Probing positive selection in the human male germline



R. Spencer Tong

Green Templeton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2024

Declarations

I declare that the work within this thesis is my own unless otherwise stated.

Acknowledgements

I first acknowledge my supervisors, Professors Anne Goriely and Marella de Bruijn. Anne – thank you for your supervision; the rigour with which you approach science will have lasting effects on my career. Marella – thank you for your perspective and commitment to my well-being, which ensured that I continued the journey.

I would also like to thank my colleagues in the Clinical Genetics group, but especially Professor Steve Twigg and Drs. Geoffrey Maher and Eduardo Calpena for their time and generosity.

Oxford and the DPhil would not have been the same without the day-to-day comedy and support of my friends at the WIMM. Thanks are owed to Drs. Ali Preston, Joe Frost, Becky Tooze, Shiyan Tang, Maria Neofytu, Dale Maxwell, Chloe Tubman, Julius Bannister, Malitha Ratnaweera, Rebecca Roddan, Sanja Brolih, and Professor Hal Drakesmith. I look forward to our next trip to the White Hart.

To my people – Drs. Mandy Tseng, Jamie Shenk, David Grant, Richard Surgenor, Martha Kirby, Viorel Tataru, Jim Chesterfield, and Alex Bluestone. I'll miss biking through the city together, trudging back from Iffley, and sitting in the gardens at GTC.

To Mom and Dad – everything that I have been able to achieve has been made possible because of your sacrifices and love. I will never be able to truly show you how much I appreciate what you have done for me and what you continue to do for me. I will continue to reach because I stand on your shoulders.

To Nathan – your belief in me has shown me how to believe in myself. I cherish our relationship and shared worldview.

To Jess – Oxford brought us to together, and for that I will always be grateful. As I finish writing, I am overwhelmed by the partnership that you've shown me during this chapter. I thank you for your patience on my worst days and the hope that you give me for the future. Thank you for the birthdays spent in the lab, for putting up with work on vacation, and for dreaming with me. I am so grateful that we get to do this life together. I love you.

Abstract

Probing positive selection in the human male germline

R. Spencer Tong
Green Templeton College, University of Oxford

Michaelmas Term 2024

De novo mutations (DNMs) are a significant contributor to genetic disease, accounting for over 40% of developmental disorders (DDs). The majority of DNMs are of paternal origin, and the number of transmitted DNMs increases linearly with paternal age. For a subset of DNMs associated with a defined group of dominant disorders, the relationship is nonlinear. Mutations can occur up to 1000-fold above the genome-wide rate and the incidence of disease increases supralinearly with paternal age. These mutations have been shown to arise spontaneously in spermatogonial stem cells (SSCs) and confer a selective advantage by hijacking critical signalling pathways involved in SSC regulation. Over time, through a process termed selfish spermatogonial selection, mutant SSCs clonally expand along the seminiferous tubules in which they originate, resulting in elevated mutation levels.

To date, all known selfish mutations occur in 12 genes encoding components of the receptor tyrosine kinase (RTK)-Ras signaling pathway. As positive selection is not anticipated to be limited to RTK-Ras dysregulation, this thesis investigated whether other mutations, genes, or pathways might also be subject to selfish behaviour.

To identify candidate selfish mutations for experimental inquiry, it was hypothesized that positive selection would manifest as DNM recurrence in clinically affected populations. Recurrent DNMs identified in a study of >31,000 parent-offspring trios were systematically evaluated to determine the most promising candidates, including double immunofluorescent labeling of testis sections for select candidate genes with established spermatogonial markers.

Following the prioritization of candidate mutations across 10 candidate genes, rhAmpSeq was adapted to perform a discovery screen and investigate whether these mutations were detectable in the testes of aged men. To improve assay sensitivity, unique molecular identifiers (UMIs) were incorporated, and their utility was assessed. This study detected variants in several novel genes, including those functioning outside of the RTK-Ras pathway.

Among these were two distinct variants in *SMAD4* associated with Myhre syndrome (MS). To assess whether MS *SMAD4* mutations are positively selected in bulk sperm, an ultra-sensitive technique was adapted to quantify MS mutations at the p.I500 codon. MS *SMAD4* mutations at p.I500 were found to be elevated in sperm but not in blood, and their frequency increased with donor age, consistent with positive selection.

Taken together, this work provides direct evidence that selfish spermatogonial selection can be driven by genes operating outside of the canonical RTK-Ras pathway and expands the known landscape of positively selected genes in the male germline.

Publications

Wood, K. A.*, Tong, R. S.*, Motta, M.*, Cordeddu, V., Scimone, E. R., Bush, S. J., Maxwell, D. W., Giannoulatou, E., Caputo, V., Traversa, A., Mancini, C., Ferrero, G. B., Benedicenti, F., Grammatico, P., Melis, D., Steindl, K., Brunetti-Pierri, N., Trevisson, E., Wilkie, A.O.M., Lin, A.E., Cormier-Daire, V., Twigg, S.R.F, Tartaglia, M., Goriely, A. (2024). SMAD4 mutations causing Myhre syndrome are under positive selection in the male germline. *The American Journal of Human Genetics*, *111*(9), 1953–1969. https://doi.org/10.1016/j.ajhg.2024.07.006

*These authors contributed equally to this work.

Table of Contents

Chapter 1 – Introduction *10*

1.1 – The origins of *de novo* mutations and human disease **11**

1.2 – The human germline point mutation rate **13**

1.2.1 – Non-random distribution of germline mutations 14

1.2.2 – Mutational signatures provide insights into the mechanisms underlying germline mutations 15

1.3 – Paternal bias in the origin of *de novo* mutations **17**

1.4 – Replication differences between the male and female germlines **19**

1.5 – An overview of spermatogenesis **22**

1.5.1 – Organization of the human testis 22

1.5.2 – The traditional model of spermatogenesis in humans 24

1.5.3 – scRNAseq reveals a more complex model of spermatogenesis than previously held 26

1.6 – Paternal Age Effect Disorders **29**

1.7 – Direct measurement of PAE mutations in sperm **34**

1.7.1 – Multiple nucleotide substitutions are a characteristic of selfish spermatogonial selection 40

1.7.2 – Positive selection of mutant spermatogonia links PAE disorders and tumorigenesis 41

1.8 – Investigation of PAE mutation levels and distributions in the testes **44**

1.9 – Immunopositive tubules and the high-throughput discovery of pathogenic DNMs in the male germline **47**

1.10 – Dysregulated RTK-Ras signalling underlies selfish spermatogonial selection **51**

1.11 – RTK-Ras signalling regulates SSC homeostasis **56**

1.11.1 – Selfish mutations exert their effects through the MAPK and PI3K/AKT pathways 57

1.11.2 – scRNAseq reveals complex signalling networks regulate SSC homeostasis in humans 60

1.12 – Positive selection of mutations in somatic tissues **61**

1.13 – Challenges of studying the human germline **64**

1.14 – Aims **65**

Chapter 2 – Materials and Methods *67*

2.1 – Ethics approval **67**

2.2 – Materials **67**

2.2.1 – Testes samples 67

2.2.2 – FFPE samples 68

2.2.3 – Sperm and blood samples 68

2.3 – Immunofluorescence **70**

2.3.1 – Tissue sectioning 70

2.3.2 – Double immunofluorescent labelling 70

2.3.3 – Imaging of IF slides 71

2.4 – Preparation of genomic DNA from testicular tissue **72**

2.4.1 – Dissection 72

2.4.2 – DNA extraction and quantification 72

2.5 – rhAmpSeq **72**

2.5.1 – Panel design 72

2.5.2 - Serial dilution of patient-derived DNA 73

2.5.3 - RNase H2-dependent PCR (rhPCR) ... 73
2.5.4 – Library preparation ... 74

2.6 – SMAD4 REDPCR ... 75
2.6.1 – Titration-reconstruction of MS *SMAD4* DNA ... 75
2.6.2 – *NsiI* RED and mutant selection ... 75
2.6.3 – PCR amplification of *NsiI*-resistant DNA ... 76
2.6.4 – Library preparation and next-generation sequencing ... 77
2.6.5 - Engineering a *NsiI*-resistant genomic integrant by CRISPR-Cas9 ... 78

2.7 – Bioinformatics analyses ... 79
2.7.1 – Variant annotation of DNMs ... 79
2.7.2 – Sequencing data processing ... 80
2.7.3 – UMI correction ... 80
2.7.4 – Statistical prioritization of low frequency variants ... 81

Chapter 3 – Identification and prioritization of candidate genes for selfish selection ... 85

3.1 – Prioritization of candidate *de novo* mutations from >31K parent-offspring trios ... 86

3.2 – Expression of candidate selfish genes in SSCs ... 90
3.2.1 – Assessing expression using an integrated scRNAseq atlas of human spermatogonia ... 90
3.2.2 – Corroboration of expression via the Human Protein Atlas ... 93
3.2.3 – Double immunofluorescence of select candidates with spermatogonial markers ... 96

3.3 – Analysis of candidate selfish genes: function, mechanism of mutation, and Paternal Age Effect criteria ... 100
3.3.1 – Phosphofurin acidic cluster sorting protein 1 (*PACS1*) ... 101
3.3.2 – Protein phosphatase 2 regulatory subunit B'delta *(PPP2R5D)* ... 102
3.3.3 – Mothers against decapentagplegic homolog 4 *(SMAD4)* ... 104
3.3.4 – Phosphofurin acidic cluster sorting protein 2 (*PACS2*) ... 108
3.3.5 – Activity-dependent neuroprotective protein (*ADNP*) ... 110
3.3.6 – Methyl-CpG-binding protein (*MECP2*) ... 112
3.3.7 – NatA catalytic subunit 10 (*NAA10*) ... 114
3.3.8 – Protein phosphatase 1 catalytic subunit beta (*PPP1CB*) ... 116
3.3.9 – Cyclin-dependent kinase 13 (*CDK13*) ... 118
3.3.10 – Casein kinase 2α subunit *(CSNK2A1)* ... 120
3.3.11 – Forkhead box P (Fox) transcription factor family protein 1 (*FOXP1*) ... 121
3.3.12 – C-terminal binding protein (*CTBP1*) ... 123
3.3.13 – Leucine rich repeat scaffold protein (*SHOC2*) ... 125
3.3.14 – Snf2-related CBP activator protein (*SRCAP*) ... 126
3.3.15 – Asparagine-linked glycosylation 13 homology (*ALG13)* ... 128
3.3.16 – Glutamate receptor, ionotropic, N-methyl, D-aspartate, subunit 2B (*GRIN2B*) ... 130
3.3.17 – Guanine nucleotide-binding protein subunit beta-1 (*GNB1*) ... 131

3.4 – Prioritizing candidate DNMs for selfish behaviour ... 136

3.5 – Discussion ... 140

Chapter 4 – Discovery of novel selfish mutations in the aged human testes ... 144

4.1 – Introduction ... 144

4.2 –Proof-of-concept: determining optimal rhAmpSeq cycling conditions ... 147
4.2.1 – Sequencing coverage and UMI distribution ... 149
4.2.2 – Analysis of serial dilutions to determine rhAmpSeq sensitivity ... 151
4.2.3 – Characterization of error rates across substitution type and cycling conditions ... 153
4.2.4 – Determination of rhPCR cycling conditions ... 156

4.3 – Investigation of selfish candidate mutations in aged human testes ... 156

4.3.1 – Experimental design and strategy to call low frequency variants 156
4.3.2 – rhAmpSeq Pilot study 160
4.3.3 – Discovery screen in aged human testes 174
4.4 – Discussion 197
Chapter 5 – Investigation of pathogenic SMAD4 mutations in sperm 203
5.1 – Introduction 203
5.2 – Results 204
5.2.1 – Strategy to detect ultra-rare *SMAD4* mutations in sperm 204
5.2.2 – Assessing the sensitivity of the SMAD4 REDPCR assay 206
5.2.3 - A pilot study of SMAD4 mutation prevalence in sperm 206
5.2.4 – Generation of an *NsiI*-resistant genomic integrant 209
5.2.5 – Titration-reconstruction experiment using Myhre syndrome DNA 210
5.2.6 – Reproducibility of the SMAD4 REDPCR assay 212
5.2.7 - Quantification of SMAD4 c.1494–1499 mutation levels in sperm and blood 215
5.3 – Discussion 221
Chapter 6 - Conclusion and future directions 226
6.1 – Identification of candidate selfish mutations 226
6.2 – Discovery of novel candidate mutations in aged testes 228
6.3 – *SMAD4* mutations at p.I500 undergo positive selection in sperm 230
6.4 – Future directions 231
6.5 – Concluding remarks 232
Appendix 233
Bibliography 242

List of Figures and Tables

Figure 1.1: The origin and timing of *de novo* mutations. (12)
Figure 1.2: Patterns of single base substitutions (SBS) reveal distinct mutational processes. (16)
Figure 1.3: Sequencing of parent-offspring trios reveal a paternal bias in *de novo* mutations. (19)
Figure 1.4: Anatomy of the human testis and seminiferous tubules. (23)
Figure 1.5: The classical model of spermatogenesis is organized by nuclear morphology. (25)
Figure 1.6: Unsupervised clustering of spermatogonial single-cell transcriptomes. (27)
Figure 1.7: Epidemiological parental age effects in Apert syndrome. (30)
Figure 1.8: Comparison of experimentally determined PAE mutation levels in sperm relative to donor age. (36)
Figure 1.9: Strategies to study pathogenic DNMs directly within the testes. (46)
Figure 1.10: Immunopositive tubules can be traced across several millimeters, indicative of clonal expansion along seminiferous tubules. (50)
Figure 1.11: Overview of the RTK-Ras signaling cascade. (52)
Figure 1.12: scRNAseq of spermatogonia implicates a complex network of signalling pathways that may be involved in SSC homeostasis. (61)
Figure 1.13: Seborrheic keratoses harbour activating mutations in the RTK-Ras pathway. (63)

Figure 2.1: Original CRISPR-Cas9 design to generate a *NsiI*-resistant genomic integrant. (79)

Figure 3.1: UMAP visualization of gene expression in human spermatogonia for candidate and established selfish genes. (91)
Figure 3.2: Immunohistochemical (IHC) staining patterns of candidate and established selfish genes in human testes. (94)
Figure 3.3: Multiplexed immunofluorescence (IF) validation of select candidate gene expression in spermatogonia. (98)
Figure 3.4: Flow chart illustrating the strategy to prioritize candidate selfish genes. (139)

Figure 4.1: rhAmpSeq strategy to develop multiplexed assays for variant detection in testes biopsies. (146)
Figure 4.2: Overview of rhAmpSeq proof-of-concept experiment to determine optimal cycling conditions for rare variant detection. (148)
Figure 4.3: Sequencing coverage across proof-of-concept experiment. (149)
Figure 4.4: UMI group distribution and size. (150)
Figure 4.5: Serial dilution of DNA from 9 probands. (152)
Figure 4.6: rhAmpSeq error rates by cycling conditions and across amplicons. (154)
Figure 4.7: rhAmpSeq error rates with and without UMI correction. (155)
Figure 4.8: Overview of rhAmpSeq strategy to investigate candidate selfish mutations in aged testis biopsies. (159)
Figure 4.9: Sequencing coverage in the rhAmpSeq pilot study. (161)
Figure 4.10: UMI group distribution in the rhAmpSeq pilot study. (163)
Figure 4.11: Sensitivity of the rhAmpSeq pilot study. (167)
Figure 4.12: Mutation plots for three control variants in previously screened testis biopsies (Tes4, 90 years). (169)
Figure 4.13: Sequencing coverage in the rhAmpSeq discovery screen. (176)

Figure 4.14: Sensitivity of the rhAmpSeq discovery screen. (179)
Figure 4.15: Example mutation plots of variants in control selfish genes identified in the rhAmpSeq discovery screen. (182)
Figure 4.16: Mutation plots of variants in candidate selfish genes identified in the rhAmpSeq discovery screen. (191)
Figure 4.17: Spatial distribution of all variants identified in a rhAmpSeq discovery screen of testes biopsies from three aged individuals. (194)

Figure 5.1: Overview of the *SMAD4* Restriction Enzyme Digest Polymerase Chain Reaction (REDPCR) assay. (205)
Figure 5.2: Establishing the limit of detection of the *SMAD4* REDPCR assay and determining whether MS *SMAD4* mutations are detectable in sperm. (208)
Figure 5.3: Sanger sequencing traces of the *NsiI*-resistant genomic integrant and a reference blood control demonstrate the location of the CRISPR-Cas9 edit. (209)
Figure 5.4: The *SMAD4* REDPCR assay is sensitive to ~10^{-6} for mutations at *SMAD4* c.1494–1499. (211)
Figure 5.5: Mutation level estimates made with the *SMAD4* REDPCR assay are consistent across technical and biological replicates. (213)
Figure 5.6: Cumulative mutation levels at *SMAD4* c.1494–1499 in a cohort of 57 sperm and 5 blood samples. (217)
Figure 5.7: Quantification of all 18 substitutions at SMAD4 c.1494–1499 in a cohort of 57 sperm and 5 blood samples (24.8–75.9 years). (219)
Figure 5.8: Correlation between mutation levels for the 9 reciprocal substitutions within the *NsiI* restriction site. (220)

Table 1.1: Summary of established PAE genes and associated disorders with accompanying evidence (33)
Table 2.1: Overview of testes samples (67)
Table 2.2: Overview of sperm and control samples (68)
Table 2.3: Antibodies used in immunofluorescent experiments (71)
Table 2.4: Reference table for the *SMAD4* REDPCR serial dilution experiments (75)
Table 2.5: Oligonucleotide sequences (82)
Table 3.1: Comparison of prioritized recurrent *de novo* mutations identified at the amino acid-level vs. the nucleotide-level (88)
Table 3.2: Evidence for selfish potential across prioritized candidate genes based on mutation patterns and PAE characteristics (133)
Table 3.3: Stratification of candidate and control selfish genes through an integrated scoring system. (138)

Table 4.1: rhAmpSeq selfish candidate panel. (157)
Table 4.2: Selfish mutations identified in previously-screened testis biopsies from an aged man (Tes4, 90 years) (168)
Table 4.3: Novel variants identified in the rhAmpSeq pilot study. (172)
Table 4.4: Mutational events detected in the rhAmpSeq discovery screen by amplicon (181)
Table 4.5: Variants identified by rhAmpSeq screening of testes biopsies from three aged individuals (63 – 76 years) (184)

Chapter 1 – Introduction

The relationship between parental age and human genetic disease has been the subject of scientific inquiry for over a century. A central finding of the field is that as parental age increases so too does the incidence of some sporadic congenital disorders, prompting the hypothesis that mutant germ cells accumulate over time. The popularly known example is Down syndrome, caused by *de novo* trisomy of chromosome 21, which increases in frequency with maternal age due to higher numbers of aneuploid oocytes. Through next-generation sequencing of parent-offspring trios, we now know that the majority of *de novo* mutations arise on the paternal allele (i.e., in spermatogonia), and that the number of *de novo* mutations in the offspring increases linearly with the age of the father. This has been attributed to the continuous replication of spermatogonial stem cells (SSCs) to maintain spermatogenesis throughout male reproductive life.

This thesis focuses specifically on a subset of mutations that arise in SSCs and result in their clonal expansion in the testis. Proposed to be driven by positive or "selfish" selection, this class of mutations has been termed paternal age effect (PAE) mutations to reflect the strong association between the incidence of the cognate disorder and advanced paternal age. PAE disorders share three defining characteristics: (1) an exclusive (or near exclusive) paternal origin of mutation, (2) a statistically significant epidemiological paternal age effect, where fathers are 2–7 years older than the average age of fatherhood, and (3) a high apparent germline mutation rate (up to three orders of magnitude higher than the genome-wide rate). To date, evidence for selection has been established in 12 genes, each of which encodes components of the RTK-Ras signaling cascade. By applying both novel and established techniques, this thesis will explore whether there are additional mutations and genes that exhibit selfish behaviour.

1.1 – The origins of *de novo* mutations and human disease

Spontaneous or *de novo* mutations are newly occurring mutations from generation to generation and are a critical substrate for human evolution. *De novo* mutations can arise in either parental germline and are transmitted to their offspring during reproduction, or otherwise occur as post-zygotic mutations during early embryonic development thereby resulting in mosaicism (**Figure 1.1**). Mutations arising during parental gametogenesis are estimated to account for ~89% of the DNMs detected in the offspring, while parental and offspring post-zygotic mosaicism contribute ~4% and ~7% of mutations respectively (Acuna-Hidalgo et al., 2015; Rahbari et al., 2016; Acuna-Hidalgo et al., 2016). Mutations occurring in the soma of the parents are, by contrast, not passed onto their offspring. *De novo* mutations can take a variety of forms with single-nucleotide variants (SNVs; also known as point mutations) being the most prevalent. Whole genome or exome sequencing studies of parent-offspring trios have shown that, on average (and depending on the parental age at conception), ~70 point mutations occur per offspring genome, whereas small insertions and deletions (indels) and copy number variations (CNVs; defined as deletions or duplications that are 1 kb or greater) account for ~3 and ~0.02 events per genome, respectively (Veltman and Brunner, 2012). Given their relative prevalence, this thesis will focus on *de novo* point mutations (DNMs).

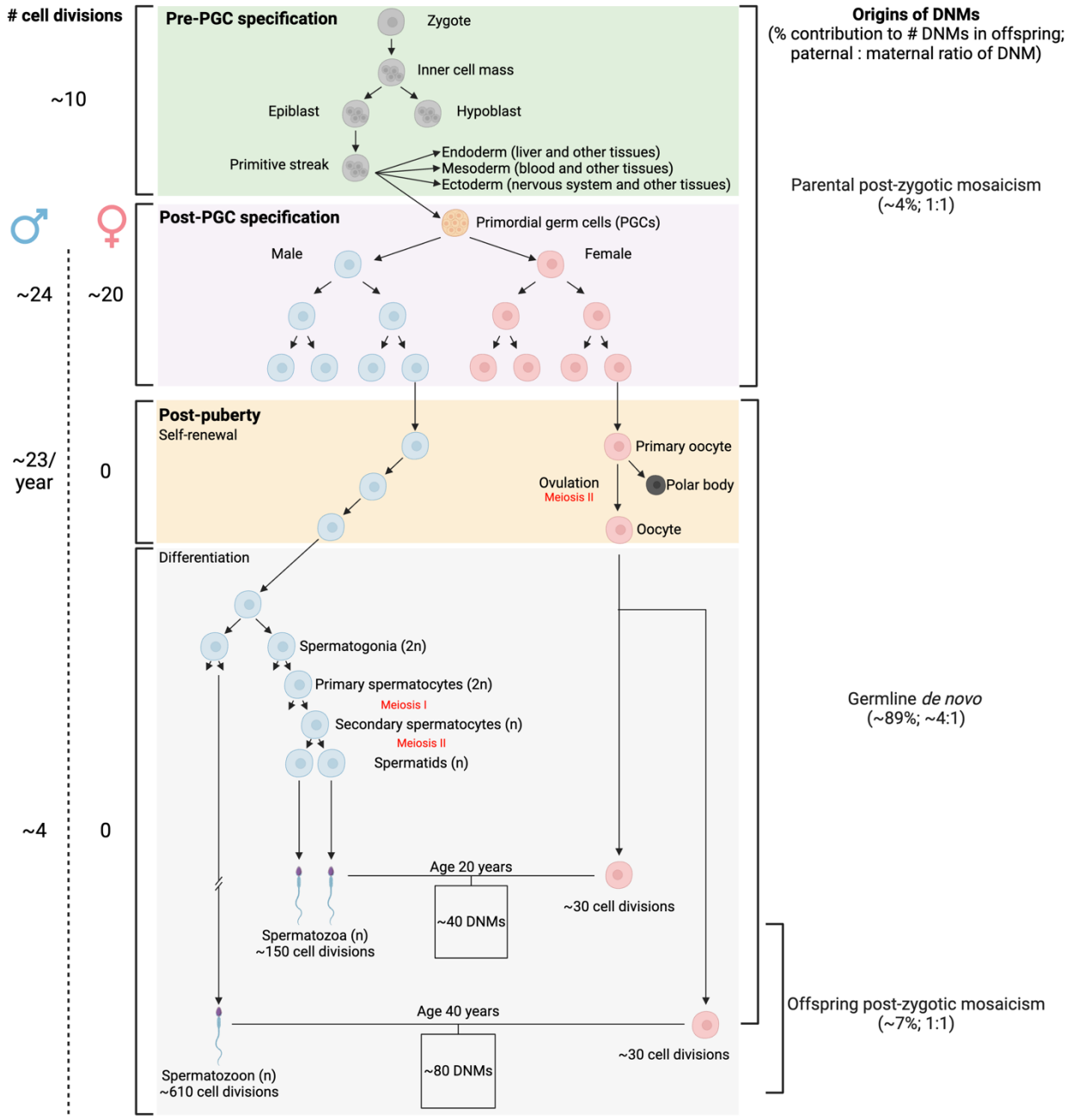


Figure 1.1: The origin and timing of *de novo* mutations. Depiction of the developmental trajectory of the human germline and the accumulation of *de novo* mutations (DNMs) throughout gametogenesis. The figure is divided into three key stages (bolded): pre-primordial germ cell (PGC) specification, post-PGC specification, and post-puberty. The pre-PGC stage is highlighted in green and shows early embryonic development from a zygote to PGC formation encompassing approximately 10 cell divisions. The post-PGC stage is highlighted in purple and demonstrates the divergence of male and female germ cell development, with males undergoing approximately 24 divisions and females about 20 divisions before birth. The post-puberty stage, depicted by yellow and gray sections, illustrate the striking contrast between male and female gametogenesis. While females maintain arrested primary oocytes until ovulation (at which point primary oocytes

complete meiosis II to become oocytes), males continuously undergo spermatogonial stem cell (SSC) cell divisions throughout life at a rate of approximately 23 per year. The figure quantifies the number of cell divisions and resulting DNMs at parental ages of 20 and 40 years. The figure also categorizes the origins of DNMs (right) into three types: parental post-zygotic mosaicism, germline *de novo*, and (offspring) post-zygotic mosaicism, providing their respective contributions to the number of DNMs detected in the offspring as well as the paternal:maternal ratios of mutation. Adapted from Rahbari et al., 2016; Acuna-Hidalgo et al., 2016; Goriely, 2016; Goldmann et al., 2019. Figure created with BioRender.

DNMs are a major driver of human genetic disease, with large-scale sequencing studies estimating that ~42–48% of individuals with severe developmental disorders carry a pathogenic DNM in a coding gene (Deciphering Developmental Disorders Study, 2017; Martin et al., 2018). Accounting for parental age, the estimated birth prevalence of monoallelic developmental disorders ranges from 1 in 448 (0.22%) for 20-year-old parents to 1 in 213 (0.47%) for 45-year-old parents (Deciphering Developmental Disorders Study, 2017), translating to nearly 400,000 of the 140 million annual births worldwide. This represents a substantial burden to public healthcare systems, especially given the intensive lifelong support often required by affected individuals. In the United States, for instance, individuals with intellectual or developmental disabilities command two to five times higher annual expenditure than the average Medicaid beneficiary (i.e. older individuals with physical or other disabilities), with total annual costs exceeding $23 billion (United States Government Accountability Office, 2023).

1.2 – The human germline point mutation rate

As DNMs primarily arise in the parental germline, understanding the mechanisms, rates, and patterns of human germline mutation is of fundamental importance. Early estimates of the human germline point mutation rate came from studies of individual genetic loci known to be implicated in severe Mendelian diseases, yielding indirect estimates of 1.3–2 x 10^{-8} per nucleotide per

generation (Kondrashov, 1988; Kondrashov, 2002; Lynch, 2009). The advent of next-generation sequencing has since transformed these efforts and offers a genome-wide view into how and when DNMs arise. Direct measurements through short-read whole-genome sequencing (WGS) of parent-offspring trios have placed the germline mutation rate between 1–1.3 x 10^{-8} per nucleotide per generation, which translates to the approximately ~70 DNMs per generation (Roach et al., 2010; Michaelson et al., 2012; Kong et al., 2012; Rahbari et al., 2016; Goldman et al., 2016). More recent developments – such as the concurrent use of long-read sequencing to uncover repetitive and complex genomic regions, as well as the completion of a telomere-to-telomere reference human genome – have revised estimates up to ~192 *de novo* mutations per generation (Nurk et al., 2022; Noyes et al., 2022; Porubsky et al., 2024).

1.2.1 – Non-random distribution of germline mutations

The frequency of germline mutation varies based on the type of mutation and the local nucleotide composition. Despite there being eight possible transversions and four possible transitions, approximately two thirds of DNMs are transition mutations (Nachman and Crowell, 2000; Roach et al., 2010; Kong et al., 2012; Goldmann et al., 2016). This is driven in large part by differences in mutation rates at CpG dinucleotides, where cytosines are predominantly methylated at the fifth position, resulting in 5-methylcytosine (5mC). Spontaneous deamination of 5mC to thymine (C>T; G>A on the complementary strand) can become incorporated if it is not repaired prior to DNA replication (Cooper and Krawczak, 1989). Transitions at CpG sites occur >10x more frequently than transitions at non-CpGs, and account for ~20% of detected DNMs despite CpGs comprising only 2% of the human genome (Nachman and Crowell, 2000; Kong et al., 2012; Ségurel et al., 2014). Transversions at CpG dinucleotides also occur at elevated (>2x) rates than at non-CpG

sites. While the exact mechanisms are not wholly understood, this phenomenon is thought to be driven, at least in part, by a universal bias favouring mutations that decrease GC content; mutations that change strong (G/C) bases to weak (A/T) ones occur ~2x more frequently for both transitions and transversions (Kong et al., 2012; Lynch, 2010).

1.2.2 – Mutational signatures provide insights into the mechanisms underlying germline mutations

In addition to the single nucleotide changes and the dinucleotide contexts, the broader sequence environment also influences mutation rates, as evidenced by the study of mutational signatures. Mutational signatures were first described by applying mathematical models to 21 breast cancer whole-genomes and considered the frequency of mutations across the trinucleotide contexts of single base substitutions (Nik-Zainal et al., 2012). Six classes of base substitutions (C>A, C>G, C>T, T>A, T>C, and T>G; the pyrimidine of the base pair is used to denote the mutated base) with 16 possible trinucleotide contexts exist, yielding 96 possible mutated trinucleotides.

Mutational signatures are thought to be reflective of specific mutational processes and their impact on the genome. For example, single base substitution (SBS) signature 4 is characterized predominantly by C>A (G>T) transversions (**Figure 1.2A**) and shows a transcriptional strand bias, with more mutations occurring on the transcribed strand (Alexandrov et al., 2013). SBS signature 4 (SBS4) is commonly observed in lung, esophageal, head and neck, and liver cancers, all of which have been strongly linked to tobacco use. SBS4 is also consistent with findings that polycyclic aromatic hydrocarbons (a prominent class of carcinogens in tobacco smoke) induce aromatic DNA adducts that give rise to G>T transversions (Pfeiffer et al., 2002), together supporting SBS4 as a genomic imprint of tobacco smoke (Alexandrov et al., 2013).

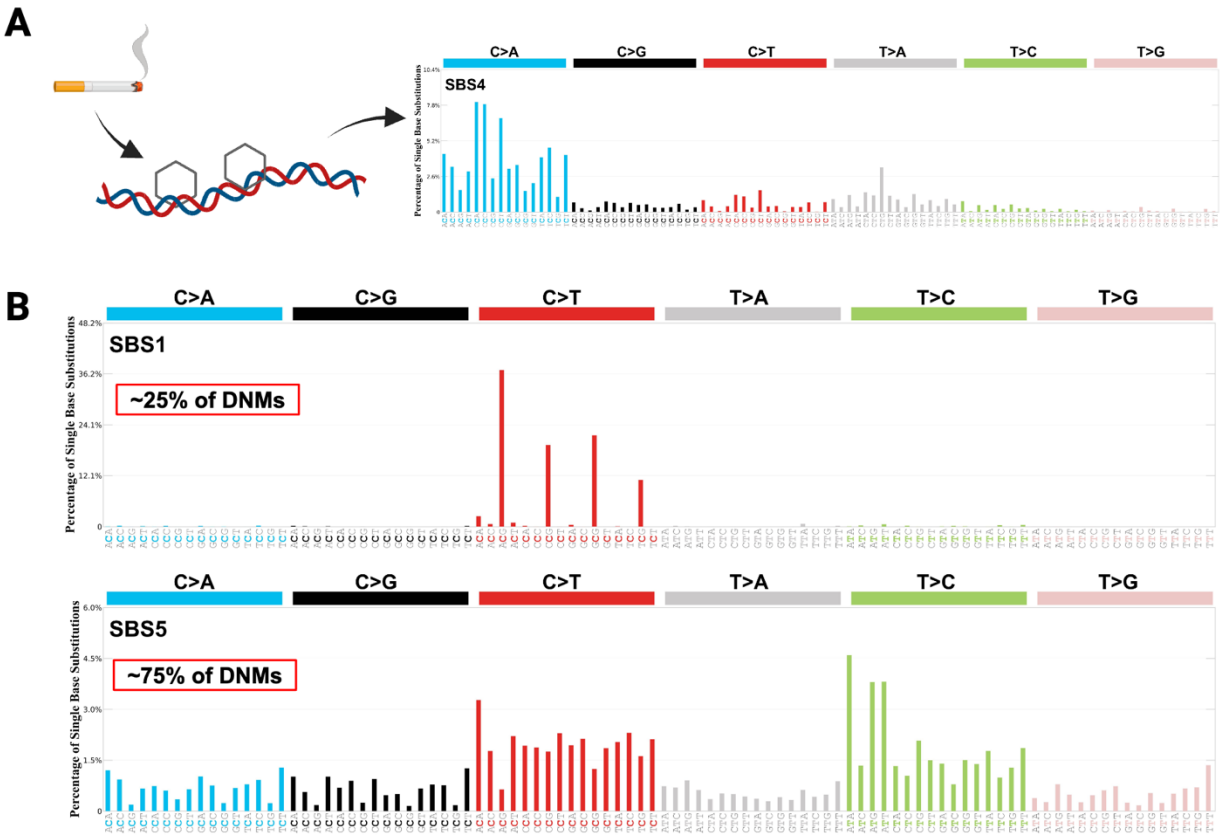


Figure 1.2: Patterns of single base substitutions (SBS) reveal distinct mutational processes. Each mutational signature is represented by the relative frequency across 96 possible trinucleotide contexts, grouped by the 6 base substitution types (coloured). The y-axis represents the percentage of mutations attributed to each trinucleotide within the signature. **(A)** SBS4 reflects the effect of tobacco-induced aromatic lesions on genomic DNA, leading to an overrepresentation of C>A (G>T) mutations. **(B)** *De novo* mutations can be deconvoluted into two distinct mutational signatures: SBS1 and SBS5 which account for 25% and 75% of DNMs, respectively. Figures of mutational signatures adapted from COSMIC v101 (cancer.sanger.ac.uk) and created with BioRender.

In the latest efforts to characterize mutational signatures in cancer, ~85 million somatic mutations from ~24K cancers were analyzed revealing 49 SBS signatures, as well as 11 double base substitution, 4 clustered base substitution, and 17 small insertion and deletion signatures (Alexandrov et al., 2020). Mutational signatures for structural variation, copy number, and RNA single base substitutions have also been established (Li et al., 2020; Steele et al., 2022; Martínez-Ruiz et al., 2023).

This framework of defining mutational signatures has been applied to parent-offspring trio sequencing data to better understand the mutational processes that underly germline mutations. By investigating SBS signatures across 109 trios, it was shown that SBS1 and SBS5 accounted for 25% and 75% of DNMs, respectively (**Figure 1.2B**) (Rahbari et al., 2016). SBS1 is characterized by C>T (G>A) transitions at NpCpG trinucleotides (where N can represent any base), whereas SBS5 is largely characterized by both C>T substitutions (without preference for CpG dinucleotides) and T>C (A>G) transitions. The mechanisms driving SBS5 are incompletely understood but it is thought to be driven by spontaneous deamination of both cytosine (becoming uracil and read as thymine) and adenine (becoming hypoxanthine and read as guanine) (Alexandrov et al., 2015). A transcriptional strand bias was also observed for the T>C mutations, potentially suggesting that these mutations may be derivative of DNA adducts repaired by transcription-coupled repair. In cancer, SBS1 and SBS5 are considered "clock-like" mutational signatures as their prevalence generally increases with the age of diagnosis (Alexandrov et al., 2015).

1.3 – Paternal bias in the origin of *de novo* mutations

While sequencing parent-offspring trios has provided insights into the frequency and distribution of germline mutations, it also has underlined important differences in the number of mutations contributed by each parent. By leveraging nearby informative single-nucleotide polymorphisms (SNPs), phased haplotypes of DNMs can be constructed and consistently show that ~80% of DNMs are paternally derived (**Figure 1.3A**) (Kong et al., 2012; Rahbari et al., 2016; Goldmann et al., 2016; Jónsson et al., 2017). Strikingly, the number of DNMs in the offspring increases monotonically with ~1–3 additional DNMs per year of the father's age (**Figure 1.3B**). By contrast,

the maternal contribution is less pronounced and estimated to be far lower at <0.4 additional DNMs per year of maternal age. This paternal bias in DNM transmission results in an approximate doubling of genic DNMs detected in the offspring of 40-year-old fathers compared to those of 20-year-old fathers (Francioli et al., 2015).

This male bias in mutation rate had first been proposed through studying the birth patterns of hemophilia, a recessive X-linked disorder. By studying pedigrees, Haldane found that sporadic hemophilia occurred much more frequently in males with heterozygous, carrier mothers than from homozygous, normal mothers (Haldane, 1946). As the maternal grandfathers were unaffected, it was reasoned that the most probable origin of the causative mutation was the germline of the maternal grandfather. Based on this analysis, Haldane inferred that the male germline mutation rate was possibly 10-fold higher than that of the female germline mutation rate (Haldane, 1946; Crow, 2000). Direct evidence for male-biased mutation was later provided by comparing the rates of substitutions between the Y chromosome (representative of the male mutation rate) and an autosome (chromosome 3; reflecting mutations arising in both male and female germlines) across humans and higher primates (Makova and Li, 2002). This allowed for the isolation of a male-to-female mutation rate ratio (α) of 5.25 (compared to the observed α of ~4 demonstrated by trio sequencing) strongly favouring male-driven evolution in hominoids.

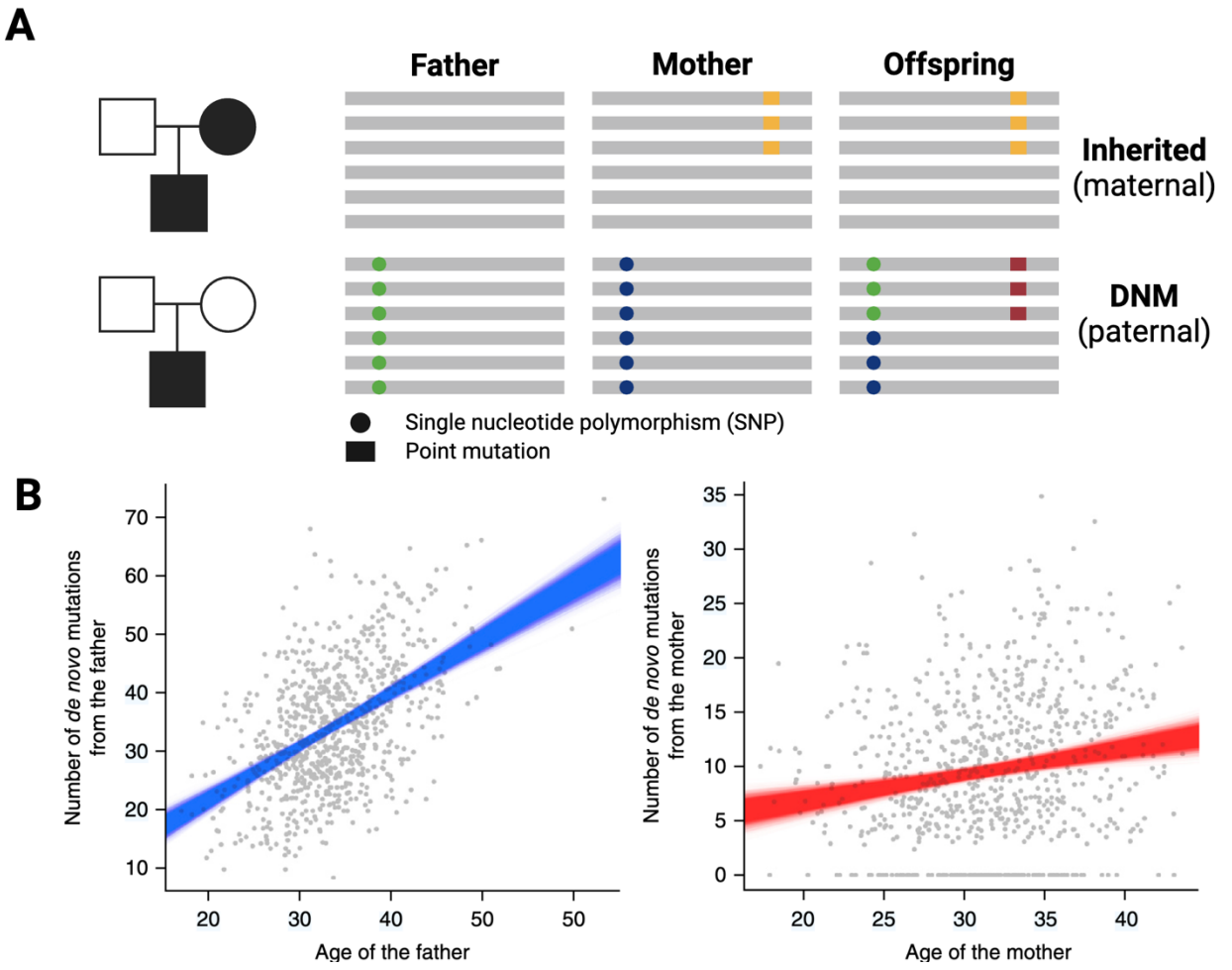


Figure 1.3: Sequencing of parent-offspring trios reveal a paternal bias in *de novo* mutations. **(A)** Representative pedigrees are accompanied by associated results from next-generation sequencing. Inherited variants will be found in both the offspring and carrier parent. The top trio illustrates a maternally derived inherited variant. *De novo* mutations can be phased by interrogating nearby informative SNP and constructing haplotypes. The bottom trio illustrates a paternally derived DNM. Figure created with BioRender. **(B)** Scatter plots showing the relationship between parental age and the number of phased DNMs in offspring from whole-genome sequencing of 816 trios and ~7200 phased DNMs. In this study, there was ~1 additional DNM per year of the father's age and ~1 additional DNM per four years of maternal age. Figure from Goldmann et al., 2016.

1.4 – Replication differences between the male and female germlines

The prevailing view is that the paternal bias in DNM transmission reflects the differences in germ cell divisions between spermatogonia and oogonia. During early human embryogenesis, there are ~10 cell divisions prior to the specification of primordial germ cells (PGCs) (**Figure 1.1**) (Rahbari

et al., 2016). PGCs require an additional ~20 divisions to complete female gametogenesis to form the complete pool of primary oocytes, which are arrested in meiotic prophase I. Male gametogenesis is completed through an additional ~24 divisions to colonize the naïve seminiferous tubules and establish the initial pool of prespermatogonia, which are mitotically arrested.

At the onset of puberty, the subsequent developmental paths of male and female germ cells diverge significantly. In males, SSCs begin self-renewing at a rate of approximately 23 divisions per year (once every 16 days) and require an additional 4 divisions to produce mature sperm (**Figure 1.1**) (Heller and Clermont, 1963; Rahbari et al., 2016). This translates to 150 cumulative cell divisions by the age of 20 versus 610 divisions at the age of 40 (Hurst and Ellegren, 1998). By contrast, oocytes only undergo one additional round of meiosis during ovulation to prepare for fertilization. The number of cell divisions for the average set of parents (e.g., ~30 years) differs 14-fold with ~430 for the father and ~30 for the mother.

However, if all *de novo* mutations are a result of DNA replication, then the calculated frequency of cell divisions (~14-fold difference for 30-year-old parents) is not proportional to the observed patterns of spontaneous mutation accumulation ($\alpha \sim= 4–5$). One key assumption that has been questioned is the model of human spermatogenesis, first proposed by Heller and Clermont through radiolabelling seminiferous epithelium in the 1960s, which posits a constant rate of post-pubertal SSC divisions (23/year) (Heller and Clermont, 1963; Ségurel et al., 2014; Goldmann et al., 2019). Insights into the ordered nature of the human spermatogenesis have been gleaned from the mouse, which has also been considered to have a hierarchal model of spermatogenesis, albeit with a shorter seminiferous epithelial cycle of 8.6 days compared to ~16 days in humans. Long-term *in vivo* tracing of labelled SSCs in mice has challenged our understanding of SSC dynamics: under a hierarchical model, long-lived mouse SSCs divide asymmetrically, and the number and

size of labelled clones would be expected to be stable, reflective of the original fraction of labelled SSCs at induction (Klein et al., 2010). Instead, it has been shown that the entire SSC compartment undergoes rapid turnover (<2 weeks on average) and that clonal expansion and extinction occurs stochastically. The fate of one SSC is coupled to that of its neighbours such that the proliferation of one clone occurs with the loss of another, with the overall population of SSCs maintained through a model of neutral drift. More recent data from studies of human spermatogonia have also suggested that SSCs exhibit developmental plasticity and may develop in adaptation to the broader SSC niche (further discussed in **sections 1.5.3 and 1.11**). Together these data in mice and humans suggest that human spermatogenesis may be a more complex and dynamic process than previously thought.

Although modest, there is a maternal age effect of an additional <0.4 DNMs per year that must be driven by non-replicative mechanisms. One such mechanism was highlighted through whole-genome sequencing of ~1300 parent-offspring trios showing that clustered DNMs (cDNMs; DNMs on the same chromosome of the same individual within 20 kilobases of each other) are equally distributed between maternal and paternal alleles and exhibit a stronger association with maternal age than paternal age (Goldmann et al., 2018). Maternally derived cDNMs are characterized by a preponderance of C>G transversions and are predominantly detected in genomic regions prone to meiotic gene conversions (i.e., chromosomes 8, 9, and 16). Additionally, maternal cDNMs additionally exhibit a similar mutational profile to SBS signature 3, which has been associated with defective homologous recombination-based DNA damage repair (Alexandrov et al., 2013). Together, these data suggest that decreasing efficiency of double-strand break repair leading to deregulated recombination may underly the contributions of the ageing maternal germline (Goldmann et al., 2016; Goldmann et al., 2018). Although the discrepancies

between the number of male and female cell divisions and the mutation rate ratio remain incompletely understood, replication-driven errors are nevertheless still considered to drive the majority of the observed male mutation bias.

1.5 – An overview of spermatogenesis

A thorough understanding of the observed male bias in mutation rate requires more than just recognizing the differences in the number of cell divisions between male and female germlines and necessitates a deeper investigation into the biology of normal human spermatogenesis and the processes governing SSC regulation.

1.5.1 – Organization of the human testis

The human testis is divided into ~250–300 lobules, each containing 1–3 highly convoluted seminiferous tubules, which serve as the functional units of spermatogenesis. Each tubule measures ~60 cm in length and ~180 μm in diameter (**Figure 1.4**). Seminiferous tubules connect to straight tubules (tubuli recti), which release mature spermatozoa into the rete testis, from where they are delivered into the excurrent ducts of the epididymis (Weinbauer et al., 2010; Holstein et al., 2003). Seminiferous tubules are surrounded by the lamina propria, which consists of an inner basal membrane and 5–7 outer layers of myofibroblasts (peritubular cells) embedded in extracellular matrix (Davidoff et al., 1990). The lamina propria is essential for maintaining the structural integrity of the tubule and coordinates peristaltic contractions of the tubule to transport immotile spermatozoa to the rete testis. Outside of the lamina propria lies the interstitial compartment which contains testosterone-producing Leydig cells, as well as the vasculature, immune cells, nerves, and connective tissues (Weinbauer et al., 2010).

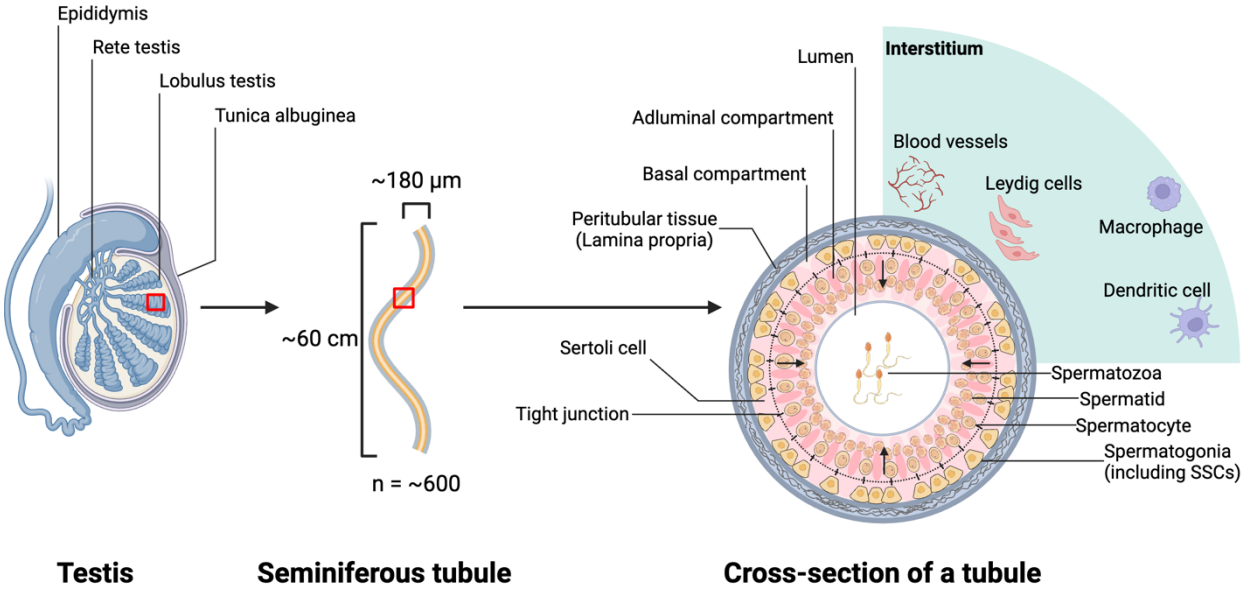


Figure 1.4: Anatomy of the human testis and seminiferous tubules. From left-to-right is an increasingly granular view of the testis to a cross-section of a seminiferous tubule. The testis is enclosed by the tunica albuginea and is organized into lobules containing the seminiferous tubules. The most undifferentiated germ cells, including spermatogonial stem cells (SSCs), are found in the basal compartment and develop inwards in the adluminal compartment. Mature spermatozoa are released into the lumen where they are eventually emptied into the rete testis and ejaculated through the epididymis. Support cells such as Sertoli cells span the entire epithelium and form tight junctions, creating the blood-testis barrier. The interstitium, located between tubules, is denoted in blue and contains critical cells such as Leydig cells responsible for testosterone production, as well as the vasculature and immune cells. Figure created with BioRender.

Within the tubule walls, the seminiferous epithelium is radially organized by the developmental stages of various germ cells: spermatogonia reside along the basal membrane, whereas spermatocytes and spermatids are located in the adluminal compartment. Mature spermatozoa are released from the seminiferous epithelium into the tubular lumen. Alongside the developing germ cells are the Sertoli cells, essential somatic cells that span the basement membrane to the lumen and provide support through nutritional transport, waste management, cell-to-cell communication. Sertoli cells are connected by tight junctions which form the blood-testis barrier (BTB); developing germ cells pass through this barrier to enter the adluminal compartment

where the BTB helps maintain immune privilege (França et al., 2016). Spermatogenesis is further supported by other structures within the broader SSC niche including the surrounding peritubular myoid cells which orchestrate contractility and sperm transport, as well as the secretion of key signaling factors such as glial cell line-derived neurotrophic factor (GDNF) (Mayerhofer, 2013).

1.5.2 – The traditional model of spermatogenesis in humans

Historically, spermatogonia have been categorized into three groups by their nuclear morphology, as determined by haematoxylin and eosin staining: type A_{dark} (Ad), type A_{pale} (Ap), and type B (Clermont, 1966). Type Ad spermatogonia are characterized by darkly stained nuclei containing dust-like chromatin and unstained, vacuole-like cavities that lack nucleoli, and are considered the "reserve" population of quiescent stem cells (**Figure 1.5A**). In contrast, type Ap spermatogonia present with lightly-stained nuclei that contain granulated chromatin and distinctive nucleoli attached to the nuclear membrane; Ap spermatogonia balance self-renewal and differentiation. Lastly, type B spermatogonia are defined by nuclei containing lightly stained chromatin and punctuated with darkly-stained granulations, as well as a centrally-located nucleolus. Type B spermatogonia commit to meiosis and give rise to spermatocytes (Neto et al., 2016).

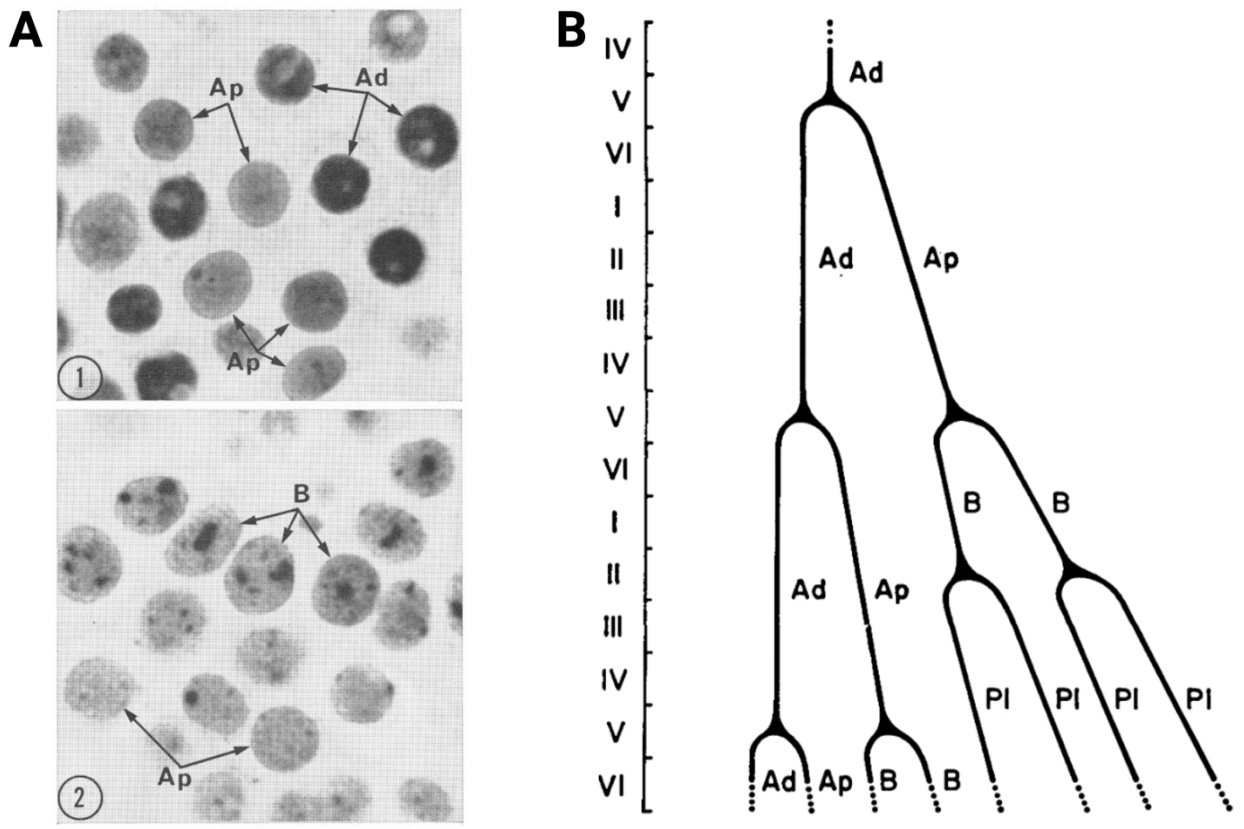


Figure 1.5: The classical model of spermatogenesis is organized by nuclear morphology. (A) Representative microphotographs of hematoxylin and eosin (H&E) staining of spermatogonia depicting the three morphologically distinct subtypes: types A_{dark} (Ad), A_{pale} (Ap), and B spermatogonia. Type Ad and Ap spermatogonia are indicated in the top panel, whereas type Ap and B spermatogonia are indicated in the bottom panel. **(B)** The hierarchal model of spermatogenesis which depicts Ad self-renewal and proposes a 1:1:2:4 ratio of spermatogonial development from Ad spermatogonia into preleptotene spermatocytes (Pl), which then enter meiosis I (not shown). Figures from Clermont Y, 1966.

By quantifying the ratio of spermatogonia and spermatocytes, Clermont proposed a model of spermatogenesis in which Ad spermatogonia divide asymmetrically to produce equal numbers of Ad and Ap spermatogonia. Type Ap spermatogonia can either self-renew or differentiate, producing two type B spermatogonia which each mitotically divide into two primary spermatocytes. These primary spermatocytes then enter meiosis I, beginning with the preleptotene stage where DNA replication occurs. They progress through several stages of prophase I, including

leptotene, zygotene, pachytene, and diplotene, during which critical events like the chromosome condensation and crossing over take place (Holstein et al., 2003). Following diakinesis, the primary spermatocyte completes meiosis I producing two secondary spermatocytes which then enter meiosis II, resulting in four haploid spermatids. In sum, each primary spermatocyte undergoes two meiotic divisions resulting in four spermatids, totalling eight spermatids from the original pair of type B spermatogonia (**Figure 1.5B**).

Despite clear histological differences, molecular profiling of Ad and Ap spermatogonia has failed to identify molecular markers that entirely distinguish the two populations (von Kopylow, 2012). Instead, a handful of markers have been identified only in subpopulations of the two morphologically distinct cell types, prompting the proposal that the nuclear morphologies may not represent distinct stem cell populations but rather different stages of the cell cycle (i.e., Ad corresponding to G0; Ap corresponding to G1/S/G2/M). This finding has been replicated by RNA sequencing (RNAseq) of laser-captured microdissected Ad and Ap cells which failed to find any differentially expressed genes between the two populations (Jan et al., 2017).

1.5.3 – scRNAseq reveals a more complex model of spermatogenesis than previously held

Recent advancements in single-cell RNA sequencing (scRNAseq) have enabled more granular classification of spermatogonia based on their gene expression patterns. Clustering analyses of single-cell transcriptomes across undifferentiated SSCs and differentiating spermatogonia have uncovered numerous distinct transcriptional states, each corresponding to a specific developmental stage in the progression from SSC self-renewal to spermatogenic commitment (Guo et al., 2017; Guo et al., 2018; Sohni et al., 2019; Di Persio et al., 2021; Bush et al., 2024). A recent effort to integrate all publicly available scRNAseq datasets of human testicular samples (60.4K cells from

29 individuals aged 14–66 years) yielded a comprehensive transcriptional map that accounted for all major germ and somatic cells (Bush et al., 2024). From this dataset, a high-resolution "human spermatogonial atlas" comprising ~4.4K cells from 9 high-quality, postpubescent testes samples was constructed and revealed that the earliest transcriptomic states represent undifferentiated spermatogonia and are categorized by three cell clusters: state 0 (characterized by *EGR4*+ and *PIWIL4*+ expression), state 0A/1 (*FGFR3*+, *GFRA1*+), and state 0B (*NANOS2*+) (**Figure 1.6A**) (Bush et al., 2024). RNA velocity analysis, which can infer cell fate trajectory, shows that cells in state 0 are transcriptionally undirected and likely to represent a reserve population (**Figure 1.6B**).

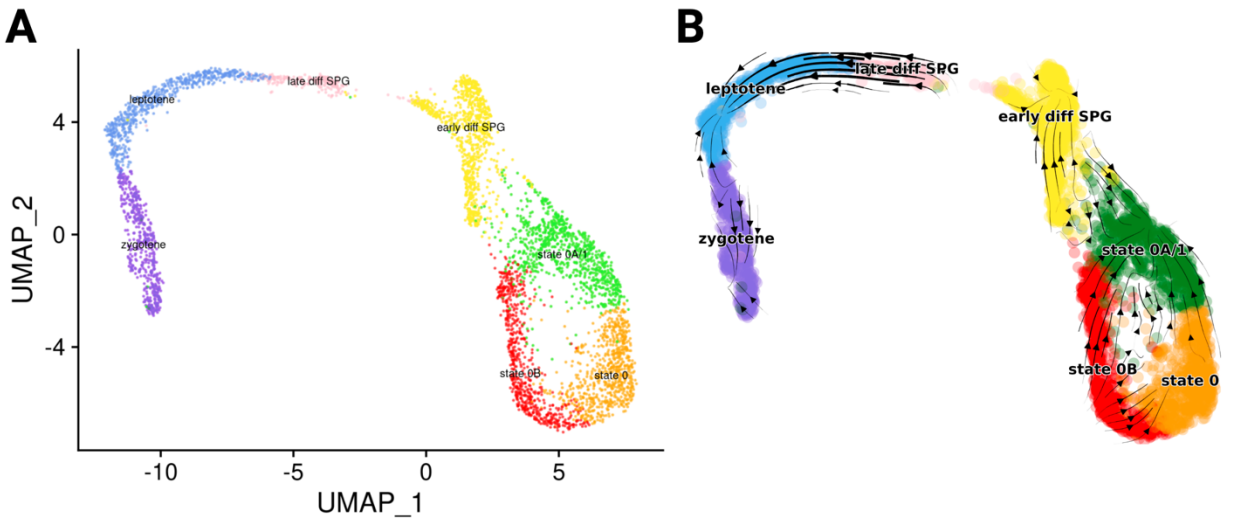


Figure 1.6: Unsupervised clustering of spermatogonial single-cell transcriptomes. (A) Unsupervised clustering of ~4400 single-cell transcriptomes of spermatogonia depicted by a UMAP plot. Each dot represents a single cell, with colours indicating the distinct transcriptional states during spermatogonial self-renewal and development. Undifferentiated spermatogonia are found in the ring of three states: 0, 0A/1 and 0B. Differentiating spermatogonia are represented by two clusters of early differentiating spermatogonia (early diff SPG) and late differentiating spermatogonia (late diff SPG). Primary spermatocytes that are undergoing meiosis I were also included and are indicated by their stage in prophase I (leptotene or zygotene). **(B)** The same UMAP plot with inferred developmental trajectories from RNA velocity analysis overlaid. Vectors indicate the predicted progression of cells through the different states of spermatogenesis. Figures from Bush et al., 2024.

However, contrary to the traditional belief in the existence of a quiescent SSC pool (Ad spermatogonia), state 0 cells do not exhibit markers of G0 quiescence, and instead appear to be predominantly positioned in either the S or G2/M phases of the cell cycle. This suggests a more active and dynamic reserve population than previously thought. In fact, the majority of all undifferentiated spermatogonia are positioned in either the S or G2/M phases, which also challenges prior hypotheses that the nuclear morphology differences between Ad and Ap spermatogonia reflect stages of the cell cycle (von Kopylow, 2012). Cells in state 0A/1 are oriented towards differentiation but have downward vectors from the cell cluster representing early differentiating spermatogonia, suggesting that these cells are capable of either progressing towards either differentiation or reverting to a less differentiated state. Cells in state 0B appear to be bidirectionally plastic with vectors both towards differentiation and back towards state 0. State 0B is characterized by a suppressive transcriptional program, notably exemplified by the expression of NANOS2. In mice, expression of NANOS2 is necessary to maintain a pool of undifferentiated, self-renewing SSCs (Sada et al., 2009; Bush et al., 2024). As this transcriptional profile was broadly associated with the upregulation of genes involved in growth arrest and DNA repair, state 0B has been hypothesized to maintain SSCs in a mitotically poised state, potentially to facilitate DNA damage repair prior to differentiation (Bush et al., 2024).

The developmental plasticity exhibited by undifferentiated spermatogonia is thought to allow cells to commit and decommit to transitioning between transcriptomic states, potentially in response to or in communication with the surrounding niche (Guo et al., 2017; Guo et al., 2018), and while discrete transcriptional states have been bioinformatically imposed onto like cells, these processes likely represent a dynamic continuum (Bush et al., 2024). The spermatogonial differentiation trajectory continues with four subsequent clusters of differentiating cells

representing early differentiating spermatogonia (*DMRT1*+, *KIT*+, *MKI67*+), late differentiating spermatogonia (*STRA8*+), leptotene spermatocytes (*SYCP3*+), and zygotene spermatocytes (*SPATA8*+). These results challenge Clermont's two-state Ad (quiescent) and Ap (dividing) model of spermatogenesis which is unidirectional and occurs in stepwise fashion. Moreover, these data point to an intricate process of SSC homeostasis with implications for how mutations may arise and propagate in the male germline.

