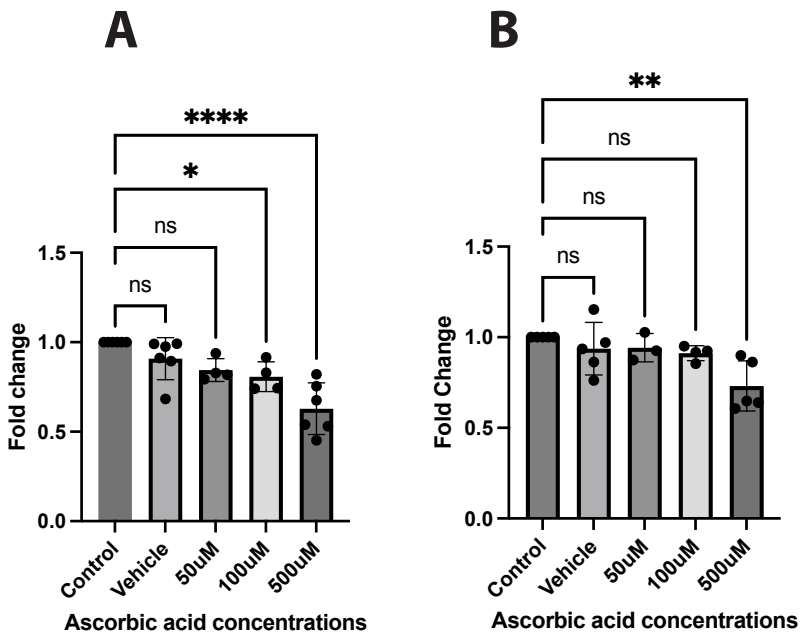


Figure 72: Day 7 proliferation assays for both QGP1 (Panel A) and BON1 (Panel B) cells. * p<0.05, **p<0.01, ****p<0.0001.

The dose of 100µM of AA was subsequently taken forward, as this is the upper limit of which AA levels may be achieved in humans with oral administration. DNA 5hmC was assessed at days 1-7 in QGP1 and BON1 cells. QGP1 DNA 5hmC is shown in **Figure 73**, and over time, the average DNA 5hmC is higher than the vehicle at each time point.

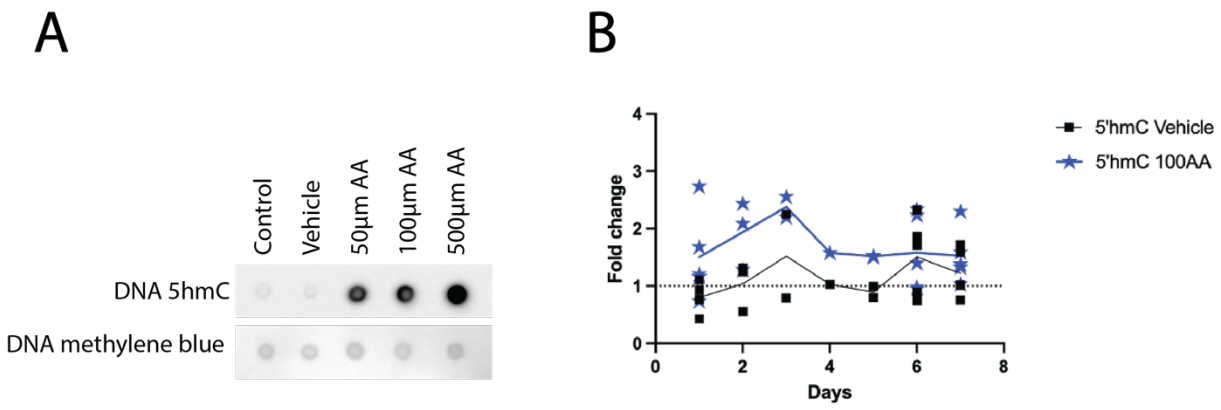


Figure 73: Panel A) An example of DNA 5hmC dot blot in QGP1 cells at day 2, treated with 50µM, 100µM and 500µM of AA at day 0. Panel B) DNA 5hmC in QGP1 cells treated with one dose of AA taken over a period of 7 days. DNA 5hmC in 100µM AA treated cells (blue stars) compared to vehicle (black boxes). Ascorbic acid (AA).

When compared to BON1 cells, initially there is an increase in DNA 5hmC at day 1, which subsequently decreased and returned to baseline by day 2 **(Figure 74)**.

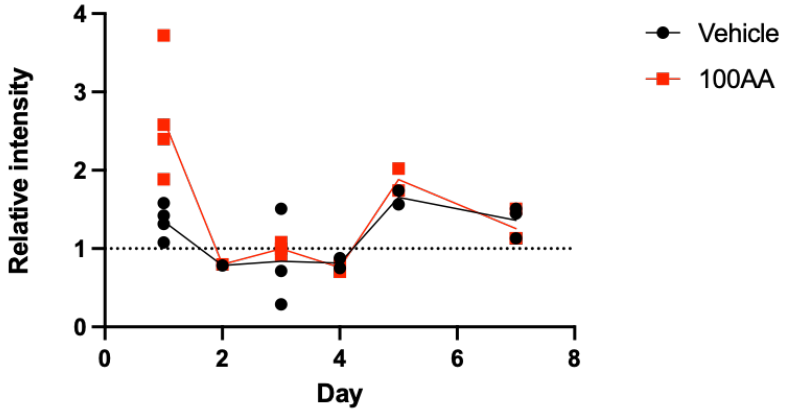


Figure 74: DNA 5hmC in BON1 cells treated with one dose of 100µM AA taken over a period of 7 days. DNA 5hmC in 100µM AA treated cells (red boxes) compared to vehicle (black circles).

All time points were combined for each cell type and DNA 5hmC levels compared **(Figure 75)**. QGP1 cells have higher DNA 5hmC across the 7 days than BON1 cells ($p<0.01$), which may

reflect NF-PNETs (e.g. QGP1 cells) having less DNA 5hmC compared to insulinomas (BON1 cells).

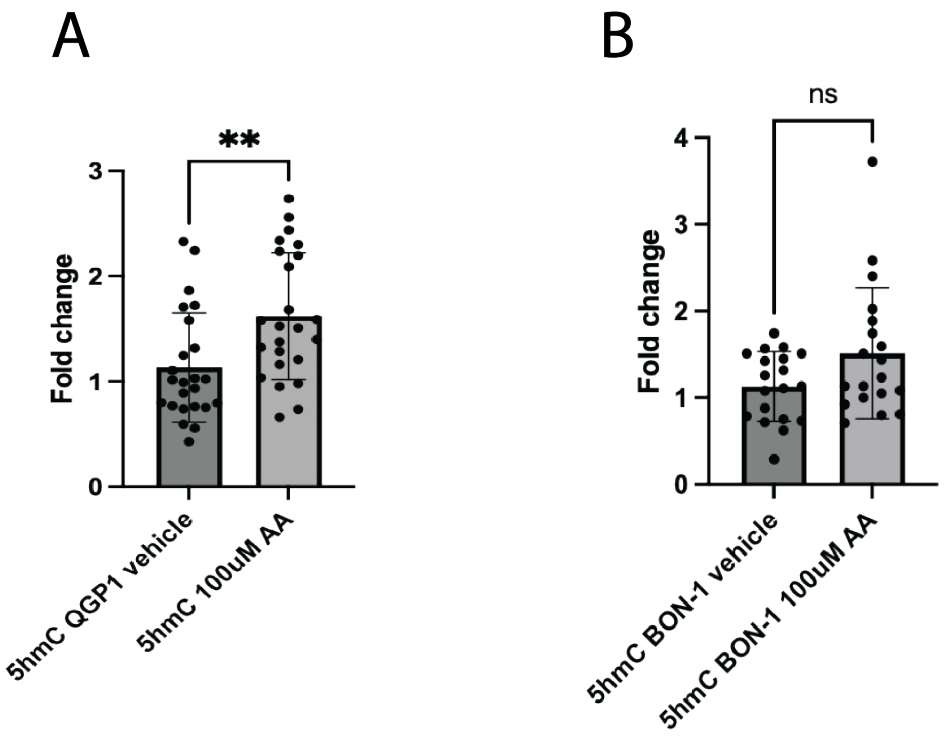


Figure 75: Panel A) Overall DNA 5hmC taken over days 1-7 in QGP1 and Panel B) BON1 cells treated with one dose of 100μM of AA. ns: not significant; **p<0.01.

4.4.2.4 Ex-vivo mouse study investigating the effect of Ascorbic acid on DNA5hmC and cell viability.

To determine DNA 5hmC in normal mouse islets and insulinomas, and to compare the effect of AA on mouse islets, an ex-vivo mouse study was undertaken. Mice were genotyped and bred to age 6-8 months.



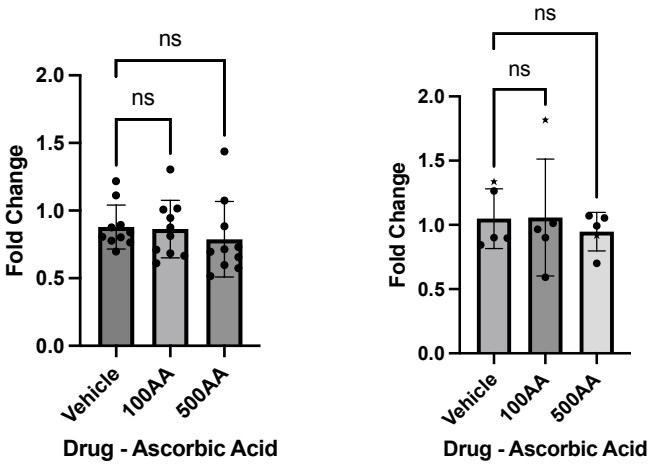


Figure 76: Cell viability assay for mouse islets treated with increasing doses of AA at day 3 (Panel A) and day 5 (Panel B).

Cell viability assay was performed at days 3 and 5 for cultured mouse islets treated with either 100μM or 500μM of AA. There was no change in the cell viability between treated and untreated mice **(Figure 76).** DNA extraction was attempted from n=4 samples, with insufficient DNA retrieved for DNA 5hmC analysis via dot blot.

A summary of the effects of ascorbic acid on *ex vivo* $Men1^{L/L}$*/RIP2 Cre*-ER mouse insulinomas, and *in vitro* QGP1 and BON1 cells lines, and *in vivo* octreotide treatment on $Men1^{L/L}$*/RIP2 Cre*-ER mice on DNA 5hmC is shown in **Table 42**.

Table 42: DNA 5hmC in *Men1*$^{L/L}$*/RIP2 Cre*-ER mice and QGP1 and BON1 cells

	***Ex vivo* mouse islets**	**BON1**	**QGP1**
Ascorbic acid treatment			
Cell viability with 100μM AA	↔	↔	↓
Cell viability with 500μM AA	↔	↓	↓
DNA 5hmC	NA[a]	↔	↑
Octreotide treatment	***In vivo* mouse islets**	**BON1**	**QGP1**
DNA 5hmC (IHC)	↓	NA	NA

[a]DNA could not be extracted in adequate amounts to undertaken DNA 5hmC dot blots from *ex vivo* islets

In summary, QGP1 cells (derived from a human somatostatinoma) are used as a NF-PNET cell line, whereas BON1 cell lines, are for insulinomas. In previous Chapters 2 and 3, it was shown that NF-PNETs have less DNA 5hmC than insulinomas, and it may be that BON1 cell lines retain their ability to maintain a DNA 5hmC mark, whereas QGP1 cells have lost this ability. Additionally, DNA 5hmC may re-express important genes at the lower AA concentration of 100μM, as opposed to the possible cytotoxic effect seen at higher AA doses due to the generation of hydrogen peroxide.

4.4.3 Discussion

Mice treated with high dose octreotide had ~3 fold decrease in DNA 5hmC when compared to vehicle. This suggests that treatment with SSAs may affect the DNA hydroxymethylome, and this needs further study in human PNET cell lines and PNETs. However, BON1 cell lines express low to no levels of SSTR2, the main SSTR by which SSAs (octreotide and lanreotide) exert their anti-proliferative and anti-secretory effect. Therefore, using unaltered BON1 cell lines is not a good model to assess DNA hydroxymethylome changes that may be seen with SSAs. There are 12 neuroendocrine neoplasm (NEN) cell lines reported to study PNETs *in vitro.* [285] This includes 8 cell lines derived from human NF-PNETs and four insulinomas (2 derived from a mouse, one from a rodent and one from a canine). 58% (7 /12) expression SSTR2 and include, SPNE1 (derived from a 44-year-old female with a NF-PNET (Ki67 30%)[286], NT3 (derived from a male with a well-differentiated PNET (Ki67 15-20%),[287] NT-18P (primary PNET), NT-18LM (liver metastasis), NT-36 (PNET recurrence) all derived from a 57-year-old male patient with a G3 NF-PNET),[288] CanINS (derived from a canine insulinoma)[289] and INS-1 (derived from a rat β cell).[290] One of the major barriers identified at the recent NET Models Meeting 2024 was access to appropriate cell lines, as there is no obligation for researchers to deposit cell lines into a publicly available repository (e.g. American Type Culture Collection (ATCC)), the slow growing nature of PNETs and risk of contamination during culture and limited access to surgical PNET specimens.[285] An alternative pathway, is to transfect SSTR2 into BON1 or QGP1 cell lines. Therefore, the further assessment of the effect of SSAs on the DNA hydroxymethyome in PNET cell lines will require access or generation of the appropriate *in vitro* model.

Next, the effect of AA was assessed on the DNA hydroxymethylome in BON1 and QGP1 cell lines. Cells were both treated with AA for 7 days, and a viability assay was undertaken. AA had a much greater effect on QGP1 cells when compared to BON1. The effect of 500μM of AA on cells may be due to the cytotoxic effect of AA as with higher doses, AA works by inducing hydrogen peroxide, leading to cell death, as opposed to lower doses, which may work predominantly by altering the DNA methylome and as cofactors for other enzymes (for example TET enzymes) and epigenetic regulators (e.g. histones demethylases). Additionally, QGP1 cells are used as an *in vitro* NF-PNET model and in previous chapters 2 and 3, I have shown that NF-PNETs have lower DNA 5hmC than normal islets and insulinomas. Therefore, using AA to increase the DNA 5hmC in NF-PNET, thereby restoring this DNA mark, may improve cancer phenotype as shown by a decrease in cell viability at day 7, which is more pronounced in QGP1 cells than in BON1 (i.e. insulinoma) cell line. Another explanation may be that as QGP1 cells proliferate more slowly, the effect of AA on DNA 5hmC may be present for a longer period, whereas, given the more rapid proliferation with BON1 cells, the effect of AA may disappear. However, this has possible therapeutic implications for the treatment (or adjunctive treatment) for NF-PNETs and suggests that in patients with NF-PNETs, increasing their oral AA intake may help improve their DNA hydroxymethylome and possibly slow tumour growth. Although not assessed in this thesis, patients with MEN1-associated PNETs have increased tumour DNA methylation when compared to non-MEN1 patients (with menin present). AA works as a demethylating agent, and it is hypothesised that this may help tumour phenotypes in patients with MEN1. Further studies with the appropriate cell models will help to answer this important question.

AA effect was also reviewed in an *ex vivo* study with islets retrieved from a MEN1 mouse model. Cell viability of *ex vivo* mouse insulinomas was assessed at days 3 and 5 after AA treatment, with no difference in cell viability identified. There were a few limitations for the *ex vivo* mouse islet study, including 1) mice still express the gulonolactone (L-) oxidase (*Gulo)*, which is obsolete in humans, and they are able to generate endogenous AA, therefore, mice are not dependent on exogenous vitamin C to help maintain their hydroxymethylome 2) there was variability due to the number of islets in each 96 well and further replicates would be needed and 3) Insufficient DNA was extracted from the available islets to assess DNA 5hmC.

5. Conclusions

In summary, I have generated a new model to classify *MEN1* missense variants by combining three separate variant effect predictors (REVEL, SPDV and ESCOTT). I have generated cut-off levels for the new *MEN1* model and have annotated a look-up table of >4000 *MEN1* missense variants. This will provide important information to clinicians looking after patients with MEN1 and asymptomatic patients who present with *MEN1* missense variants and may help guide shared decision-making regarding the need for MEN1 tumour surveillance screening. As *MEN1* is involved in DNA methylome regulation, further investigation into DNA modifications in PNETs was undertaken, and focused on the less commonly studied epigenetic mark, DNA 5hmC, which has an opposite function to DNA methylation. I have shown that DNA 5hmC is enriched in normal islet cells and is lost in PNETs, with lower levels of DNA 5hmC seen in NF-PNETs compared to insulinomas. DNA 5hmC may be a marker of cell differentiation and suggests that NF-PNETs (most having lost the ability to secrete clinically relevant peptides), may be less differentiated when compared to insulinomas and normal islet cells. Given this difference in DNA 5hmC seen between PNET subtypes, AA was used to modulate the DNA hydroxymethylome. AA had a greater effect on increasing DNA 5hmC in QGP1 cells (*in vitro* model for NF-PNETs) than in BON1 (*in vitro* model for insulinoma) cells and AA had a greater effect on cell viability in QPG1 cells after 7 days of treatment. This identifies important potential treatment options for patients with NF-PNETs, as supplementing their diet (or advice to increase foods containing vitamin C) may help to improve cancer phenotype and/or slow tumour progression. Whether vitamin C is also helpful to slow tumour growth in patients with MEN1 is unclear, however, given that loss of MEN1 leads to increased tumour DNA methylation, it is possible that vitamin C treatment may also benefit this group of patients.

6. Future studies

6.1 *MEN1* missense variants

The new *MEN1* missense variant model generated was based upon *MEN1* variants that were classified as likely pathogenic or pathogenic by ACMG criteria. The model excluded *MEN1* VUSs which have been reported in the literature to possibly cause MEN1 and that had been classified as VUS based on ACMG criteria. Therefore, there is still a grey area surrounding the pathogenicity of some of these variants and validation of the model against these VUS should be undertaken. Additionally, further model generation and/or optimisation can be undertaken with ongoing collaboration with Dr Lukas Gerasimavicius and Professor Joe Marsh, from the MRC Human Genetics Unit in Edinburgh, given their expertise in this area. The current model has been validated against a cohort of patients from Geisinger. However, one of the limitations was that there were no likely pathogenic or pathogenic *MEN1* variants in this cohort, which hadn't been used to generate the original model. Therefore, novel pathogenic variants in other cohorts (e.g. *MEN1* variants reported in the UMD-MEN1 database) could be used for further model validation. An important consideration is also model validation against other ethnic backgrounds, as most *MEN1* variants used for the model generation and validation were from people of white background. Therefore, this model should be further validated with patients of other ethnic backgrounds (e.g. Chinese, South Asian or African) to assess model accuracy, as genetic modifiers may also affect the model's performance. Additionally, this model can be further refined as novel pathogenic *MEN1* variants are reported.

6.2 DNA hydroxymethylation in normal islets and PNETs

6.2.1 DNA hydroxymethylation in normal islet biology

6.2.1.1 DNA 5hmC in each islet subtype

This thesis has shown that normal islet cells are heavily hydroxymethylated. One aspect that this thesis does not address is what is the function of this DNA mark that is so important for normal islet biology. I would like to further assess whether DNA 5hmC is islet subtype specific or is a general feature of all islet cells. This could be further investigated by Fluorescence-activated cell sorting (FACs) sorting islets cells and then analysing islet subtype DNA 5hmC by methylationEPIC array. This would require a relatively high number of normal islets, to ensure that enough islet subtype specific DNA was obtained to undergo both bisulfite and oxidative-bisulfite conversion. Additionally, it would be useful to determine the expression data (mRNA) and genetic landscape (DNA SNVs) for each normal pancreatic islet subtype and correlating these genetic and epigenetic changes.

6.2.1.2 DNA 5hmC and insulin secretion

Results from Chapter 3 report that in normal pancreatic tissue Insulin signalling is a significantly hydromethylated pathway. Therefore, could DNA 5hmC play a role in insulin secretion from the β cell, and or have any relevance to the pathogenesis of type 2 diabetes (T2DM). There are a few cell lines that have been used to model insulin secretion in normal β cells including endo β cells (EndoC-βH1 and EndoC-βH2).[291,292] These cell lines could be treated with and without AA, and grown in different concentrations of glucose-containing serum. Changes in DNA 5hmC could then be assessed by DNA dot blot, and insulin assays undertaken to assess for any

changes in insulin secretion. This could provide pilot data for further research projects looking at DNA 5hmC in normal islet biology and in conditions mimicking islets pathology.

6.2.2 DNA hydroxymethylation in PNETs

6.2.2.1 Correlating genetic and epigenetic changes in PNETs

In this thesis, I have investigated expression data (mRNA) and epigenetic data (DNA methylation and hydroxymethylation) of insulinomas and PNETs. Going forward it would be useful to correlate these data with genetic changes (DNA mutations) for each PNET subtype. This could then be compared to that of normal pancreatic islet subtypes to get a broad understanding of the aberrant genetic and epigenetic changes seen in PNETs and ultimately use this data to explore other treatment options.

6.2.2.2 AA treatment on gene expression in PNET cell lines

In Chapter 4, I have shown that AA increased DNA 5hmC which was more pronounced and sustained in QGP1 cells than BON1. AA used at lower doses (100μM) also lead to a decrease in cell viability. The next steps are to extract RNA from treated and untreated QGP1 cells and to analyse gene expression changes by RNA-seq. This will provide further information regarding what AA is doing at the transcription level. The utility of repeated doses of AA on cell viability and DNA 5hmC could also be assessed, to give an idea about how long the effect of AA lasts in QGP1 cells. DNA 5hmC changes seen with AA treatment could also be assessed using the methylationEPIC array in different cell lines.

6.2.2.3 Assessing DNA 5hmC in a PNET mouse model treated with AA

One of the current limitations of using our mouse model of MEN1 (*Men1*$^{L/L}$*/RIP2 Cre*-ER mice), is that this mouse expresses the enzyme Gulo, and these mice are endogenously able to generate vitamin C, whereas humans are reliant on dietary supplements of vitamin C. Therefore, a mouse model of MEN1 with an additional *Gulo* knock-out could be used to more accurately investigate the effect of vitamin C on DNA 5hmC in MEN1-associated PNETs. Studies could include supplementing different concentrations of AA to mouse drinking water, and then 1) assessing the proliferation rates of mouse islet tumours (using bromodeoxyuridine (BrDU)) and 2) analysing the level of DNA 5hmC in mouse insulinomas by IHC. Another limitation by using AA in a mouse model of MEN1, is that mice usually develop insulinomas, which in humans have higher levels of baseline DNA 5hmC. Generation of a mouse model with a NF-PNET (e.g. *MEN1* α cell knockout mouse), may be more clinically relevant.

6.2.2.4 Assessment of somatostatin analogues as DNA 5hmC modulators

In Chapter 4, I have shown that *Men1*$^{L/L}$*/RIP2 Cre*-ER mice treated with high doses of octreotide have lower DNA 5hmC when compared to untreated mice. One possible mechanism for this finding is that octreotide may decrease intracellular cAMP, and previous studies have shown that lowering intracellular cAMP leads to an increase in DNA 5hmC.[91] Using a mouse model for MEN1 may not be directly comparable as mice express different SSTR receptors on islets cells. Mice have been shown to express SSTR5 in β cells predominantly and SSTR2 in α cells[293], whereas the reverse is seen in humans with β cells predominantly expressing SSTR2 and α cells, SSTR5. SSA can be assessed in PNET cell models, however, the most commonly available cell lines BON1 and QGP1 do not express SSTR2. Therefore, either obtaining other

PNET cell lines that express SSTR2 or transfecting in SSTR2 to BON1 or QGP1 cells is needed. Preliminary work has shown that BON1 (but not QGP1) cells could transiently express SSTR2 using the pCMV6-Entry Mammalian Vector with C-terminal Myc-DDY Tag **(Figure 77)**.