1.6 – Paternal Age Effect Disorders

Long before molecular studies demonstrated that paternal age is the predominant factor in determining the DNM load in offspring, epidemiological studies of spontaneous disorders had indicated a role for advanced paternal age. Beginning in 1912, Weinberg noted that the incidence of achondroplasia (OMIM: #100800), the commonest form of skeletal dysplasia, was linked to birth order, such that the last-born child in a family was most at-risk (Weinberg, 1912). Four decades later, Penrose demonstrated that achondroplasia fathers were between 5.4–6.8 years older than fathers in the general population and proposed that the observations made by Weinberg were mainly driven by advanced paternal age (Penrose, 1955). Penrose originally attributed this paternal age effect to copy errors accumulating during spermatogonial cell division. However, a striking statistical analysis of parental age for sporadic cases across 17 autosomal dominant disorders showed that not all paternal age effects are equal (Risch et al., 1987). Risch's approach involved using observed/expected (O/E) ratios which compared the number of observed cases for both parents to the number of expected cases based on the age distribution of parents in the general population. By analyzing O/E ratios, it was shown that two groups of disorders existed: those whose incidence increased linearly with paternal age (compatible with a copy error model of

mutation acquisition), and those which increased non-linearly (proposed to be exponential) (Risch et al., 1987). Amongst disorders whose incidence increased supralinearly with paternal age included achondroplasia and Apert (OMIM: #101200), Crouzon (OMIM: #123500), and Pfeiffer (OMIM: #101600) syndromes. Disorders in this group had paternal O/E ratios that rose more rapidly than a linear model of mutation acquisition would predict, especially at older ages (**Figure 1.7A**). At the time, however, Risch could not rule out a modest maternal age effect and estimated that mothers contributed ~30% of mutations to the disorders in the non-linear group (**Figure 1.7B**).

The identification of the causative mutations underlying these disorders allowed for the molecular discernment of parental age effects. As an exemplar, Apert syndrome was shown to be caused by point mutations affecting neighbouring residues in *FGFR2* (RefSeq: NM_000141): c.755C>G (p.S252W) and c.758C>G (p.P253R) (Wilkie et al., 1995; Park et al., 1995). These findings paved the way for parent-of-origin studies in 57 cases of Apert syndrome where it was shown that the mutated allele was exclusively derived from the father (α), thereby providing a biological basis for the documented paternal bias (Moloney et al., 1996).

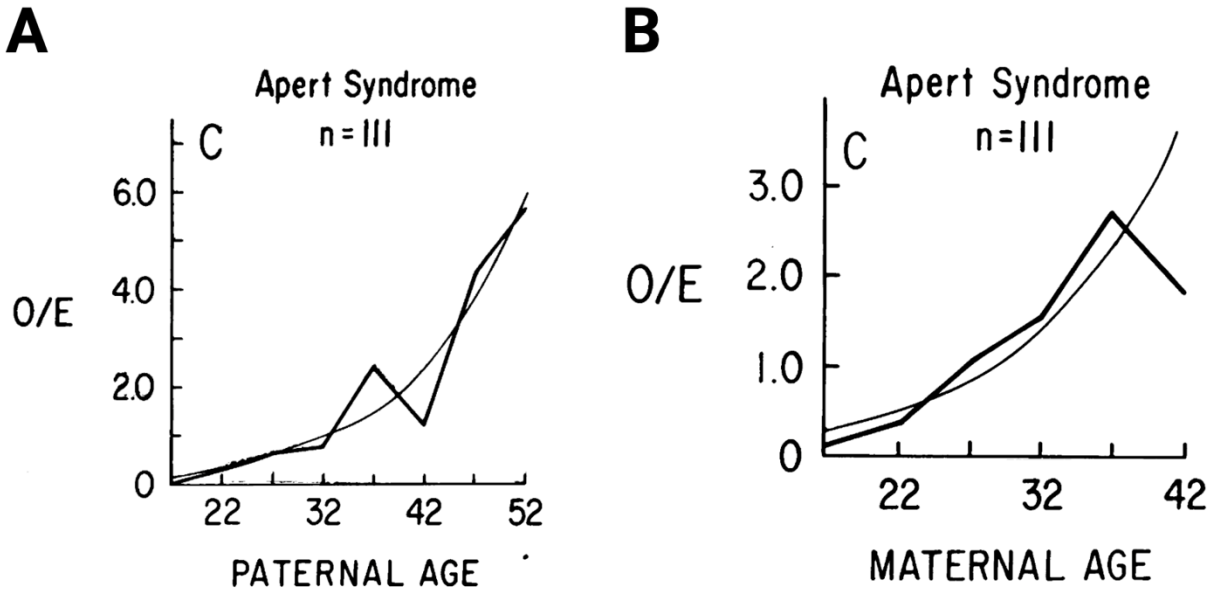


Figure 1.7: Epidemiological parental age effects in Apert syndrome. Apert syndrome is one disorder identified by Risch whose incidence increased non-linearly with the age of the parent. This was determined by comparing observed/expected (O/E) ratios (y-axis) of parental age distributions in five-year intervals (x-axis) against the general population. The thick black lines indicate observed cases whereas the thin black lines denote fitted exponential models. **(A)** The paternal O/E ratio in Apert syndrome rises sharply after 30 years and reach a high O/E ratio at older ages. **(B)** The maternal O/E ratio in Apert syndrome also rises non-linearly but is more modest than the paternal age effect. As the molecular basis of Apert syndrome had not been defined, the maternal contribution to disease incidence could not be ruled out at the time and was estimated to be ~30%. Figure from Risch et al., 1987.

While Risch could not fully tease apart parental age effects – likely because older fathers tend to have older wives, and vice versa – these results indicated that only the age of the father contributes to the sharp increase in incidence, and the increasing maternal age is artefactual. Indeed, Apert fathers, like achondroplasia fathers, were also statistically older (~4.6–10 years) than fathers in the general population (Blank, 1959; Erickson and Cohen, 1974). Importantly, the narrow mutational spectrum combined with birth estimates of Apert syndrome (Czeizel et al., 1993; Cohen et al., 1992; Tolarova et al., 1997) imply an apparent germline mutation rate that is up to three orders of magnitude higher than genome-wide estimates for base substitutions (~10^{-5}–10^{-6} vs. ~10^{-8}) (**section 1.2**).

It is instructive to also compare the characteristics of disorders whose incidence increase linearly with paternal age. One such disorder is Treacher Collins syndrome (OMIM: #154500), which is caused by loss-of-function mutations in *TCOF1* (RefSeq: NM_001135243) (The Treacher Collins Syndrome Collaborative Group, 1996). Parental origin of mutation was determined in 10 sporadic cases showing a modest paternal bias ($\alpha = 2.3$), and the mean paternal age (26.8 years) did not differ from fathers in the general Brazilian population (27.9 years) (Splendore et al., 2003). These results emphasize the distinction between a paternal bias in mutation and a statistical paternal age effect, indicating that one does not beget the other.

These distinctive characteristics thus formed the basis for a strict set of criteria shared by a small group of dominant disorders – termed Paternal Age Effect (PAE) disorders – each exhibiting: (1) an extreme bias in the paternal origin of mutation ($\alpha > 20$), (2) a statistical paternal age effect (2–7 years older than the age-matched general population), and (3) a high apparent germline mutation rate ($\geq 10^{-6}$ for individual mutations) (Goriely and Wilkie, 2012). To date, 9 disorders caused by point mutations in 6 genes have fulfilled all criteria set forth: *FGFR2* (Apert, Crouzon, and Pfeiffer syndromes), *FGFR3* (RefSeq: NM_000142; achondroplasia and Muenke syndrome, OMIM: #602849), *HRAS* (RefSeq: NM_001130442; Costello syndrome, OMIM: #218040), *PTPN11* (RefSeq: NM_002834; Noonan syndrome, OMIM: #163950), and *RET* (RefSeq: NM_020630; multiple endocrine neoplasia types 2A and 2B (MEN2A, MEN2B), OMIM: #171400 and #162300). The evidence supporting their designation as PAE disorders is summarized in **Table 1.1.**

Table 1.1: Summary of established PAE genes and associated disorders with accompanying evidence.
Adapted from Goriely and Wilkie, 2012.

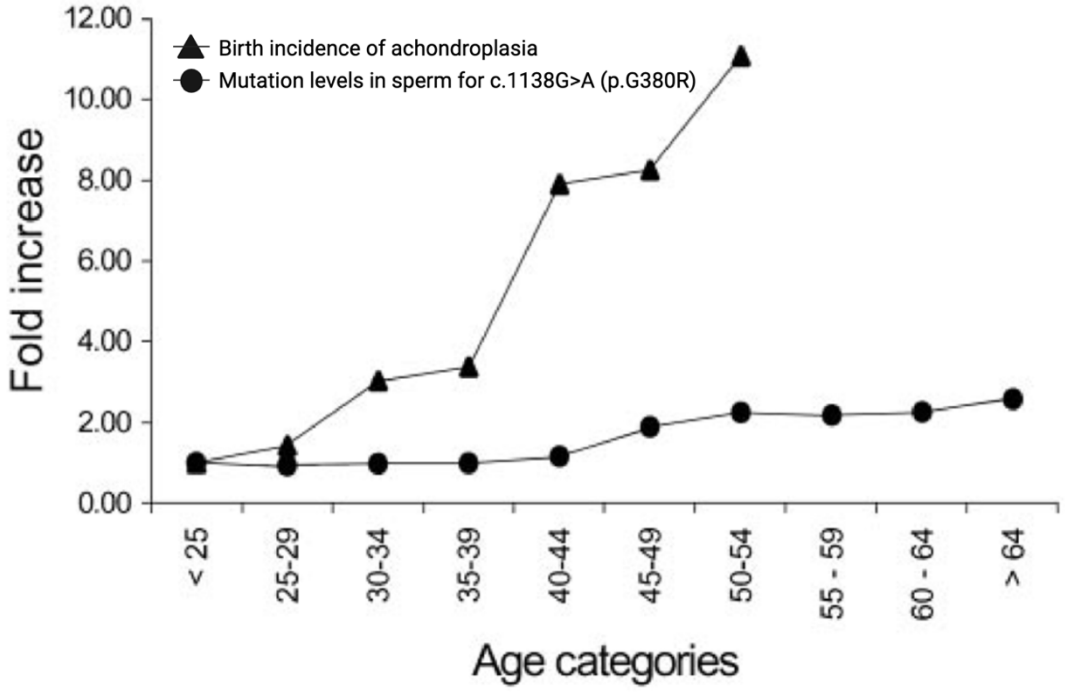
Gene	cDNA mutation (Amino acid change)	Germline disorder	Paternal mutations (n)	Maternal mutations (n)	Mean paternal age (n reports; range; excess years)	Birth prevalence	Experimental detection in male germline	References
FGFR2	c.755C>G (p.S252W) c.758C>G (p.P253R)	Apert syndrome (OMIM: #101200)	75	0	33.2 years (NR; NR; +2.5 years)	1/65,000	Sperm Testis	Moloney et al., 1996 Tolarova et al., 1997 Cohen et al., 1993 Goriely et al., 2003 Goriely et al., 2005 Qin et al., 2007 Choi et al., 2008 Maher et al., 2016
	Most commonly: p.C342Y/R/S	Crouzon/Pfeiffer syndromes (OMIM: #123500 and #101600)	22	0	34.5 years (30; 22-60; +4.1 years)	<1/65,000	Sperm Testis	Glaser et al., 2000 Goriely et al., 2003 Goriely et al., 2005 Maher et al., 2018 Rutland et al., 1995
FGFR3	c.1138G>A (p.G380R) c.1138G>C (p.G380R)	Achondroplasia (OMIM: #100800)	40	0	36.3 years (47; NR; +4.6 years) 37.2 years (31; NR; +9.7 years)	1/17,000– 1/28,000	Sperm Testis	Wilkin et al., 1998 Oriolli et al., 1995 Horton et al., 2007 Waller et al., 2008 Tiemann-Boege et al., 2002 Shinde et al., 2013
	c.749C>G (p.P250R)	Muenke syndrome (OMIM: #602849)	10	0	34.7 years (13; NR; +4.1 years)	1/30,000	Sperm	Rannan-Eliya et al., 2004 Kruszka et al., 2016 Salazar et al., 2022
	Most commonly: c.742C>T (p.R248C)	Thanatophoric dysplasia type I (OMIM: #187600)	-	-	33.6 years (30; NR; +2.6 years)	1/33,000– 1/60,000	Testis	Orioli et al., 1995 Vajo et al., 2000 Waller et al., 2008 Goriely et al., 2009 Maher et al., 2016 Salazar et al. 2022
	c.1928A>G (p.K650E)	Thanatophoric dysplasia type II (OMIM: #187601)	-	-	36.4 years (34; NR; +7.6 years)		Sperm Testis	
HRAS	Most commonly: c.34G>A (p.G12S)	Costello syndrome (OMIM: #218040)	42	2	36.2 years (18; NR; +4-7 years) 38 years (9; 29-48; +5.1 years)	1/380,000	Sperm Testis	Sol-Church et al., 2006 Schulz et al., 2007 Lurie, 1994 Gripp et al., 2011 Zampino et al. 2007 Giannoulatou et al., 2013
PTPN11	Most commonly: c.922A>G (p.N308D)	Noonan syndrome (OMIM: #163950)	14	0	35.6 years (14; NR; +6.1 years)		Testes	Tartaglia et al., 2004 Yoon et al., 2013 Maher et al., 2016 Eboreime et al., 2016
RET	Most commonly: c.1900T>C (p.C634R)	Multiple endocrine neoplasia type 2A (OMIM: #171400)	11	0	39.3 years (11; 29-57; +8.7 years)	1/40,000	Testis	Schuffenecker et al., 1997 Ralph, 2018 Mathiesen et al., 2018
	Most commonly: c.2753T>C (p.M918T)	Multiple endocrine neoplasia type 2B (OMIM: #162300)	25	1	33 years (25; 22-52; +3.1 years)	1/1,000,000	Testis	Carlson et al., 1994 Kitamura et al., 1995 Choi et al., 2012 Mathiesen et al., 2017

1.7 – Direct measurement of PAE mutations in sperm

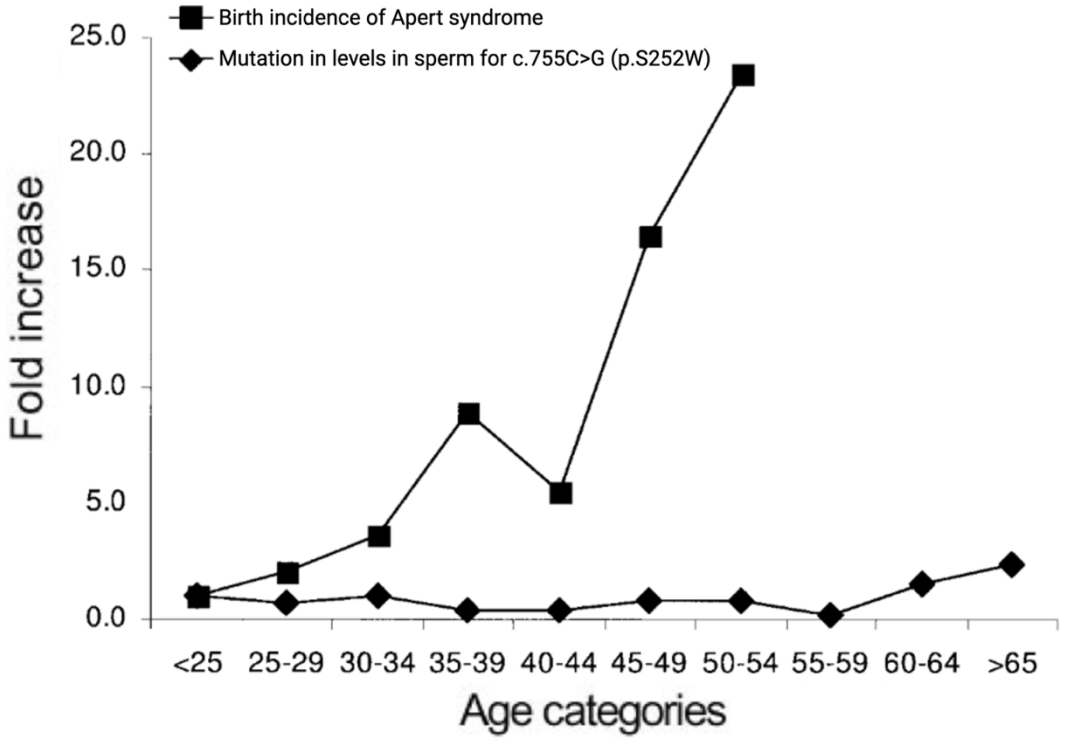
As the incidence of PAE disorders could not be explained by a linear model of mutation acquisition, other non-replicative mechanisms were proposed, including an age-related decrease in DNA polymerase fidelity or repair efficiency, the presence of mutational hotspots, and changing rates of spermatogonial cell divisions during ageing (Vogel and Rathenberg, 1975; Crow, 2000). A selective advantage conferred to mutant spermatogonia had also been proposed (Moloney et al., 1996). To characterize the underlying mechanisms of PAE mutations, numerous efforts were undertaken to directly measure these mutations in sperm. Tiemann-Boege first developed a sensitive assay (to ~1/26,000, or 38.5 mutants per million) for the *FGFR3* c.1138G>A mutation associated with achondroplasia (Tiemann-Boege et al., 2002; Wilkie, 2005). Designed to only detect a specific substitution, primer-mismatch PCR and a single *MspI* restriction digest were combined to enrich for mutant DNA, followed by allele-specific PCR to quantify mutant templates. Mutation levels in sperm were estimated by comparing against standards carrying varying amounts of spiked-in DNA from an achondroplasia cell line. 1 μg (~300,000 haploid genomes) of DNA from 118 sperm samples from men spanning a range of ages (18–80 years) was screened. Mutation levels were flat between the ages of 18–40 years and 55–80 years but experienced a ~2-fold increase between 40 and 55 years (**Figure 1.8A**). These changes in c.1138G>A mutation levels were incompatible with either a linear or non-linear model of mutation acquisition (Tieman-Boege et al., 2002). Next, Glaser used allele-specific PCR to detect the two Apert mutations *FGFR2* c.755C>G and c.758C>G in 150 ng of DNA (~45,000 haploid genomes) from 146 sperm samples (22–80 years) (Glaser et al., 2003). Mutation frequency did not increase monotonically with age, and instead was flat until 60 years when it rapidly rose (**Figure 1.8B**). In contrast, the incidence of Apert syndrome starts to inflect at a paternal age >30 years (Risch et al.,

1987; **Figure 1.7**). The results of these first two surveys in sperm had significant technical issues which limited their applicability (Wilkie, 2005). The assay developed by Tiemann-Boege was only able to identify mutations (expected to be detected at an average level of 2.1×10^{-5} based on birth prevalence) above a background level of 3.9×10^{-5} and mutations below this threshold were called at background, thereby obfuscating a discernable paternal age effect. Meanwhile, the allele-specific PCR assay used by Glaser only queried ~45,000 haploid genomes, an inadequate number of molecules to reliably detect mutations expected to be in sperm ~1/48,000 (Wilkie, 2005).

A ***FGFR3*** **c.1138G>A (p.G380R)**



B ***FGFR2*** **c.755C>G (p.S252W)**



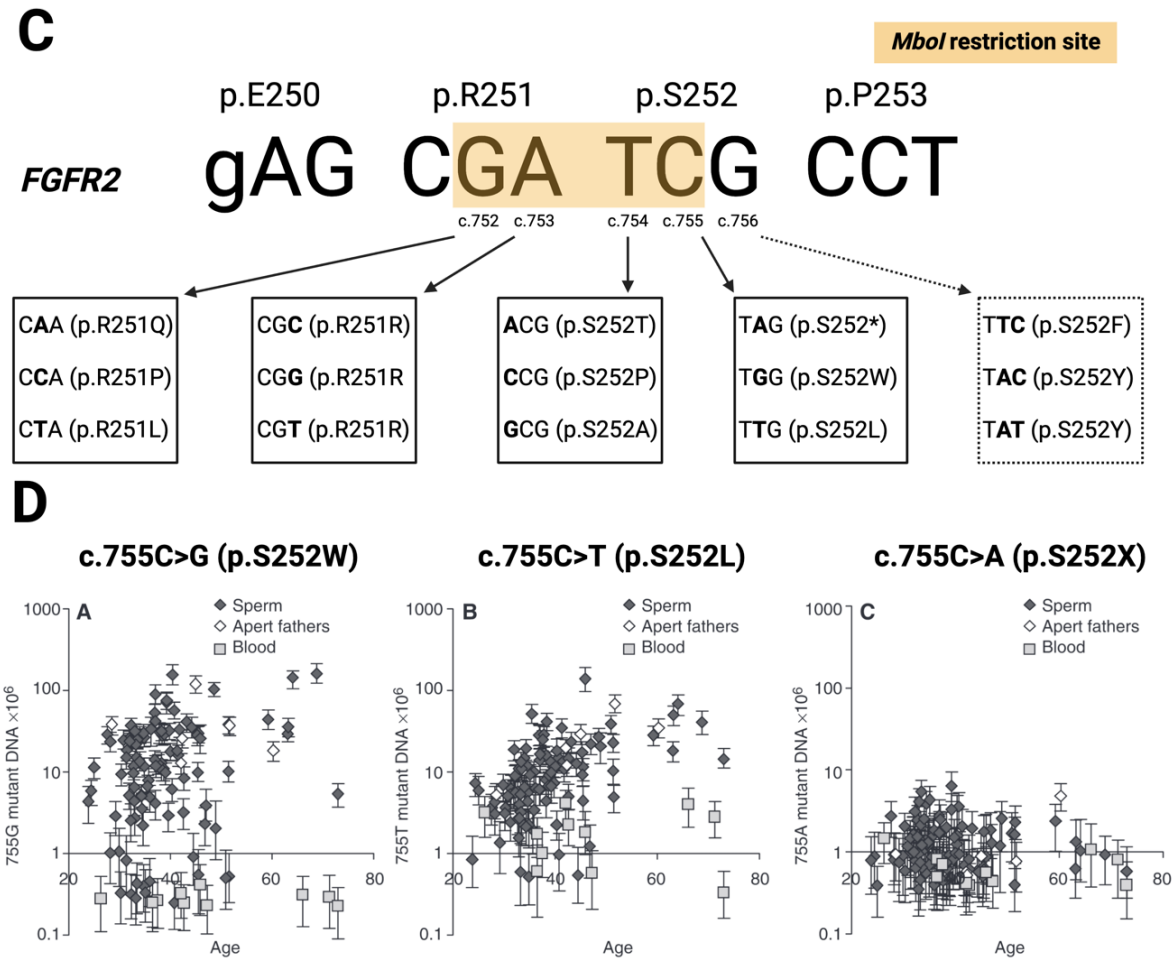


Figure 1.8: Comparison of experimentally determined PAE mutation levels in sperm relative to donor age. The results of the first two studies investigating mutation levels in sperm by **(A)** Tiemann-Boege et al. (2002) for *FGFR3* c.1138G>A and **(B)** Glaser et al. (2003) for *FGFR2* c.755C>G (p.S252W) are included here for contextual purposes. Both studies show discrepancies between observed mutation levels and reported birth incidences for achondroplasia and Apert syndrome, respectively, attributed to technical limitations (discussed in **section 1.7**). **(C)** Target space for the *FGFR2* REDPCR assay by Goriely. The *MboI* restriction site (yellow box) at *FGFR2* c.752_755 is used to cleave wild-type sequences and enrich for mutant DNA. Codons and cDNA positions are annotated above and below the nucleotide sequence, respectively, and uppercase letters denote exonic bases whereas lowercase denotes intronic. All 12 potential DNA substitutions (bolded) within the *MboI* site are indicated (solid boxes) beneath the nucleotide sequence along with the associated amino acid changes. Select pathogenic double mutations affecting the *MboI* restriction site are outlined (dashed box). Adapted from Goriely et al., 2005. **(D)** Mutation levels measured by REDPCR for all potential substitutions at *FGFR2* c.755 in sperm from healthy men (black diamonds), Apert fathers (white diamonds), and control blood (gray squares) plotted against donor age (24–73 years). Age-dependent increases observed for *FGFR2* c.755C>G (p.S252W) and c.755C>T (p.S252L) in sperm, both encoding gain-of-function mutations. No enrichment is seen for *FGFR2* c.755C>A, which encodes an early stop codon. Mutation levels for the Apert

mutation were within 1.6-fold of expectations based on birth prevalence. Figure from Goriely et al., 2003.

In contrast to these first two studies, Goriely developed an ultrasensitive technique that could reliably measure mutations to ~10^{-6} (1 mutant molecule per million) at *FGFR2* c.752_755 (Goriely et al., 2003). By combining *MboI* restriction enzyme digest (RED) to cleave wild-type DNA at the *FGFR2* c.752_755 restriction site, followed by gel purification of *MboI*-resistant sequences, PCR amplification was done to enrich for intact mutant sequences. After a second *MboI* digestion and nested PCR amplification, pyrosequencing was paired with statistical modeling for quantification. Specifically, 10 μg of DNA (~3M haploid genomes) from each sperm sample was spiked with known quantities of DNA from a proband harbouring a unique *FGFR2* mutation, facilitating quantification. REDPCR enriched and allowed for the measurement of all mutant sequences at the *MboI* site *FGFR2* c.752_755 (**Figure 1.8C**). In this initial study, the authors focused on measuring substitutions at *FGFR2* c.755, a CpG dinucleotide, which included: c.755C>G (p.S252W) the most prevalent mutation in Apert syndrome, c.755C>T (p.S252L) which had been reported in both the milder Crouzon syndrome and in healthy adults (Oldridge et al., 1997), as well as c.755C>A (p.S252X) which had not been previously reported and encoded an early stop codon. Landmark analysis of mutation levels in the sperm of 99 healthy men (average age 38.7 years; range 23.8–73 years) showed that both *FGFR2* c.755C>G and c.755C>T reached higher levels in sperm (c.755C>G, mean 2.3 x 10^{-5}, maximum 1.6 x 10^{-4}; c.755C>T, mean 1.3 x 10^{-5}, maximum 1.4 x 10^{-4}) than blood (n = 11; average age 47.7 years; range 26.3–73 years), in which mutation levels for both substitutions were measured at or below the limit of detection of 10^{-6} (c.755C>G, mean 0.3 x 10^{-6}, maximum 0.4 x 10^{-6}; c.755C>T, mean 2.0 x 10^{-6}, maximum 4.1 x 10^{-6}) (**Figure 1.8D**). Mutation levels in sperm for both *FGFR2* c.755C>G (p.S252W) and

c.755C>T (p.S252L) significantly increased with the age of the donor, and the predicted birth prevalence of Apert syndrome was within ~1.6-fold of observed values from an age-matched control population. In contrast, mutation levels for *FGFR2* c.755C>A did not reach high levels in either tissue (mean 1.4 x 10^{-6}, maximum 6.3 x 10^{-6}), nor did mutation levels increase with age. Screening of sperm from 6 unaffected Apert fathers (average age 45.0 years; range 28.5–60.3 years) additionally demonstrated that mutation levels were detected within the range of values from healthy donors (c.755C>G, mean 4.2 x 10^{-5}, maximum 1.2 x 10^{-4}) implying that these men are sampled from the general population.

By relying on a common SNP located in a nearby intronic region (rs2071616G>A), Goriely haplotyped sperm samples heterozygous for the SNP using allele-specific PCR and showed that *FGFR2* c.755C>G mutations were characterized by significantly more allelic skewing than c.755C>T. Under a neutral model of acquisition (e.g. high copy-error at this locus), statistical modelling dictated that the more prevalent *FGFR2* c.755C>G mutation would instead be expected to exhibit less skewing. Furthermore, that mutation levels for both FGFR2 c.755C>G and c.755C>T increased with the age of the sperm donor collectively suggested that only biological models incorporating selection could explain all observed patterns (Goriely et al., 2003). Supporting this hypothesis, an analysis of FGF ligand binding showed that FGFR2 p.S252W alters FGF2 binding by decreasing the rate of dissociation from the mutant receptor and resulted in a six-fold apparent increase in ligand affinity, whereas FGFR2 p.S252L exhibited wild-type kinetics (Anderson et al., 1998). As the relative strength of the FGFR2 p.S252W gain-of-function effect was reflected by the more severe phenotype of Apert syndrome, it was proposed that the *FGFR2* c.755C>G (p.S252W) mutation conferred a stronger selective advantage to the mutant SSC than

FGFR2 c.755C>T (p.S252L), explaining the higher mutation frequency observed despite the higher mutability of the C>T transition at a CpG site (Goriely et al., 2003).

Analysis of the three other nucleotides within the original *MboI* restriction site (*FGFR2* c.752_755) provided important context for the findings at *FGFR2* c.755 (Goriely et al., 2005). Firstly, like *FGFR2* c.755C, c.752G belongs to a CpG dinucleotide (**Figure 1.8C**) which provided a lens to interrogate mutation levels for the c.755C>T transition; were elevated mutation levels only a result of higher mutability, a similar pattern of mutation would be expected for the c.752G>A mutation. Secondly, as all substitutions at *FGFR2* c.753A are synonymous and expected to be selectively neutral, increased mutation frequencies for these substitutions would not be expected. Lastly, as mutations at *FGFR2* c.754T result in amino acid substitutions that are structurally distinct changes to the Apert and Crouzon mutations, enhanced FGF binding was also not expected. Indeed, analysis of 95 sperm and 11 control blood samples showed that elevated mutation levels in sperm were restricted to the two mutations associated with germline disorders: *FGFR2* c.755C>G (p.S252W) and c.755C>T (p.S252L).

1.7.1 – Multiple nucleotide substitutions are a characteristic of selfish spermatogonial selection

While investigating *FGFR2* mutations in sperm, the authors also described a child with Apert syndrome-like features carrying a heterozygous double mutation: *FGFR2* c.755_756CG>TC (p.S252F). Analysis of the parents showed that the clinically unaffected father was heterozygous for the c.755C>T (p.S252L) mutation, which he himself had inherited from his clinically unaffected mother. The presence of the additional *FGFR2* c.756G>C mutation in the proband suggests that the transversion sequentially arose *de novo* in the sperm of a healthy father on the

allele carrying the c.755C>T mutation. Supporting this interpretation, an instance of this particular tandem mutation had been detected in the sperm of an unrelated individual, in which ~20% of *FGFR2* c.755C>T sequences carried the double mutation, indicating that the *FGFR2* c.756G>C mutation was likely to have been acquired in stepwise fashion (Goriely et al., 2005).

These double mutations introduced a large aromatic side chain, similar to the canonical Apert mutation *FGFR2* p.S252W. Given this, the authors predicted that a mutation encoding tyrosine could emulate the properties of the two reported mutations but would require two separate substitutions: *FGFR2* c.755_756CG>AT or AC (**Figure 1.8C**). Retrospective analysis of the REDPCR data revealed a sperm sample carrying high levels ($>10^{-5}$) of the double mutation, *FGFR2* c.755_756CG>AC (p.S252Y). Although double substitutions are expected to independently arise extremely infrequently ($\sim10^{-11}$) (Kondrashov, 2002), multiple nucleotide substitutions resulting in rare, complex mutations could arise through sequential acquisition and become enriched in sperm due to cumulative fitness effects. This process of protein-driven selection was termed selfish spermatogonial selection (Goriely et al., 2003; Wilkie, 2005).

1.7.2 – Positive selection of mutant spermatogonia links PAE disorders and tumorigenesis

The proposal that mutant spermatogonia with gain-of-function properties are positively selected would explain the sharp increase in incidence with paternal age and high apparent mutation rate observed in several PAE disorders. This model of mutation acquisition is reminiscent of processes underlying tumorigenesis in which positive selection of mutated stem cells is an established concept (Nowell, 1976), and sequential mutations are required for oncogenic transformation (Knudson, 1971). A striking parallel was noticed where gain-of-function mutations matching those associated with several PAE disorders had been identified in human cancers. For instance,

activating mutations in *FGFR3* known to cause thanatophoric dysplasia type II (TDII; OMIM: #187601) had been identified in cell lines and primary tumor samples across multiple myeloma and bladder cancer (Chesi et al., 1997; Cappellen et al., 1999). These findings suggested that a possible consequence of mutant spermatogonia could also be the generation of small neoplasms within the testes, potentially leading to tumorigenesis (Hansen et al., 2005). To investigate a possible link between mutant spermatogonia and tumorigenesis in the testes, Hansen screened 28 testicular germ cell tumor samples for *FGFR2* mutations across regions known to harbour activating mutations associated with germline disorders: none of the testicular germ cell tumours were found to carry activating mutations in *FGFR2*.

However, testicular germ cell tumors may not be an appropriate model for understanding the connection between selfish spermatogonia and tumorigenesis. Testicular tumors are the most common malignancy in young men (15–40 years) and are thought to originate from precursor lesions of the developing testis called carcinoma *in situ* (CIS) (Beyer et al., 2012; Sonne et al., 2009). CIS cells resemble arrested gonocytes that do not differentiate into spermatogonia, but nevertheless persist in the postnatal testes. As selfish spermatogonia are thought to accumulate with paternal age, the etiology of testicular germ cell tumors is incompatible with a model of mutation acquisition that occurs over time. In contrast, spermatocytic tumours (once known as spermatocytic seminomas) are rare testicular tumors that generally present in older adults (mean age at diagnosis 53.6 years) and resemble adult spermatogonia (Eble, 1994; Rajpert-De Meyts et al., 2003; Giannoulatou et al., 2017). By sequencing mutation hotspots across *FGFR1*, *FGFR2*, and *FGFR3* in 30 spermatocytic tumours, Goriely found that two tumours harboured the same *FGFR3* c.1948A>G (p.K650E) mutation associated with thanatophoric dysplasia type II (TDII; OMIM: #187601) (Bellus et al., 2000). Five additional seminomas were shown to carry mutations

at *HRAS* p.Q61 (c.181C>A (p.Q61K) and c.182A>G (p.Q61R)) which is a commonly mutated residue in cancer, but not Costello syndrome which is caused by mutations at p.G12 (Goriely et al., 2009). Spermatocytic tumours generally exhibited elevated levels of HRAS or FGFR3 expression compared to normal testis tissue, indicating a role for activating FGFR-Ras signaling in the origin of these tumours. Strikingly, the 7 individuals with either FGFR3 or HRAS mutation-positive tumours were on average older (74.9 years) than those with mutation-negative cases (57.6 years).

The REDPCR method was then adapted to detect substitutions at the *FGFR3* p.K650 codon. REDPCR screening of 78 sperm samples (mean age 40.2 years; range 22.1–73.9 years) and 8 blood samples (mean age 52 years; range 36.5–73 years) showed that mutation levels for c.1948A>G (p.K650E) were elevated in sperm (mean 3.3×10^{-5}; maximum 2.1×10^{-4}), but not blood (mean 6.2×10^{-6}; maximum 1.8×10^{-5}) and were significantly correlated with the age of the donor (Goriely et al., 2009). Activating mutations reported in other dominant disorders, such as acanthosis nigricans and hypochondroplasia, were also detected at elevated levels in a handful of donors, whereas silent or selectively neutral mutations were not. In accordance with prior observations of mutations in *FGFR2* and Apert syndrome, mean mutation levels in sperm for *FGFR3* c.1948A>G were within 1.4-fold of the birth prevalence of thanatophoric dysplasia (~2.4×10^{-5}). Mutation levels for all substitutions were highly correlated with the number of cases of both germline disorders and reported somatic mutations in bladder cancer. Together, these data indicate that the same mechanisms underly mutations of the soma (eventually resulting in tumorigenesis), and those of the germline (resulting in heritable disorders in the next generation).

1.8 – Investigation of PAE mutation levels and distributions in the testes

Independent efforts to measure PAE mutations directly in the testes have reinforced the conclusions of the REDPCR sperm studies. By systematically dissecting donor testes across a range of ages (19–62 years), mutation frequencies and distributions for the two Apert mutations, *FGFR2* c.755C>G (p.S252W) and c.758C>G (p.P253R), were measured using bi-pyrophosphorolysis-activated polymerization (Bi-PAP-A). This technique, a highly sensitive variation of allele-specific PCR, uses a primer with a 3' end blocked by a modified nucleotide to inhibit DNA synthesis of wild-type DNA and permits quantification of individual nucleotide changes to ~10^{-6} (Qin et al., 2007; Choi et al., 2008). These studies demonstrated a striking age-related difference in mutation frequencies. For the *FGFR2* c.758C>G mutation, the average level was between 1–7.3 x 10^{-4} for older individuals (45–62 years), but several orders of magnitude lower in the testes of younger men (19 and 23 years): 5.8–6.5 x 10^{-7}. Similarly, the *FGFR2* c.755C>G mutation was detected between 6.7 x 10^{-5} and 6.8 x 10^{-4} in older testes compared to 9.4 x 10^{-7} and 1.6 x 10^{-5} in younger testes. The spatial distribution of mutational loads also differed dramatically between older and younger donors (**Figure 1.9A**). For older donors, 95% of the *FGFR2* c.758C>G mutational load was localized to just 0.5–6.3% of testes pieces whereas younger donors had a more diffuse distribution of 7.3–14.1% of pieces. Mutational loads for *FGFR2* c.755C>G showed a consistent pattern where older donors carried 95% of their mutational load in 3.7–13% of pieces as compared to younger donors with 10.4–18.2% of pieces.

Importantly, this highly concentrated distribution of mutations in older testes allowed for the statistical rejection of a hot spot model of mutation acquisition, which would predict more evenly distributed mutation levels throughout the testis. Quantitative modelling instead suggested that the high frequency and clustered distribution of *FGFR2* mutations could be explained by a germline selection model that incorporated occasional symmetric divisions of mutant

spermatogonia. In this model, mutated SSCs were estimated to undergo symmetric divisions with a probability *p* of 0.01, or one of every hundred divisions is symmetric (occurring approximately once every 4 years). Simulations with this model reproduced the observed mutational frequencies (10^{-5}–10^{-4}) and highly concentrated distribution of mutational foci in the aged testes, and could also explain the increase in disease incidence with paternal age reported for Apert syndrome; this model of germline selection could not be statistically rejected. Notably, mutation frequencies measured in testes pieces matched those found in epididymal (mature) sperm samples from the same donors (median difference of ~1.8x), supporting a view that selection likely occurs at the level of diploid premeiotic cells versus cell competition between haploid sperm (Qin et al., 2007).

Similar patterns of mutation have been found by studying the frequency and distribution of the MEN2B mutation, *RET* c.2943T>C (p.M918T) (Choi et al., 2012), the most prevalent mutation in Noonan syndrome, *PTPN11* c.922A>G (p.N308D) (Yoon et al., 2013), and the achondroplasia mutation, *FGFR3* c.1138G>A (p.G380R) (Giudicelli et al., 2008; Shinde et al., 2013). Together, these data have supported the proposed model whereby positive selection of infrequent mutation drives the age-related clonal expansion of mutant spermatogonia.

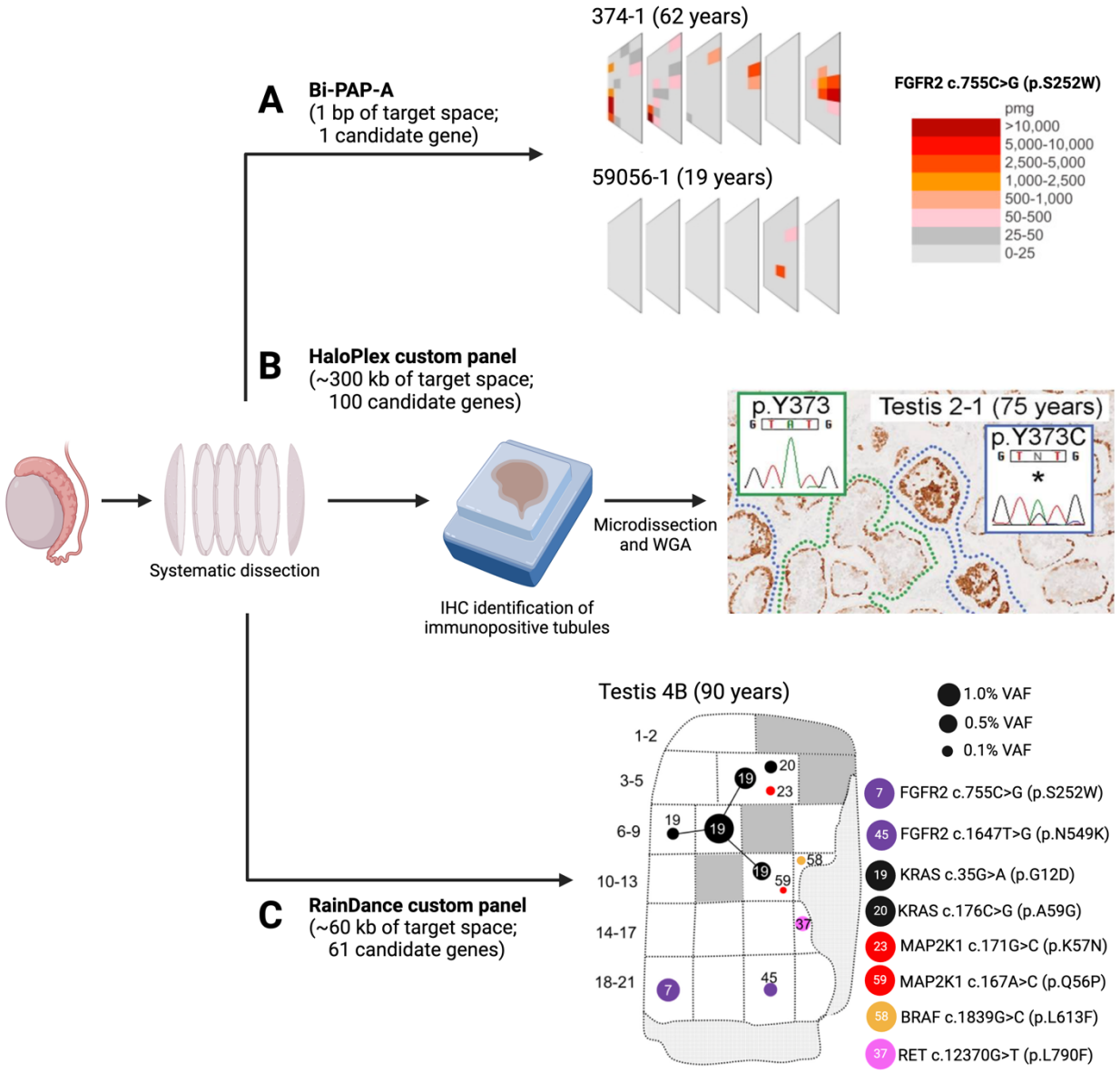


Figure 1.9: Strategies to study pathogenic DNMs directly within the testes. Systematic dissection of a testis into approximately six discrete slices. **(A)** Testis slices are biopsied into ~32 biopsies and screened using Bi-PAP-A. Here, the spatial distribution of mutation levels for the Noonan syndrome-associated mutation, *FGFR2* c.755C>G (p.S252W), is reported from a younger (19 years; top) and an older (62 years; bottom) donor. Biopsies are coloured by their mutational load (per million genomes; pmg), as indicated in the figure legend. Each testis slice from the older donor has multiple mutational foci carrying high mutation levels in a non-randomly distributed manner. Adapted from Choi et al., 2008. **(B)** Immunohistochemistry of FFPE sections permits the identification of immunopositive tubules which are characterized by enhanced MAGEA4 staining and abnormal morphology, such as multiple layers of spermatogonia along the basal lamina. Immunopositive tubules were isolated by laser capture microdissection, followed by whole-genome amplification (WGA) of DNA, and sequencing with a HaloPlex panel of >100 genes. Wild-type seminiferous tubules are outlined in green and are accompanied by a Sanger sequencing

trace, while immunopositive tubules containing the pathogenic *FGFR3* c.1118A>G (p.Y373C) mutation, associated with TD type I, are outlined in blue. Adapted from Maher et al., 2016. **(C)** Spatial distribution of selfish mutations detected in a testis section from a 90-year-old man. Mutations are represented by numbered circles, with circle size corresponding to variant allele fraction (VAF) as indicatied in the legend. Each mutation is colour-coded and labelled with its gene, nucleotide change, and resulting amino acid change. Adapted from Maher et al., 2018. Figure created with BioRender.

1.9 – Immunopositive tubules and the high-throughput discovery of pathogenic DNMs in the male germline

DNA analysis of systematically dissected testes revealed clustered foci of mutations but inherently precluded the direct visualization of mutant clones within their tissue of origin. Reasoning that there were visual correlates of these clonal expansions, Lim developed immunohistochemical (IHC) methods to screen testes sections for expression across a panel of 7 antibodies, each corresponding to either distinct states during spermatogonial development or signalling pathway activation (Lim et al., 2012). Through the staining of serial testes sections from older men (n = 6; mean age 72.8 years), Lim found infrequent seminiferous tubules that were characterized by pronounced MAGEA4 staining (a germ cell marker) and abnormal cellular morphology. Designated "immunopositive tubules," these MAGEA4+ cells often formed multiple layers along the basal lamina of seminiferous tubules and exhibited elevated FGFR3 and pAKT expression, indicative of enhanced Ras and PI3K/AKT signaling. 3D reconstruction of immunopositive tubules from serial testes sections showed that several tubules could be traced across neighbouring sections (**Figure 1.10**), extending over several millimeters of seminiferous epithelium in some cases. If immunopositive tubules extending ~5 mm along a seminiferous tubule (and potentially up to ~5 cm by connecting physically close immunopositive tubules) (**section 1.5.1**) were to represent individual clones, this would in theory produce mutant spermatozoa at a frequency of ~10^{-6}–10^{-5}, compatible with direct measurements of PAE mutations in sperm (Maher et al., 2014).

To enable the *in situ* discovery of pathogenic DNMs, Maher further developed these methods by using laser-capture microdissection to isolate immunopositive tubules from 14 donor testes (mean age 72 years; range 39–90 years) and assaying whole-genome amplified DNA using HaloPlex sequencing panels of >100 candidate genes reported in cancer and in germline disorders (Maher et al., 2016). Compared with the normal adjacent tubules, immunopositive tubules not only exhibited impaired spermatogenesis but, in the case of one elderly male (90 years) with severe testicular atrophy, also contributed to an apparent survival advantage of mutant spermatogonia. Despite the broad target space, 11 distinct gain-of-function mutations were detected across only five genes (*FGFR2*, *FGFR3, HRAS, KRAS,* and *PTPN11*), corresponding to both oncogenic and germline mutations reported in TDI (OMIM: #187600), TDII, and Apert, Crouzon/Pfeiffer, and Noonan syndromes (**Figure 1.9B**). It was noted that the majority of these mutations (8/11) resulted in perinatal lethal disorders or had only been reported in cancer, indicating that while the severity of these mutations yielded strong immunoreactivity and clonal expansion in the testes, they may be incompatible with the production of mature sperm or offspring viability. Indeed, mildly activating mutations such as the achondroplasia mutation, *FGFR3* c.1338G>A (p.G380R), were not detected despite representing the most mutable nucleotide in the genome (Bellus et al., 1995), likely because of less pronounced immunoreactivity. Together, these results demonstrated that clonal expansions of selfish spermatogonia are readily identifiable in their tissue of origin and support a model in which mutant spermatogonia are positively selected and expand one-dimensionally along seminiferous tubules. This model is consistent with *in vivo* live tracing of labeled spermatogonia which demonstrates that stem cells can become clonally enriched along the mouse seminiferous tubule over time due to neutral drift (Nakagawa et al., 2007; Klein et al., 2010).

In the most recent discovery screen of selfish mutations in the male germline, Maher paired systematic dissection of the testes (similar to the first surveys of individual PAE mutations in the testes; **section 1.8**) with high-throughput next-generation sequencing (**Figure 1.9C**) (Maher et al., 2018). In this screen, 276 biopsies from five testes (mean age 73 years; range 34–90 years) were screened using a custom RainDance panel covering 500 coding regions across 61 candidate genes (Ras-MAPK signaling pathway members, oncogenes, and genes in other signaling pathways associated with dominant disorders). Although most mutations were detected in five known PAE genes (*FGFR2, FGFR3, KRAS, PTPN11, RET)*, clonal expansions were also detected in *BRAF* (RefSeq: NM_004333)*, CBL* (RefSeq: NM_005188)*, MAP2K1* (RefSeq: NM_002755)*, MAP2K2* (RefSeq: NM_030662)*, RAF1* (RefSeq: NM_002880)*,* and *SOS1* (RefSeq: NM_005633).

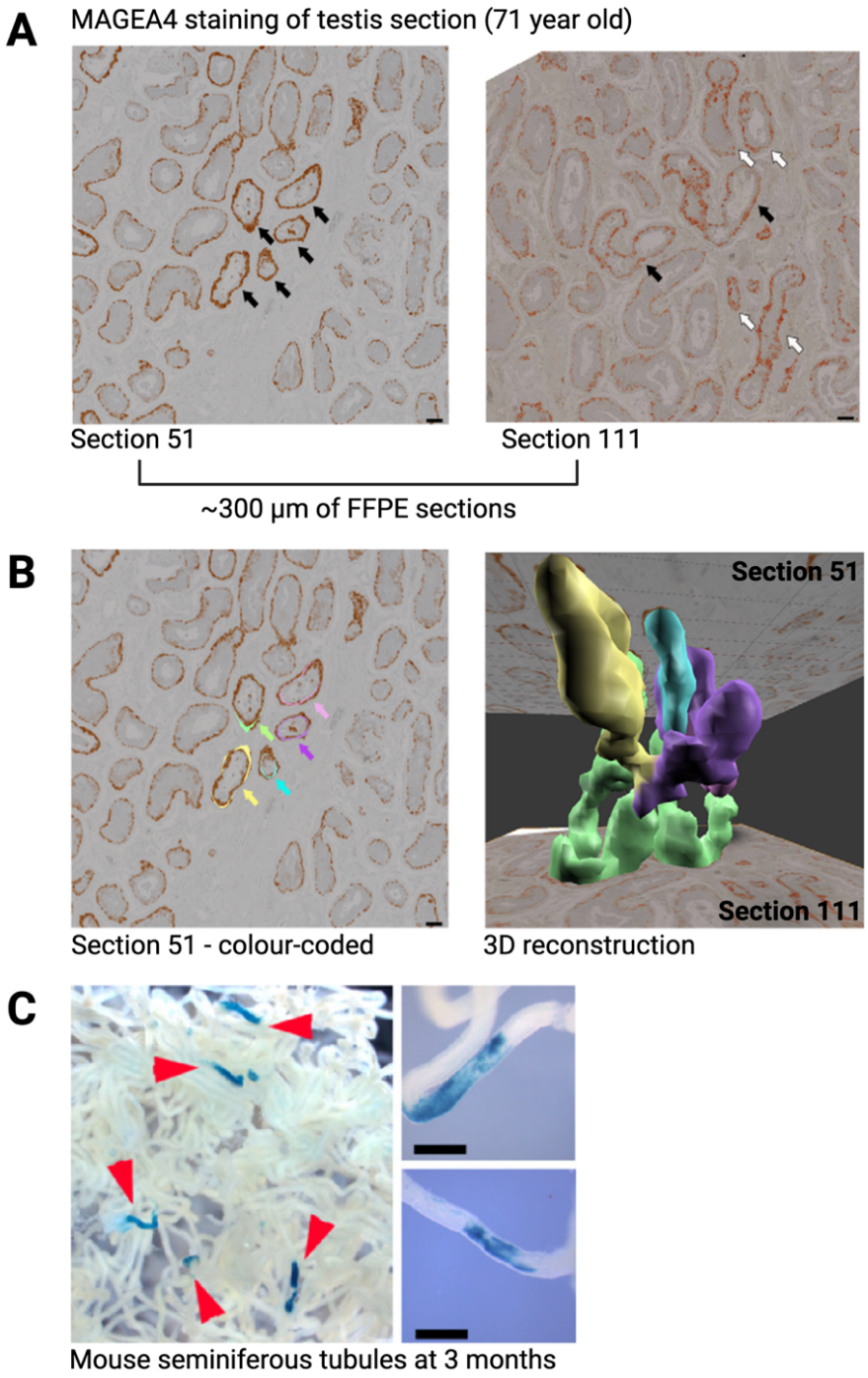


Figure 1.10: Immunopositive tubules can be traced across several millimeters, indicative of clonal expansion along seminiferous tubules. (A) MAGEA4 staining of testis sections from a 71-year-old man. Two sections (51 and 111) are shown and are separated by ~300 µm of formalin-fixed paraffin-embedded (FFPE) tissue. Black arrows indicate immunopositive tubules that were traced between sections 51 and 111, whereas white arrows indicate new immunopositive tubules to that FFPE section. **(B)** Colour-coded version of section 51 highlighting five immunopositive tubules that are 3D reconstructed between sections 51 and 111. Four of five coloured tubules (blue,

green, purple, and yellow) were shown to be contiguous. **(C)** Left: Mouse seminiferous tubules at 3 months of age with red arrows indicating tubules in which X-gal labelled spermatogonia have colonized discrete sections of the tubule, driven by neutral drift. Right: Higher magnification of individual tubules sections with expanded labelled spermatogonia. Adapted from Lim et al., 2012 and figure from. Klein et al., 2010.

1.10 – Dysregulated RTK-Ras signalling underlies selfish spermatogonial selection

Each of the identified genes exhibiting evidence of selfish selection notably encodes a component of the canonical RTK-Ras signalling cascade. The RTK-Ras cascade is activated through several receptor tyrosine kinases (RTKs) following ligand binding by growth factors on the cell membrane. One such class of growth factor is the fibroblast growth factor (FGF) family in which 22 FGF ligands associate with four FGF receptors (FGFRs) encoded by the *FGFR1, FGFR2, FGFR3,* and *FGFR4* genes (Xie et al., 2020). Canonical FGF binding to the extracellular domain of monomeric FGFRs triggers receptor dimerization, allowing for the autophosphorylation of the FGFR intracellular tyrosine kinase domains (**Figure 1.11**). This leads to the phosphorylation of the major FGFR kinase substrate, fibroblast growth factor receptor substrate 2 alpha (FRS2α), which permits the sequential recruitment of the adaptor protein, growth factor receptor-bound protein 2 (GRB2), and the guanine nucleotide exchange factor, Son of Sevenless (SOS), which facilitates the exchange of GDP for GTP on Ras. FRS2α also interacts with other key proteins such as Casitas B-lineage lymphoma (CBL), GRB2-associated binding protein 1 (GAB1), and Src homology region 2 domain-containing phosphatase-2 (SHP2; encoded by the *PTPN11* gene) which help regulate signal transduction (Teven et al., 2014).

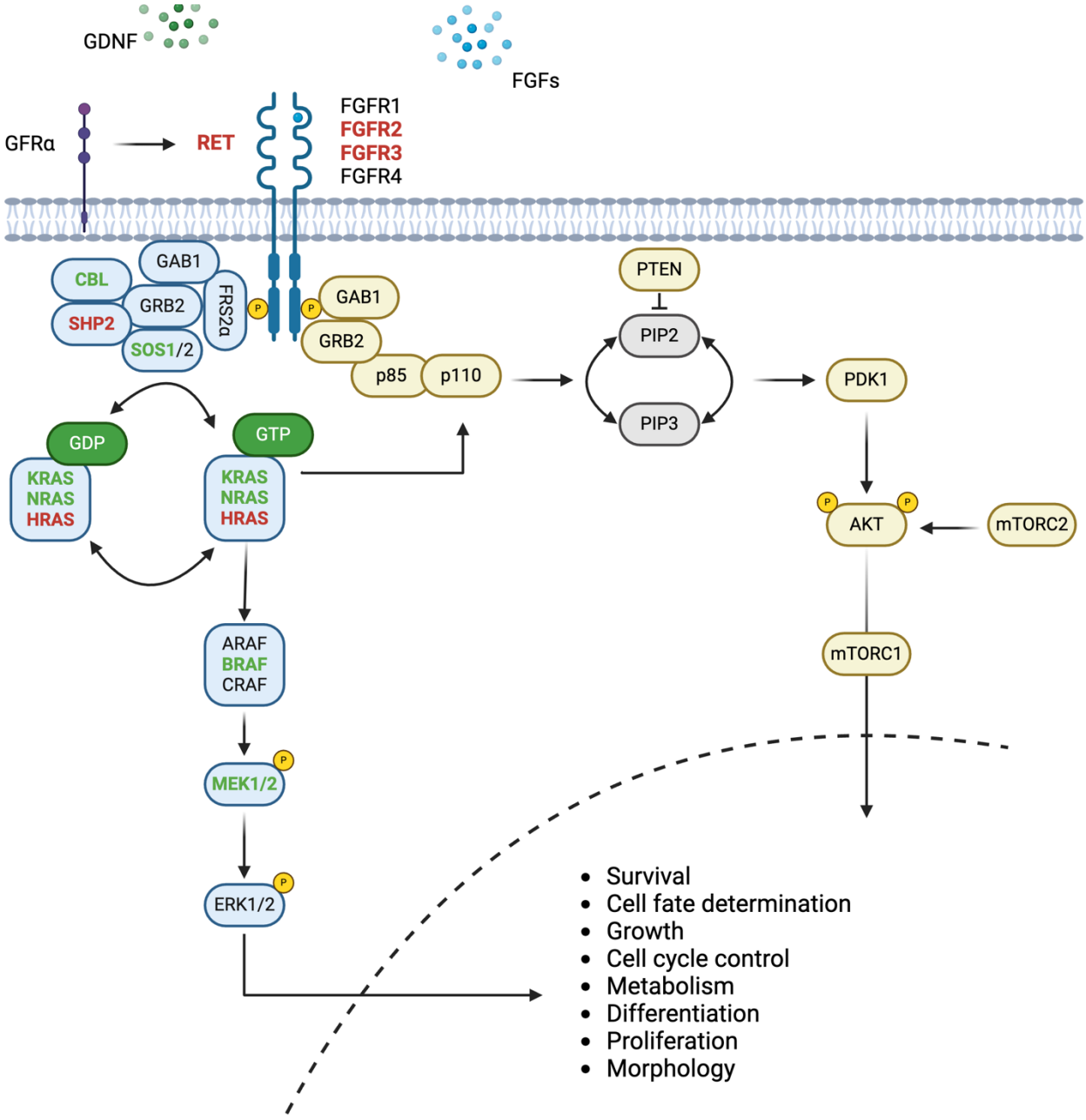


Figure 1.11: Overview of the RTK-Ras signaling cascade. This figure illustrates the key components of the RTK-Ras pathway, highlighting those known to be associated with PAE disorders (red) and those exhibiting evidence of selfish spermatogonial selection (green). Relevant signaling pathways downstream of RTK activation are divided into two groups: MAPK (blue) and PI3K/AKT (yellow). A detailed description of how signals propagate down these pathways is available in **section 1.10**. Adapted from Punekar et al., 2022; Goriely and Wilkie, 2012. Figure created with BioRender.

In addition to the FGFR family, the rearranged during transfection (RET) receptor is another RTK that binds a range of growth factors such as the glial cell line-derived neurotrophic factors (GDNFs) when in complex with GDNF-family receptor-α (GFRα), a ligand-binding co-receptor found along the cell membrane (Arighi et al., 2005). The GDNF and GFRα complex facilitates RET dimerization, allowing for RET autophosphorylation of tyrosine residues within the intracellular domain. This process mediates downstream signaling by providing docking sites for a host of key adaptor proteins, such as FRS2 and SHC (Src homology 2 domain containing). These proteins facilitate the recruitment of GRB2 (which can also directly bind to phosphorylated RET) and SOS, leading to the subsequent activation of Ras (Takahashi, 2022). Activated GTP-bound Ras recruits the Raf protein kinases to the cell membrane where binding facilitates Raf activation through phosphorylation, relief of autoinhibition, and Raf dimerization (Terrell and Morrison, 2019). Activated Raf phosphorylates MAP/ERK kinase (MEK), which in turn phosphorylates extracellular signal-regulated kinase (ERK) at key tyrosine and threonine residues. This results in ERK activation and its subsequent translocation to the nucleus, where it exerts pleiotropic effects on transcription factors to regulate gene expression and cellular proliferation and survival (Dillon et al., 2021).

As well as facilitating MAPK signal transduction, the adaptor protein GRB2 also regulates the phosphoinositide 3-kinase (PI3K)/AKT signalling cascade by complexing with GAB1. Following localization to regions along the cell membrane enriched with phosphatidylinositol 4,5-biphosphate (PIP2), phosphorylated tyrosine residues on GAB1 recruit the p85 regulatory subunit of PI3K (Wöhrle et al., 2009; **Figure 1.11**). Phosphorylation of PIP2 by the p110 catalytic subunit of PI3K generates phosphatidylinositol-3,4,5-triphosphate (PIP3) which in turn binds to proteins with a pleckstrin homology (PH) domain, such as 3-phosphoinositide-dependent protein kinase-1

(PDK1) and AKT. Incidentally, GTP-bound Ras can also directly bind to the p110 subunit and promote PI3K catalytic activity (Castellano and Downward, 2011). Once localized at the cell membrane, PDK1 phosphorylates AKT at p.T308 to activate the kinase, whose activity is further enhanced following phosphorylation at p.S473 by mechanistic target of rapamycin complex 2 (mTORC2) (He et al., 2021; Krygowska and Castellano, 2018). Phosphorylated AKT (pAKT) is then able to exert pleiotropic effects regulating cell survival, growth, proliferation, and differentiation.

Biochemical analyses of PAE mutations show that all are activating mutations conferring a range of gain-of-function properties such as altered ligand and substrate affinity, specificity, and residence times (Anderson et al., 1998; Ibrahimi et al., 2001; Yasuda et al., 2012; Bocciardi et al., 1997), as well as ligand-independent receptor activation (Naski et al., 1996; Cranston et al., 2006). These mutations interfere with normal skeletal and bone development by perturbing chondrocyte and osteoblast morphogenesis, which underlies the shared phenotypes of craniosynostosis, syndactyly, and short stature across PAE disorders (Wang et al., 1999; Wang et al., 2005; Yin et al., 2008; Di Rocco et al., 2014). While not universal, some conditions such as Noonan and Costello syndromes are additionally characterized by cardiovascular abnormalities, intellectual disability, and developmental delay. These phenotypes have been recapitulated in mouse models and are attributed to dysregulated Ras-ERK activity (Araki et al., 2004; Wang et al., 2005; Lee et al., 2014). Importantly, these shared phenotypes can be partially to completely rescued through MEK and ERK inhibition *in vivo*, indicating that the effects of these mutations are substantially mediated through the MAPK pathway (Shukla et al., 2007; Yin et al., 2008; Chen et al., 2010; Lee et al., 2014).

The role of dysregulated PI3K/AKT signaling in the etiology of PAE disorders is less clear and has been the subject of conflicting findings. Rosenberger first demonstrated that fibroblasts from individuals with Costello syndrome (CS) exhibited elevated levels of pAKT and altered PI3K signaling dynamics compared to controls (Rosenberger et al., 2008). By contrast, levels of both phosphorylated MEK and ERK did not differ, suggesting a specific AKT-mediated effect. A later study by Carpentieri investigated the mechanisms underlying the metabolic dysfunction in Costello syndrome showed that Costello fibroblasts were characterized by altered glucose uptake and energy homeostasis (Carpentieri et al., 2022). However, contrary to earlier findings, Costello fibroblasts had similar levels of pAKT to controls and instead showed elevated levels of phosphorylated AMPK and p38. Although the application of a dual p38/PI3K inhibitor could rescue these phenotypes, the specific contribution of the PI3K/AKT pathway to Costello syndrome etiology remains unclear. Additional evidence supporting a potential role for PI3K/AKT signaling in PAE disorders comes from studies on Noonan syndrome with multiple lentigines (NSML; formerly known as LEOPARD syndrome). NSML is clinically related to Noonan syndrome, is also defined by mutations in *PTPN11* (predominantly at p.Y279 and p.T468), and is notably characterized by hypertrophic cardiomyopathy (Tartaglia et al., 2006). Through *in vivo* administration of a specific AKT inhibitor, Wang demonstrated that AKT antagonism could ameliorate the associated cardiomyopathy in a NSML mouse model, indicating that at least some of the clinical manifestations observed in certain PAE disorders (such as the cardiovascular complications in Noonan syndrome) may be mediated by the PI3K/AKT pathway (Wang et al., 2017).

1.11 – RTK-Ras signalling regulates SSC homeostasis

Just as dysregulated RTK-Ras signaling contributes to the multi-faceted phenotypes of PAE disorders, this cascade is also a critical regulator of SSC homeostasis. GDNF is an essential growth factor for spermatogenesis *in vitro* and the concentration of GDNF determines the balance of murine SSC cell fate (Meng et al., 2000). Low levels of the ligand favour SSC differentiation (and the depletion of the SSC pool), and overexpression of GDNF results in clusters of undifferentiated spermatogonia, primarily through the activation of PI3K/AKT signaling (Meng et al., 2000; Lee et al., 2007). FGF2 is also considered an indispensable growth factor for the initiation and maintenance of spermatogenesis in culture. FGF2 regulates SSC self-renewal through the phosphorylation of mitogen-activated protein kinase kinase 1 (MAP2K1, also known as MEK1) and collaborates with AKT to modulate essential transcription factors (e.g. *BCL6B* and *ETV5*) to drive SSC proliferation (Oatley et al., 2006; Ishii et al., 2012; Takashima et al., 2015).

In addition to being essential components for initiating and maintaining murine spermatogenesis *in vivo*, the FGF growth factors also play a role in dictating the organization of the SSC niche. For example, GFP-labeling of SSCs has shown that undifferentiated spermatogonia preferentially localize near the vasculature (Yoshida et al., 2007) and that the density of these GFRα1-expressing cells linearly correlates with Fgf5 gene dosage *in vivo* (Kitadate et al., 2019). FGF5, secreted by lymphatic endothelial cells, promotes the proliferation of GFRα1+ SSCs in a concentration-dependent manner, upregulating genes involved in stem cell maintenance and downregulating genes associated with differentiation. Moreover, SSCs express FGF receptors and internalize FGF5 – suggesting direct consumption of the growth factor – and other FGFs (such as FGF4 and FGF8) were also shown to contribute additively to stem cell regulation in a similar manner. This led to the proposed model of mitogen competition wherein SSCs compete for limited FGFs, which regulates SSC density and distribution along the seminiferous tubules in an open

niche system, providing a mechanism for SSCs to sense the local concentration of mitogens and adjust their fate to maintain homeostasis.

It is also important to consider how these regulatory mechanisms may change over time. The ageing process appears to impact the SSC niche, potentially altering the abundance or balance of signalling factors – such as those in the RTK-Ras pathway – that maintain SSC homeostasis. For example, Ryu followed germ-cell labelled mice and found that fertility declined after 12 months, with all males becoming infertile by 20 months, which corresponded to reductions in testes weight and spermatogenesis (Ryu et al., 2006). By transplanting testes cells from young (2–12 months), ageing (14–18 months), and old (20–24) months) mice, a dramatic age-related decline in the number of stem cells recovered in the recipients 3 months later was noted. However, as older SSCs retained their ability to produce normal-sized stem cell colonies (suggestive of their self-renewal capabilities), it was suggested that the decline in stem cell activity was a function of a declining stem cell niche. Accordingly, RT-PCR analysis of ageing testes showed that both GDNF and GFRα1 decreased in expression over time. As the SSC niche ages, the selective pressure to resident stem cells may increase (e.g. decreasing availability of signalling molecules), potentially creating an environment in which SSCs carrying selfish mutations are able to outcompete their peers to survive.

1.11.1 – Selfish mutations exert their effects through the MAPK and PI3K/AKT pathways

Experimental models have provided valuable insights into the effects of dysregulated RTK-Ras signalling on SSCs. For example, Martin used lentiviral transduction to introduce the Apert syndrome-associated FGFR2 p.S252W mutation into murine SSCs (Martin et al., 2014). Relative to wild-type SSCs, FGFR2 p.S252W SSCs were characterized by higher expression of *Etv5*, a

transcription factor known to be associated with SSC self-renewal, and were additionally shown to be more sensitive to *in vitro* growth factors FGF2 and GDNF, consistent with prior findings showing that the Apert mutation altered FGF2 binding by decreasing the rate of ligand dissociation (**section 1.7**; Anderson et al., 1998). Under restricted ligand conditions, FGFR2 p.S252W SSCs induced greater phosphorylation of both MAPK and AKT than wild-type SSCs, suggesting involvement of both signalling cascades. Moreover, co-culturing fluorescently tagged FGFR2 p.S252W and wild-type SSCs allowed for the characterization of cell competition between the two populations: under reduced FGF culture conditions, FGFR2 p.S252W SSCs were able to outcompete wild-type cells independent of the initial ratio of starting mutant and wild-type cells. Transplantation of cultured wild-type and FGFR2 p.S252W SSCs into busulfan-treated mice extended these observations by showing that mutant FGFR2 p.S252W SSCs formed significantly more spermatogonia colonies *in vivo*, indicative of greater stem cell activity.

SSCs transfected with the HRAS p.G12V mutation – known to cause a rarer and more severe form of Costello syndrome (Giannoulatou et al., 2013) – have been shown to be able to self-renew *in vitro* without standard FGF growth factors in media, and also show enhanced proliferative capabilities relative to wild-type SSCs (Lee et al., 2009). The application of PI3K, AKT, and MEK inhibitors abrogated this proliferative advantage, indicating that this effect was mediated by both PI3K/AKT and Ras-MAPK signalling. Additionally, the transplantation of HRAS p.G12V SSCs into congenitally infertile mice resulted in germ cell colonies and functional spermatogenesis, but notably led to the formation of seminomatous tumours in all recipient mice. Just as strongly activating mutations were found in spermatocytic tumours in aged men (**section 1.7.2**; Goriely et al., 2009), and overexpressed GDNF induced the formation of clusters of

undifferentiated spermatogonia (**section 1.10.1**; Meng et al., 2000), overactive Ras was proposed to underly the excessive SSC self-renewal resulting in the *in vivo* tumor formation.

In a separate study, Yamada induced germline-specific expression of HRAS p.G12V through tamoxifen-inducible Cre-mediated recombination at the endogenous HRAS locus (Yamada et al., 2019). Through long-term lineage tracing, a modest (~9%) increase in the allele fraction of HRAS p.G12V sperm was observed between 4 and 12 months (as opposed to no change in wild-type mice), but there was no increase in the proportion of HRAS p.G12V offspring. By performing *in vitro* cell competition assays to compare wild-type and HRAS p.G12V SSCs over 5 passages, the authors concluded that HRAS p.G12V SSCs do not have a competitive advantage over wild-type SSCs because the parental line was not completely replaced over a span of ~4–5 weeks.

The seemingly conflicting results between the studies by Lee and Yamada may be explained by gene dosage and expression levels of mutant HRAS p.G12V. Lee used lentiviral vectors to introduce HRAS p.G12V into SSCs prior to transplantation, potentially resulting in multiple copy insertions. In contrast, the Cre-recombination method used by Yamada to activate a single copy of HRAS p.G12V at its endogenous locus would likely result in more physiological expression levels. This difference in approach could account for the enhanced proliferation and tumor formation observed by Lee, suggesting a dose-dependent effect. It's additionally important to note that both studies have limitations in fully recapitulating the Paternal Age Effect. The study by Yamada, while more physiologically relevant, was conducted in young mice over a relatively short period (up to 14 months), which may neither recapitulate an ageing murine SSC niche (in which fertility began to decrease only after 12 months) (Ryu et al., 2006), nor may it adequately model the decades-long process in humans. Furthermore, neither study fully accounted for other

age-related changes in the SSC niche such as the declining expression of GDNF and GFRα (Ryu et al., 2006). This could potentially alter the competitive dynamics between wild-type and mutant SSCs, a factor not fully explore by either HRAS p.G12V study. Taken together, these data indicate selfish mutations likely exert their effect through activation of both the Ras-MAPK and PI3K/AKT pathways. However, the interplay between these related cascades and their relative contributions, as well as how they interact with the ageing SSC niche requires additional investigation.

1.11.2 – scRNAseq reveals complex signalling networks regulate SSC homeostasis in humans

The mechanisms governing SSC homeostasis in humans remain incompletely understood, as evidenced by the inability to maintain long-term *in vitro* culture of human SSCs under mouse culture conditions (Zheng et al., 2014; Medrano et al., 2016). scRNAseq has implicated a more complex model of spermatogenesis than previously thought (**section 1.5.3**) and has also provided unprecedented insights into the signaling pathways that may regulate SSC homeostasis in humans. For example, Guo used scRNAseq to identify differentially expressed genes between undifferentiated and differentiated spermatogonial populations. In addition to confirming the importance of established pathways in murine SSC homeostasis (e.g. FGF and GDNF), scRNAseq highlighted members of the WNT, BMP, PDGF, INTEGRIN/TSPAN, NOTCH/HES1 pathways as upregulated in undifferentiated SSCs (**Figure 1.12**) (Guo et al., 2017).

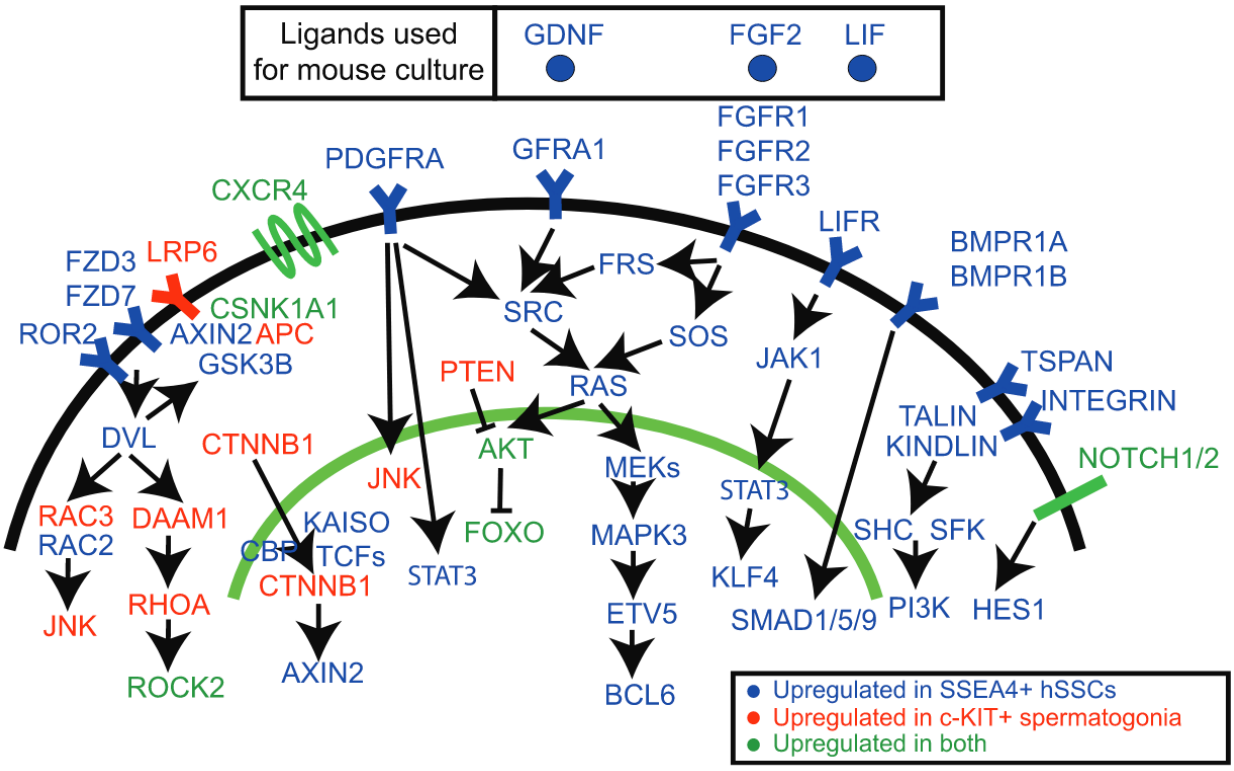


Figure 1.12: scRNAseq of spermatogonia implicates a complex network of signalling pathways that may be involved in SSC homeostasis. Boxed at the top of the figure are key growth factors used in mouse SSC culture. scRNAseq identified differential expression of numerous members of the above signaling cascades; genes are colour-coded by the population of spermatogonia in which they are upregulated. Figure from Guo et al., 2017.

1.12 – Positive selection of mutations in somatic tissues

Selfish mutations of the male germline have also been identified as age-related clonal expansions in other tissue contexts. Seborrheic keratoses (SK) are benign skin tumors that present as verrucous brown plaques ranging up to several centimeters in diameter and are highly prevalent in older adults, with 80–100% of people >50 years displaying SKs (**Figure 1.13A**) (Hafner et al., 2006). By sequencing SK lesions from 65 older adults (mean age 60.1 years), it was found that the majority (57%) harboured activating mutations in *FGFR3* that are routinely identified in the germline disorders TDI and TDII, as well as severe achondroplasia with developmental delay and acanthosis nigricans (SADDAN; OMIM: #616482) (**Figure 1.13B**) (Hafner et al., 2007).

Individuals with lesions carrying mutations in *FGFR3* were statistically older than those with *FGFR3* wild-type SKs (64.4 vs. 54.5 years). It was noted that despite having the highest genome-wide mutation rate, the achondroplasia mutation *FGFR3* c.1138G>A (p.G380R) was not represented amongst SKs and it was proposed that the selective advantage conferred by this mildly activating mutation was not sufficient to induce skin tumors (Hafner et al., 2006). These results are strikingly reminiscent of observations made while studying selfish spermatogonia where individuals with *FGFR3* and *HRAS*-mutant spermatocytic tumors were older than those with wild-type tumors, and only the strongest activating mutations (associated with the most severe clinical phenotypes) resulted in discernable, immunoreactive tubules (Goriely et al., 2009; Maher et al., 2016). Extending the link between selfish mutations and positive selection of somatic mutations across tissue contexts, oncogenic mutations in *HRAS* (including mutations both at p.G12 and p.Q61) as well as other members of the MAPK and PI3K/AKT were found in a minority of SK lesions (**Figure 1.13B**), with evidence of elevated levels of phosphorylated AKT and ERK compared to normal skin (Hafner et al., 2010).

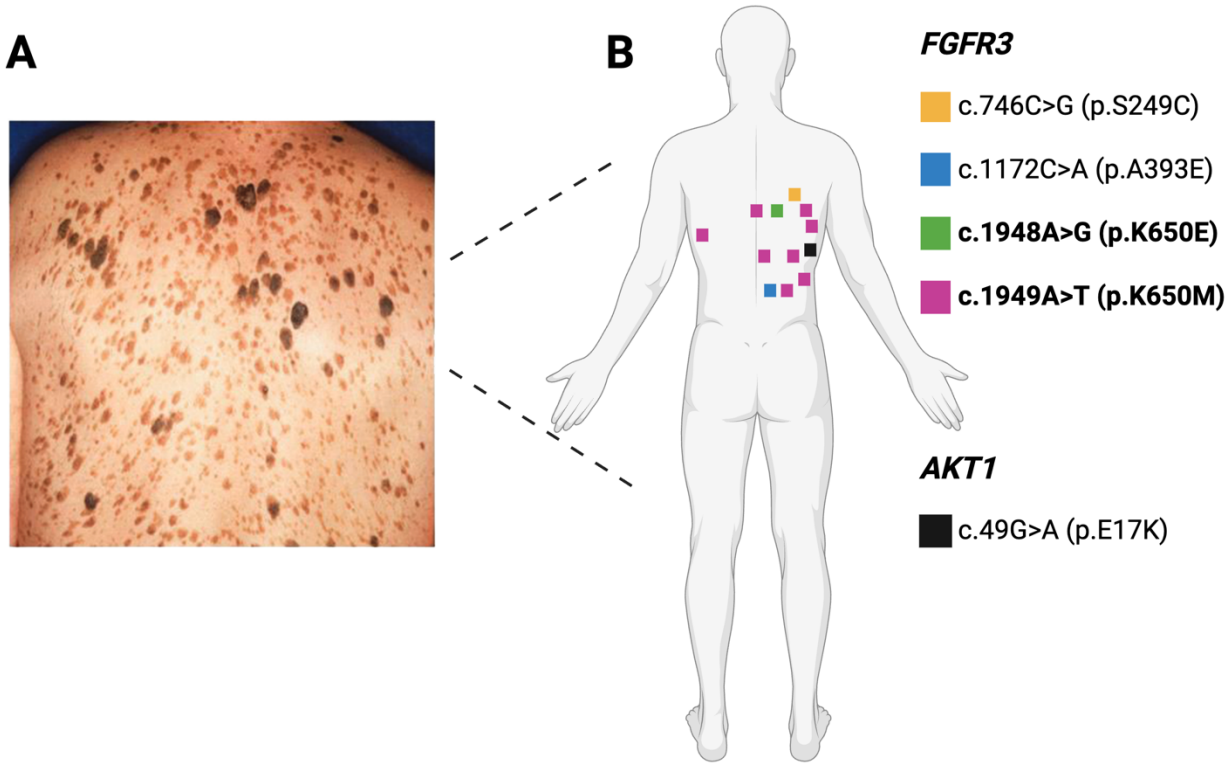


Figure 1.13: Seborrheic keratoses harbour activating mutations in the RTK-Ras pathway. **(A)** Schematic representation of the human body showing the distribution of activating mutations in *FGFR3* and *AKT1* found in seborrheic keratoses (SKs). The coloured squares indicate the locations and types of mutations detected. Bolded mutations have been reported as germline mutations: *FGFR3* p.K650E in TDII and p.K650M in SADDAN, while unbolded variants correspond to mutations reported in both bladder (*FGFR3*) and breast cancers (*AKT1*). Adapted from Hafner et al., 2010. **(B)** The inset image is an illustrative example of SK lesions on a patient's back (note: this image does not correspond to the mutation distribution shown in the schematic). Adapted from Hafner et al., 2006. Figure created with BioRender.

In fact, positive selection of somatic mutations in healthy tissues is a ubiquitous process and has been extensively studied in the blood (Genovese et al., 2014; Jaiswal et al., 2014), skin (Martincorena et al., 2015), bladder (Lawson et al., 2020), esophagus (Martincorena et al., 2018), colorectum (Lee-Six et al., 2019), endometrium (Moore et al., 2020), liver (Brunner et al., 2019), and stomach (Coorens et al., 2024). Mutations accumulate in cancer-associated genes and, through clonal expansion, can increase the risk of oncogenic transformation by providing a more

susceptible pool of preneoplastic cells (Martincorena and Campbell, 2015). However, positive selection of stem cells harbouring driver mutations in healthy tissues does not exclusively drive oncogenesis and has been associated with numerous other diseases such as atherosclerosis (Fuster et al., 2017), inflammatory bowel disease (Olafsson et al., 2020), and chronic liver disease (Ng et al., 2021). Although each tissue type exhibits a unique landscape of driver mutations, the same clock-like mutational signatures SBS1 and SBS5 (**section 1.2.2**) that account for all germline mutations are also identified in all healthy and diseased somatic tissues, albeit to varying degrees (Moore et al., 2021).

1.13 – Challenges of studying the human germline

The same mutational processes appear to operate in both the germline and soma, as evidenced by similar mutational signatures across tissues. However, germline mutation rates have been reported to be more than an order of magnitude lower than those of the soma, implying different degrees of genome maintenance (Milholland et al., 2017; Moore et al., 2020). While protection against mutation may reduce the risk of transmission, this feature presents technical challenges when studying the germline.

The headwinds of a lower germline mutation rate are exacerbated by the architecture of the testes where – despite forming a tightly packed and convoluted network – each seminiferous tubule is its own functional unit rather than a single continuous tubule (**section 1.5.1**) (Nakata et al., 2015). In contrast, stem cell niches such as the bone marrow compartment allow for the regular migration of hematopoietic stem cells (HSCs) to and from peripheral blood (Wright et al., 2001), which could allow mutant hematopoietic stem cells (HSCs) to clonally expand relatively unrestricted. This is exemplified by findings along distinct segments of the gastrointestinal tract

where the oesophagus and intestine of 60-year-olds are occupied by ~60% and ~1% clones, respectively (Coorens et al., 2024). These differences are thought to be driven in part by differences in epithelial architecture: the oesophagus has a continuous epithelium whereas the intestinal tract is characterized by individual crypts which limit the expansion of clones beyond individual crypts (Coorens et al., 2024; Olafsson et al., 2020). Indeed, studies of clonal hematopoiesis demonstrate that mutations can be commonly detected at high VAFs (>0.035) in healthy older individuals (~5.6% of individuals in their 60s, with mutational prevalence increasing each decade), >2 orders of magnitude higher than the maxima of PAE mutation levels detected in sperm (Jaiswal et al., 2014; **section 1.7**). Efforts to overcome the dilutive effects of screening bulk sperm by systemically dissecting testes biopsies have improved sensitivity, but even in the most comprehensive discovery screen to date, the average VAF of prioritized variants identified by next-generation sequencing was ~0.54% VAF (range 0.06–2.95% VAF) (Maher et al., 2018).

1.14 – Aims

Activating mutations leading to the dysregulation of the RTK-Ras signaling pathway have been established as a central feature of selfish spermatogonial selection. However, in theory, any gene or signaling pathway involved in the regulation of SSC homeostasis could be subject to positive selection (Goriely and Wilkie, 2012; Maher et al., 2018). The aims of this thesis are therefore to probe the landscape of DNMs that are positively selected in the male germline.

As known selfish mutations may represent the "tip of the iceberg," a rationally designed target space of genomic sites will be designed from the hypothesis that selection is reflected by recurrent DNMs displaying characteristics consistent with known selfish mutations (**section 3**). The suitability of rhAmpSeq to multiplex these prioritized targets for the discovery of pathogenic

DNMs in aged testes will be assessed (**section 4**). Finally, an established ultra-sensitive technique will be adapted to investigate whether ultra-rare mutations are detectable in sperm (**section 5**).

Chapter 2 – Materials and Methods

2.1 – Ethics approval

Ethical approval for the use of anonymized human testicular and sperm tissue was provided by the Oxfordshire Research Ethics Committee A (C03.07: Receptor tyrosine kinases and germ cell development: detection of mutations in normal testis, testicular tumors and sperm) and the North East – Newcastle & North Tyneside 1 Research Ethics Committee (22/NE/0025; NEW_MUTS: New Mutations and Male Germline Development).

2.2 – Materials

2.2.1 – Testes samples

Fresh adult testis samples were obtained with the appropriate consent from post-mortem cases at the John Radcliffe Hospital (Oxford, UK) or from a commercial tissue bank, Medical Solutions plc (Peterborough, UK) (**Table 2.1**). Testes from two donors (Tes9 and 30) had been previously dissected into transverse slices and stored at -80°C. Genomic DNA had been previously extracted and quantified from Tes4 biopsies (Maher et al., 2018).

Table 2.1: Overview of testes samples

Testis sample ID	Source	Donor age (years)	Biopsies assayed in this thesis (n)
4	Medical Solutions plc (Peterborough, UK)	90	4
9	John Radcliffe Hospital (Oxford, UK)	73	84
30	John Radcliffe Hospital (Oxford, UK)	63	48
33	John Radcliffe Hospital (Oxford, UK)	76	83

2.2.2 – FFPE samples

Formalin-fixed paraffin-embedded (FFPE) testis blocks were obtained from a 63-year-old man following orchidectomy for a hernia. Four tissue blocks (BBP2167, BBP2169, BBP2171, and BBP2173) containing normal testicular tissue were stored at room temperature prior to sectioning.

2.2.3 – Sperm and blood samples

Sperm and blood samples had previously been obtained from healthy volunteers (~24.8–75.9 years) and genomic DNA was stored at -20°C (**Table 2.2**).

Table 2.2: Overview of sperm and control samples. Samples are organized by age and tissue type.

Sample ID	Tissue Type	Age (Years)
1236	Blood	24.76
1240	Blood	26.41
1063C	Blood	39.06
1008	Blood	45.51
1197	Blood	45.51
1239	Blood	57.01
987	EBV	49.00
1047	Sperm	27.22
1050	Sperm	27.33
1016	Sperm	27.67
1037	Sperm	27.98
935	Sperm	28.72
716	Sperm	29.94
901	Sperm	30.26
921	Sperm	31.49
1015	Sperm	31.57
923	Sperm	31.64
1010	Sperm	32.1
950	Sperm	32.12
1030	Sperm	32.49
920	Sperm	32.58
1027	Sperm	33.24

904	Sperm	33.51
1039	Sperm	33.75
888	Sperm	33.88
1052	Sperm	34.81
1055	Sperm	35.05
1046	Sperm	35.11
925	Sperm	35.35
915	Sperm	35.39
1038	Sperm	35.76
413	Sperm	35.86
406	Sperm	36.38
1025	Sperm	36.93
1023	Sperm	36.99
1056	Sperm	37.10
869	Sperm	37.25
885	Sperm	37.34
1048	Sperm	37.40
917	Sperm	37.84
903	Sperm	38.28
779	Sperm	38.38
1012	Sperm	38.57
898	Sperm	38.84
868	Sperm	39.31
872	Sperm	40.63
958	Sperm	41.08
1057	Sperm	41.28
860	Sperm	41.62
565	Sperm	41.70
1022	Sperm	41.85
409	Sperm	42.09
1040	Sperm	42.14
780	Sperm	42.39
942	Sperm	42.47
931	Sperm	43.16
960	Sperm	43.21
361	Sperm	43.66
1026	Sperm	43.79
1033	Sperm	43.92
916	Sperm	44.12
306	Sperm	45.76
307	Sperm	45.78
726	Sperm	46.34
963	Sperm	46.89
836	Sperm	46.93
879	Sperm	47.09

1031	Sperm	47.59
763	Sperm	48.03
784	Sperm	49.56
766	Sperm	50.27
1028	Sperm	50.63
424	Sperm	52.91
948	Sperm	53.73
526	Sperm	54.17
809	Sperm	55.92
788	Sperm	56.21
1124	Sperm	58.57
799	Sperm	59.48
1018	Sperm	59.88
278	Sperm	60.59
294	Sperm	71.77
545	Sperm	75.86

2.3 – Immunofluorescence

2.3.1 – Tissue sectioning

Formalin-fixed paraffin-embedded (FFPE) tissue blocks (**section 2.2.2**) were pre-chilled at -20°C prior to being cut on a microtome into ultrathin sections (~5 μm). Using forceps, cut sections were floated on the surface of a heated water bath (~40–45°C) to remove wrinkles. Sections were then mounted onto glass slides and were allowed to dry overnight at room temperature.

2.3.2 – Double immunofluorescent labelling

Slides were heated at 55°C for 1 hour, then deparaffinized in three 5-minute changes of HistoClear (National Diagnostics). Tissue sections were rehydrated through graded ethanol changes (100% (n =3), 95% (n =2), 70% (n = 1), 5 minutes each) and rinsed with running distilled water for 5 minutes. Heat-induced epitope retrieval was performed in 10 mM sodium citrate buffer (pH 6) for 30 minutes. Tissues were permeabilized in phosphate buffered saline (PBS) (Gibco) with 0.1% Tween (PBST) for 1 hour and non-specific binding was subsequently blocked with SuperBlock (Thermo Fisher Scientific) for 1 hour. Tissues were incubated overnight with two primary antibodies (one

candidate gene and one known germ cell marker; **Table 2.3**) at 4°C, after which slides were washed with PBST. Fluorophore-conjugated secondary antibodies were applied for 2 hours at room temperature in darkness. 4',66-diamidino-2-phenylindole (DAPI) was used for nuclear counterstaining. Both primary and secondary antibodies were diluted in SignalBoost Immunoreaction Enhancer Kit (MilliporeSigma). After a final wash with PBST, slides were mounted using Vectashield (Vector Laboratories).

2.3.3 – Imaging of IF slides

Images were acquired using a Zeiss LSM 900 confocal microscope with 405 nm (DAPI), 488 nm, and 555 nm laser lines using Smart settings. Images processing and analysis were performed using ImageJ software.

Table 2.3: Antibodies used in immunofluorescent experiments.

Antibody type	Antibody	Host	Dilution	Manufacturer
Primary	PACS1	Rabbit	1:200	NovusBio
Primary	PPP2R5D	Rabbit	1:200	NovusBio
Primary	SMAD4	Rabbit	1:200	Abcam
Primary	MAGEA4	Mouse	1:1000	Prof. Spagnoli
Primary	GFRα1	Goat	1:400	R&D Systems
Secondary	Anti-Rabbit IgG (Alexa Fluor 488)	Chicken	1:500	Invitrogen
Secondary	Anti-Goat IgG (Alexa Fluor 555)	Donkey	1:500	Invitrogen
Secondary	Anti-Mouse IgG (Alexa Fluor 594)	Donkey	1:500	Invitrogen
Nuclear stain	DAPI	N/A	1:2000	Roche

2.4 – Preparation of genomic DNA from testicular tissue

2.4.1 – Dissection

Fresh post-mortem testes were dissected in chilled PBS (Gibco) on ice. After removing the extraneous tissues (e.g. the tunica albuginea, rete testis, and epididymis), the testis was cut into transverse slices (~3–5 μm thick) which were either frozen at -80°C or formalin-fixed for paraffin embedding. Testis slices from two donors (Tes9 and 33) were thawed in chilled PBS (Gibco) and dissected into 88–90 biopsies per testis (~20–60 mm^3). Tes30, slice G (Tes30G) had previously been dissected into 51 biopsies (35–100 mm^3) and genomic DNA was extracted and stored at -20°C (Sahar Ghaffari).

2.4.2 – DNA extraction and quantification

To improve yield, genomic DNA was extracted from individual testes biopsies using a modified protocol which combined the use of steel beads and the TissueLyser III (Qiagen) to pulverize tissue and the QIAamp DNA Mini Kit (Qiagen). DNA was eluted in 50 μL of elution buffer, quantified using the Qubit dsDNA BR assay Kit and fluorometer (Thermo Fisher Scientific), and stored at -20°C.

2.5 – rhAmpSeq

2.5.1 – Panel design

A rhAmpSeq panel was constructed to multiplex 20 primer sets targeting genomic regions across 11 candidate and 9 control loci in 15 genes totaling ~4.1 kb. Target-specific (TS) primers were designed to provide overlapping coverage of all loci of interest and to have a melting point ≥60°C, recommended by the manufacturer (Integrated DNA Technologies) for optimal RNase H2 activity.

TruSeq adaptors were appended to the 5’ ends of both forward and reverse TS primers to allow for indexing, as well as a 5 bp random degenerate unique molecular index (dUMI) to enable bioinformatic error correction. dUMIs were customized to be any random nucleotide except for the complementary base to the five bases immediately upstream of the target-specific primers. An RNA base and 3' blocking moiety, which prevented primer annealing and extension until cleavage by activated RNase H2, were appended to the 3’ end of each TS primer (oligonucleotide sequences are listed in **Table 2.5**). rhPCR primers were reconstituted and diluted into separate forward and reverse multiplexed primer pools at a final concentration of 250 nM per primer unless otherwise indicated.

2.5.2 - Serial dilution of patient-derived DNA

Genomic DNA (~5–200 ng) from 13 probands (11 diagnosed with known PAE disorders, 2 with a candidate disorder; **Table 4.1**) were mixed with ~1.7 μg of DNA derived from control blood to dilute the original heterozygous mutations to initial VAFs ranging from 0.001–0.06. This mixture underwent four rounds of 2-fold serial dilution into genomic DNA from control blood. This resulted in five distinct mutation levels for each sample, excluding the blood-only control. For example, a mutation starting at 0.01 VAF would be tested at 0.01, 0.005, 0.0025, 0.00125 and 0.000625 VAFs.

2.5.3 - RNase H2-dependent PCR (rhPCR)

PCR amplification was carried out in a volume of 20 μL using 2X rhAmp Genotyping Master Mix (Integrated DNA Technologies), forward and reverse primer pools at a final concentration of 25 nM per primer, and 50 ng of genomic DNA. To query enough molecules (100 ng of DNA; ~30K

molecules), two separate 20 μL PCR reactions were pooled following amplification, representing one technical replicate. PCR cycling was performed using a T100 Thermal Cycler (Bio-Rad) and consisted of a 10-minute RNase H2 activation step at 95°C, followed by 8 cycles of denaturation at 95°C for 15 seconds and annealing/extension at 60°C for 4 minutes, followed by an RNase H2 heat inactivation step at 99.5°C for 15 minutes. Each sample was assayed in three technical replicates (3 x 100 ng of DNA), constituting 6 separate rhPCR reactions. Each 40 μL pooled rhPCR reaction was purified using 1.5X AMPure XP beads (Beckman Coulter) in a 100 μL mixture, and DNA was eluted in a final volume of 8 μL.

2.5.4 – Library preparation

PCR barcoding was carried out in a 20 μL reaction using 2X iProof MasterMix (Bio-Rad), the entire 8 μL rhPCR eluate, and custom P5/P7 primers, each containing Illumina TruSeq adapters and an 8 bp sample-specific barcode (**Table 2.5**), at a final concentration of 0.5 μM. PCR cycling was performed at 98°C for 30 seconds, and 22 cycles of 98°C for 10 seconds, 60°C for 30 seconds, and 72°C for 30 seconds, followed by a final extension step of 72°C for 2 minutes.

Sequencing libraries were accordingly constructed by pooling one technical replicate from each sample in equimolar amounts. Pooled libraries were purified using 1.5X AMPure XP beads (Beckman Coulter) and the final library size and concentration were measured with the High Sensitivity D1000 ScreenTape on the 4200 TapeStation System (Agilent) and Qubit dsDNA HS Kit (Thermo Fisher Scientific), respectively. Libraries were diluted to a final concentration of 10 nM and couriered to a third-party sequencing service (Novogene) for ultra-deep sequencing on a NovaSeq instrument (Illumina). Demultiplexed sequencing files were downloaded from Novogene.

2.6 – SMAD4 REDPCR

2.6.1 – Titration-reconstruction of MS *SMAD4* DNA

Genomic DNA from up to two individuals with Myhre syndrome (heterozygous for *SMAD4* c.1498A>G and c.1499T>C) were used in a titration-reconstruction experiment in which 10 µg (~3 million haploid genomes) from control blood was mixed with a dilution series of Myhre syndrome DNA across six mutation levels (**Table 2.4**). Each sample was assayed in technical replicates using genomic DNA from the blood of two healthy donors.

Table 2.4: Reference table for the *SMAD4* REDPCR serial dilution experiments. For each mutation level tested, the associated amount of genomic DNA spiked in from each Myhre syndrome (MS) proband, as well as the number of haploid genomes, is outlined.

MS mutation level	MS spike-in DNA (ng)	MS spike-in DNA (haploid genomes)
0	0	0
10^{-6}	0.02	3
3×10^{-6}	0.06	10
10^{-5}	0.2	30
3×10^{-5}	0.6	100
10^{-4}	2	300

2.6.2 – *NsiI* RED and mutant selection

The DNA concentration of each biological sample (blood or sperm) (**Table 2.2**) was measured in three technical replicates using the Qubit dsDNA HS assay Kit (Thermo Fisher Scientific). Initial proof-of-concept experiments (**sections 5.2.2 and 5.2.3**) were performed without replicates and without the addition of the *NsiI*-resistant genomic integrant; all other aspects of the protocol remained the same.

Triplicate samples, each containing 10 µg of genomic DNA mixed with 0.6 ng of NsiI-resistant genomic integrant (final concentration of 3×10^{-5}), were subjected to double restriction digest using 10X CutSmart buffer (New England Biolabs), 40 units of NsiI-HF (New England

Biolabs), and 40 units of PstI-HF (New England Biolabs) in a final volume of 100 µL for 4 hours at 37°C. Each sample was prepared in in technical triplicates. Wild-type sequences at SMAD4 c.1494-1499 were digested into two fragments of 844 and 484 bp, whereas mutant sequences at SMAD4 c.1494_1499 were left intact at 1328 bp. Each restriction digest was gel electrophoresed overnight at 4°C flanked by two lanes containing the λ DNA-BstEII digest ladder (New England Biolabs) in a 1% TAE agarose gel without ethidium bromide. In the morning, λ DNA-BstEII digest ladder lanes were excised and stained with ethidium bromide (1:20,000 dilution) for 20 minutes. Ladder lanes were visualized on a UV transilluminator and the 1264 and 1371 marker bands were injected with a syringe containing loading dye for identification. Following replacement back into the original gel, a gel slice was excised between the two 1264 and 1371 bp markers from each sample lane, and gel purified in a final volume of 25 µL using the Gel DNA Recovery Kit (Zymo Research).

2.6.3 – PCR amplification of *NsiI*-resistant DNA

To amplify intact mutant sequences at SMAD4 c.1494–1499, an initial PCR (PCR1) amplification to produce a 278 bp amplicon was performed in a final volume of 40 µL using the full 25 µL of eluate, 5X Q5 Reaction Buffer and 1 unit of Q5 High-Fidelity DNA Polymerase (New England Biolabs), 200 µM dNTPs, and 0.5 µM of each SMAD4 PCR1 primer (**Table 2.5**). PCR cycling was performed at 98°C for 2 minutes, 25 cycles of 98°C for 10 seconds, 68°C for 30 seconds, and 72°C for seconds, and a final extension at 72°C for 2 minutes. To deplete remaining wild-type DNA, an additional restriction digest was performed by mixing the full PCR1 product using 10X CutSmart Buffer (New England Biolabs) and 20 units of NsiI-HF (New England Biolabs) in a final volume of 100 µL and incubated for 1 hour at 37 °C.

A second, nested PCR (PCR2) generating a 164 bp fragment (208 bp including CS adaptors) was performed in a final volume of 75 µL using 5 µL of the prior NsiI restriction digest, 10X DreamTaq Buffer and 1 unit of DreamTaq DNA Polymerase (Thermo Fisher Scientific), 100 µM of dNTPs, and 0.1 µM of each SMAD4 CS-tagged PCR2 primer, which each had common sequence (CS) adaptors appended to the 5' end to enable barcoding (**Table 2.5**). PCR2 cycling conditions were 95°C for 5 minutes, followed by 20 cycles of 95°C for 30 seconds, 64°C for 30 seconds, and 72°C for 30 seconds, followed by a final extension step at 72°C for 5 min.

2.6.4 – Library preparation and next-generation sequencing

To prepare libraries for Illumina sequencing, a barcoding PCR was performed in a final volume of 20 µL using 2X iProof HF Master Mix (Bio-Rad) and 0.4 µM of Access Array barcodes (Fluidigm). The barcoding PCR cycling conditions were as follows: 98°C for 30 seconds, 8 cycles of 98°C for 10 seconds, 68 °C for 30 seconds, 72 °C for 30 seconds, and 72°C for 2 min. Libraries were pooled in equimolar amounts and digested with 1 unit of *NsiI*-HF (New England Biolabs) for 1 hour at 37°C. The entire digested product was gel electrophoresed on a 2% TAE agarose gel. *NsiI*-resistant DNA was excised and gel purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research) in a final volume of 20 µL. Library concentration and size were measured with the Qubit dsDNA HS Kit (Thermo Fisher Scientific) and the 4200 TapeStation System (Agilent). Libraries were sequenced on a MiSeq platform (Illumina) using custom primers at final concentrations of 0.5 µM (read 1: CS1; read 2: CS2; index: CS2rc) (**Table 2.5**).

2.6.5 - Engineering a *NsiI*-resistant genomic integrant by CRISPR-Cas9

A 20 bp single-guide RNA (sgRNA) targeting SMAD4 c.1488_1504 (**Table 2.5**) was cloned into the pSpCas9(BB)-2A-GFP (PX458) plasmid by leveraging dual BbsI sites, as previously described (Doench JG et al., 2016; Ran FA et al., 2013), and the construct was confirmed by Sanger sequencing. The construct was electroporated into RPE-1 FRT/TR cells (Ximbio) by mixing three biological replicates of 2.5 x 10^5 RPE-1 FRT/TR cells with 0.4 μg of construct, 1 μg of single-stranded oligodeoxynucleotide (ssODN) donor, and P3 Primary Cell Full Electroporation Buffer (Lonza) in a final volume of 20 μL. The ssODN was designed to disrupt the PAM site at c.1482–1484, introduce an *MboI* site (GATC) at c.1491–1494 (which would be used to screen clones), and abolish the NsiI site at c.1494 by a conservative p.L498F change (**Figure 2.1**). The EA-104 program on the 4D-Nucleofector System (Lonza) was used to electroporate cells, which were then re-plated in 1 mL of pre-warmed Dulbecco's modified Eagle's medium (DMEM; Thermo Fisher Scientific) containing 10% fetal bovine serum (FBS; Sigma), 1% of L-glutamine (Thermo Fisher Scientific), 1% Pen-Strep (Thermo Fisher Scientific), 1% sodium pyruvate (Thermo Fisher Scientific), and 1% non-essential amino acids (NEAA; Thermo Fisher Scientific). Cells were allowed to recover for 2 days in humidified incubators with 5% CO_2 at 37°C. Cells were then trypsinized with 0.25% Trypsin-EDTA (1X) (Thermo Fisher Scientific) and single-cell sorted into 96-well plates for GFP+ by fluorescence-activated cell sorting (FACS) on a Sony MA900 machine.

Cells were cultured for an additional 2 – 4 weeks in DMEM media mix. To screen clones for NsiI-resistance, genomic DNA was extracted using the QuickExtract Kit (Lucigen) and PCR amplified using 10X DreamTaq Buffer and 1 unit of DreamTaq DNA Polymerase (Thermo Fisher Scientific), 100 μM of dNTPs, and 0.1 μM of each SMAD4_integrant primer in a final volume of 25 μL (**Table 2.5**). PCR2 cycling conditions were 95°C for 5 minutes, followed by 30 cycles of

95°C for 30 seconds, 64°C for 30 seconds, and 72°C for 30 seconds, followed by a final extension step at 72°C for 5 min. Amplified product was then digested in a volume of 20 µL using 10X Buffer 3.1 (New England Biolabs) and 1 unit of NsiI (New England Biolabs) for 1 hour at 37°C. Restriction digests were run on a 3% TAE agarose gel and visualized on a Gel Doc XR+ (Bio-Rad). NsiI-resistant clones were expanded in T75 flasks (Thermo Fisher Scientific) for 1 week. Cells were trypsinized and genomic DNA was extracted using the Blood & Cell Culture DNA Kit (Qiagen) and eluted in a final volume of 50 µL.

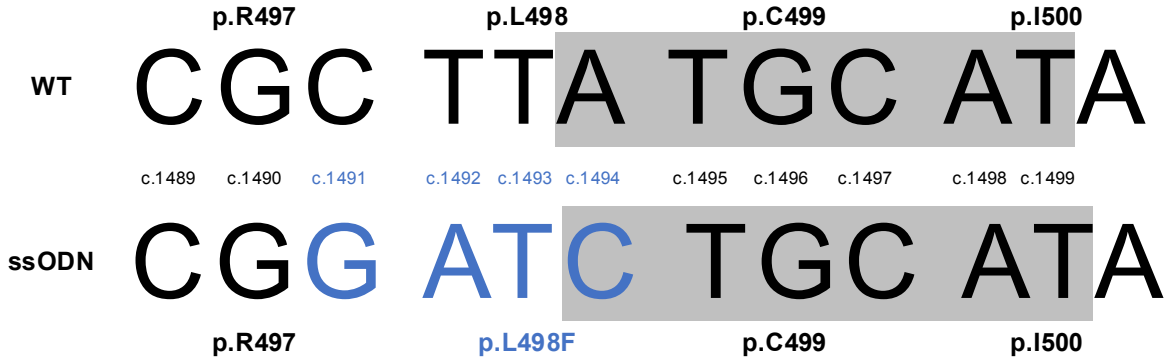


Figure 2.1: Original CRISPR-Cas9 design to generate a *NsiI*-resistant genomic integrant. A portion of the ssODN (bottom; **Table 2.5**) is shown, which was designed to introduce an *MboI* site (blue; GATC) at c.1491–1494, and thereby abolish the *NsiI* site (grey box) at c.1494 through a conservative p.L498F substitution. The p.R497 codon was to remain unchanged.

2.7 – Bioinformatics analyses

2.7.1 – Variant annotation of DNMs

Prioritized recurrent DNMs from **section 3**, as well as variants identified in the rhAmpSeq discovery screen (**section 4**), were annotated using the Ensembl Variant Effect Predictor (McLaren et al., 2016) according to the following RefSeq transcript IDs: FGFR2: NM_000141, RET: NM_020975, HRAS (NM_005343), *PACS1* (NM_018026), *PPP2R5D* (NM_006245), *SMAD4* (NM_005359), *PACS2* (NM_001100913), *ADNP* (NM_001282531), *MAP2K1* (NM_002755),

MECP2 (NM_004992), *NAA10* (NM_003491), *PPP1CB* (NM_002709), PTPN11 (NM_002834), *CDK13* (NM_003718), *CSNK2A1* (NM_177559), *FGFR3* (NM_000142), *FOXP1* (NM_001349338), *CTBP1* (NM_001328), *SRCAP* (NM_006662), *ALG13* (NM_001099922), *GNB1* (NM_002074), *GRIN2B* (NM_000834).

2.7.2 – Sequencing data processing

Demultiplexed files were processed using Amplimap v.0.4.19, a bioinformatic pipeline that employs a custom Snakemake workflow to automate processing of sequencing data (Koelling N et al., 2019), with parameters "pileups -ncores 10". Reads were trimmed and quality control was performed using Trimmomatic v.0.39 and FastQC v.0.11.9. Reads shorter than 20 bp were discarded and primers were removed. Reads were aligned to the reference genome (GRCh38/hg38) using BWA-MEM v.0.7.17-r1188 and post-alignment processing was done with SAMtools v1.12. Pileups were generated using SAMtools mpileup v.1.17, retaining only bases with a minimum Phred quality score of 30. All rhAmpSeq data from Novogene. as well as the sequencing data from the REDPCR survey of 57 blood and sperm samples (**section 5.2.7**), were processed by Dr. Stephen Bush.

2.7.3 – UMI correction

Bioinformatic error correction was performed using the UMI-tools package (Smith T et al., 2017). Each 5 bp UMI was extracted from the start of both read pairs. FASTQs were then processed using "group" mode (--method=unique), clustering reads that shared the exact same UMI and mapping location. Consensus sequences were generated if there was ≥50% agreement, otherwise an "N"

was assigned; any consensus with ≥5% 'N' across its length was discarded. UMI correction was performed by Dr. Stephen Bush.

2.7.4 – Statistical prioritization of low frequency variants

Following initial filtering for sample failure (requiring median coverage ≥1000x), binomial modelling was performed for all genomic positions with median coverage ≥5000x across retained samples. At each position, all three possible non-reference changes (e.g. A>C, A>G, A>T) were independently tested. Samples containing spiked-in mutations (i.e., serial dilutions as well as control testis biopsies in the pilot study) were excluded from the background error-rate calculation. Multiple hypothesis testing was accounted for using a Bonferroni correction, and variants were only called if they attained a corrected $p < 0.05$ and were supported by ≥5 variant reads in at least 2 technical replicates. To filter out systematic artefacts such as error-prone loci, base calls that were present in a blood control or within serial dilution samples at non-spiked positions were excluded. Calls with VAFs ≥3% were excluded as they were more likely to represent inherited SNPs rather than true *de novo* events. These scripts can be found in this online repository: https://github.com/rspencertong/

Table 2.5: Oligonucleotide sequences. Organized by the methods in which they are used. Lower case letters indicate universal sequencing adapters (**section 2.5**, TruSeq; **sections 2.6**, Common Sequence (CS)). For primers used in **section 2.5**, underlined degenerate bases through the IUPAC nucleotide code system represent 5 bp dUMIs. TS primers are followed by a 3' RNA base ("r") and blocking moiety ("/3SpC3/") which is enzymatically cleaved by activated RNase H2. For primers used in all sections, "N" is used to represent sample-specific barcodes which are attached during library preparation. Locked nucleic acids are indicated by "+N" (where N can be any base). The indicated product size is only representative of the targeted genomic region.

Gene or target	Method section	Genomic coordinates (hg38)	Forward primer sequence	Reverse primer sequence	Product size (bp)	Annealing temperature (°C)
ADNP	2.5	chr20: 50892466-50892692	acactctttccctacacgacgctcttccgatctBHDBBCATGACCGCCTCAACTATCACTCTGCrATCTAC/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctBBDBBCAGGCTCTTCAGGCTTCTCTTCAAAGAAGrCTGGGA/3SpC3/	227	60
ALG13	2.5	chrX: 111684965 - 111685103	acactctttccctacacgacgctcttccgatctVHVBHGTGCAGGAAGCTGTTTGGAGACTrCTGGAT/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctDBVBDCTGCAGGTACAATAGAAGAGATGACCrCTCTTA/3SpC3/	139	60
CDK13	2.5	chr7: 40045943 - 40046107	acactctttccctacacgacgctcttccgatctHBDBHCTCATGGAGGGTCTGGATTATTGTCrATAAGT/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctDDHDDGCTACGGTTTCCAGTCTATTCTCTCrCATGTT/3SpC3/	165	60
CSNK2A1	2.5	chr20: 492230 - 492413	acactctttccctacacgacgctcttccgatctDBBBBCTGTGCCTGCCCTTCTGTTCTTACrCTGATT/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctDBHVHTGGGTAGCTTAGAGAAAGATAAGATTGrCTCATC/3SpC3/	184	60
FGFR2	2.5	chr10: 121520078 - 121520324	acactctttccctacacgacgctcttccgatctVDVHVATTCATGGGGCCACAGTGTTATTTCAAAGrGTGTCT/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctDVHVBAACCTTGCAGACAAACTCTACGTCTCrCTCCGT/3SpC3/	247	60
FGFR3.1	2.5	chr4: 1804303 – 1804506	acactctttccctacacgacgctcttccgatctDBHHDCAGGCCTCAACGCCCATGTCTTTGCrAGCCGT/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctHBHDHGGAAGCGGGAGATCTTGTGCACGrGTGGGC/3SpC3/	204	60
FGFR3.2	2.5	chr4: 1805554 – 1804745	acactctttccctacacgacgctcttccgatctDDHVHCACAGACGATGCCACTGACAAGGACrCTGTCC/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctDDVHHGGACACGGGCTCCTCAGACGrGGCTGG/3SpC3/	192	60
FGFR3.3	2.5	chr4: 1806042 - 1806249	acactctttccctacacgacgctcttccgatctDBDDDCTTCCCCAGTGCATCCACAGGGACrCTGGCA/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctHHHBHGGTGTGGGAAGGCGGTGTTGGrCGCCAA/3SpC3/	208	60
GNB1	2.5	chr1: 1806367 - 1806590	acactctttccctacacgacgctcttccgatctVHDDHCTGCTGCCTTCCCTATCCTGTrATTACC/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctVVHVBCACTTTCTGTGTTGTGCGGTACTrGATAGC/3SpC3/	224	60

HRAS	2.5	chr11: 533759 - 533965	acactctttccctacacgacgctcttccgatctHHVDDCTGAGCCCTGTCCTCCTGCAGGrATTCCA/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctDBDHHGGTTCACCTGTACTGGTGGATGTCCTCrAAAAGT/3SpC3/	207	60
PACS1	2.5	chr11: 66211057-66211289	acactctttccctacacgacgctcttccgatctDBHBBAGACTGTGTGTCCTCCAGAACCrCCTCAT/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctBBBHHCCAACAGTCTGGAGACTGTGTTCrCTCTCT/3SpC3/	233	60
PPP1CB	2.5	chr2: 28776842 - 28776986	acactctttccctacacgacgctcttccgatctVVBVDGTTTGTCAGTACGAGGATGTCGTCrCAGGAT/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctVVVBDATACCACAAATTTTCAGCGGTGCTTrCCAATA/3SpC3/	145	60
PPP2R5D.1	2.5	chr6: 43007166 - 43007399	acactctttccctacacgacgctcttccgatctHDDDBGGTGGAGCTCTAACTGGCCCTACCrCCTCAC/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctBBBBHGCACGCCACTGAGGGGATGTTCCrCAAGGT/3SpC3/	234	60
PPP2R5D.2	2.5	chr6: 43009218 - 43009389	acactctttccctacacgacgctcttccgatctBHDDDCCATTTCCAGGTGAGACTCCAACrCTAGCT/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctDVDHHGCAGCATTGTCACTTATCAGGCTrCATGAA/3SpC3/	172	60
PTPN11.1	2.5	chr12: 112450269 - 112450518	acactctttccctacacgacgctcttccgatctDVBBVGTAACTCTTTATTTGTCCCCTTGCCTCCrCTTTCG/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctBVBVVACTGACCTTTCAGAGGTAGGATCTGCrACAGTA/3SpC3/	250	60
PTPN11.2	2.5	chr12: 112477612 - 112477820	acactctttccctacacgacgctcttccgatctBBHDBGTCCAGGACTTATGTGACCGTGrGTCTCA/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctDDVVBAAGTTACTTTCAGGACATGAGGAAGGrATTTAT/3SpC3/	209	60
RET.1	2.5	chr10: 43118337 - 43118557	acactctttccctacacgacgctcttccgatctDBHHBGCGATCGTTTGCAACCTGCTCTGTGrCTGCAA/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctHHBHBAAGAGGGAGAACAGGGCTGTATGGrAGCCCG/3SpC3/	221	60
RET.2	2.5	chr10: 43121786 - 43122032	acactctttccctacacgacgctcttccgatctHDVDDAGCCCCTTCAAAGATGTGTGTGGCCArGTTCTC/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctHBHBHCAACACCCACACTTACACATCACTTTGCrGTGGTC/3SpC3/	247	60
SHOC2	2.5	chr10: 110964267 - 110964431	acactctttccctacacgacgctcttccgatctVVVVVTGATTGTGTAGGATCTTTGTCTCTTCATrCTTTGT/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctDVVVVCTCTTTCCTTGGCTGATGGTACTTTGrGGATC/3SpC3/	165	60
SMAD4	2.5	chr18: 51078191 - 51078400	acactctttccctacacgacgctcttccgatctBVVVBGAATGTAGGGAGGATGGGAAGAGATCACCrCTGTCG/3SpC3/	gtgactggagttcagacgtgtgctcttccgatctHHHDDCGGTGTAAGTGAATTTCAATCCAGCAAGrGTGTTA/3SpC3/	209	60
P5/P7	2.5	-	AATGATACGGCGACCACCGAGATCTNNNNNNNNacactctttccctacacgacgctcttccgatct	CAAGCAGAAGACGGCATACGAGATNNNNNNNNgtgactggagttcagacgtgtgctcttccgatct	-	60
SMAD4 sgRNA (top)	2.6	-	aaactTCGTCGCTTATGCATACTCAc		-	-

SMAD4 sgRNA (bottom)	2.6	-	caccgTGAGTATGCATAAGCGACGa		-	-
ssODN	2.6	-	AATCCGGTCCCCAGCCTTTCACAAAACTCATCCTGAGTATGCAGATCCGACGAAGGTCATCAACACCAATTCCAGCAGCAGCTGACAGACCTA		-	-
SMAD4_integrant	2.6	chr18: 51078211 - 51078398	acactgacgacatggttctacaGAGATCACCCTGTCCCTCTGATG	tacggtagcagagacttggtctGTGTAAGTGAATTTCAATCCAGCAA	188	64
SMAD4 PCR1	2.6	chr18: 51078192 - 51078469	AATGTAGGGAGGATGGGAAGAGATCACC	CCTCAGTCTAAAGGTTGTGGGTCTGCAA	278	68
SMAD4 PCR2	2.6	chr18: 51078241 - 51078404	acactgacgacatggttctacaAATCTTTTCTGTTAGGTCTGTCAGC	tacggtagcagagacttggtctGGCCCGGTGTAAGTGAATTT	164	64
Access Array barcodes	2.6	-	AATGATACGGCGACCACCGAGATCTacactgacgacatggttctaca	AAGCAGAAGACGGCATACGAGATNNNNNNNNNNtacggtagcagagacttggtct	-	60
CS1	2.6	-	A+CA+CTG+ACGACATGGTTCTACA	-	-	-
CS2	2.6	-	T+AC+GGT+AGCAGAGACTTGGTCT	-	-	-
CS2rc	2.6	-	A+GAC+CA+AGTCTCTGCTACCGTA	-	-	-

Chapter 3 – Identification and prioritization of candidate genes for selfish selection

In theory, mutations affecting any gene or pathway involved in spermatogonial stem cell homeostasis that alter the balance between self-renewal and differentiation could result in positive selection. Published discovery screens for selfish mutations have used broad approaches targeting established selfish genes, genes encoding RTK/Ras-MAPK pathway components, genes with narrow mutational spectra suggesting gain-of-function effects in developmental disorders, and known oncogenic drivers (Maher et al., 2016; Maher et al., 2018). Despite screening large genomic regions (>300 kb across >100 genes in microdissected tubules and ~67 kb across 71 genes in testes biopsies), novel non-Ras-MAPK mutations have remained elusive, suggesting that alternative approaches to candidate prioritization may be required. Preliminary evidence from a prior doctoral thesis (Ralph, 2018) using a sequencing panel originally designed to detect known and novel drivers of clonal hematopoiesis (104 loci across 74 genes; ~8 kb) (Acuna-Hidalgo et al., 2017) suggests that non-Ras-MAPK mutations may indeed be elevated in the male germline, but these results have yet to be validated.

To address this challenge, this chapter describes a systematic approach to identify novel candidate mutations that may undergo positive selection in the male germline. Large-scale sequencing data were first analyzed to identify recurrent *de novo* mutations. Next, candidate gene expression in SSCs was evaluated using complementary methods. Their functional potential to influence SSC fitness – and alignment with established PAE characteristics – was assessed through a review of the literature. Finally, these lines of evidence were integrated into a custom scoring system to prioritize the most promising candidates for experimental inquiry.

3.1 – Prioritization of candidate *de novo* mutations from >31K parent-offspring trios

As positive selection of mutant spermatogonia underlies the high apparent mutation rate observed in PAE disorders, it was hypothesized that selfish spermatogonial selection should be reflected by their recurrence as *de novo* mutations in clinically affected populations. To explore this hypothesis, data from the largest-to-date parent-offspring trio study of individuals with severe developmental disorders were analyzed. This study (hereafter "the Kaplanis study") utilized whole-exome sequencing (WES) to detect ~45.2K coding and splicing DNMs across ~31.1K parent-offspring trios (Kaplanis et al., 2020). The Kaplanis study comprised three patient cohorts: GeneDx, a diagnostic company (~18.8K trios), the Deciphering Developmental Disorders study (DDD; ~9.9K trios), and Radboud University Medical Center (RUMC; ~2.4K trios). Inclusion into the study required referral for genetic testing (GeneDx), the presence of an undiagnosed developmental disorder (DDD), or unexplained intellectual disability (RUMC). As patients with recognizable disorders that would receive a positive diagnosis through standard targeted sequencing were excluded, this dataset was relatively enriched for novel DNMs and thus well-suited for discovery.

In their study, Kaplanis identified 736 recurrent (≥2 observations) single nucleotide variants, of which 16 were highlighted as highly recurrent (observed ≥9 times) and proposed to confer a selective advantage in the germline (Kaplanis et al., 2020). However, because selection operates at the protein level, and different nucleotide changes can result in identical amino acid changes, a re-analysis of these data was undertaken. Candidate selfish mutations were therefore identified by their effect on the cognate amino acid. Furthermore, to both maximize discovery potential and capture functionally related mutations, DNMs affecting the same residue were grouped regardless of the resulting codon change. Although the precise functional effect and

selective advantage depends on the specific amino acid change (**section 1.7**), different amino acid substitutions at the same codon often result in the same or similar clinical phenotypes, as exemplified by multiple PAE disorders, including Noonan and Costello syndromes (Tartaglia et al., 2005; Giannoulatou et al., 2013). This strategy has also been applied in cancer genomics, where the discovery of positively selected oncogenic mutations has relied on the identification of recurrent changes at the same or even neighbouring residues (Lawrence et al., 2014).

Publicly available DNA base calls were annotated using the Ensembl Variant Effect Predictor (McLaren et al., 2016), resulting in a total of 863 recurrent DNMs. Synonymous mutations, which do not alter the protein sequence and are therefore unlikely to confer a selective advantage, accounted for 34/863 (3.9%) of DNMs (average 2.2 observations; range 2–6 observations). The recurrence of synonymous variants likely reflects elevated mutation rates at mutational hotspots and may confound the effects of selection (Poon et al., 2021). Therefore, to distinguish germline selection from hypermutability, DNMs were prioritized if they were observed ≥7 times, a threshold exceeding the maximum number of observations seen for any synonymous mutation. Additionally, to account for their elevated mutability, transition mutations occurring at CpG dinucleotides were prioritized only if they were observed ≥9 times. By applying these criteria, 22 DNMs (2.5%) across 20 genes met these criteria and were shortlisted for further assessment (**Table 3.1**). Despite actively excluding patients with known and readily diagnosable disorders, three of the prioritized DNMs were located in established selfish genes (*FGFR3*, *MAP2K1*, and *PTPN11*), several of which had previously been detected as mutational foci in testes biopsies, thus lending support to the approach (Yoon et al., 2013; Maher et al., 2018; Salazar et al., 2022).

Table 3.1: Comparison of prioritized recurrent *de novo* mutations identified at the amino acid-level vs. the nucleotide- level

Gene	Genomic position[1] (hg19/GRCh37)	cDNA mutation[2] (Amino acid change)	CpG transition[3]	DNM count (n) Amino acid-level[4]	DNM count (n) Nucleotide-level[5]
PACS1	Chr 11: 65978677	c.607C>T (p.R203W)	Yes	36	36
PPP2R5D	Chr 6:42975003	c.592G>A (p.E198K)	No	22	22
SMAD4	Chr 18:48604676	c.1498A>G (p.I500V); n = 21	No	22	21
	Chr 18:48604677	c.1499T>C (p.I500T); n = 1			
PACS2	Chr 14: 105834449	c.625G>A (p.E209K)	Yes	13	13
ADNP	Chr 20: 50892556	c.2156dup (p.Y719*); n = 1	No	11	8
	Chr 20: 50892557	c.2157C>G (p.Y719*); n = 8			
	Chr 20: 50892557	c.2157C>A (p.Y719*); n = 2			
MAP2K1[6]	Chr 15: 66729181	c.389A>G (p.Y130C)	No	11	11
MECP2	Chr X: 1532996777	c.502C>T (p.R168*)	Yes	11	10
NAA10	Chr X: 153197863	c.247C>T (p.R83C)	Yes	11	11
PPP1CB	Chr 2: 28999810	c.146C>G (p.P49R)	No	11	11
PTPN11[6]	Chr 12: 112915523	c.922A>G (p.N308D); n = 9	No	11	9
	Chr 12: 112915524	c.923A>G (p.N308S); n = 2			
CDK13	Chr 7: 40085606	c.2525A>G (p.N842S)	No	10	10
CSNK2A1	Chr 20: 472926	c.593A>G (p.K198R)	No	10	10
FGFR3[6]	Chr 4: 1807371	c.1619A>G (p.N540S); n = 3	No	10	7
	Chr 4: 1807371	c.1620G>C (p.N540K); n = 7			

Gene	Genomic position[1] (hg19/GRCh37)	cDNA mutation[2] (Amino acid change)	CpG transition[3]	DNM count (n) Amino acid-level[4]	DNM count (n) Nucleotide-level[5]
FOXP1	Chr 3: 71021818	c.1540C>T (p.R514C); n =1	Yes	10	9
	Chr 3: 71021817	c.1541G>A (p.R514H); n = 9			
CTBP1	Chr 4: 1206816	c.1024C>T (p.R342W)	Yes	9	9
MECP2	Chr X: 153296882	c.397C>T (p.R133C); n = 8	Yes	9	8
	Chr X: 153296881	c.398G>A (p.R133H); n = 1			
SHOC2	Chr 10: 112724120	c.4A>G (p.S2G)	No	9	9
SMAD4	Chr 18: 48604664	c.1486C>T (p.R496C)	Yes	9	9
SRCAP	Chr 16: 30748664	c.7303C>T (p.R2435*)	Yes	9	9
ALG13	Chr X: 110928268	c.320A>G (p.N107S)	No	8	8
GNB1	Chr 1: 1737942	c.239T>C (p.I80T); n = 6	No	7	6
		c.239T>G (p.I80S); n = 1			
GRIN2B	Chr 12: 13720098	c.2450G>C (p.G820A); n = 6	No	7	6
		c.2450G>T (p.G820V); n = 1			

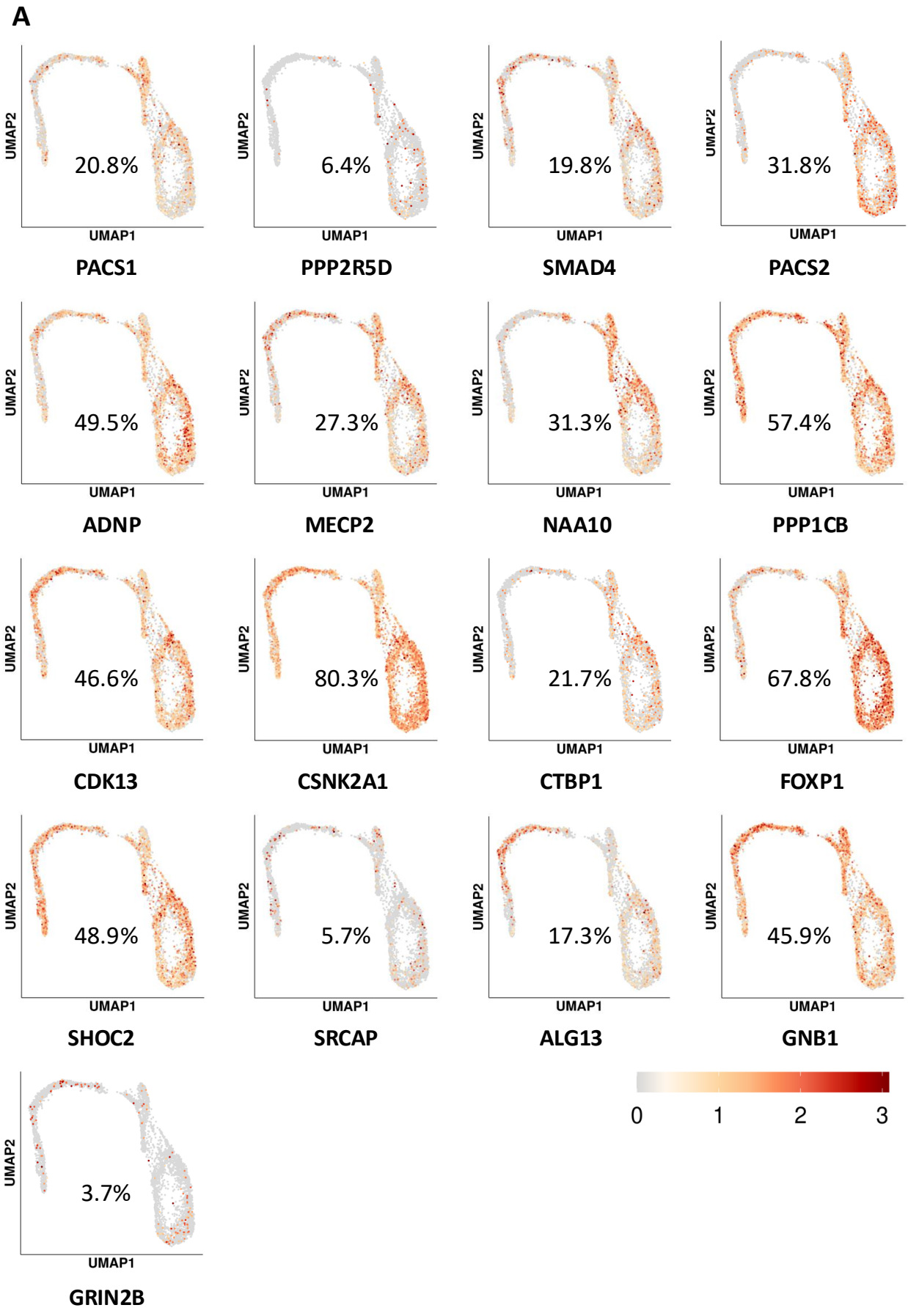
[1] Multiple genomic positions are shown for variants affecting the same amino acid residue.
[2] See **section 2.7.1** for the RefSeq transcript IDs used in variant annotation. For genes with multiple variants affecting the same amino acid residue, the number of nucleotide-level observations is also reported.
[3] CpG status affects mutation rates and was considered when setting prioritization thresholds (≥7 observations for non-CpG variants vs. ≥9 observations for CpG transitions)
[4] Amino acid-level counts represent aggregated nucleotide variants affecting the same amino acid residue.
[5] Nucleotide-level counts refer to the original, ungrouped counts of individual DNA base changes, as reported in Kaplanis et al., 2020.
[6] Established selfish gene.