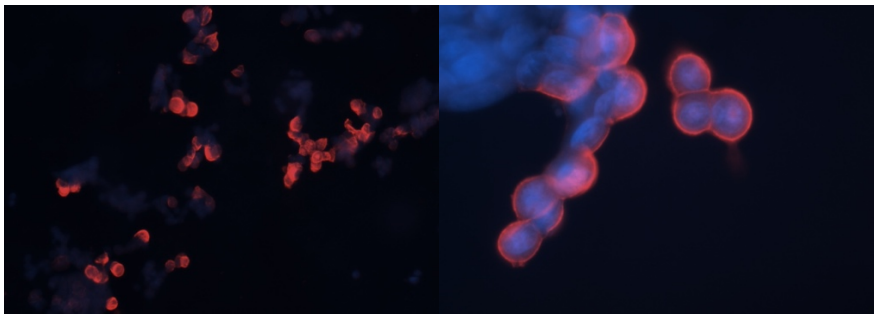


Figure 77: Immunocytochemistry for the SSTR2 flag antibody showing expression of SSTR2 receptor at the cell membrane. Blue represents the nuclei and red represents the SSTR2 receptor.

SSAs have both anti-secretory and anti-proliferative effects. Teasing out which intracellular pathways are responsible for these two effects would be useful, as this could to determine if additional treatment with AA would be beneficial or counterproductive. One of the side effects of SSAs is pancreatic exocrine insufficiency (PEI) and an increased risk of T2DM. Further studies looking at whether AA could help alleviate the symptoms of PEI or decrease the risk of T2DM without detrimentally affecting its anti-proliferative action is needed.

7. References

1. Innes JT, Carey LC. Normal pancreatic dimensions in the adult human. Am J Surg. 1994;167(2):261–3.

2. Pandol S. The Exocrine Pancreas. San Rafael (CA): Morgan & Claypool Life Sciences; 2010.

3. Pandiri AR. Overview of exocrine pancreatic pathobiology. Toxicol Pathol. 2014;42(1):207–16.

4. Jolles S. Paul Langerhans. J Clin Pathol. 2002;55(4):243.

5. Jörgens V. Paul Langerhans: The Man Who Discovered the Islets. In: Jörgens V, Porta M, Joergens V, Porta M, editors. Unveiling Diabetes - Historical Milestones in Diabetology. 29: S.Karger AG; 2020. p. 0.

6. Ionescu-Tirgoviste C, Gagniuc PA, Gubceac E, et al. A 3D map of the islet routes throughout the healthy human pancreas. Sci Rep. 2015;5:14634.

7. Lawlor N, George J, Bolisetty M, et al. Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. Genome Res. 2017;27(2):208–22.

8. Brissova M, Fowler MJ, Nicholson WE, et al. Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy. J Histochem Cytochem. 2005;53(9):1087–97.

9. Cabrera O, Berman DM, Kenyon NS, Ricordi C, Berggren PO, Caicedo A. The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. Proc Natl Acad Sci U S A. 2006;103(7):2334–9.

10. Ichii H, Inverardi L, Pileggi A, et al. A Novel Method for the Assessment of Cellular Composition and Beta-Cell Viability in Human Islet Preparations. American Journal of Transplantation. 2005;5(7):1635–45.

11. Jensen R, Ito T. Gastrinoma. South Dartmouth (MA): MDText.com; 2023. Available from: https://www.ncbi.nlm.nih.gov/books/NBK279075/.

12. Park W, Chawla A, O'Reilly EM. Pancreatic Cancer: A Review. JAMA. 2021;326(9):851–62.

13. Sonbol MB, Mazza GL, Mi L, et al. Survival and Incidence Patterns of Pancreatic Neuroendocrine Tumors Over the Last 2 Decades: A SEER Database Analysis. The Oncologist. 2022.

14. Ehehalt F, Saeger HD, Schmidt CM, Grützmann R. Neuroendocrine tumors of the pancreas. Oncologist. 2009;14(5):456–67.

15. Yao JC, Hassan M, Phan A, et al. One hundred years after "carcinoid": epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. J Clin Oncol. 2008;26(18):3063–72.

16. Liu X, Chen B, Chen J, Su Z, Sun S. The incidence, prevalence, and survival analysis of pancreatic neuroendocrine tumors in the United States. Journal of Endocrinological Investigation. 2023;46(7):1373–84.

17. Dasari A, Shen C, Halperin D, et al. Trends in the Incidence, Prevalence, and Survival Outcomes in Patients With Neuroendocrine Tumors in the United States. JAMA Oncol. 2017;3(10):1335–42.

18. Ryan DP, Hong TS, Bardeesy N. Pancreatic adenocarcinoma. N Engl J Med. 2014;371(22):2140–1.

19. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA Cancer J Clin. 2021;71(1):7–33.

20. Romanet P, Mohamed A, Giraud S, et al. UMD-MEN1 Database: An Overview of the 370 MEN1 Variants Present in 1676 Patients From the French Population. J Clin Endocrinol Metab. 2019;104(3):753–64.

21. de Laat JM, van der Luijt RB, Pieterman CR, et al. MEN1 redefined, a clinical comparison of mutation-positive and mutation-negative patients. BMC Med. 2016;14(1):182.

22. Crona J, Skogseid B. GEP- NETS UPDATE: Genetics of neuroendocrine tumors. Eur J Endocrinol. 2016;174(6):R275–90.

23. Capurso G, Falconi M, Panzuto F, et al. Risk factors for sporadic pancreatic endocrine tumors: a case-control study of prospectively evaluated patients. Am J Gastroenterol. 2009;104(12):3034–41.

24. Ben Q, Zhong J, Fei J, et al. Risk Factors for Sporadic Pancreatic Neuroendocrine Tumors: A Case-Control Study. Sci Rep. 2016;6:36073.

25. Haugvik SP, Hedenstrom P, Korsaeth E, et al. Diabetes, smoking, alcohol use, and family history of cancer as risk factors for pancreatic neuroendocrine tumors: a systematic review and meta-analysis. Neuroendocrinology. 2015;101(2):133–42.

26. Permert J, Ihse I, Jorfeldt L, von Schenck H, Arnqvist HJ, Larsson J. Pancreatic cancer is associated with impaired glucose metabolism. The European journal of surgery = Acta chirurgica. 1993;159(2):101–7.

27. Song S, Wang B, Zhang X, et al. Long-Term Diabetes Mellitus Is Associated with an Increased Risk of Pancreatic Cancer: A Meta-Analysis. PLOS ONE. 2015;10(7):e0134321.

28. Gallo M, Ruggeri RM, Muscogiuri G, Pizza G, Faggiano A, Colao A. Diabetes and pancreatic neuroendocrine tumours: Which interplays, if any? Cancer Treatment Reviews. 2018;67:1–9.

29. Lavezzi E, Brunetti A, Smiroldo V, et al. Case Report: New CDKN1B Mutation in Multiple Endocrine Neoplasia Type 4 and Brief Literature Review on Clinical Management. Front Endocrinol (Lausanne). 2022;13:773143.

30. Occhi G, Regazzo D, Trivellin G, et al. A Novel Mutation in the Upstream Open Reading Frame of the CDKN1B Gene Causes a MEN4 Phenotype. PLOS Genetics. 2013;9(3):e1003350.

31. Seabrook A, Wijewardene A, De Sousa S, et al. MEN4, the MEN1 Mimicker: A Case Series of three Phenotypically Heterogenous Patients With Unique CDKN1B Mutations. The Journal of clinical endocrinology and metabolism. 2022;107(8):2339–49.

32. Lines KE, Nachtigall LB, Dichtel LE, et al. Multiple Endocrine Neoplasia Type 1 (MEN1) Phenocopy Due to a Cell Cycle Division 73 (CDC73) Variant. Journal of the Endocrine Society. 2020;4(11).

33. Petignot S, Daly AF, Castermans E, et al. Pancreatic Neuroendocrine Neoplasm Associated with a Familial MAX Deletion. Horm Metab Res. 2020;52(11):784–7.

34. Tirosh A, Sadowski SM, Linehan WM, et al. Association of VHL Genotype With Pancreatic Neuroendocrine Tumor Phenotype in Patients With von Hippel–Lindau Disease. JAMA Oncology. 2018;4(1):124–6.

35. Keutgen XM, Hammel P, Choyke PL, Libutti SK, Jonasch E, Kebebew E. Evaluation and management of pancreatic lesions in patients with von Hippel-Lindau disease. Nat Rev Clin Oncol. 2016;13(9):537–49.

36. Libutti SK, Choyke PL, Alexander HR, et al. Clinical and genetic analysis of patients with pancreatic neuroendocrine tumors associated with von Hippel-Lindau disease. Surgery. 2000;128(6):1022–7;discussion 7–8.

37. Weisbrod AB, Kitano M, Thomas F, et al. Assessment of tumor growth in pancreatic neuroendocrine tumors in von Hippel Lindau syndrome. J Am Coll Surg. 2014;218(2):163–9.

38. Landry JP, Schertz KL, Chiang YJ, et al. Comparison of Cancer Prevalence in Patients With Neurofibromatosis Type 1 at an Academic Cancer Center vs in the General Population From 1985 to 2020. JAMA Netw Open. 2021;4(3):e210945.

39. Geurts JL. Inherited syndromes involving pancreatic neuroendocrine tumors. J Gastrointest Oncol. 2020;11(3):559–66.

40. Mowrey K, Northrup H, Rougeau P, et al. Frequency, Progression, and Current Management: Report of 16 New Cases of Nonfunctional Pancreatic Neuroendocrine Tumors in

Tuberous Sclerosis Complex and Comparison With Previous Reports. Frontiers in Neurology. 2021;12.

41. Scarpa A, Chang DK, Nones K, et al. Whole-genome landscape of pancreatic neuroendocrine tumours. Nature. 2017;543(7643):65–71.

42. Ahmad S, Naber MR, Giles RH, Valk GD, van Leeuwaarde RS. Diagnostic and management strategies for pNETs in Von Hippel-Lindau: a systematic review. Endocr Relat Cancer. 2021;28(3):151–60.

43. Evans LM, Geenen KR, O'Shea A, Hedgire SS, Ferrone CR, Thiele EA. Tuberous sclerosis complex-associated nonfunctional pancreatic neuroendocrine tumors: Management and surgical outcomes. Am J Med Genet A. 2022;188(9):2666–71.

44. Thakker RV, Newey PJ, Walls GV, et al. Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN1). The Journal of clinical endocrinology and metabolism. 2012;97(9):2990–3011.

45. Chandrasekharappa SC, Guru SC, Manickam P, et al. Positional cloning of the gene for multiple endocrine neoplasia-type 1. Science. 1997;276(5311):404–7.

46. Larsson C, Skogseid B, Oberg K, Nakamura Y, Nordenskjöld M. Multiple endocrine neoplasia type 1 gene maps to chromosome 11 and is lost in insulinoma. Nature. 1988;332(6159):85–7.

47. Huang J, Gurung B, Wan B, et al. The same pocket in menin binds both MLL and JUND but has opposite effects on transcription. Nature. 2012;482(7386):542–6.

48. Stevenson M, Lines KE, Thakker RV. Molecular Genetic Studies of Pancreatic Neuroendocrine Tumors: New Therapeutic Approaches. Endocrinol Metab Clin North Am. 2018;47(3):525–48.

49. Gallo A, Cuozzo C, Esposito I, et al. Menin uncouples Elk-1, JunD and c-Jun phosphorylation from MAP kinase activation. Oncogene. 2002;21(42):6434–45.

50. Agarwal SK, Kester MB, Debelenko LV, et al. Germline mutations of the MEN1 gene in familial multiple endocrine neoplasia type 1 and related states. Hum Mol Genet. 1997;6(7):1169–75.

51. Yokoyama A, Somervaille TCP, Smith KS, Rozenblatt-Rosen O, Meyerson M, Cleary ML. The Menin Tumor Suppressor Protein Is an Essential Oncogenic Cofactor for MLL-Associated Leukemogenesis. Cell. 2005;123(2):207–18.

52. Taguchi R, Yamada M, Horiguchi K, et al. Haploinsufficient and predominant expression of multiple endocrine neoplasia type 1 (MEN1)-related genes, MLL, p27Kip1 and p18Ink4C in endocrine organs. Biochem Biophys Res Commun. 2011;415(2):378–83.

53. Thakker RV. Multiple endocrine neoplasia type 1 (MEN1) and type 4 (MEN4). Mol Cell Endocrinol. 2014;386(1-2):2–15.

54. Chamberlain CE, Scheel DW, McGlynn K, et al. Menin determines K-RAS proliferative outputs in endocrine cells. J Clin Invest. 2014;124(9):4093–101.

55. Jiao Y, Shi C, Edil BH, et al. DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. Science. 2011;331(6021):1199–203.

56. Hendy GN, Kaji H, Sowa H, Lebrun JJ, Canaff L. Menin and TGF-beta superfamily member signaling via the Smad pathway in pituitary, parathyroid and osteoblast. Horm Metab Res. 2005;37(6):375–9.

57. Kaji H, Canaff L, Lebrun JJ, Goltzman D, Hendy GN. Inactivation of menin, a Smad3-interacting protein, blocks transforming growth factor type beta signaling. Proc Natl Acad Sci U S A. 2001;98(7):3837–42.

58. Knudson AG. Mutation and Cancer: Statistical Study of Retinoblastoma. Proceedings of the National Academy of Sciences. 1971;68(4):820–3.

59. Ryland GL, Doyle MA, Goode D, et al. Loss of heterozygosity: what is it good for? BMC Medical Genomics. 2015;8(1):45.

60. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW. Cancer genome landscapes. Science. 2013;339(6127):1546–58.

61. Schwartz CJ, Khorsandi N, Blanco A, Mukhtar RA, Chen YY, Krings G. Clinicopathologic and genetic analysis of invasive breast carcinomas in women with germline CHEK2 variants. Breast Cancer Res Treat. 2024;204(1):171–9.

62. Iyevleva AG, Aleksakhina SN, Sokolenko AP, et al. Somatic loss of the remaining allele occurs approximately in half of CHEK2-driven breast cancers and is accompanied by a border-line increase of chromosomal instability. Breast Cancer Research and Treatment. 2022;192(2):283–91.

63. Lim BWX, Li N, Mahale S, et al. Somatic inactivation of breast cancer predisposition genes in tumors associated with pathogenic germline variants. J Natl Cancer Inst. 2023;115(2):181–9.

64. Lejonklou MH, Barbu A, Stålberg P, Skogseid B. Accelerated proliferation and differential global gene expression in pancreatic islets of five-week-old heterozygous Men1 mice: Men1 is a haploinsufficient suppressor. Endocrinology. 2012;153(6):2588–98.

65. Perren A, Anlauf M, Henopp T, et al. Multiple Endocrine Neoplasia Type 1 (MEN1): Loss of One MEN1 Allele in Tumors and Monohormonal Endocrine Cell Clusters But Not in Islet Hyperplasia of the Pancreas. The Journal of clinical endocrinology and metabolism. 2007;92(3):1118–28.

66. Lemos MC, Thakker RV. Multiple endocrine neoplasia type 1 (MEN1): analysis of 1336 mutations reported in the first decade following identification of the gene. Hum Mutat. 2008;29(1):22–32.

67. Brandi ML, Pieterman CRC, English KA, et al. Multiple endocrine neoplasia type 1 (MEN1): recommendations and guidelines for best practice. The Lancet Diabetes & Endocrinology.

68. Thakker RV. Multiple endocrine neoplasia type 1 (MEN1) and type 4 (MEN4). Molecular and Cellular Endocrinology. 2014;386(1):2–15.

69. Brandi ML, Gagel RF, Angeli A, et al. Guidelines for diagnosis and therapy of MEN type 1 and type 2. J Clin Endocrinol Metab. 2001;86(12):5658–71.

70. Faggiano A, Ferolla P, Grimaldi F, et al. Natural history of gastro-entero-pancreatic and thoracic neuroendocrine tumors. Data from a large prospective and retrospective Italian epidemiological study: the NET management study. J Endocrinol Invest. 2012;35(9):817–23.

71. Antongiulio F, Valeria R, Luisa C, Annamaria C. Hereditary Neuroendocrine Tumor Syndromes. In: Monica F, editor. Hot Topics in Endocrine and Endocrine-Related Diseases. Rijeka: IntechOpen; 2013. p. Ch. 1.

72. Chan CS, Laddha SV, Lewis PW, et al. ATRX, DAXX or MEN1 mutant pancreatic neuroendocrine tumors are a distinct alpha-cell signature subgroup. Nat Commun. 2018;9(1):4158.

73. Yates CJ, Newey PJ, Thakker RV. Challenges and controversies in management of pancreatic neuroendocrine tumours in patients with MEN1. Lancet Diabetes Endocrinol. 2015;3(11):895–905.

74. Missiaglia E, Dalai I, Barbi S, et al. Pancreatic endocrine tumors: expression profiling evidences a role for AKT-mTOR pathway. J Clin Oncol. 2010;28(2):245–55.

75. Tirosh A, Mukherjee S, Lack J, et al. Distinct genome-wide methylation patterns in sporadic and hereditary nonfunctioning pancreatic neuroendocrine tumors. Cancer. 2019;125(8):1247–57.

76. Ohki R, Saito K, Chen Y, et al. PHLDA3 is a novel tumor suppressor of pancreatic neuroendocrine tumors. Proc Natl Acad Sci U S A. 2014;111(23):E2404–13.

77. Cao Y, Gao Z, Li L, et al. Whole exome sequencing of insulinoma reveals recurrent T372R mutations in YY1. Nature Communications. 2013;4(1):2810.

78. Hong X, Qiao S, Li F, et al. Whole-genome sequencing reveals distinct genetic bases for insulinomas and non-functional pancreatic neuroendocrine tumours: leading to a new classification system. Gut. 2020;69(5):877–87.

79. Seto E, Shi Y, Shenk T. YY1 is an initiator sequence-binding protein that directs and activates transcription in vitro. Nature. 1991;354(6350):241–5.

80. Rezai-Zadeh N, Zhang X, Namour F, et al. Targeted recruitment of a histone H4-specific methyltransferase by the transcription factor YY1. Genes & development. 2003;17(8):1019–29.

81. Annunziato A. DNA Packaging: Nucleosomes and Chromatin. Nature Education 2008;1(1):26.

82. Beacon TH, Delcuve GP, Lopez C, et al. The dynamic broad epigenetic (H3K4me3, H3K27ac) domain as a mark of essential genes. Clin Epigenetics. 2021;13(1):138.

83. Hyun K, Jeon J, Park K, Kim J. Writing, erasing and reading histone lysine methylations. Exp Mol Med. 2017;49(4):e324.

84. English KA, Thakker RV, Lines KE. The role of DNA methylation in human pancreatic neuroendocrine tumours. Endocr Oncol. 2023;3(1):e230003.

85. Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. Nature. 2013;502(7472):472–9.

86. Moore LD, Le T, Fan G. DNA Methylation and Its Basic Function. Neuropsychopharmacology. 2013;38(1):23–38.

87. Ren R, Horton JR, Zhang X, Blumenthal RM, Cheng X. Detecting and interpreting DNA methylation marks. Curr Opin Struct Biol. 2018;53:88–99.

88. Du Q, Luu PL, Stirzaker C, Clark SJ. Methyl-CpG-binding domain proteins: readers of the epigenome. Epigenomics. 2015;7(6):1051–73.

89. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature. 2010;466(7310):1129–33.

90. Tahiliani M, Koh KP, Shen Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009;324(5929):930–5.

91. Camarena V, Sant DW, Huff TC, et al. cAMP signaling regulates DNA hydroxymethylation by augmenting the intracellular labile ferrous iron pool. Elife. 2017;6.

92. Kinney S, Chin H, Vaisvila R, et al. Tissue-specific Distribution and Dynamic Changes of 5-Hydroxymethylcytosine in Mammalian Genomes. The Journal of biological chemistry. 2011;286:24685–93.

93. Yuan Z, Sánchez Claros C, Suzuki M, et al. Loss of MEN1 activates DNMT1 implicating DNA hypermethylation as a driver of MEN1 tumorigenesis. Oncotarget. 2016;7(11):12633–50.

94. Skvortsova K, Masle-Farquhar E, Luu P-L, et al. DNA Hypermethylation Encroachment at CpG Island Borders in Cancer Is Predisposed by H3K4 Monomethylation Patterns. Cancer Cell. 2019;35(2):297–314.e8.

95. Bird AP. CpG-rich islands and the function of DNA methylation. Nature. 1986;321(6067):209–13.

96. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. Nature Reviews Genetics. 2002;3(6):415–28.

97. Valinluck V, Tsai H-H, Rogstad DK, Burdzy A, Bird A, Sowers LC. Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). Nucleic Acids Research. 2004;32(14):4100–8.

98. Valinluck V, Sowers LC. Endogenous Cytosine Damage Products Alter the Site Selectivity of Human DNA Maintenance Methyltransferase DNMT1. Cancer Research. 2007;67(3):946–50.

99. Spruijt Cornelia G, Gnerlich F, Smits Arne H, et al. Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. Cell. 2013;152(5):1146–59.

100. Zhou T, Xiong J, Wang M, et al. Structural Basis for Hydroxymethylcytosine Recognition by the SRA Domain of UHRF2. Molecular Cell. 2014;54(5):879–86.

101. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. The Behaviour of 5-Hydroxymethylcytosine in Bisulfite Sequencing. PLOS ONE. 2010;5(1):e8888.

102. Li J, Wu X, Zhou Y, et al. Decoding the dynamic DNA methylation and hydroxymethylation landscapes in endodermal lineage intermediates during pancreatic differentiation of hESC. Nucleic Acids Res. 2018;46(6):2883–900.

103. Ehrlich M. DNA hypomethylation in cancer cells. Epigenomics. 2009;1(2):239–59.

104. Eden A, Gaudet F, Waghmare A, Jaenisch R. Chromosomal Instability and Tumors Promoted by DNA Hypomethylation. Science. 2003;300(5618):455–.

105. Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R. DNA hypomethylation leads to elevated mutation rates. Nature. 1998;395(6697):89–93.

106. Takai D, Yagi Y, Habib N, Sugimura T, Ushijima T. Hypomethylation of LINE1 retrotransposon in human hepatocellular carcinomas, but not in surrounding liver cirrhosis. Jpn J Clin Oncol. 2000;30(7):306–9.

107. Stricker I, Tzivras D, Nambiar S, et al. Site- and grade-specific diversity of LINE1 methylation pattern in gastroenteropancreatic neuroendocrine tumours. Anticancer Res. 2012;32(9):3699–706.

108. Choi I-S, Estecio MRH, Nagano Y, et al. Hypomethylation of LINE-1 and Alu in well-differentiated neuroendocrine tumors (pancreatic endocrine tumors and carcinoid tumors). Modern Pathology. 2007;20(7):802–10.

109. Stefanoli M, La Rosa S, Sahnane N, et al. Prognostic relevance of aberrant DNA methylation in g1 and g2 pancreatic neuroendocrine tumors. Neuroendocrinology. 2014;100(1):26–34.

110. Marinoni I, Wiederkeher A, Wiedmer T, et al. Hypo-methylation mediates chromosomal instability in pancreatic NET. Endocrine-Related Cancer. 2017;24(3):137–46.

111. Chalitchagorn K, Shuangshoti S, Hourpai N, et al. Distinctive pattern of LINE-1 methylation level in normal tissues and the association with carcinogenesis. Oncogene. 2004;23(54):8841–6.

112. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860–921.

113. Deininger P. Alu elements: know the SINEs. Genome Biology. 2011;12(12):236.

114. Dammann R, Schagdarsurengin U, Liu L, et al. Frequent RASSF1A promoter hypermethylation and K-ras mutations in pancreatic carcinoma. Oncogene. 2003;22(24):3806–12.

115. House MG, Herman JG, Guo MZ, et al. Aberrant hypermethylation of tumor suppressor genes in pancreatic endocrine neoplasms. Ann Surg. 2003;238(3):423–31; discussion 31–2.

116. Arnold CN, Sosnowski A, Schmitt-Graff A, Arnold R, Blum HE. Analysis of molecular pathways in sporadic neuroendocrine tumors of the gastro-entero-pancreatic system. Int J Cancer. 2007;120(10):2157–64.

117. Malpeli G, Amato E, Dandrea M, et al. Methylation-associated down-regulation of RASSF1A and up-regulation of RASSF1C in pancreatic endocrine tumors. BMC Cancer. 2011;11:351.

118. Muscarella P, Melvin WS, Fisher WE, et al. Genetic alterations in gastrinomas and nonfunctioning pancreatic neuroendocrine tumors: an analysis of p16/MTS1 tumor suppressor gene inactivation. Cancer Res. 1998;58(2):237–40.