3.2 – Expression of candidate selfish genes in SSCs

For mutations to influence positive selection in the male germline, they must occur in genes expressed in self-renewing SSCs. While established selfish genes are known to be expressed in SSCs where they regulate diverse aspects of stem cell homeostasis (**section 1.11**), there are limited studies directly assessing the expression or role of the 17 novel candidate genes in SSCs.

3.2.1 – Assessing expression using an integrated scRNAseq atlas of human spermatogonia

The integrated human spermatogonial atlas (**section 1.5.3**; Bush et al., 2024) offers an unbiased resource to address this gap and assess scRNAseq expression patterns directly in the cells of interest. Despite the apparent bidirectional plasticity of spermatogonia, where early differentiating cells can revert to an undifferentiated state, this analysis focused specifically on states 0, 0A/1, and 0B – which represent undifferentiated spermatogonia with self-renewal capacity that together comprised ~50.1% of the ~4.4K cells in the atlas (Bush et al., 2024). To account for the high rates of false positives in scRNAseq, genes were considered expressed only if detected in ≥10% of undifferentiated spermatogonia based on published thresholds (Karlsson et al., 2021; Chen and Zhou, 2018). Gene expression patterns were visualized to examine the distribution of each candidate and established selfish gene across spermatogonial states (**Figure 3.1**). 14 of 17 candidate genes were expressed in undifferentiated spermatogonia (average 34.2% of cells; range 3.7–80.3% of cells) (**Figure 3.1A**). All known selfish genes (12/12) were expressed in these states (average 31.3% of cells; range 12.3 –63.3% of cells) (**Figure 3.1B**).


A
UMAP2
UMAP1
20.8%
PACS1
6.4%
PPP2R5D
19.8%
SMAD4
31.8%
PACS2
49.5%
ADNP
27.3%
MECP2
31.3%
NAA10
57.4%
PPP1CB
46.6%
CDK13
80.3%
CSNK2A1
21.7%
CTBP1
67.8%
FOXP1
48.9%
SHOC2
5.7%
SRCAP
17.3%
ALG13
45.9%
GNB1
3.7%
GRIN2B
0
1
2
3

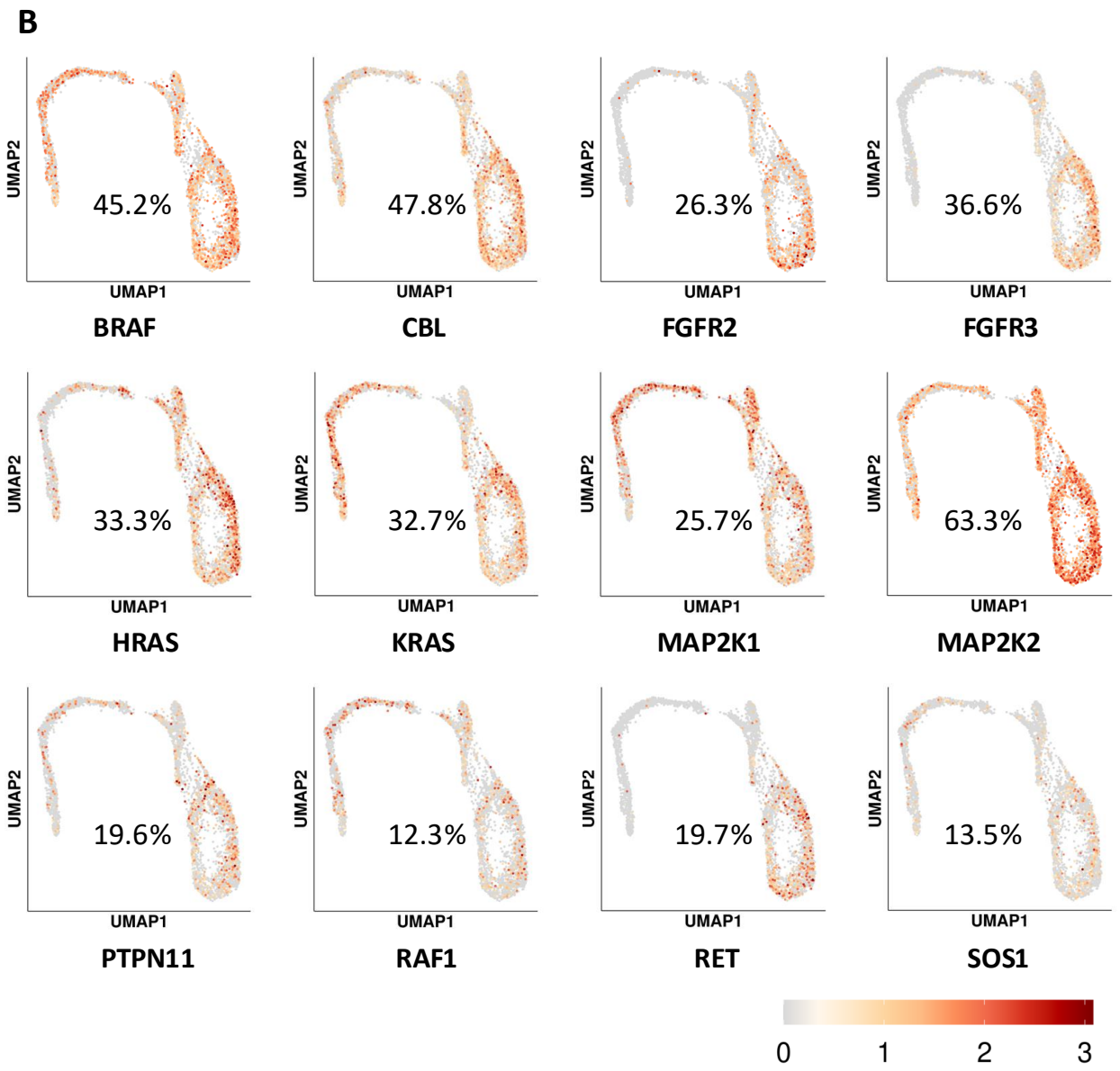


Figure 3.1: UMAP visualization of gene expression in human spermatogonia for candidate and established selfish genes. Gene expression data were derived from single-cell RNA sequencing (scRNAseq) analysis of human spermatogonia (Bush et al., 2024). Each dot represents a single cell, with colours indicating the normalized gene expression levels from 0 (no expression; grey) to 3 (high expression; dark red) as shown by the colour gradient below each plot. Panels **(A)** and **(B)** show expression patterns for candidate and established selfish genes, respectively. Expression in undifferentiated spermatogonia was assessed by focusing on the three earliest transcriptomic states (states 0, 0A/1, and 0B) represented by the ring in each UMAP plot (refer to **Figure 1.6**). The percentage in each plot indicates the proportion of cells in transcriptomically-defined undifferentiated spermatogonia expressing that gene, with genes considered expressed if detected in ≥10% of these cells.

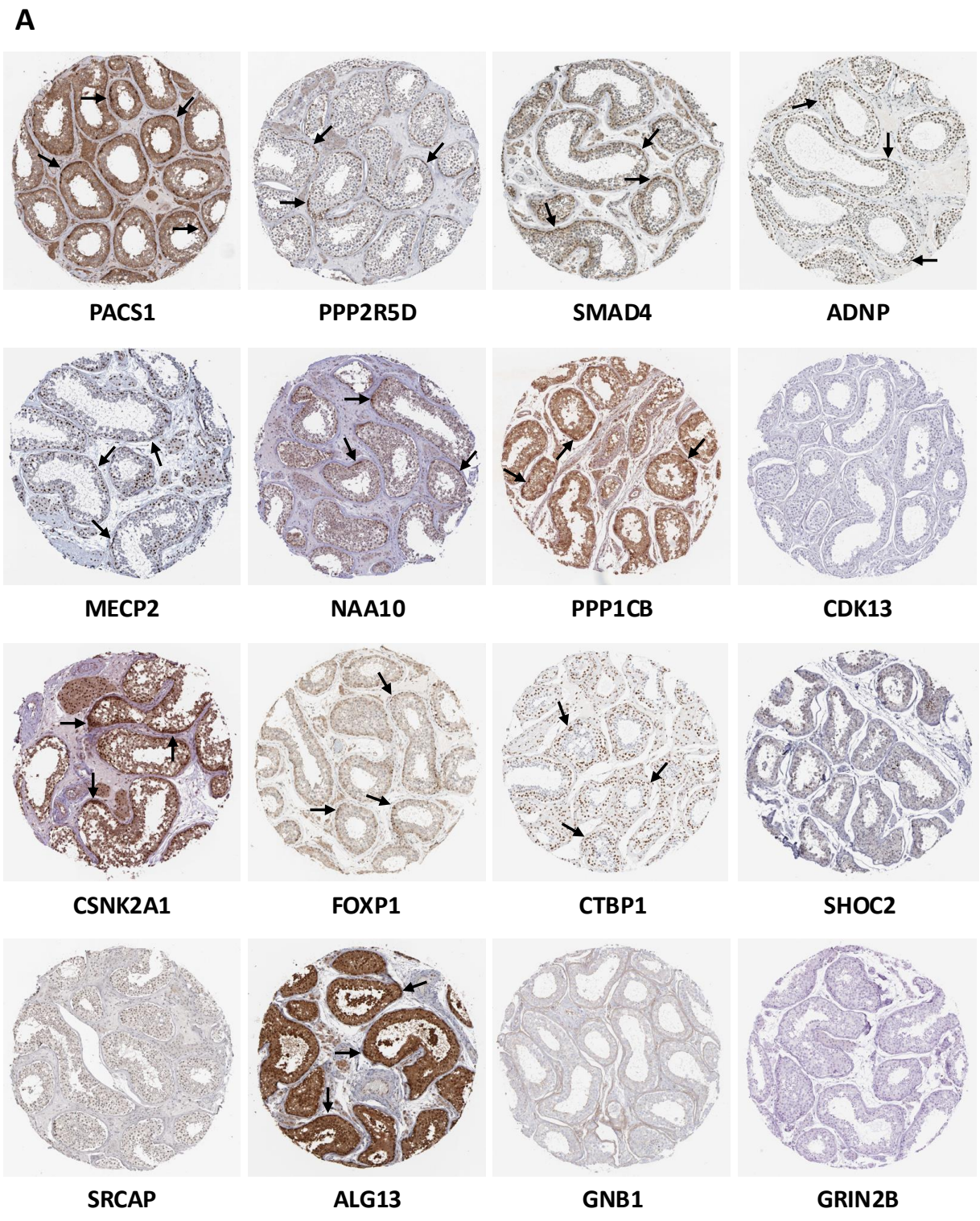
3.2.2 – Corroboration of expression via the Human Protein Atlas

Single-cell RNAseq provides unprecedented resolution to characterize complex spermatogonial populations but does not directly offer information about protein expression. Antibody-based approaches such as immunohistochemistry (IHC) can provide this additional layer of insight into the expression of candidate genes in SSCs. The tubular structure, characterized by its radial organization of developing germ cells from the basement membrane to the lumen (**section 1.5.1**), offers a useful spatial framework to understand expression patterns at the protein level.

Analysis of immunohistochemical staining patterns from the Human Protein Atlas (HPA) v23.0 database, an online resource containing millions of IHC images from >27.5K antibodies targeting >17.2K genes (87% of human protein-coding genes) across normal tissues (Thui and Lindkog, 2018), was performed to validate scRNAseq findings. As SSCs are found in the basal compartment along the basal membrane, protein expression patterns of candidate genes were assessed in cells along the periphery of seminiferous tubules. For each gene with reported antibody staining (PACS2 being the exception), between 1–18 testis tissue sections (average 5.7) were manually examined. Excluding *PACS2*, 10/16 candidate and 7/12 established selfish genes exhibited expression in cells along the basal lamina, suggestive of protein expression in SSCs (**Figure 3.2**).

A comparison between single-cell expression data in the human spermatogonial atlas and IHC staining patterns in HPA v23.0 revealed partial concordance between the two techniques. There was agreement between both techniques on 12/16 candidate and 9/12 known selfish genes. Whereas 3/18 candidate and 3/12 known selfish genes were expressed by scRNAseq but not by IHC, only 1/18 candidate and 0/12 selfish genes were expressed by IHC but not scRNAseq. These differences could reflect technical limitations such as tissue processing conditions, tissue quality,

antibody specificity and sensitivity, or could instead indicate biological factors such as post-transcriptional regulation.



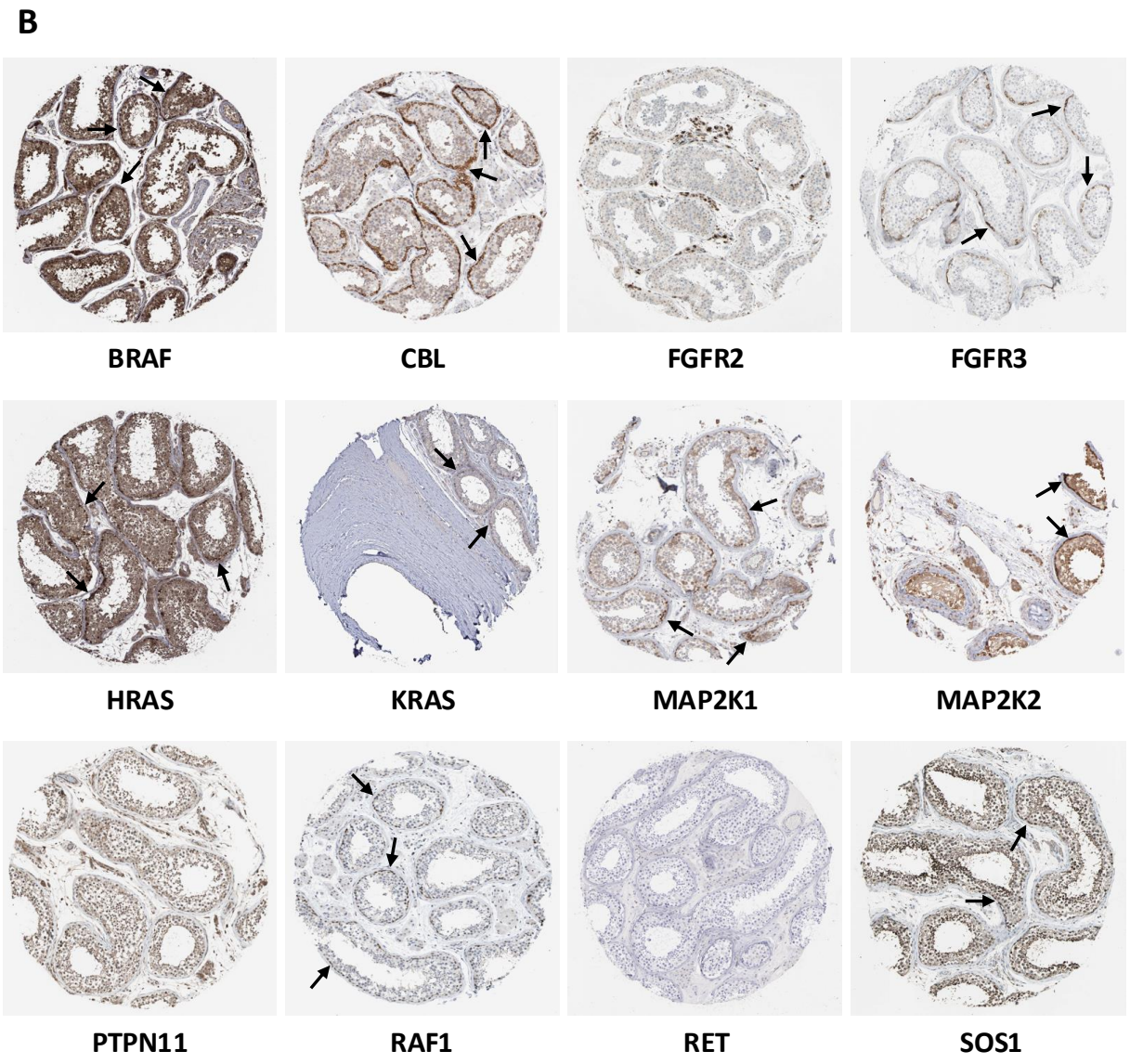


Figure 3.2: Example immunohistochemical (IHC) staining patterns of candidate and established selfish genes in human testes. Representative images from the Human Protein Atlas v23.0 database showing protein expression in formalin-fixed paraffin-embedded (FFPE) testis sections. Panels **(A)** and **(B)** show staining patterns for candidate and established selfish genes, respectively. Black arrows highlight positive staining along the basal membrane of seminiferous tubules, where undifferentiated spermatogonia are located. Images without arrows indicate that detectable expression was not observed.

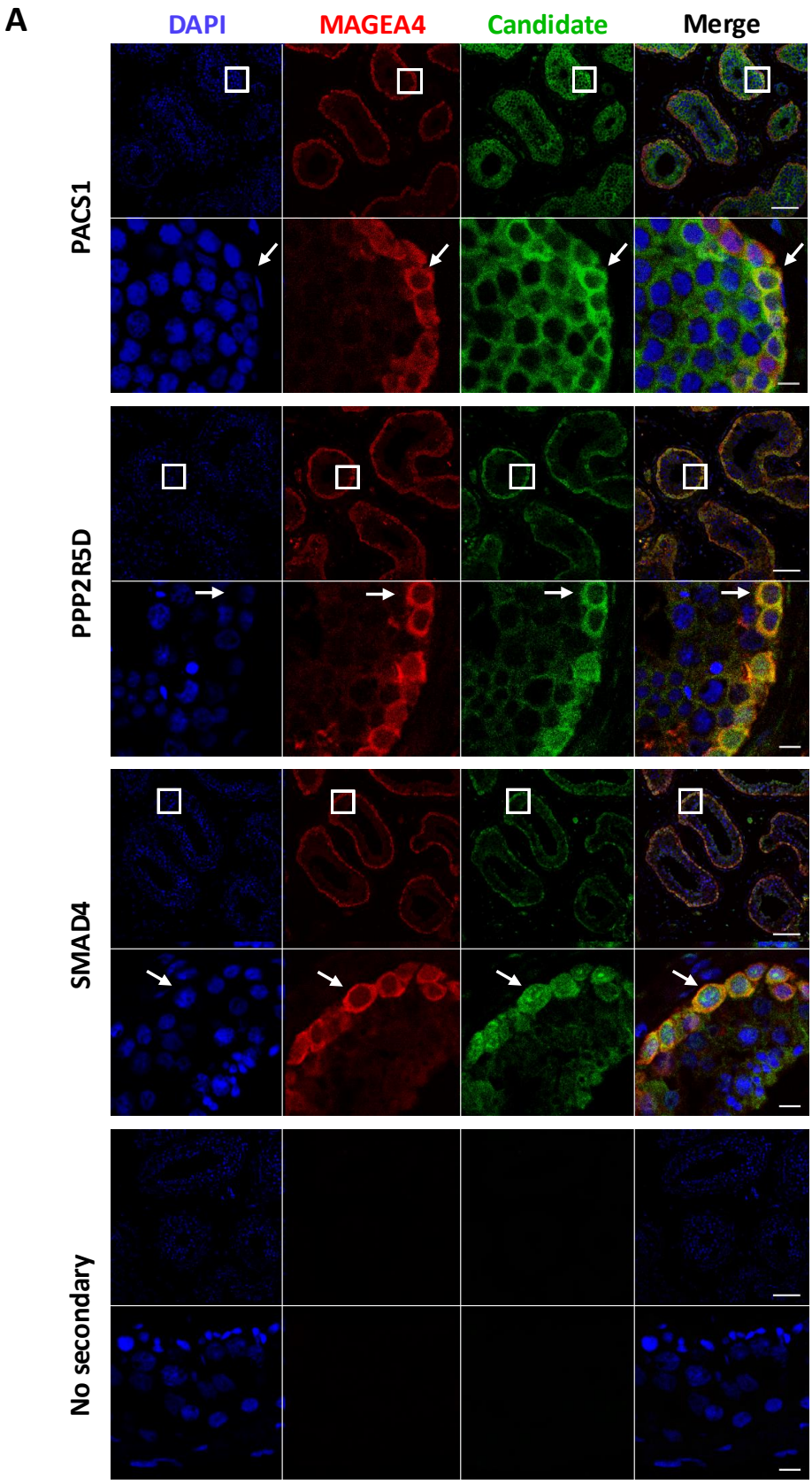
3.2.3 – Double immunofluorescence of select candidates with spermatogonial markers

Multiplexed immunofluorescence (IF) permits *in situ* analysis of protein localization and has been used to characterize expression across the different transcriptomic states of spermatogenesis (Guo et al., 2017). Here, the expression patterns of the three most recurrently mutated genes from the Kaplanis study were examined: PACS1, PPP2R5D, and SMAD4 (**Table 3.1**). While PACS1 and SMAD4 were both identified as expressed in SSCs at the RNA level by scRNAseq and at the protein level by IHC, PPP2R5D met expression criteria only by IHC. Using IF, testes sections were immunolabelled with antibodies against each candidate gene alongside two reference markers: MAGEA4, known for its expression in premeiotic germ cells (Aubry et al., 2001), and GFRA1, which is specifically expressed in a subset of undifferentiated spermatogonia (Guo et al., 2018; Bush et al., 2024).

MAGEA4 immunolabelling revealed continuous expression along the basal compartment, supporting the notion that MAGEA4 is broadly expressed in spermatogonia (**Figure 3.3A**). By contrast, GFRA1 exhibited a more punctuate pattern of expression, indicative of marking a narrower subset of undifferentiated spermatogonia (**Figure 3.3B**). Although each candidate antigen demonstrated co-expression with MAGEA4 and GFRA1-positive cells along the periphery of seminiferous tubules, there were noticeable differences in their respective patterns of expression. Compared with MAGEA4 and GFRA1, PACS1 displayed broad expression that extended beyond the basal membrane layer and could be observed at varying intensities in cells approaching the lumen, suggesting expression in spermatocytes and spermatids. On the other hand, PPP2R5D and SMAD4 exhibited more selective expression patterns with signal largely restricted to the periphery of the seminiferous tubules.

Comparison with DAPI-labelled nuclei allows for characterization of subcellular localization of the candidate antigens. Whereas PACS1 and PPP2R5D do not substantially co-localize with DAPI, suggesting a cytoplasmic or membrane-associated distribution, SMAD4 expression frequently overlaps with DAPI, consistent with the understanding that SMAD4 continuously shuttles between the nucleus and cytoplasm to regulate transcription (Pierreux et al., 2000; **section 3.3.3**).

Multiplexed IF results were consistent with IHC findings for all three candidate genes, which indicated expression in cells along the periphery of seminiferous tubules (**Figure 3.2A**). To investigate whether multiplexed IF reflected gene expression at the RNA level, the broader human testicular transcriptional map – which contains scRNAseq expression data for all testicular cell types and from which the spermatogonial atlas was constructed (**section 1.5.3**; Bush et al., 2024) – was analyzed. PACS1 expression was confirmed across all major germ cell types, whereas SMAD4 expression was limited to spermatogonia and only one of four clusters of spermatids. PPP2R5D did not meet the ≥10% expression criteria in any cell population, yet still showed the highest relative expression in undifferentiated spermatogonia, aligning with the IHC and IF data.


A
DAPI
MAGEA4
Candidate
Merge
PACS1
PPP2R5D
SMAD4
No secondary

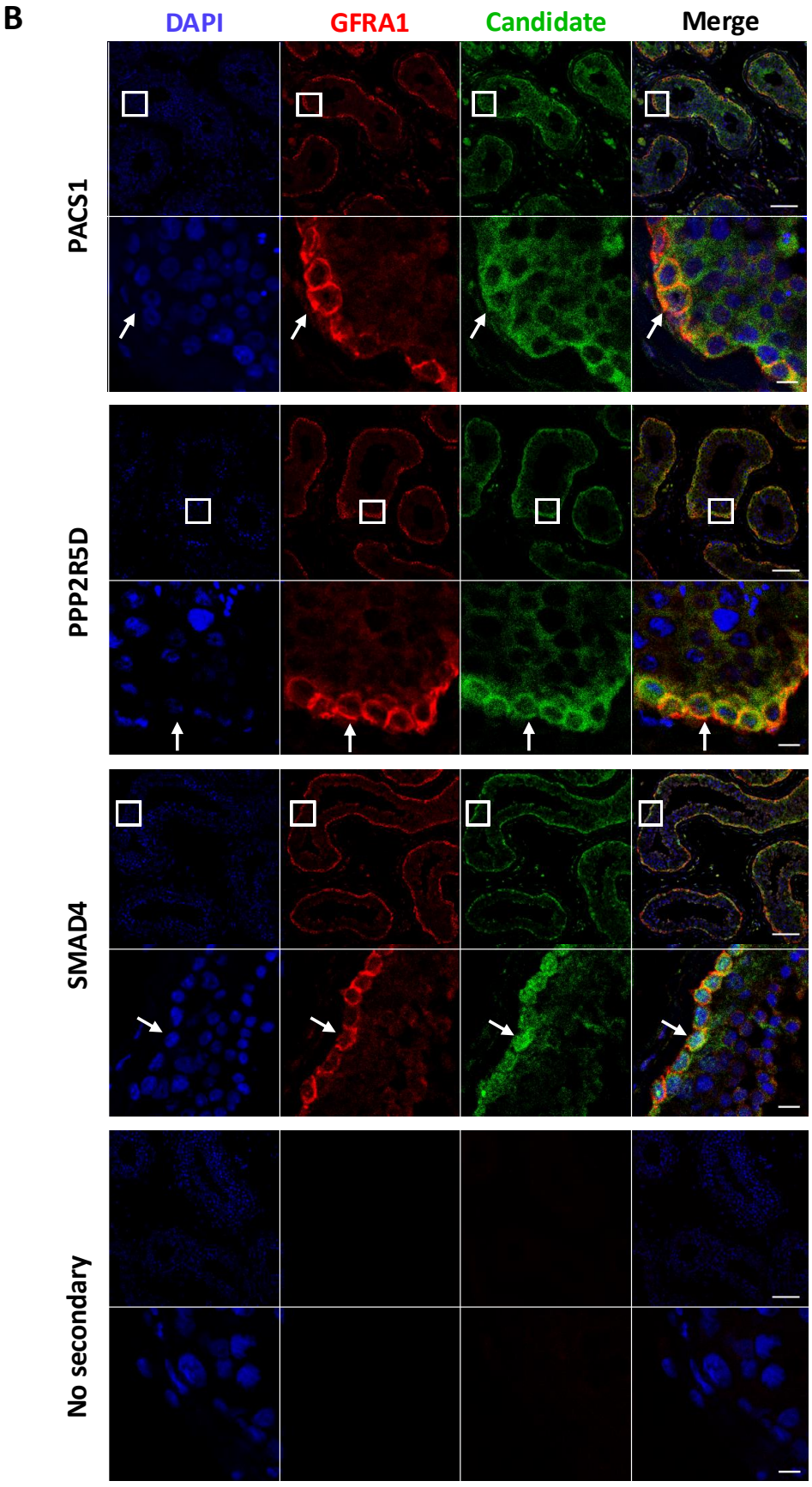

B
DAPI
GFRA1
Candidate
Merge
PACS1
PPP2R5D
SMAD4
No secondary

Figure 3.3: Multiplexed immunofluorescence (IF) validation of select candidate gene expression in spermatogonia. Testis sections were immunolabeled with candidate antigens (green) together with either MAGEA4 (red), a marker for premeiotic germ cells, or GFRA1 (red), a marker for a subset of undifferentiated spermatogonia, and counterstained with DAPI (blue) to visualize nuclei. Candidate genes are presented first in combination with **(A)** MAGEA4 and then with **(B)** GFRA1. Each set of images consists of four panels (DAPI, reference marker, candidate antigen, and composite) and includes low-magnification overviews (top; scale bar = 100 μm) and high-magnification insets (bottom; scale bar = 10 μm) to highlight antigen localization relative to the basal compartment of the seminiferous tubules. White boxes on the low-magnification images indicate the regions shown at higher magnification, and white arrows in the high-magnification insets highlight exemplar spermatogonia expressing both candidate antigen and reference marker. A row of negative controls (no secondary antibody applied) is included to show background signal levels.

3.3 – Analysis of candidate selfish genes: function, mechanism of mutation, and Paternal Age Effect criteria

For mutations to confer a selective advantage in the male germline, the affected genes must not only be expressed in self-renewing SSCs but must also participate in signaling pathways that regulate SSC homeostasis. Additionally, the specific mutations must alter protein function in ways that enhance stem cell fitness. To prioritize candidate genes most likely to undergo selfish selection, this section examines each candidate gene through three critical lenses: its molecular functions, signaling pathway involvement, and the mechanistic consequences of the identified mutations. For example, identifying the same or similar variants in databases such as COSMIC v101 (https://cancer.sanger.ac.uk/cosmic) may suggest that these mutations confer a fitness advantage, analogous to what might occur in SSCs. Together, these lines of evidence may indicate potential mechanisms for influencing SSC behavior and how protein function might be altered to confer advantage. Although these data remain limited in the literature, this section also considers evidence related to the three established PAE criteria: advanced paternal age, an exclusive (or near-exclusive) paternal bias in mutation origin, and a high apparent mutation rate. While the main text

focuses on the most informative and interpretable evidence, a comprehensive overview of each shortlisted variant is provided in **Table 3.2**.

3.3.1 – Phosphofurin acidic cluster sorting protein 1 (*PACS1*)

PACS1 encodes a trans-Golgi network traffic regulator that sorts and localizes cytosolic proteins to the trans-Golgi membrane (Wan et al., 1998). Through the recognition of acidic cluster sorting motifs on cargo proteins by the furin-binding region (FBR), and in complex with clathrin adaptor proteins such as AP1, PACS1 mediates key cellular processes such as the retrieval of cargo proteins from endosomes in the trans-Golgi network (Scott et al., 2003). PACS1 most notably mediates the trafficking of the proprotein convertase furin, which is involved in the proteolytic activation of growth factors such as BMP/TGF-β family members (Cui et al., 2001; Youker et al., 2015). The activity of PACS1 is regulated at the p.S278 residue, where phosphorylation by casein kinase 2 (CK2; **section 3.3.10**) promotes cargo binding, and dephosphorylation by protein phosphatase 2A (PP2A; **section 3.3.2**) reduces this interaction (Scott et al., 2003).

De novo mutations in *PACS1* cause Schuurs-Hoeijmakers syndrome (SHMS; OMIM: #615009) which is characterized by distinct facial dysmorphia, structural abnormalities of the heart and brain, epilepsy, and intellectual delay (Schuurs-Hoeijmakers et al., 2012). SHMS exhibits a highly clustered mutational spectrum, with nearly all cases caused by the highly recurrent c.607C>T (p.R203W) mutation. The PACS1 p.R203W mutation occurs at a CpG dinucleotide, which is in close proximity to the CK2 binding motif within the FBR, suggesting that it may impact interactions and the trafficking of cargo proteins (Gadzicki et al., 2015). Functional studies have shown that the PACS1 p.R203W mutation may exert its effects through a gain-of-function mechanism, such as promoting interactions and potentiating activity of the cytoplasmic histone

deacetylase, HDAC6, which disturbs proper development of the Golgi apparatus (Villar-Pazos et al., 2023); the mutation also disrupts the expression of genes involved in ion channel transport and synaptic transmission, which are thought to underly impaired electrical activity such as prolongated burst duration (Rylaarsdam et al., 2024). Injection of human *PACS1* c.607C>T mRNA into zebrafish embryos results in impaired craniofacial development, potentially due to improper migration and specification of neural crest cells (Schuurs-Hoeijmakers et al., 2012).

PACS1 does not have an established role in the regulation of SSCs but is expressed in undifferentiated spermatogonia at the RNA level (**Figure 3.1**) and in both MAGEA4 and GFRA1-expressing cells at the protein level (**Figure 3.2; Figures 3.3A/B**). CK2 and PP2A, which modulate PACS1 activity, also regulate the MAPK, AKT, and WNT pathways – signalling networks implicated in SSC homeostasis (Silverstein et al., 2002; Kuo et al., 2008; Thompson et al., 2018; Guo et al., 2017). Moreover, as furin mediates the maturation of several BMP ligands, such as BMP4, PACS1 may indirectly affect BMP signalling, another pathway upregulated in undifferentiated spermatogonia. Despite these connections, no direct evidence yet shows that PACS1 actively modulates these critical signaling pathways in SSCs.

3.3.2 – Protein phosphatase 2 regulatory subunit B'delta *(PPP2R5D)*

PPP2R5D encodes the δ isoform of B56, a highly-conserved regulatory subunit of protein phosphatase 2A (PP2A) involved in determining substrate specificity and subcellular targeting (McCright et al., 1996; Hertz et al., 2016). The PP2A holoenzyme consists of scaffold (A) and catalytic (C) subunits, each encoded by two genes, which form a heterodimeric core that binds one of 15 possible regulatory (B) subunits (Sandal et al., 2021). PP2A is a ubiquitous serine/threonine

protein phosphatase in eukaryotes that regulates essential processes such as cell division, development, and growth (Seshacharyulu et al., 2013; Lillo et al., 2014).

De novo missense mutations in *PPP2R5D* cause Jordan syndrome (JS; OMIM: #616355), a dominant disorder characterized by intellectual disability, hypotonia, and macrocephaly (Houge et al., 2015). The majority of DNMs in JS, including the highly recurrent PPP2R5D c.592G>A (p.E198K) mutation (detected 22 times in the Kaplanis parent-offspring trio study), cluster within a highly conserved acidic loop that mediates contact with the PP2A scaffolding (Aα) and catalytic (Cα) subunits, encoded by *PPP2CA* and *PPP2R1A*, respectively (Seshacharyulu et al., 2013). *In silico* structural analysis predicts that these mutations disrupt holoenzyme formation by introducing an electrostatic change near contact regions between the Aα and Cα subunits (Houge et al., 2015). This mutant conformation prevents the dephosphorylation of several known B56-specific substrates such as the GSK-3β p.S9 protein kinase, thereby potentially disrupting downstream signalling processes. Despite not meeting the recurrence threshold for prioritization, the c.1258G>A (p.E420K) mutation (observed 6 times in the Kaplanis study) is also frequently reported in Jordan syndrome. The p.E420 position is located outside of the acidic loop, but nevertheless faces the catalytic subunit near the active site of the holoenzyme, suggesting that the functional effects of this mutation may be similar to those clustered within the loop (Shang et al., 2016; Smolen et al., 2023). Both PPP2R5D p.E198K and p.E420K mutations increase phosphorylation of ERK1/2 at the p.T202/Y204 sites, indicative of enhanced MAPK signaling (Smolen et al., 2023). While both mutations induce hyperphosphorylation of RPS6, a downstream target of mTORC1, only p.E420K exhibits constitutive AKT activation, suggesting an alternative mechanism for p.E198K cells, potentially involving the MAPK pathway (Papke et al., 2021; Smolen et al., 2023). As PPP2R5D KO mice are cognitively normal but develop movement

defects, mutations in *PPP2R5D* are thought to act in a dominant-negative manner (Louis et al., 2011). PPP2R5D also contributes to the tumor suppressive function of PP2A; mice lacking PPP2R5D are viable but develop spontaneous tumour formation, thought to be driven by dysregulation of c-Myc, a potent oncogene, mediated by the loss of GSK-3β (Lambrecht et al., 2020).

Although *PPP2R5D* does not have an established role in SSC homeostasis, the PP2A catalytic subunit *PPP2CA* is essential for normal spermatogenesis, as knockout mice exhibit disordered spermatogonia and are unable to undergo meiosis resulting in infertility (Pan et al., 2015). While *PPP2R5D* was not expressed in undifferentiated spermatogonia at the RNA level by scRNAseq (**Figure 3.1**), IHC and multiplexed IF analyses showed that PPP2R5D was expressed in both MAGEA4 and GFRA1-positive cells (**Figure 3.2; Figures 3.3C/D**). Through direct dephosphorylation of MEK1/2 and ERK1/2, AKT, and *β*-catenin, PP2A has diverse roles influencing critical signalling pathways implicated in SSC regulation (Silverstein et al., 2002; Kuo et al., 2008; Thompson et al., 2018). Jordan syndrome mutations in *PPP2R5D* have been specifically shown to activate MAPK and AKT signalling, established mechanisms by which known selfish mutations confer a selective advantage to SSCs. Consistent with this, the c.592G>A (p.E198K) mutation was previously detected at 0.67% VAF in a single testis biopsy from a 90-year-old man, suggesting that such mutations may become enriched over time (Ralph, 2018) (**Table 3.2**).

3.3.3 – Mothers against decapentagplegic homolog 4 *(SMAD4)*

SMAD4 is the central regulator of the bone morphogenetic protein (BMP) and transforming growth factor-beta (TGF-β) signaling pathways. The SMAD proteins are responsible for the transduction

of extracellular signals to the the nucleus and are categorized into three groups: the receptor-regulated SMADs (R-SMADs), the common mediator SMAD (co-SMAD; *SMAD4*), and the inhibitor SMADs (I-SMADs) (Attisano et al., 2000). Ligand binding to type II BMP/TGF-β receptors leads to recruitment and phosphorylation of type I receptors, which then phosphorylate the R-SMADs (*SMAD1*, *SMAD*5, and *SMAD8* transduce BMP signals, whereas *SMAD2* and *SMAD3* mediate TGF-β signals). Signal propogation and subsequent transcriptional control is dependent on the formation of R-SMAD and SMAD4 heterotrimeric complexes, a process which is mediated by the Mad Homology 2 (MH2) domain, one of two globular domains characteristic of the SMAD proteins (Ross and Hill, 2008). SMAD6 and SMAD7 are the I-SMADs that negatively regulate BMP and TGF-β signaling, respectively, by directly competing with R-SMADs for type I receptor interaction as well as potentially impeding complex formation between the R-SMADs and SMAD4. The I-SMADs can additionally recruit E3 ligases and other proteins to degrade and dephosphorylate receptors (Miyazawa et al., 2002). Following R-SMAD/SMAD4 hetero-oligomerization, the complexes are then shuttled into the nucleus, where a B-hairpin structure in the Mad Homology 1 (MH1) domain recognizes and binds to 5 base pair (bp) DNA motifs (CAGAC) called Smad-binding elements (SBEs), which helps direct transcriptional control of TGF-β and BMP target genes (Massagué and Wotton, 2000).

SMAD4 is a putative tumor suppressor and was originally discovered for its involvement in pancreatic ductal adenocarcinoma where 30% of cases carry homozygous deletions (Hahn et al., 1996). Mutations in *SMAD4* have now been detected in numerous cancers with COSMIC reporting 1633 unique point mutations across all 12 coding regions of *SMAD4* in 46 different histologies. Genetic deletions of *SMAD4* have been shown to promote tumor growth and metastasis (Ding et al., 2000) and both nonsense and missense point mutations resulting in

inactivated TGF-β signaling (Xu et al., 2000; Lanauze et al., 2021). In addition to somatic *SMAD4* mutations driving oncogenesis, germline loss-of-function (LOF) mutations in *SMAD4* cause two additional autosomal dominant conditions: juvenile polyposis syndrome (JPS; OMIM: 174900) and JPS-hereditary hemorrhagic telangiectasia (JPHT; OMIM: #175050). Approximately 23% of JPS cases are caused by heterozygous loss-of-function *SMAD4* deletions, small indels and point mutations throughout the gene (Chow and Macrae, 2005). JPS is characterized by the presence of polyps along the gastrointestinal tract which cause gastrointestinal bleeding, anemia, and an elevated risk for colorectal cancer (Brosens et al., 2011). JPHT shares clinical features from both JPS and hereditary hemorrhagic telangiectasia (HHT; OMIM: #187300), a disorder characterized by multi-organ vascular dysplasia which can lead to both hemorrhage and stroke (Williams et al., 2012). Although the majority of HHT cases are caused by mutations in other TGF-β family members, such as *ENG* and *ACVRL1*, JPHT has a similar spectrum of heterozygous *SMAD4* LOF mutations to JPS (Gallione et al., 2009; Iyer et al., 2010).

In contrast to the wide-range of LOF mutations involved in both JPS/JPHT and cancer, MS *SMAD4* mutations affect only two residues (p.R496 and p.I500) located in exon 12 within the MH2 domain. A review of all 146 published cases of *de novo* MS to date, as well as those reported in genomic databases, indicate that the distribution of MS SMAD4 mutations is as follows: c.1486C>T (p.R496C), 16.4%; c.1498A>G (p.I500V), 62.3%; c.1499T>C (p.I500T), 19.9%; c.1500A>G (p.I500M), 1.4% (**Supplementary Table S1**). The highly-clustered nature of these mutations suggests a distinct mechanism of action and consequence on the *SMAD4* protein from haploinsufficiency (Caputo et al., 2012; Lelieveld et al., 2017). The two p.496 and p.I500 codons are in close proximity to two known sites of *SMAD4* monoubiquitination: p.K507 and p.K519 (Morén et al., 2003; Dupont et al., 2009). Mutagenesis of these two residues have shown reduced

monoubiquitination, impaired R-SMAD hetero-oligomerization (Dupont et al., 2009), and defective TGF-β signaling (Morén et al., 2003). *In silico* structural analyses of the MS *SMAD4* mutations suggest that the conformational changes introduced by these variants could affect proper SMAD4 ubiquitination at p.K519, thereby leading to perturbed signalling due to increased SMAD4 levels (Caputo et al., 2012; Caputo et al., 2014). Indeed, analyses of MS patient fibroblasts have shown decreased ubiquitination of SMAD4 vs. wild-type controls, enhanced levels of total Smad and phosphorylated SMAD2/3 (pSMAD2/3) and pSMAD1/5/8, and impaired transcriptional activation of TGF-β and BMP target genes (Le Goff et al., 2012). These data have led others to propose that MS *SMAD4* mutations operate through a gain-of-function mechanism whereby *SMAD4* mutations perturb normal ubiquitination, which leads to increased SMAD4 levels, ultimately resulting in a loss of BMP/ TGF-β transcriptional control (Piccolo et al., 2014; Lin et al., 2016).

More recently, however, Alankarage et al. proposed an alternative model wherein MS *SMAD4* mutations exert their effects via a dominant-negative mechanism. The authors showed that SMAD4 I500V not only has reduced transactivation ability compared to WT SMAD4, it also acts in a dominant-negative manner on co-transfected WT SMAD4 to reduce its transcriptional activity (Alankarage et al., 2022). Their data suggested that mutant SMAD4 retains the ability to dimerize and subsequently sequesters functional wild-type SMAD4 and phosphorylated R-SMADs in transcriptionally defunct homodimers and heterotrimeric complexes, respectively. They proposed that the observed perturbation of ubiquitination by Le Goff et al. is instead a compensatory response to rescue intracellular levels of functional WT SMAD4. Moreover, like Le Goff et al., they showed a loss of TGF-β and BMP transcriptional activity and additionally demonstrate that SMAD4 I500V can directly bind to and dysregulate critical SMAD4 transcription

factors. Therefore, future studies will be needed to resolve these conflicting data and outline the exact mechanisms by which these *SMAD4* mutations lead to the MS phenotype.

SMAD4 does not have a known role in SSC maintenance, but BMP pathway members are highly expressed in self-renewing human SSCs (Guo et al., 2017) and a role for TGF-β signaling in spermatogonial development has been suggested in chickens (Zuo et al., 2017). *SMAD4* is expressed in undifferentiated spermatogonia at both the RNA level by scRNAseq (**Figure 3.1**) and in GFRA1 and MAGEA4-expressing cells at the protein level (**Figure 3.2; Figures 3.3E/F**). There are 61 reports of MS paternal age with a mean (± SD) age of 36.1 years (2.5) (range 22–49 years) and 25 reports of MS maternal age with a mean (± SD) age of 32.0 years (0.6) (range 22–42 years) (**Supplementary Table S1**). MS *SMAD4* mutations have been detected in four biopsies in the testes of men aged 71 - 90 years (Ralph, 2018).

3.3.4 – Phosphofurin acidic cluster sorting protein 2 (*PACS2*)

PACS2 encodes a multifunctional sorting connector located at mitochondria-associated membranes (MAMs) – critical contact points between the endoplasmic reticulum (ER) and mitochondria – that mediates the trafficking of cargo proteins within the trans-Golgi network (Li et al., 2020). Similar to its paralogue PACS1, the PACS2 structure includes a FBR at the N-terminus that is responsible for binding acidic clusters on cargo proteins, a disordered middle region (MR) that carries a nuclear localization signal and an AKT phosphorylation site, and a C-terminal region (CTR) (Thomas et al., 2017). PACS2 is regulated by phosphorylation by AKT at p.S437 within the MR and has documented roles in mediating both pro-apoptotic and anti-apoptotic responses, as well as in maintaining the integrity and function of MAMs (Li et al., 2020).

When phosphorylated, PACS2 interacts with and distributes proteins such as calnexin and polycystin-2 to the MAMs on the ER, both of which are required for normal Ca^{2+} signalling.

De novo mutations in *PACS2* cause Developmental and Epileptic Encephalopathy 66 (DEE66; OMIM: #618067) which is additionally characterized by gross motor delay, neuroimaging abnormalities, and facial dysmorphism (Olson et al., 2018; Chou et al., 2023). Virtually all reported cases of DEE66 carry the c.625G>A (p.E209K) mutation, which occurs at a CpG dinucleotide and was detected 13 times in the Kaplanis parent-offspring trio study. The p.E209 residue neighbours a predicted autoregulatory (AR) domain; the same AR domain in PACS1 reversibly binds to the FBR to regulate client protein interactions and additionally contains the site of CK2 phosphorylation (Scott et al., 2003). Therefore, *PACS2* c.625G>A (p.E209K) mutations are believed to disrupt the ability of the autoregulatory domain to facilitate interactions with client proteins, as reflected by the increased affinity that PACS2 p.E209K mutants have for known cargo proteins (Olsen et al., 2018; Zang et al., 2022). Introduction of this mutation alters the phosphorylation of PACS2 at the neighbouring p.S207/208/211 residues by CK2, thereby disrupting PACS2 complex formation with PKD2 which regulates Ca^{2+} transport from the ER to the mitochondria (Truong et al., 2023). Cell lines carrying this mutation exhibit reduced Ca^{2+} uptake by mitochondria as well as impaired MAM formation and integrity; mutant primary glutamatergic neurons have elevated levels of neurotransmitters at presynaptic terminals. Together, these molecular consequences are thought to underlie the epileptic phenotype of DEE66.

PACS2 does not have an established role in SSC regulation but was detected in undifferentiated spermatogonia at the RNA level by scRNAseq (**Figure 3.1**). Without a suitable antibody against PACS2, protein-level validation by IHC or IF was not performed. Although PACS2 is an AKT substrate, further study is required to demonstrate whether PACS2 or the

p.E209K mutation can reciprocally influence the AKT signaling cascade or other key pathways involved in SSC maintenance.

3.3.5 – Activity-dependent neuroprotective protein (*ADNP*)

ADNP encodes a putative transcription factor that plays crucial roles in neurogenesis, as well as neuronal growth, and development (Mandel et al., 2008). Mice with complete loss of ADNP are embryonic lethal due to defects in neural tube closure and brain formation, whereas heterozygous ADNP knockout mice are viable but exhibit learning and memory deficits (Pinhasov et al., 2003; Vulih-Shultzman et al., 2007). ADNP functions as part of the SWI/SNF (BAF) complex, directly interacting with key complex components such as SMARCA4, ARID1A, and SMARCC2, which is responsible for regulating gene expression through chromatin remodeling (Helsmoortel et al., 2014). Amongst its diverse functions, ADNP notably modulates Wnt signaling by directly interacting with β-catenin and preventing its degradation by the ubiquitin-proteasome pathway (Sun et al., 2020).

Pathogenic *de novo* mutations in *ADNP* have been associated with a range of developmental disorders, including Helsmoortel-Van der Aa syndrome (HVDAS; OMIM: #615873), a dominant disorder characterized by intellectual delay, global developmental delay, and autism spectrum disorder (ASD) (Helsmoortel et al., 2014). ADNP has been estimated to account for at least 0.17% of ASD cases and is considered one of the most prevalent ASD-associated genes known. HVDAS mutations are located in the last exon of ADNP and are primarily frameshift or nonsense variants, resulting in a premature termination codon (Helsmoortel et al., 2014). The most recurrent mutation in HVDAS is ADNP p.Y719*, which was detected 10 times in the Kaplanis parent-offspring trio study. ADNP p.Y719 is located within the nuclear localization

signal (NLS) sequence that regulates binding to importin proteins. Indeed, cells transfected with ADNP p.Y719* exhibit impaired nuclear transport and cytoplasmic accumulation (Cappuyns et al., 2018). Mutant mRNA – but not mutant protein – is detectable in HVDAS patient-derived samples, indicating escape from nonsense-mediated decay and supports a model of haploinsufficiency in which a single functional ADNP allele is insufficient for normal development (Helsmoortel et al., 2014; Cappuyns et al., 2018). Nevertheless, several observations challenge haploinsufficiency as being the sole mechanism of mutation. For example, individuals with p.Y719* have a more severe phenotype than other recurrent ADNP mutations such as p.R730*, possibly reflective of the fact that while ADNP p.Y719* has impaired subcellular localization, ADNP p.R730* retains an intact NLS allowing for nuclear import. This has led to speculation that cytoplasmic ADNP p.Y719* sequesters ADNP binding partners, thus interfering with wild-type ADNP function in a novel dominant-negative manner, in addition to a simple model of haploinsufficiency (Cappuyns et al., 2018). Functional studies of HVDAS patient-derived tissue exhibited downregulated Wnt signaling which is thought to underlie the neurodevelopmental delay phenotype (D'Incal et al., 2024).

ADNP does not have a known role in regulating SSC homeostasis but is expressed in undifferentiated spermatogonia both at the RNA level by scRNAseq (**Figure 3.1**) and at the protein level by IHC (**Figure 3.2**). In mice, upregulation of the Wnt/*β*-catenin pathway governs the transition of undifferentiated spermatogonia from a stem-like state to a progenitor state, and Wnt inhibition has therefore been proposed to preserve stemness (Mäkelä and Hobbs, 2019). HVDAS mutations such as ADNP p.Y719*, which downregulate Wnt signaling, may therefore influence SSC dynamics by promoting self-renewal. In a prior discovery screen, the *ADNP* c.2187C>T (p.R730Q) mutation was detected in two testes biopsies from an older man (71 years) (Ralph,

2018). Although this mutation occurs at the same residue as the recurrent *ADNP* p.R730* mutation, potentially implying functional similarities, the consequence on the ADNP protein structure is expected to be fundamentally different without the introduction of a premature stop codon.

Evidence for paternal age effects in HVDAS is limited. While one HVDAS case series reported a mean paternal age of 32.1 years (n = 65) versus a mean maternal age of 29.8 years (n = 67), no comparison against an age-matched population was performed (Van Dijck et al., 2019) (**Table 3.2**).

3.3.6 – Methyl-CpG-binding protein (*MECP2*)

MECP2 encodes a nuclear protein that primarily functions as a reader of the DNA methylome and global regulator of gene expression. MECP2 binds to methylated DNA sequences, specifically methylated CG dinucleotides (mCG) and mCAC, through interaction with the methyl-CpG binding domain (MBD) (Tillotson and Bird, 2020). Once bound, MECP2 is able to recruit several binding partners to mediate transcriptional repression. One such partner is the NCoR1/2 co-repressor complex, to which MECP2 binds through the transcriptional repression domain (TRD) and regulates gene expression through deacetylation of histone proteins via a core subunit, HDAC3 (Tillotson et al., 2021; Tillotson and Bird, 2020). MECP2 can also regulate gene expression post-transcriptionally by interfering with the ability of DGCR8, a central component of the nuclear miRNA processing machinery, to complex with Drosha. This complex is responsible for processing primary miRNA transcripts into mature precursor miRNAs, including those involved in neuronal development and plasticity (Cheng et al., 2014).

De novo mutations in *MECP2* cause Rett syndrome (RTT; OMIM: #312750), a severe X-linked neurological disorder that primarily affects females and is characterized by cognitive and

motor delay, irregular hand movements, ataxia, as well as autonomic dysfunction (Ip et al., 2018). C>T transitions at a handful of CpG dinucleotides in MECP2 underlie RTT; both missense and nonsense mutations are primarily located in the MBD and TRD, disrupting proper DNA binding and interactions with co-repressor complexes, respectively. Located in the TRD, *MECP2* c.502C>T (p.R168*) is among the most prevalent mutations in RTT and was detected 11 times in the Kaplanis study. p.R168* results in a truncated protein that retains the ability to bind to methylated DNA but loses the ability to interact with corepressors such as SIN3A resulting in aberrant gene expression (Matijevic et al., 2009). By contrast, the c.397C>T (p.R133C) mutation is located in the MBD, and has been shown to reduce methyl-CpG binding affinity and has less severe clinical phenotypes (Krishnaraj et al., 2017). Whereas mice with global deletion of MECP2 had shortened survival and recapitulated a range of RTT features, including impaired locomotion, tremors, and breathing difficulties, mice with MECP2 expression restricted to the nervous system lacked an overt RTT phenotype indicating that the central nervous system is the primary site of action for MECP2 (Ross et al., 2016).

MECP2 is not known to regulate SSCs but was found to be expressed in undifferentiated spermatogonia at the RNA level by scRNAseq (**Figure 3.1**) and at the protein level by IHC (**Figure 3.2**). Functional studies indicate that MECP2 mutations in RTT directly modulate key signalling pathways involved in SSC homeostasis. RTT mouse models have demonstrated that MECP2 deficiency results in the upregulation of miR-214 in self-renewing neural progenitors, which in turn reduces PTEN expression and results in AKT pathway activation (Mellios et al., 2018). Brain organoids harbouring MECP2 mutations also exhibit high levels of phosphorylated SMAD1/5 (**section 3.3.3**) as well as upregulation of genes involved in the BMP pathway.

MECP2 mutations in RTT are predominantly of paternal origin, with one case series demonstrating that 26/27 RTT cases (96.3%) were paternally derived (Trappe et al., 2001). However, RTT fathers in this same case series were not older than the German general population, a finding that has been replicated in a much larger survey (n = 1226 RTT probands) (Fang et al., 2023). This, together with recent evidence indicating that MECP2 germline mosaicism is common (~28%) in the sperm of both RTT fathers and healthy adult males (Zhang et al., 2019; Wen et al., 2023), indicates that positive selection is unlikely to underlie the high rate of recurrence in RTT.

3.3.7 – NatA catalytic subunit 10 (*NAA10*)

N-alpha-acetylation is one of the most common protein modifications in eukaryotes and involves the transfer of an acetyl group from acetyl-coenzyme A (Ac-CoA) to the N-terminus of the protein substrate, a process which is catalysed by N-terminal acetyltransferases (NATs) (Arnesen et al., 2009). *NAA10* encodes the catalytic subunit of NatA, the major NAT complex responsible for ~40% of N-acetylations in humans. The active NatA complex is formed once NAA10 binds to its auxiliary subunit NAA15; NAA10 is responsible for transferring the acetyl group from Ac-CoA to the protein substrate (Cheng et al., 2019; Ree et al., 2019). N-terminal acetylation is involved in mediating a range of cellular processes such as protein folding and protein-protein interactions that are essential in normal neurogenesis and bone formation (Holmes et al., 2014; Lee et al., 2018)

Mutations in the X-linked *NAA10* gene cause a developmental disorder with a range of disease severity, collectively termed NAA10-related N-terminal acetylation deficiency (NAA deficiency). The most severe form was originally described in hemizygous males from two unrelated families with the c.109T>C (p.S37P) mutation, initially called Ogden syndrome (Rope et al., 2011). While this mutation was lethal for male probands, carrier mothers had no or mild

symptoms, presumably due to X-chromosome inactivation. Subsequently, numerous other mutations in the N-acetyltransferase domain of *NAA10* were described in both males and females with a milder phenotype. Among these, the *de novo* c.247C>T (p.R83C) mutation is highly recurrent and was detected 11 times in the Kaplanis study. Individuals with this mutation are characterized by intellectual and developmental delay, growth retardation, muscular hypotonia, cardiac anomalies, and heterogeneous facial dysmorphia (Ree et al., 2019; Saunier et al., 2016). The c.247C>T (p.R83C) mutation occurs at a CpG dinucleotide within the Ac-CoA binding site of NAA10, suggesting a potential effect on Ac-CoA binding and transfer. However, functional studies have yielded conflicting results. Whereas Saunier found that p.R83C reduced the catalytic capacity of monomeric NAA10, Cheng studied the p.R83C mutation in the context of a fully assembled NatA complex and showed that the mutation increased the catalytic activity of the complex (Saunier et al., 2016; Cheng et al., 2019). Cheng proposed that the p.R83C mutation may enhance enzymatic activity by promoting NAA15 binding and improving the positioning of Ac-CoA for catalysis.

NAA10 does not have a known role in SSC regulation but was found to be expressed in undifferentiated human spermatogonia at the RNA level by scRNAseq (**Figure 3.1**) and at the protein level by IHC (**Figure 3.2**), as well as in pre-meiotic spermatogonia in mice (Lee et al., 2018). Although a role in spermatogenesis has not been elucidated, NAA10 has been shown to acetylate β-catenin, a key component of Wnt/β-catenin signaling which is upregulated in human SSCs (Guo et al., 2017). Site-directed mutagenesis at the site of auto-acetylation, p.K136, disrupts downstream activation of β-catenin and subsequent tumorigenesis in xenograft models of lung cancer; however, it remains unclear how mutations within the Ac-CoA binding site (such as p.83C) influence the Wnt pathway (Seo et al., 2010). Although a novel variant, *NAA10* c.293A>G

(p.M98T), was detected at 0.82% VAF in a single testis biopsy from a 57-year-old man (Ralph, 2018), mutations at this residue have not been reported in *NAA10*-related deficiency, nor in COSMIC.

The majority of *de novo* cases of *NAA10*-related deficiency are female, with the largest case series reporting 19 affected females and 1 affected male with *de novo* missense mutations in *NAA10* (Cheng et al., 2019). While this would imply a paternal bias in the origin of mutation approaching that of classical PAE disorders (typically $\propto > 20$) (**Table 1.1**), this interpretation is complicated by the fact that male probands have a more severe clinical phenotype and experience early or prenatal death (Saunier et al., 2016; Rope et al., 2011), an observation that is well supported by functional studies of NAA10 knockout mice exhibiting both neonatal and embryonic lethality (Lee et al., 2017; Kweon et al., 2021). The true incidence of males with *NAA10*-deficiency is therefore likely underreported due to limitations in clinical ascertainment, thus challenging an extreme paternal bias in the origin of mutation.

3.3.8 – Protein phosphatase 1 catalytic subunit beta (*PPP1CB*)

PPP1CB encodes the beta catalytic subunit of protein phosphatase 1 (PP1), a ubiquitous serine/threonine phosphatase that catalyzes the majority of protein dephosphorylation events in eukaryotes (Bollen et al., 2010). PP1 has two other catalytic subunits: alpha, encoded by *PPP1CA*, and gamma, encoded by *PPP1CC*. The PP1 subunits contain short docking motifs (4 – 8) that bind to PP1-interacting proteins (PIPs) to determine the substrate specificity of the phosphatase. PP1 interacts with ~200 distinct PIPs to regulate a range of cellular processes including cell cycle progression and protein and mRNA processing, and has been implicated in cancer, cardiometabolic, and neurologic diseases (Bollen et al., 2010; Ferreira et al., 2019). PP1 regulates

numerous key signalling pathways and notably complexes with SHOC2 (**section 3.3.12**) and MRAS to dephosphorylate the inhibitory sites of RAF kinases (ARAF p.S214, BRAF p.S365, and CRAF p.S259) leading to MAPK activation (Rodriguez-Viciana et al., 2006).

De novo mutations in *PPP1CB* cause Noonan-like syndrome with loose anagen hair 2 (NS-LAH2; OMIM: #617506), which is characterized by macrocephaly, developmental delay, irregular hair development, and characteristic low-set ears (Gripp et al., 2016). Mutations within the catalytic core of PPP1CB (p.P41 – P269) – which is essential for phosphatase activity – underlie NS-LAH2, with the c.146C>G (p.P49R) mutation as the most common (Ma et al., 2016). The p.P49R mutation was observed 11 times in the Kaplanis study and is located at the PP1 interface between SHOC2 and MRAS. Functional studies have shown that the p.P49R mutation increases the affinity and stability of the ternary complex, likely leading to MAPK activation through enhanced RAF dephosphorylation, maintaining it in an active state (Young et al., 2018; Hauseman et al., 2022).

Although PPP1CB does not have a known role in regulating spermatogenesis, PP1-interacting proteins have been implicated in murine SSC survival and maturation. One such PP1-interacting protein is NIPP1 which, when deleted postnatally, results in a Sertoli-cell only phenotype as germ cell proliferation decreases in concert with an increase in apoptosis (Ferreira et al., 2017). Moreover, the deletion of PPP1CC of male germ cells in mice results in reduced sperm count and abnormal morphology, as well as infertility (Sinha et al., 2013). PPP1CB was found to be expressed in undifferentiated human spermatogonia at the RNA level by scRNAseq (**Figure 3.1**) and at the protein level by IHC (**Figure 3.2**). In addition to the direct evidence that the PPP1CB p.P49R mutation activates the MAPK pathway, PPP1CB also regulates AKT

phosphorylation at p.T450 (Xiao et al., 2010), indicating possible mechanisms for influencing SSC homeostasis (Ma et al., 2017).

There are six cases of NS-LAH2 caused by mutations in *PPP1CB* with documented paternal age (mean 34 years; range 24–43 years) (**Table 3.2**). Although parent-of-origin studies for NS-LAH2 are lacking, it is worth noting that PTPN11 mutations in the related Noonan syndrome are exclusively found on the paternal allele (Tartaglia et al., 2004).

3.3.9 – Cyclin-dependent kinase 13 (*CDK13*)

CDK13 is a member of the cyclin-dependent serine/threonine protein kinase family that regulates gene expression at both the transcriptional and translational levels. CDK13 binds to cyclin K (CCNK) to form an active kinase complex that is best known for phosphorylating key residues (p.S2 and p.S5) within the C-terminal domain (CTD) of RNA polymerase II (Pol II), which regulates transcriptional elongation and pre-mRNA processing (Greifenberg et al., 2016; Fan et al., 2020). CDK13 has also been shown to engage with and phosphorylate proteins involved in translation initiation, such as components of the eIF4F complex (Wu et al., 2023). Both genetic knockdown and CDK13-specific inhibition decreased phosphorylation of eIF4F complex subunits and led to decreased synthesis of proteins such as the Myc oncogene – which has been shown to regulate SSC self-renewal in mice – which impaired tumor growth *in vivo.*

De novo mutations in the protein kinase domain (p.705 – 998) of *CDK13* cause congenital heart defects, dysmorphic facial features, and intellectual development disorder (CHDFIDD; OMIM: #617360). The c.2525A>G (p.N842S) variant was observed ten times in the Kaplanis parent-offspring trio study and is also the most recurrently reported mutation in CHDFIDD (Hamilton and Mohnish, 2019). *In silico* modelling of this mutation predicts a structural change to

the ligand binding pocket of CDK13, compromising ATP binding and ultimately disrupting its ability to phosphorylate Pol II (Hamilton et al., 2018). CCNK binding is also predicted to remain intact allowing CDK13 p.N842S to sequester CCNK and prevent activation of wild-type CDK13, suggestive of a dominant-negative mechanism of action. Indeed, individuals with CDK13 loss-of-function mutations have been reported in the gnomAD exome database, indicating that haploinsufficiency can be tolerated. It has been additionally noted that CDK13 mutations such as p.N842S occur at highly conserved positions across protein kinases; CHDFIDD mutations at equivalent positions in kinases such as RAF1 result in loss of catalytic activity and dominant disorder, Noonan syndrome type 5 (OMIM: #611553) (Hamilton et al., 2018).

Although CDK13 does not have a known role in SSC homeostasis, both CDK13 and CCNK are highly expressed in mouse SSCs, but not in differentiated spermatogonia (Xiang et al., 2014). In humans, CDK13 is expressed in undifferentiated spermatogonia at the RNA level by scRNAseq (**Figure 3.1**) but was not detected at the protein level in testes by IHC in the HPA (**Figure 3.2**). Pol II, whose activity is mediated by CDK13, undergoes dynamic changes during murine spermatogenesis, and its disruption affects spermatocyte and spermatid development (Kaye et al., 2024). Additionally, POLR2A, the largest Pol II subunit, has been shown to regulate gene expression in SSCs (Sisakhtnezhad and Heshmati, 2017). Consistent with a loss of catalytic activity, CDK13 depletion enriches for RTK signaling components (including FGF19 and PIK3CD), and CDK12/CDK13 degradation results in AKT pathway activation (Greifenberg et al., 2016; Chang et al., 2024), together indicating potential mechanisms by which CDK13 mutations may influence SSC homeostasis.

3.3.10 – Casein kinase 2α subunit *(CSNK2A1)*

CSNK2A1 encodes the α catalytic subunit of CK2, a ubiquitous tetrameric serine/threonine kinase comprised of two catalytic (α and/or α', encoded by *CSNK2A2*) and two regulatory (β) subunits (*CSNK2B*) (Borgo et al., 2021). While the β subunit contributes to holoenzyme stability and substrate specificity, the catalytic subunits mediate the phosphorylation of hundreds of protein substrates. CK2 regulates several key signalling pathways. For example, CK2 phosphorylates two serine residues (p.S244 and p.T246) within the nuclear translocation signal (NTS) of ERK1/2 to create a negatively charged patch, which then permits binding to importin 7, a nuclear import protein (Plotnikkov et al., 2019). CK2 also phosphorylates AKT at p.S129, enhancing its activity, and phosphorylates PTEN to inhibit phosphatase activity, thereby amplifying PI3K/AKT signaling in a two-pronged manner (Borgo et al., 2021). Components of the Wnt pathway are also positively regulated by CK2 phosphorylation, including direct phosphorylation of *β*-catenin at p.T393 which enhances the stability of the protein, and increases transcriptional activity (Song et al., 2003). CK2 is known to regulate a wide array of biological processes during development such as neurogenesis, cardiogenesis, spermatogenesis, and osteogenesis (Halloran et al., 2022).

De novo mutations in *CSNK2A1* cause Okur-Chung neurodevelopmental syndrome (OCNDS; OMIM #617062), a dominant developmental disorder characterized by intellectual disability, developmental delay, hypotonia, short stature, and facial dysmorphia (Okur et al., 2016). DNMs cluster within the protein kinase domain (p.Y39–p.F324) of *CSNK2A1*, with the commonest being c.593A>G (p.K198R). The p.K198R mutation was detected 10 times in the Kaplanis parent-offspring trio study and is located within the P+1 loop of the activation segment that mediates substrate recognition (Werner et al., 2022). OCNDS mutations in *CSNK2A1* were initially thought to cause a loss of function, as early studies demonstrated reduced kinase activity (Dominguez et al., 2021). More recent studies of the p.K198R mutation show that $CK2^{K198R}$ is an

active kinase but displays altered substrate specificity, such as decreasing the preference for threonine phosphorylation and potentially increasing the preference for tyrosine phosphorylation, which would represent a notable gain of function for a ubiquitous serine/threonine kinase (Caefer et al., 2022; Werner et al., 2022). Although specific tyrosine residues have not been noted, OCNDS mouse models carrying the p.K198R mutation show increased phosphorylation of AKT p.S473, indicative of PI3K/AKT activation (Cruz-Gamero et al., 2024).

CSNK2A1 does not have a known role in the maintenance of SSCs but has been shown to be highly and specifically expressed in murine spermatogonia (Xu et al., 1999). Whereas expression of CSNK2A1 decreases as cells differentiate, CSNK2A2 is preferentially expressed in the later stages of spermatogenesis. Moreover, deficiency of *CSNK2A2* has been shown to dysregulate spermatogenesis, resulting in globozoospermia and infertility in mice (Escalier et al., 2003; Xu et al., 1999). In humans, CSNK2A1 was found to be expressed in undifferentiated spermatogonia at both the RNA level by scRNAseq (**Figure 3.1**) and at the protein level by IHC (**Figure 3.2**). Given that AKT, MAPK, and Wnt are critical pathways for SSC homeostasis, the capacity of CK2 to modulate these signalling networks suggests that OCNDS mutations in CSNK2A1 may also influence SSC dynamics.

3.3.11 – Forkhead box P (Fox) transcription factor family protein 1 (*FOXP1*)

FOXP1 belongs to the forkhead family of transcription factors which have an established role in the regulation of cell fate and differentiation across tissues (Li et al., 2004). FOXP1 was first identified as a transcriptional repressor in lung-specific gene expression (Shu et al., 2001) and has since been established as a regulator of embryogenesis, immune development, cardiogenesis, and neurogenesis (Braccioli et al., 2017; Rousso et al., 2008; Hu et al., 2006). FOXP1 binds to DNA

through its Forkhead DNA-binding domain, targeting specific gene promoter and enhancer regions, and recruiting co-repressor proteins such as CTBP1 (**section 3.3.12**) (Rocca et al., 2017). CTBP1 recruits LSD1 and HDAC2 to form a co-repressor complex to regulate histone modification (Shu et al., 2001). DNA binding activity requires the formation of either FOXP1 homodimers (mediated by the leucine zipper domain) or heterodimerization with other FOXP family members (Li et al., 2004).

De novo mutations in *FOXP1* cause FOXP1 syndrome, which is characterized by intellectual disability, speech and language impairment, and global developmental delay (Meerschaut et al., 2017; Le Fevre et al., 2013). The *FOXP1* c.1541G>A (p.R514H) mutation is the most recurrent FOXP1 syndrome mutation (reported 9 times in the Kaplanis study). Along with the c.1540C>T (p.R514C) mutation (observed once), p.R514H affects a CpG dinucleotide within the Forkhead DNA-binding domain (Sollis et al., 2016). This position is highly conserved across 49/50 FOX proteins with the equivalent residue in FOXP2 (p.R553) located in the DNA recognition helix, therefore mutations are thought to disrupt DNA binding. Consistent with this, functional studies show that the p.514C mutation disrupts the transcriptional repression activity of FOXP1. Although haploinsufficiency has been proposed as the underlying mechanism of action (Le Fevre et al., 2013), the mutant FOXP1 protein retains the ability to interact with wild-type FOXP1 and sequesters them in nuclear and cytoplasmic aggregates, presumably abolishing transcriptional repression due to disrupted subcellular localization, in a dominant-negative manner (Sollis et al., 2016).

FOXP1 does not have an established role in SSC regulation but exhibits selective expression in GFRA1+ type A spermatogonia in mice (Prokai et al., 2020). In humans, FOXP1 was detected in undifferentiated spermatogonia at the RNA level by scRNAseq (**Figure 3.1**) and

at the protein level by IHC (**Figure 3.2**). FOXP1 plays important regulatory roles in other stem cell populations, promoting symmetric self-renewal in neural stem cells (Pearson et al., 2020) and maintaining the stem cell pool in hematopoietic stem cells (Naudin et al., 2017). While mutations like p.R514C that disrupt FOXP1 function might be expected to impair stem cell maintenance based on FOXP1's roles in other stem cell types, extrapolating FOXP1's function in SSCs from other tissues is complicated by its context-dependent regulation of key pathways. For example, FOXP1 enhances the expression of ALK5, a type I BMP/TGF-β receptor, to promote osteoblast differentiation but represses ALK5 in cardiac endothelial cells during remodelling (Zhang et al., 2022; Liu et al., 2019).

3.3.12 – C-terminal binding protein (*CTBP1*)

CTBP1 functions as a transcriptional repressor by interacting with DNA-binding transcription factors through PXDLS motifs (where X can represent any amino acid) within their promoter regions (Chinnadurai, 2002). Upon binding, CTBP1 recruits a suite of chromatin-modifying enzymes including histone deacetylases (HDAC1/2), histone methyltransferases (G9A and GLA), and histone H3-K4 demethylases (LSD1) to facilitate gene silencing through chromatin remodelling. Although the exact mechanisms remain unclear, CTBP1 appears to play a critical role in embryonic development as CTBP1-null mice are viable, but display impaired muscular and skeletal development, and often experience early death (Hildebrand and Soriano, 2002). CTBP1 notably regulates the Wnt pathway by binding to PXDLS motifs on the TCF-4 transcription factor to repress its activity; binding by *β*-catenin displaces the CTBP1 repressor complex and transcriptional repression (Chinnadurai, 2002).

A single *de novo* mutation in CTBP1 causes Hypotonia, ataxia, developmental delay, and tooth enamel defect syndrome (HADDTS; OMIM: #617915), which additionally presents with intellectual delay (Beck et al., 2016). The *CTBP1* c.1024C>T (p.R342W) variant occurs at a CpG dinucleotide located in the PXDLS binding cleft at the C-terminus of the CTBP1 protein and was therefore predicted to disrupt interactions with chromatin-modifying partners, thus altering gene expression. Given the highly restricted mutational spectrum of HADDTS, and because loss-of-function mutations have been detected in unaffected individuals, the p.R342W mutation is thought to act in a dominant-negative or gain-of-function manner (Beck et al., 2016). Functional studies of the p.R342W mutation confirmed this hypothesis, showing reduced interactions across known repressor complex cofactors and deregulation of gene expression across multiple cellular processes (Beck et al., 2019; Vijayalingam et al., 2019). For example, genes involved in neurodevelopment (e.g. synapse formation, axon guidance) were notably downregulated, while pro-apoptotic genes were upregulated. Interestingly, Gene Ontology analysis of identified TGF-β signalling among the most downregulated gene sets in p.R342W cells compared to wild-type cells (Beck et al., 2019).

CTBP1 does not have a known role in SSC regulation but is expressed in undifferentiated spermatogonia at the RNA level by scRNAseq (**Figure 3.2**) and at the protein level by IHC (**Figure 3.3**). Although CTBP1 is known to regulate the Wnt pathway, functional studies of the p.R342W mutation did not report significant changes in Wnt pathway gene expression. While down-regulation of TGF-β signaling was noted, how CTBP1 p.R342W might influence SSC homeostasis remains unresolved.

3.3.13 – Leucine rich repeat scaffold protein (*SHOC2*)

SHOC2 is a scaffolding protein that primarily contains 20 tandem leucine-rich repeat (LRR) motifs, which are known to mediate protein-protein interactions. Although early studies suggested that SHOC2 binds to Ras proteins via its LRR motifs (Sieburth et al., 1998), more recent studies demonstrate that SHOC2 displays high specificity for MRAS, likely mediated by non-LRR switch regions (Rodriguez-Viciana et al., 2006; Young et al., 2018). SHOC2 forms a SHOC2-MRAS-PPP1CB ternary complex in a GTP-dependent manner; upon GTP hydrolysis, MRAS undergoes a conformational change that introduces steric hindrance at the switch regions, impairing complex formation and negatively regulating pathway activity (Liau et al., 2022). The phosphatase complex dephosphorylates inhibitory RAF sites (**section 3.3.8**), displacing inhibitory 14-3-3 proteins and facilitating Ras-RAF binding (Rodriguez-Viciana et al., 2006).

De novo mutations in *SHOC2* cause Noonan-like syndrome with loose anagen hair 1 (NS-LAH1; OMIM: # 607721) which presents with the characteristic Noonan-like facial phenotype, short stature, and a short and webbed neck (Mazzanti et al., 2003). Notably, PPP1CB mutations result in a related Noonan-like phenotype with overlapping clinical features that reflect their shared role in MAPK pathway regulation. The most frequent NS-LAH1 mutation, *SHOC2* c.4A>G (p.S2G), was identified in 9 cases in the Kaplanis study. The p.S2G mutation introduces a novel N-myristolyation site that drives constitutive localization to the plasma membrane by acting as a lipid anchor (Cordeddu et al., 2009). Wild-type SHOC2 is typically distributed in the cytoplasm and nucleus, only transiently moving to the plasma membrane upon growth factor receptor stimulation. In contrast, because MRAS resides at the plasma membrane, the p.S2G mutation promotes continuous binding and complex formation with PPP1CB. The p.S2G mutation also impairs nuclear import of SHOC2 (a process involving LRR motifs), allowing prolonged cytoplasmic ERK phosphorylation and establishing a gain-of-function mechanism that sustains

pathway activation (Motta et al., 2016). Additionally, while SHOC2 p.S2G does not directly activate AKT signalling, it can potentiate AKT-mediated ERK1/2 activation upon EGF stimulation (Wilson et al., 2024).

SHOC2 does not have a known role in spermatogenesis but is expressed in undifferentiated spermatogonia as shown by scRNAseq (**Figure 3.1**) and IHC (**Figure 3.2**). Given the clear regulatory role that SHOC2 plays in MAPK signalling, there is a clear mechanistic rationale for the activating p.S2G mutation to influence SSC homeostasis. Consistent with this, paternal age has been reported for seven NS-LAH1 cases (mean 39.1 years; range 28–52 years), suggestive of advanced paternal age (**Table 3.2**).

3.3.14 – Snf2-related CBP activator protein (*SRCAP*)

SRCAP encodes a subunit of an ATP-dependent chromatin remodeler, the SRCAP-contained remodelling complex (SRCAP-C). Comprised of an actin-related protein (ARP) module and an ATPase-containing motor module (Yu et al., 2024), SRCAP-C catalyzes the ATP-dependent exchange of the canonical H2A with the H2A.Z variant histone proteins. The nucleosome, the fundamental subunit of chromatin, is composed of two copies each of H2A and H2B, which lie on the nucleosome surface, and H3 and H4 form the core (Gaimo et al., 2019). Incorporation of histone variants such as H2A.Z influences several processes including DNA repair, transcription, and cell cycle control (Colino-Sanguino et al., 2022). SRCAP most notably regulates CBP, a ubiquitous acetyltransferase that coordinates transcriptional responses to numerous key signalling cascades including the calcium, NFκB, and Notch pathways (Dancy and Cole, 2015).

De novo mutations in *SRCAP* cause Floating-Harbor syndrome (FHS; OMIM: #136140) which is characterized by a triangular face with a bulbous nose and large ears, speech and language

delay, and skeletal and growth defects (Nikkel et al., 2013). FHS mutations in *SRCAP* cluster in the terminal 34th exon and are all truncating (Hood et al., 2012). Two recurrent truncating mutations dominate the mutational spectrum: *SRCAP* c.7303C>T (p.R2435*) which was detected 9 times in the Kaplanis study, and c.7330C>T (p.R2444*), which did not meet the threshold for recurrence (n = 4), both occurring at CpG dinucleotides (Hood et al., 2012). The highly clustered nature of exon 34 mutations and the specificity of the phenotype, as well as the presence of healthy individuals with structural deletions encompassing SRCAP (Hood et al., 2012) suggest that these mutations act through a dominant-negative mechanism. These truncating mutations eliminate the C-terminus (which contains AT-hook motifs essential for DNA binding) and may disrupt normal binding to chromatin through competition with wild-type SRCAP or form inactive heterodimers resulting in a loss-of-function (Messina et al., 2016). SRCAP knockout models exhibit reduced H2A.Z deposition in promoter regions of target genes, resulting in dysregulated gene expression (Wong et al., 2007), and although depletion of H2A.Z can activate the Notch pathway (Giaimo et al., 2018), this effect has not been reported in FHS-related functional studies. Individuals with FHS exhibit DNA methylation signatures enriched for terms involved in DNA recombination, DNA repair, and chromosome organization, underscoring SRCAP's role in chromatin regulation (Rots et al., 2021).

SRCAP has no established function in the regulation of SSCs and is not expressed in undifferentiated spermatogonia at the RNA-level by scRNAseq (**Figure 3.1**) or at the protein-level by IHC (**Figure 3.2**). Its key substrate, H2A.Z, appears to influence chromosomal reorganization during spermatogenesis, but is notably absent in murine spermatogonia, only becoming detectable in spermatocytes (Greaves et al., 2006). Since SSC expression is a key requirement for selfish

spermatogonial selection, it is unlikely that the high recurrence observed in SRCAP p.R2435* mutations is driven by a fitness advantage in mutant SSCs.

3.3.15 – Asparagine-linked glycosylation 13 homology (*ALG13)*

ALG13 encodes the catalytic subunit of a hetero-oligomeric UDP-N-acetylglucosamine transferase, an enzyme critical for N-linked glycosylation, a common post-translational modification where an oligosaccharide is covalently attached to an asparagine residue (Averback et al., 2008). ALG13 engages ALG14, which serves as a membrane anchor, and catalyzes one step of lipid-linked oligosaccharide (LLO) synthesis on the surface of the endoplasmic reticulum. Specifically, ALG13 mediates transfer of the sugar donor N-acetylglucosamine (GlcNAc) onto dolichol, a lipid anchor, which is then translocated across the ER membrane into the lumen for further assembly. Once assembled, the oligosaccharide is transferred to select asparagine residues on the nascent protein in the ER lumen (Gao et al., 2005; Ng et al., 2020). N-linked glycosylation has been linked to several critical cellular processes including protein folding and stability, and cell signalling. For example, LRP6 – a Wnt co-receptor – requires N-glycosylation at multiple sites to ensure proper protein folding, cell surface trafficking, and modulation of receptor activity (Xu et al., 2023). Similarly, GDNF, the GFRA1 ligand, requires N-glycosylation for proper protein folding and cleavage of pro-GDNF during maturation (Cintrón-Colón et al., 2020). While this modification is not wholly required for RET/GFRα1 receptor activation, unglycosylated GDNF binds to the receptor complex with lower affinity and results in less efficient activation (Piccinini et al., 2012).

De novo mutations in *ALG13* cause an X-linked disorder of asparagine-linked (N-linked) glycosylation (ALG13-CDG; OMIM: #300884), which is characterized by global developmental

delay, seizures and epilepsy, intellectual disability, and hypotonia (Ng et al., 2020). The *ALG13* c.320A>G (p.N107S) mutation was observed eight times in the Kaplanis parent-offspring trio study and is the most recurrent mutation in ALG13-CDG and primarily affects females. The precise mechanism by which DNMs in *ALG13* result in the ALG13-CDG phenotype is not fully understood, however *in silico* analysis of the ALG13 protein indicates that the p.N107S mutation may affect the structure of the active site and therefore the affinity or specificity of the mutant Alg13 protein for the developing glycan. Functional studies suggest that the ALG13 p.N107S mutation results in an almost complete loss of glycosyltransferase enzymatic activity disrupting LLO synthesis yet, surprisingly, ALG13-CDG patients exhibit normal glycosylation patterns in the serum challenging a simple haploinsufficiency model (Galama et al., 2017). ALG13-CDG mouse models recapitulate this finding, with no obvious manifestations of N-glycosylation defects (Gao et al., 2019). Instead, hyperphosphorylation of mTOR at p.S2448, indicative of upregulated AKT signalling, has been proposed to underly the epileptic phenotype.

ALG13 does not have a known role in SSC regulation, but ALG13 copy number variants and indels have been associated with abnormal spermatogenesis and infertility (Halder et al., 2017; Malcher et al., 2022). ALG13 is expressed in undifferentiated spermatogonia at the RNA level by scRNAseq (**Figure 3.1**) and at the protein level by IHC (**Figure 3.2**). Although N-glycosylation plays important regulatory roles in several signalling pathways associated with SSC homeostasis, direct evidence that the AKT pathway is upregulated offers a direct mechanism to affect SSC homeostasis. Parent-of-origin studies are lacking, but notably, 2 of 42 *de novo* ALG13-CDG cases in the largest case series (Ng et al., 2021) were male, implying an extreme paternal bias ($\propto = 21$). for the X-linked ALG13 mutations. However, unlike the male infants with Ogden syndrome (due

to mutations in *NAA10*), early death has not observed, thus complicating ascertainment. Nevertheless, the X-linked nature of these mutations complicate interpretation.

3.3.16 – Glutamate receptor, ionotropic, N-methyl, D-aspartate, subunit 2B (*GRIN2B*)

GRIN2B encodes the GluN2B subunit of N-methyl-D-aspartate receptors (NMDADRs), glutamate-gated ion channel receptors that mediate excitatory neurotransmission in the central nervous system (Hansen et al., 2017). NMDARs are heterotetrametric complexes that consist of one GluN1 subunit with either GluN2 subunits or a mixture of GluN2 and GluN3 subunits; four distinct GluN2 subunits (GluN2A–D) and two GluN3 subunits (GluN3A–B) are known (Paoletti et al., 2013). Among these, the GluN2 subunits are the major determinants of the receptor's functional heterogeneity, spatiotemporal expression, and binding specificity. Specifically, GluN2B-containing receptors recruit unique signalling molecules, such as CaMKII, essential for long-term potentiation and memory formation (Paoletti et al., 2013). Loss of GluN2B impairs synaptic plasticity, reduces dendritic density, and leads to learning and memory defects in mice (Brigman et al., 2010).

De novo mutations in *GRIN2B* cause Developmental and Epileptic Encephalopathy 27 (DEE27; OMIM: #616139) which is characterized by severe intellectual delay, autism spectrum disorder, epilepsy, hypotonia, and loss of muscle control (Platzer et al., 2017). Mutations affecting the *GRIN2B* p.G280 residue were detected in the Kaplanis parent-offspring trio study 7 times (c.2459G>C (p.G820A), n = 6; c.2459G>T (p.G820V), n = 1). The p.G820 residue is located in the M4 region of the transmembrane domain (TMD), which forms the ion channel pore (Hu et al., 2016; Shah et al., 2022). While the precise mechanism by which the p.G820A mutation contributes to disease etiology is incompletely understood, *in silico* analyses suggest it may enhance protein

stability, while other mutations in the TMD of GluN2B have been shown to disrupt ion channel gating (Swanger et al., 2016).

GRIN2B does not have a known role in SSC homeostasis and although related NMDAR subunits GluN1 and GluN2A appear to mediate ERK activation in murine spermatogonia, a role for GluN2B has not been reported (Santillo et al., 2013). Consistent with this, undifferentiated human spermatogonia do not express GRIN2B at the RNA level by scRNAseq (**Figure 3.1**) or at the protein level by IHC (**Figure 3.2**). Given that GRIN2B expression is absent in SSCs, it is unlikely that GRIN2B mutations confer a selective advantage in the male germline.

3.3.17 – Guanine nucleotide-binding protein subunit beta-1 (*GNB1*)

GNB1 encodes the Gβ1 protein which complexes with the Gα and Gγ subunits to form heterotrimeric G proteins. Ligand binding to G protein-coupled receptors (GPCR) initiates interactions with the G protein complex, prompting the Gα subunit to exchange GDP for GTP, thus acting as a molecular switch (Clapham, 1997). In the active GTP-bound state, Gα dissociates from the Gβγ dimer to interact with downstream effector proteins. The Gβγ dimer serves multiple functions, including membrane anchoring to facilitate interactions with GPCRs and regulation of Gα affinity for GDP, GTP, and receptors (Clapham, 1997). Following G protein activation and Gα dissociation, Gβγ can directly modulate several effector molecules including ERK1/2 and PI3Kγ. Gβγ binds directly to the catalytic subunit of PI3Kγ, p110γ, which forms a dimer with the p101 regulatory subunit (Nürnberg and Beer-Hammer, 2019). Activated PI3Kγ can translocate to the plasma membrane, stimulate PIP3 production and Akt activation (Nürnberg and Beer-Hammer, 2019), and recruit ERK1/2 to the Golgi apparatus for phosphorylation, resulting in MAPK activation (Khater et al., 2020).

De novo mutations in *GNB1* cause Mental Retardation, Autosomal Dominant 42 (MRD42; OMIM: #616973), a developmental disorder characterized by global developmental delay, hypotonia, seizures, ophthalmological manifestations, and growth defects (Petrovski et al., 2016). DNMs cluster within exons 6 and 7 at the N-terminal interface that makes contact with the Gα subunit (Ford et al., 1998). The most recurrent DNMs affect the *GNB1* p.I80 codon (~25% of cases) and were documented 7 times in the Kaplanis parent-offspring trio study: *GNB1* c.239T>C (p.I80T) was detected six times and c.239T>A (p.I80N) once. At least 40 additional mutations have been identified, with ~90% occurring in exons 6 and 7 (Da Silva et al., 2021). Activating mutations in *GNB1*, including the recurrent p.I80T and p.I80N variants, have also been described in various hematologic malignancies. These mutations exhibit reduced binding to Gα but not Gγ subunits, thus liberating the Gβγ dimer and constitutively activating downstream signalling pathways (Yoda et al., 2015). GNB1 mutations (including those at p.I80) result in elevated phosphorylation of AKT, MEK, and ERK, indicating activation of AKT and MAPK pathways in a gain-of-function manner.

Although GNB1 does not have a known role in SSC homeostasis, GPCRs are known to regulate stem cell dynamics in multiple contexts, including pluripotent stem cells during murine development (Chevalier et al., 2021) and in prespermatogonia in drosophila (Malpe et al., 2020), and control critical signaling pathways involved in SSC homeostasis. GNB1 is expressed in undifferentiated spermatogonia at the RNA level by scRNAseq (**Figure 3.1**) but was not detected at the protein level by IHC (**Figure 3.2**). The ability for MRD42 GNB1 mutations to activate both MAPK and AKT pathways suggest a possible mechanism by which mutations could influence SSC homeostasis.

Table 3.2: Evidence for selfish potential across prioritized candidate genes based on mutation patterns and PAE characteristics

Gene	cDNA mutation (Amino acid change)	Germline disorder	Likely mechanism[1]	COSMIC v101[2]	Paternal mutations (n)	Maternal mutations (n)	Mean paternal age (n reports; range)	High apparent mutation rate[3]	References
PACS1	c.607C>T (p.R203W)	Schuurs-Hoeijmakers syndrome (OMIM: #615009)	GOF	4 (4, 5)	1	0	40.7 years (6; 26–51 years)	-	Gadziki et al., 2015 Miyake et al., 2017 Pefkianaki et al., 2018 Abbassi et al., 2023 Corona-Rivera et al., 2023 Brown et al., 2023
PPP2R5D	c.592G>A (p.E198K)	Jordan syndrome (OMIM: #616355)	Dominant-negative	3 (3, 3)	0	0	-	Testis	Ralph, 2018
SMAD4[4]	c.1486C>T (p.R496C)	Myhre syndrome (OMIM: #139210)	GOF	7 (7, 30)	0	0	36.1 years (61; 22–49 years)	Testis	Ralph, 2018 Full references in **Supplementary Table S1**
	c.1498A>G (p.I500V)			-				Testis	
	c.1499T>C (p.I500T)			-				-	
PACS2	c.625G>A (p.E209K)	Developmental and Epileptic Encephalopathy 66 (OMIM: #618067)	GOF	1 (1, 1)	0	0	-	-	-
ADNP	c.2156dup (p.Y719*)	Helsmoortel-Van der Aa syndrome (OMIM: #615873)	Dominant-negative; LOF	-	0	0	32.1 years (65; NR)	-	Van Dijck et al., 2019
	c.2157C>G (p.Y719*)			1 (1, 1)					
	c.2157C>A (p.Y719*)			-					
MAP2K1	c.389A>G (p.Y130C)	Cardiofacialcutaneous syndrome (OMIM: #6152679)	GOF	4 (5, 8)	0	0	-	-	Maher et al., 2018
MECP2[4]	c.397C>T (p.R133C)	Rett syndrome (OMIM: #312750)	LOF	-	26	1	31.3 years (26; 23–42 years) NS vs. German population	Sperm	Trappe et al., 2001 Zhang et al., 2019 Wen et al., 2023
	c.398G>A (p.R133H)			-				-	
	c.502C>T (p.R168*)			-				Sperm	

Gene	cDNA mutation (Amino acid change)	Germline disorder	Likely mechanism[1]	COSMIC v101[2]	Paternal mutations (n)	Maternal mutations (n)	Mean paternal age (n reports; range)	High apparent mutation rate[3]	References
NAA10	c.247C>T (p.R83C)	*NAA10*-related deficiency (OMIM: #300855)	GOF	2 (2, 6)	0	0	-	-	-
PPP1CB	c.146C>G (p.P49R)	Noonan-like syndrome with loose anagen hair 2 (OMIM: #617506)	GOF	-	0	0	33.2 years (6; 24–43 years)	-	Gripp et al., 2016 Zambrano et al., 2016 Zhou et al., 2020
PTPN11	c.922A>G (p.N308D)	Noonan syndrome (OMIM: #163950)	GOF	9 (9, 16)	14	0	35.6 years (14; NR) +6.1 years vs. British population	Testis	Tartaglia et al., 2004 Yoon et al., 2013
	c.923A>G (p.N308S)			4 (4, 16)					
CDK13	c.2525A>G (p.N842S)	Congenital heart defects, dysmorphic facial features, and intellectual development disorder (OMIM: #617360)	Dominant-negative	-	0	0	-	-	-
CSNK2A1	c.593A>G (p.K198R)	Okur-Chung neurodevelopmental syndrome (OMIM: #617062)	GOF	-	0	0	36.5 years (2; 31–42 years)	-	Trinh et al., 2017 Blanc et al., 2024
FGFR3	c.1619A>G (p.N540S)	Hypochondroplasia (OMIM: #146000)	GOF	-	0	0	33.8 years (17; NR) +4 years vs. US population	-	Walker et al., 1971 Maher et al., 2018 Salazar et al., 2022
	c.1620C>G (p.N540K)			-				Testis Sperm	
FOXP1	c.1540C>T (p.R514C)	FOXP1 syndrome (OMIM: #605515)	Dominant-negative	-	1	0	39 years (1; NR)	-	Jay et al., 2019 O'Roak et al., 2011
	c.1541G>A (p.R514H)			-					
CTBP1	c.1024C>T (p.R342W)	Hypotonia, ataxia, developmental delay, and tooth staining (OMIM: #617915)	Dominant-negative	-	0	0	-	-	-
SHOC2	c.4A>G (p.S2G)	Noonan-like syndrome with loose anagen hair 1 (OMIM: #607721)	GOF	1 (1, 1)	0	0	39.1 years (7; 28–52 years)	-	Hoban et al., 2012 Gripp et al., 2013 Zmolikova et al., 2014

Gene	cDNA mutation (Amino acid change)	Germline disorder	Likely mechanism[1]	COSMIC v101[2]	Paternal mutations (n)	Maternal mutations (n)	Mean paternal age (n reports; range)	High apparent mutation rate[3]	References
SRCAP	c.7303C>T (p.R2435*)	Floating-Harbor syndrome (OMIM: #136140)	Dominant-negative	0 (0, 1)	0	0	33.9 years (13; 28–44 years)	-	Hood et al., 2012
ALG13	c.320A>G (p.N107S)	Developmental and Epileptic Encephalopathy 36 (OMIM: #300884)	GOF	-	0	0	29 years (1; NR)	-	Hamici et al., 2017
GNB1	c.239T>C (p.I80T)	Mental Retardation, Autosomal Dominant 42 (OMIM: #616973)	GOF	8 (8, 16)	0	0	50 years (1; NR)	-	Szczaluba et al., 2018
	c.239T>G (p.I80S)			3 (3, 16)					
GRIN2B	c.2450G>C (p.G820A)	DEE27 (OMIM: #613970)	GOF	1 (1, 2)	0	0	-	-	-
	c.2450G>T (p.G820V)			-					

[1] Inferred or proposed mechanism of mutation (gain-of-function, GOF; loss-of-function, LOF) based on available functional evidence.
[2] Counts and from the COSMIC v101 database; counts are grouped as: (n of specific DNA change, n of specific acid substitution, total variants affecting that residue).
[3] Reported mutations limited to variants affecting the same amino acid residue detected in sperm or testes tissue.
[4] DNMs affecting multiple amino acid residues in the same gene are grouped as they result in the same dominant disorder.

NR = not reported; NS = not significant

3.4 – Prioritizing candidate DNMs for selfish behaviour

Recent developments in next-generation sequencing have enabled high-throughput discovery screens in the male germline (Maher et al., 2018). However, there are economic trade-offs between the number of targets that can be evaluated and the depth of sequencing for each target. To optimize resource allocation and direct screening efforts efficiently, an evidence-based scoring framework was developed to prioritize DNMs with the strongest evidence of selfish potential.

Mechanistic evidence was scored from 0 to 2 points based on demonstrated modulation of key signalling pathways involved in SSC homeostasis. Although all known selfish mutations act through a gain-of-function mechanism to activate MAPK or AKT signalling, how other mutations (such as those operating through a loss-of-function mechanism) as well as how other pathways influence SSC fitness remains uncertain. Direct pathway modulation that could feasibly increase SSC fitness warranted 2 points; evidence of interaction without clear modulation received 1 point. Moreover, considering the parallels between selfish selection and oncogenesis, mutations identified at the same residue in the COSMIC v101 database were noted. Although not all established selfish mutations appear in COSMIC (**Table 3.2**), and reported mutations can potentially be passenger mutations, variants encoding any mutant residue were assigned 1 point.

Expression in self-renewing SSCs was scored from 0 to 2 points based on integrated evidence from multiple complementary approaches: (1) detection by scRNAseq in transcriptomic states representing undifferentiated spermatogonia, (2) IHC staining in cells along the basal membrane of seminiferous tubules in the HPA, (3) multiplexed IF validation with established SSC markers for select genes, and (4) supporting evidence from the literature. Full (2) points were awarded for consistent evidence across (≥2) methods, whereas a partial (1) score reflected mixed

or limited evidence. Given technical limitations and the dynamic nature of gene expression, lack of detection by any single method was not considered definitive evidence of non-expression.

Alignment with established PAE features was assessed for three defining characteristics. Supportive evidence, even if not definitive, was considered – for example, elevated paternal age was noted if the mean paternal age across at least five case reports was ≥33 years (based on the lower bound of established PAE disorders; **Table 1.1**). A paternal bias in mutation origin was considered supportive if either the observed or implied ratio of paternal to maternal mutations was greater than 20. A high apparent mutation rate could be established either through disease prevalence estimates or direct detection in either sperm or testis tissue. While absence of evidence was not penalized given limited data availability, definitive evidence against PAE criteria resulted in a (-3) point penalty for each violated feature, reflective of their importance.

The scoring system was calibrated using the three detected selfish mutations as positive controls, which scored between 5–8 total points. MECP2, which scored 3 total points, served as a negative control since its high recurrence rate is driven by germline mosaicism rather than positive selection. Based on this calibration, candidate genes scoring >3 total points were prioritized; 10/17 candidate genes met these criteria and were taken forward (**Figure 3.4**).

Table 3.3: Stratification of candidate and control selfish genes through an integrated scoring system.

Gene[1]	Expression in self-renewing SSCs[2]	SSC signalling pathway regulation[3]	COSMIC v101[4]	Advanced paternal age[5]	Paternal bias in origin of mutation[6]	High apparent mutation rate[7]	Total
PACS1	2	1	1	1	-	-	5
PPP2R5D	2	2	1	-	-	1	6
SMAD4	2	2	1	1	-	1	7
PACS2	1	1	1	-	-	-	3
ADNP	2	2	1	-	-	-	5
MAP2K1	2	2	1	-	-	-	5
MECP2	2	2	-	-3	1	1	3
NAA10	2	1	1	-	-3	-	1
PPP1CB	2	2	-	1	-	-	5
PTPN11	2	2	1	1	1	1	8
CDK13	2	2	-	-	-	-	4
CSNK2A1	2	2	-	-	-	-	4
FGFR3	2	2	-	1	-	1	6
FOXP1	2	1	-	-	-	-	3
CTBP1	2	1	-	-	-	-	3
SHOC2	1	2	1	1	-	-	5
SRCAP	-	1	1	1	-	-	3
ALG13	2	2	-	-	-	-	4
GNB1	1	2	1	-	-	-	4
GRIN2B	-	-	1	-	-	-	1

[1] Gene symbols in bold text indicate prioritized candidate or established PAE genes.
[2] (0 – 2 points) Concordance between ≥2 methodologies, or supportive evidence from the literature, of expression in SSCs.
[3] (0 – 2 points) 1 point for direct evidence of interaction with key signalling pathways involved in SSC homeostasis, and 2 points for modulating such pathways in a manner compatible with positive selection.
[4] (0 – 1 point) Presence of mutation affecting specific residue reported in COSMIC v101.
[5] (0 – 1 point) Mean paternal age ≥32 years across at least five case reports.
[6] (0 – 1 point) Paternal to maternal origin of mutation ($\propto > 20$), observed or implied.
[7] (0 – 1 point) Detection of any individual candidate mutation in either sperm or testis tissue.

Violation of any of the cardinal PAE characteristics resulted in a score of (-3 points).

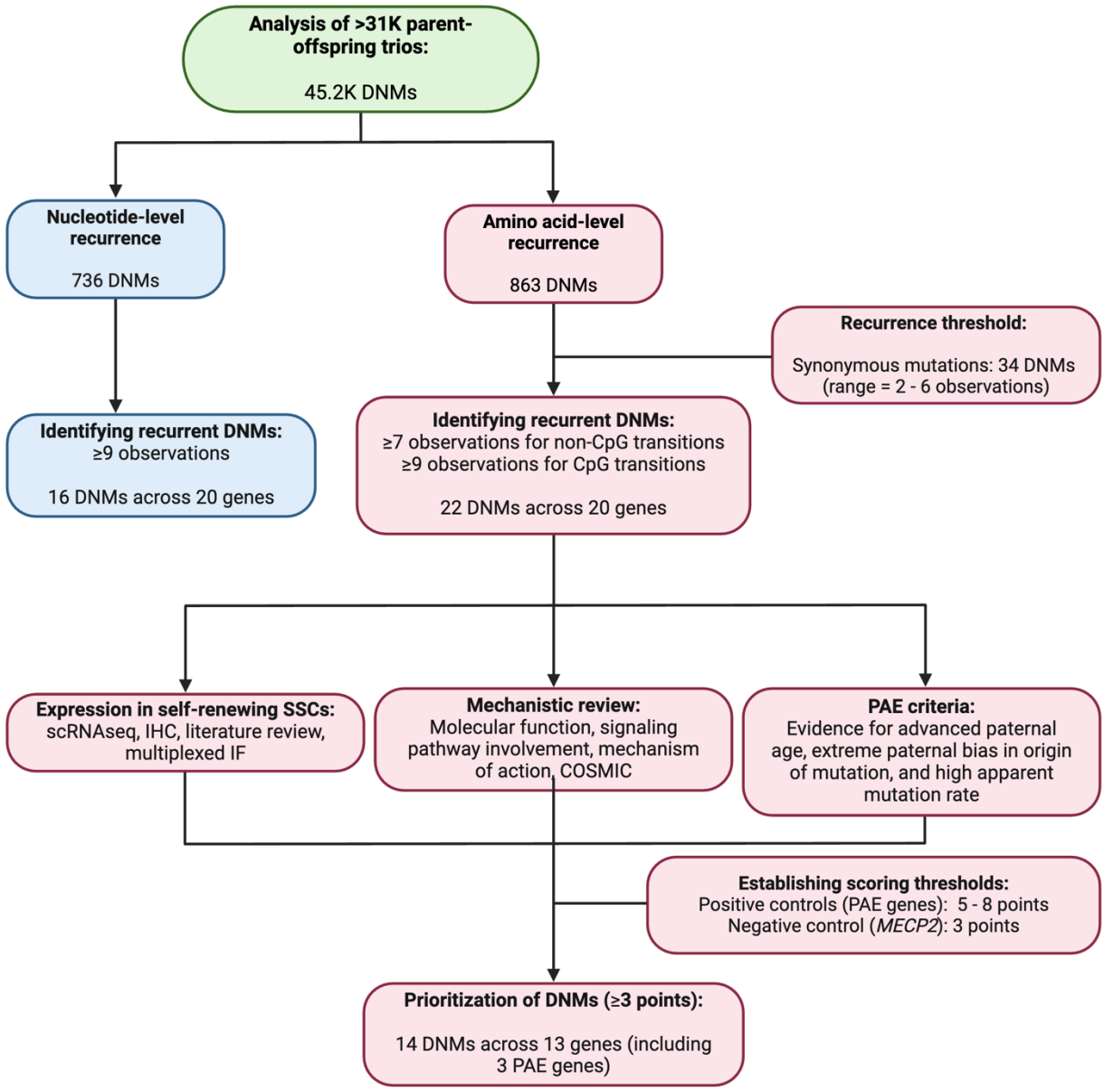


Figure 3.4: Strategy to prioritize candidate selfish genes. Recurrent DNMs from whole-exome sequencing of >31K parent-offspring trios were identified at the nucleotide level (blue) (Kaplanis et al., 2022). In this study, DNMs were analyzed for recurrence at the amino acid level (red), evaluated against broad discriminative criteria, and scored against thresholds established by internal controls to prioritize candidate selfish mutations. Figure created with BioRender.

3.5 – Discussion

In this chapter, a systematic approach was undertaken to identify and prioritize candidate mutations that may undergo positive selection in the male germline. By analyzing DNMs from >31K parent-offspring trios, it was hypothesized that recurrence would reflect underlying germline selection in mutant SSCs. By first filtering for mutations that exceeded the upper bound of synonymous mutations, this approach conservatively sought to only to capture mutations above background. The decision to re-analyze DNMs at the amino acid level, rather than the nucleotide level, facilitated the capture of distinct nucleotide changes that affected the same residue, reflective of operative processes during selection. For example, the Kaplanis study highlighted 16 nucleotide-level mutations observed ≥9 times; however, an amino-acid level approach would have revealed three currently prioritized DNMs (including an established selfish mutation in *FGFR3*) in their initial analysis (**Table 3.1**). In the current analysis, substitutions affecting two genes – one of which was subsequently prioritized – would not have been identified without an amino acid-level approach.

An important aspect of this approach was the integration of numerous lines of evidence to evaluate selfish potential into a comprehensive scoring framework. In addition to recurrence as DNMs, candidate genes were evaluated and scored for their molecular functions, expression in self-renewing SSCs, mechanistic evidence that their mutations could influence SSC homeostasis – such as through modulation of established and associated signalling pathways involved in SSC regulation – and alignment with known PAE characteristics. Calibration against established selfish genes and an internal negative control (*MECP2*) ensured that scoring was not purely theoretical, but also reflective of known selfish biology. Together, this multi-faceted approach allowed for the exclusion of variants with high recurrence but little to no mechanistic plausibility to influence selection.

Although all known selfish mutations operate through a gain-of-function mechanism within the RTK-Ras signalling pathway, the identified candidate mutations have a broader range of putative mutational mechanisms, including LOF and dominant-negative mechanisms. For example, the nonsense variant *ADNP* p.Y719*, the most recurrent mutation in Helsmoortel-Van der Aa syndrome (Helsmoortel et al., 2014), disrupts ADNP's nuclear localization signal and causes mislocalization to the cytoplasm (Cappuyns et al., 2018). In addition to haploinsufficiency, this mutation may sequester ADNP binding partners and act in a dominant-negative manner (**section 3.3.5**), thereby interfering with ADNP's ability to stabilize *β*-catenin and resulting in downregulated Wnt signalling (D'Incal et al., 2024). In mice, Wnt signalling mediates SSC homeostasis and its loss promotes self-renewal over differentiation, thus providing a possible mechanism by which a LOF or dominant-negative mutation can result in positive selection of spermatogonia (Mäkelä and Hobbs, 2019). This phenomenon has been observed in other tissue contexts such as blood, where loss-of-function or dominant-negative mutations in *DNMT3A* – the most recurrently mutated gene in clonal hematopoiesis – result in reduced methylation at enhancer regions that govern self-renewal genes, and subsequently bias hematopoietic stem cell fate to favour self-renewal over differentiation (Challen and Goodell, 2020). Therefore, in order to interrogate candidate mutations in an unbiased manner, mutations were not scored specifically on the basis of activating RTK-Ras signalling, but on their overall mechanistic rationale regardless of the mechanism of mutation.

This chapter also directly characterized the expression of three candidate genes in human spermatogonia using multiplexed IF. This approach not only validated expression patterns inferred from the integrated scRNAseq atlas of human spermatogonia (Bush et al., 2024) and the HPA (Thui and Lindkog, 2018), but also demonstrated *in situ* co-expression of these candidate genes in

spermatogonia (MAGEA4+ cells) and in a subset of undifferentiated spermatogonia (GFRA1+ cells). These results helped overcome discrepancies observed between scRNAseq and IHC, as PPP2R5D expression was not detected by scRNAseq, despite being evident by IHC. This raises the question around the sensitivity of scRNAseq detection thresholds, but also confirms the value of an integrated approach, where multiple methods were leveraged to confirm expression of candidate genes in SSCs.

Mutations affecting 14 amino acid residues across 13 genes met a threshold of evidence to be considered strong candidates for selfish selection. Among them were activating mutations in five genes encoding components of the RTK-Ras pathway (including three established PAE genes), an established mechanism by which selfish mutations influence SSC fitness. Although the prioritization of mutations in this pathway was validating, the set of mutations prioritized was inherently prone to self-reinforcement – mutations affecting RTK-Ras activity were prioritized, in part, because modulation of this pathway was specifically scored. Nevertheless, the decision to score for activity in RTK-Ras pathways, as well as those implicated by scRNAseq, was a necessary limitation given our current understanding of germline selection (Goriely and Wilkie, 2012; Guo et al., 2017).

There were several other limitations to this prioritization framework that should be noted. The focus on severely affected clinical populations inherently selects for highly pathogenic mutations that may promote clonal expansion most strongly. However, mutations resulting in more subtle clinical phenotypes, or even positive traits such as increased cognition, would naturally be excluded from this analysis (Goriely and Wilkie, 2012; Goriely et al., 2013). The scoring system developed here also relied on thresholds (e.g. ≥33 years for advanced paternal age) that were empirically established, chosen based on prior knowledge and practical constraints. For example,

although transitions occurring at CpGs are ~10x more common than at non-CpG dinucleotides, it was not feasible to only prioritize CpG transitions observed ≥70 times (relative to the minimum threshold). This framework also awarded points for the presence of a somatic mutation affecting the same reside in COSMIC, as potential evidence of a selective advantage. However, the biological impact of this was unclear as only two of three prioritized selfish mutations were reported in COSMIC.

Despite these limitations, this approach provides a comprehensive and reproducible framework for identifying candidate selfish mutations in future genomic data sets. The value of this approach for defining an experimental target space will be explored in the subsequent chapters.

Chapter 4 – Discovery of novel selfish mutations in the aged human testes

4.1 – Introduction

De novo mutations in 10 prioritized candidate selfish genes were hypothesized to undergo positive selection in the male germline (**section 3**). The principal goal of this chapter was to perform a discovery screen in aged human testes to determine whether these candidate selfish mutations are identifiable at elevated levels. Previous studies have demonstrated that selfish mutations are detectable in bulk sperm at variant allele frequencies (VAFs) ranging from 10^{-6} to 10^{-3} with mutation frequencies increasing with the age of the donor (**section 1.7**). Although existing methods to screen sperm are highly sensitive, they are limited by the amount of genomic space that can be queried (1–6 base pairs).

To overcome the technical burden of screening for ultra-rare mutations in bulk sperm, several strategies have been developed (**section 1.9**). One strategy involves the systematic dissection of testicular tissue into discrete biopsies, each containing cross-sections of multiple seminiferous tubules (Qin et al., 2007; Choi et al., 2008; Choi et al., 2012; Shinde et al., 2013; Yoon et al., 2013; Eboreime et al., 2016; Maher et al., 2018; Eboreime et al., 2022). The pairing of testicular dissection with massively parallel multiplex PCR and next-generation sequencing has demonstrated that mutational foci are detectable between ~0.0005–0.03 VAF in individual biopsies (Maher et al., 2018). Despite these advances, the reliable detection of low-frequency variants remains challenging by the inherent error rate of NGS (~0.001–0.01 VAF) (Kennedy et al., 2014).

To improve sensitivity for rare variant detection and enable a robust screen of candidate selfish mutations, rhAmpSeq was adapted to facilitate multiplex targeting of candidate and control loci. A major challenge for multiplexed PCR assays is primer-dimer formation, which compete for

PCR resources and diminish the capacity to amplify rare alleles (Yang et al., 2020). RNase H2-dependent PCR (rhPCR) exploits the unique ability of RNase H2 to recognize and cleave a single nucleotide from RNA:DNA hybrid substrates (Cerritelli and Crouch, 2009) by employing specialized PCR primers with an RNA base and blocking moiety appended to the 3' end of the primer (Dobosy et al., 2021). Perfect heteroduplex formation is required for RNase H2-mediated cleavage of the duplex to initiate amplification (**Figure 4.1**), which has been shown to eliminate primer dimers and increase the sensitivity of allele-specific PCR by 45-fold (Dobosy et al., 2021).

Prior efforts had been undertaken to incorporate unique molecular identifiers (UMIs) – short, random oligonucleotide sequences used to enable bioinformatic correction of errors introduced during library preparation or sequencing (Smith et al., 2017; Schmitt et al., 2012) – into rhPCR primers to further improve the sensitivity of the assay (Drs. Stephen Bush and Katherine Wood) (**Figure 4.1**). UMIs were deliberately designed to be 5 bp in length (2 x NNNNN) and degenerate to the 5' sequence upstream of the primer, as complementarity could bias primer binding and influence amplification and quantification of rare variants (Schmitt et al., 2012). Implementing this degenerate design reduced the theoretical UMI complexity from 4^{10} (~1.05 million unique sequences) to 3^{10} (~59,000 sequences).

Using previously generated rhAmpSeq data, cycling conditions optimized for rare variant detection were determined. A custom sequencing panel was then designed to target both prioritized candidate loci and mutational hotspots in established selfish genes was developed, and the assay was assessed through a pilot study. Finally, 186 testis biopsies from three aged men (average age 70.7 years; range 63–76 years) were screened to assess whether novel candidate mutations exhibit selfish behaviour in the male germline.

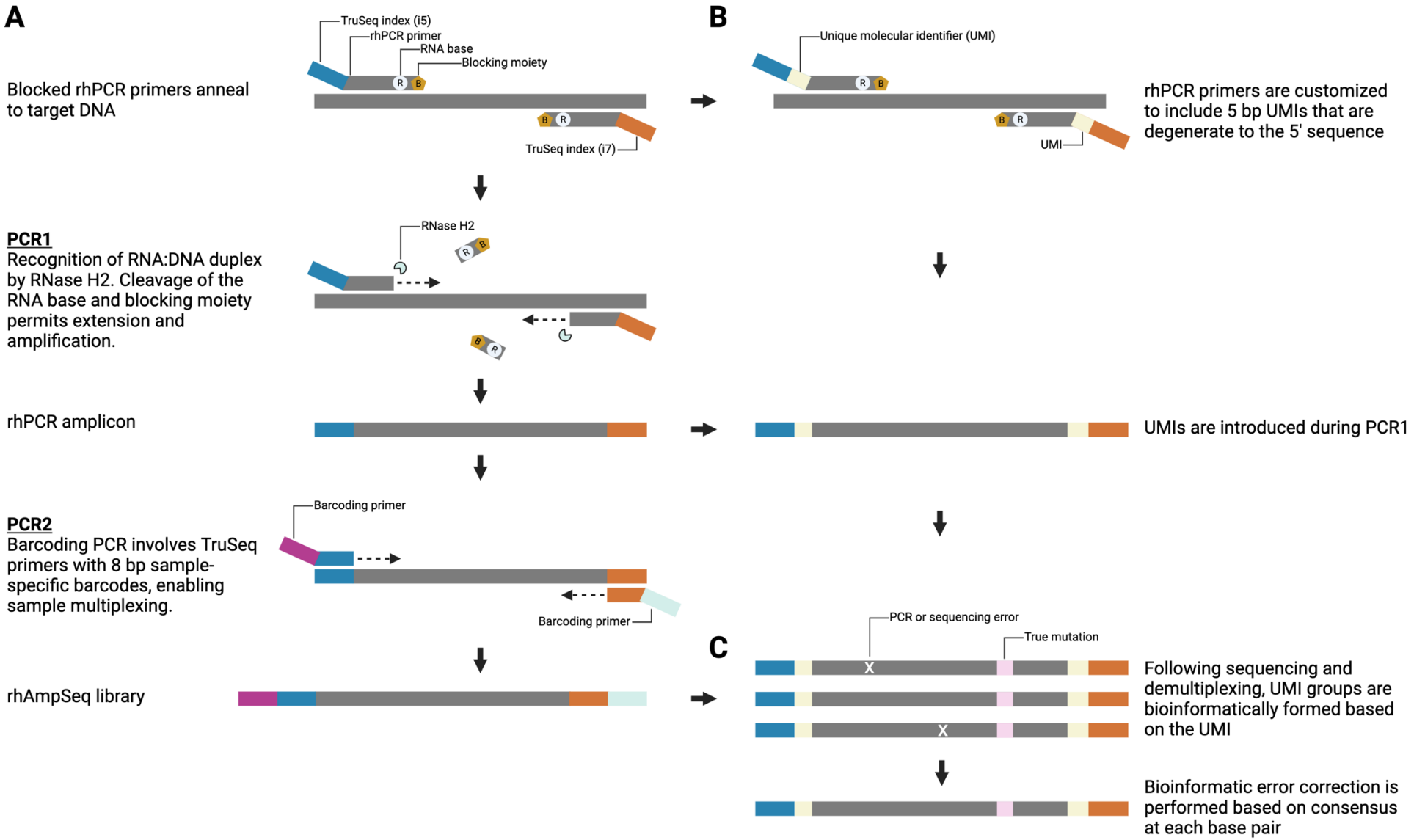


Figure 4.1: rhAmpSeq strategy to develop multiplexed assays for variant detection in testes biopsies. (A) Schematic of the rhAmpSeq workflow in which blocked rhPCR primers are activated following perfect hybridization and cleavage by RNase H2. This permits primer extension and the initiation of PCR1 (dotted arrows). rhPCR amplicons are then barcoded during PCR2, which permits sample multiplexing. **(B)** Customization of rhPCR primers to incorporate UMIs (2 x NNNNN; degenerate to the 5' sequence of the rhPCR primer) during PCR1. The library preparation workflow is otherwise the same. **(C)** Following sequencing and sample demultiplexing, reads are bioinformatically grouped into UMI groups which allows for an error-corrected consensus sequence to be generated. Figure created with BioRender. Adapted from IDT.

4.2 –Proof-of-concept: determining optimal rhAmpSeq cycling conditions

Because UMIs are integrated into the rhPCR primers used in PCR1, they are introduced at the initial amplification stage (**Figure 4.1**). PCR1 must balance a sufficient number of cycles to produce adequate template material for library preparation with the risk of UMI overwriting, where daughter strands originating from the same parental molecule are tagged with different UMIs. UMI overwriting can result in the over-deduplication of reads and inaccurately represent the original molecule abundance (Saunders K et al., 2020).

Patient-derived DNA from 9 probands had been serially diluted 2-fold into control blood DNA, resulting in mutation levels of 0.000625–0.005 VAF (corresponding to a range of 1 in 1600 to 1 in 200) (Dr. Katherine Bush) (**Figure 4.2**). Using a sequencing panel of 9 amplicons (1984 bp), 20 total libraries were each prepared from 100 ng of DNA (~30,000 haploid genomes) using four distinct rhPCR cycling protocols that had been empirically determined (8-18X, 8-14X, 14-18X, 14-22X; which correspond to the number of PCR1 and PCR 2 cycles, **Figure 4.1**).

Libraries were sequenced on an Illumina MiSeq, and sequencing data was processed using Amplimap, a custom Snakemake workflow to automate alignment to the human genome (hg38) and production of sequencing pileups (Dr. Stephen Bush) (**section 2.4.6**) (Koelling et al., 2019). UMIs were extracted, grouped, and consensus sequences were formed using UMI-tools (Smith et al., 2017).

rhAmpSeq proof-of-concept panel (9 amplicons, 7 genes)

rhPCR amplicon	Patient-derived DNA
FGFR2	c.755C>T (p.S252W)
FGFR3	c.1948A>G (p.K650E)
HRAS	-
PACS1	-
PPP2R5D	-
PTPN11.1	c.181G>A (p.D61N); c.215C>G (p.A72G); c.228G>C (p.E139D)
PTPN11.2	c.416G>C (p.E139D)
PTPN11.3	c.1520A>G (p.M504V)
SMAD4	c.1498A>G (p.I500V); c.1499T>C (p.I500T)

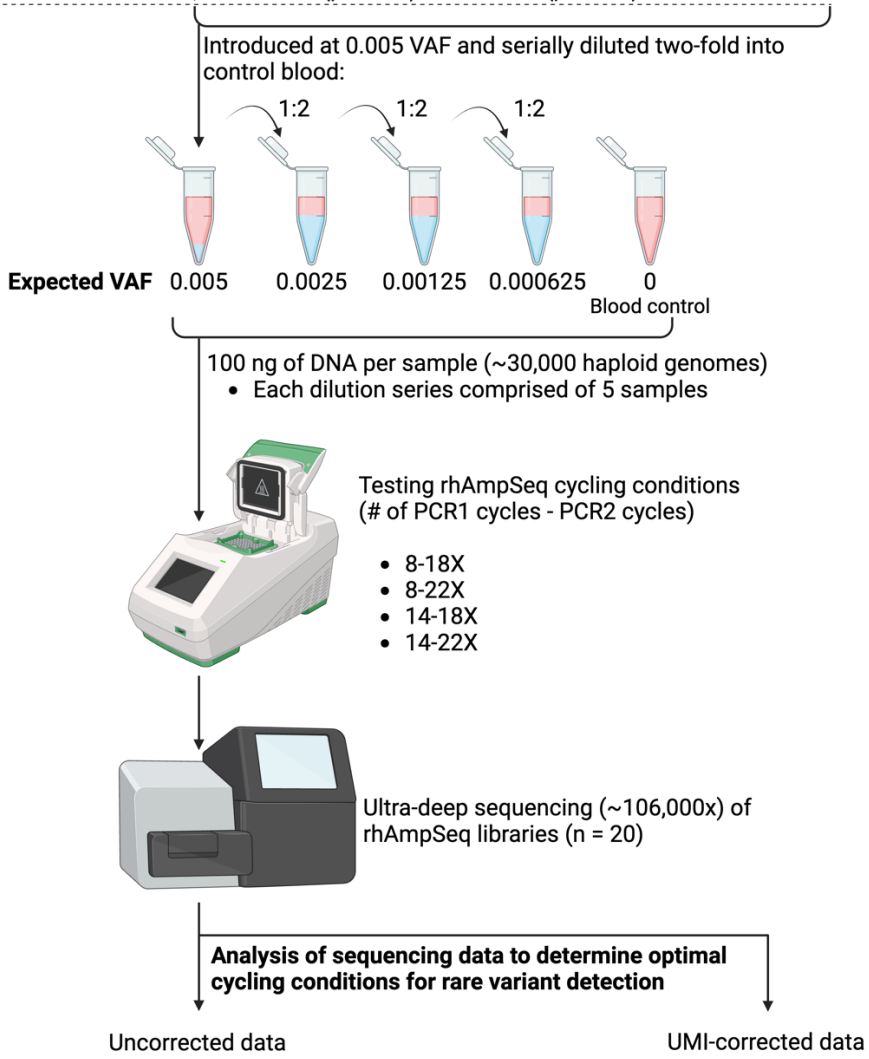


Figure 4.2: Overview of rhAmpSeq proof-of-concept experiment to determine optimal cycling conditions for rare variant detection. Patient-derived DNA from 8 probands were serially diluted 2-fold into control blood DNA between 0.000625–0.500 VAF. Each serial dilution (n = 5 samples) was prepared using four different cycling conditions and sequenced on an Illumina MiSeq. In this work (bolded), sequencing data was analyzed to characterize the sensitivity of rhAmpSeq and to understand the effect of cycling conditions on UMI formation and error rates. Figure created with BioRender.

4.2.1 – Sequencing coverage and UMI distribution

Following primer trimming, 1516 unique positions were retained and the median sequencing coverage across all 20 samples was 87,321x, which increased with the total number of PCR cycles used (**Figure 4.3A**). Across the 9 amplicons, there was an ~8-fold difference in the median coverage between the best and worst covered regions, with 8 of 9 spiked-in mutations occurring in amplicons with median coverage of 116,195x (**Figure 4.3B**).

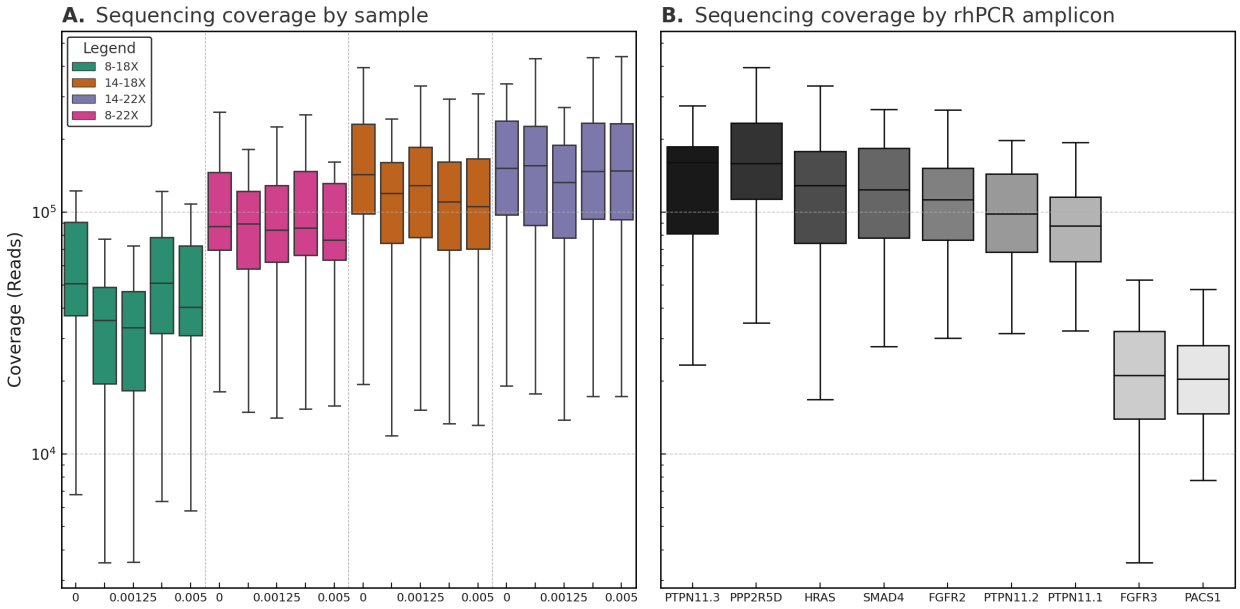
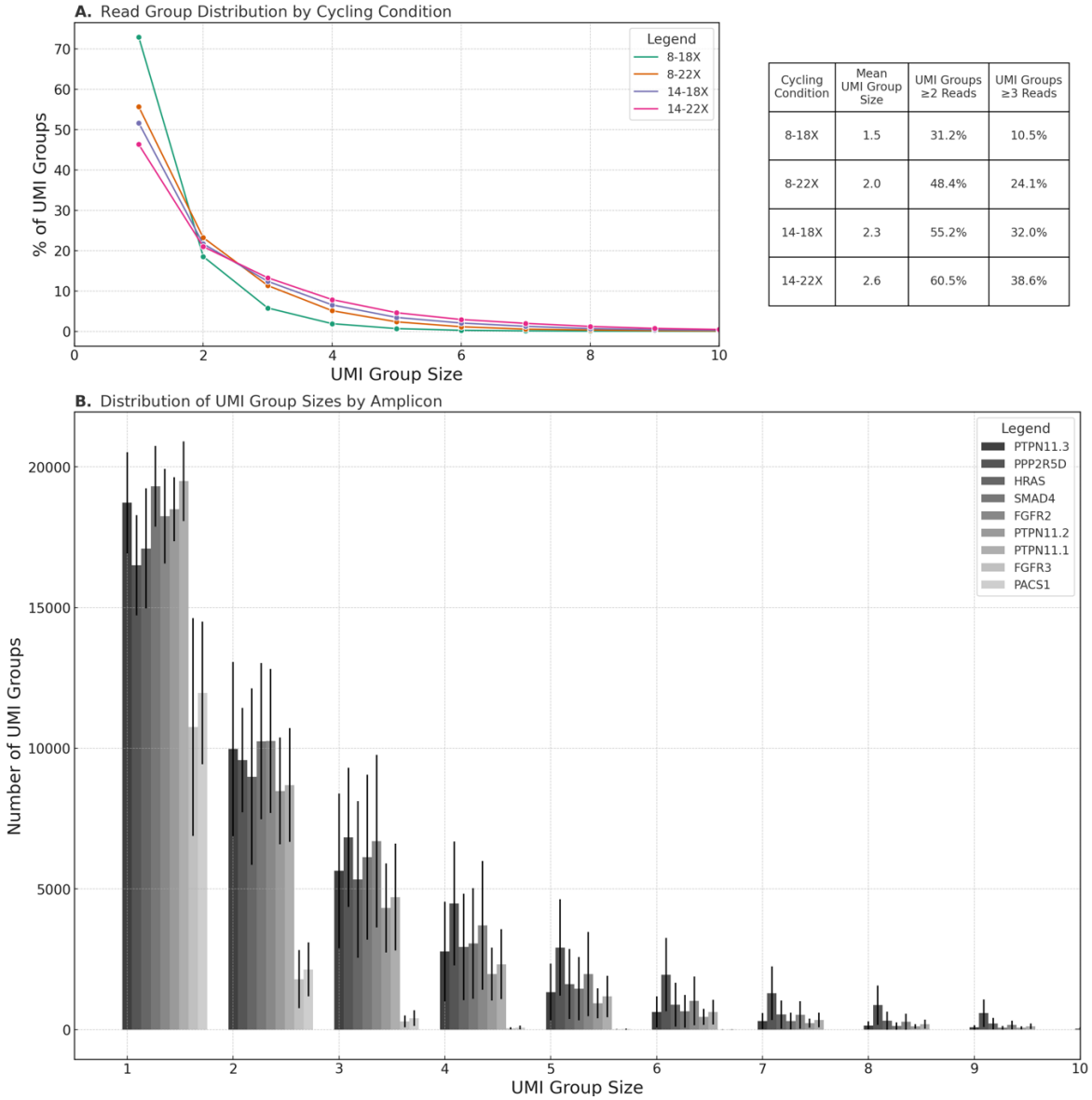


Figure 4.3: Sequencing coverage across proof-of-concept experiment. **(A)** Sequencing coverage organized by cycling condition. Within each group, samples are organized by their dilution level. **(B)** Sequencing coverage by amplicon, organized by median coverage.