119. Serrano J, Goebel SU, Peghini PL, Lubensky IA, Gibril F, Jensen RT. Alterations in the p16INK4a/CDKN2A tumor suppressor gene in gastrinomas. J Clin Endocrinol Metab. 2000;85(11):4146–56.

120. Bartsch DK, Kersting M, Wild A, et al. Low Frequency of p16^{INK4a} Alterations in Insulinomas. Digestion. 2000;62(2-3):171–7.

121. Campana D, Walter T, Pusceddu S, et al. Correlation between MGMT promoter methylation and response to temozolomide-based therapy in neuroendocrine neoplasms: an observational retrospective multicenter study. Endocrine. 2018;60(3):490–8.

122. Chan AO-O, Kim SG, Bedeir A, Issa J-P, Hamilton SR, Rashid A. CpG island methylation in carcinoid and pancreatic endocrine tumors. Oncogene. 2003;22(6):924–34.

123. Liu L, Broaddus RR, Yao JC, et al. Epigenetic alterations in neuroendocrine tumors: methylation of RAS-association domain family 1, isoform A and p16 genes are associated with metastasis. Mod Pathol. 2005;18(12):1632–40.

124. Schmitt AM, Pavel M, Rudolph T, et al. Prognostic and predictive roles of MGMT protein expression and promoter methylation in sporadic pancreatic neuroendocrine neoplasms. Neuroendocrinology. 2014;100(1):35–44.

125. Ushiku H, Yamashita K, Kawamata H, et al. Homeobox-Only Protein Expression Is a Critical Prognostic Indicator of Pancreatic Neuroendocrine Tumor and Is Regulated by Promoter DNA Hypermethylation. Pancreas. 2016;45(9):1255–62.

126. Wild A, Ramaswamy A, Langer P, et al. Frequent methylation-associated silencing of the tissue inhibitor of metalloproteinase-3 gene in pancreatic endocrine tumors. J Clin Endocrinol Metab. 2003;88(3):1367–73.

127. Ban X, Mo S, Lu Z, et al. Expression and methylation status of MMR and MGMT in well-differentiated pancreatic neuroendocrine tumors and potential clinical applications. Endocrine. 2022;77(3):538–45.

128. Conemans EB, Lodewijk L, Moelans CB, et al. DNA methylation profiling in MEN1-related pancreatic neuroendocrine tumors reveals a potential epigenetic target for treatment. Eur J Endocrinol. 2018;179(3):153–60.

129. Liu B, Tang LH, Liu Z, et al. alpha-Internexin: a novel biomarker for pancreatic neuroendocrine tumor aggressiveness. J Clin Endocrinol Metab. 2014;99(5):E786–95.

130. Cros J, Hentic O, Rebours V, et al. MGMT expression predicts response to temozolomide in pancreatic neuroendocrine tumors. Endocr Relat Cancer. 2016;23(8):625–33.

131. Dejeux E, Olaso R, Dousset B, et al. Hypermethylation of the IGF2 differentially methylated region 2 is a specific event in insulinomas leading to loss-of-imprinting and overexpression. Endocrine-Related Cancer. 2009;16(3):939–52.

132. Zhang C, Jin J, Xie J, et al. The Clinical Features and Molecular Mechanisms of ACTH-secreting Pancreatic Neuroendocrine Tumors. J Clin Endocrinol Metab. 2020;105(11).

133. Evans JS, Beaumont J, Braga M, et al. Epigenetic potentiation of somatostatin-2 by guadecitabine in neuroendocrine neoplasias as a novel method to allow delivery of peptide receptor radiotherapy. Eur J Cancer. 2022;176:110–20.

134. Boons G, Vandamme T, Ibrahim J, et al. PDX1 DNA Methylation Distinguishes Two Subtypes of Pancreatic Neuroendocrine Neoplasms with a Different Prognosis. Cancers. 2020;12(6):1461.

135. Di Domenico A, Pipinikas CP, Maire RS, et al. Epigenetic landscape of pancreatic neuroendocrine tumours reveals distinct cells of origin and means of tumour progression. Communications Biology. 2020;3(1):740.

136. Lakis V, Lawlor RT, Newell F, et al. DNA methylation patterns identify subgroups of pancreatic neuroendocrine tumors with clinical association. Communications Biology. 2021;4(1):155.

137. Simon T, Riemer P, Jarosch A, et al. DNA methylation reveals distinct cells of origin for pancreatic neuroendocrine carcinomas and pancreatic neuroendocrine tumors. Genome Medicine. 2022;14(1):24.

138. Yachida S, Totoki Y, Noë M, et al. Comprehensive Genomic Profiling of Neuroendocrine Carcinomas of the Gastrointestinal System. Cancer Discovery. 2022;12(3):692–711.

139. Dubois F, Bergot E, Zalcman G, Levallet G. RASSF1A, puppeteer of cellular homeostasis, fights tumorigenesis, and metastasis-an updated review. Cell Death Dis. 2019;10(12):928.

140. Song MS, Song SJ, Kim SY, Oh HJ, Lim DS. The tumour suppressor RASSF1A promotes MDM2 self-ubiquitination by disrupting the MDM2-DAXX-HAUSP complex. EMBO J. 2008;27(13):1863–74.

141. Wang J, Hong Y, Shao S, Zhang K, Hong W. FFAR1-and FFAR4-dependent activation of Hippo pathway mediates DHA-induced apoptosis of androgen-independent prostate cancer cells. Biochemical and Biophysical Research Communications. 2018;506(3):590–6.

142. Kranzhöfer DK, Gilsbach R, Grüning BA, Backofen R, Nührenberg TG, Hein L. 5'-Hydroxymethylcytosine precedes loss of CpG methylation in enhancers and genes undergoing activation in cardiomyocyte maturation. PloS one. 2016;11(11):e0166575.

143. Cui X-L, Nie J, Ku J, et al. A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. Nature communications. 2020;11(1):6161.

144. Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome biology. 2011;12:1–8.

145. Pastor WA, Pape UJ, Huang Y, et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature. 2011;473(7347):394–7.

146. Wu H, D'Alessio AC, Ito S, et al. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes & development. 2011;25(7):679–84.

147. Uribe-Lewis S, Stark R, Carroll T, et al. 5-hydroxymethylcytosine marks promoters in colon that resist DNA hypermethylation in cancer. Genome biology. 2015;16:1–15.

148. Dong Z-R, Zhang C, Cai J-b, et al. Role of 5-hydroxymethylcytosine level in diagnosis and prognosis prediction of intrahepatic cholangiocarcinoma. Tumor Biology. 2015;36(4):2763–71.

149. Kuang L, Zhang J, Li Y, Wang Q, Liu J, Zhang B. Association of tet methylcytosine dioxygenase 2 and 5-hydroxymethylcytosine in endometrioid adenocarcinoma and its clinical significance. BMC Women's Health. 2024;24(1):188.

150. Fu Y-L, Wu Y-H, Cao D-H, et al. Increased 5-hydroxymethylcytosine is a favorable prognostic factor of Helicobacter pylori-negative gastric cancer patients. World Journal of Gastrointestinal Oncology. 2022;14(7):1295.

151. Sharma AE, Olivas A, Parilla M, et al. Epigenetic Dysregulation of 5-hydroxymethylcytosine in Well-Differentiated Pancreatic Neuroendocrine Tumors. Appl Immunohistochem Mol Morphol. 2022;30(2):e11–e5.

152. Alberge JB, Magrangeas F, Wagner M, et al. DNA hydroxymethylation is associated with disease severity and persists at enhancers of oncogenic regions in multiple myeloma. Clin Epigenetics. 2020;12(1):163.

153. Sjöström M, Zhao SG, Levy S, et al. The 5-Hydroxymethylcytosine Landscape of Prostate Cancer. Cancer Research. 2022;82(21):3888–902.

154. Bhattacharyya S, Pradhan K, Campbell N, et al. Altered hydroxymethylation is seen at regulatory regions in pancreatic cancer and regulates oncogenic pathways. Genome Res. 2017;27(11):1830–42.

155. Song CX, Yin S, Ma L, et al. 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. Cell Res. 2017;27(10):1231–42.

156. Mandel P, Metais P. [Nuclear Acids In Human Blood Plasma]. C R Seances Soc Biol Fil. 1948;142(3-4):241–3.

157. Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. Nat Rev Cancer. 2011;11(6):426–37.

158. Thierry AR, El Messaoudi S, Gahan PB, Anker P, Stroun M. Origins, structures, and functions of circulating DNA in oncology. Cancer Metastasis Rev. 2016;35(3):347–76.

159. Grande E, Capdevila J, Castellano D, et al. Pazopanib in pretreated advanced neuroendocrine tumors: a phase II, open-label trial of the Spanish Task Force Group for Neuroendocrine Tumors (GETNE). Ann Oncol. 2015;26(9):1987–93.

160. Oberg K, Krenning E, Sundin A, et al. A Delphic consensus assessment: imaging and biomarkers in gastroenteropancreatic neuroendocrine tumor disease management. Endocr Connect. 2016;5(5):174–87.

161. Modlin IM, Kidd M, Bodei L, Drozdov I, Aslanian H. The Clinical Utility of a Novel Blood-Based Multi-Transcriptome Assay for the Diagnosis of Neuroendocrine Tumors of the Gastrointestinal Tract. Official journal of the American College of Gastroenterology | ACG. 2015;110(8):1223–32.

162. Liu H, Tang T, Zhang H, et al. Identification of a 5-Hydroxymethylation Signature in Circulating Cell-Free DNA for the Noninvasive Detection of Colorectal Cancer. J Oncol. 2022;2022:3798741.

163. Walter T, van Brakel B, Vercherat C, et al. O6-Methylguanine-DNA methyltransferase status in neuroendocrine tumours: prognostic relevance and association with response to alkylating agents. Br J Cancer. 2015;112(3):523–31.

164. Booth MJ, Ost TW, Beraldi D, et al. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. Nat Protoc. 2013;8(10):1841–51.

165. Booth MJ, Branco MR, Ficz G, et al. Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. Science. 2012;336(6083):934–7.

166. Yu M, Hon Gary C, Szulwach Keith E, et al. Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. Cell. 2012;149(6):1368–80.

167. Nestor CE, Meehan RR. Hydroxymethylated DNA Immunoprecipitation (hmeDIP). In: Stockert JC, Espada J, Blázquez-Castro A, editors. Functional Analysis of DNA and Chromatin. Totowa, NJ: Humana Press; 2014. p. 259–67.

168. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics. 2016;8(3):389–99.

169. Geeleher P, Hartnett L, Egan LJ, Golden A, Raja Ali RA, Seoighe C. Gene-set analysis is severely biased when applied to genome-wide methylation data. Bioinformatics. 2013;29(15):1851–7.

170. Affinito O, Palumbo D, Fierro A, et al. Nucleotide distance influences co-methylation between nearby CpG sites. Genomics. 2020;112(1):144–50.

171. Pavel M, Baudin E, Couvelard A, et al. ENETS Consensus Guidelines for the management of patients with liver and other distant metastases from neuroendocrine neoplasms of foregut, midgut, hindgut, and unknown primary. Neuroendocrinology. 2012;95(2):157–76.

172. Rinke A, Wittenberg M, Schade-Brittinger C, et al. Placebo-Controlled, Double-Blind, Prospective, Randomized Study on the Effect of Octreotide LAR in the Control of Tumor Growth in Patients with Metastatic Neuroendocrine Midgut Tumors (PROMID): Results of Long-Term Survival. Neuroendocrinology. 2017;104(1):26–32.

173. Strosberg J, El-Haddad G, Wolin E, et al. Phase 3 Trial of (177)Lu-Dotatate for Midgut Neuroendocrine Tumors. N Engl J Med. 2017;376(2):125–35.

174. Bertoni I, Soares MF, Roberts ISD, Connor T. Radiation nephropathy is associated with a glomerular thrombotic microangiopathy and progression to end-stage kidney disease. Clinical Kidney Journal. 2023;16(9):1534–7.

175. Geenen L, Nonnekens J, Konijnenberg M, Baatout S, De Jong M, Aerts A. Overcoming nephrotoxicity in peptide receptor radionuclide therapy using [177Lu]Lu-DOTA-TATE for the treatment of neuroendocrine tumours. Nuclear Medicine and Biology. 2021;102-103:1–11.

176. Raymond E, Dahan L, Raoul JL, et al. Sunitinib malate for the treatment of pancreatic neuroendocrine tumors. N Engl J Med. 2011;364(6):501–13.

177. Yao JC, Fazio N, Singh S, et al. Everolimus for the treatment of advanced, non-functional neuroendocrine tumours of the lung or gastrointestinal tract (RADIANT-4): a randomised, placebo-controlled, phase 3 study. Lancet. 2016;387(10022):968–77.

178. Cullinane C, Waldeck K, Kirby L, et al. Enhancing the anti-tumour activity of (177)Lu-DOTA-octreotate radionuclide therapy in somatostatin receptor-2 expressing tumour models by targeting PARP. Sci Rep. 2020;10(1):10196.

179. Lines KE, Stevenson M, Filippakopoulos P, et al. Epigenetic pathway inhibitors represent potential drugs for treating pancreatic and bronchial neuroendocrine tumors. Oncogenesis. 2017;6(5):e332.

180. Guenter R, Aweda T, Carmona Matos DM, et al. Overexpression of somatostatin receptor type 2 in neuroendocrine tumors for improved Ga68-DOTATATE imaging and treatment. Surgery. 2020;167(1):189–96.

181. Shah RR. Safety and Tolerability of Histone Deacetylase (HDAC) Inhibitors in Oncology. Drug Saf. 2019;42(2):235–45.

182. Carvalho KC, Cunha IW, Rocha RM, et al. GLUT1 expression in malignant tumors and its use as an immunodiagnostic marker. Clinics (Sao Paulo). 2011;66(6):965–72.

183. Monfort A, Wutz A. Breathing-in epigenetic change with vitamin C. EMBO Rep. 2013;14(4):337–46.

184. Caplin ME, Pavel M, Ćwikła JB, et al. Lanreotide in metastatic enteropancreatic neuroendocrine tumors. N Engl J Med. 2014;371(3):224–33.

185. Caplin ME, Pavel M, Phan AT, et al. Lanreotide autogel/depot in advanced enteropancreatic neuroendocrine tumours: final results of the CLARINET open-label extension study. Endocrine. 2021;71(2):502–13.

186. Concolino P, Costella A, Capoluongo E. Multiple endocrine neoplasia type 1 (MEN1): An update of 208 new germline variants reported in the last nine years. Cancer Genet. 2016;209(1-2):36–41.

187. Dershem R, Gorvin CM, Metpally RPR, et al. Familial Hypocalciuric Hypercalcemia Type 1 and Autosomal-Dominant Hypocalcemia Type 1: Prevalence in a Large Healthcare Population. Am J Hum Genet. 2020;106(6):734–47.

188. Dewey FE, Murray MF, Overton JD, et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science. 2016;354(6319):aaf6814.

189. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nature Genetics. 2014;46(3):310–5.

190. Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet. 2016;99(4):877–85.

191. Chen Y, Lu H, Zhang N, Zhu Z, Wang S, Li M. PremPS: Predicting the impact of missense mutations on protein stability. PLoS Comput Biol. 2020;16(12):e1008543.

192. Pejaver V, Urresti J, Lugo-Martinez J, et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. Nature Communications. 2020;11(1):5918.

193. Shihab HA, Gough J, Cooper DN, et al. Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. Human Mutation. 2013;34(1):57–65.

194. Carter H, Douville C, Stenson PD, Cooper DN, Karchin R. Identifying Mendelian disease genes with the Variant Effect Scoring Tool. BMC Genomics. 2013;14(3):S3.

195. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nature Methods. 2010;7(4):248–9.

196. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Research. 2003;31(13):3812–4.

197. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. 2015;31(16):2745–7.

198. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Research. 2011;39(17):e118–e.

199. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nature Methods. 2010;7(8):575–6.

200. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome Res. 2009;19(9):1553–61.

201. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput Biol. 2010;6(12):e1001025.

202. Garber M, Guttman M, Clamp M, Zody MC, Friedman N, Xie X. Identifying novel constrained elements by exploiting biased substitution patterns. Bioinformatics. 2009;25(12):i54–62.

203. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 2010;20(1):110–21.

204. Siepel A, Bejerano G, Pedersen JS, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 2005;15(8):1034–50.

205. Tekpinar M, David L, Henry T, Carbone A. PRESCOTT: a population aware, epistatic and structural model accurately predicts missense effect. medRxiv. 2024:2024.02.03.24302219.

206. Jagota M, Ye C, Albors C, et al. Cross-protein transfer learning substantially improves disease variant prediction. Genome Biol. 2023;24(1):182.

207. Kellogg EH, Leaver-Fay A, Baker D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. Proteins: Structure, Function, and Bioinformatics. 2011;79(3):830–8.

208. Leaver-Fay A, Tyka M, Lewis SM, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol. 2011;487:545–74.

209. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583–9.

210. Lloyd R, Osamura R, Kloppel G. WHO classification of tumours of endocrine organs: IARC; 2017.

211. Eyres M, Lanfredini S, Xu H, et al. TET2 Drives 5hmc Marking of GATA6 and Epigenetically Defines Pancreatic Ductal Adenocarcinoma Transcriptional Subtypes. Gastroenterology. 2021;161(2):653–68.e16.

212. Tian Y, Morris TJ, Webster AP, et al. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. Bioinformatics. 2017;33(24):3982–4.

213. Loyfer N, Magenheim J, Peretz A, et al. A DNA methylation atlas of normal human cell types. Nature. 2023;613(7943):355–64.

214. Gujar H, Liang JW, Wong NC, Mozhui K. Profiling DNA methylation differences between inbred mouse strains on the Illumina Human Infinium MethylationEPIC microarray. PLoS One. 2018;13(3):e0193496.

215. Liu Y, Bhagwate A, Winham SJ, et al. Quality control recommendations for RNASeq using FFPE samples based on pre-sequencing lab metrics and post-sequencing bioinformatics metrics. BMC Med Genomics. 2022;15(1):195.

216. Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 2010;38(6):1767–71.

217. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biology. 2010;11(2):R14.

218. Evers BM, Townsend CM, Jr., Upp JR, et al. Establishment and characterization of a human carcinoid in nude mice and effect of various agents on tumor growth. Gastroenterology. 1991;101(2):303–11.

219. Kaku M, Nishiyama T, Yagawa K, Abe M. Establishment of a carcinoembryonic antigen-producing cell line from human pancreatic carcinoma. Gann. 1980;71:596–601.

220. Souren NY, Fusenig NE, Heck S, et al. Cell line authentication: a necessity for reproducible biomedical research. Embo j. 2022;41(14):e111307.

221. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association. 1958;53(282):457–81.

222. Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological). 1972;34(2):187–202.

223. Girden ER. ANOVA: Repeated measures. Thousand Oaks, CA, US: Sage Publications, Inc; 1992. vi, 77–vi, p.

224. An K. Sulla determinazione empirica di una legge didistribuzione. Giorn Dell'inst Ital Degli Att. 1933;4:89–91.

225. Smirnov NV. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. Bull Math Univ Moscou. 1939;2(2):3–14.

226. Kim H-Y. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Restor Dent Endod. 2013;38(1):52–4.

227. Hippel Pv. Skewness. In: Lovric M, editor. International Encyclopedia of Statistical Science. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 1340–2.

228. DeCarlo LT. On the meaning and use of kurtosis. Psychological Methods. 1997;2(3):292–307.

229. Teschendorff AE, Marabita F, Lechner M, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013;29(2):189–96.

230. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics. 2016;32(2):286–8.

231. Maksimovic J, Oshlack A, Phipson B. Gene set enrichment analysis for genome-wide DNA methylation data. Genome Biol. 2021;22(1):173.

232. Zhao M, Sun J, Zhao Z. TSGene: a web resource for tumor suppressor genes. Nucleic Acids Research. 2012;41(D1):D970–D6.

233. Ghosh R, Harrison SM, Rehm HL, Plon SE, Biesecker LG. Updated recommendation for the benign stand-alone ACMG/AMP criterion. Hum Mutat. 2018;39(11):1525–30.

234. Tian Y, Pesaran T, Chamberlin A, et al. REVEL and BayesDel outperform other in silico meta-predictors for clinical variant classification. Sci Rep. 2019;9(1):12752.

235. Livesey BJ, Marsh JA. Interpreting protein variant effects with computational predictors and deep mutational scanning. Dis Model Mech. 2022;15(6).

236. Gerasimavicius L, Livesey BJ, Marsh JA. Correspondence between functional scores from deep mutational scans and predicted effects on protein stability. Protein Science. 2023;32(7):e4688.

237. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. Nucleic Acids Research. 2005;33(suppl_2):W382–W8.

238. Caswell RC, Owens MM, Gunning AC, Ellard S, Wright CF. Using Structural Analysis In Silico to Assess the Impact of Missense Variants in MEN1. J Endocr Soc. 2019;3(12):2258–75.

239. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Research. 2017;46(D1):D1062–D7.

240. Shah N, Hou YC, Yu HC, et al. Identification of Misclassified ClinVar Variants via Disease Population Prevalence. Am J Hum Genet. 2018;102(4):609–19.

241. Lines KE, Javid M, Reed AAC, et al. Genetic background influences tumour development in heterozygous Men1 knockout mice. Endocr Connect. 2020;9(5):426–37.

242. Cordes MH, Davidson AR, Sauer RT. Sequence space, folding and protein design. Curr Opin Struct Biol. 1996;6(1):3–10.

243. Beesley LJ, Salvatore M, Fritsche LG, et al. The emerging landscape of health research based on biobanks linked to electronic health records: Existing resources, statistical challenges, and potential opportunities. Stat Med. 2020;39(6):773–800.

244. Pieterman CRC vLR, van den Broek MFM. Multiple Endocrine Neoplasia Type 1. . Feingold KR AB, Blackman MR, et al.,, editor2021.

245. Marini F, Giusti F, Brandi ML. Multiple endocrine neoplasia type 1: extensive analysis of a large database of Florentine patients. Orphanet J Rare Dis. 2018;13(1):205.

246. Kozak M. Structural features in eukaryotic mRNAs that modulate the initiation of translation. J Biol Chem. 1991;266(30):19867–70.

247. Kuper WFE, van Alfen C, van Eck L, et al. The c.1A > C start codon mutation in CLN3 is associated with a protracted disease course. JIMD Rep. 2020;52(1):23–7.

248. Nelakurti DD, Pappula AL, Rajasekaran S, Miles WO, Petreaca RC. Comprehensive Analysis of MEN1 Mutations and Their Role in Cancer. Cancers (Basel). 2020;12(9).

249. Borkin D, He S, Miao H, et al. Pharmacologic inhibition of the Menin-MLL interaction blocks progression of MLL leukemia in vivo. Cancer Cell. 2015;27(4):589–602.

250. Martincorena I, Raine KM, Gerstung M, et al. Universal Patterns of Selection in Cancer and Somatic Tissues. Cell. 2017;171(5):1029–41.e21.

251. Parveen N, Dhawan S. DNA Methylation Patterning and the Regulation of Beta Cell Homeostasis. Frontiers in Endocrinology. 2021;12:651258.

252. Peng D, Ge G, Gong Y, et al. Vitamin C increases 5-hydroxymethylcytosine level and inhibits the growth of bladder cancer. Clin Epigenetics. 2018;10(1):94.

253. Chen K, Zhang J, Guo Z, et al. Loss of 5-hydroxymethylcytosine is linked to gene body hypermethylation in kidney cancer. Cell Res. 2016;26(1):103–18.

254. Lines KE, Vas Nunes RP, Frost M, Yates CJ, Stevenson M, Thakker RV. A MEN1 pancreatic neuroendocrine tumour mouse model under temporal control. Endocr Connect. 2017;6(4):232–42.

255. Grillo F, Bruzzone M, Pigozzi S, et al. Immunohistochemistry on old archival paraffin blocks: is there an expiry date? J Clin Pathol. 2017;70(11):988–93.

256. Kyriakopoulos G, Mavroeidi V, Chatzellis E, Kaltsas GA, Alexandraki KI. Histopathological, immunohistochemical, genetic and molecular markers of neuroendocrine neoplasms. Ann Transl Med. 2018;6(12):252.