Although the mean UMI group size increased with the number of PCR cycles ($p < 0.001$ for all comparisons, Wilcoxon rank-sum test), these differences were numerically small; under all conditions, the mean UMI group size was <3 (**Figure 4.4A**). Singletons (UMI groups with only one read) were the most common read group across all cycling protocols. This presented a technical challenge for UMI-based error correction, which typically requires ≥3 reads per UMI group to generate a reliable consensus sequence, with larger UMI groups showing a further

reduction in error rates (Schmitt et al., 2012). However, requiring ≥3 reads per UMI group in this dataset would discard the majority of reads (~61–90%) (**Figure 4.4A**), which would severely reduce the effective coverage and poses a challenge for rare variant detection. Therefore, to balance the benefits of error correction with the need to retain enough data for rare variant calling, a threshold of ≥2 reads was chosen to preserve data yield (**section 2.2**).



Cycling Condition	Mean UMI Group Size	UMI Groups ≥2 Reads	UMI Groups ≥3 Reads
8-18X	1.5	31.2%	10.5%
8-22X	2.0	48.4%	24.1%
14-18X	2.3	55.2%	32.0%
14-22X	2.6	60.5%	38.6%

Figure 4.4: UMI group distribution and size. (A) Effect of rhPCR cycling conditions on the distribution of UMI group size. **(B)** Distribution of UMI groups by number of reads per group across amplicons.

4.2.2 – Analysis of serial dilutions to determine rhAmpSeq sensitivity

To assess the sensitivity of rhAmpSeq with and without UMI correction, analysis of serially diluted DNA was undertaken. Across the four cycling conditions, the mean observed VAFs were within ~0.9-fold of expectations (range 0.2–2.5) (**Figure 4.5**). Following UMI correction, effective coverage decreased ~76.7% (range 74.3–85.9%) on average for each queried position (**Figure 4.5**), and observed VAFs were, on average, within 0.4-fold of expectations (range 0–2.0).

To assess sensitivity, each dilution (averaged across all replicates) was compared against the blood control to determine the lowest VAF distinguishable from background. For uncorrected reads, spiked-in mutations were distinguishable from blood in 5/9 and 8/9 dilutions at 0.000625 and 0.00125 VAF, respectively (**Figure 4.5**). Following UMI-correction, mutations at 0.000625 and 0.00125 VAF were discernable in 3/9 and 5/9 dilutions, respectively.

Detection limits differed by substitution type. For transitions, spiked-in mutations were detectable at 0.000625 VAF in 1/5 uncorrected dilutions and 0/5 UMI-corrected dilutions, respectively. By contrast, spiked-in transversions were detectable at 0.000625 VAF in 4/4 uncorrected and 3/4 UMI-corrected dilutions.

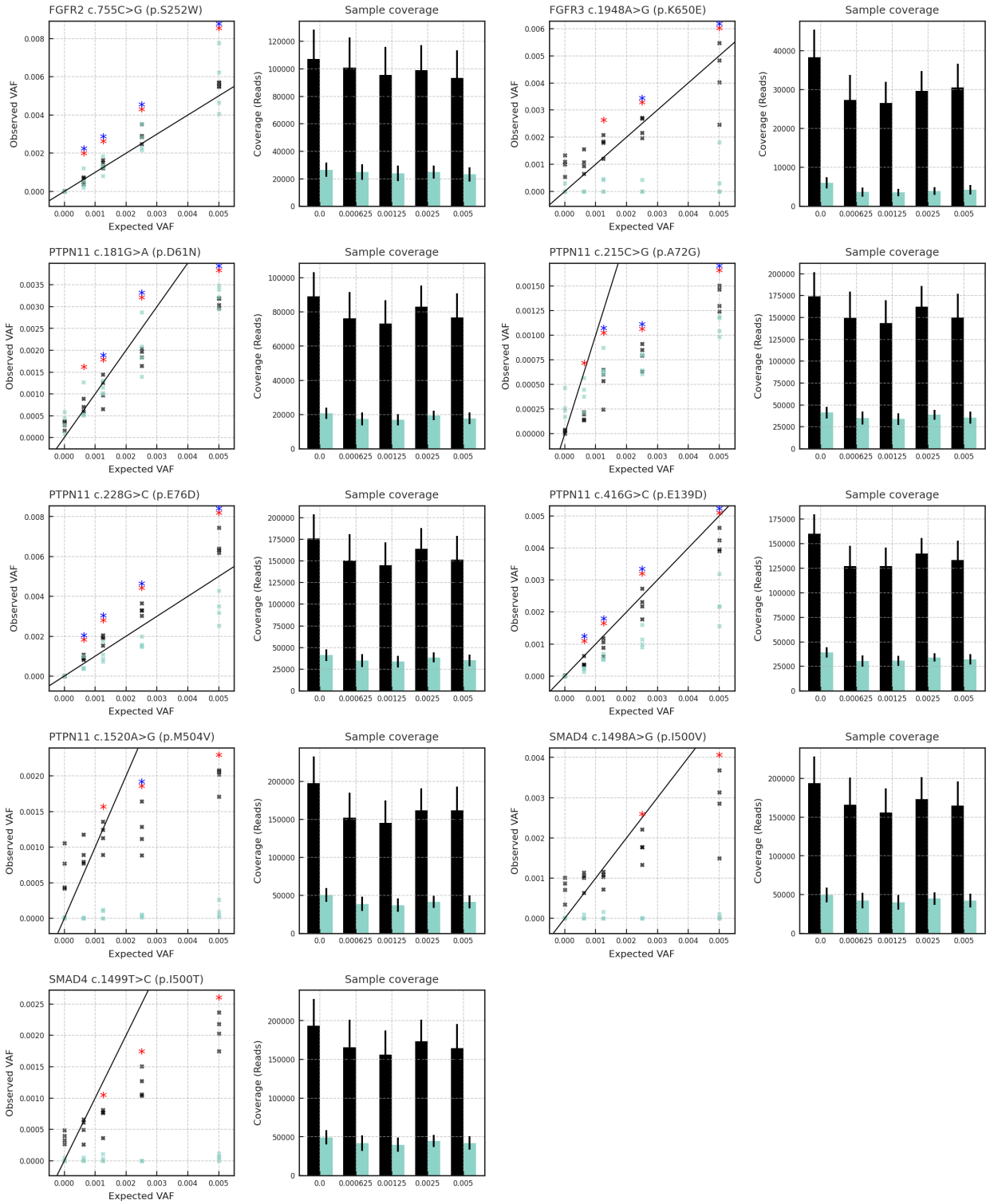


Figure 4.5: Serial dilution of DNA from 9 probands. Observed mutation levels for 4 samples prepared under different PCR cycling conditions (8-18X, 8-22X, 14-18X, 14-22X) are individually plotted for each dilution (black objects, uncorrected; teal objects; UMI-corrected). The solid black line serves as a reference to compare observed data with expected values. For each dilution, the mean VAF of 4 samples is compared against the blood-only control. Asterisks indicate that the

dilution level is statistically distinguishable ($p < 0.05$; Wilcoxon-rank sum test) from control (red, uncorrected; blue, UMI-corrected). Each scatter plot is accompanied by a bar plot indicating the mean (± SD) coverage for each dilution across 4 samples at the given position.

4.2.3 – Characterization of error rates across substitution type and cycling conditions

An analysis of error rates across cycling conditions, before and after UMI correction, was undertaken. Error rates increased with the total number of PCR cycles ($p < 0.01$ for all comparisons; Wilcoxon rank-sum test) (**Figure 4.6A**). Cycles sharing the same number of PCR1 cycles (e.g. 8-22X vs. 8-18X) differed by only ~1.03-fold whereas conditions differing in PCR1 cycles (e.g. 8-18X vs. 14-18X) showed a ~1.32-fold difference. When the error rates for each of the 9 amplicons were compared pairwise across cycling conditions (54 total comparisons), two patterns emerged: first, conditions sharing the same PCR1 cycle count were largely similar, differing in only 3 of 18 comparisons ($p < 0.05$; Wilcoxon signed-rank test) and second, conditions that differed in PCR1 cycles were significantly different across all 36 comparisons ($p < 10^{-10}$; Wilcoxon rank-sum test) (**Figure 4.6B**). Error rates also differed by substitution type, with transition rates ~25-fold higher than transversion rates (1.9×10^{-3} vs. 7.7×10^{-5}, respectively), explaining the observed difference in sensitivity between substitution types (**Figure 4.7A**).

UMI-based error correction significantly decreased the mean error rates for all cycling conditions ($p < 10^{-308}$; Wilcoxon rank-sum test) by ~0.82-fold (range 0.76–0.90) (**Figure 4.7B**). Although UMI-corrected error rates were numerically similar across all cycling conditions (range 0.00167–0.00174 VAF), the 8-22X cycling conditions produced the lowest UMI-corrected error rate ($p < 0.05$; Wilcoxon rank-sum test).

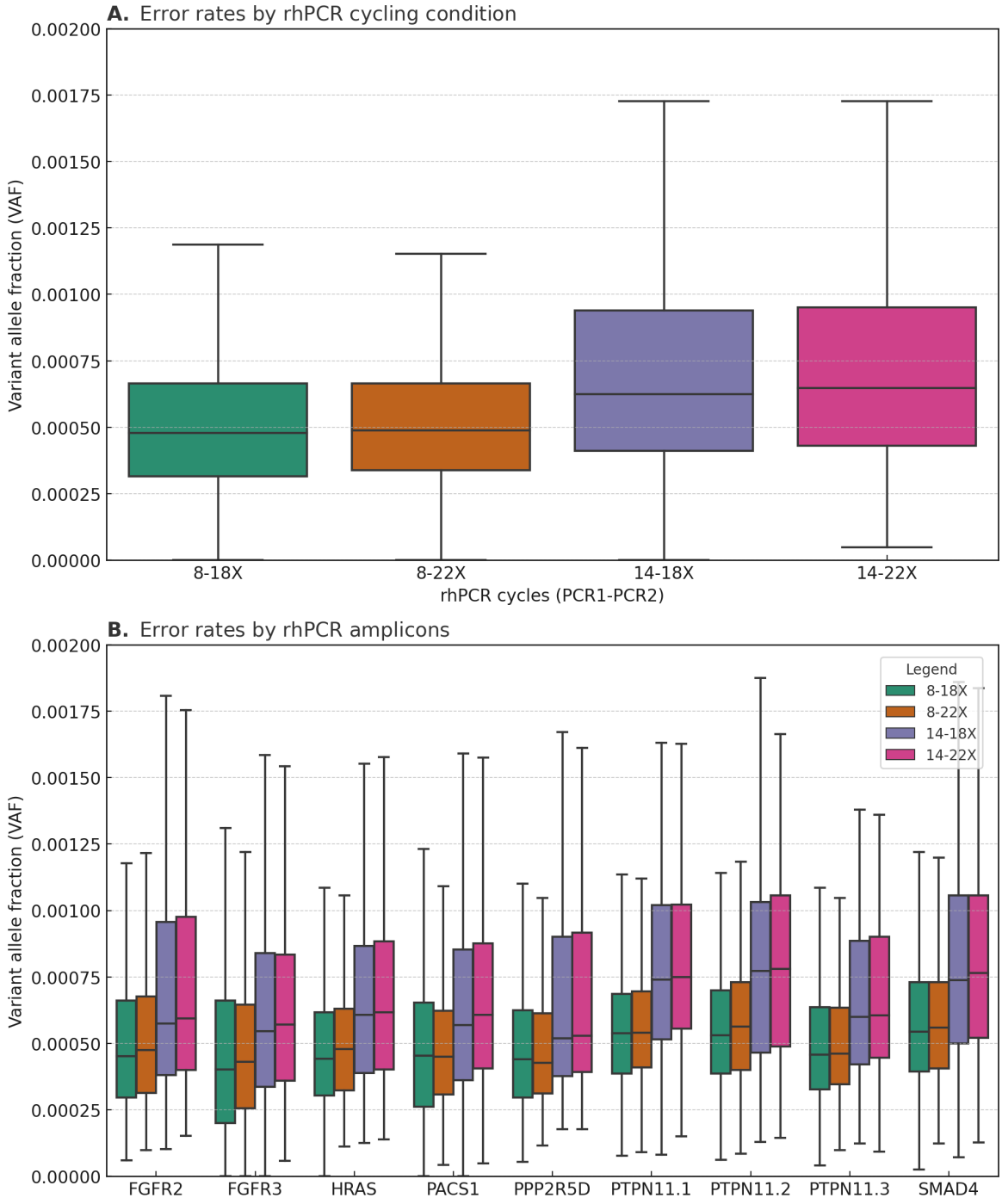


Figure 4.6: rhAmpSeq error rates by cycling conditions and across amplicons. (A) Uncorrected error rates across all positions organized by the four different rhPCR cycling conditions (PCR1-PCR2). Outliers are not visualized. **(B)** Uncorrected error rates by rhPCR amplicon.

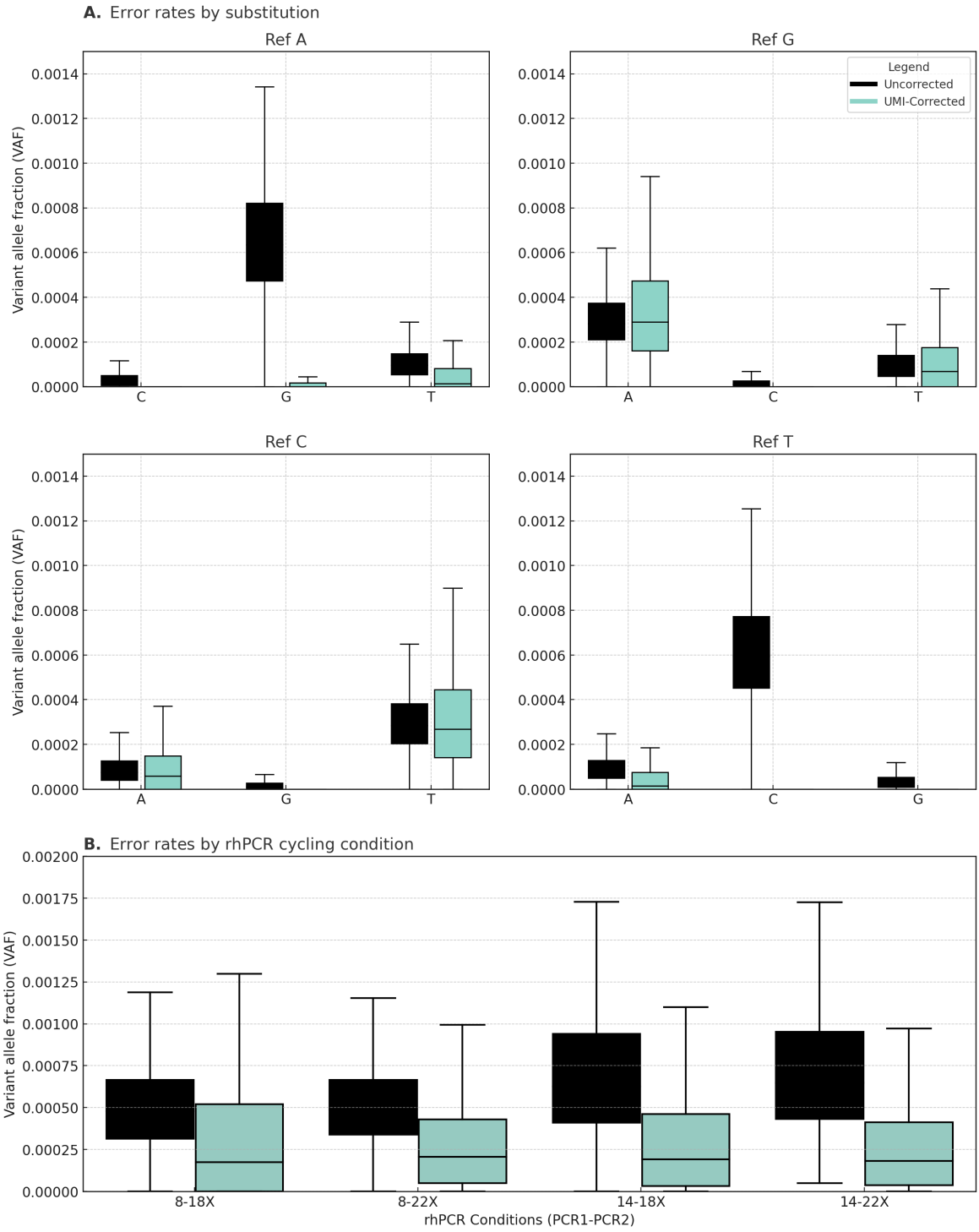


Figure 4.7: rhAmpSeq error rates with and without UMI correction. (A) Error rates by substitution. Each panel shows box plots of VAF for a specific reference base (black, uncorrected; teal; UMI-corrected). Boxes span IQR with the median indicated by a horizontal line. Outliers are not visualized. **(B)** Aggregated error rates across all substitutions by rhPCR cycling condition.

4.2.4 – Determination of rhPCR cycling conditions

Although UMI correction reduced raw error rates, no improvement in the overall sensitivity for spiked-in mutations was observed. In several cases, and in spite of high effective coverage, UMI correction led to a loss of signal. A combination of fewer PCR1 cycles (8X) and more PCR2 cycles (22X) was selected to balance minimizing the introduction of errors and maximizing potential UMI group size. This compromise preserved the option for UMI-based error correction, without fully depending on it.

4.3 – Investigation of selfish candidate mutations in aged human testes

4.3.1 – Experimental design and strategy to call low frequency variants

A custom sequencing panel was designed to include 20 rhPCR primer sets with UMIs. This panel covered ~4.1 kb of genomic space and consisted of 11 rhPCR primer sets targeting mutations across 10 candidate selfish genes (51.4% of total coverage), as well as 9 rhPCR primer sets targeting known mutational loci across 6 established selfish genes to serve as positive controls (**Table 4.1**). Because *PPP2R5D* mutations associated with Jordan syndrome were deemed to be a high priority (**Table 3.3**), an additional amplicon covering the p.E420 locus was included to achieve ~99% coverage of known Jordan syndrome mutations (Levine and Chung, 2023).

Testes tissue from three aged individuals (average 70.6 years; range 63–76 years) was systematically dissected into 229 small biopsies (~20–100 mm^3 each), of which 215 were included in this work (**section 2.4.1**). Because selfish selection is an age-related process with mutational frequencies increasing with donor age, tissues from aged men were screened to maximize the likelihood of discovery. Whereas prior studies typically dissected testis slices into biopsies sized ~60–180 mm^3 (Maher et al., 2018; Eboreime et al., 2022), dissecting smaller pieces in this study afforded a ~1.6–2.8-fold increase in resolution for detecting rare mutations. Moreover, all samples

were screened in technical triplicates to enhance reliability, since true mutations should be reproducibly observed, similar to the concept behind UMIs.

Table 4.1: rhAmpSeq selfish candidate panel. List of amplicons targeting prioritized candidate loci and known mutational hotspots. If a serially diluted mutation was introduced, the associated genomic position and expected range of VAFs is indicated.

Amplicon	Select loci of interest	Patient-derived DNA	Genomic position (hg38)	Expected range of VAFs
ADNP	p.Y719	-	-	-
ALG13	p.N107	-	-	-
CDK13	p.N842	-	-	-
CSNK2A1	p.K198	-	-	-
FGFR2	p.S252; p.P253	c.755C>G (p.S252W)	chr10: 121520163	0.000625–0.01
		c.755C>T (p.S252L)	chr10: 121520163	0.00375–0.06
		c.755_756CG>TC (p.S252F)	chr10: 121520162-121520163	0.00125–0.02
		c.755_757CGC>TCT (p.S252F; p.P253S)	chr10: 121520161-121520163	0.000625–0.01
		c.758C>G (p.P253R)	chr10: 121520160	0.000625–0.01
FGFR3.1	p.G380	-	-	-
FGFR3.2	p.N540	-	-	-
FGFR3.3	p.K650	c.1948A>C (p.K650Q)	chr4: 1806162	0.000625–0.01
		c.14948A>G (p.K650E)	chr4: 1806162	0.000625–0.01
		c.1949A>T (p.K650M)	chr4: 1806163	0.000625–0.01
GNB1	p.I80	-	-	-
HRAS	p.Q61	c.182A>G (p.Q61R)	chr11: 533874	0.000625–0.01
		c.184G>T (p.Q61H)	chr11: 533873	0.000625–0.01
PACS1	p.R203	-	-	-
PPP1CB	p.P49	-	-	-
PPP2R5D.1	p.E198	-	-	-
PPP2R5D.2	p.E420	-	-	-
PTPN11.1	p.D61; p.A72; p.E76	-	-	-
PTPN11.2	p.N308	-	-	-
RET.1	p.L790	-	-	-
RET.2	p.M918	c.2753T>C (p.M918T)	chr10: 43121968	0.000625–0.01
SHOC2	p.S2	-	-	-
SMAD4	p.R496; p.I500	c.1498A>G (p.I500V)	chr18: 51078306	0.0000625–0.001
		c.1499T>C (p.I500T)	chr18: 51078307	0.0000625–0.001

A custom statistical model was developed to detect low-frequency variants by identifying non-reference outliers at each genomic position (**section 2.7.3**). Briefly, for each genomic position, the background error rate is estimated for each base substitution through a binomial framework. As there are upwards of 7.5 million total hypotheses that could be tested (~4.1 kb across 215

biopsies in triplicate for all 3 non-reference bases), multiple comparisons are accounted for by applying a Bonferroni correction. Detection of low-frequency variants also requires sufficient coverage depth to ensure that true variants can be distinguished from noise. Based on binomial statistics, a minimum internal target of 5000x median coverage per position was set to reliably call mutations at 0.001 (1 in 1000) with a minimum of 5 supporting reads at ~90% confidence (Spencer et al., 2014). Variants were only called if identified in at least two independent replicates from the same testis biopsy.

As calling low frequency mutations is prone to artefacts arising during library preparation or sequencing, each experiment also included a series of serial dilutions from 13 probands across 5 amplicons at initial VAFs ranging from 0.001 to 0.06 (**Table 4.1**). Although most mutations (9 of 13) were introduced at an expected starting VAF of 0.01, a wider range of starting VAFs was explored depending on the availability of genomic DNA (0.001) and to characterize a broader dynamic range (0.02–0.06) than in the proof-of-concept experiment. These diluted samples were prepared and sequenced under identical conditions as the testis biopsies, thereby allowing for an empirical estimate of the limit of detection (LOD) for each run. The LOD was defined as the lowest VAF at which the majority of known spiked mutations were consistently detected across replicates.

Following statistical prioritization, variants that were prioritized in all three replicates were classified as Tier 1, whereas those detected in exactly two replicates were categorized as Tier 2. Based on a dynamic limit of detection, variants that were detected below the LOD, were moved to a lower confidence tier such that any Tier 2 variant detected under the LOD would be deprioritized, representing a higher likelihood of amplification bias or sequencing artefact.

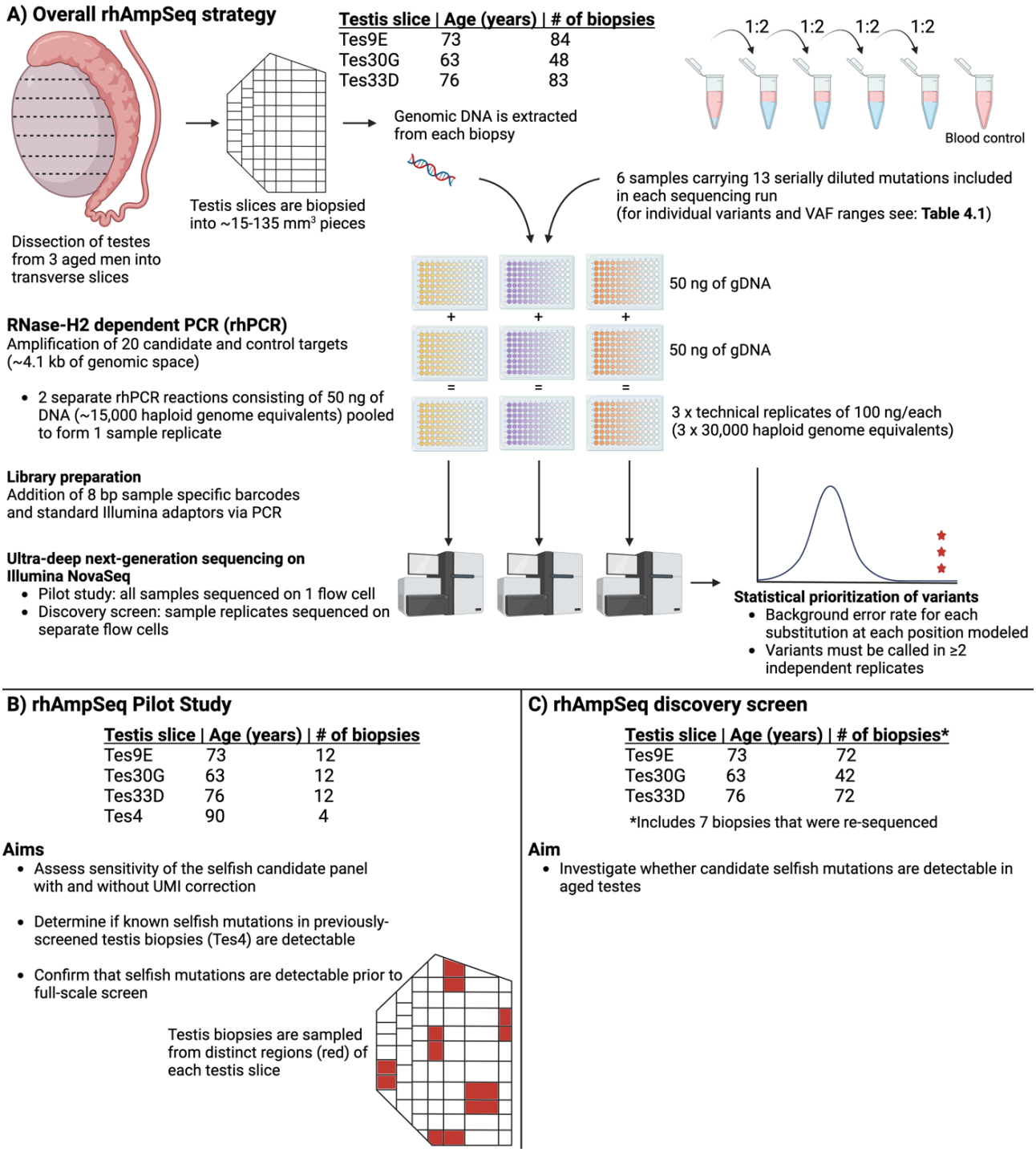


Figure 4.8: Overview of rhAmpSeq strategy to investigate candidate selfish mutations in aged testis biopsies. (A) Overall strategy to dissect and screen testes biopsies alongside serially diluted DNA in technical triplicates (3 x 100 ng). After ultra-deep sequencing, a custom statistical approach is applied to identify variants that are consistently detected as outliers in at least two independent replicates. Summary and scope of the **(B)** pilot study and **(C)** full discovery screen.

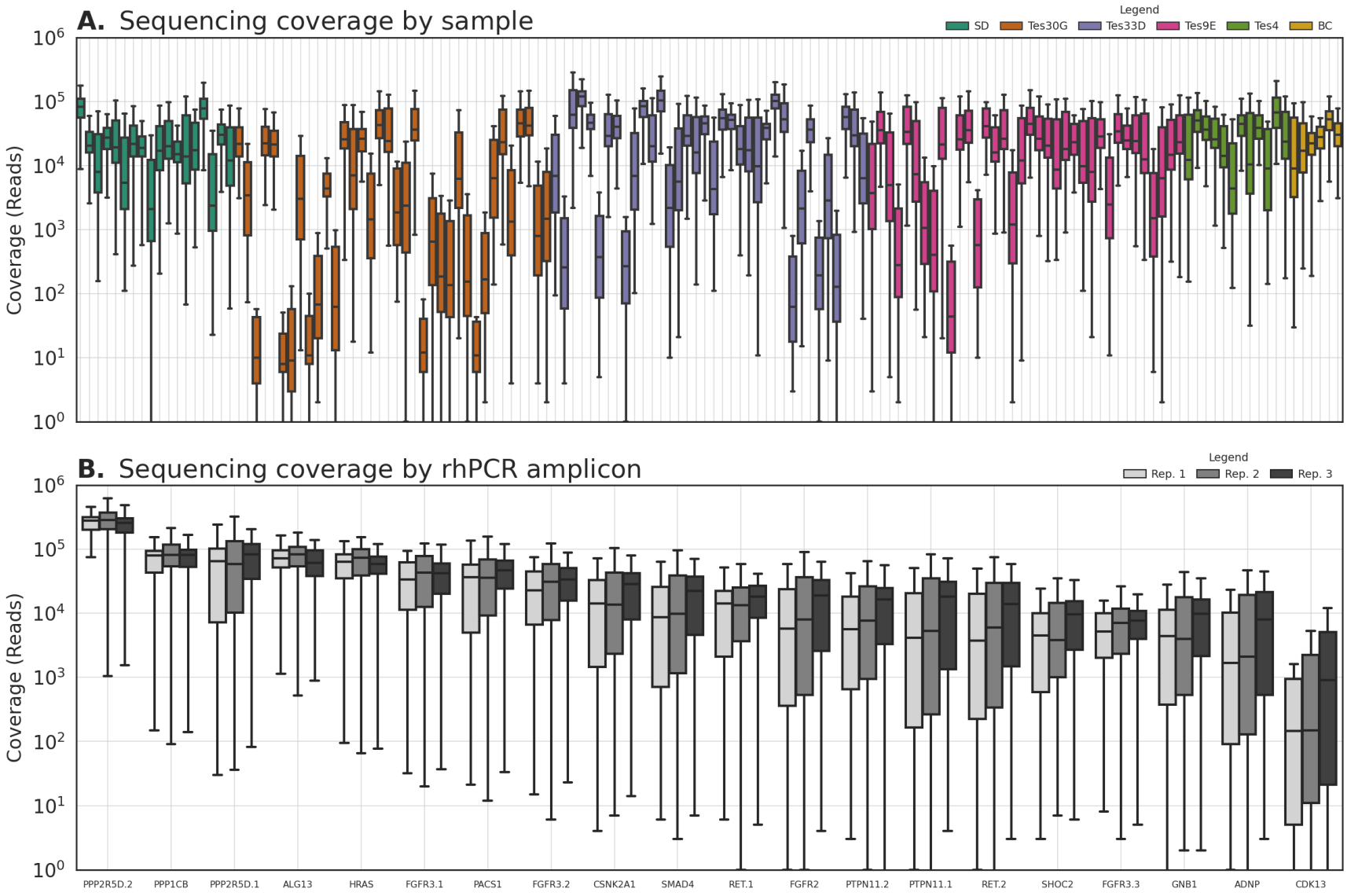
4.3.2 – rhAmpSeq Pilot study

To evaluate panel performance, a pilot study was conducted with 36 of 215 testis biopsies (16.7%) and 2 blood controls, 6 serial dilution samples, as well as 4 biopsies harbouring known selfish mutations identified in a prior discovery screen (Maher et al., 2018). That study reported no detectable selfish mutations in an entire slice of an 87-year-old donor, potentially due to age-related hypospermatogenesis. Therefore, this pilot study also sought to confirm the suitability of the testes tissue and verify that known selfish mutations could be detected before proceeding with a full discovery screen. As selfish mutations are expected to exhibit localized clonal expansions in the testes of aged men, these 36 biopsies were deliberately chosen to sample distinct geographical regions of each slice to increase the likelihood of detection (**Figure 4.8**).

Using the cycling conditions determined in **section 4.2**, rhAmpSeq libraries were prepared for all 48 samples in technical triplicates, which were sequenced on an Illumina NovaSeq. Sequencing data was processed using Amplimap, and UMIs were processed using UMI-tools (Dr. Stephen Bush) (**sections 2.7.2 and 2.7.3**).

4.3.2.1 – Sequencing coverage and UMI analysis

Following primer trimming, the original 4083 bp target space was reduced to 3089 unique callable bases. 24 of 144 samples (16.7%) did not meet a median coverage threshold of <1000x used to define sample failure and were therefore excluded from further analysis. The median coverage across 120 samples (83.3%) passing filtering was 20,611x (**Figure 4.9A**). Coverage varied across amplicons with approximately 4.5 billion of ~17.3 billion callable bases concentrated in a single amplicon, PPP2R5D.2 (**Figure 4.9B**). Despite the skew in coverage, 2974 (96.3%) of the original 3089 callable positions achieved median coverage >5000x (**Figure 4.9C**).


A. Sequencing coverage by sample
Legend
SD
Tes30G
Tes33D
Tes9E
Tes4
BC
Coverage (Reads)
10^0
10^1
10^2
10^3
10^4
10^5
10^6
B. Sequencing coverage by rhPCR amplicon
Legend
Rep. 1
Rep. 2
Rep. 3
Coverage (Reads)
PPP2R5D.2
PPP1CB
PPP2R5D.1
ALG13
HRAS
FGFR3.1
PACS1
FGFR3.2
CSNK2A1
SMAD4
RET.1
FGFR2
PTPN11.2
PTPN11.1
RET.2
SHOC2
FGFR3.3
GNB1
ADNP
CDK13

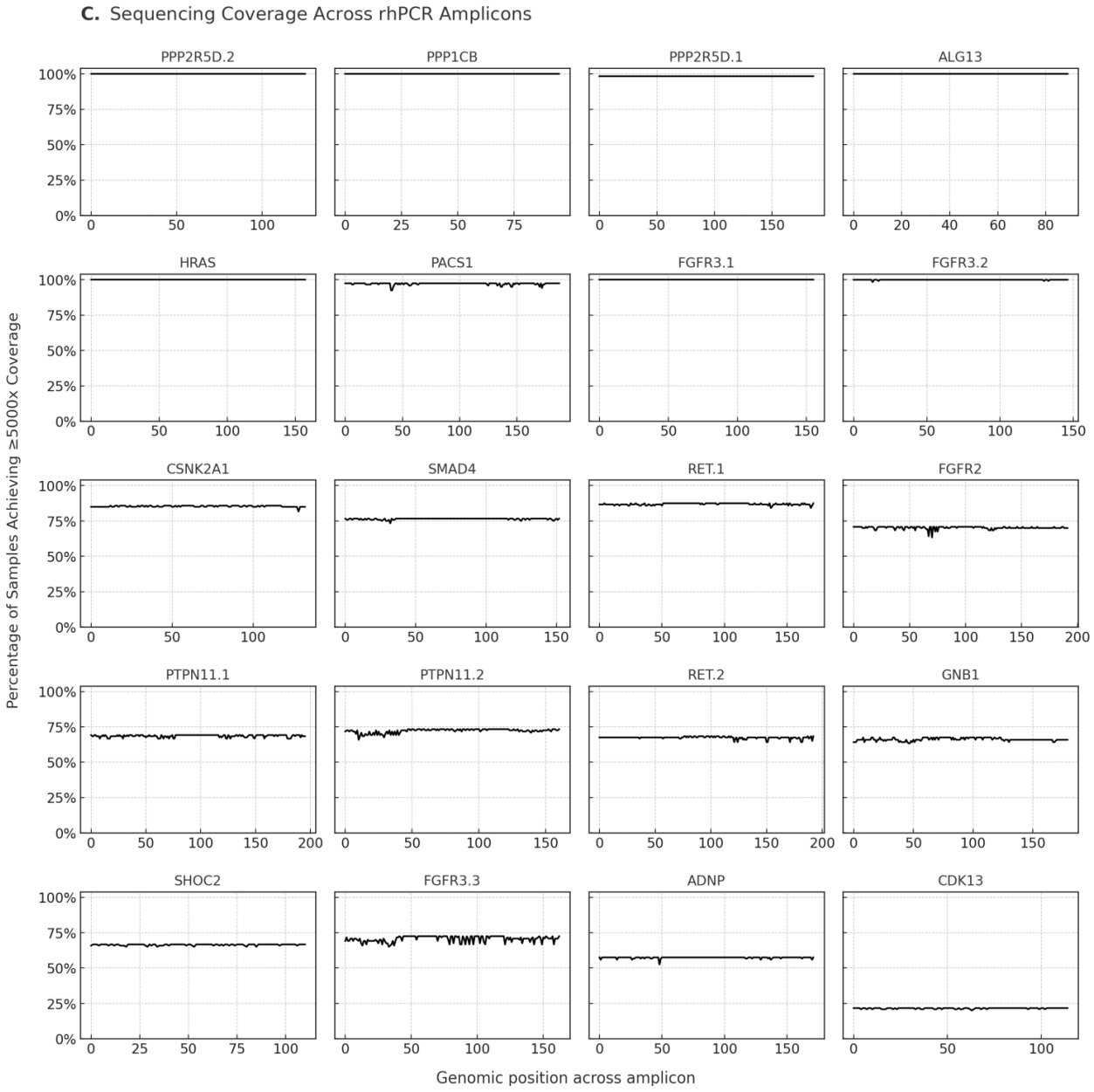


Figure 4.9: Sequencing coverage in the rhAmpSeq pilot study. In this figure, bar plots depicting median coverage and IQR for all libraries (n = 576) are colour-coded by tissue source material. **(A)** Coverage across all amplicons is depicted for each sample grouped by sample type, as well as for each **(B)** rhAmpSeq amplicon organized in descending order. **(C)** Percentage of samples at achieving ≥5000x coverage at each position in each amplicon.

Following UMI correction, the median effective coverage was 5725x per sample, and 1547 positions (50.1%) maintained median coverage >5000x. The distribution of UMI group size was

within observed values in the proof-of-concept experiment (**Figure 4.10A**), indicating that the increased amplicon complexity did not alter UMI formation. The mean UMI group size (2.11) in this pilot study was consistent with previously observed values (2.17) despite the approximately 5-fold lower median coverage (**Figure 4.10B**).

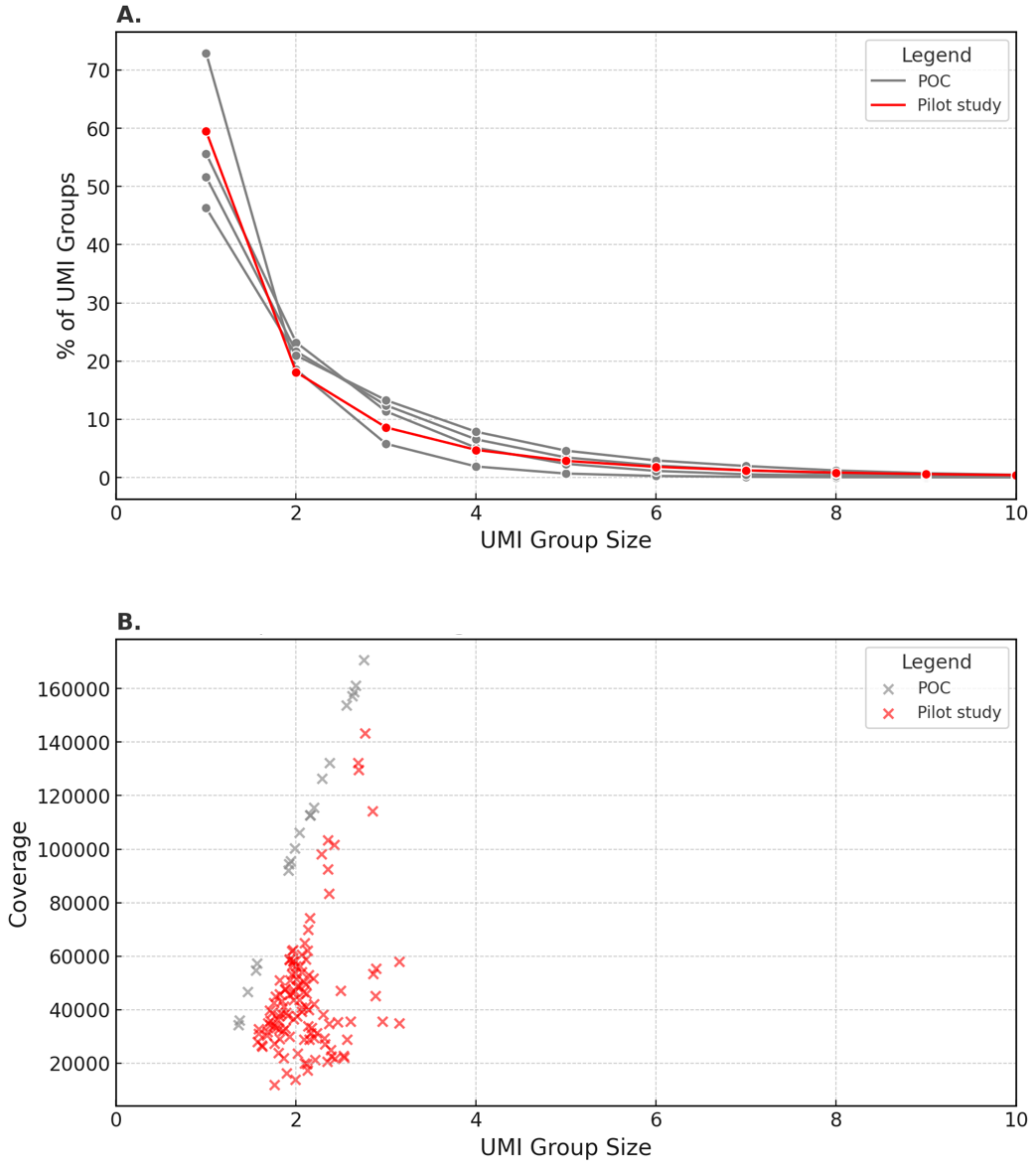


Figure 4.10: UMI group distribution in the rhAmpSeq pilot study. (A) UMI group-size distribution vs. proof-of-concept study. **(B)** UMI group size vs. coverage.

4.3.2.2 – Statistical prioritization of variants

Statistical prioritization identified 796 variant calls across 424 genomic positions. To eliminate likely artefacts, any base calls present in the blood control or observed at non-spiked positions in the serial dilution samples were removed, reducing this set to 618 variant calls (335 genomic positions). Only variant calls occurring in at least 2 technical replicates were retained, resulting in 180 base calls across 24 positions, which corresponded to 67 variants (46 in Tier 1, 21 in Tier 2).

To maximize data retention following UMI-correction, parameters for statistical modeling were relaxed for minimum median coverage per position (5000x to 1000x). This increased the total number of variants calls from 195 (151 genomic positions) to 341 (171 unique positions). Using relaxed parameters, there were 110 calls across 19 unique positions (vs. 47 calls and 11 positions), which corresponded to 43 variants (24 in Tier 1, 19 in Tier 2).

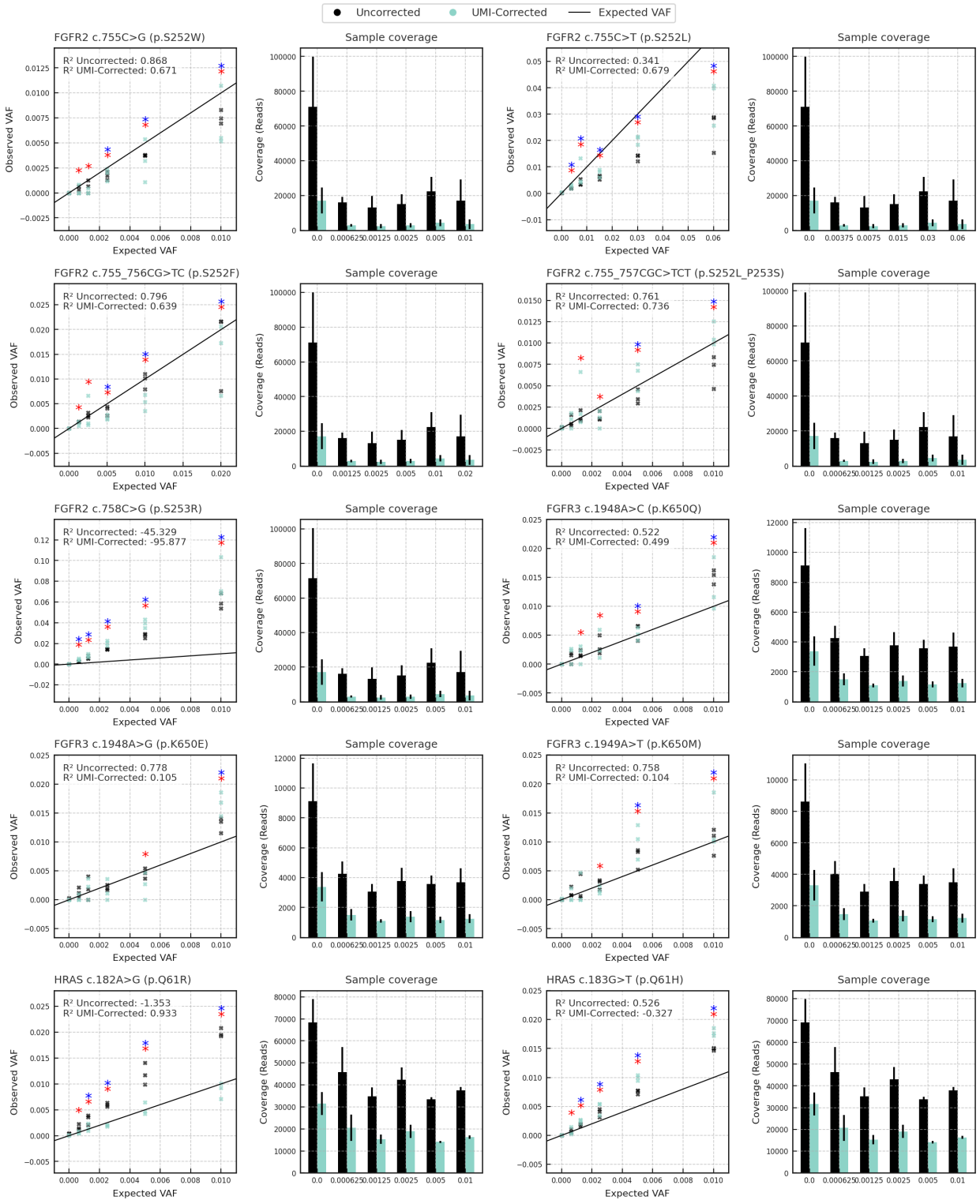
4.3.2.3 – Sensitivity of rhAmpSeq the pilot study

Thirteen serially diluted mutations were analyzed to characterize the sensitivity of the experiment. For uncorrected data, VAFs among replicates fell within ~1.6-fold of each other (range 1.0–14.4), and mean VAFs across replicates were within ~1.3-fold (range 0.3–6.3) of expected values (**Figure 4.11**). For UMI-corrected data, there were too many zero values to reliably estimate variability between sample replicates; mean VAFs across replicates were within ~0.7-fold (range 0–4.2) of expected values. In one case, FGFR2 c.758C>G, measurements were ~6-fold overestimated at all dilutions, which likely stemmed from inaccurate DNA quantitation. This outlier was retained in **Figure 4.11** but excluded from subsequent analyses to avoid inflating estimates of the limit-of-detection (LOD).

Of the 67 detected variants across UMI-uncorrected data, 48 (37 Tier 1 and 11 Tier 2) corresponded to the spiked-in mutations from serial dilutions. For the two mutations spiked in at

0.02 and 0.06 VAF, all dilutions were distinguishable down to 0.00125 and 0.00375 VAFs, respectively (**Figure 4.11**). The two SMAD4 mutations were only called at the starting dilution of 0.001 VAF and, together with the eight mutations spiked in at 0.01 VAF (excluding the *FGFR2* c.758C>G outlier), dilutions to ~0.0005–0.000625 and ~0.001–0.00125 VAF were called in 3 of 10 and 7 of 10 samples, respectively. At the lowest dilution, 2 of 4 transversions were detected, whereas only 1 of 6 transitions were called. By contrast, at ~0.001–0.00125 VAF, 3 of 4 transversions and 4 of 6 transitions were called. The LOD for this experiment was therefore set at 0.0005 VAF for transversions and 0.001 VAF for transitions. Following LOD filtering, 13 of the remaining 19 non-spiked-in variants were retained (8 in Tier 1 and 5 in Tier 2).

To compare sensitivity before and after UMI correction, it was most instructive to assess highly covered positions. For example, for both spiked-in mutations in *HRAS* (one transition, c.182A>G (p.Q61R), and one transversion, c.183G>T (p.Q61H)), median effective coverage ~16,858x, range 9354–38,663x), only dilutions down to 0.00125 VAF were called. At the 0.000625 VAF dilution, only 0 of 3 and 1 of 3 replicates, respectively, were called. By contrast, without UMI correction, both spiked-in *HRAS* mutations were called to 0.000625 VAF as Tier 1 variants. Similarly, between spiked-in mutations in *SMAD4* (median effect coverage 8512x, range 624–54,760x), only the starting dilution (0.001 VAF) was called for c.1499T>C. Taken together, the LOD following UMI-correction – even for positions with high effective coverage – was limited to ~0.001 VAF.


Uncorrected
UMI-Corrected
Expected VAF
FGFR2 c.755C>G (p.S252W)
R² Uncorrected: 0.868
R² UMI-Corrected: 0.671
Sample coverage
FGFR2 c.755C>T (p.S252L)
R² Uncorrected: 0.341
R² UMI-Corrected: 0.679
Sample coverage
FGFR2 c.755_756CG>TC (p.S252F)
R² Uncorrected: 0.796
R² UMI-Corrected: 0.639
Sample coverage
FGFR2 c.755_757CGC>TCT (p.S252L_P253S)
R² Uncorrected: 0.761
R² UMI-Corrected: 0.736
Sample coverage
FGFR2 c.758C>G (p.S253R)
R² Uncorrected: -45.329
R² UMI-Corrected: -95.877
Sample coverage
FGFR3 c.1948A>C (p.K650Q)
R² Uncorrected: 0.522
R² UMI-Corrected: 0.499
Sample coverage
FGFR3 c.1948A>G (p.K650E)
R² Uncorrected: 0.778
R² UMI-Corrected: 0.105
Sample coverage
FGFR3 c.1949A>T (p.K650M)
R² Uncorrected: 0.758
R² UMI-Corrected: 0.104
Sample coverage
HRAS c.182A>G (p.Q61R)
R² Uncorrected: -1.353
R² UMI-Corrected: 0.933
Sample coverage
HRAS c.183G>T (p.Q61H)
R² Uncorrected: 0.526
R² UMI-Corrected: -0.327
Sample coverage
Observed VAF
Expected VAF
Coverage (Reads)

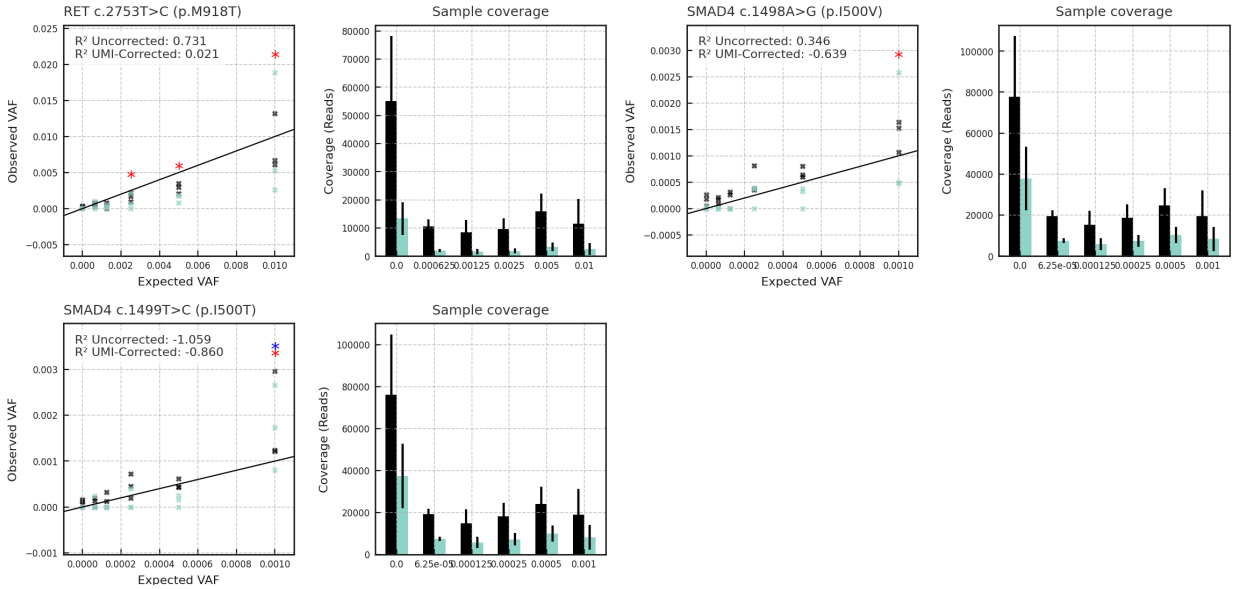


Figure 4.11: Sensitivity of the rhAmpSeq pilot study. Observed mutation levels across 13 serially diluted mutations were individually plotted for all 3 replicates (black, uncorrected; teal, UMI-corrected). Dilutions called by statistical prioritization are indicated by asterisks (red, uncorrected; blue, UMI-corrected). Accompanying bar plots indicate the mean (± SD) coverage for each dilution across replicates at the given position. For each mutation, the coefficient of determination (R^2) is plotted for both uncorrected and UMI-corrected data.

4.3.2.4 – Detection of control variants in previously screened testes biopsies

Previously screened testis biopsies (Tes4, 90 years) harbouring known selfish mutations in *FGFR3*, *PTPN11*, and *RET* were included as positive controls (range 0.0015–0.0047 VAF). Following a late change to the control amplicons, an additional mutation, *PTPN11* c.1504T>C (p.S502T) located in Tes4D-25 at ~0.0025 VAF, lay outside of the reconfigured panel. From 13 retained variants from UMI-uncorrected data, all 3 control variants were each detected in Tier 1 (**Table 4.2**). Measured VAFs across technical replicates were within ~1.4 to 2.7-fold of each other, and mean VAFs across replicates were within ~1.0 to 1.6-fold of previously reported values (**Figure 4.12**). UMI-correction also yielded detection of all 3 control variants, but only the *FGFR3* c.1620C>A (p.N540) variant was prioritized in Tier 1 (**Table 4.2**). VAFs across technical replicates were within ~1.3 to 6-fold of each other, and the mean VAFs across replicates were within ~1.1 to 1.5-fold of previously reported values (**Figure 4.12**).

Table 4.2 – Selfish mutations identified in previously-screened testis biopsies from an aged man (Tes4, 90 years)

Variant	Gene	Genomic position (hg38/GRCh38)	cDNA mutation (AA change)	Germline disorder	Sample (Tes; Biopsy)	Variant reads	Total reads	VAF	p-value	Expected VAF[1]
2	FGFR3	chr 4: 1805644	c.1620C>A (p.N540K)	Hypochondroplasia (OMIM: #146000)	4C; 6	62	43450	0.0014	1.4E-78	0.0018 (range 0.0015–0.0020)
						37	31807	0.0012	6.0E-42	
						43	49260	0.00087	3.2E-44	
						UMI variant reads	**UMI total reads**	**UMI VAF**	**p-value**	
						25	20394	0.0012	4.1E-21	
						20	14724	0.0014	4.6E-18	
						25	23014	0.0011	2.5E-21	
						Variant reads	**Total reads**	**VAF**	**p-value**	
3	RET	chr 10: 43118458	c.2370G>C (p.L790F)	Multiple endocrine neoplasia IIA (OMIM: #171400)	4F; 1	28	27043	0.0010	3.2E-46	0.0015 (range 0.0012–0.0018)
						88	35829	0.0025	1.1E-194	
						26	29094	0.00089	4.3E-43	
						UMI variant reads	**UMI total reads**	**UMI VAF**	**p-value**	
						9	10909	0.00083	6.3E-07	
						39	15268	0.0026	5.7E-64	
						5	11816	0.00042	NS	
						Variant reads	**Total reads**	**VAF**	**p-value**	
1	PTPN11	chr 12: 112477719	c.922A>G (p.N308D)	Noonan syndrome (OMIM: #163950)	4D; 17	16	5335	0.0030	1.4E-09	0.0047 (range 0.0034–0.0069)
						48	14830	0.0032	3.8E-39	
						18	3919	0.0046	1.5E-14	
						UMI variant reads	**UMI total reads**	**UMI VAF**	**p-value[2]**	
						4	1763	0.0023	NS	
						8	5540	0.0014	4.4E-04	
						7	1208	0.0058	3.7E-07	

[1] Expected VAFs were averaged across 2–3 independent sequencing calls from Maher et al., 2018.
[2] Relaxing statistical prioritization parameters (e.g. reducing minimum variant reads from 5 to 4) did not influence results.

NS = not significant

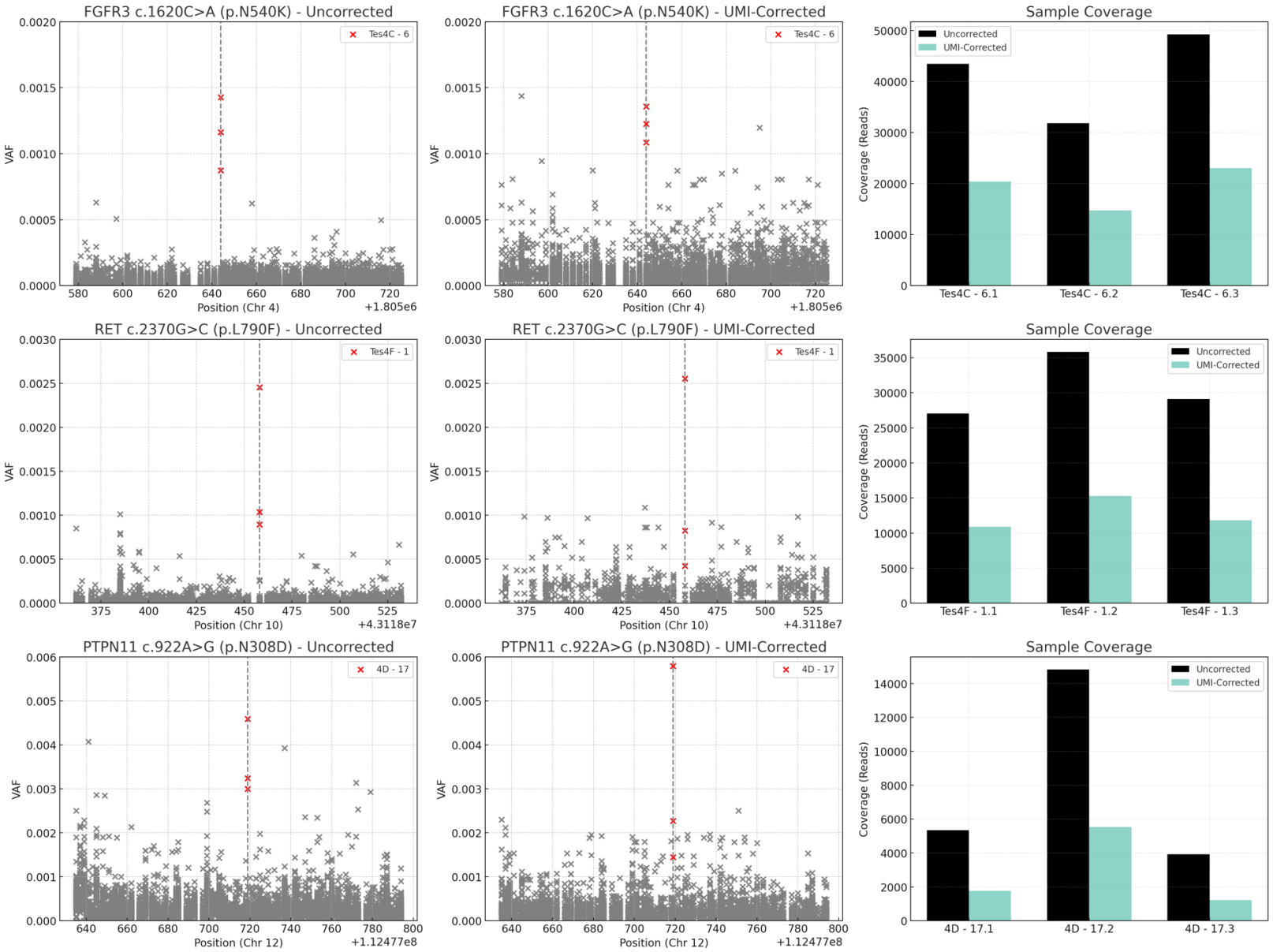


Figure 4.12: Mutation plots for three control variants in previously screened testis biopsies (Tes4, 90 years). *FGFR3* c.1620C>A (p.N540K, *PTPN11* c.922A>G (p.N308D), *RET* c.2370G>C (p.L790F). VAFs for the respective substitutions are plotted along the respective amplicons for both uncorrected and UMI-corrected reads. A dotted vertical line denotes the position of interest, with the three replicates of biopsy sample highlighted in red. Sample coverage is indicated in a bar plot on the right for both sets of samples.

4.3.2.5 – Detection of novel variants in the pilot study

36 biopsies from 3 aged men (Tes9E, 73 years; Tes30G, 63 years; Tes33D, 76 years) were assayed to determine whether selfish variants were detectable prior to proceeding with a larger-scale analysis. Of these 36 biopsies, 29 (80.1%) were represented by at least 2 technical replicates following filtering for sample failure. While the 4 biopsies from Tes4 (90 years) were included in this pilot study as positive controls, they were also assessed for novel variants.

Across uncorrected data, 10 novel variants (5 in Tier 1 and 5 in Tier 2) were detected across these 40 testis biopsies (**Table 4.3**). There were 6 variants called in 3 established selfish genes (*FGFR3*, *HRAS*, and *PTPN11*) and all corresponded to known mutations in PAE disorders. Importantly, known selfish variants were detected in testis biopsies from all slices: Tes9E (n = 2), Tes30G (n =1), and Tes33D (n = 1), supporting further study. The most frequently observed variant was *PTPN11* c.215C>G (p.A72G) in two biopsies (50 and 60) from Tes9E (73 years) as a Tier 1 and 2 variant (VAF range 0.005–0.0092) (variants P5 and P10). This variant is a commonly observed germline mutation in Noonan syndrome (Tartaglia et al., 2005) and has been detected as mutational foci in the testis in prior studies (Maher et al., 2018; Eboreime et al., 2022). In Tes30G-17, the *FGFR3* c.1138G>A (p.G380R) was detected as a Tier 2 variant (~0.0052 VAF), although it was noted that 1 of 3 replicates experienced sample failure (variant P8). This variant is one of the canonical selfish mutations and underlies achondroplasia (Bellus et al., 1995). In Tes33D, *PTPN11* c.181G>C (p.Y63C) was detected as a Tier 2 variant at ~0.0029 VAF, and it was also noted that the third replicate had low coverage at the position (1425x) which may have precluded its identification (variant P9, **Table 4.3**). This variant is among the most recurrent mutations in Noonan syndrome (Tartaglia et al., 2005). The *HRAS* c.179G>A (p.G60D) variant, a known mutation in Costello syndrome (Gripp et al., 2019) was initially detected as a Tier 1 variant at

~0.0009 VAF in the Tes4D-25 control biopsy but was reclassified as a Tier 2 variant following the establishment of the limit of detection (variant P7).

The remaining **4** variants (3 in Tier 1, 1 in Tier 2) were detected across 3 different candidate genes: *PPP2R5D* (n=2), *PACS1* (n=1), and *CSNK2A1* (n=1). One Tier 1 variant, *PPP2R5D* c.598G>A (p.E200K), was detected at ~0.0016 VAF in a control biopsy, Tes4C-6 (variant P1). *PPP2R5D* was not covered in the study by Maher et al., but the p.E200K variant is the second most recurrent mutation in Jordan syndrome (Levine and Chung, 2023). Another Tier 1 variant, c.627A>C (p.H209P), was also identified in *PPP2R5D* at ~0.00064 VAF in Tes33D-16 (variant P2). This variant not been associated with Jordan syndrome but is located in the same conserved loop as p.E200 that mediates contact with other PP2A subunits (**section 3.3.2**) and has been reported in the ClinGen database (Pawliczek et al., 2018). A synonymous variant in *PACS1*, c.522G>A (p.K184K), was also detected as a Tier 1 variant in the same control biopsy (Tes4C-6) harbouring the pathogenic *PPP2R5D* c.598G>A (p.E200K) and *FGFR3* c.1620C>A (p.N540K) variants, potentially suggestive of a passenger mutation (variant P6). Lastly, *CSNK2A1* c.521T>C (p.I174R) was detected as a Tier 2 variant at ~0.0011 VAF in Tes30G-2. While this particular variant has not been reported in Okur-Chung neurodevelopmental syndrome, mutations at this residue (e.g. c.522A>G (p.I174M)) have been reported (Owen et al., 2018).

A subset of the above variants was detected following UMI correction. Of the 5 Tier 1 variants detected (median 0.0019 VAF; range 0.00065–0.0092), 4 were detected (3 as Tier 1, and 1 as a Tier 2 variant) following UMI correction. Similarly, of the 5 Tier 2 variants detected, 4 were detected (all Tier 2). The 8 detected variants following UMI correction had median VAFs of ~0.0033 (range 0.00092–0.0033) whereas the 2 undetected variants had median VAFs of 0.00078 (range 0.00065–0.00091), below the implied UMI-corrected LOD (**section 4.3.2.3**).

Table 4.3: Novel variants identified in the rhAmpSeq pilot study. 36 testis biopsies from 3 aged men (average 70.7 years; range 63–76 years) including 4 control biopsies from a 90-year-old man.