257. Hofland J, Refardt JC, Feelders RA, Christ E, de Herder WW. Approach to the Patient: Insulinoma. The Journal of clinical endocrinology and metabolism. 2023;109(4):1109–18.

258. Ionescu-Tirgoviste C, Gagniuc PA, Gubceac E, et al. A 3D map of the islet routes throughout the healthy human pancreas. Scientific Reports. 2015;5(1):14634.

259. Cejas P, Drier Y, Dreijerink KMA, et al. Enhancer signatures stratify and predict outcomes of non-functional pancreatic neuroendocrine tumors. Nat Med. 2019;25(8):1260–5.

260. Papaspyropoulos A, Bradley L, Thapa A, et al. RASSF1A uncouples Wnt from Hippo signalling and promotes YAP mediated differentiation via p73. Nature Communications. 2018;9(1):424.

261. Tomita T, Masuzaki H, Noguchi M, et al. GPR40 gene expression in human pancreas and insulinoma. Biochemical and Biophysical Research Communications. 2005;338(4):1788–90.

262. Briscoe CP, Tadayyon M, Andrews JL, et al. The Orphan G Protein-coupled Receptor GPR40 Is Activated by Medium and Long Chain Fatty Acids *. Journal of Biological Chemistry. 2003;278(13):11303–11.

263. Rady B, Liu J, Huang H, et al. A FFAR1 full agonist restores islet function in models of impaired glucose-stimulated insulin secretion and diabetic non-human primates. Front Endocrinol (Lausanne). 2022;13:1061688.

264. Bailey CJ. Could FFAR1 assist insulin secretion in type 2 diabetes? The Lancet. 2012;379(9824):1370–1.

265. Houthuijzen JM. For Better or Worse: FFAR1 and FFAR4 Signaling in Cancer and Diabetes. Molecular Pharmacology. 2016;90(6):738–43.

266. Serfling E, Jasin M, Schaffner W. Enhancers and eukaryotic gene transcription. Trends in Genetics. 1985;1:224–30.

267. Leibiger IB, Leibiger B, Berggren PO. Insulin signaling in the pancreatic beta-cell. Annu Rev Nutr. 2008;28:233–51.

268. Ardestani A, Maedler K. The Hippo Signaling Pathway in Pancreatic β-Cells: Functions and Regulations. Endocrine Reviews. 2017;39(1):21–35.

269. Cebola I, Rodríguez-Seguí SA, Cho CHH, et al. TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors. Nature Cell Biology. 2015;17(5):615–26.

270. George NM, Day CE, Boerner BP, Johnson RL, Sarvetnick NE. Hippo signaling regulates pancreas development through inactivation of Yap. Mol Cell Biol. 2012;32(24):5116–28.

271. Gao T, Zhou D, Yang C, et al. Hippo signaling regulates differentiation and maintenance in the exocrine pancreas. Gastroenterology. 2013;144(7):1543–53, 53.e1.

272. Itoh Y, Kawamata Y, Harada M, et al. Free fatty acids regulate insulin secretion from pancreatic β cells through GPR40. Nature. 2003;422(6928):173–6.

273. Bartoov-Shifman R, Ridner G, Bahar K, Rubins N, Walker MD. Regulation of the Gene Encoding GPR40, a Fatty Acid Receptor Expressed Selectively in Pancreatic β Cells*. Journal of Biological Chemistry. 2007;282(32):23561–71.

274. Masuho I, Kise R, Gainza P, et al. Rules and mechanisms governing G protein coupling selectivity of GPCRs. Cell Reports. 2023;42(10):113173.

275. Petersen JE, Pedersen MH, Dmytriyeva O, et al. Free fatty acid receptor 1 stimulates cAMP production and gut hormone secretion through Gq-mediated activation of adenylate cyclase 2. Mol Metab. 2023;74:101757.

276. Mariot P, Vanoverberghe K, Lalevée N, Rossier MF, Prevarskaya N. Overexpression of an α_{1H} (Ca_v3.2) T-type Calcium Channel during Neuroendocrine Differentiation of Human Prostate Cancer Cells *. Journal of Biological Chemistry. 2002;277(13):10824–33.

277. Garber SS, Hoshi T, Aldrich RW. Regulation of ionic currents in pheochromocytoma cells by nerve growth factor and dexamethasone. J Neurosci. 1989;9(11):3976–87.

278. Monteith GR, Prevarskaya N, Roberts-Thomson SJ. The calcium–cancer signalling nexus. Nature Reviews Cancer. 2017;17(6):373–80.

279. GURLO T, GEORGIA S, DHAWAN S. DNA Hydroxymethylation Regulates Beta-Cell Maturation and Expansion. Diabetes. 2018;67(Supplement_1).

280. Körner M, Christ E, Wild D, Reubi JC. Glucagon-like peptide-1 receptor overexpression in cancer and its impact on clinical applications. Frontiers in Endocrinology. 2012;3.

281. Nauck MA, Vilsbøll T, Gallwitz B, Garber A, Madsbad S. Incretin-Based Therapies: Viewpoints on the way to consensus. Diabetes Care. 2009;32(suppl_2):S223–S31.

282. Strosberg J, Kvols L. Antiproliferative effect of somatostatin analogs in gastroenteropancreatic neuroendocrine tumors. World J Gastroenterol. 2010;16(24):2963–70.

283. Tentler JJ, Hadcock JR, Gutierrez-Hartmann A. Somatostatin acts by inhibiting the cyclic 3',5'-adenosine monophosphate (cAMP)/protein kinase A pathway, cAMP response element-binding protein (CREB) phosphorylation, and CREB transcription potency. Mol Endocrinol. 1997;11(7):859–66.

284. Tronci L, Serreli G, Piras C, et al. Vitamin C Cytotoxicity and Its Effects in Redox Homeostasis and Energetic Metabolism in Papillary Thyroid Carcinoma Cell Lines. Antioxidants (Basel). 2021;10(5).

285. Ear PH, Marinoni I, Dayton T, et al. NET Models Meeting 2024 white paper: the current state of neuroendocrine tumour research models and our future aspirations. Endocrine Oncology. 2024;4(1):e240055.

286. Lou X, Ye Z, Xu X, et al. Establishment and characterization of the third non-functional human pancreatic neuroendocrine tumor cell line. Human Cell. 2022;35(4):1248–61.

287. Benten D, Behrang Y, Unrau L, et al. Establishment of the First Well-differentiated Human Pancreatic Neuroendocrine Tumor Model. Molecular Cancer Research. 2018;16(3):496–507.

288. Viol F, Sipos B, Fahl M, et al. Novel preclinical gastroenteropancreatic neuroendocrine neoplasia models demonstrate the feasibility of mutation-based targeted therapy. Cellular Oncology. 2022;45(6):1401–19.

289. Capodanno Y, Altieri B, Elders R, Colao A, Faggiano A, Schrader J. Canine insulinoma as a model for human malignant insulinoma research: Novel perspectives for translational clinical studies. Translational Oncology. 2022;15(1):101269.

290. Asfari M, Janjic D, Meda P, Li G, Halban PA, Wollheim CB. Establishment of 2-mercaptoethanol-dependent differentiated insulin-secreting cell lines. Endocrinology. 1992;130(1):167–78.

291. Tsonkova VG, Sand FW, Wolf XA, et al. The EndoC-βH1 cell line is a valid model of human beta cells and applicable for screenings to identify novel drug target candidates. Mol Metab. 2018;8:144–57.

292. Blanchi B, Taurand M, Colace C, et al. EndoC-βH5 cells are storable and ready-to-use human pancreatic beta cells with physiological insulin secretion. Mol Metab. 2023;76:101772.

293. Kailey B, van de Bunt M, Cheley S, et al. SSTR2 is the functionally dominant somatostatin receptor in human pancreatic β- and α-cells. Am J Physiol Endocrinol Metab. 2012;303(9):E1107–16.

8. Supplementary Appendix

8.1 Supplementary Tables

Table 43: Quality control data after all samples were uploaded into R

	Total number of CpGs	Detection p value >0.01	Beadcount <3 in at least 5% of samples	Non-Cg probes	SNPs identified in Zhou et al., 2016	Multiple locations (Nordlund et al.,)	Location X or Y chromosome
Normal pancreatic tissue (n=7)	727,322	10,817	13,553	2,838	95,509	11	15,868
5hmC analysis in all PNETs (n=9)	723,389	14,497	14,749	2,819	94,740	11	15,713
5hmC insulinomas (n=4)	734,439	9,592	6,543	2,865	96,327	11	16,141
5hmC NF-PNETs (n=5)	730,927	11,099	9,298	2,816	95,994	19	16,085

Table 44: Significantly hydroxymethylated genes in normal pancreatic tissue using R package "GoMeth'

Genes								
AACS	ACSM5	AFG3L2	ALDOC	APH1B	ASAH1	ATXN1	BDKRB1	CACNA1S
AANAT	ACSM6	AFMID	ALG1	APIP	ASAH2	ATXN10	BECN1	CACNA2D1
AARS1	ACSS1	AFP	ALG11	APOA1	ASAP1	ATXN1L	BET1	CACNA2D2
AARS2	ACSS2	AGAP1	ALG14	APOA4	ASAP2	ATXN2	BHLHA15	CACNA2D4
AASDHPPT	ACTA2	AGAP3	ALG2	APOA5	ASAP3	ATXN3	BHLHE40	CACNB1
AASS	ACTG1	AGER	ALG6	APOBEC3H	ASH1L	AUH	BHLHE41	CACNB2
ABAT	ACTN1	AGK	ALG9	APOC3	ASIC4	AVPR1B	BHMT2	CACNB3
ABCA1	ACTN2	AGL	ALK	APOH	ASPA	AXIN1	BICD1	CACNB4
ABCA2	ACTN3	AGPAT1	ALOX5	APP	ASPH	AXIN2	BICD2	CACNG1
ABCA4	ACTN4	AGPAT3	ALPG	APPL1	ASPSCR1	AZI2	BIK	CALCR
ABCA5	ACTR2	AGPAT4	ALPL	AQP1	ASRGL1	B2M	BIN1	CALD1
ABCA9	ACTR3	AGPS	ALS2	AQP2	ASS1	B3GALNT1	BIRC6	CALM2
ABCB9	ACTR3C	AGRN	AMBRA1	AQP4	ATF6	B3GALNT2	BLNK	CALML4
ABCC1	ACVR1	AGRP	AMD1	AQP5	ATF6B	B3GALT1	BLVRA	CALML5
ABCC10	ACVR1B	AGT	AMDHD1	AQP9	ATG10	B3GALT2	BLVRB	CALML6
ABCC12	ACVR2A	AHCTF1	AMDHD2	ARAP1	ATG13	B3GALT4	BMF	CALR
ABCC3	ACVR2B	AHCY	AMFR	ARF1	ATG16L2	B3GALT5	BMI1	CAMK1D
ABCC5	ACY1	AHCYL1	AMHR2	ARF3	ATG2A	B3GLCT	BMP2	CAMK2A
ABCC9	ADA	AHCYL2	AMOTL1	ARF6	ATG2B	B3GNT2	BMP2K	CAMK2B
ABCD3	ADAM10	AHNAK	AMOTL2	ARFGAP1	ATG4D	B3GNT4	BMP8A	CAMK2D
ABCG1	ADAM17	AHNAK2	AMPD3	ARFGAP3	ATG5	B3GNT5	BMPR1A	CAMK2G
ABHD16A	ADAR	AHR	AMY2B	ARFGEF1	ATG7	B3GNT7	BMPR1B	CAMKK1
ABHD5	ADCY10	AIFM2	ANAPC13	ARFGEF2	ATM	B4GALNT3	BNIP1	CAMKK2
ABI1	ADCY3	AIP	ANAPC4	ARHGAP10	ATP1A1	B4GALT1	BNIP3	CAMKMT
ABL1	ADCY4	AJUBA	ANGPT1	ARHGAP17	ATP1B1	B4GALT2	BNIP3L	CAMP
ABL2	ADCY5	AK2	ANGPTL3	ARHGAP35	ATP2A1	B4GALT4	BOC	CANT1
ABLIM1	ADCY6	AK4	ANK1	ARHGAP5	ATP2A2	BAAT	BPGM	CANX
ABLIM2	ADCY9	AKAP13	ANK3	ARHGDIA	ATP2A3	BABAM1	BPNT2	CAPN1
ACAA1	ADGRB1	AKR1A1	ANPEP	ARHGEF1	ATP2B1	BABAM2	BRAF	CAPN2
ACACA	ADH1C	AKR1B1	ANXA11	ARHGEF11	ATP2B4	BACE1	BRCA2	CAPNS1
ACACB	ADH5	AKR1C4	ANXA2	ARHGEF12	ATP4A	BAD	BTC	CAPZA1
ACADS	ADH6	AKR7L	AOX1	ARHGEF18	ATP4B	BAG2	BTD	CAPZA2
ACADSB	ADH7	AKT1	AP1B1	ARHGEF28	ATP5F1A	BAIAP2	BTG2	CAPZB
ACAP2	ADI1	AKT1S1	AP1G1	ARHGEF4	ATP5F1B	BAIAP2L1	BTG3	CARD10
ACAP3	ADIPOR1	AKT2	AP1G2	ARHGEF7	ATP5F1C	BAIAP3	BTG4	CARD14
ACAT2	ADIPOR2	AKT3	AP1M1	ARID1A	ATP5F1D	BARD1	BTRC	CARD9
ACBD3	ADK	ALAS1	AP1S3	ARID1B	ATP5MC2	BATF2	BUB1B	CARM1
ACER3	ADM	ALB	AP2A1	ARID2	ATP5MC3	BATF3	BUB3	CARNS1
ACLY	ADM2	ALCAM	AP2A2	ARL8A	ATP5MF	BBOX1	BVES	CARS1
ACO1	ADO	ALDH18A1	AP3B1	ARL8B	ATP5PB	BCAR1	C1GALT1	CARS2
ACOT7	ADORA2B	ALDH1A2	AP3B2	ARNT	ATP5PD	BCAT1	C1QA	CASP10
ACOX2	ADORA3	ALDH1A3	AP3D1	ARNT2	ATP6V0C	BCAT2	C3	CASP2
ACOX3	ADPGK	ALDH1L1	AP3M1	ARNTL	ATP6V0E1	BCKDHB	CA12	CASP4
ACP1	ADRM1	ALDH1L2	AP3M2	ARPC1A	ATP6V1B2	BCL10	CA7	CASP5
ACP5	ADSL	ALDH2	AP3S1	ARPC1B	ATP6V1C1	BCL2	CAB39	CASP7
ACSF3	ADSS1	ALDH3A1	AP3S2	ARPC2	ATP6V1C2	BCL2L1	CACNA1B	CASP8
ACSL1	ADSS2	ALDH4A1	AP4S1	ARRB1	ATP6V1D	BCL2L11	CACNA1C	CASP9
ACSL3	AFDN	ALDH9A1	APC	ARRB2	ATP6V1G1	BCL3	CACNA1D	CASR
ACSL5	AFF1	ALDOA	APC2	ARSB	ATP6V1H	BCL6	CACNA1G	CAST
ACSL6	AFF4	ALDOB	APEX1	AS3MT	ATP7B	BDH1	CACNA1H	CASTOR1

Genes								
CAV2	CDK14	CHRNB1	COA3	CREB1	CTSO	DCHS2	DIO2	DTX3
CBLB	CDK2	CHRNB3	COA5	CREB3L1	CTSS	DCK	DIS3L	DTX3L
CBR1	CDK5	CHRND	COL11A1	CREB3L2	CTSZ	DCLRE1C	DISP1	DUOX1
CBR4	CDK6	CHST1	COL11A2	CREB5	CTTN	DCP1A	DLD	DUSP1
CBS	CDK7	CHST10	COL12A1	CREBBP	CUL1	DCP1B	DLG1	DUSP10
CCDC12	CDK9	CHST11	COL13A1	CREM	CUL3	DCPS	DLG2	DUSP16
CCDC6	CDKN1A	CHST12	COL14A1	CRHR1	CUL5	DCTD	DLG4	DUSP2
CCDC88C	CDKN1B	CHST6	COL15A1	CRHR2	CUL7	DCTN1	DLG5	DUSP3
CCL15	CDKN1C	CHSY1	COL16A1	CRK	CWC15	DCTN2	DLGAP1	DUSP4
CCL2	CEBPB	CHSY3	COL17A1	CRKL	CXCL10	DCTN4	DLL1	DUSP5
CCL21	CEBPG	CIR1	COL18A1	CRLS1	CXCL12	DCTN6	DLL4	DUSP6
CCL28	CENPS	CISH	COL1A1	CRPPA	CXCL16	DCTPP1	DLX5	DUSP7
CCN4	CEPT1	CKAP4	COL26A1	CRTC1	CXCL2	DDC	DMGDH	DUSP8
CCND2	CERK	CKB	COL27A1	CRTC2	CXCR5	DDO	DMXL1	DUT
CCND3	CERS2	CKMT2	COL28A1	CRTC3	CXCR6	DDOST	DMXL2	DVL1
CCNE1	CERS4	CLCF1	COL3A1	CRY1	CYB5R1	DDX19B	DNAH1	DVL2
CCNH	CERS5	CLCN2	COL4A1	CRY2	CYB5R3	DDX39B	DNAH12	DVL3
CCNT2	CERS6	CLCNKB	COL4A2	CRYAB	CYBA	DDX42	DNAH14	DYNC1I1
CCS	CETP	CLDN1	COL6A3	CRYL1	CYFIP1	DDX5	DNAH2	DYNC1I2
CD2	CFI	CLDN10	COL9A2	CS	CYFIP2	DDX58	DNAH6	DYNC1LI2
CD200R1L	CFL2	CLDN14	COL9A3	CSF1	CYLD	DDX6	DNAH7	DYNC2H1
CD226	CFLAR	CLDN15	COLGALT1	CSF2	CYP17A1	DEGS1	DNAH8	DYNLL1
CD276	CGN	CLDN20	COLGALT2	CSF3R	CYP19A1	DEGS2	DNAJA2	DYNLRB1
CD2AP	CGNL1	CLDN22	COMT	CSGALNACT1	CYP26B1	DEPTOR	DNAJA3	DYNLT1
CD34	CHAD	CLDN4	COP1	CSGALNACT2	CYP27A1	DERA	DNAJB1	E2F2
CD36	CHCHD10	CLDN6	COQ5	CSNK1A1	CYP2C18	DERL1	DNAJB12	E2F3
CD37	CHD4	CLIP1	COQ6	CSNK1D	CYP2D7	DERL2	DNAJB2	E2F4
CD38	CHD8	CLK4	COX10	CSNK1E	CYP2E1	DERL3	DNAJC3	E2F5
CD44	CHDH	CLN5	COX14	CSNK1G1	CYP2R1	DGKA	DNAJC5	EBAG9
CD48	CHEK1	CLOCK	COX15	CSNK1G2	CYP2W1	DGKD	DNASE2B	ECH1
CD55	CHEK2	CLP1	COX19	CSNK2B	CYP3A4	DGKE	DNM1	ECHDC1
CD58	CHERP	CLPB	COX20	CSTF1	CYP46A1	DGKG	DNM1L	ECI1
CD59	CHIT1	CLTA	COX4I1	CTBP1	CYP51A1	DGKI	DNM2	EDARADD
CD68	CHKA	CLTB	COX6C	CTBP2	CYSLTR2	DGKZ	DNM3	EDC3
CD80	CHMP1A	CLTC	COX7A1	CTF1	CYTH1	DGLUCY	DNMT3A	EDC4
CD81	CHMP1B	CLU	COX7B2	CTLA4	CYTH3	DGUOK	DNMT3B	EDEM1
CD82	CHMP2A	CMBL	CPA1	CTNNA1	DAAM1	DHCR24	DNTT	EDN2
CD9	CHMP2B	CNDP2	CPEB1	CTNNA3	DAAM2	DHCR7	DOCK1	EEF1A1
CDC14A	CHMP3	CNGB1	CPEB2	CTNNB1	DAB2IP	DHDDS	DOCK4	EEF1G
CDC14B	CHMP4A	CNOT1	CPEB3	CTNNBIP1	DAG1	DHFR2	DOLK	EEF2
CDC16	CHMP7	CNOT10	CPEB4	CTNND1	DAGLA	DHH	DOT1L	EEF2K
CDC25A	CHPF	CNOT2	CPLX4	CTNND2	DAGLB	DHRS11	DPF3	EFNA1
CDC25B	CHPT1	CNOT3	CPOX	CTNS	DAPK1	DHRS3	DPM2	EFNB2
CDC27	CHRM1	CNOT4	CPSF1	CTPS1	DAPK2	DHX15	DPP4	EGFR
CDC34	CHRM3	CNOT6	CPSF6	CTRB2	DAPK3	DHX16	DPYD	EGLN1
CDC40	CHRM4	CNOT6L	CPT1A	CTSB	DARS1	DHX36	DPYS	EHD2
CDC42	CHRM5	CNOT8	CPT2	CTSC	DAXX	DIAPH1	DROSHA	EHD4
CDC5L	CHRNA2	CNOT9	CR2	CTSD	DAZAP1	DIAPH3	DSG2	EHMT1
CDH1	CHRNA6	CNTF	CRAT	CTSH	DBI	DICER1	DSP	EHMT2
CDH17	CHRNA9	COA1	CRCP	CTSL	DCAF1	DIO1	DTX1	EIF2AK1

Genes								
EIF2AK2	EPHB4	FAAP100	FIG4	G6PC1	GDF3	GNPTAB	GUSB	HMGA2
EIF2AK4	EPHX1	FAAP24	FIP1L1	G6PC3	GDF6	GNRH1	GXYLT1	HMGCL
EIF2B3	EPHX2	FABP3	FIS1	GAB1	GDF9	GNRH2	GXYLT2	HMGCS1
EIF2S1	EPN1	FADS1	FKBP4	GAB2	GDPD3	GNS	GYG1	HMGCS2
EIF3H	EPN2	FAF1	FKBP5	GABARAPL2	GFPT1	GOSR1	H2AC11	HMOX1
EIF4A3	EPN3	FAH	FKRP	GABBR1	GGA2	GOSR2	H3-3A	HMOX2
EIF4B	EPOR	FAN1	FLCN	GABRB3	GGPS1	GOT1	H3-3B	HNF1A
EIF4E	EPS15	FANCA	FLI1	GABRG3	GGT1	GP1BB	H6PD	HNF1B
EIF4EBP2	EPS15L1	FANCE	FLNB	GABRR2	GGT6	GP9	HACD2	HNF4A
EIF4G1	ERBB2	FANCF	FLNC	GABRR3	GHR	GPAT4	HACD3	HNF4G
EIF4G2	ERBB3	FANCL	FLOT1	GAL3ST1	GHRL	GPD1	HACL1	HNMT
EIF4G3	ERBIN	FAR1	FMO2	GALC	GIPR	GPD2	HAMP	HNRNPA2B1
EIF6	ERC1	FARP2	FMO3	GALE	GIT1	GPER1	HBS1L	HNRNPA3
ELAVL1	ERCC1	FARS2	FNBP1L	GALK1	GIT2	GPLD1	HCLS1	HNRNPC
ELK4	ERCC4	FARSA	FNIP2	GALK2	GLB1	GPR156	HDAC1	HNRNPU
ELL2	ERCC5	FAT4	FOLR1	GALM	GLCE	GPR161	HDAC10	HNRNPUL1
ELL3	ERN1	FBL	FOLR2	GALNS	GLG1	GPSM1	HDAC11	HOMER1
ELMO1	ESAM	FBP1	FOS	GALNT1	GLI3	GPT2	HDAC2	HOMER3
ELMO2	ESD	FBXL3	FOSL1	GALNT10	GLO1	GPX2	HDAC3	HOXA1
ELMO3	ESPL1	FBXO2	FOXA3	GALNT11	GLP1R	GPX3	HDAC4	HPN
ELN	ESR2	FBXO22	FOXO1	GALNT12	GLRA1	GPX4	HDAC5	HPSE2
ELOB	ETHE1	FBXO32	FOXO3	GALNT15	GLS	GRB10	HDAC7	HRC
ELOC	ETNK2	FBXO43	FOXP1	GALNT17	GLUD1	GRB2	HDDC2	HS6ST1
ELOVL2	ETS1	FBXO5	FPGS	GALNT18	GLUL	GRHPR	HDDC3	HS6ST3
ELOVL3	ETS2	FBXO6	FPGT	GALNT2	GLYAT	GRID1	HEPHL1	HSD11B1
ELOVL5	ETV1	FBXW11	FRAS1	GALNT3	GMDS	GRIN2C	HERC1	HSD11B2
ELOVL6	ETV5	FBXW7	FRAT1	GALNT4	GMPPA	GRIN2D	HERC3	HSD17B1
ELOVL7	ETV6	FBXW8	FREM2	GALNT6	GMPR	GRK2	HERC4	HSD17B12
EME1	ETV7	FCGR2A	FRMD1	GALNT7	GMPS	GRK4	HEXA	HSD17B2
EML4	EVC	FCSK	FRMD6	GALNTL6	GNA12	GRK5	HEXD	HSD17B4
ENAH	EVC2	FDFT1	FRS2	GANAB	GNA13	GRK7	HGD	HSD17B8
ENGASE	EVL	FDPS	FSCN1	GANC	GNA14	GRM3	HGS	HSD3B7
ENO1	EWSR1	FER	FTH1	GAPDH	GNAI1	GRM5	HGSNAT	HSP90AA1
ENOPH1	EXOC2	FES	FTL	GAR1	GNAI2	GRM8	HHAT	HSP90AB1
ENOSF1	EXOC4	FGF1	FUCA1	GARS1	GNAL	GSK3B	HHATL	HSPA1A
ENPP1	EXOC7	FGF10	FUCA2	GATA3	GNAO1	GSN	HIBADH	HSPA1L
ENPP4	EXOSC6	FGF18	FURIN	GATA4	GNAQ	GSPT1	HIBCH	HSPA2
ENPP6	EXOSC7	FGF21	FUS	GATB	GNAS	GSTO1	HIF1A	HSPA4
ENTPD4	EXOSC9	FGF22	FUT1	GATC	GNAT1	GSTP1	HIP1	HSPA4L
ENTPD6	EXT1	FGF7	FUT2	GBA2	GNAT2	GTF2A1	HIPK1	HSPB2
ENTPD8	EXT2	FGF9	FUT4	GBE1	GNB1	GTF2B	HIPK2	HSPBP1
EOGT	EXTL3	FGFR1	FXN	GBF1	GNB2	GTF2E2	HIPK3	HSPG2
EP300	EZH1	FGFR2	FXYD1	GBP1	GNB5	GTF2F2	HJV	HTR2B
EPAS1	EZH2	FGFR3	FYCO1	GBP2	GNE	GTF2H4	HK1	HTR4
EPHA1	EZR	FGFR4	FYN	GCK	GNG12	GTF2H5	HK2	HTT
EPHA2	F10	FGG	FZD4	GCLC	GNG3	GTF2I	HKDC1	HYAL1
EPHA3	F11	FGGY	FZD5	GCLM	GNG7	GTF2IRD1	HLA-C	HYAL2
EPHA6	F11R	FHIT	FZD8	GCNT1	GNL3	GUCY1A1	HLA-DMA	HYAL3
EPHB2	F2	FHL2	FZD9	GCNT4	GNMT	GUCY1B1	HLA-DOA	HYI
EPHB3	FAAH	FHOD1	FZR1	GDF15	GNPDA1	GUK1	HLA-E	HYOU1

Genes								
IARS2	IMPA2	ITGB8	KITLG	LDHA	LTA	MAPK1	MET	MITF
ICA1	IMPDH1	ITPK1	KLB	LDHAL6B	LTB4R	MAPK10	MFN1	MKNK1
ICAM1	INHBC	ITPKB	KLC1	LDHB	LTB4R2	MAPK13	MFN2	MKNK2
IDNK	INMT	ITPKC	KLC2	LDHD	LTBP1	MAPK14	MGAT1	MLH1
IDO2	INPP1	ITPR1	KLF3	LDLR	LTBR	MAPK6	MGAT4A	MLKL
IDUA	INPP4A	ITPR2	KLF5	LDLRAP1	LTC4S	MAPK8	MGAT4B	MLLT1
IFI30	INPP4B	ITPR3	KLK1	LEFTY1	LYL1	MAPK8IP1	MGAT5	MLLT3
IFIT1	INPP5A	JAG1	KLKB1	LEFTY2	LYN	MAPK8IP3	MGLL	MLST8
IFIT1B	INPP5B	JAG2	KMT2A	LEPR	LYNX1	MAPKAP1	MGRN1	MLX
IFNAR2	INPP5D	JAK1	KMT2C	LGMN	LYPD6	MAPKAPK2	MGST1	MLXIP
IFNGR1	INPP5F	JAM3	KMT2D	LGR4	LYPLA1	MAPKAPK5	MGST2	MLYCD
IFNGR2	INPP5J	JARID2	KMT2E	LHPP	MACROH2A2	MAPT	MIB1	MMAB
IFNK	INPP5K	JMJD1C	KMT5B	LIFR	MAD1L1	MARCHF6	MICA	MMP14
IFNL1	INPPL1	JMJD7-PLA2G4B	KMT5C	LIG1	MAFA	MARCKS	MINPP1	MMP15
IFNLR1	INSR	JUP	KNG1	LIG3	MAFB	MARCKSL1	MIOX	MMP2
IFT57	INVS	KAT2A	KPNA1	LIMD1	MAG	MARVELD2	MIR101-1	MMP24
IGF1	IP6K1	KAT2B	KPNA2	LIMK2	MAGI1	MARVELD3	MIR103A1	MMP7
IGF1R	IP6K2	KAT6A	KPNA3	LIN54	MAGI2	MASP1	MIR103A2	MNAT1
IGF2R	IPMK	KCND3	KPNA4	LIPA	MAGI3	MASP2	MIR103B1	MOB1A
IGSF11	IPO4	KCNE1	KPNA6	LIPC	MAML1	MAT1A	MIR125A	MOB1B
IHH	IPO5	KCNE2	KRAS	LIPE	MAML2	MAT2B	MIR145	MOCOS
IKBKB	IPO8	KCNE3	KRIT1	LITAF	MAML3	MATR3	MIR150	MOCS1
IKBKE	IPO9	KCNJ1	KRT36	LLGL1	MAN1A1	MAVS	MIR15B	MOGAT1
IL11RA	IQCE	KCNJ11	KRT39	LLGL2	MAN1B1	MAX	MIR16-2	MPC1
IL15	IQGAP1	KCNJ13	KSR1	LMAN2	MAN1C1	MBD4	MIR193B	MPC2
IL17D	IQGAP2	KCNJ14	KSR2	LMBRD1	MAN2A1	MBOAT1	MIR194-1	MPDZ
IL17F	IQSEC1	KCNJ16	KYAT1	LMNA	MAN2A2	MBOAT7	MIR195	MPG
IL17RA	IQSEC3	KCNJ4	KYAT3	LMO2	MANBA	MCAT	MIR199A2	MPI
IL17RB	IRAG1	KCNJ5	KYNU	LMO7	MAP1B	MCCC1	MIR200A	MPL
IL17RE	IRAK2	KCNK10	LAMA1	LNPEP	MAP1LC3A	MCCC2	MIR21	MPST
IL19	IRF1	KCNK3	LAMA4	LPAR2	MAP1LC3B2	MCL1	MIR214	MPZL1
IL1B	IRF3	KCNMA1	LAMA5	LPAR3	MAP2K1	MCM2	MIR215	MRAS
IL1R1	IRF5	KCNMB3	LAMB1	LPAR6	MAP2K3	MCM3	MIR23B	MRC2
IL1RL2	IRF7	KCNN2	LAMB2	LPCAT1	MAP2K4	MCM5	MIR25	MRI1
IL20	IRF9	KCNN3	LAMB3	LPGAT1	MAP2K5	MCM7	MIR26B	MRPL1
IL20RA	IRS1	KCNQ1	LAMB4	LPIN1	MAP2K6	MCU	MIR27B	MRPL12
IL22RA2	IRS2	KCNQ3	LAMC1	LPIN2	MAP2K7	MDH2	MIR29C	MRPL13
IL23A	ITCH	KCNQ4	LAMP1	LPIN3	MAP3K1	MDN1	MIR30B	MRPL14
IL23R	ITGA1	KDELR3	LAPTM4A	LRIG2	MAP3K10	ME3	MIR30C2	MRPL16
IL27	ITGA10	KDM1A	LAPTM4B	LRP1	MAP3K11	MECOM	MIR32	MRPL2
IL4I1	ITGA2	KDM3A	LARGE1	LRP5	MAP3K12	MED12L	MIR331	MRPL20
IL4R	ITGA3	KDM3B	LARGE2	LRP6	MAP3K14	MED13L	MIR335	MRPL22
IL6	ITGA6	KEAP1	LARS2	LRPAP1	MAP3K2	MED16	MIR345	MRPL23
IL6R	ITGA7	KIDINS220	LATS1	LRRC4	MAP3K20	MED27	MIR34B	MRPL24
IL6ST	ITGA9	KIF22	LATS2	LRRK2	MAP3K3	MEF2A	MIR34C	MRPL32
IL9	ITGAE	KIF3A	LBR	LRTOMT	MAP3K4	MEF2C	MIR451A	MRPL33
ILK	ITGAV	KIF5B	LCAT	LSM2	MAP3K5	MEF2D	MIRLET7A3	MRPL35
IMMP1L	ITGB1	KIF5C	LCK	LSM7	MAP4K2	MEIS1	MIRLET7B	MRPL4
IMMP2L	ITGB4	KIF7	LCLAT1	LSP1	MAP4K3	MEN1	MIRLET7D	MRPS11
IMPA1	ITGB5	KIT	LDB1	LSS	MAP4K4	MEP1B	MIRLET7E	MRPS14

Genes								
MRPS16	NAMPT	NELFA	NPNT	OAS3	PAK6	PDE4D	PHGDH	PLCB4
MRPS18A	NANP	NELFB	NPPA	OAT	PALS1	PDE5A	PHKG1	PLCD1
MRPS6	NANS	NELFE	NPPC	OCLN	PAN3	PDE6A	PHLPP1	PLCE1
MRPS9	NAPA	NEO1	NPRL2	ODC1	PANK1	PDE6G	PHOSPHO2	PLCG1
MSH3	NAPEPLD	NEU1	NPRL3	OGA	PANK2	PDE6H	PHYH	PLCG2
MSH6	NARS1	NEU3	NR1D1	OGDH	PANK3	PDE7A	PHYKPL	PLCZ1
MSI2	NARS2	NF1	NR1H3	OGG1	PANX1	PDE7B	PI4K2A	PLD1
MSTN	NAT1	NF2	NR3C1	OPA1	PAOX	PDE8A	PI4KA	PLD2
MT1A	NBN	NFATC1	NR3C2	OPRM1	PAPOLA	PDGFA	PI4KB	PLD3
MT1G	NCALD	NFATC2	NR4A1	OPTN	PAPOLG	PDGFB	PIAS1	PLEKHG5
MT2A	NCBP1	NFATC3	NR4A3	OR2F2	PAPSS1	PDGFC	PIAS2	PLEKHM2
MTHFD1L	NCK1	NFE2L2	NR5A2	OR2H1	PAPSS2	PDGFD	PIDD1	PLIN1
MTHFD2	NCK2	NFKB1	NRBF2	OR2H2	PAQR5	PDGFRA	PIGC	PLIN2
MTHFS	NCKAP1	NFKBIB	NRCAM	OR4D5	PAQR7	PDGFRB	PIGF	PLIN4
MTMR2	NCL	NFKBIE	NRF1	OR51A2	PAQR8	PDHX	PIGG	PLIN5
MTMR4	NCOA1	NFKBIZ	NRG1	OR51S1	PARD3	PDIA3	PIGL	PLK2
MTNR1A	NCOA2	NFYA	NRG2	OR52W1	PARD6B	PDIA6	PIGR	PLN
MTOR	NCOA3	NFYB	NRP1	OR5C1	PARN	PDK4	PIGS	PLOD1
MTR	NCOA4	NFYC	NRXN1	OR5D13	PARP1	PDPK1	PIGU	PLOD2
MUC2	NCOA7	NGEF	NRXN2	OR6M1	PARP3	PDSS2	PIGV	PLPP1
MUC5B	NCOR1	NGF	NRXN3	OR6T1	PARVA	PDXK	PIGW	PLPP2
MUTYH	NCOR2	NGFR	NSD1	OR8D1	PAX6	PECR	PIK3C2A	PLPP3
MVB12B	NDC1	NGLY1	NSD2	ORAI1	PAX8	PEDS1	PIK3C2B	PLPP4
MVD	NDST1	NHEJ1	NSD3	ORAI2	PBRM1	PEDS1-UBE2V1	PIK3CA	PLPP5
MVK	NDST2	NHLRC1	NSF	ORAI3	PBX1	PELO	PIK3CB	PLXNA1
MX1	NDUFA12	NIT2	NSFL1C	ORC4	PBX3	PEMT	PIK3CD	PLXNA2
MYC	NDUFA13	NLGN2	NSMAF	ORC5	PC	PER1	PIK3R1	PLXNA4
MYD88	NDUFA4	NLK	NT5C2	OS9	PCBD1	PER2	PIK3R5	PLXNB2
MYH10	NDUFA4L2	NLN	NT5C3A	OSBPL5	PCBP2	PER3	PIK3R6	PML
MYH14	NDUFA6	NLRC4	NT5C3B	OSMR	PCCA	PEX10	PIKFYVE	PMM1
MYH9	NDUFA7	NLRP6	NTN1	OXA1L	PCCB	PEX12	PILRB	PMVK
MYL12B	NDUFAF2	NMB	NTN3	OXCT1	PCGF2	PEX14	PINK1	PNLDC1
MYLIP	NDUFAF3	NMD3	NTN4	P2RX1	PCGF3	PEX2	PIP4K2A	PNN
MYLK	NDUFAF5	NME1-NME2	NTNG2	P2RX4	PCGF5	PEX26	PIP5K1A	PNP
MYLK2	NDUFAF6	NME2	NTPCR	P2RY1	PCGF6	PEX5	PIP5K1C	PNPLA2
MYO10	NDUFAF8	NME6	NTRK1	P2RY11	PCK1	PEX5L	PIPOX	PNPLA7
MYO1B	NDUFC2	NMNAT2	NUMB	P2RY14	PCK2	PEX6	PISD	PNPO
MYO1C	NDUFS1	NMNAT3	NUMBL	P4HA1	PCNA	PEX7	PKM	PNPT1
MYO1D	NDUFS2	NMUR1	NUP133	P4HA2	PCSK9	PFKFB2	PKN1	PODXL
MYO1E	NDUFS4	NOD1	NUP188	P4HB	PCYT1A	PFKFB3	PLA1A	POFUT1
MYO1F	NDUFS5	NODAL	NUP210	PABPC1	PCYT2	PFKL	PLA2G12A	POFUT2
MYO5A	NDUFV2	NOP58	NUP210L	PABPC4	PDCD1	PFN2	PLA2G3	POLD1
MYO5B	NDUFV3	NOS1AP	NUP58	PABPN1	PDCD4	PFN4	PLA2G4D	POLD3
MYO5C	NECTIN1	NOTCH1	NUP62	PAFAH1B1	PDCD6IP	PGAM1	PLA2G4E	POLD4
MYO6	NECTIN3	NOTCH2	NUP85	PAFAH1B2	PDE11A	PGAM5	PLA2G7	POLE3
NA	NECTIN4	NOTCH4	NUP93	PAH	PDE1A	PGLS	PLAAT3	POLE4
NACA	NEDD4	NPAS2	NUP98	PAICS	PDE1B	PGM1	PLB1	POLR1A
NADK	NEDD4L	NPC2	NVL	PAK1	PDE1C	PGM2	PLCB1	POLR1D
NADK2	NEIL1	NPL	NXF1	PAK2	PDE4A	PGM3	PLCB2	POLR1F
NAGLU	NEK7	NPLOC4	OAS1	PAK4	PDE4B	PGS1	PLCB3	POLR1G

Genes								
POLR2A	PPP2R3A	PROC	PTPRS	RAMP1	RFC3	RPL21	S1PR3	SEMA6C
POLR2B	PPP2R3C	PRPF38B	PTTG1	RANBP17	RFC4	RPL22	SACM1L	SEMA6D
POLR2E	PPP2R5C	PRPF8	PUF60	RAP1A	RFNG	RPL24	SAE1	SEMA7A
POLR2F	PPP2R5D	PRPH	PUM1	RAP1GAP	RFXAP	RPL26	SAG	SENP2
POLR2G	PPP2R5E	PRR5	PUM2	RAPGEF1	RGL2	RPL3	SAR1A	SEPTIN11
POLR2H	PPP3CA	PRUNE1	PVR	RAPGEF2	RGMA	RPL30	SARDH	SEPTIN4
POLR2I	PPP3CC	PSAP	PXMP2	RAPGEF3	RGMB	RPL31	SARS1	SEPTIN5
POLR2L	PPP3R1	PSD3	PXMP4	RAPGEF4	RGS14	RPL37	SAT2	SEPTIN7
POLR2M	PPP4R3A	PSD4	PXN	RAPGEF5	RGS3	RPL37A	SAV1	SEPTIN8
POLR3B	PPT2	PSEN1	PYCARD	RAPGEF6	RHEB	RPL3L	SCARB1	SEPTIN9
POLR3D	PRDM16	PSEN2	PYCR1	RARA	RHOA	RPL4	SCARB2	SERINC3
POLR3E	PRDM2	PSIP1	PYCR2	RARB	RHOB	RPL5	SCD	SERINC5
POLR3F	PRDM6	PSMA1	PYCR3	RASA1	RHOBTB1	RPL7A	SCD5	SERPINA5
POLR3G	PRDX1	PSMA4	PYGB	RASA3	RHOD	RPL9	SCN1B	SERPINB6
POLR3GL	PRDX5	PSMB2	PYGM	RASAL1	RHOH	RPN1	SCN3A	SERPINC1
POLR3H	PRDX6	PSMB5	PYM1	RASAL2	RHOQ	RPP25	SCN4B	SERPING1
POMC	PREB	PSMB6	RAB10	RASSF1	RHOT1	RPS12	SCN9A	SESN1
POMGNT1	PREP	PSMB7	RAB11FIP1	RASSF4	RICTOR	RPS15	SCNN1A	SESN2
POMGNT2	PREX1	PSMB8	RAB11FIP2	RASSF5	RIF1	RPS18	SCO2	SESN3
POMK	PRH1	PSMB9	RAB11FIP3	RASSF6	RILP	RPS2	SCP2	SETD1A
POMT1	PRICKLE1	PSMC1	RAB11FIP4	RB1	RIMKLA	RPS23	SCRIB	SETD1B
POMT2	PRICKLE2	PSMC4	RAB11FIP5	RBBP4	RIMKLB	RPS24	SCTR	SETD2
POP5	PRIM2	PSMD1	RAB14	RBCK1	RIMS1	RPS29	SDC1	SETD7
POU2F1	PRKAA1	PSMD13	RAB1A	RBKS	RIN1	RPS3	SDC2	SETDB1
POU2F2	PRKAB1	PSMD7	RAB1B	RBL1	RIOK2	RPS3A	SDC3	SETDB2
POU5F1	PRKAB2	PSMD8	RAB27B	RBM25	RIPK1	RPS6	SDC4	SETMAR
PPARA	PRKACA	PSMD9	RAB2A	RBP2	RMI1	RPS6KA1	SDHD	SF3A1
PPARD	PRKAG1	PSME4	RAB33B	RBPJ	RNASEH2A	RPS6KA2	SDS	SF3A2
PPCDC	PRKAG2	PSPH	RAB35	RBPJL	RNASEH2B	RPS6KA4	SDSL	SF3B1
PPIF	PRKAR1A	PSTPIP1	RAB3A	RBX1	RNASEL	RPS6KA5	SEC11A	SF3B2
PPIL2	PRKAR1B	PTAFR	RAB4A	RCAN1	RND1	RPS6KB1	SEC13	SF3B3
PPIP5K2	PRKAR2A	PTCH1	RAB5A	RCAN2	RNF31	RPSA	SEC22B	SF3B5
PPM1A	PRKAR2B	PTDSS1	RAB5C	RCHY1	RNF43	RPTOR	SEC23B	SFN
PPM1B	PRKCA	PTDSS2	RAB7A	RCOR1	RNGTT	RRAGC	SEC24B	SFRP5
PPP1CA	PRKCD	PTEN	RAB8B	RDH10	RNPS1	RRAGD	SEC24C	SGCB
PPP1CB	PRKCE	PTGDS	RABEP1	RDH11	RO60	RRAS2	SEC24D	SGK1
PPP1R12A	PRKCH	PTGER3	RAC1	RDH12	ROBO3	RRBP1	SEC31A	SGK2
PPP1R12B	PRKCI	PTGER4	RACK1	RDX	ROCK1	RRM2	SEC61A1	SGK3
PPP1R12C	PRKCQ	PTH2R	RAD1	REC8	ROCK2	RRM2B	SEC61A2	SGMS1
PPP1R14A	PRKCSH	PTK2	RAD21	REG4	ROR1	RSAD2	SEH1L	SGMS2
PPP1R15A	PRKCZ	PTPA	RAD23A	RELB	ROR2	RTN3	SELE	SGO1
PPP1R1B	PRKD1	PTPN11	RAD51B	RELN	RORA	RTN4	SELENOI	SGPP1
PPP1R3C	PRKD2	PTPN2	RAD51C	RELT	RORC	RUBCN	SELENOS	SGSH
PPP1R3D	PRKD3	PTPN6	RAD51D	RET	RPA1	RUFY1	SEMA3B	SH2B1
PPP1R3E	PRKDC	PTPRC	RAD52	RETSAT	RPA2	RUNX2	SEMA3C	SH2B2
PPP2CA	PRKG1	PTPRD	RAF1	REV1	RPA3	RUVBL1	SEMA4B	SH2B3
PPP2CB	PRKN	PTPRF	RAG2	REV3L	RPIA	RXRA	SEMA4C	SH2D2A
PPP2R1B	PRLR	PTPRJ	RALA	REXO1	RPL13	RXRB	SEMA4G	SH3BP2
PPP2R2A	PRMT1	PTPRM	RALB	REXO2	RPL13A	RYK	SEMA5A	SH3GL1
PPP2R2D	PRNP	PTPRN2	RALGDS	RFC1	RPL17	RYR2	SEMA6A	SH3GLB1