Tier	Variant	Sample (Tes; Biopsy)	Gene	Genomic position (hg38/GRCh38)	cDNA mutation (AA change)	COSMIC[1] v101	Germline disorder	gnomAD[2] v4.1.0	Variant reads	Total reads	VAF	p-value
1	P1	4C; 6	PPP2R5D	Chr6: 43007271	c.598G>A (p.E200K)	-	Jordan syndrome (OMIM: #616355)	-	181	83115	0.0022	3.6E-95
									51	56423	0.00090	7.0E-08
									177	102732	0.0017	2.5E-77
								UMI	101	38249	0.0026	3.1E-64
									26	26068	0.0010	7.6E-04
									75	46570	0.0016	2.9E-32
1	P2	33D; 16	PPP2R5D	Chr6: 43007299	c.627A>C (p.H209P)	-	-	-	94	190190	0.00049	5.2E-11
									303	383640	0.00079	1.2E-88
									109	166719	0.00065	1.6E-22
								UMI	15	75279	0.00020	NS
									17	105575	0.00016	NS
									19	69236	0.00027	NS
1	P3	4C; 6	PACS1	Chr11: 66211151	c.552G>A (p.K184K)	-	-	6.20E-7	101	55633	0.0018	5.9E-79
									108	39763	0.0027	1.0E-102
									91	68930	0.0013	5.5E-59
								UMI	53	26886	0.0020	2.9E-40
									55	18232	0.0030	5.7E-53
									47	33369	0.0014	6.5E-30
1	P4	33D; 41	PTPN11	Chr12: 112450361	c.181G>C (p.D61H)	9 (9, 169) Leukemia	Noonan syndrome (OMIM: #163950)	-	19	1592	0.012	4.5E-18
									332	43360	0.0077	0.0E+00
									355	45418	0.0078	0.0E+00
								UMI	5	189	0.026	NS
									107	10075	0.011	4.4E-93
									100	10677	0.0093	1.4E-81
1	P5	9E; 50	PTPN11	Chr12: 112450395	c.215C>G (p.A72G)	4 (4, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	74	21697	0.0034	6.7E-37
									31	7481	0.0041	2.9E-14
									153	42949	0.0036	2.6E-83
								UMI	21	9186	0.0023	2.7E-08
									10	2693	0.0037	1.6E-03
									47	20098	0.0023	1.7E-24

2	P6	30G; 2	CSNK2A1	Chr20: 492354	c.521T>C (p.I174R)	-	-	-	1	1469	0.00068	NS
									53	44570	0.0012	1.3E-45
									47	49676	0.00095	1.6E-35
								UMI	0	412	0	NS
									10	20293	0.00049	7.6E-09
									4	22911	0.00018	NS
2	P7	4D; 25	HRAS	Chr 11: 533877	c.179G>A (p.G60D)	3 (3, 6) Upper aerodigestive tract; Stomach	-	-	99	113675	0.00087	1.1E-24
									68	79165	0.00086	2.7E-15
									187	181659	0.0010	3.1E-60
								UMI	44	51228	0.00086	1.02E-07
									36	36678	0.00098	4.07E-07
									36	73290	0.00090	8.8E-15
2	P8	30G; 17	FGFR3	Chr 4: 1804392	c.1138G>A (p.G380R)	37 (38, 41) Bladder; Nasopharynx ; Skin	Achondroplasia (OMIM: #100800)	4.34E-6	116	23450	0.0049	8.9E-84
									-	-	-	-
									71	15384	0.0046	1.2E-47
								UMI	57	9868	0.0058	3.4E-39
									-	-	-	-
									40	5936	0.0067	9.6E-29
2	P9	33D; 41	PTPN11	Chr 12: 112450368	c.188GA>G (p.Y63C)	60 (60, 169) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	140	38686	0.0036	1.1E-86
									145	40322	0.0036	1.4E-89
									2	1425	0.0014	NS
								UMI	15	10095	0.0015	1.64E-08
									16	10695	0.0015	1.99E-09
									-	-	-	-
2	P10	9E; 60	PTPN11	Chr 12: 112450395	c.215C>G (p.A72G)	4 (4, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	66	15683	0.0042	7.2E-38
									203	34931	0.0058	2.2E-151
									-	-	-	-
								UMI	22	6253	0.0035	1.0E-12
									68	15580	0.0044	1.9E-54
									-	-	-	-

[1] Counts and tumour types from the COSMIC v101; counts are grouped as: (n of specific DNA change, n of specific acid substitution, total variants affecting that residue).
[2] Allele frequencies from the gnomAD database.
Empty values are denoted by with a '-'and indicate sample failure.
NS = not significant.

4.3.2.6 – Panel improvements and exclusion of UMIs

A key limitation in the pilot study was coverage skew, with the PPP2R5D.2 amplicon consuming ~27.7% of the sequencing run. Although equimolar concentrations of rhPCR primers were recommended (IDT), relative rhPCR primer concentrations were adjusted before proceeding to the full-scale analysis to improve coverage uniformity. Despite these skew, overall assay sensitivity remained high: known mutations were recovered between ~0.0005–0.001 VAF. Variants in selfish genes were also detected in testis biopsies from each slice, supporting the suitability of these tissues for further inquiry.

UMI correction did not improve assay sensitivity. Across two serial dilutions, UMI correction reduced the ability of rhAmpSeq to detect spiked-in mutations at or below ~0.001 VAF. Although known selfish mutations in control biopsies were still detected following UMI correction, in two of three instances, UMI correction resulted in loss of signal in 1 of 3 replicates, even under relaxed statistical conditions. For these reasons, UMI correction was excluded from the full discovery screen to preserve overall sensitivity.

4.3.3 – Discovery screen in aged human testes

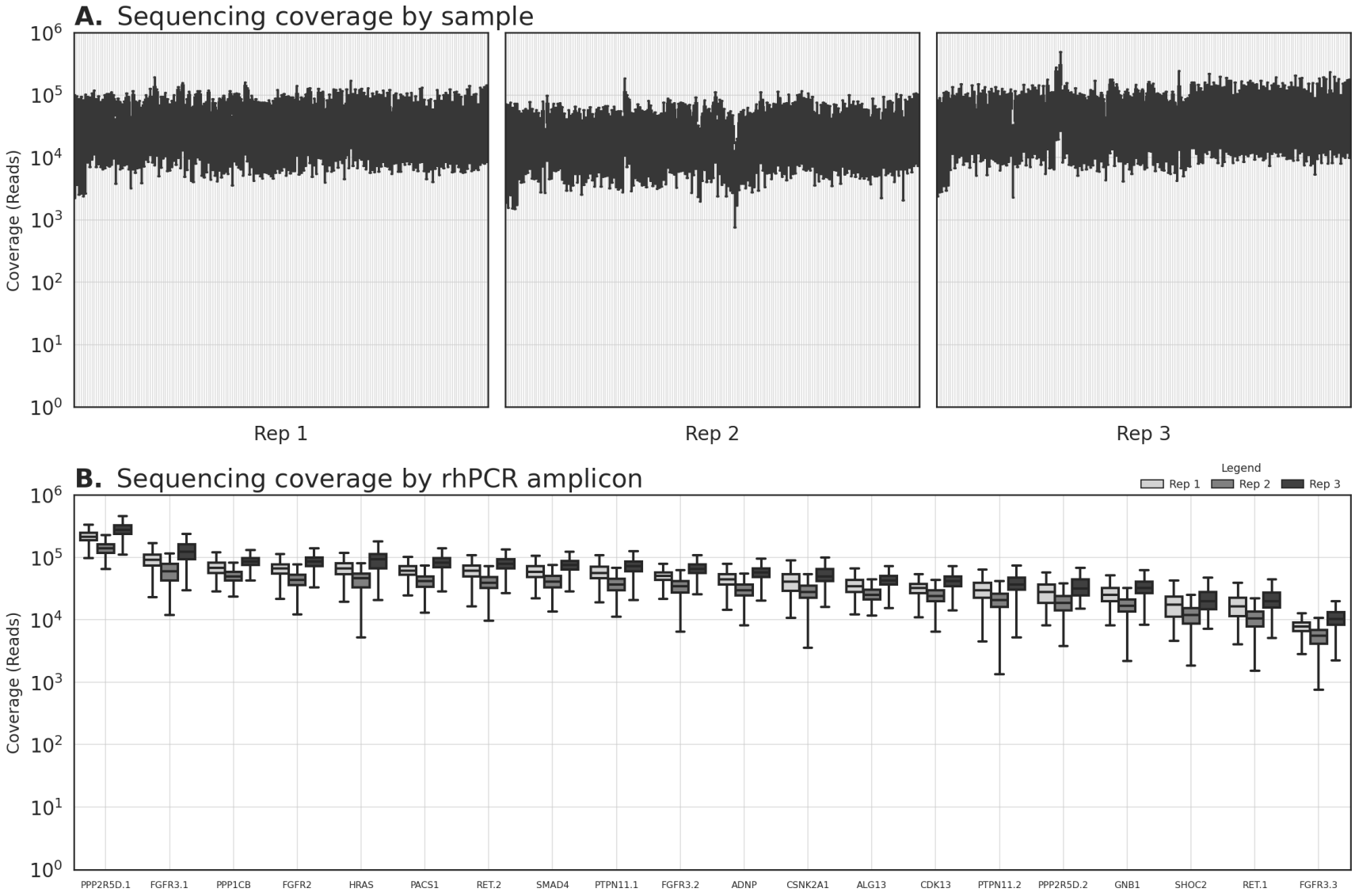
A total of 186 testes biopsies, including resequencing of the 7 biopsies that experienced sample failure in at least two replicates in the pilot study (**section 4.3.2.1**), and the same 6 samples harbouring the 13 serially diluted DNA were processed in technical triplicates (3 x 30,000 haploid genome equivalents), resulting in 576 total samples. Libraries comprised of one technical replicate for each of the 192 samples were sequenced on separate NovaSeq runs to distribute potential batch effects and reduce the risk of sequencing run-specific artefacts skewing the overall results (**Figure 4.8B**).

4.3.3.1 – Sequencing coverage

Median coverage across 576 samples was 44,445x and was consistent between replicates with no sample failure (**Figure 4.13A**). The median coverage varied by ~27-fold between the lowest and highest covered amplicons (FGFR3.3 and PPP2R5D.1; range 7600–205,371x) (**Figure 4.13B**). Moreover, 100% and 94.7% of the 3089 genomic positions had median coverage >5000x and >10,000x, respectively (**Figure 4.13C**).

4.3.3.2 – Statistical prioritization of variants

Statistical prioritization identified 3487 outliers distributed across 676 unique positions. Base calls present in non-spiked positions in the serial dilution samples were removed, reducing this set to 1877 variant calls across 639 genomic loci. Following filtering for calls that were observed across ≥2 technical replicates, 380 calls across 48 positions were made, corresponding to 137 variants (108 in Tier 1, 29 in Tier 2)


A. Sequencing coverage by sample
Coverage (Reads)
10^6
10^5
10^4
10^3
10^2
10^1
10^0
Rep 1
Rep 2
Rep 3
B. Sequencing coverage by rhPCR amplicon
Legend
Rep 1
Rep 2
Rep 3
Coverage (Reads)
PPP2R5D.1
FGFR3.1
PPP1CB
FGFR2
HRAS
PACS1
RET.2
SMAD4
PTPN11.1
FGFR3.2
ADNP
CSNK2A1
ALG13
CDK13
PTPN11.2
PPP2R5D.2
GNB1
SHOC2
RET.1
FGFR3.3

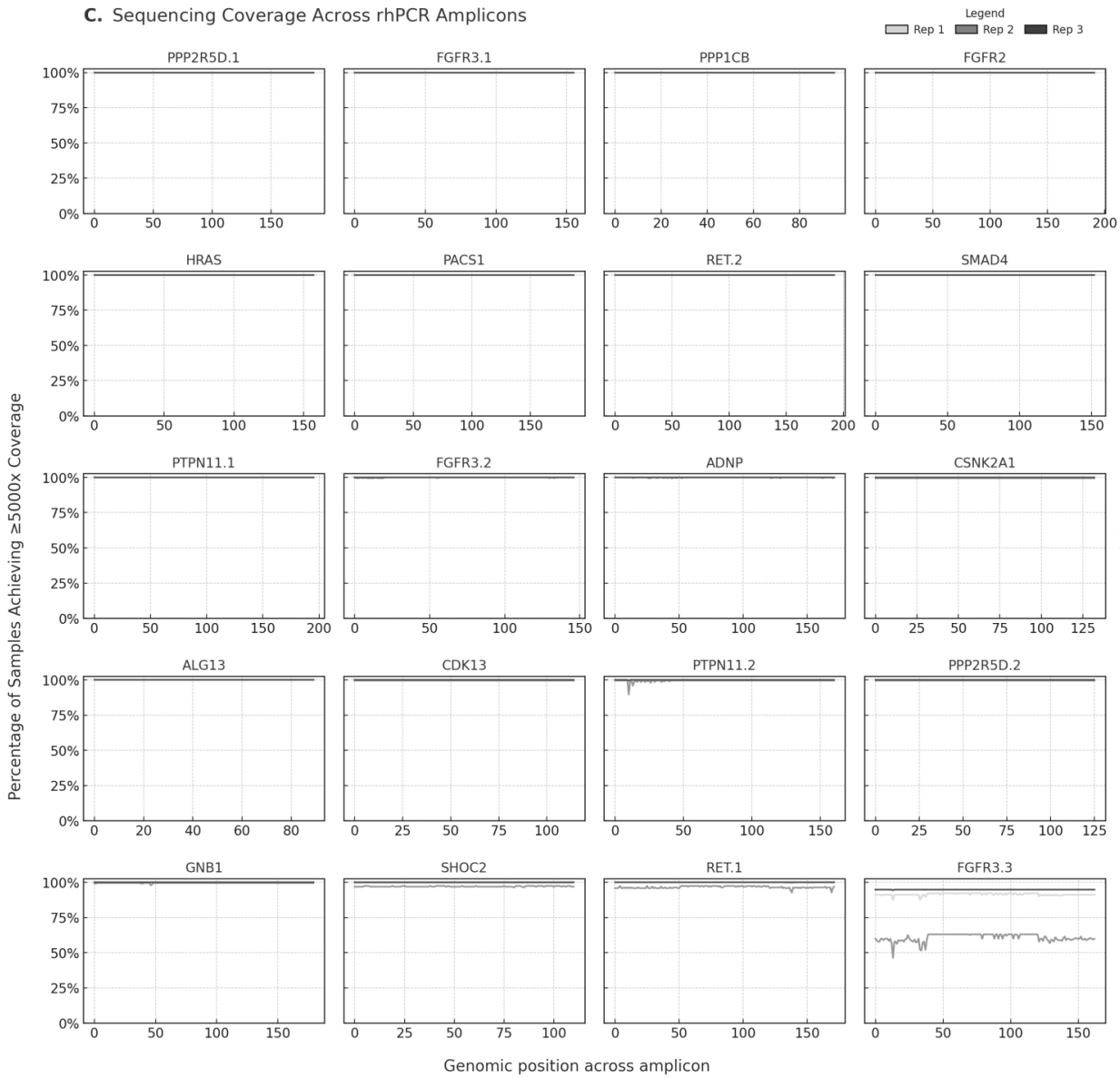


Figure 4.13: Sequencing coverage in the rhAmpSeq discovery screen. (A) Sequencing coverage for each sample, organized according to technical replicates. **(B)** Sequencing coverage for each rhAmpSeq amplicon, grouped by sequencing run. **(C)** Percentage of genomic positions within each amplicon that achieve at least 5000x coverage.

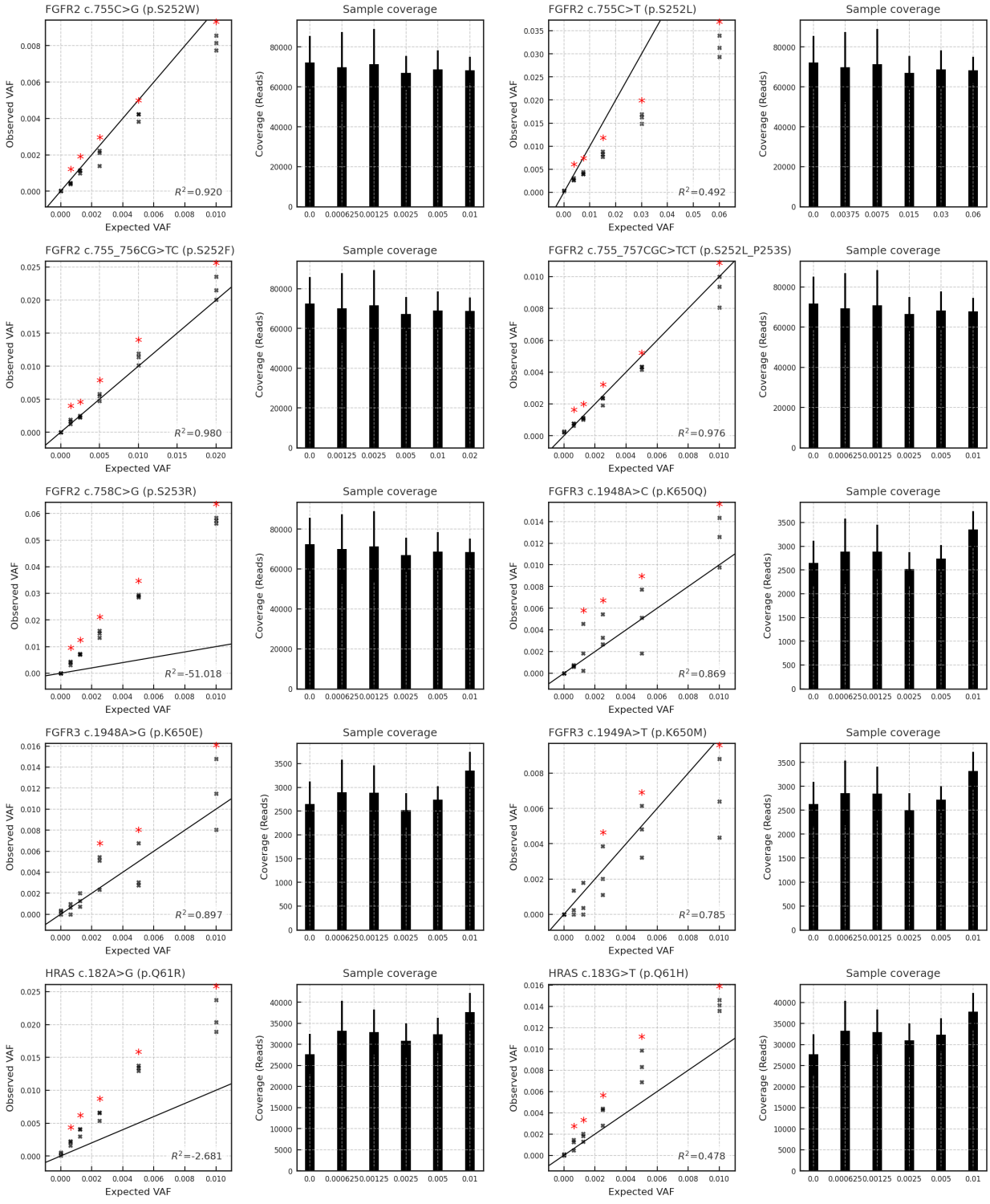
4.3.3.3 – Sensitivity of the rhAmpSeq discovery screen

Observed VAFs among replicate groups were within ~1.3-fold of each other (range 1–18.1) and the mean replicate VAFs were within ~1.0-fold of expected values (range 0.5–6.3) (**Figure 4.14**). As before, the *FGFR2* c.758C>G mutation was excluded to prevent inflation of the limit of

detection. The exclusion of this mutation resulted in mean replicate VAFs that were within ~1.0-fold of expectations (range 0.5–3.3), and ~1.2-fold (range 0.7–3.3) at lower VAFs (<0.0025)

Of 137 total variants, 55 corresponded to spiked-in mutations. All dilutions down to 0.00125 and 0.00375 VAF were detectable for mutations introduced at 0.02 and 0.06 VAF, respectively. Mutations were identifiable in 5 of 10 and 8 of 10 cases down to ~0.0005-0.000625 VAF and 0.001–0.00125 VAF, respectively. The limiting mutations occurred in the FGFR3.3 amplicon, which had low median coverage across serially diluted samples (~2848x) (**Figure 4.13C**). At higher read depths (>10,000x), mutations were detectable in 6 of 7 (2 of 2 transversions; 4 of 5 transitions) and 7 of 7 cases down to 0.0005–0.000625 VAF and 0.001–0.00125 VAF, respectively. Although the 0.00025 VAF dilutions for both SMAD4 mutations were significant by calling criteria, the replicate measurements at these dilutions were highly scattered and many observed values exceeded the range at 0.0005 VAF (**Figure 4.14**).

Given this, a broad limit of detection for both transitions and transversions of 0.0005 VAF (1 in 2000) was set to filter variant calls. Following filtering for the LOD, 71 of 82 non-serially diluted variants were retained (56 Tier 1 and 15 Tier 2).


FGFR2 c.755C>G (p.S252W)
Observed VAF
Expected VAF
R^2=0.920
Sample coverage
Coverage (Reads)
FGFR2 c.755C>T (p.S252L)
R^2=0.492
FGFR2 c.755_756CG>TC (p.S252F)
R^2=0.980
FGFR2 c.755_757CGC>TCT (p.S252L_P253S)
R^2=0.976
FGFR2 c.758C>G (p.S253R)
R^2=-51.018
FGFR3 c.1948A>C (p.K650Q)
R^2=0.869
FGFR3 c.1948A>G (p.K650E)
R^2=0.897
FGFR3 c.1949A>T (p.K650M)
R^2=0.785
HRAS c.182A>G (p.Q61R)
R^2=-2.681
HRAS c.183G>T (p.Q61H)
R^2=0.478

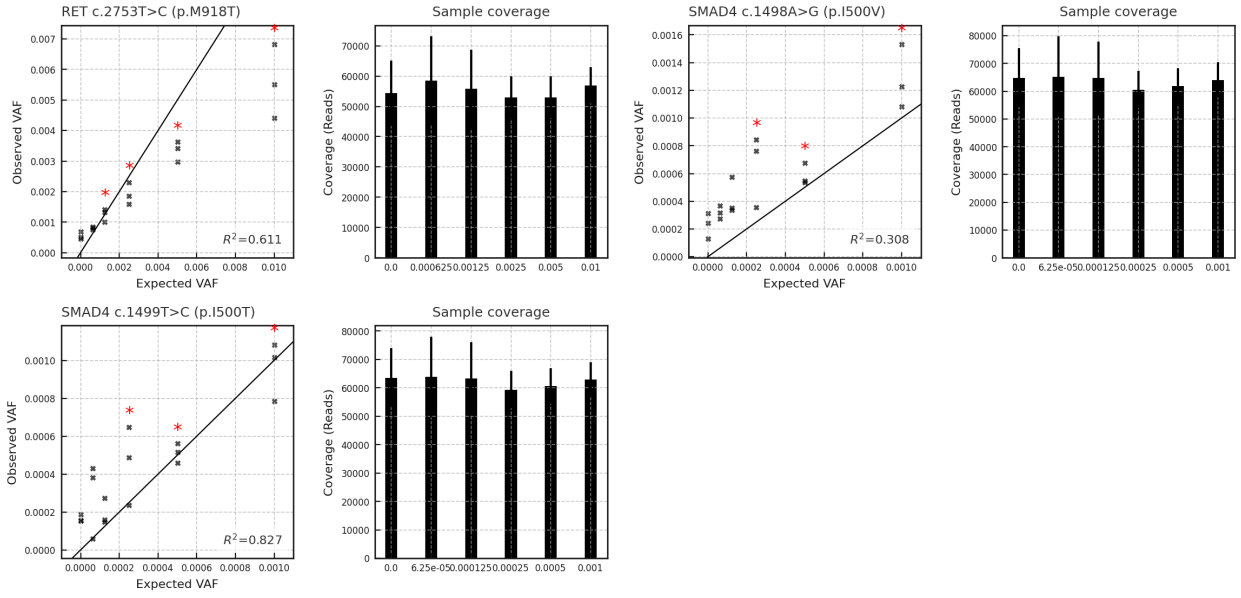


Figure 4.14: Sensitivity of the rhAmpSeq discovery screen. Observed mutation levels across 13 serially diluted mutations were individually plotted for all 3 replicates, with the solid black line serving as reference for expected VAFs. If a given dilution was called through statistical prioritization, it is highlighted with a red asterisk. Each scatter plot has an accompanying bar plot indicating the mean (± SD) coverage for each dilution across replicates at the given position. For each mutation, the coefficient of determination (R^2) is plotted.

4.3.2.4 – Detection of novel variants across aged testes biopsies

Across the 186 testes biopsies, the 71 detected variants were made across 65 biopsies. Tier 1 variants were detected at higher frequencies (n = 56; median 0.0018 VAF, range 0.00053–0.0087) than those in Tier 2 (n = 15; median 0.00095 VAF; range 0.00020–0.0020) (p = 0.0003, Wilcoxon rank-sum test). Variants in *PTPN11* were the most commonly observed in this screen (34 total mutational events corresponding to 14 distinct variants), followed by *FGFR3* (10 mutational events corresponding to 7 distinct variants). Together, 37 distinct variants across 14 of 20 amplicons were detected (**Table 4.4**).

Table 4.4: Mutational events detected in the rhAmpSeq discovery screen by amplicon

Amplicon	Select loci of interest	Mutational events	Distinct variants
ADNP	p.Y719	1	1
ALG13	p.N107	-	-
CDK13	p.N842	-	-
CSNK2A1	p.K198	5	2
FGFR2	p.S252; p.P253	3	3
FGFR3.1	p.G380	6	4
FGFR3.2	p.N540	3	2
FGFR3.3	p.K650	1	1
GNB1	p.I80	-	-
HRAS	p.Q61	3	1
PACS1	p.R203	1	1
PPP1CB	p.P49	5	1
PPP2R5D.1	p.E198	3	3
PPP2R5D.2	p.E420	-	-
PTPN11.1	p.D61; p.A72; p.E76	29	10
PTPN11.2	p.N308	5	3
RET.1	p.L790	-	-
RET.2	p.M918	3	3
SHOC2	p.S2	-	-
SMAD4	p.R496; p.I500	3	2

53 of 71 variants (27 distinct) occurred in known selfish genes and were detected at a median VAF of 0.0016 (range 0.00020–0.0087) (**Table 4.5**). It was reassuring that 15 distinct variants that were prioritized as Tier 1 variants had previously been experimentally associated with selfish selection. These included canonical examples of selfish mutations such as *FGFR3* c.1138G>A (p.G380R) in achondroplasia and *PTPN11* c.922A>G (p.N308D) in Noonan syndrome (**Figure 4.15**). The majority of variants (24 of 27) in known selfish genes corresponded to known recurrent mutations in germline disorders or were located at known cancer hotspots. Of the three remaining, two were located in *FGFR3* and one was located in *RET*. The first, *FGFR3* c.1181C>T (p.T394M), does not have an established association with a germline disorder, however mutations at this amino acid residue (c.1180A>T (p.T394S)) have been described in rare cases of achondroplasia (Zhu et al., 2003). The second, *FGFR3* c.1173G>A (p.A391A), is a synonymous variant (Tier 2) located at a CpG dinucleotide within a recurrently mutated residue in

Crouzon syndrome with acanthosis nigricans (OMIM: #612247) (Meyers et al., 1995). The p.A391A variant was detected in the same biopsy (Tes30G-12) as the c.188A>G (p.Y63C) variant but at roughly 2x lower VAF, which could suggest a passenger mutation. Lastly, the c.2754G>C (p.M918I) variant was detected as a Tier 1 variant in Tes33D-25 and although it does not have an established association with a germline disorder, mutations at this residue (p.M918T) underlie virtually all cases of multiple endocrine neoplasia type 2B.

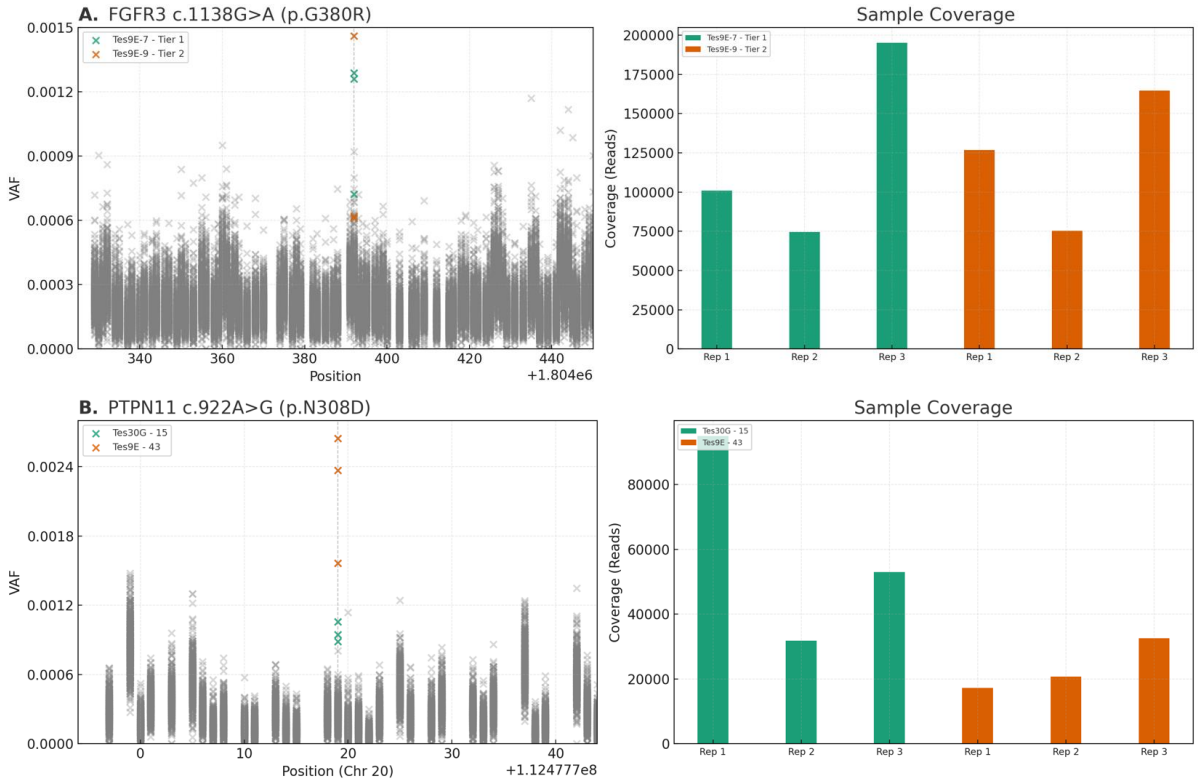


Figure 4.15: Example mutation plots of variants in control selfish genes identified in the rhAmpSeq discovery screen. For each variant, the VAF of all three replicates is plotted with colours representing replicates from the same biopsy. Label legends indicate which biopsies are called in which tiers. Bar plots indicating individual replicate coverage at the position are included.

Several of the same variants were common mutational events in the testes of different men, such as the *FGFR3* c.1620C>G (p.N540K) variant, which was detected as Tier 1 variants in individual biopsies of two men (Tes30G; 63 years and Tes33D; 76 years) (variant 7, **Table 4.5**). This variant is most recurrent mutation in hypochondroplasia, has been detected in prior discovery

screens in the testis and sperm (Maher et al., 2018; Salazar et al., 2022), and was a prioritized DNM (**section 3.4**). Similarly, three distinct *PTPN11* variants – c.188A>G (p.Y63C), c.214G>A (p.A72T), and c.922A>G (p.N308D) – were each detected across multiple biopsies and testis slices. For instance, c.188A>G (p.Y63C) was present in biopsies from Tes33D (n=3), Tes30G (n=1), and Tes9E (1); c.214G>A (p.A72T) was detected in Tes9E (n=3) and Tes33D (n=1); while c.922A>G (p.N308D) was identified in Tes30G (n =1) and Tes9E (n=1). Whereas the variants resulting in p.Y63C and p.N308D are among the most recurrent mutations in Noonan syndrome (the latter being a prioritized control DNM), the p.A72T mutation is a common somatic mutation in leukemia (Tartaglia and Gelb, 2005; Tartaglia et al., 2006).

The 10 distinct variants occurring in candidate selfish genes were detected at lower frequencies than those in established selfish genes (median 0.00078 VAF; range 0.00053–0.0038) ($p = 0.00019$, Wilcoxon rank-sum test) (**Figure 4.16**). Variants were detected across 5 candidate genes: *CSNK2A1* (4 events across 2 distinct variants), *PPP1CB* (5 events corresponding to 1 variant), *PPP2R5D* (3 events corresponding to 3 variants), *SMAD4* (3 events corresponding to 2 variants), *ADNP* (1 variant), *PACS1* (1 variant). 6 of 10 variants (1 each in *CSNK2A1, PACS1, PPP1CB,* and *PPP2R5D*, and 2 in *SMAD4*) had been specifically prioritized on the basis of DNM recurrence and their alignment with known selfish characteristics. 2 additional variants, *ADNP* c.2213C>G (p.S738*) and *PPP2R5D* c.598G>A (p.E200K), the latter of which was also detected in the pilot study, have established links to germline disorders associated with prioritized, recurrent DNMs (**Table 4.4**). The two remaining variants are *PPP2R5D* c.620G>C (p.W207S) (**Figure 4.16D**) and *CSNK2A1* c.575A>C (p.V192G) (**Figure 4.16H**) and do not have known disease associations, but the former occurs at a recurrently mutated residue in Jordan syndrome (Levine and Chung, 2023).

Table 4.5: Variants identified by rhAmpSeq screening of testes biopsies from three aged individuals (63 – 76 years)

Tier	Variant	Sample (Tes; Biopsy)	Gene	Genomic position (hg38/GRCh38)	cDNA mutation (AA change)	COSMIC v101	Germline disorder	gnomAD	Variant reads	Total reads	VAF	p-value
1	1	30G; 19	CSNK2A1	Chr 20: 492282	c.593A>G (p.K198R)	-	Okur-Chung neurodevelopmental syndrome (OMIM: #617062)	1.24E-6	40	40255	0.00099	1.8E-18
									47	34217	0.0014	2.5E-28
									51	57525	0.00089	2.37E-22
1	1	30G; 23	CSNK2A1	Chr 20: 492282	c.593A>G (p.K198R)	-	Okur-Chung neurodevelopmental syndrome (OMIM: #617062)	1.24E-6	42	46138	0.00091	3.6E-18
									31	30919	0.0010	1.4E-13
									22	49535	0.00044	4.5E-02
1	1	30G; 30	CSNK2A1	Chr 20: 492282	c.593A>G (p.K198R)	-	Okur-Chung neurodevelopmental syndrome (OMIM: #617062)	1.24E-6	39	48008	0.00081	6.1E-15
									29	37113	0.00078	1.0E-9
									46	65196	0.00071	6.8E-16
1	2[1]	30G; 11	FGFR2	Chr 10: 121520160	c.758C>G (p.P253R)	20 (20, 25) Endometrium; Breast; Prostate	Apert syndrome (OMIM: #101200)	-	112	62699	0.0018	1.6E-165
									75	58117	0.0013	3.4E-99
									97	68629	0.0014	3.9E-133
1	3[1]	9E; 74	FGFR3	Chr 4: 1804372	c.1118A>G (p.Y373C)	488 (488, 489) Bladder; Skin; Leukemia	Thanatophoric dysplasia type I (OMIM: #187600)	-	160	131642	0.0012	1.5E-63
									66	72146	0.00091	1.8E-17
									136	177870	0.00076	7.9E-32
1	3[1]	9E; 75	FGFR3	Chr 4: 1804372	c.1118A>G (p.Y373C)	488 (488, 489) Bladder; Skin; Leukemia	Thanatophoric dysplasia type I (OMIM: #187600)	-	122	137581	0.00089	3.91E-34
									58	60136	0.00096	5.8E-16
									236	205399	0.0011	1.4E-90
1	4[1]	9E; 7	FGFR3	Chr 4: 1804392	c.1138G>A (p.G380R)	37 (38, 41) Bladder; Nasopharynx; Skin	Achondroplasia (OMIM: #100800)	4.34E-6	130	100932	0.0013	6.5E-38
									94	74649	0.0013	3.2E-27
									141	195182	0.00072	2.0E-17
1	5	33D; 54	FGFR3	Chr 4: 1804435	c.1181C>T (p.T394M)	3 (3, 6) Intestine	-	1.41E-4	67	91531	0.00073	2.2E-05
									48	68196	0.00070	2.8E-02
									160	136836	0.0012	3.22E-40
1	6[1]	30G; 15	FGFR3	Chr 4: 1805636	c.1612A>G (p.I538V)	0 (0, 2) Leukemia; Intestine	Hypochondroplasia (OMIM: #146000)	1.67E-5	74	56116	0.0013	1.8E-33
									78	46951	0.0017	2.2E-42
									109	83863	0.0013	1.6E-49
1	7[1]	9E; 3	FGFR3	Chr 4: 1805644	c.1620C>G (p.N540K)	-	Hypochondroplasia (OMIM: #146000)	1.24E-6	89	45870	0.0019	1.1E-116
									88	46000	0.0019	8.0E-115
									156	80603	0.0019	5.2E-207

1	7[1]	33D; 66	FGFR3	Chr 4: 1805644	c.1620C>G (p.N540K)	-	Hypochondroplasia (OMIM: #146000)	1.24E-6	194	62113	0.0031	4.3E-298
									133	45626	0.0029	2.4E-199
									200	62108	0.0032	4.8E-310
1	8[1]	9E; 20	FGFR3	Chr 4: 1806162	c.1948A>G (p.K650E)	67 (67, 128) Bladder; Skin; Leukemia	Thanatophoric dysplasia type I (OMIM: #187600)	-	38	7353	0.0052	1.8E-32
									77	7999	0.0096	8.3E-90
									115	14588	0.0079	1.3E-124
1	9[1]	33D; 64	HRAS	Chr 11: 533874	c.182A>G (p.Q61R)	414 (524, 1035) Thyroid; Urinary tract; Salivary gland	-	-	264	69582	0.0038	1.6E-152
									130	43653	0.0030	2.8E-61
									289	93030	0.0031	4.6E-145
1	9[1]	33D; 81	HRAS	Chr 11: 533874	c.182A>G (p.Q61R)	414 (524, 1035) Thyroid; Urinary tract; Salivary gland	-	-	92	30745	0.0030	1.7E-42
									94	20045	0.0047	6.5E-60
									184	43387	0.0042	2.6E-113
1	10	9E; 37	PPP1CB	Chr 2: 28776944	c.146C>G (p.P49R)	-	Noonan-like syndrome with loose anagen hair 2 (OMIM: #617506)	-	19	50540	0.00038	6.7E-12
									27	43066	0.00063	9.8E-23
									48	76302	0.00063	8.2E-46
1	10	9E; 38	PPP1CB	Chr 2: 28776944	c.146C>G (p.P49R)	-	Noonan-like syndrome with loose anagen hair 2 (OMIM: #617506)	-	19	45011	0.00042	8.3E-13
									47	63274	0.00074	4.6E-48
									78	104242	0.00075	2.8E-82
1	10	9E; 48	PPP1CB	Chr 2: 28776944	c.146C>G (p.P49R)	-	Noonan-like syndrome with loose anagen hair 2 (OMIM: #617506)	-	22	66591	0.00033	3.3E-13
									36	49299	0.00073	9.8E-36
									36	66264	0.00054	2.8E-31
1	10	9E; 62	PPP1CB	Chr 2: 28776944	c.146C>G (p.P49R)	-	Noonan-like syndrome with loose anagen hair 2 (OMIM: #617506)	-	28	49293	0.00057	4.8E-24
									21	41978	0.00050	5.1E-16
									66	68663	0.00096	4.3E-76
1	10	9E; 71	PPP1CB	Chr 2: 28776944	c.146C>G (p.P49R)	-	Noonan-like syndrome with loose anagen hair 2 (OMIM: #617506)	-	38	56103	0.00068	1.0E-36
									23	42727	0.00054	1.4E-18
									65	82324	0.00079	1.8E-69
1	11	33D; 62	PPP2R5D	Chr 6: 43007271	c.598G>A (p.E200K)	-	Jordan syndrome (OMIM: #616355)	-	543	237158	0.0023	0.0E+00
									322	147823	0.0022	6.3E-210
									526	286122	0.0018	6.2E-310
1	12	9E; 31	PPP2R5D	Chr 6: 43007293	c.620G>C (p.W207S)	-	Jordan syndrome (OMIM: #616355)	-	460	204028	0.0023	0.0E+00
									235	89011	0.0026	0.0E+00
									721	334192	0.0022	0.0E+00

1	13[1]	9E; 78	PTPN11	Chr 12: 112450346	c.166A>G (p.I56V)	0 (0, 2) Breast; Stomach	Noonan syndrome (OMIM: #163950)	1.24E-6	117	52525	0.0022	2.2E-35
									66	37075	0.0018	2.2E-13
									174	70589	0.0025	1.5E-60
1	14	9E; 6	PTPN11	Chr 12: 112450361	c.181G>A (p.D61N)	15 (15, 169) Leukemia, Intestine	Noonan syndrome (OMIM: #163950)	6.20E-7	66	43098	0.0015	6.7E-32
									49	39453	0.0012	7.9E-19
									145	81781	0.0018	1.0E-82
1	15[1]	9E; 53	PTPN11	Chr 12: 112450361	c.181G>T (p.D61Y)	60 (60, 169) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	164	51802	0.0032	8.6E-232
									113	30145	0.0037	8.4E-167
									282	68155	0.0041	0.0E+00
1	16	33D; 6	PTPN11	Chr 12: 112450368	c.188A>G (p.Y63C)	6 (6, 6) Leukemia	Noonan syndrome (OMIM: #163950)	8.68E-6	471	55429	0.0085	0.0E+00
									330	37783	0.0087	0.0E+00
									463	52689	0.0088	0.0E+00
1	16	33D; 26	PTPN11	Chr 12: 112450368	c.188A>G (p.Y63C)	6 (6, 6) Leukemia	Noonan syndrome (OMIM: #163950)	8.68E-6	59	67963	0.00087	7.3E-07
									65	38487	0.0017	3.9E-22
									77	80726	0.00095	2.8E-12
1	17	9E; 9	PTPN11	Chr 12: 112450385	c.205G>A (p.E69K)	64 (64, 70) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	55	44265	0.0012	1.4E-20
									28	25313	0.0011	2.3E-07
									82	61394	0.0013	1.0E-34
1	18[1]	9E; 23	PTPN11	Chr 12: 112450391	c.211T>C (p.F71L)	13 (29, 30) Leukemia; Intestine	Noonan syndrome (OMIM: #163950)	6.20E-7	123	50818	0.0024	2.6E-126
									111	34047	0.0033	1.4E-127
									113	55215	0.0020	8.4E-108
1	19	9E; 22	PTPN11	Chr 12: 112450394	c.214G>A (p.A72T)	68 (68, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	115	57070	0.0020	5.9E-54
									64	38534	0.0017	9.0E-24
									116	59015	0.0020	2.3E-53
1	19	9E; 34	PTPN11	Chr 12: 112450394	c.214G>A (p.A72T)	68 (68, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	372	50076	0.0074	0.0E+00
									266	35607	0.0075	5.7E-270
									545	70769	0.0077	0.0E+00
1	19	9E; 35	PTPN11	Chr 12: 112450394	c.214G>A (p.A72T)	68 (68, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	128	56438	0.0023	3.2E-66
									67	29798	0.0022	1.1E-32
									101	50625	0.0020	1.5E-46
1	20[1]	9E; 51	PTPN11	Chr 12: 112450395	c.215C>G (p.A72G)	4 (4, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	55	61689	0.00090	3.6E-10
									51	32009	0.0016	1.9E-19
									65	67012	0.00097	1.7E-14

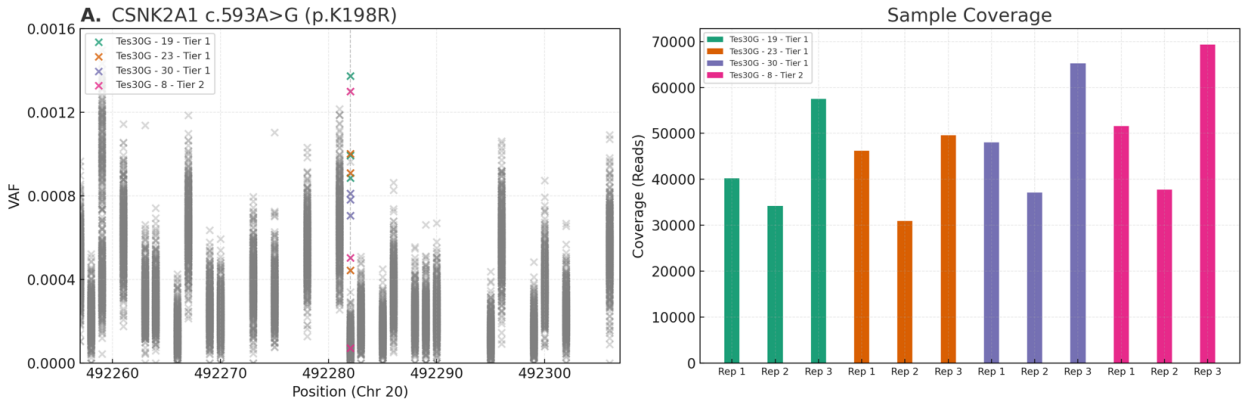
1	20[1]	9E; 52	PTPN11	Chr 12: 112450395	c.215C>G (p.A72G)	4 (4, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	173	54050	0.0032	1.1E-121
									65	25479	0.0026	1.5E-37
									274	76480	0.0036	2.9E-207
1	20[1]	9E; 53	PTPN11	Chr 12: 112450395	c.215C>G (p.A72G)	4 (4, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	384	52743	0.0073	0.0E+00
									162	30874	0.0052	1.6E-146
									406	68789	0.0059	0.0E+00
1	20[1]	9E; 54	PTPN11	Chr 12: 112450395	c.215C>G (p.A72G)	4 (4, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	48	57442	0.00084	2.1E-07
									41	32876	0.0012	3.1E-11
									62	71551	0.00087	2.1E-11
1	20[1]	9E; 55	PTPN11	Chr 12: 112450395	c.215C>G (p.A72G)	4 (4, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	343	44946	0.0076	0.0E+00
									189	24616	0.0077	1.1E-201
									461	63270	0.0073	0.0E+00
1	20[1]	9E; 56	PTPN11	Chr 12: 112450395	c.215C>G (p.A72G)	4 (4, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	61	77695	0.00079	2.8E-09
									26	28224	0.00089	2.1E-02
									98	88748	0.0011	4.8E-28
1	20[1]	9E; 57	PTPN11	Chr 12: 112450395	c.215C>G (p.A72G)	4 (4, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	60	47699	0.0013	2.0E-18
									55	28785	0.0019	5.0E-25
									100	62121	0.0016	3.9E-42
1	20[1]	9E; 58	PTPN11	Chr 12: 112450395	c.215C>G (p.A72G)	4 (4, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	103	43129	0.0024	5.2E-59
									67	35829	0.0019	9.1E-31
									97	65000	0.0015	4.7E-38
1	20[1]	9E; 62	PTPN11	Chr 12: 112450395	c.215C>G (p.A72G)	4 (4, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	120	38920	0.0031	1.7E-81
									69	27677	0.0025	1.8E-39
									127	54175	0.0023	1.1E-72
1	20[1]	9E; 63	PTPN11	Chr 12: 112450395	c.215C>G (p.A72G)	4 (4, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	123	33518	0.0037	2.7E-92
									102	32674	0.0031	3.0E-69
									193	49956	0.0039	6.9E-151
1	20[1]	9E; 64	PTPN11	Chr 12: 112450395	c.215C>G (p.A72G)	4 (4, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	99	35895	0.0028	3.4E-62
									54	28396	0.0019	2.1E-24
									132	143805	0.0027	1.0E-83
1	20[1]	9E; 65	PTPN11	Chr 12: 112450395	c.215C>G (p.A72G)	4 (4, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	64	26547	0.0024	1.7E-35
									65	28754	0.0023	1.8E-34
									103	47358	0.0022	2.9E-55

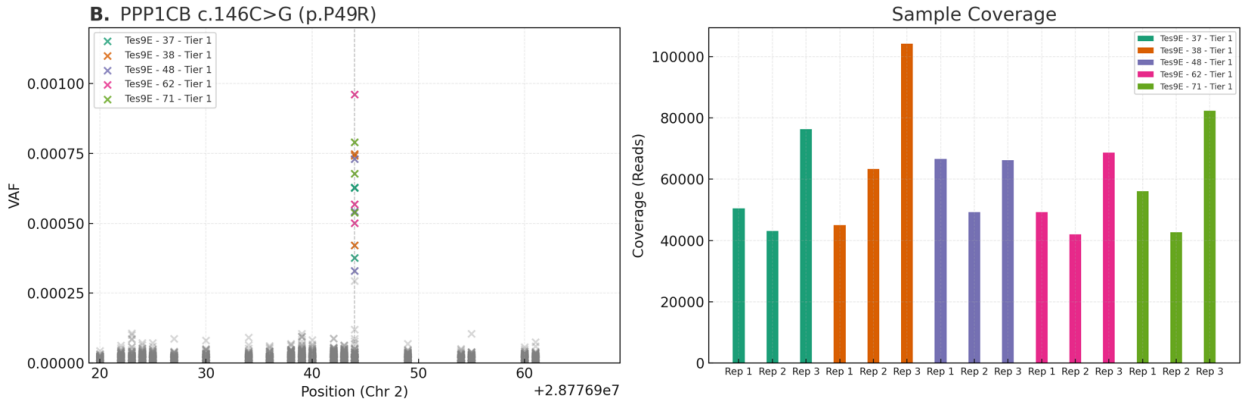
1	20[1]	9E; 66	PTPN11	Chr 12: 112450395	c.215C>G (p.A72G)	4 (4, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	128	45472	0.0028	1.6E-82
									106	30266	0.0035	5.0E-77
									236	74501	0.0032	2.7E-166
1	21	9E; 68	PTPN11	Chr 12: 112450408	c.228G>C (p.E76D)	0 (0, 293) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	1.24E-6	81	46139	0.0018	2.1E-118
									53	30062	0.0018	2.8E-76
									157	69098	0.0023	2.6E-250
1	22	9E; 30	PTPN11	Chr 12: 112450416	c.236A>G (p.Q79R)	6 (6, 7) Leukemia	Noonan syndrome (OMIM: #163950)	1.24E-6	83	63105	0.0013	1.8E-10
									59	41980	0.0014	4.9E-14
									129	81090	0.0016	1.3E-40
1	23	30G; 48	PTPN11	Chr 12: 112477651	c.854T>C (p.F285S)	14 (14, 15) Leukemia; Brain; Breast	Noonan syndrome (OMIM: #163950)	-	170	47611	0.0036	5.2E-123
									57	18153	0.0031	6.6E-36
									153	37526	0.0041	1.3E-118
1	24[1]	30G; 15	PTPN11	Chr 12: 112477719	c.922A>G (p.N308D)	9 (9, 16) Leukemia; Brain; Prostate	Noonan syndrome (OMIM: #163950)	7.44E-6	84	95009	0.00088	2.0E-15
									30	31784	0.00094	1.2E-03
									56	52997	0.0011	3.2E-12
1	24[1]	9E; 43	PTPN11	Chr 12: 112477719	c.922A>G (p.N308D)	9 (9, 16) Leukemia; Brain; Prostate	Noonan syndrome (OMIM: #163950)	7.44E-6	27	17278	0.0016	1.4E-07
									49	20695	0.0024	1.3E-23
									86	32511	0.0026	1.2E-49
1	25	9E; 67	PTPN11	Chr 12: 112477720	c.923A>C (p.N308T)	3 (3, 16) Leukemia; Liver; Brain	Noonan syndrome (OMIM: #163950)	-	38	21856	0.0017	5.1E-42
									25	12883	0.0019	1.6E-27
									50	31963	0.0016	3.2E-54
1	26[1]	30G; 33	RET	Chr 10: 43121968	c.2753T>C (p.M918T)	514 (514, 891) Thyroid; Adrenal gland; Breast	Multiple endocrine neoplasia, type IIB (OMIM: #162300)	-	93	61001	0.0015	5.4E-17
									91	40407	0.0023	3.2E-28
									205	83539	0.0025	2.4E-74
1	27	33D; 25	RET	Chr 10: 43121969	c.2754G>C (p.Met918Ile)	-	-	1.24E-6	77	65869	0.0012	1.9E-120
									41	36666	0.0011	1.4E-61
									105	94783	0.0011	3.9E-163
1	28	30G; 32	RET	Chr 10: 43121970	c.2755G>C (p.A919P)	0 (0, 4) Cervix; Lung; Intestine; Thyroid	Multiple endocrine neoplasia, type IIB (OMIM: #162300)	2.48E-6	83	46240	0.0018	1.2E-128
									43	31003	0.0014	4.4E-60
									63	62738	0.0010	4.5E-81
1	29	30G; 33	SMAD4	Chr 18: 51078294	c.1486C>T (p.R496C)	7 (7, 30) Intestine, Skin	Myhre syndrome (OMIM: #139210)	4.34E-6	73	60392	0.0012	7.1E-15
									62	39598	0.0016	2.0E-17
									165	78520	0.0021	1.0E-69

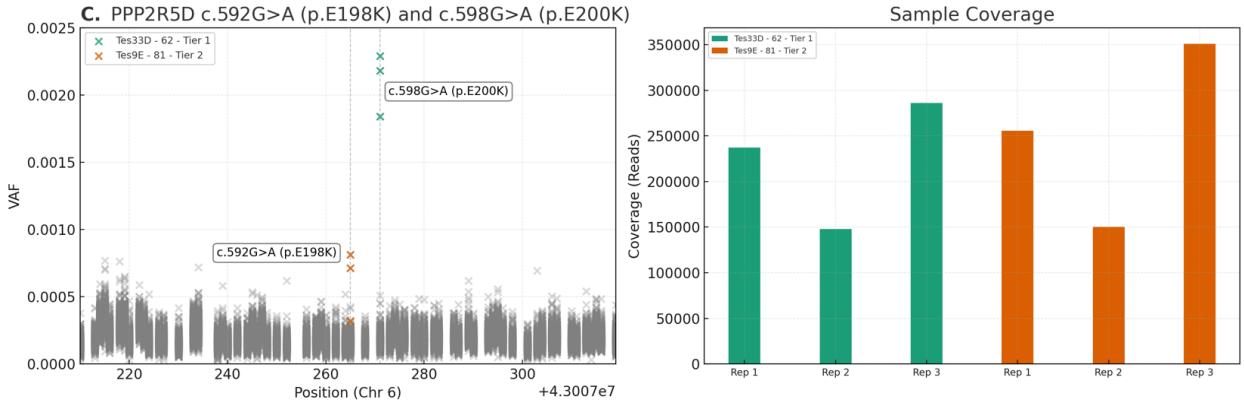
1	29	30G; 47	SMAD4	Chr 18: 51078294	c.1486C>T (p.R496C)	7 (7, 30) Intestine, Skin	Myhre syndrome (OMIM: #139210)	4.34E-6	103	34179	0.0030	2.9E-56
									119	28688	0.0041	1.4E-80
									100	24098	0.0041	3.7E-67
1	30	30G; 44	SMAD4	Chr 18: 51078306	c.1498A>G (p.I500V)	-	Myhre syndrome (OMIM: #139210)	6.19E-7	44	63034	0.00070	6.3E-05
									32	41227	0.00078	1.5E-03
									54	75865	0.00071	3.1E-07
2	31	33D; 47	ADNP	Chr 20: 50892501	c.2213C>G (p.S738*)	-	Helsmoortel-Van der Aa syndrome (OMIM: #615873)	-	15	72625	0.00021	8.6E-07
									15	57858	0.00026	3.6E-08
									16	120185	0.00013	9.7E-05
2	1	30G; 8	CSNK2A1	Chr 20: 492282	c.593A>G (p.K198R)	-	Okur-Chung neurodevelopmental syndrome (OMIM: #617062)	1.24E-6	26	51606	0.00050	1.3E-09
									49	37721	0.00130	6.1E-34
									5	69344	0.000072	NS
2	32	30G; 31	CSNK2A1	Chr 20: 492300	c.575A>C (p.V192G)	-	-	-	33	37339	0.00088	4.3E-44
									31	28056	0.0011	1.8E-44
									5	46601	0.00011	NS
2	33[1]	30G; 45	FGFR2	Chr 10: 121520163	c.755C>T (p.S252L)	8 (8, 118) Endometrium; Breast; Ovarian; Cervical	Crouzon/Pfeiffer syndromes (OMIM: #123500 and #101600)	4.03E-5	76	91489	0.00083	1.8E-07
									32	35083	0.00091	NS
									65	80347	0.00081	2.3E-05
2	34[1]	33D; 18	FGFR2	Chr 10: 121520163	c.755C>G (p.S252W)	8 (8, 118) Endometrium; Breast; Ovarian; Cervical	Apert syndrome (OMIM: #101200)	4.03E-5	43	88772	0.00048	2.7E-31
									21	45725	0.00046	7.9E-13
									37	93237	0.00040	2.1E-23
2	4[1]	9E; 9	FGFR3	Chr 4: 1804392	c.1138G>A (p.G380R)	37 (38, 41) Bladder; Nasopharynx; Skin	Achondroplasia (OMIM: #100800)	3.72E-6	185	126736	0.0015	1.3E-55
									46	75208	0.00061	NS
									102	164762	0.00062	4.2E-04
2	35	30G; 12	FGFR3	Chr 4: 1804427	c.1173G>A (p.A391A)	3 (3, 42) Bladder, Skin	-	3.72E-6	80	123315	0.00065	2.4E-02
									40	96472	0.00041	NS
									93	149079	0.00062	1.4E-02
2	9[1]	33D; 78	HRAS	Chr 11: 533874	c.182A>G (p.Q61R)	414 (524, 1035) Thyroid; Urinary tract; Salivary gland	-	-	69	55161	0.0013	2.9E-10
									25	36215	0.00069	NS
									76	77074	0.00099	1.3E-06
2	36	33D; 27	PACS1	Chr 11: 66211206	c.607C>T (p.R203W)	4 (4, 5)	Schuurs-Hoeijmakers syndrome	6.20E-7	47	54188	0.00087	5.7E-05
									29	29936	0.00097	1.6E-02

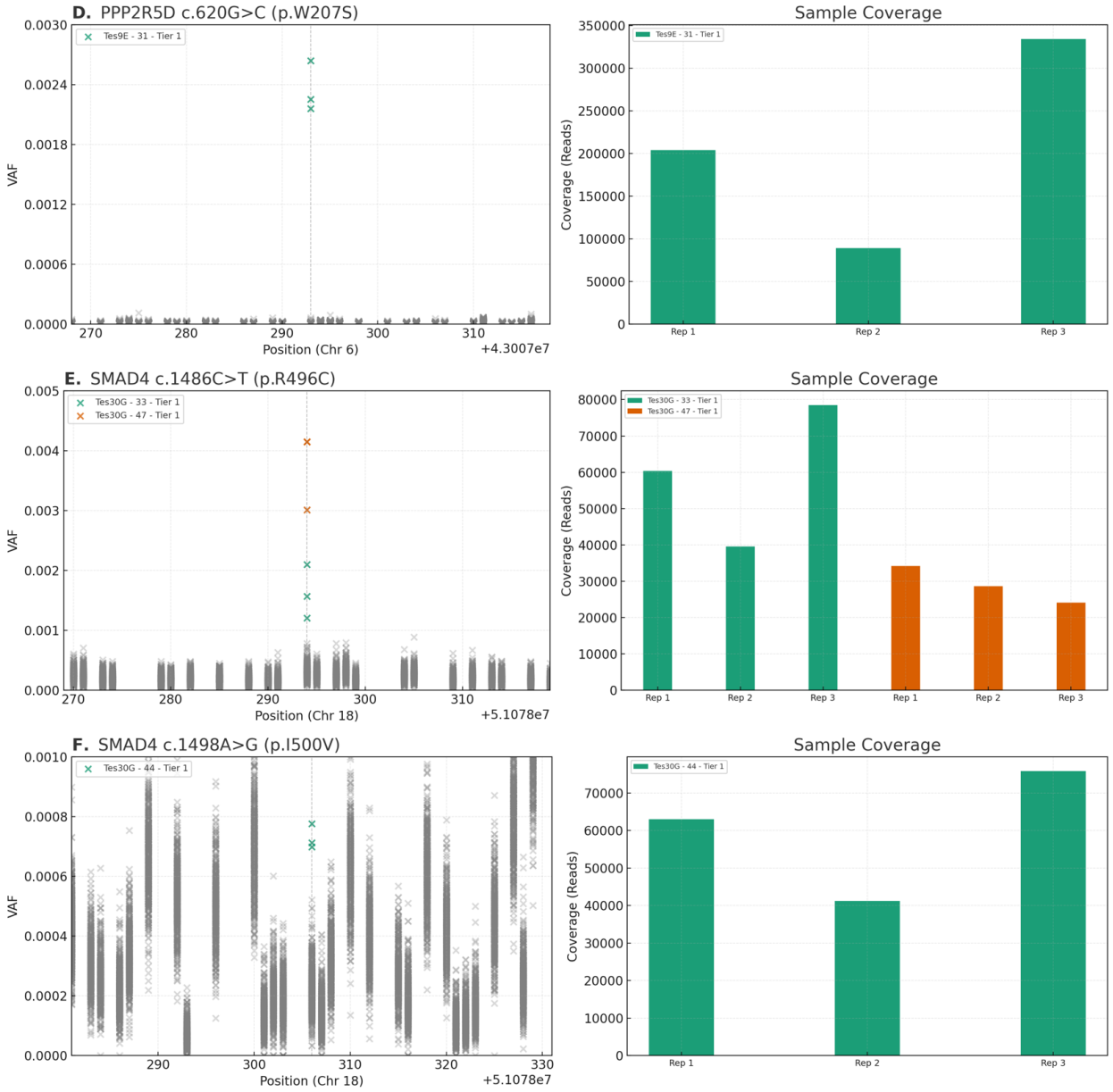
						Skin; Breast; Ovary	(OMIM: #615009		32	60847	0.00053	NS
2	37	9E; 81	PPP2R5D	Chr 6: 43007265	c.592G>A (p.E198K)	3 (3, 3) Lung, Head and neck	Jordan syndrome (OMIM: #616355)	-	208	255606	0.00081	4.2E-65
									49	150332	0.00032	NS
									250	350980	0.00071	8.7E-68
2	15	9E; 46	PTPN11	Chr 12: 112450361	c.181G>T (p.D61Y)	60 (60, 169) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	27	54668	0.00049	4.7E-14
									30	49073	0.00061	1.3E-18
									24	71966	0.00033	2.2E-08
2	16[1]	30G; 12	PTPN11	Chr 12: 112450368	c.188A>G (p.Y63C)	6 (6, 6) Leukemia	Noonan syndrome (OMIM: #163950)	8.68E-6	84	81876	0.0010	1.3E-15
									33	58294	0.00057	NS
									95	79734	0.0012	1.1E-22
2	16[1]	9E; 61	PTPN11	Chr 12: 112450368	c.188A>G (p.Y63C)	6 (6, 6) Leukemia	Noonan syndrome (OMIM: #163950)	8.68E-6	30	22423	0.0013	1.2E-05
									36	26942	0.0013	1.4E-07
									25	55177	0.00045	NS
2	19	33D; 13	PTPN11	Chr 12: 112450394	c.214G>A (p.A72T)	68 (68, 196) Leukemia; Brain	Noonan syndrome (OMIM: #163950)	-	69	82289	0.00084	4.2E-10
									35	47098	0.00074	NS
									63	92936	0.00068	4.3E-05
2	23	30G; 42	PTPN11	Chr 12: 112477651	c.854T>C (p.F285S)	14 (14, 15) Leukemia; Brain; Breast	Noonan syndrome (OMIM: #163950)	-	49	29191	0.0017	1.8E-18
									31	19732	0.0016	1.6E-09
									10	34008	0.00029	NS

[1] Variants that have been previously documented as clonal expansions in the testis or sperm. Relevant references: Choi et al., 2008, Maher et al., 2016, Shinde et al., 2013, Maher et al., 2018; Goriely et al., 2009, Eboreime et al., 2022, Yoon et al., 2013, Choi et al., 2012.