Genes								
SH3GLB2	SLC27A1	SMAD3	SORBS1	ST3GAL2	SYNGAP1	TF	TNFSF18	TUBA1B
SHANK2	SLC27A3	SMAD4	SORT1	ST3GAL3	SYNJ2	TFAP4	TNIP1	TUBA1C
SHARPIN	SLC27A5	SMAD5	SOS1	ST3GAL4	SYNPO	TFDP1	TNNC1	TUBA4A
SHC1	SLC29A1	SMAD6	SOS2	ST3GAL5	SYT1	TFEB	TNPO1	TUBB
SHC4	SLC29A2	SMAD7	SOX9	ST6GALNAC1	SYVN1	TFF1	TNPO3	TUBG2
SHMT1	SLC29A3	SMAD9	SP1	ST6GALNAC2	TAB2	TFG	TNR	TUSC3
SHMT2	SLC2A1	SMAP1	SPART	ST6GALNAC3	TAC4	TFPI	TNXB	TXK
SHOC2	SLC2A2	SMAP2	SPATA13	ST6GALNAC4	TAF10	TGFB3	TOB2	TXNDC12
SIAH1	SLC38A1	SMARCA2	SPATA2	ST6GALNAC6	TAF11	TGFBR1	TOLLIP	TXNDC5
SIK2	SLC38A2	SMARCA4	SPATA2L	STAG1	TAF13	TGFBR2	TOMM40L	TXNL4A
SIL1	SLC38A3	SMARCB1	SPATA5	STAG3	TAF15	TGIF1	TOP2B	TXNRD1
SIPA1	SLC38A9	SMARCC1	SPCS2	STAR	TAF4	TGIF2	TP53	TXNRD2
SIPA1L1	SLC39A10	SMARCC2	SPDYC	STARD3	TAF5L	TH	TP53BP2	TYMP
SIPA1L2	SLC39A11	SMARCD1	SPG11	STAT1	TAF6	THBS1	TP53I3	TYMS
SIPA1L3	SLC39A14	SMARCD2	SPG21	STAT2	TAF6L	THBS3	TP53INP2	U2SURP
SIRPA	SLC39A2	SMARCD3	SPHK2	STAT3	TAF8	THOC6	TP63	UAP1
SIRT1	SLC39A4	SMC3	SPI1	STAT4	TALDO1	THOC7	TPCN1	UBA2
SIRT4	SLC39A5	SMG6	SPINT1	STAT5A	TAOK2	THPO	TPCN2	UBA6
SIRT6	SLC39A6	SMG7	SPOP	STAT5B	TAOK3	THRA	TPK1	UBA7
SIRT7	SLC39A7	SMNDC1	SPOPL	STEAP2	TAP1	THRB	TPM1	UBC
SKAP1	SLC39A8	SMOX	SPTBN2	STEAP3	TAP2	THSD4	TPM3	UBE2B
SKIL	SLC3A2	SMPD1	SPTLC2	STIM1	TAPBP	TICAM1	TPM4	UBE2D1
SKIV2L	SLC40A1	SMUG1	SQOR	STIM2	TARS3	TICAM2	TPMT	UBE2D2
SLC10A1	SLC44A1	SMURF1	SQSTM1	STIP1	TAS1R1	TIFA	TRA2A	UBE2D3
SLC11A1	SLC44A2	SMURF2	SRC	STK10	TAX1BP1	TIRAP	TRA2B	UBE2E1
SLC11A2	SLC44A3	SMYD2	SRD5A3	STK11	TBC1D1	TJP1	TRADD	UBE2E2
SLC12A2	SLC44A4	SMYD3	SREBF1	STK3	TBC1D15	TJP2	TRAF1	UBE2E3
SLC12A7	SLC46A1	SNAI1	SRGAP1	STMN1	TBC1D4	TK1	TRAF3	UBE2F
SLC15A1	SLC4A2	SND1	SRP14	STRADA	TBK1	TK2	TRAF3IP2	UBE2G1
SLC16A3	SLC4A4	SNF8	SRP68	STT3A	TBKBP1	TKT	TRAF5	UBE2H
SLC17A5	SLC4A5	SNRNP40	SRP9	STT3B	TBL1XR1	TLE1	TRAF6	UBE2I
SLC19A1	SLC51A	SNRNP70	SRPRB	STX11	TBPL1	TLE2	TRAK2	UBE2J2
SLC1A2	SLC6A1	SNRPA	SRR	STX12	TBXAS1	TLE3	TRAM1	UBE2L6
SLC1A5	SLC6A12	SNRPB	SRSF1	STX17	TCF3	TLN2	TRAP1	UBE2M
SLC22A1	SLC6A19	SNRPC	SRSF10	STX18	TCF7	TLR5	TRIB3	UBE2N
SLC22A4	SLC6A4	SNRPD1	SRSF3	STX19	TCF7L1	TLR6	TRIM5	UBE2O
SLC22A5	SLC6A9	SNU13	SRSF4	STX1A	TCF7L2	TM7SF2	TRIP12	UBE2Q1
SLC22A7	SLC7A1	SNUPN	SRSF5	STX7	TCIRG1	TMBIM6	TRIP6	UBE2W
SLC23A1	SLC7A11	SNW1	SRSF7	STX8	TCOF1	TMED10	TRIT1	UBE2Z
SLC25A10	SLC7A5	SNX1	SS18	SUCLG1	TDO2	TMPRSS2	TRPM2	UBE3A
SLC25A17	SLC7A7	SNX18	SS18L1	SUCLG2	TEAD1	TNF	TRPM6	UBE3B
SLC25A20	SLC8A3	SNX2	SSH1	SUFU	TEAD3	TNFRSF10A	TRPM7	UBE4A
SLC25A29	SLC9A1	SNX4	SSH2	SULT1A1	TEAD4	TNFRSF19	TRPV3	UBE4B
SLC25A4	SLC9A3R1	SNX5	SSR2	SULT1E1	TEC	TNFRSF1A	TRPV4	UBOX5
SLC26A1	SLC9A3R2	SNX9	SSR3	SUMF1	TECPR1	TNFRSF1B	TRRAP	UBQLN1
SLC26A3	SLIT3	SOAT1	SSTR5	SUMO3	TECR	TNFRSF21	TSC2	UBR4
SLC26A4	SLURP1	SOAT2	SSU72	SUMO4	TELO2	TNFRSF4	TSHR	UBR5
SLC26A6	SLX4	SOCS5	SSX2IP	SUOX	TENT4A	TNFSF10	TTC37	UBXN1
SLC26A7	SMAD1	SOCS7	ST14	SUPT3H	TENT4B	TNFSF12	TTN	UBXN6
SLC26A9	SMAD2	SOD1	ST3GAL1	SYMPK	TERT	TNFSF13B	TTR	UCHL1

Genes							
UCK2	UPF2	VANGL2	VPS41	WNT5B	ZBTB17	ZNF30	ZNF718
UCKL1	UPF3A	VAPA	VPS4A	WWC1	ZCCHC7	ZNF302	ZNF737
UGCG	UPP2	VARS1	VTCN1	WWP1	ZEB1	ZNF311	ZNF74
UGDH	UQCR11	VARS2	VTI1A	WWP2	ZEB2	ZNF317	ZNF764
UGGT1	UQCRFS1	VASP	VTI1B	WWTR1	ZFHX3	ZNF331	ZNF766
UGGT2	UQCRQ	VAV2	VWF	XAB2	ZFP1	ZNF34	ZNF783
UGP2	UROS	VAV3	WARS1	XDH	ZFP14	ZNF354B	ZNF79
UGT1A1	USE1	VCAM1	WASF2	XPC	ZFPM2	ZNF429	ZNF791
UGT1A10	USP1	VCL	WASF3	XPO1	ZFYVE1	ZNF43	ZNF8
UGT1A3	USP15	VDAC3	WASHC4	XPO4	ZFYVE16	ZNF436	ZNF836
UGT1A4	USP25	VDR	WASL	XPO6	ZFYVE27	ZNF454	ZNF860
UGT1A5	USP30	VEGFA	WDR33	XPO7	ZFYVE9	ZNF490	ZNF91
UGT1A6	USP39	VEGFB	WDR43	XRCC2	ZMPSTE24	ZNF492	ZNRF3
UGT1A7	USP7	VHL	WDR7	XRCC4	ZNF100	ZNF516	
UGT1A8	USP8	VIM	WDR82	XRCC5	ZNF107	ZNF517	
UGT1A9	UST	VIPR1	WFS1	XRN1	ZNF124	ZNF529	
UGT2A3	UTP14C	VIPR2	WHAMM	XXYLT1	ZNF136	ZNF563	
UGT8	UTS2B	VLDLR	WIPF1	XYLB	ZNF169	ZNF57	
ULK1	UVRAG	VMP1	WIPF3	XYLT1	ZNF195	ZNF570	
ULK2	UXS1	VPS16	WIPI1	YAP1	ZNF212	ZNF589	
UMPS	VAC14	VPS18	WIPI2	YBX3	ZNF222	ZNF619	
UNC5B	VAMP2	VPS28	WNT10A	YOD1	ZNF26	ZNF641	
UNG	VAMP5	VPS33A	WNT11	YWHAG	ZNF267	ZNF670	
UPB1	VAMP8	VPS37B	WNT4	YWHAQ	ZNF268	ZNF682	
UPF1	VANGL1	VPS37C	WNT5A	ZBTB16	ZNF3	ZNF714	

Table 45: Significantly hydroxymethylated genes in insulinomas using R package 'GoMeth'

Genes								
AAK1	ALPK3	ATG5	C3orf67-AS1	CDRT4	CRADD	DHX15	ENTPD5	FEZ2
ABCA1	ANAPC10	ATG7	C4orf36	CDYL2	CREBBP	DHX40	EPB41	FGGY
ABCA5	ANAPC16	ATL3	C5	CELF1	CREBL2	DICER1	EPB41L2	FHDC1
ABCC1	ANGPTL2	ATP11A	C6orf106	CERS6	CRTC3	DIDO1	EPC1	FHIT
ABCF3	ANKH	ATP6V1C1	C6orf129	CFAP126	CRYGC	DIS3L2	EPN2	FIP1L1
ABHD2	ANKRD11	ATP8A1	C6orf154	CFLAR	CRYZL1	DISP1	EPN2-IT1	FKBP11
ABI1	ANKRD12	ATP9A	C8orf86	CGREF1	CSGALNACT1	DLEU1-AS1	EPS15L1	FKBP1A
ABI2	ANKRD40	ATP9B	C9orf3	CHAMP1	CSNK1G1	DLEU7	ERBB2IP	FKBP1A-SDCBP2
ABL1	ANKRD46	ATXN1	CACNA1D	CHCHD3	CSRP1	DLG5	ERC2	FKBP5
ABLIM1	ANKRD6	ATXN7	CACNB2	CHD2	CTAGE5	DLGAP4	ERCC6L2	FLJ10213
ABLIM2	ANKS1A	ATXN7L1	CADM1	CHD4	CTBP2	DNHD1	ERGIC1	FLNB
ACAD10	ANP32B	AUTS2	CAMK2D	CHD6	CTDSPL	DNM3	ERRFI1	FNBP1L
ACBD6	ANXA11	AVIL	CANX	CHD7	CTNNA1	DNMT3A	ETNK1	FNDC3A
ACOT7	AP1S3	AVL9	CAPRIN1	CHD8	CTNNA3	DOCK10	ETS1	FNDC3B
ACOX2	AP2A2	B3GALT6	CAPZB	CHD9	CTNNB1	DOCK3	ETS2	FOXK2
ACSL1	AP3B1	B4GALT1	CASKIN2	CHKA	CUEDC1	DOCK9	EWSR1	FOXN3
ACSL3	APBB2	B4GALT5	CASZ1	CHN2	CUGBP1	DRAIC	EXOC1	FOXO1
ACTN1	APOL6	BAT2L1	CBFA2T2	CHST8	CUL1	DSE	EXOC4	FOXO3
ACTN4	APP	BBIP1	CBLB	CHTOP	CUL4A	DSP	EXOSC2	FOXO6
ACTR3	ARG2	BBX	CC2D2A	CLASP2	CUX1	DST	EXT1	FOXP1
ACVR1	ARGLU1	BCL2L1	CCDC12	CLDN20	CYFIP1	DTNA	EYS	FRAS1
ACVR2A	ARHGAP10	BCL6	CCDC149	CLN6	CYSLTR2	DTNB	EZR	FRS2
ADAM10	ARHGAP12	BCL9	CCDC23	CLOCK	CYTH1	DTNBP1	EZR-AS1	FRY
ADAM9	ARHGAP24	BLOC1S5	CCDC50	CLSTN1	D2HGDH	DUSP15	FAAH	FRYL
ADARB1	ARHGAP26	BMP2K	CCDC57	CMAS	DAAM1	DUSP16	FADS3	FSIP1
ADCY9	ARHGAP35	BMPR1A	CCDC67	CMBL	DACH1	DYM	FAM100A	FTO
ADD3	ARHGEF12	BRD1	CCDC88A	CMIP	DAG1	DYNC1I2	FAM107B	FUBP1
ADD3-AS1	ARHGEF17	BRE	CCDC97	CMTM4	DAGLB	DYNC2H1	FAM114A1	FUCA2
ADIPOR2	ARHGEF28	BRIX1	CCNL2	CMTM7	DAPK1	DYRK1A	FAM131C	FURIN
ADK	ARHGEF3	BRUNOL4	CCNT2	CMTM8	DAPP1	EBAG9	FAM160B1	G2E3
AFF1	ARID1A	BTBD7	CCNY	CNIH3	DARS	ECE1	FAM164C	GAB1
AFF4	ARID1B	BTBD9	CCNYL1	CNKSR3	DBR1	ECHDC2	FAM175B	GAB2
AGAP1	ARID2	BTD	CCSER1	CNNM3	DCBLD1	EDC4	FAM184A	GAK
AGAP3	ARMC9	BZW2	CD247	CNOT1	DCK	EFCAB2	FAM20B	GALK2
AGBL4	ARNTL	C10orf107	CD2AP	CNOT2	DCP1B	EFCAB8	FAM69A	GALNT2
AGFG1	ARPC2	C10orf11	CD44	CNOT4	DCPS	EFNA3	FAM78B	GALNT3
AGL	ARSB	C10orf118	CD59	CNST	DCTN1	EFR3A	FAM8A1	GAPVD1
AGPS	ART4	C10orf76	CD82	COA1	DCTPP1	EGLN1	FARP1	GATA6
AGRP	ASAP1	C11orf49	CDC123	COBLL1	DDAH1	EHMT1	FARS2	GATSL3
AIFM2	ASAP3	C11orf53	CDC14A	COG3	DDHD1	EIF3H	FAT3	GCC1
AIG1	ASB1	C11orf71	CDC42BPB	COL27A1	DDX10	EIF4G3	FBXL17	GCK
AK5	ASB4	C11orf74	CDC42EP3	COLQ	DDX17	ELF1	FBXL3	GCOM1
AKAP1	ASCL5	C14orf145	CDC7	COMMD10	DDX59	ELL2	FBXO18	GDE1
AKAP11	ASH1L	C15orf52	CDC73	COPG2	DDX60	ELMO1-AS1	FBXO22	GDPD3
AKAP13	ASPH	C16orf52	CDGAP	CPA5	DECR1	ELMSAN1	FBXO32	GGA2
AKAP2	ASXL1	C16orf72	CDH17	CPEB3	DENND1A	ELOVL5	FBXO34	GGNBP2
AKIRIN2	ATAD1	C19orf35	CDIP1	CPEB4	DEPDC5	ELP4	FBXW11	GHRLOS
ALDH1A1	ATF6B	C1GALT1	CDK14	CPNE1	DGKD	EMID1	FBXW4	GIGYF1
ALG2	ATF7IP	C20orf196	CDK6	CPQ	DGKZ	EML3	FBXW7	GIGYF2
ALG9	ATG13	C3orf49	CDKAL1	CPSF6	DHRS3	ENAH	FEM1B	GIPR

Genes								
GLG1	HS3ST1	KIAA1958	LOC100128288	MAGI2	MIR4505	N4BP1	NYNRIN	PHYKPL
GLIS3	HSD17B8	KIAA2026	LOC100129148	Magmas	MIR5094	N4BP2L2	OBFC1	PIAS1
GLO1	HSP90AB1	KIF13A	LOC100130357	MAMDC2	MIR5096	NAA60	ODF3L2	PICALM
GLT8D1	HSPA4L	KIF16B	LOC100130691	MAML2	MIR548N	NAB1	ODZ3	PIK3AP1
GLYR1	IFIH1	KIF18B	LOC100286844	MAML3	MIR591	NAP1L4	ODZ4	PIK3C2B
GMDS-AS1	IFIT3	KIF1B	LOC100287036	MAN1C1	MIR644	NAV1	OLA1	PIK3C3
GNAQ	IFNGR1	KIFC1	LOC100288798	MAP1B	MIR6752	NAV2	ONECUT2	PIK3IP1
GNAS	IFT122	KIRREL3	LOC100419583	MAP2K4	MITF	NAV2-AS2	OPA1-AS1	PITPNC1
GNPDA1	IGF1R	KLF7	LOC100506023	MAP3K1	MKL2	NBEAL1	ORAI1	PITPNM1
GNRH1	IL20RA	KLHDC10	LOC100507002	MAP3K13	MKLN1	NCALD	OSBPL8	PJA2
GOLGB1	IL6ST	KLHDC5	LOC100507487	MAP3K9	MLH3	NCBP1	OSBPL9	PKM
GPC6	IMMP2L	KLHL2	LOC101927066	MAP4	MLL5	NCK1	OSGEPL1	PKN2
GPD1L	IMPDH1	KLHL20	LOC101927438	MAP4K3	MLLT10	NCK2	OTUD4	PKP4
GPD2	INO80D	KLHL29	LOC101927701	MAP4K4	MLLT4	NCKAP5	OTUD7B	PLA1A
GPR137B	INPP1	KLHL41	LOC101927780	MAP4K5	MMD	NCOA2	OXA1L	PLAC8L1
GPR161	INPP4A	KLHL5	LOC101927787	MAPK1	MMP16	NCOA3	PACS2	PLCB4
GPR56	INPP4B	KMT2A	LOC101927817	MAPK8IP1	MOBKL1A	NCOR2	PALM2	PLCL2
GPRIN3	INTS6	KRBA1	LOC101928008	MAPK8IP3	MON2	NDFIP2	PANK2	PLCXD2
GRAP2	INVS	KRT7	LOC101928461	MAPKAP1	MORF4L1	NDUFS4	PANK4	PLD1
GRB2	IQSEC1	KRT80	LOC101929450	MAPKBP1	MOV10	NEDD4L	PAPD5	PLEC1
GRHL2	IRF2	KSR1	LOC101929541	MARCH10	MPP5	NEDD9	PARD3	PLEKHA1
GRHL3	IST1	KTI12	LOC101929633	MARCH6	MPPE1	NEK1	PARN	PLEKHA5
GRIN2A	ITCH	KTN1	LOC101929710	MARCH8	MPZL1	NEK7	PARP8	PLEKHA6
GSE1	ITGA3	LAMA1	LOC101929717	MAST2	MR1	NFASC	PARP9	PLEKHG5
GSK3B	ITGB1	LAMA2	LOC102467214	MATR3	MRPL1	NFAT5	PARVA	PLEKHM3
GSTCD	ITPKB	LAMB3	LOC102723344	MB21D2	MRPL14	NFIA	PAX8-AS1	PLK2
GTF2A1	ITPKB-IT1	LAMC1	LOC102723439	MBD2	MRPL23	NFIA-AS1	PBX1	PLS1
GTF2B	ITSN2	LANCL1	LOC102723824	MBNL1	MRPL32	NFIC	PCDH1	PLSCR1
GTF2E2	IVNS1ABP	LARP1	LOC102724552	MBNL2	MRPS28	NFIL3	PCDH7	PLXNA2
GTF2IRD1	JAK1	LARS2	LOC102724784	MBP	MRPS6	NFKBIZ	PCNXL2	PLXNB1
HBS1L	JARID2	LASS5	LOC283856	MBTPS1	MSANTD2	NFU1	PCNXL4	PLXNC1
HDAC2	JAZF1	LASS6	LOC285484	MCAT	MSH2	NFYB	PCTP	POLD3
HDAC4	JMJD1C	LCAT	LOC440839	MCC	MSH3	NIPSNAP3B	PDE3A	POLE4
HDAC9	JPH1	LCOR	LOC441869	MCF2L2	MSI2	NKIRAS2	PDE4B	POU2F1
HDHD2	KALRN	LDB3	LOC651250	MCM3AP	MSL2	NKTR	PDE4D	PPAP2B
HDLBP	KANSL1	LDLRAD4	LOC728730	MCM9	MSMO1	NME6	PDE8A	PPARD
HECTD1	KAT6A	LEMD3	LOC730668	MCTP2	MT1E	NMNAT3	PDGFD	PPFIA1
HECTD4	KAT6B	LEPROTL1	LPCAT3	MDC1	MTA1	NMT2	PDHX	PPFIBP1
HERC3	KCCAT198	LGR5	LPIN2	MDFIC	MTHFD1L	NOC2L	PDIA3P	PPIG
HEXA	KCCAT211	LIMA1	LPP	ME3	MTMR14	NOL4L	PDLIM5	PPM1G
HFE	KCNJ13	LIMCH1	LRBA	MED13L	MTMR2	NOLC1	PDS5B	PPM1H
HGFAC	KCNK7	LIMK2	LRCH1	MEF2A	MTMR3	NOSTRIN	PDZD8	PPP1CA
HIF1A	KCNMA1	LIMS1	LRIG1	MEF2D	MTSS1	NPAS3	PEBP4	PPP1R12B
HIGD1A	KCNQ1	LINC-PINT	LRIG3	MEGF9	MTSS1L	NPEPL1	PELI2	PPP1R14C
HINT2	KCNQ1OT1	LINC00523	LRP4	MEI1	MTUS1	NPM1	PELO	PPP1R3B
HIPK2	KCTD3	LINC00662	LRP5	MEIS1	MTUS2	NPTN	PEMT	PPP1R9A
HIPK3	KCTD6	LINC00673	LRP6	MEIS2	MX1	NR1D2	PET117	PPP2R1B
HISPPD1	KCTD7	LINC00862	LRRC1	MEMO1	MXI1	NR3C1	PEX11A	PPP2R2A
HIVEP2	KDELC2	LINC00870	LRRC32	MEP1B	MYADM	NR3C2	PFDN4	PPP2R3A
HIVEP3	KDM2A	LINC01004	LRRC8B	MFHAS1	MYBPHL	NRBF2	PFKFB2	PPP2R5A
HLTF	KDM4C	LINC01089	LRRC8D	MGA	MYL12B	NRCAM	PGCP	PPP2R5E
HMBOX1	KDM5B	LINC01213	LUZP6	MGMT	MYLIP	NRIP1	PHACTR2	PPP4C
HMG20A	KDM6B	LINC01229	LYPLA1	MICAL3	MYLK	NRSN2	PHACTR4	PPP4R2
HMGCL	KIAA0240	LINC01270	LYRM9	MIEF1	MYO1B	NSMCE2	PHC3	PPP6R3
HMGCLL1	KIAA0247	LINC01353	LYST	MIPOL1	MYO9A	NT5C2	PHF17	PRCC
HMGN3	KIAA0922	LINC01355	MACF1	MIR1306	MYOF	NUDC	PHF20	PRDM11
HMOX1	KIAA1239	LINC01358	MACROD2	MIR3134	MYRF	NUDT14	PHF3	PRDM2
HOOK1	KIAA1324	LMNA	MAFF	MIR3163	MYRIP	NUDT3	PHLDB2	PREX1
HPCAL1	KIAA1522	LMO7	MAGI1	MIR3692	MYST3	NUFIP1	PHLPP1	PRH1-PRR4
HPS1	KIAA1715	LOC100128071	MAGI1-AS1	MIR3960	MYT1L	NUMB	PHTF2	PRICKLE2

Genes									
PRKAG2	RBM12B	RPH3AL	SGK269	SNORA47	STARD3NL	THRAP3	TSHZ2	WASF3	ZNF609
PRKCA	RBM15	RPIA	SGMS1	SNORA70G	STAT5B	THUMPD3-AS1	TSPAN3	WDR1	ZNF638
PRKCE	RBM33	RPL24	SGSM1	SNORD113-2	STAU2	TINAGL1	TSTD2	WHSC1L1	ZNF644
PRKCI	RBM47	RPN2	SH2D4A	SNORD17	STIM1	TJP1	TTBK1	WNK1	ZNF710
PRKD2	RBM9	RPP14	SH2D6	SNORD81	STIM2	TLK1	TTC23L	WNT4	ZNF827
PRKD3	RBMS2	RPS15A	SH3BP5	SNRNP40	STK24	TM4SF1	TTC39B	WRNIP1	ZNF830
PROX1	RBPMS	RPS18	SH3GLB1	SNTB2	STK33	TMCC3	TTPAL	WSB1	ZNRF2
PRPF39	RC3H2	RPS6KA5	SH3PXD2A	SNX14	STK38L	TMEM131	TTR	WWOX	ZRANB1
PRR14L	RCAN1	RPTOR	SH3RF1	SNX29	STK39	TMEM17	TULP4	WWP1	ZSWIM6
PRR7-AS1	RCC1	RRAS2	SHF	SNX3	STT3B	TMEM30A	TXNDC11	WWP2	ZYX
PSAP	RCCD1	RREB1	SHFM1	SNX5	STX17	TMEM37	TXNDC15	XPNPEP1	ZZEF1
PSIP1	RCHY1	RRM2B	SHPRH	SNX9	STX7	TMEM41B	TYMP	XYLT1	ZZZ3
PSMD13	RDH12	RSAD1	SHROOM3	SOBP	STXBP1	TMEM62	UBE2D1	YARS2	
PSMD9	RDH14	RSL24D1	SIAH1	SOCS5	STXBP5	TMEM63A	UBE2E1	YLPM1	
PSME4	REEP3	RUVBL1	SIK2	SOCS6	STXBP5-AS1	TMEM64	UBE2E3	YWHAG	
PTAR1	REPIN1	SAMD4A	SIK3	SORBS1	STXBP6	TMEM91	UBE2F	YWHAQ	
PTBP3	REPS1	SASH1	SIPA1L1	SORL1	SUCLG2	TMEM99	UBE2H	YWHAZ	
PTDSS1	RER1	SATB1	SIPA1L3	SOS1	SUFU	TMOD2	UBE2L3	YY1AP1	
PTK2	RERE	SBF2	SKI	SOX13	SUPT3H	TMOD3	UBE2N	ZBED3-AS1	
PTPLAD1	REST	SC5D	SKIL	SOX6	SWAP70	TNFRSF19	UBE2W	ZBTB1	
PTPN12	REV1	SCAF11	SLC12A3	SP3	SYNE1	TNFRSF1A	UBE3A	ZBTB16	
PTPN4	REV3L	SCAF8	SLC1A2	SPAG9	SZRD1	TNFSF10	UBE3C	ZBTB2	
PTPRJ	RFC1	SCAPER	SLC20A2	SPATA13	TACC2	TNRC6B	UBR1	ZBTB20	
PTPRM	RFWD2	SCFD1	SLC25A46	SPATA18	TADA1	TNS1	UBR3	ZBTB40	
PUM1	RFX3	SCNN1A	SLC29A1	SPATA5	TADA2B	TNS3	UBR5	ZC3H11A	
QKI	RGL1	SCP2	SLC2A13	SPATA7	TAF4	TOM1	UBXN7	ZC3H12D	
QSER1	RGS3	SCYL3	SLC2A2	SPECC1L	TANC1	TOM1L2	UCK2	ZCCHC14	
RAB10	RHEB	SDCBP	SLC31A2	SPEN	TANGO6	TOMM20	UFSP1	ZCCHC7	
RAB12	RHNO1	SEC14L1	SLC35G2	SPG11	TAP2	TOP1	UGCG	ZDHHC13	
RAB1A	RHOBTB1	SEC14L2	SLC37A1	SPIDR	TAX1BP1	TOX3	UGP2	ZEB2	
RAB26	RHOU	SEC22C	SLC38A10	SPIRE1	TBC1D1	TP53I11	UHRF2	ZFAND3	
RAB30	RIN2	SEC24B	SLC38A11	SPOPL	TBC1D10A	TPCN1	UMAD1	ZFAND6	
RAB5A	RINL	SEC24C	SLC3A2	SPPL2A	TBC1D12	TPD52	UNC93B1	ZFHX3	
RAB7A	RIPK2	SEC31A	SLC44A1	SPPL3	TBC1D14	TPP2	UPF3A	ZFP1	
RABGAP1L	RMND1	SEC63	SLC4A7	SPRED2	TBC1D16	TPRG1L	URGCP	ZFYVE26	
RABL3	RNF10	SECISBP2L	SLC6A6	SPTBN1	TBC1D22A	TPST1	URI1	ZFYVE27	
RAD51B	RNF13	SELT	SLC7A1	SPTBN4	TBC1D2B	TRA2B	USP10	ZHX2	
RAD51L1	RNF130	SEMA4B	SLC8A1	SPTSSA	TBCD	TRAF1	USP21	ZMIZ1-AS1	
RAI1	RNF138	SEMA5A	SLC8A1-AS1	SREBF1	TBL1XR1	TRAF3IP2	USP54	ZMYND11	
RALGPS2	RNF139	SENP6	SLMO2	SREK1	TCEANC2	TRAF5	USP6NL	ZNF12	
RALY	RNF149	SEPT11	SMAD1	SRGAP1	TCF12	TRAM1	UST	ZNF148	
RANBP2	RNF168	SERINC5	SMAD3	SRGAP2	TCF3	TRAPPC3	UTRN	ZNF155	
RANBP3	RNF170	SERPINB1	SMAD6	SRPK2	TCF7L2	TRAPPC9	UVRAG	ZNF219	
RANBP9	RNF185	SERPINB6	SMAD9	SRSF5	TDRD3	TRIM13	UXS1	ZNF318	
RAP1GAP2	RNF19A	SERPINE2	SMAP2	SS18	TEAD1	TRIM26	VAMP2	ZNF319	
RAP1GDS1	RNF216	SESN1	SMARCA2	SSBP3	TET2	TRIM32	VARS2	ZNF384	
RAP2A	RNGTT	SETBP1	SMARCA5	SSH1	TEX14	TRIM39	VEPH1	ZNF407	
RAPGEF1	RNU6-71P	SETD2	SMCHD1	SSH2	TFEB	TRIM52	VNN3	ZNF410	
RARRES3	RNU6-79P	SETD3	SMCR5	SSTR3	TGFBR3	TRIO	VOPP1	ZNF428	
RASAL2	ROR1	SEZ6	SMIM13	SSX2IP	THADA	TRUB1	VPS13B	ZNF43	
RASSF5	RORA	SF3A1	SMIM18	ST3GAL3	THAP3	TSC22D2	VTI1A	ZNF516	
RASSF6	RORA-AS1	SFT2D2	SMYD3	STAG1	THCAT158	TSEN15	VWA9	ZNF592	
RBFOX2	RPE	SFTPB	SND1	STARD13	THRA	TSHZ1	WAC	ZNF608	

Table 46: DESeq2 gene counts comparing insulinomas to NF-PNETs, shown adjusted p<0.01

Gene Name	Base mean	Log2(FC)	StdErr	P-value	Adj. p value	Type
ASB9	58.7785291	5.97914716	0.78248792	2.15E-14	4.22E-10	protein_coding
KCNH6	227.129783	-3.8424998	0.52100848	1.64E-13	1.61E-09	protein_coding
STEAP3	97.7088068	3.63509818	0.54000704	1.68E-11	8.22E-08	protein_coding
RP11-688I9.4	34.783824	5.99313936	0.88935677	1.60E-11	8.22E-08	lincRNA
IGFBP3	394.263393	-4.6020679	0.71013113	9.14E-11	3.58E-07	protein_coding
TMEM108	77.9752924	3.96921822	0.63615386	4.39E-10	1.44E-06	protein_coding
INS	3873.62804	5.61792844	0.91599978	8.62E-10	2.41E-06	protein_coding
PTPRT	438.198606	-4.2011918	0.69618246	1.59E-09	3.90E-06	protein_coding
GC	369.655517	-5.2110251	0.8822833	3.50E-09	6.24E-06	protein_coding
LRFN2	33.3812311	4.72546405	0.79857883	3.27E-09	6.24E-06	protein_coding
EGLN3	19.9063783	-4.3989622	0.7440059	3.37E-09	6.24E-06	protein_coding
USH1C	195.639278	-3.821927	0.66769299	1.04E-08	1.70E-05	protein_coding
CLVS2	54.5007265	5.3473157	0.94483971	1.52E-08	2.29E-05	protein_coding
LINC00514	32.3372131	2.97564646	0.5310019	2.10E-08	2.94E-05	lincRNA
SERPINA1	2452.24269	-4.6755561	0.84432921	3.07E-08	4.01E-05	protein_coding
PFKFB2	636.608039	2.96146794	0.54169278	4.58E-08	5.61E-05	protein_coding
IRX2	132.565288	-4.9456217	0.91273199	6.01E-08	6.93E-05	protein_coding
G6PC2	512.882551	3.57501034	0.66336561	7.08E-08	7.71E-05	protein_coding
DLK1	192.138198	4.82974234	0.90188647	8.55E-08	8.44E-05	protein_coding
CHRNA3	65.1102597	4.26227432	0.79610242	8.61E-08	8.44E-05	protein_coding
C1orf127	153.563532	3.26804616	0.61442499	1.04E-07	9.75E-05	protein_coding
MAFA	383.991905	4.37883456	0.8254382	1.13E-07	0.00010047	protein_coding
PLCH2	428.763651	3.32221746	0.62847444	1.25E-07	0.00010301	protein_coding
GLP1R	176.166913	4.20949897	0.79659142	1.26E-07	0.00010301	protein_coding
ARHGEF3	120.932001	1.89922993	0.36438037	1.87E-07	0.00014631	protein_coding
RP11-279F6.1	23.2386079	-4.3719285	0.84324781	2.16E-07	0.00016318	lincRNA
CLDN11	42.4325332	-3.6982188	0.72571396	3.47E-07	0.00025189	protein_coding
RP11-672L10.2	19.1946557	4.66418741	0.92029894	4.02E-07	0.00027156	processed_transcript
CYP4F3	170.827326	-4.4873021	0.88438691	3.90E-07	0.00027156	protein_coding
REG1P	205.8721	5.03937529	1.00904959	5.91E-07	0.00037362	pseudogene
GPR119	37.1078746	-5.0840213	1.01786772	5.89E-07	0.00037362	protein_coding
GCGR	122.594839	3.62607513	0.74261987	1.05E-06	0.00064063	protein_coding
CADM1	398.388771	2.02724342	0.41652783	1.13E-06	0.00067301	protein_coding
WNT9A	28.8235694	2.51471489	0.52566208	1.72E-06	0.00099121	protein_coding
HADH	196.350872	3.04689837	0.63882273	1.85E-06	0.00100534	protein_coding
MBD6	904.119291	1.17348267	0.24601433	1.84E-06	0.00100534	protein_coding
TTR	9938.96465	-3.8421499	0.80795586	1.98E-06	0.00104928	protein_coding
INS-IGF2	64.4972537	4.52035508	0.95804966	2.38E-06	0.00121581	protein_coding
GPR179	121.450732	-3.3092807	0.70188557	2.42E-06	0.00121581	protein_coding

FEV	51.0689515	-3.9088633	0.83160507	2.60E-06	0.00127006	protein_coding
GLT1D1	43.4482784	-4.0330175	0.85886809	2.66E-06	0.00127006	protein_coding
ENTPD3	111.018117	2.6623232	0.56828378	2.80E-06	0.00127717	protein_coding
GRIN2A	93.0548166	3.29251139	0.70279754	2.80E-06	0.00127717	protein_coding
CYP3A5	967.596399	-3.8374926	0.83256052	4.04E-06	0.00180024	protein_coding
PCDH7	249.175827	3.54014626	0.76931539	4.19E-06	0.00182555	protein_coding
CAMK2G	203.396013	-1.9716502	0.42939941	4.40E-06	0.0018292	protein_coding
ATN1	1294.42212	1.02887074	0.22394245	4.34E-06	0.0018292	protein_coding
PSG8	24.0518106	4.68409745	1.02098784	4.48E-06	0.0018292	protein_coding
WSCD2	122.264261	3.44605018	0.75389054	4.85E-06	0.0019416	protein_coding
MAMLD1	192.19514	-2.7541644	0.60549224	5.40E-06	0.0021168	protein_coding
RN7SKP255	65.2909452	1.81048195	0.39981529	5.95E-06	0.00228541	misc_RNA
LDLRAP1	274.189646	1.84979565	0.4100004	6.43E-06	0.00237859	protein_coding
RP1-80N2.3	17.2361665	-4.5244415	1.00262921	6.40E-06	0.00237859	lincRNA
POU6F2	81.2138777	-3.0873305	0.68510496	6.59E-06	0.00239394	protein_coding
TGFBR3	153.525866	2.52599189	0.56281284	7.18E-06	0.00256046	protein_coding
LRAT	21.1406601	-3.969785	0.88807776	7.82E-06	0.00270243	protein_coding
ABCB1	102.824497	-2.8204217	0.63110686	7.86E-06	0.00270243	protein_coding
HSPB6	109.628369	3.36418037	0.76629819	1.13E-05	0.00382784	protein_coding
ZBTB20-AS1	46.199341	1.99865756	0.45639516	1.19E-05	0.003957	antisense
RP11-216M21.1	10.772429	-3.8919283	0.90122148	1.57E-05	0.00513215	pseudogene
RP11-305F18.1	14.7541362	-4.0776263	0.94542225	1.61E-05	0.00517488	lincRNA
DRD1	10.6994897	-4.2176303	0.97923986	1.65E-05	0.00523088	protein_coding
UST	67.594665	3.28376566	0.7630908	1.68E-05	0.00523702	protein_coding
KBTBD11	147.292799	-2.7582431	0.64169853	1.72E-05	0.00527038	protein_coding
KCNH1	32.5011775	2.99189943	0.70057374	1.95E-05	0.00577292	protein_coding
KCNK17	128.931684	2.97697926	0.69806834	2.00E-05	0.00577292	protein_coding
PIGA	35.8706208	2.41804216	0.56679062	1.99E-05	0.00577292	protein_coding
ELAVL3	113.394976	-3.7397662	0.87581409	1.95E-05	0.00577292	protein_coding
AC003080.4	12.4445994	-3.9041786	0.92813208	2.59E-05	0.00726302	pseudogene
IGF2-AS	40.0866391	4.13079946	0.98160561	2.57E-05	0.00726302	antisense
SERPINE2	114.590118	-2.4055858	0.57562762	2.93E-05	0.00803286	protein_coding
RP11-688I9.2	23.7492818	4.21246903	1.00843113	2.95E-05	0.00803286	antisense
LSAMP-AS1	16.8134067	4.11093227	0.98767138	3.15E-05	0.00835042	antisense
SOGA2	301.679589	2.12089535	0.50956493	3.15E-05	0.00835042	protein_coding
MUC6	5667.44145	3.05891223	0.73548287	3.20E-05	0.00835228	protein_coding
CTD-2561B21.5	15.0645211	-3.8851184	0.93678524	3.36E-05	0.0086774	antisense
CYGB	40.545186	-2.702972	0.65300393	3.48E-05	0.00886886	protein_coding
CRYM	13.8223969	-4.0420839	0.98078826	3.77E-05	0.0094694	protein_coding
TDRG1	8.63900911	4.17337467	1.01540261	3.96E-05	0.00981443	antisense

Table 47: Common hypermethylated genes in insulinomas and NF-PNETs

Genes							
A4GALT	ADSL	BAIAP2	CCDC167	CSTB	DIRC2	ETHE1	GATA2
AAAS	ADSSL1	BCAR3	CCDC33	CSTF1	DLG1	ETS1	GATA2-AS1
AACS	AEBP2	BCAT1	CCDC88C	CTB-12A17.2	DLG5	ETV5	GATA4
AATK	AF146191.4	BCL11A	CCNDBP1	CTB-147N14.6	DMTF1	EXT2	GDF1
ABALON	AFG3L2	BCL2	CCNF	CTB-43E15.2	DNAJA4	EYA4	GFOD1
ABCA3	AGAP3	BCL2L1	CD72	CTBP1	DNAJB6	EZR	GID4
ABHD11	AGFG1	BCL2L2-PABPN1	CDH15	CTC-338M12.4	DNAJC10	EZR-AS1	GIPR
ABHD8	AHCYL1	BCL6	CDH22	CTC-467M3.1	DPCD	FADS6	GJA3
ABR	AHRR	BEST3	CDH4	CTC-467M3.3	DRAM2	FAM110A	GLI4
AC002116.7	AK4	BRF1	CDK18	CTD-2026K11.3	DSC2	FAM110B	GLIS3
AC005076.5	AK7	BRWD1	CDK2	CTD-2105E13.16	DSCAS	FAM129B	GLRX
AC005237.4	AKT1	BRWD1-AS2	CDK2AP1	CTD-2132N18.3	DTWD2	FAM149B1	GM2A
AC006026.13	AL022344.7	BSN	CDK2AP2	CTD-2194D22.3	DUOX1	FAM159A	GNA11
AC006372.4	AL512506.1	BSN-AS2	CDK5R1	CTD-2213F21.2	DUOXA1	FAM171A2	GNA12
AC007326.9	ALDH1A2	BTRC	CDKN1B	CTD-2291D10.4	DUSP1	FAM171B	GNB2L1
AC007405.4	ALDOA	C11orf49	CELSR1	CTD-2323K18.1	DUSP18	FAM179A	GNE
AC007405.6	ALOX5AP	C14orf166	CEP112	CTD-2330J20.2	DYNC1H1	FAM20B	GNG12
AC008060.7	ANKIB1	C14orf169	CEP78	CTD-2515O10.5	DYRK2	FAM214B	GNG12-AS1
AC013733.3	ANKRD11	C16orf70	CEPT1	CTD-2532D12.4	DZIP3	FAM221A	GNG4
AC016747.3	ANKRD20A8P	C1orf198	CERS1	CTD-2545M3.6	E2F5	FAM222A	GNG7
AC017101.10	ANKRD30BL	C1orf210	CHAD	CTD-2555C10.3	EBF3	FAM64A	GNL3
AC026401.1	ANO10	C1S	CHCHD7	CTD-2600O9.1	ECD	FAM81A	GNPTAB
AC068533.7	ANO7	C2CD4D	CHID1	CTD-2623N2.5	EEF2	FANK1	GNRHR2
AC073934.6	AP003176.1	C2orf71	CHKB	CTD-3131K8.2	EFCC1	FASTKD5	GPALPP1
AC090340.1	AP1S1	C2orf74	CHST1	CTD-3138B18.4	EFHD2	FBRSL1	GPN1
AC092171.4	AP2A2	C5orf51	CHST10	CTDNEP1	EFNA3	FBXL18	GPR124
AC092687.4	APBB2	C8orf76	CHST12	CYB5D2	EGR3	FBXL19	GPR157
AC092755.4	ARFGAP2	C9orf116	CKAP5	CYFIP2	EHBP1L1	FBXL22	GPR161
AC098820.3	ARG2	C9orf163	CLASP2	CYHR1	EIF2AK2	FBXO21	GPR19
AC099684.1	ARHGAP9	CABLES1	CLASRP	DCAF5	EIF2AK4	FBXW8	GPR56
AC104667.3	ARHGEF10L	CACNA1H	CLDN6	DCTN1	EIF2B5	FFAR1	GPR75
AC105402.4	ARHGEF26	CACNA2D2	CLIP3	DDA1	EIF2S1	FGF12	GPR75-ASB3
AC110781.3	ARHGEF26-AS1	CACNB2	CLN8	DDX31	EIF3F	FGFR2	GPRC5C
AC135048.13	ARHGEF3	CADM1	CLTA	DEGS1	ELAVL4	FGFRL1	GPSM1
AC137932.6	ARID1B	CAMK1D	CMC1	DENND1A	ELF3	FHIT	GRAMD3
AC140912.1	ARPC2	CAMKK1	CMIP	DENND1B	ELFN2	FIBCD1	GRIK3
ACAA1	ASXL2	CAMSAP1	CMPK2	DENND1C	ELMO3	FLCN	GS1-24F4.2
ACACA	ATG16L1	CAPN10	CNIH4	DENND5A	ELMOD1	FLJ20021	GSE1
ACADM	ATOH8	CAPN10-AS1	CNKSR1	DGCR5	ELP5	FNDC9	GTF2A2
ACBD3	ATP2B2	CARS2	CNTN5	DGCR6	EML2	FOXCUT	GTF2H5
ACLY	ATP5SL	CASC1	COL18A1	DGKG	ENPP1	FOXP1	H3F3A
ACOT7	ATP6V0D1	CASP9	COMMD6	DHX32	EPB41	FOXP4	HACE1
ACSF2	ATP6V1D	CASZ1	COMMD9	DHX36	EPB42	FPGS	HCN2
ACTL6A	ATP7B	CBX7	COMT	DHX37	EPHA5	FRAT1	HCP5
ACVR2A	ATP8B2	CCAR2	CORO7-PAM16	DHX38	EPHA5-AS1	FRG1	HDAC4
ADCK3	ATPAF2	CCDC121	CPEB2	DICER1	EPHB4	FYN	HDLBP
ADCY4	ATXN1	CCDC122	CPXM1	DICER1-AS1	ERC2	FZD2	HEATR4
ADCYAP1	ATXN10	CCDC142	CRAMP1L	DIO3	ERCC1	FZD4	HEATR5A
ADD2	AURKA	CCDC148	CRCP	DIO3OS	ESPN	GABRB1	HECTD4
ADM	AZIN1	CCDC151	CRYBB2P1	DIP2C	ESRP2	GALNT2	HERC2

Genes						
HES2	KCNIP3	LMTK2	MIER1	NDUFB4	PC	POLH
HIBADH	KCNQ1	LMX1A	MIR1247	NDUFV3	PCDHGA1	PPARGC1B
HIC1	KCNQ1-AS1	LPCAT3	MIR141	NEDD4L	PCDHGA10	PPIL2
HIP1R	KCTD1	LPGAT1	MIR200C	NEO1	PCDHGA11	PPL
HIST1H2BD	KIAA0430	LRP5	MIR330	NES	PCDHGA12	PPP1R8
HIST1H3E	KIAA0930	LRRC14	MIR3619	NEUROG1	PCDHGA2	PPP1R9B
HIST1H3I	KIAA1161	LRRC23	MIR6090	NFATC1	PCDHGA3	PPP2R5C
HKDC1	KIAA1522	LRRC24	MIR663A	NFKB2	PCDHGA4	PPP3CA
HLX	KIAA1524	LRRC4B	MIR663AHG	NIN	PCDHGA5	PRDM11
HNF4A	KIAA1614	LRRFIP1	MIR663B	NIPAL2	PCDHGA6	PRDM6
HNRNPF	KIF25	LSM3	MIR744	NKX2-2	PCDHGA7	PRDM8
HOOK2	KIF5C	LTBP3	MIR7845	NKX2-2-AS1	PCDHGA8	PRKAG2
HOXA-AS2	KIFC2	LTBP4	MIRLET7BHG	NKX2-8	PCDHGA9	PRKAR1B
HOXA10	KIFC3	LTF	MKNK1	NKX6-1	PCDHGB1	PROSER3
HOXA10-HOXA9	KLC4	LUZP1	MKNK2	NLRP3	PCDHGB2	PRR35
HOXA3	KLF7	LYRM5	MLLT1	NMRK2	PCDHGB3	PRUNE2
HOXA9	KLHDC4	LYST	MLLT10	NNT	PCDHGB4	PSMG4
HOXB-AS3	KLHL21	MAD1L1	MNX1	NNT-AS1	PCDHGB5	PTCH2
HOXB3	KLHL24	MAEA	MNX1-AS2	NOL11	PCDHGB6	PTGES3P2
HOXB6	KREMEN2	MAGEF1	MOB3C	NOP58	PCDHGB7	PTPN12
HOXD11	KRIT1	MAML3	MPC1	NPAS3	PCDHGC3	PTPN9
HP1BP3	KRTAP5-AS1	MAP1LC3A	MPP5	NRIP1	PCDHGC4	PTPRN2
HPCAL1	LACC1	MAP2K4	MRGBP	NRSN2-AS1	PCMT1	PVRL2
HRASLS5	LAMC2	MAP4	MROH6	NTHL1	PDCD6	PWWP2B
HS6ST1	LBX2	MAPK15	MRPL2	NTM	PDE4A	PXDC1
HSPB6	LBX2-AS1	MAPK8IP2	MRPS2	NTN4	PDZD2	QRSL1
HSPBAP1	LDLRAD4	MAPK8IP3	MRPS9	NUDT21	PEMT	QSOX2
HVCN1	LDLRAP1	MAPKAPK5	MSI1	NUDT7	PEX10	QTRT1
ICMT	LEKR1	MAPKAPK5-AS1	MSRA	NUFIP1	PEX11B	R3HCC1L
IGF1R	LHPP	MAPRE2	MSX1	NUP43	PGBD2	RAB11FIP3
IGF2BP1	LIMD1-AS1	MAPT	MTA3	OBFC1	PGLS	RAB17
IGF2BP2	LINC00298	MAPT-AS1	MTMR14	OGFOD1	PHF13	RAB31
IGSF21	LINC00299	MAPT-IT1	MTPN	ONECUT2	PHLDA3	RAB43
IGSF9B	LINC00620	MARS	MYD88	OTX1	PHRF1	RAB5C
IL20RA	LINC00856	MAST1	MYO10	OVCH2	PI4KB	RAB8A
ING2	LINC00886	MAT2B	MYO18A	PABPC4	PIEZO1	RABGGTB
INPP5A	LINC00910	MAU2	MYO1E	PABPN1	PIP5K1C	RAD17
IRF4	LINC00960	MBNL2	MYRIP	PACS1	pk	RAD51AP2
IRX4	LINC01096	MBOAT2	MYT1L	PACS2	PKDCC	RAI1
IRX6	LINC01097	MCF2L	N4BP2L1	PAM16	PKP4	RALGAPA2
IST1	LINC01184	MCM10	NA	PAOX	PLAG1	RANBP2
ISY1-RAB43	LINC01237	MCOLN2	NA	PAPLN	PLCH2	RAP1B
ITM2C	LINC01336	MED27	NAP1L4	PARD3	PLEKHA1	RAP1GAP2
ITPK1	LINC01534	MEF2BNB	NAPA	PARD6G	PLEKHA2	RARRES2P1
ITPKB	LINC01579	MEF2BNB-MEF2B	NAPA-AS1	PARK2	PLEKHM2	RASGRF1
ITPR2	LL21NC02-21A1.1	MEIS2	NAT8L	PARN	PLEKHM3	RASSF5
KAZN	LMF1	METTL16	NAT9	PARP11	PLOD1	RBCK1
KB-1507C5.2	LMO3	MGRN1	NCOR2	PARP14	PMEL	RCBTB1
KCNC2	LMO7	MGST1	NDE1	PBRM1	PMEPA1	RCC1
KCNG3	LMO7-AS1	MICA	NDUFAF4	PBX3	PMPCA	RCOR1

Genes				
RCOR2	RP11-377G16.2	RP11-973D8.4	SH3RF3	SPATA4
RCSD1	RP11-383H13.1	RP13-452N2.1	SH3TC1	SPDYA
REC8	RP11-386M24.3	RP3-455J7.4	SHANK2	SPIB
RFESD	RP11-386M24.6	RP3-461F17.3	SHANK3	SPSB1
RFX1	RP11-396C23.4	RP3-467K16.4	SHB	SPSB2
RFX5	RP11-399E6.1	RP4-569M23.5	SHF	SPTBN1
RFXANK	RP11-399K21.11	RP4-613B23.1	SIL1	SSBP4
RGL2	RP11-404E16.1	RP4-647C14.2	SIPA1	ST5
RGL3	RP11-415A20.1	RP4-683L5.1	SIPA1L1	ST8SIA1
RGS9	RP11-443B7.1	RP4-791C19.1	SIPA1L3	STARD13
RHOBTB3	RP11-445F12.1	RP5-1033H22.2	SLC12A2	SUCLG1
RIC1	RP11-459E5.1	RP5-1063M23.1	SLC12A9	SUN1
RIN2	RP11-45A12.2	RP5-881L22.5	SLC13A5	SUV420H2
RMDN3	RP11-45M22.4	RP5-930J4.2	SLC17A7	SYK
RMI2	RP11-465N4.4	RP5-956O18.2	SLC1A2	SYN3
RN7SKP114	RP11-472N13.2	RPH3AL	SLC1A5	SYNGAP1
RNF135	RP11-473C18.3	RPL14	SLC23A2	SYNPR
RNF152	RP11-482M8.1	RPL37A	SLC25A28	SYT7
RNF220	RP11-484D2.2	RPS10L	SLC25A48	SYTL2
RP1-168L15.5	RP11-493E12.3	RPS12	SLC2A2	TACC2
RP1-37E16.12	RP11-501C14.5	RPS27L	SLC35E4	TACSTD2
RP1-4G17.5	RP11-510N19.5	RPS6KA2	SLC38A1	TADA2A
RP1-59D14.5	RP11-519G16.2	RRP1	SLC43A2	TAF4
RP1-63G5.5	RP11-521D12.5	RSAD2	SLC45A4	TAF9
RP1-91J24.3	RP11-540D14.8	RTN4IP1	SLC47A1	TBX2
RP11-1070N10.3	RP11-544M22.13	RTN4RL1	SLC4A10	TBX2-AS1
RP11-122D10.1	RP11-552D8.1	SACM1L	SLC6A1	TCF20
RP11-126K1.6	RP11-566K11.4	SALL1	SLC6A20	TCF7
RP11-126K1.9	RP11-566K19.6	SBF2	SLC7A8	TEKT2
RP11-149I9.2	RP11-5G9.5	SCD5	SLC9A3R1	TESK1
RP11-158I13.2	RP11-60L3.6	SCGB1B2P	SLFN12L	THAP7
RP11-162G10.5	RP11-613M10.9	SCMH1	SLITRK3	THAP7-AS1
RP11-16E12.1	RP11-662I13.3	SDC4	SMAP2	TIMP3
RP11-176H8.1	RP11-664D1.1	SDCCAG3	SNAP47	TIPIN
RP11-207C16.4	RP11-672L10.2	SDK1	SND1	TJP2
RP11-219B4.7	RP11-672L10.3	SEC16A	SNED1	TLE6
RP11-227H15.4	RP11-6N17.4	SEL1L	SNHG3	TLX2
RP11-26F2.2	RP11-713D19.1	SELM	SNORD101	TMCC2
RP11-26L20.3	RP11-730K11.1	SEPHS1P6	SNORD45C	TMCO3
RP11-278A23.2	RP11-736K20.6	SEPHS2	SNTA1	TMEM104
RP11-27K13.3	RP11-748C4.1	SERAC1	SNUPN	TMEM131
RP11-288I21.1	RP11-756J15.3	SERHL	SNX27	TMEM143
RP11-29G8.3	RP11-762H8.3	SFSWAP	SNX29	TMEM178A
RP11-307C19.2	RP11-779O18.3	SGK223	SOCS1	TMEM183A
RP11-321C24.1	RP11-806H10.4	SGSM2	SOCS3	TMEM201
RP11-351I24.1	RP11-809N15.3	SH2D4A	SORBS2	TMEM219
RP11-353N14.5	RP11-83J21.3	SH2D5	SORL1	TMEM241
RP11-358B23.1	RP11-848P1.2	SH3BP5	SORT1	TMEM87B
RP11-358M14.2	RP11-85A1.3	SH3GLB2	SOX12	TMEM9B
RP11-367N14.2	RP11-934B9.3	SH3PXD2B	SP2	TMEM9B-AS1

Genes				
TMTC4	TUB	VOPP1	ZCCHC10	ZNF567
TNFRSF12A	TUBB3	VTI1A	ZCCHC14	ZNF575
TOLLIP	TXNL4B	WBP2NL	ZCCHC24	ZNF639
TOP1MT	U2	WDR25	ZDHHC20	ZNF663P
TOR1B	U47924.27	WDR4	ZEB2	ZNF669
TOX2	U47924.29	WDR78	ZFAND4	ZNF730
TPGS1	U47924.6	WFIKKN2	ZFP1	ZNF784
TRAF3	UBA52P5	WRB	ZHX1-C8orf76	ZNF821
TRAPPC12	UBAP2	WT1	ZIC2	ZNF846
TRAPPC9	UBE2F	WT1-AS	ZMIZ1	ZNF852
TRIM3	UBE2F-SCLY	WT1-AS_1	ZMYND8	ZNF90
TRIOBP	UBE3C	WT1-AS_2	ZNF212	ZSWIM4
TRPA1	UBOX5	WT1-AS_3	ZNF215	ZXDC
TRPM4	UBTF	WWOX	ZNF248	ZZEF1
TSC2	UCHL3	XKR5	ZNF281	
TSPAN17	UHRF1	XPC	ZNF354A	
TSPAN18	ULK1	XPO4	ZNF397	
TSPAN3	URAD	XPO5	ZNF45	
TSSC1	URAHP	XXbac-BPG294E21.9	ZNF503	
TSTD1	USP3-AS1	YWHAQ	ZNF503-AS2	
TTC17	UTRN	ZBTB38	ZNF521	
TTC21B	VAC14	ZBTB47	ZNF544	
TTC31	VAX2	ZBTB9	ZNF550	
TTF2	Vax2os1_3	ZC3H3	ZNF561	
TTR	VGLL2	ZC3H7B	ZNF561-AS1	

8.2 R Code

##Load libraries##

```
library(missMethyl)
library(ChAMP)
library(readr)
library(ggplot2)
```

##Load samples##

```
amyLoad <- champ.load("/Users/katherineenglish/Dropbox/informatics/Combined oxBS data only", arraytype="EPIC")
```

##Normalise samples##

```
amyNorm<-champ.norm(beta=amyLoad$beta, arraytype="EPIC")
champ.QC(beta=amyLoad$beta, pheno=amyLoad$pd$Sample_Group)
```

##SVD samples##

```
amyLoadpda <- as.data.frame(amyLoad$pd)
amyLoadpda$Slide <- as.factor(amyLoadpda$Slide)
amyNormdfa <- as.data.frame(amyNorm)
covariate_sheet_total3 <- read_csv("~/Dropbox/informatics/covariate sheet total3.csv")
jointdataseta <- merge(amyLoadpda, covariate_sheet_total3, by = 'Sample_Name', all.x= TRUE)
```

```
amySVD <- champ.SVD(beta=amyNormdfa, pd=jointdataseta, RGEffect=TRUE)

##GOregion analysis##

group <- factor(jointdataseta$Sample_Group,levels=c("B","O"))
id <- factor(jointdataseta$ID)
design <- model.matrix(~id + group)
design

DMRcate_manual_run = function(beta_matrix,
                        pheno,
                        arraytype,
                        dist = 2,
                        mafcut = 0.05,
                        fdr = 0.05,
                        lambda = 1000,
                        C = 2)
{

  require(DMRcate)
  myMs <- logit2(beta_matrix)
  myMs <- rmSNPandCH(myMs, dist = dist, mafcut = mafcut)
  design <- model.matrix(~ pheno)

  if(arraytype == "450K"){
    myannotation <- cpg.annotate(datatype = "array",
                         fdr = 0.05,
                         myMs,
                         design = design,
                         coef = ncol(design),
                         analysis.type = "differential",
                         annotation = c(array = "IlluminaHumanMethylation450k",
                                        annotation = "ilmn12.hg19"),
                         what = "M")
  } else {
    myannotation <- cpg.annotate(datatype = "array",
                         fdr = 0.05,
                         myMs,
                         design = design,
                         coef = ncol(design),
                         analysis.type = "differential",
                         annotation = c(array = "IlluminaHumanMethylationEPIC",
                                        annotation = "ilm10b4.hg19"),
                         what = "M")
  }

  dmrcoutput <- dmrcate(myannotation, lambda = 1000, C = 2, betacutoff=0.05)
  DMR <- as.data.frame(extractRanges(dmrcoutput, genome = "hg19"))
  rownames(DMR) <- paste("DMR", 1:nrow(DMR), sep="_")

  message("DMRcate detected ", nrow(DMR)," DMRs with mafcut as = ", mafcut, ".")
  if(nrow(DMR) == 0)
    stop("No DMR detected.")

  return(DMR)
}

amyNorm1 <- as.matrix(amyNorm)

aDMR1s <- DMRcate_manual_run(amyNorm1, amyLoad$pd$Sample_Group, "EPIC")
```

If the code in red is also run (above),this will return the number of DMRs, then this section of code must be re-run after removing the red, as it will result in an error when trying to run goregion below. The code for the solution to this error was found at the website:
https://github.com/YuanTian1991/ChAMP/issues/18#issuecomment-809842467. And is shown in italics.

```
DMRcate_manual_run = function(beta_matrix,
                        pheno,
                        arraytype,
                        dist = 2,
                        mafcut = 0.05,
                        fdr = 0.05,
                        lambda = 1000,
                        C = 2)
{

    require(DMRcate)
    myMs <- logit2(beta_matrix)
    myMs <- rmSNPandCH(myMs, dist = dist, mafcut = mafcut)
    design <- model.matrix(~ pheno)

    if(arraytype == "450K"){
        myannotation <- cpg.annotate(datatype = "array",
                            fdr = 0.05,
                            myMs,
                            design = design,
                            coef = ncol(design),
                            analysis.type = "differential",
                            annotation = c(array = "IlluminaHumanMethylation450k",
                                        annotation = "ilmn12.hg19"),
                            what = "M")
    } else {
        myannotation <- cpg.annotate(datatype = "array",
                            fdr = 0.05,
                            myMs,
                            design = design,
                            coef = ncol(design),
                            analysis.type = "differential",
                            annotation = c(array = "IlluminaHumanMethylationEPIC",
                                        annotation = "ilm10b4.hg19"),
                            what = "M")
    }


    dmrcoutput <- dmrcate(myannotation, lambda = lambda, C = 2, betacutoff=0.05)

    DMR <- extractRanges(dmrcoutput, genome = "hg19")

}

aDMR2s <- DMRcate_manual_run(amyNorm1, pheno=amyLoad$pd$Sample_Group, "EPIC")

agst.region <- goregion(aDMR2s, all.cpg=rownames(amyNorm),
                    collection="KEGG", array.type="EPIC", plot.bias=TRUE)

write.csv(agst.region, "aoxBS GSEA DMRcate.csv")
```

The code for the solution to this error was found at the website:
https://github.com/YuanTian1991/ChAMP/issues/18#issuecomment-809842467. And is shown in italics.

##GOmeth analysis##

```
amyNormM <- logit2(amyNorm)
amyDMPM <- champ.DMP(beta=amyNormM, pheno= amyLoad$pd$Sample_Group,
arraytype="EPIC")
amyDMPM1 <- amyDMPM[[1]][amyDMPM[[1]]$deltaBeta>0,]
amyDMPM2 <- as.data.frame(amyDMPM1)
amyDMPM3 <- subset(amyDMPM2, amyDMPM2$adj.P.Val <0.01)
amyDMPM4 <- subset(amyDMPM3, amyDMPM3$deltaBeta >1)

aGSEAM1c <- gometh(sig.cpg=rownames(amyDMPM4), all.cpg=rownames(amyNormM), collection="KEGG",
          plot.bias=TRUE, sig.genes=TRUE)

write.csv(aGSEAM1c, "aGSEAM1c.csv")
```

##To make a scatter plot of hmc##

```
amyDMPM5 <- champ.DMP(beta=amyNorm, pheno = amyLoad$pd$Sample_Group,
arraytype="EPIC")

amergescatter <- merge(x=amyDMPM2, y=amyDMPM5, by='row.names', all.x=TRUE)

amergescatter2 <- subset(amergescatter, !is.na(t))

ascatter <- ggplot(amergescatter2, aes(x=N_to_T.deltaBeta, y=deltaBeta))
ascatter + geom_point(shape=18)
```

##To make the GSEA figures##

```
library(forcats)

aGSEAM1csubset2 <- subset(aGSEAM1c, aGSEAM1c$FDR <0.05)

aGSEAM1csubset2 <- subset(aGSEAM1csubset, aGSEAM1csubset$FDR <0.05)

aGSEAM1csubset2 <- subset(aGSEAM1csubset, aGSEAM1csubset$FDR <0.019)

aGSEAM1csubset2 %>%
       mutate(Description = fct_reorder(Description, desc(FDR))) %>%
        ggplot(aes(x=DE, y=Description, size = ave)) +
        geom_point(alpha=0.8, color="red") +
        scale_size(range = c(1, 10)) + theme_bw() + theme(axis.text = element_text(face="bold", colour="black"))

aGSEAM1csubset3 <- subset(aGSEAM1csubset, aGSEAM1csubset$FDR <0.037)
aGSEAM1csubset3 <- subset(aGSEAM1csubset3, aGSEAM1csubset3$FDR >0.029)

aGSEAM1csubset3 %>%
       mutate(Description = fct_reorder(Description, desc(FDR))) %>%
        ggplot(aes(x=DE, y=Description, size = ave)) +
        geom_point(alpha=0.8, color="red") +
        scale_size(range = c(1, 10)) + theme_bw() + theme(axis.text = element_text(face="bold", colour="black"))
```

The same code is run as above, for PNET samples. In addition:

##To make a scatter plot of hmc##

```
bmyDMPM5 <- bmyDMPM4
bmyDMPM5$CpG <- row.names(bmyDMPM5)

amyDMPM9 <- amyDMPM4
```

```
amyDMPM9$CpG <- row.names(amyDMPM9)

amyNormdfa2 <- amyNormdfa
amyNormdfa2$INS13N <- amyNormdfa2$INS13N_BS - amyNormdfa2$INS13N_ox
amyNormdfa2$NF36N <- amyNormdfa2$NF36N_BS - amyNormdfa2$NF36N_ox
amyNormdfa2$NF35N <- amyNormdfa2$NF35N_BS - amyNormdfa2$NF35N_ox
amyNormdfa2$INS9N <- amyNormdfa2$INS9N_BS - amyNormdfa2$INS9N_ox
amyNormdfa2$`11893` <- amyNormdfa2$`11893_BS` - amyNormdfa2$`11893_ox`
amyNormdfa2$`11969` <- amyNormdfa2$`11969_BS` - amyNormdfa2$`11969_ox`
amyNormdfa2$`12002` <- amyNormdfa2$`12002_BS` - amyNormdfa2$`12002_ox`

amyNormdfa3 <- amyNormdfa2[c(15:21)]

amyNormdfa4 <- as.matrix(amyNormdfa3)

heat2 <- merge(amyNormdfa4, amyDMPM9, by='row.names', all.Y=TRUE)
row.names(heat2) <- heat2$Row.names
heat4 <- subset(heat2, cgi %in% c('shore','island','shelf'))
heat3 <- heat4[c(2:8)]
heat5 <- as.matrix(heat3)
heatmap(heat5)

bmyNormdfa2 <- bmyNormdfa
bmyNormdfa2$INS13T <- bmyNormdfa2$INS13T_BS - bmyNormdfa2$INS13T_ox
bmyNormdfa2$INS9T <- bmyNormdfa2$INS9T_BS - bmyNormdfa2$INS9T_ox
bmyNormdfa2$INS23T <- bmyNormdfa2$INS23_BS - bmyNormdfa2$INS23_ox
bmyNormdfa2$INS27T <- bmyNormdfa2$INS27_BS - bmyNormdfa2$INS27_ox
bmyNormdfa2$NF36T <- bmyNormdfa2$NF36T_BS - bmyNormdfa2$NF36T_ox
bmyNormdfa2$NF35T <- bmyNormdfa2$NF35T_BS - bmyNormdfa2$NF35T_ox
bmyNormdfa2$NF24.2T <- bmyNormdfa2$NF24.2_BS - bmyNormdfa2$NF24.2_ox
bmyNormdfa2$NF34T <- bmyNormdfa2$NF34_BS - bmyNormdfa2$NF34_ox
bmyNormdfa2$`11947T` <- bmyNormdfa2$`11947_BS` - bmyNormdfa2$`11947_ox`
sheat2 <- merge(bmyNormdfa2, bmyDMPM5, by='row.names', all.Y=TRUE)
row.names(sheat2) <- sheat2$Row.names
sheat4 <- subset(sheat2, cgi %in% c('shore','island','shelf'))
sheat3 <- sheat4[c(20:28)]
sheat5 <- as.matrix(sheat3)
heatmap(sheat5)

library(reshape)
dfddd <- melt(heat5)
colnames(dfddd) <- c("x", "y","value")
ggplot(dfddd, aes(x = x, y = y, fill = value)) +
    geom_tile()
```

This output gives you a continuous scale for the heatmap – in the case for Normals its from 0 to 0.6 – compared to below. Zero is light yellow and 0.6 is dark red.

Repeated for tumours. 0 to 0.5 (dark blue to light blue i.e. light yellow to dark red).

To determine how many genes are interrogated by the methylationEPIC array##

```
View(EPIC.manifest.hg19)
r <-unique(EPIC.manifest.hg19$gene_HGNC)
write.csv(r, "r.csv")
ttrrc <- read.table("~/um12/ttrrc.csv", quote="\"", comment.char="")
View(ttrrc)
hgnc_complete_set <- read.delim("~/Downloads/hgnc_complete_set.txt")
View(hgnc_complete_set)
names(ttrrc)[names(ttrrc) == 'V1'] <- 'symbol'
ttrrc$try <- 1
merry2 <- merge(x=ttrrc, y=hgnc_complete_set, by= 'symbol', all.y=TRUE)
```

```
merry2s <- subset(merry2, !is.na(try))
table(merry2$try)
table(merry2s$locus_group)
```

To determine which CpGs are methylated in PNETs, which are actually hydroxymethylated in normal tissue##
R code:
##Hypermethylated DMPs in PNETs compared to normal, merged to 5'hmc in normal##

```
acmergec <- merge(x=cmyDMPM4, y=amyDMPM4, by='row.names', all.x=TRUE)
acmerge2c <- subset(acmergec, !is.na(t))
rownames(acmerge2c) <- acmerge2c$Row.names
acmerge3c <- acmerge2c[c(2:21)]
acSigCpGsc <- as.character(acmerge2c$Row.names)
acGSEAc <- gometh(sig.cpg=acSigCpGsc, all.cpg=rownames(cmyNorm), collection="KEGG",
                  plot.bias=TRUE, sig.genes=TRUE)
write.csv(acGSEAc, "acGSEAc.csv")
```

##Hypomethylated DMPs in PNETs compared to normal, merged to 5'hmc in normal##

```
cmyDMPM5 <- subset(cmyDMPM3, cmyDMPM3$N_to_T.deltaBeta <0)
acmergecdown <- merge(x=cmyDMPM5, y=amyDMPM4, by='row.names', all.x=TRUE)
acmerge2cdown <- subset(acmergecdown, !is.na(t))

##How to find the probe features of the CpGs on the methylationEPIC array##

data(probe.features.epic)
View(probe.features)
probe <- probe.features[c(6:9)]
dddd <- merge(ggtt, probe, by='row.names', all.x=TRUE)
dddd <- dddd[-c(45:79)]
table(cmyDMPM3$N_to_T.gene)

NT2$mean <- apply(NT2[,3:9],1,mean)
NT5 <- subset(NT2, mean >0.7)

NT5 <- subset(NT2, mean >0.7)

yty <- merge(ff22, NT5, by='ff', all.x=TRUE)
yty2 <- subset(yty, !is.na(`Ensemble ID`))
```

##How to find unique values in a data frame##

```
df2 <- as.vector(as.matrix(sig_hmc_genes))
ff <- unique(df2)
ff2 <- as.data.frame(ff)
```

Gene symbols downloaded form: https://www.biotools.fr/human/ensembl_symbol_converter

Copied them over and merged them together.

```
ff2$extra <- 1

names(fg2)[names(fg2) == 'fg'] <- 'ff'
fg3 <- subset(fg2, !is.na(ff))
```

##To get column names and to get rownames from the dataset & then delete the variables##

```
colnames(FULL) <- FULL[1,]
FULL <- FULL[-1, ]
row.names(FULL) <- FULL$Row.names
FULL2 <- FULL
FULL2 <- FULL2[-1]
FULL2 <- FULL2[-1]
```

##RNA seq code##
DE2seq output was uploaded into R and Goseq analysis completed.

```
Library(goseq)

supportedOrganisms()[supportedOrganisms()$Genome=="hg19",]

rr <- `Galaxy747.[DESeq2_result_file_on_data_732,_data_719,_and_others]`

DE_genes <- as.integer(rr$V7 <0.05)

names(DE_genes) <- rownames(rr)

DE_genes <- DE_genes[complete.cases(DE_genes)]

pwf=nullp(DE_genes,"hg19","ensGene")

GO.wall=goseq(pwf,"hg19","ensGene")

head(GO.wall)

enriched=GO.wall$category[p.adjust(GO.wall$over_represented_pvalue, method="BH")<.05]

library(GO.db)
  for(go in enriched[1:10]){
    print(GOTERM[[go]])
      cat("--------------------------------------\n")
      }
```

##To do KEGG analysis##

```
en2eg=as.list(org.Hs.egENSEMBL2EG)
eg2kegg=as.list(org.Hs.egPATH)
grepKEGG=function(id,mapkeys)
kegg=lapply(en2eg,grepKEGG,eg2kegg)
pwf=nullp(DE_genes,"hg19","ensGene")
KEGG=goseq(pwf,gene2cat=kegg)
head(KEGG)
```

Can't annotate KEGG pathways in R – therefore saved and manually added names to KEGG pathways after subsetting by significance.

```
KEGG2 <- KEGG
KEGG2 <- subset(KEGG2, KEGG2$over_represented_pvalue <0.05)
write.csv(KEGG2, "KEGG2 insulinomavs control RNA.csv"

library(tidyverse)
library(ggplot2)
library(forcats)
KEGG2.insulinomavs.control.RNA %>%
        mutate(Name = fct_reorder(Name, desc(over_represented_pvalue))) %>%
        ggplot(aes(x=numInCat, y=Name, size = Ave)) +
        geom_point(alpha=0.8, color="red") +
        scale_size(range = c(1, 10)) + theme_bw() + theme(axis.text = element_text(face="bold", colour="black"))
```