A. CSNK2A1 c.593A>G (p.K198R)
Tes30G - 19 - Tier 1
Tes30G - 23 - Tier 1
Tes30G - 30 - Tier 1
Tes30G - 8 - Tier 2
VAF
Position (Chr 20)
Sample Coverage
Coverage (Reads)
Rep 1
Rep 2
Rep 3


B. PPP1CB c.146C>G (p.P49R)
Tes9E - 37 - Tier 1
Tes9E - 38 - Tier 1
Tes9E - 48 - Tier 1
Tes9E - 62 - Tier 1
Tes9E - 71 - Tier 1
VAF
Position (Chr 2)
+2.87769e7
Sample Coverage
Coverage (Reads)


C. PPP2R5D c.592G>A (p.E198K) and c.598G>A (p.E200K)
Tes33D - 62 - Tier 1
Tes9E - 81 - Tier 2
c.598G>A (p.E200K)
c.592G>A (p.E198K)
VAF
Position (Chr 6)
+4.3007e7
Sample Coverage
Coverage (Reads)


D. PPP2R5D c.620G>C (p.W207S)
Tes9E - 31 - Tier 1
VAF
Position (Chr 6)
+4.3007e7
Sample Coverage
Coverage (Reads)
Rep 1
Rep 2
Rep 3
E. SMAD4 c.1486C>T (p.R496C)
Tes30G - 33 - Tier 1
Tes30G - 47 - Tier 1
Position (Chr 18)
+5.1078e7
F. SMAD4 c.1498A>G (p.I500V)
Tes30G - 44 - Tier 1

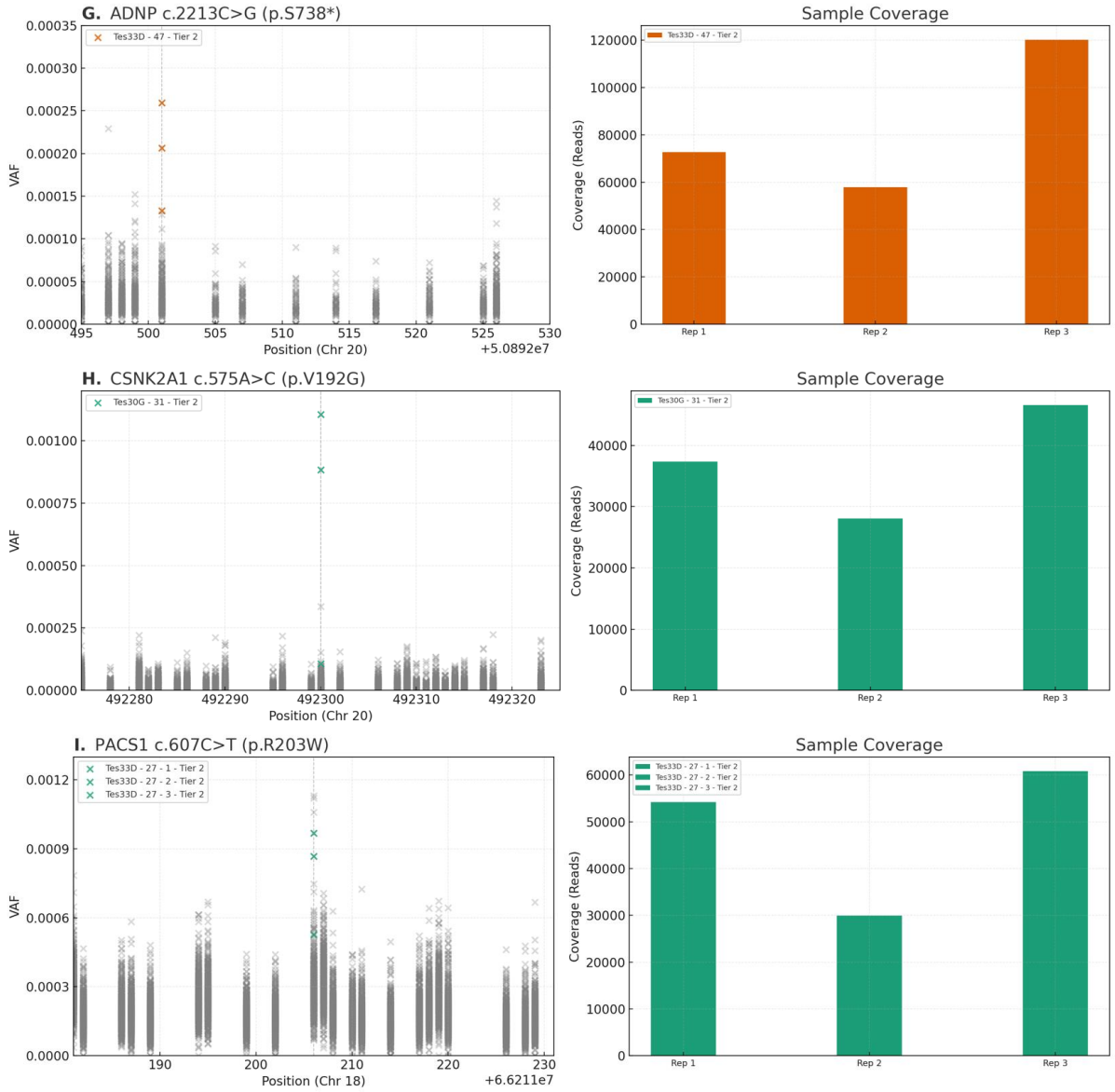
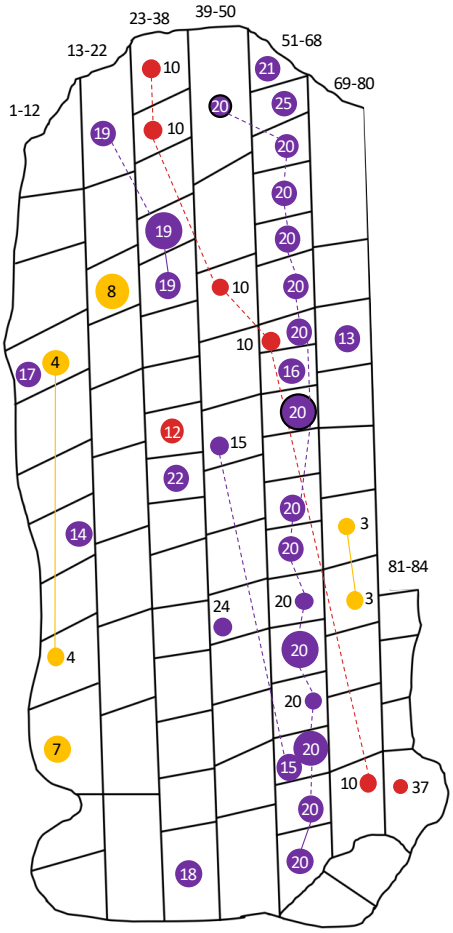


Figure 4.16: Mutation plots of variants in candidate selfish genes identified in the rhAmpSeq discovery screen. For each variant, the VAF of all three replicates is plotted with colours representing replicates from the same biopsy. Label legends indicate which biopsies are called in which tiers. Bar plots indicating individual replicate coverage at the position are included.

A | Tes9E (73 years)



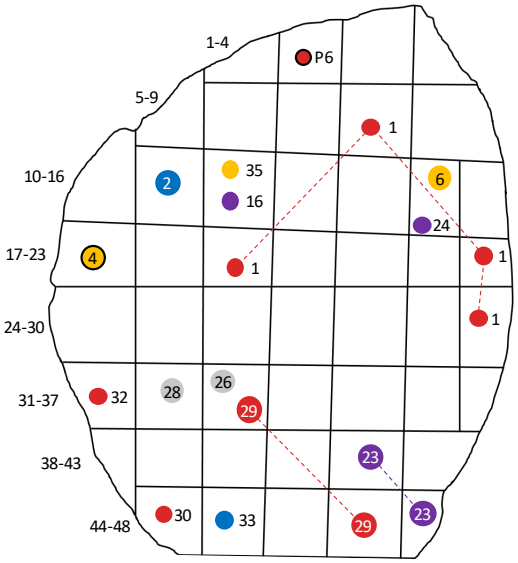
- **2** *FGFR2* c.758C>G (p.P253R)
- **33** *FGFR2* c.755C>T (p.S252L)
- **34** *FGFR2* c.755C>G (p.S252W)
- **3** *FGFR3* c.1118A>G (p.Y373C)
- **4** *FGFR3* c.1138G>A (p.G380R)
- **5** *FGFR3* c.1181C>T (p.T394M)
- **6** *FGFR3* c.1612A>G (p.I538V)
- **7** *FGFR3* c.1620C>A (p.N540K)
- **8** *FGFR3* c.1948A>G (p.K650E)
- **35** *FGFR3* c.1173G>A (p.A391A)
- **9** *HRAS* c.182A>G (p.Q61R)
- **13** *PTPN11* c.166A>G (p.I56V)
- **14** *PTPN11* c.181G>A (p.D61N)
- **P4** *PTPN11* c.181G>C (p.D61H)
- **15** *PTPN11* c.181G>T (p.D61Y)
- **16** *PTPN11* c.188A>G (p.Y63C)
- **17** *PTPN11* c.205G>A (p.E69K)
- **18** *PTPN11* c.211T>C (p.F71L)
- **19** *PTPN11* c.214G>A (p.A72T)
- **20** *PTPN11* c.215C>G (p.A72G)
- **21** *PTPN11* c.228G>C (p.E76D)
- **22** *PTPN11* c.236A>G (p.Q79R)
- **23** *PTPN11* c.854T>C (p.F285S)
- **24** *PTPN11* c.922A>G (p.N308D)
- **25** *PTPN11* c.923A>C (p.N308T)
- **26** *RET* c.2753T>C (p.M918T)
- **27** *RET* c.2754G>A/C (p.M918I)
- **28** - *RET* c.2755G>C (p.A919P)
- **1** *CSNK2A1* c.593A>G (p.K198R)
- **P6** *CSNK2A1* c.521T>C (p.I174R)
- **32** CSNK2A1 c.575T>C (p.V192G)
- **10** *PPP1CB* c.149C>G (p.P49R)
- **11** *PPP2R5D* c.598G>A (p.E200K)
- **12** *PPP2R5D* c.620G>C (p.W207S)
- **P2** *PPP2R5D* c.627A>C (p.H209P)
- **37** *PPP2R5D* c.592G>A (p.E198K)
- **29** *SMAD4* c.1486C>T (p.R496C)
- **30** *SMAD4* c.1498A>G (p.I500V)
- **31** *ADNP* c.2213C>G (p.S738*)
- **36** *PACS1* c.607C>T (p.R203W)

<0.001 VAF

0.001–0.005 VAF

>0.005 VAF

B | Tes30G (63 years)



C | Tes33D (76 years)

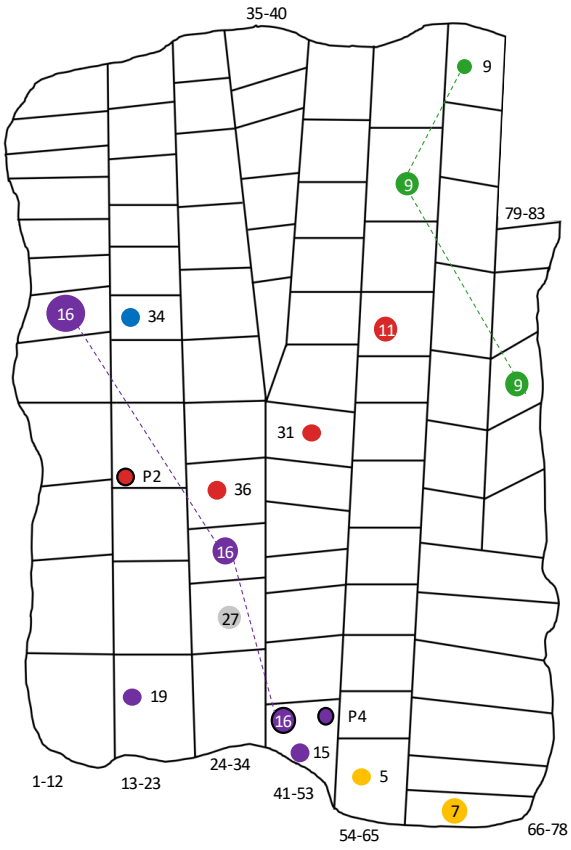


Figure 4.17: Spatial distribution of all variants identified in a rhAmpSeq discovery screen of testes biopsies from three aged individuals. The sections are labeled for each testes slice **(A)** Tes9E (73 years; 84 biopsies), **(B)** Tes30G (63 years; 48 biopsies), **(C)** Tes33D (76 years; 83 biopsies) which together comprise 215 total biopsies. Biopsy numbers can be found adjacent to corresponding rows or columns, and each variant is numbered according to **Table 4.4**. Variants that were detected in the pilot study are outlined in black, and are numbered according to the variant number in the full discovery screen (e.g. P4 is variant number 4 in **Table 4.2**). Variants that were detected in both the discovery screen are underlined and pilot study are numbered according to **Table 4.4**, and those detected only in the pilot study are numbered according to **Table 4.2** with the "P" prefix. Colours denote the associated gene in which they were detected, and the circle size represents the VAF range, as detailed in the legend. Identical variants in neighbouring biopsies are linked by a dotted line which suggest singular clonal events, possibly tracking along a seminiferous tubule, although the exact trajectory is speculative.

4.3.2.5 – Spatial distribution of variants in the testis

Mutations under positive selection are anticipated to display a highly clustered distribution (i.e. mutation foci) and the detection of the same variant in contiguous or neighbouring biopsies suggests clonal expansion along the length of the seminiferous tubule in which the SSC arose (Choi et al., 2008; Maher et al., 2016; Maher et al., 2018).

In this discovery screen, 10 distinct variants in known selfish genes were observed in more than one biopsy. The most frequently observed variant was *PTPN11* c.215C>G (p.A72G), detected in 13 largely adjoining biopsies (14 including variant P5 in the pilot study; Table 4.2) in one testis slice (Tes9E; 73 years) (range 0.00093–0.0075 VAF), consistent with clonal expansion along the length of a tubule (variant 20, Figure 4.17). The *PTPN11* c.215C>G (p.A72G) variant is a recurrently observed germline mutation in Noonan syndrome (Tartaglia and Gelb, 2005) and has been detected as clonal expansions in the testis in prior studies (Maher et al., 2018; Eboreime et al., 2022). In addition to variants reported in constitutional disorders, variants with established associations with cancer were also detected in multiple biopsies. For example, *HRAS* c.182A>G (p.Q61R) was detected in 3 neighbouring biopsies in a separate testis slice (Tes33D; 76 years), two of which were prioritized as Tier 1 variants (range 0.00092–0.0040 VAF) (variant 9, **Figure**

4.17). While germline mutations in *HRAS* are known to underlie Costello syndrome, substitutions the p.Q61 codon have not been reported. Instead, mutations at p.Q61 (including p.Q61R) are frequently reported in cancer with >1000 observations in COSMIC v101 and have been identified in spermatocytic tumours, rare tumours that present in older men (Goriely et al., 2009; **section 1.7.2**). Because the HRAS p.Q61R mutation exhibits high transformation activity *in vitro* (Der et al., 1986) but has not been reported in Costello syndrome, suggests that these mutations are too disruptive for normal embryonic development (Giannoulatou et al., 2013).

3 distinct variants in candidate genes were observed in more than one biopsy from the same testis slice. The most frequent candidate mutation was *PPP1CB* c.146C>G (p.P49R), which was prioritized as a Tier 1 variant in 5 biopsies (range 0.00053–0.00068 VAF) from the same testis slice (Tes9; 73 years). Variants occurred in largely adjacent biopsies, suggestive of a single clonal event (variant 10; **Figure 4.17**). *PPP1CB* has a known role within the RTK-Ras pathway and the c.146C>G (p.P49R) variant is the most recurrently observed mutation in Noonan-like syndrome with loose anagen hair 2 (**section 3.3.8**) – clonal expansions in this gene have not been previously detected. Secondly, the *CSNK2A1* c.593A>G (p.K198R) variant was also detected in three biopsies as a Tier 1 variant and in one biopsy as a Tier 2 variant, all within the same testis slice (Tes30G; 63 years) (variant 1, **Table 4**). Variants were detected in nearby biopsies and VAFs did not differ by more than ~1.4-fold (range 0.00077–0.0011 VAF), suggestive of clonal expansion within a tubule. The CSNK2A1 c.593A>G (p.K198R) variant is the most recurrent mutation in Okur-Chung neurodevelopmental syndrome (**section 3.3.10**) and variants in this gene have not been previously detected in the male germline. Lastly, the *SMAD4* c.1486C>T (p.R496C) variant was detected prioritized twice in Tier 1 (range 0.0016–0.0038 VAF) and was detected in two neighbouring biopsies in Tes30G (63 years), likely deriving from the same clone. The c.1486C>T

(p.R496C) variant is a recurrent mutation in Myhre syndrome (**section 3.3.3**), and this particular mutation has been previously detected in testis biopsies (Ralph, 2018).

7 biopsies across the 3 testes slices harboured more than 1 variant. As each testis biopsy contains multiple seminiferous tubules, variants in these biopsies could represent competing mutational events within the same tubule, or variants in tubules that are closely packed together.

4.4 – Discussion

In this chapter, rhAmpSeq was adapted to investigate candidate selfish mutations in the aged male germline. To provide an empirical estimate of the limit of detection in an unbiased manner, triplicate samples carrying 13 serially diluted mutations were screened alongside testes biopsies and analyzed with a custom statistical prioritization pipeline. Using this approach, rhAmpSeq successfully identified most spiked-in mutations down to ~0.0005 VAF, although sensitivity varied with assay conditions. Apart from one spiked-in mutation that was ~6-fold overestimated (*FGFR2* c.758C>G), observed mutation levels across triplicates were within ~0.5 to 3.3-fold of expectations.

In the pilot study, three testis biopsies carrying 3 known selfish mutations at ~0.0015–0.0047 VAF were also included to confirm the assay's accuracy and sensitivity under "real-world" conditions. All three control variants were recovered within ~1.0–1.6-fold of previously reported values, reinforcing the reliability of rhAmpSeq. The pilot study also showed that assay performance is highly dependent on achieving adequate coverage. For example, the spiked-in *SMAD4* mutations were reliably detected down to ~0.0005 VAF in the discovery screen (median coverage 65,129x, range 37,238–89,517x). However, in the pilot study (median coverage 22,030x, range 1849–110,474x), these same mutations were only called at ~0.001 VAF. While other factors,

such as the robustness of the error model across different sample sizes, may affect sensitivity, achieving sufficient coverage depth is key for reliable detection of low-frequency variants. Nevertheless, several variants detected in the pilot study were also detected in adjacent biopsies in the discovery screen, reinforcing the robustness of the approach and the ability of the assay to reproducibly capture low-frequency selfish mutations across independent experiments. The pilot study also allowed for the optimization of panel balance and confirmed the suitability of the tissues before conducting the larger-scale screen.

However, even at well-covered positions, sensitivity was not uniform. For example, although the *RET* c.2753 locus was covered at a median of 58,238x (range 31,591–82,330x) in the discovery screen, the c.2753T>C mutation was only detected at 0.00125 VAF. This finding aligns with known NGS error profiles, where A>G/T>C exhibit 5–10 times higher error rates than transversions (Ma et al., 2019). Despite these nuances, the LOD for the discovery screen was set at ~0.0005 VAF for both transitions and transversion mutations. This was based on observations that 2 of 2 well-covered spiked-in mutations – excluding those in the FGFR3.3 amplicon, which had lower median coverage (~2848x) – were detected at the lowest dilution levels. In testes biopsies, three transversions were detected below this threshold; two were retained as Tier 2 variants: *FGFR2* c.755C>G (p.S252W) at ~0.00047 VAF (Apert syndrome) and *ADNP* c.2213C>G (p.S738*) at ~0.0002 VAF (Helsmoortel-Van der Aa syndrome) (variants 31 and 34, **Table 4.5**). One Tier 2 variant, *RET* c.2755G>T (p.A919V) at ~0.00022 VAF in Tes33D-25, was excluded by the LOD filter despite having been previously observed as a somatic mutation in medullary thyroid carcinoma, a common feature of MEN2A/B (Dvořaková et al., 2006). This underscores a limitation in the study design: only a subset of the 13 spiked-in mutations were

transversions diluted to ~0.0005–0.000625 VAF, so the LOD for these specific substitutions may be incompletely characterized.

Although this chapter included internal quantitative controls (i.e., serial dilutions) in each experiment and applied stringent calling criteria based on statistical significance across independent technical replicates, the discrepancy between these replicate-based results and the assay's theoretical sensitivity points to another study-design limitation. Specifically, validating variant calls – both Tier 1 and Tier 2, as well as those filtered after establishing the limit of detection – via an orthogonal approach such as droplet digital PCR would have strengthened the confidence in these findings and provided a clearer view of the false-positive rate.

An initial hypothesis was that incorporating UMIs would enhance the sensitivity of rhAmpSeq. A proof-of-concept experiment showed that error rates with UMIs were significantly lower than raw error rates. Despite this, UMIs ultimately provided limited benefit for detecting rare variants in two serial dilutions and in known selfish mutations in control biopsies. Two key factors likely contributed. First, in the pilot study, effective coverage was low (median 5725x per sample), likely due to coverage skew. Analysis therefore focused on positions with high effective coverage such as *HRAS* c.182 and c.183 (median ~16,858x, range 9354–38,663x). While uncorrected reads called both dilutions at ~0.000625 VAF, UMI-corrected only called them at ~0.00125 VAF. Similar patterns were observed in the proof-of-concept, where the median coverage was ~22,370x for positions of interest. Only 5 of 9 dilutions at 0.00125 VAF were detected with UMI correction compared to 8 of 9 using uncorrected reads. For three specific mutations – *PTPN11* c.1520A>G (p.M504V), *SMAD4* c.1498A>G (p.I500V) and c.1499T>C (p.I500T) – mutant reads were lost altogether following UMI correction (**Figure 4.5**). Second, the degenerate design of the UMIs reduced overall complexity from ~1.05 million to ~59,000 unique

molecular tags. With 8 cycles of PCR1 from a starting material of ~30,000 genomes per sample, there are upwards of ~7.7 million total molecules. It reasons that many different original templates must share the same UMI, including those tagging low frequency variants. In those cases, UMI correction may have resulted in the loss of true mutant reads by merging them with wild-type reads, thereby diluting signal. Future rhAmpSeq studies incorporating degenerate UMIs are therefore recommended to adopt a minimum UMI length of 8 bp (3^{16} ~= 43 million tags) to ensure sufficient complexity.

Despite these technical limitations, rhAmpSeq enabled the identification of 71 variants (37 distinct) across 11 genes in a discovery screen of 186 testes biopsies from aged men (average 70.7 years, range 63–76 years). Most (27 of 37) occurred in established selfish genes and corresponded to known pathogenic mutations implicated in dominant disorders and cancer, many of which have been previously observed as clonal expansions in the male germline. Furthermore, across all three testes slices, there was strong evidence that variants in selfish genes exhibited behaviour consistent with positive selection (i.e., clonal expansions spanning multiple adjacent biopsies). These data reinforce the methodology applied here and support rhAmpSeq as a reliable tool for discovering variants under positive selection.

There were 10 distinct variants detected across 6 candidate genes in the discovery screen. Among these were 3 variants in 3 genes whose spatial distributions were consistent with clonal expansion along the length of a seminiferous tubule. One was the PPP1CB p.P49R variant, detected in five neighbouring biopsies in Tes9E (73 years), which is known to act in a gain-of-function manner by enhancing RTK-Ras signalling (Hauseman et al., 2022; **section 3.3.8**), consistent with the established mechanism by which selfish mutations confer a fitness advantage to spermatogonia. The other two candidate DNMs, located in *CSNK2A1* and *SMAD4*, notably do

not encode members of the RTK-Ras pathway. The *CSNK2A1* c.593A>G (p.K198R) and *SMAD4* c.1486C>T (p.R496C) variants were each detected in 4 and 2 neighbouring biopsies, respectively, from the same man Tes30G (63 years). *CSNK2A1* encodes the catalytic subunit of a ubiquitous serine/threonine kinase with critical roles in signalling pathways involved in SSC maintenance (e.g. ERK and AKT). *In vivo* modelling indicates that the p.K198R mutation activates AKT (Plotnikkov et al., 2019; Cruz-Gamero et al., 2014; **section 3.3.10**), suggesting that even though it is not in the RTK-Ras pathway, its effect may still be mediated by RTK-Ras dysregulation. In contrast, functional studies of SMAD4 p.R496C patient fibroblasts have reported elevated levels of SMAD4 and increased pSMAD2, indicative of constitutive TGF-β signalling (Piccolo et al., 2014; **section 3.3.3**). *SMAD4*-family genes are upregulated in human SSCs by scRNAseq (Guo J et al., 2017) and SMAD4 co-expression with both MAGEA4 and GFRA1 at the protein-level has been demonstrated (**Figure 3.3**). Although 2 of these 3 mutations have not been previously detected the human testis, their plausible mechanisms of action and clustered spatial distributions support the possibility genes outside of the canonical RTK-Ras pathway may also drive positive selection.

The remaining 7 distinct variants in candidate genes each occurred in only a single biopsy, complicating their interpretation. Although many selfish variants have also been observed in isolation, one important example is the *FGFR3* c.1173G>A (p.A391A) variant (~0.00056 VAF) identified in Tes30G-35. This variant is synonymous and is unlikely to be functional; it could be a passenger to the co-occuring *PTPN11* c.188A>G (p.Y63C) variant at ~0.00093 VAF in the same biopsy. By this same logic, it is possible that distinct variants in *ADNP*, *CSNK2A1*, *SMAD, PACS1*, and *PPP2R5D* also represent drift.

However, for several variants, there is evidence that argues against a purely stochastic explanation. First, for the *SMAD4* c.1498A>G (p.I500V) variant, the functionally related c.1486C>T (p.R496C) variant (also a recurrent mutation in Myhre syndrome) showed evidence of clonal expansion. Next, the *PPP2R5D* variants (two Tier 1, one Tier 2) each localize near to or within a highly conserved acidic loop, coinciding with pathogenic residues associated with Jordan syndrome. As *PPP2R5D* is in the same phosphatase family as *PPP1CB*, and pathogenic mutations in Jordan syndrome have been shown to enhance RTK-Ras signalling (Smolen et al., 2023; **section 3.3.2**), these clustered mutations more plausibly represent biologically significant events rather than stochastic drift. Moreover, as variants in candidate genes were detected at significantly lower VAFs than those in known selfish genes (0.00078 vs. 0.0016 VAF; p = 0.00019, Wilcoxon rank-sum test), this may reflect relatively milder fitness advantages conferred to mutant spermatogonia that could explain the lack of additional evidence of clonal expansion. Functional characterization of these candidate selfish mutations through competitive assays *in vitro* or through *in vivo* transplantation (**section 1.11.1**) will be required to confirm this hypothesis.

In sum, these data show that rhAmpSeq is well-suited for multiplexed target capture and the sensitive detection of rare mutations under positive selection in the male germline. These findings expand the repertoire of potential selfish mutations and support the central premise of this thesis: that studying positive selection through the the lens of recurrent DNMs in clinically affected populations is an effective approach for discovery.

Chapter 5 – Investigation of pathogenic *SMAD4* mutations in sperm

5.1 – Introduction

Two *de novo* mutations in *SMAD4,* c.1486C>T (p.R496C) and c.1498A>G (p.I500V), which correspond to the causative mutations in Myhre syndrome (MS), were identified as the highest-prioritized targets identified in an analysis of recurrent DNMs drawn from a study of >31,000 parent-offspring trios (**section 3**). These same DNMs were both subsequently identified as mutational events (Tier 1 variants) in a discovery screen of aged testis biopsies (average age 70.7 years, range 63–76 years). Notably, the c.1486C>T (p.R496C) variant was detected in two neighbouring biopsies, suggestive of clonal expansion along the length of a seminiferous tubule (**section 4**).

Prior work has demonstrated that selfish mutations can be directly detected in sperm, and that the estimated mutation levels correspond well with the birth prevalence of their associated disorders (Goriely et al., 2003; Goriely et al., 2009; Giannoulatou et al., 2013). Given the estimated birth prevalence of MS (Lin et al., 2024), MS *SMAD4* mutations would be expected to occur at approximately ~10^{-6} (1 in a million) in sperm. To investigate whether pathogenic *SMAD4* mutations are positively selected in the male germline, a validated strategy to detect ultra-rare mutations was adapted to directly detect and quantify *SMAD4* mutations in the sperm of men across a range of ages.

5.2 – Results

5.2.1 – Strategy to detect ultra-rare *SMAD4* mutations in sperm

To determine whether MS-associated mutations in *SMAD4* are detectable in sperm, an ultra-sensitive assay was adapted to target the *NsiI* restriction endonuclease site (5'-ATGCAT-3') at *SMAD4* c.1494–1499. This approach relies on a restriction enzyme digest (RED) to cleave wild-type DNA sequences and select for mutations disrupting the restriction site of interest, and PCR to subsequently enrich intact mutant sequences (REDPCR) (Goriely et al., 2003; Goriely et al., 2005; Goriely et al., 2009; Giannoulatou et al., 2013). Here, *NsiI* restriction digest leads to the equal enrichment of all 18 possible substitutions within the 6 bp restriction site, including the two commonest MS mutations: *SMAD* c.1498A>G (p.I500V) and c.1499T>C (p.I500T) (**Figure 5.1**).

A key feature of the *NsiI* site is its palindromic nature; each potential substitution within *SMAD4* c.1494–1499 has a corresponding reciprocal substitution (**Figure 5.1**). This symmetry can be leveraged to interpret variant frequencies, revealing whether certain substitutions are enriched on their own or at comparable levels with their reciprocal, thereby providing additional context for evaluating the biological significance of observed variants.

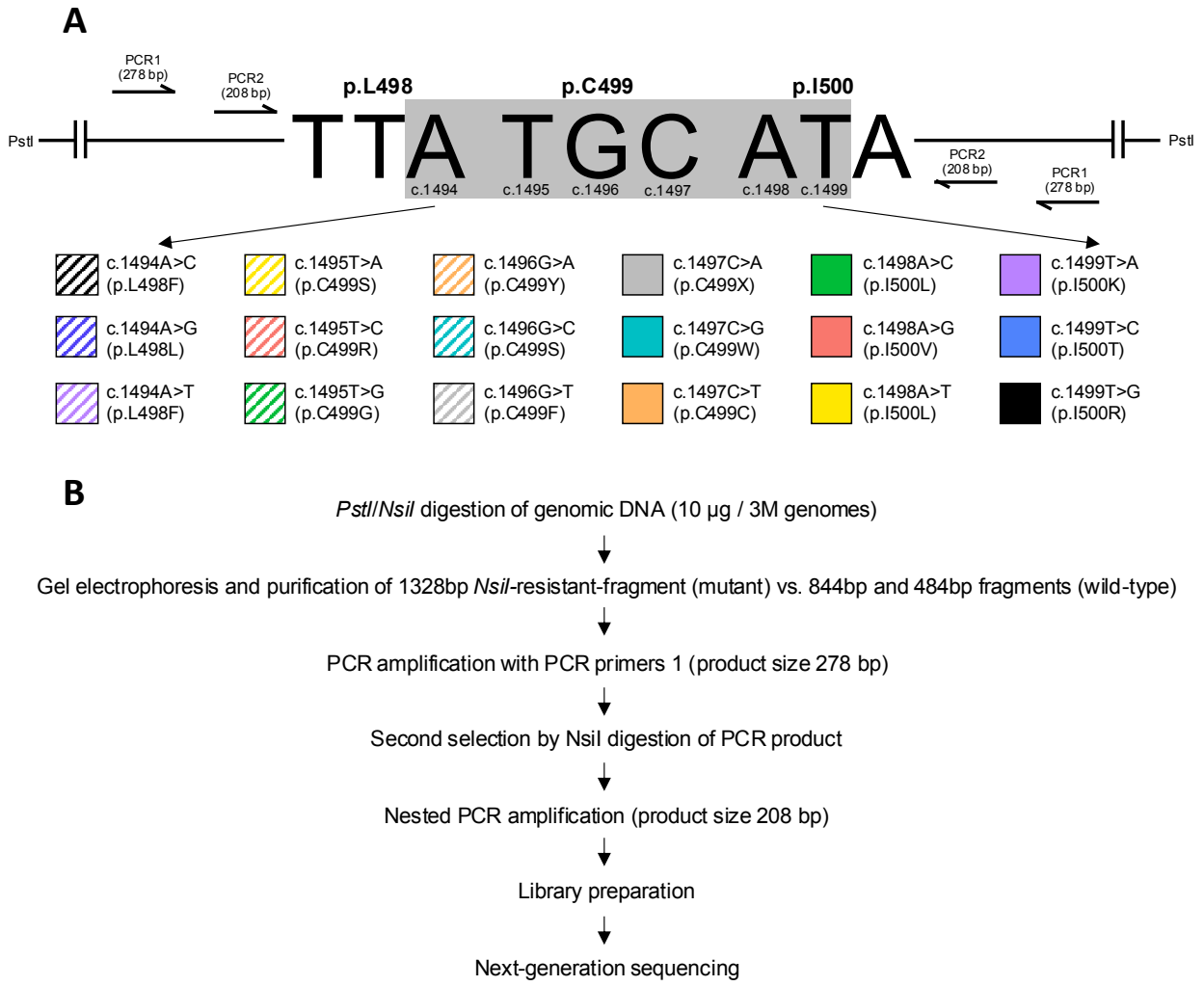


Figure 5.1: Overview of the *SMAD4* Restriction Enzyme Digest Polymerase Chain Reaction (REDPCR) assay. (A) The discriminative *NsiI* restriction enzyme site at *SMAD4* c.1494–1499 is boxed in gray and flanked by two *PstI* restriction sites, which are used to release a 1328 bp fragment; the cDNA positions (below) and amino acids (above) are also outlined. All 18 possible substitutions resistant to *NsiI*-digestion between *SMAD4* c.1494-1499 are outlined with unique colours; reciprocal substitutions within the palindromic site share the same corresponding colour with differing patterns. **(B)** Summary of the *SMAD4* REDPCR workflow. 10 µg of genomic DNA (3 million haploid genomes) per sample is digested with *NsiI* and *PstI*. Wild-type DNA is cleaved into two smaller fragments of 844 and 484 bps whereas mutant DNA remains intact at 1328 bp. Digested DNA samples are then gel electrophoresed overnight and *NsiI*-resistant DNA is excised. DNA extraction is individually performed for each gel slice. Extracted DNA is amplified with an initial PCR (278 bp) followed by an additional *NsiI* restriction digest to remove remaining WT DNA. A final, nested PCR (208 bp) is then performed to amplify mutant DNA prior to NGS library preparation.

5.2.2 – Assessing the sensitivity of the SMAD4 REDPCR assay

To establish the assay's limit of detection, genomic DNA from an individual with MS (heterozygous for c.1498A>G (p.I500V)) was spiked into 10 µg of control blood DNA (3 million genome equivalents) corresponding to mutation levels between ~10^{-6} (~3 mutant molecules) to ~10^{-4} (~300 mutant molecules). Following REDPCR enrichment and next-generation sequencing, the relative abundance of all 18 substitutions was analyzed.

The non-reference allele fraction of *SMAD4* c.1498A>G increased with the amount of spiked-in DNA and was not detected in the blood control, demonstrating the specificity of the assay (**Figure 5.2A**). At the lowest spike-in level of ~10^{-6}, c.1498A>G represented ~18% of non-reference reads, whereas at the highest spike-in level of 10^{-4}, it accounted for ~99% of non-reference reads. Two additional substitutions, c.1496G>T and c.1497C>A, which are reciprocal substitutions within the *NsiI* site, were detected at similar non-reference fractions in all samples (average ~22% and 18%, respectively) and were increasingly prevalent at low spike-in levels and in the blood control. c.1499T>G was also detected in the blood control, but represented only a minority of non-reference reads (~15%).

5.2.3 - A pilot study of SMAD4 mutation prevalence in sperm

Having demonstrated the ability of the assay to capture as few as 3 mutant molecules per 3 million genomes (~10^{-6}), a pilot study of 12 sperm samples (range 29.9–60.6 years) and 1 EBV control sample was undertaken to investigate whether mutations at *SMAD4* c.1494–1499 are detectable in sperm.

The most frequently observed substitutions were c.1496G>T and c.1497C>A, occurring in all samples at similar frequencies (average 25.1% and 22.0%, respectively) and were the only two non-reference substitutions identified in the EBV control (**Figure 5.2B**). Two other substitutions,

c.1496G>A and c.1497C>T, were not observed in the EBV control but were detected at similarly low abundance (average 1.8% and 2.5%, respectively) in 5 of 12 and 6 of 12 sperm samples, respectively.

The MS mutations, c.1498A>G and c.1499T>C, were detectable in 9 of 12 and 7 of 12 sperm samples, respectively. Across sperm samples with detectable MS mutations, c.1498A>G was detected at a higher relative abundance than c.1499T>C (average 50% vs. 15%, respectively). Three other substitutions – c.1498A>C, c.1498A>T, and c.1499T>G – were detected in sperm a total of 4 times (across 3 samples). Although these variants have not been reported in association with a germline disorder nor are they present in databases such as gnomAD or COSMIC, they all encode amino acid changes at the p.I500 codon.

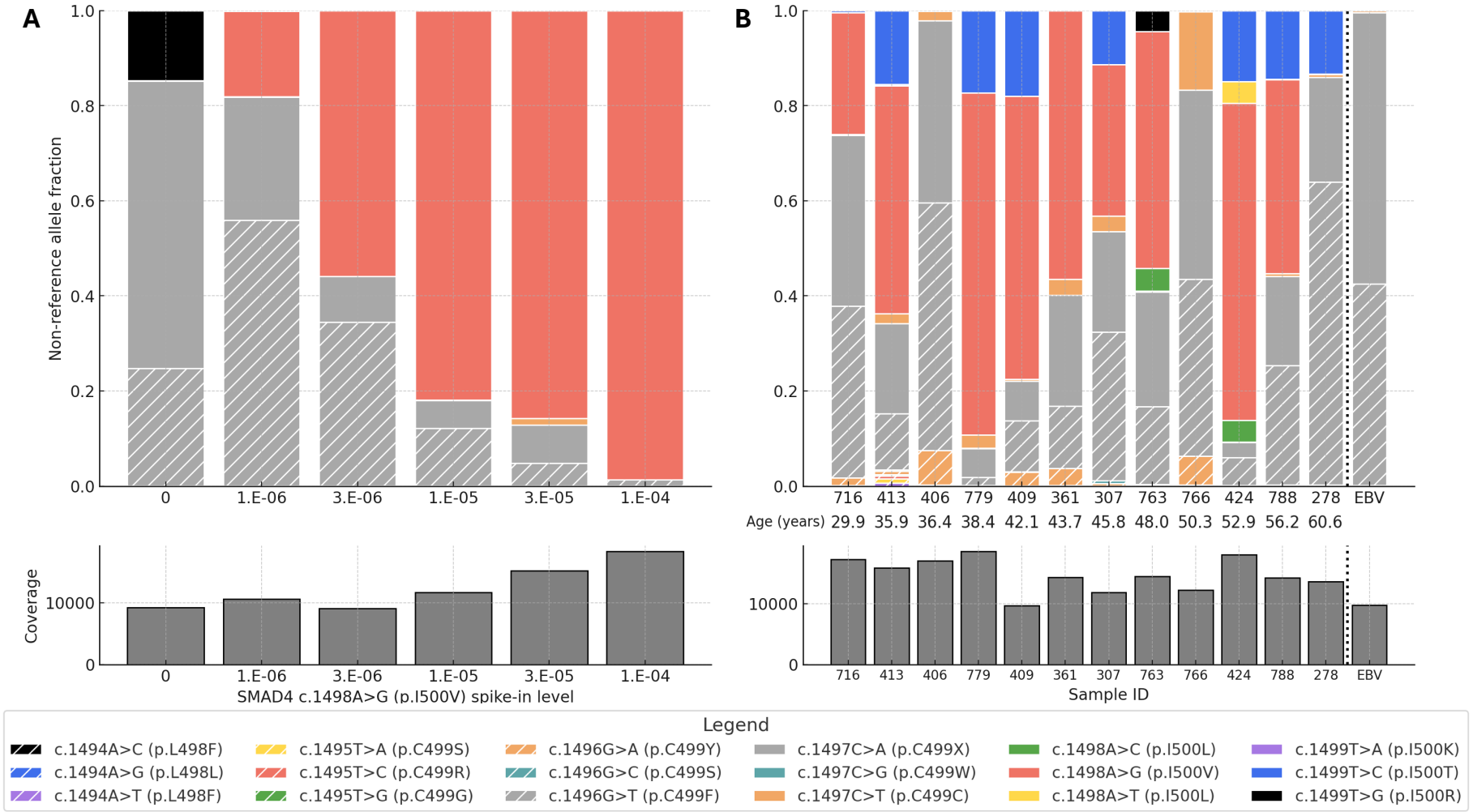


Figure 5.2: Establishing the limit of detection of the *SMAD4* REDPCR assay and determining whether MS *SMAD4* mutations are detectable in sperm. The legend outlines the 18 possible substitutions at the SMAD4 c.1494–1499 *NsiI* restriction site and is colour-coded as described in **Figure 5.1**. **(A)** Serial dilution of DNA from an MS patient heterozygous for SMAD4 c.1498A>G (p.I500V) into 10 µg (3 million haploid genomes) of blood control DNA. Non-reference reads between SMAD4 c.1494–1499 were plotted for each sample. The legend (below) outlines the 18 substitutions at SMAD4 c.1494–1499 and is colour-coded as described in Figure 2. Details outlining the spike-in level of c.1498A>G, the number of spiked in mutant molecules, and the depth of sequencing are included below the x-axis. Data is provided in **Supplementary Table S2**. Figure 4: Pilot screen to investigate whether mutations at SMAD4 c.1494–1499 are detectable in 12 sperm samples from a range of ages (29.9–60.6 years). Non-reference fractions for 18 substitutions between c.1494–1499 are plotted per sample. Each bar represents an individual sample and additional details regarding sample ID, donor age, mutational status for c.1498A>G and c.1499T>C, and sequencing depth are included below. Samples are organized by donor age and a dotted black line separates sperm samples from the EBV control sample. Data is provided in **Supplementary Table S3**.

5.2.4 – Generation of an *NsiI*-resistant genomic integrant

Given the rarity of MS, it was surprising to note that most men harboured detectable c.1498A>G and c.1499T>C mutations. To enable quantification, previous studies have relied on spiking in known quantities of genomic DNA from patients with unique mutational profiles such as double or triple mutations (Goriely et al., 2003; Giannoulatou et al., 2013). In the absence of such unique *SMAD4* signatures, a CRISPR-Cas9 strategy was employed to engineer an *NsiI*-resistant genomic integrant for use as a quantitative standard.

Although originally designed to introduce a conservative, single base change abolishing the *NsiI* site at *SMAD4* c.1494 (**section 2.6.5**), due to the closure of the MRC Weatherall Institute of Molecular Medicine because of the COVID-19 pandemic, a compromise was made to prioritize a clone harbouring a 17 bp heterozygous deletion spanning c.1478–1494 (c.1478_1494del). As the final base of c.1478_1494del occurs at the first base of the *NsiI* restriction site, the mutant allele is resistant to *NsiI* cleavage (**Figure 5.3**).

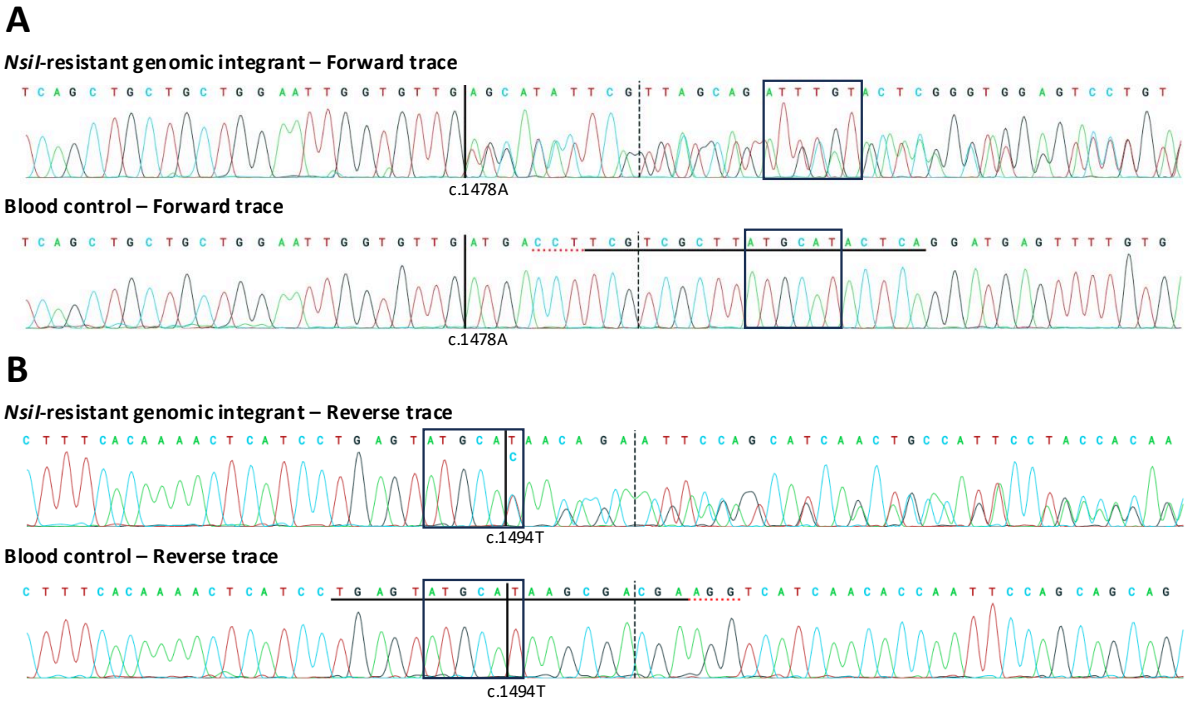


Figure 5.3: Sanger sequencing traces of the *NsiI*-resistant genomic integrant and a reference blood control demonstrate the location of the CRISPR-Cas9 edit. A red horizontal dotted line underlines the reverse complement of the PAM site, and a vertical dashed line denotes the corresponding cut site 3 bp downstream. In each trace, the *NsiI* restriction site at c.1494–1499 is boxed; in the control traces, the *NsiI* site is intact, whereas in the *NsiI* integrant traces, it includes the 17 bp heterozygous deletion (c.1478_1494del). **(A)** Forward traces: the solid vertical line at c.1478 marks the start of the c.1478_1494del, also denoted by overlapping traces. **(B)** Reverse traces: the solid vertical line at c.1494 marks the end of c.1478_1494del, abolishing the first base pair of the *NsiI* site.

5.2.5 – Titration-reconstruction experiment using Myhre syndrome DNA

To establish the limit of detection of the REDPCR assay, a titration-reconstruction experiment was performed by mixing genomic DNA from two MS probands (one heterozygous for c.1498A>G and the other for c.1499T>C) into 10 μg control blood DNA (3 million genome equivalents), resulting in mutation levels ranging from 10^{-6} (~3 mutant molecules) to 10^{-4} (~300 mutant molecules). In technical replicates, titration samples were also spiked with 100 mutant molecules of the *NsiI*-resistant genomic integrant at a final concentration of 3×10^{-5}.

Following next-generation sequencing, mutation levels were estimated by normalizing against the level of the *NsiI* integrant. For c.1498A>G, the estimated range was 0.08×10^{-6} – 61.53×10^{-6}, and for c.1499T>C, 0.04×10^{-6} – 60.41×10^{-6}. Between 10^{-5} and 10^{-4}, mutation levels for c.1498A>G and c.1499T>C were underestimated by up to 1.7-fold whereas below 10^{-5}, mutation levels were overestimated by up to 2.1-fold. Input DNA correlated with measured mutation levels for both c.1498A>G (Pearson $r > 0.99$, $p < 0.0001$) and c.1499T>C (Pearson $r > 0.99$, $p < 0.0001$) (**Figure 5.4A**). The only other substitutions detected $\geq 10^{-6}$ were c.1496G>T and c.1497C>A (range 0.24×10^{-6} – 1.47×10^{-6} and 0.95×10^{-6} – 2.96×10^{-6}, respectively), neither of which were correlated with input DNA (Pearson $r = 0.22$, $p = 0.6687$; Pearson $r = -0.16$, $p = 0.7584$, respectively).

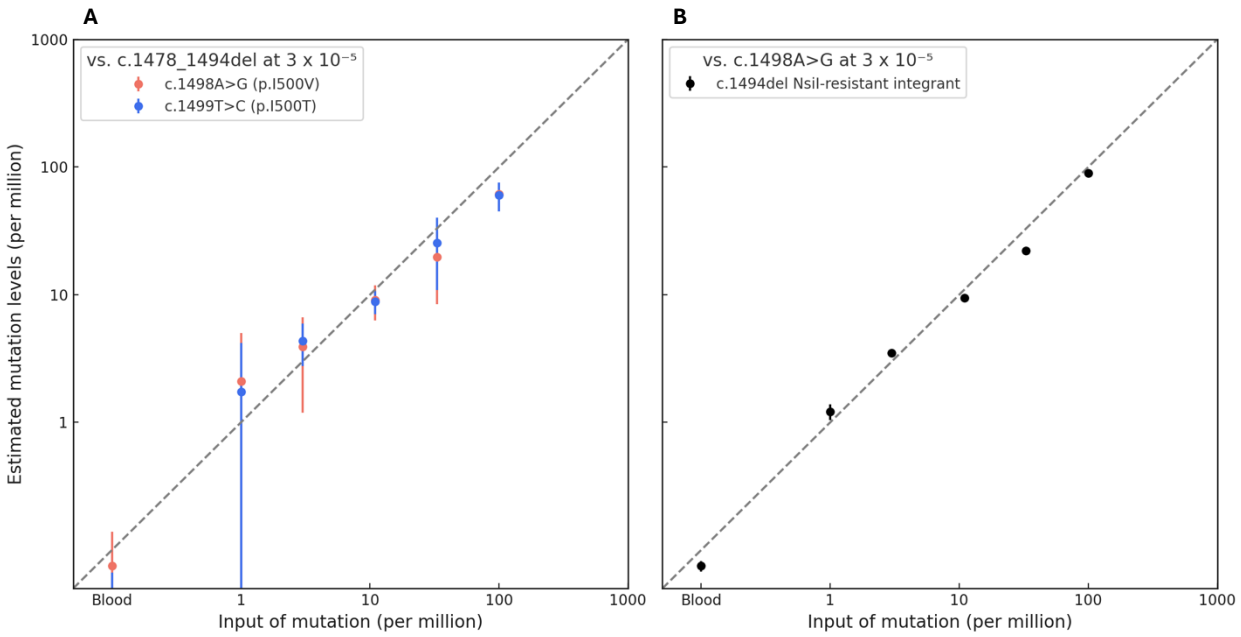


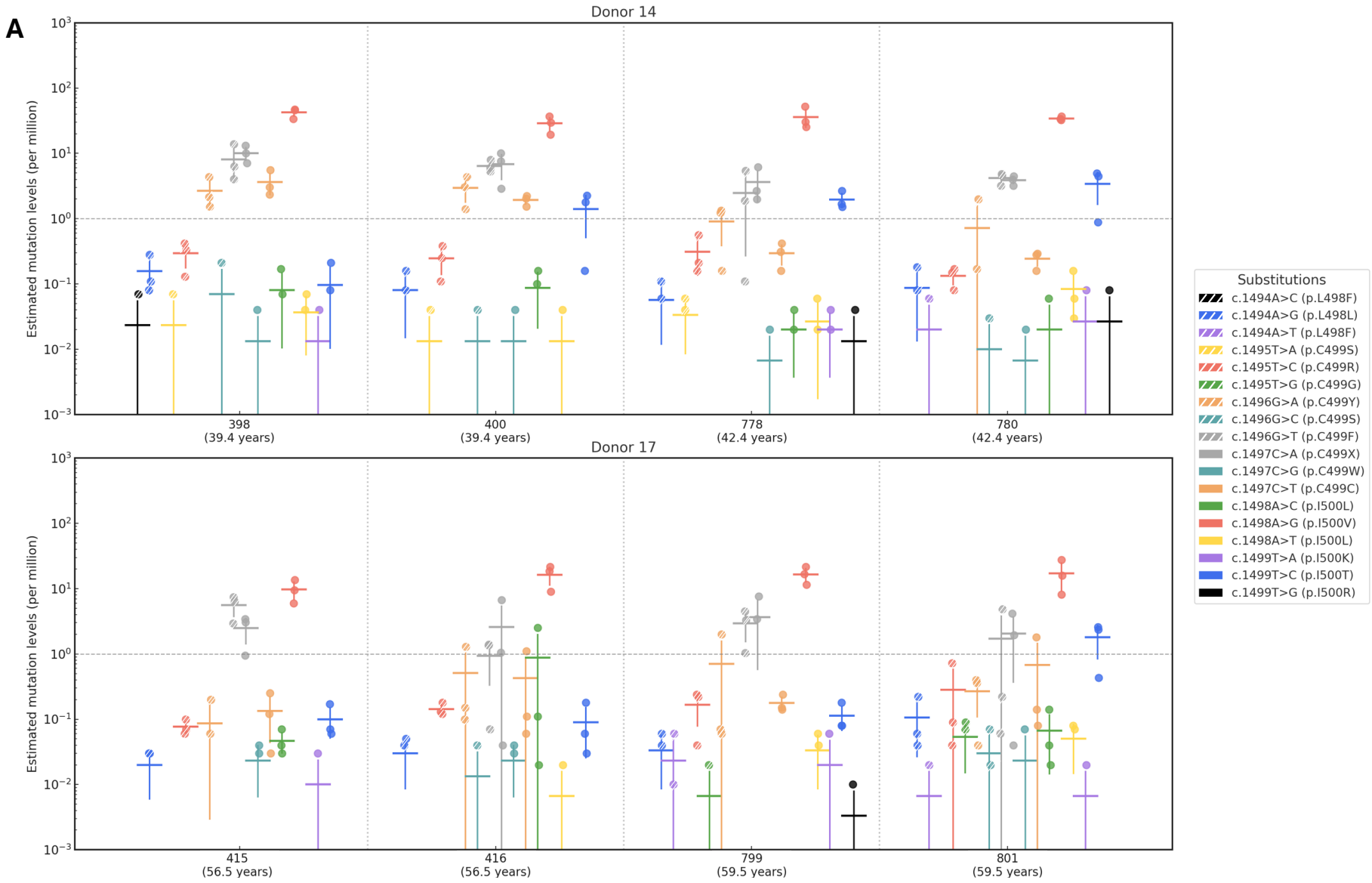
Figure 5.4: The *SMAD4* REDPCR assay is sensitive to ~10^{-6} for mutations at *SMAD4* c.1494–1499. Estimated mutation levels (mean ± SD of two technical replicates) from two titration reconstruction experiments. **(A)** DNA from two Myhre syndrome patients was serially diluted into control blood DNA and spiked with the *NsiI*-resistant genomic integrant (final concentration 3 x 10^{-5}). **(B)** DNA from the heterozygous c.1478_1494del *NsiI* integrant was serially diluted into control blood DNA, while DNA from two MS patients (heterozygous for c.1498A>G and c.1499T>C) was spiked in at 3 x 10^{-5}. Data is provided in **Supplementary Table S4**.

To further evaluate the robustness of the assay, the reverse experimental conditions were tested, in which the *NsiI* integrant was serially diluted into 10 µg of control blood DNA at levels 10^{-6} to 10^{-4}. Each sample was spiked with MS DNA from two probands (c.1498A>G and c.1499T>C) at a final concentration of 3 x 10^{-5}. Mutation levels of the titrated genomic integrant were estimated by normalizing against the level of c.1498A>G. The *NsiI* integrant was detected at a range of 0.07 x 10^{-6} to 90.09 x 10^{-6} and mutation levels were within 1.5-fold of expectations at all spike-in levels (**Figure 5.4B**). Taken together, these data demonstrate the ability of the assay to accurately measure mutation levels at *SMAD4* c.1494-1499 as low as ~10^{-6}.

5.2.6 – Reproducibility of the SMAD4 REDPCR assay

Having established the sensitivity of the *SMAD4* REDPCR assay, the reproducibility of the assay was assessed by analyzing between 2–4 biological replicates (i.e., separate sperm collection made between ~1 week to ~3 years apart) from four individuals (range 39.4–59.5 years). Each biological sample (two each from donors 21 and 27, and four each from donors 14 and 17) was then screened in technical triplicates.

Across 12 biological samples, there were 50 total substitutions detected above $\geq 10^{-6}$ and mutation level estimates were consistent across technical replicates of the same sample with a median fold-difference of 3.1 (95% CI: 2.4–4.3) (**Figure 5.5**). Measurements were also consistent between biological samples of the same individual. 21 substitutions (8 distinct) were detected above background in more than one sample of a donor and the median fold difference was ~2.78 (95% CI: 1.77–5.98). In particular, estimated mutation levels for the c.1498A>G, which was detected $\geq 10^{-6}$ in all samples (range 6.53×10^{-6} to 8.38×10^{-5}), did not differ between biological samples of the same individual ($p > 0.05$ all comparisons, Wilcoxon rank-sum test) and differed by only ~1.2-fold (95% CI: 1.00–1.77) between sperm collections up to 3 years (**Figure 5.5**). These results indicate that measurements of a single sperm sample reliably capture the mutational load of an individual.


A
Donor 14
Donor 17
Estimated mutation levels (per million)
10^{3}
10^{2}
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
398
(39.4 years)
400
(39.4 years)
778
(42.4 years)
780
(42.4 years)
415
(56.5 years)
416
(56.5 years)
799
(59.5 years)
801
(59.5 years)
Substitutions
c.1494A>C (p.L498F)
c.1494A>G (p.L498L)
c.1494A>T (p.L498F)
c.1495T>A (p.C499S)
c.1495T>C (p.C499R)
c.1495T>G (p.C499G)
c.1496G>A (p.C499Y)
c.1496G>C (p.C499S)
c.1496G>T (p.C499F)
c.1497C>A (p.C499X)
c.1497C>G (p.C499W)
c.1497C>T (p.C499C)
c.1498A>C (p.I500L)
c.1498A>G (p.I500V)
c.1498A>T (p.I500L)
c.1499T>A (p.I500K)
c.1499T>C (p.I500T)
c.1499T>G (p.I500R)

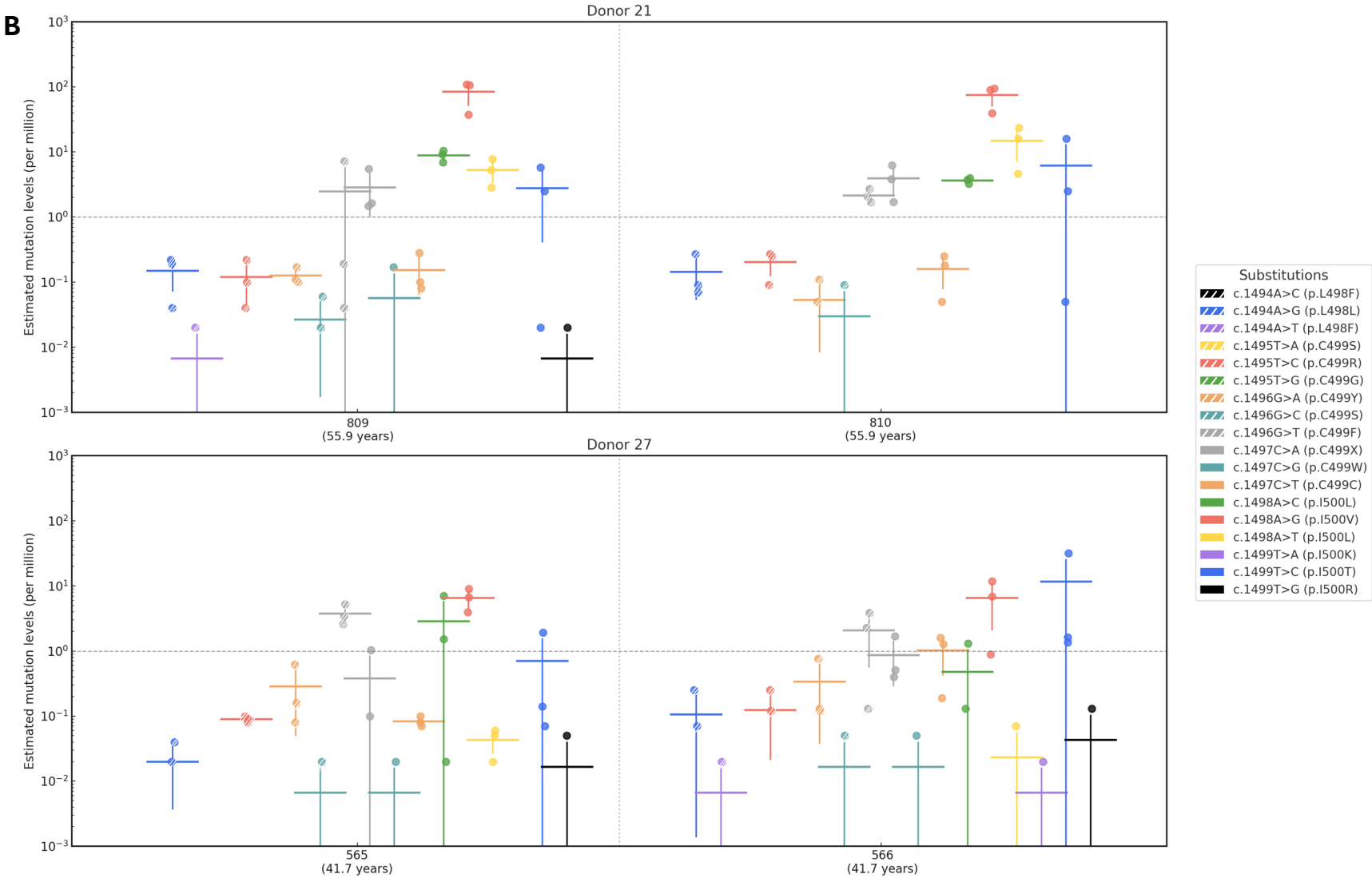


Figure 5.5: Mutation level estimates made with the *SMAD4* REDPCR assay are consistent across technical and biological replicates. Estimated mutation levels for all 18 possible substitutions within the *NsiI* restriction site (colour and hatching indicated in the legend) are plotted in order of their appearance within c.1494–1499 for each sample. Mutation levels for individual technical replicates are shown, along with the mean (± SD) across triplicates. The dotted horizontal line indicates the assay's limit of detection. **(A)** Donors 14 and 17 each had four sperm samples comprised of two groups: samples within a group were collected ~1 week apart, while samples in different groups were obtained ~3 years apart. **(B)** Donors 21 and 27 each had two samples ~1 week apart. Data is provided in **Supplementary Table S5**.

5.2.7 - Quantification of SMAD4 c.1494–1499 mutation levels in sperm and blood

Mutation levels for all 18 substitutions at *SMAD4* c.1494–1499 were measured in a cohort of 57 sperm (average age 40.7 years, range 27.2–75.9 years) and 5 blood samples (average age 39.5 years, range 24.8–57.0 years), with each sample assayed in technical triplicates.

Substitutions encoding changes at the p.L498 codon (i.e., at the first base of the c.1494–1499 *NsiI* site; **Figure 5.1**) were not detected above background in either sperm or blood. In contrast, two substitutions encoding changes at the p.C499 codon (which spans c.1495–1497), c.1496G>T and c.1497C>A, were detected in most sperm (51 of 57 and 52 of 57, respectively) and blood samples (4 of 5 and 3 of 5, respectively) (**Figure 5.6**). The average levels of c.1496G>T in sperm and blood were 3.69×10^{-6} (range 0.01×10^{-6} to 1.58×10^{-5}) and 3.16×10^{-6} (range 0.06×10^{-6} – 8.07×10^{-6}), respectively, and did not differ between tissues ($p = 0.578$, Wilcoxon rank-sum test). Similarly, the average levels of c.1497C>A in sperm and blood were 3.62×10^{-6} (range 0.00×10^{-6} – 1.49×10^{-5}) and 2.47×10^{-6} (range 0.00×10^{-6} – 6.69×10^{-6}, $n = 5$), respectively, and also did not differ between tissues ($p = 0.301$, Wilcoxon rank-sum test). Mutation levels in sperm and donor age were not correlated for either c.1496G>T (Spearman $r_s = 0.189$, $p = 0.158$) or c.1497C>A (Spearman $r_s = 0.035$, $p = 0.796$, respectively) (**Figure 5.7**). Likewise, mutation levels in blood and donor age did not correlate for either mutation (Spearman $r_s = 0.300$, $p = 0.624$; Spearman $r_s = 0.000$, $p = 1.000$, respectively).

Two other substitutions encoding changes at p.C499, c.1496G>A and c.1497C>T were detected $\geq 10^{-6}$ in a minority of samples this survey (3/57 and 10/57 sperm samples, respectively; 0/5 and 2/5 blood samples, respectively). The average levels of c.1496G>A in sperm and blood were 0.41×10^{-6} (range 0.05×10^{-6} – 2.06×10^{-6}) and 0.31×10^{-6} (range 0.01×10^{-6} – 0.67×10^{-6}), respectively, and did not differ between tissues ($p = 0.283$, Wilcoxon rank-sum test). The average

levels of c.1497C>T in sperm and blood were 0.5 x 10^{-6} (range = 0.01 x 10^{-6} – 2.03 x 10^{-6}, n = 57) and 0.61 x 10^{-6} (0.01 x 10^{-6} – 1.33 x 10^{-6}, n = 5), respectively, and did not differ between tissues (p = 0.990, Wilcoxon rank sum test). Mutation levels in sperm and donor age did not correlate for either c.1496G>A (Spearman r_s = 0.210, p = 0.116) or c.1497C>T (Spearman r_s = -0.156, p = 0.2481, respectively).

Substitutions encoding changes at the p.I500 codon exhibited a different pattern than those at p.L498 and p.C499 (**Figure 5.6**). Among them, the c.1498A>G substitution was the most prevalent, detected ≥10^{-6} in 42/57 sperm samples (73.7%) and reaching a high maximum of 1.04 x 10^{-4} in a sperm sample of a man aged 59.9 years. Across all 57 sperm donors (average age 40.7 years, range 27.2–75.9 years), the average mutation level of c.1498A>G was 1.02 x 10^{-5}, and mutation levels were positively correlated with donor age (Spearman r_s = 0.485, p = 0.0001) (**Figure 5.7**). By contrast, c.1498A>G was detected ≥10^{-6} in 1/5 blood samples and measured in blood at an average level of 0.40 x 10^{-6}, below the assay's limit of detection. Mutation levels of c.1498A>G were significantly higher in sperm vs. blood (p = 0.010, Wilcoxon rank sum test). Similarly, c.1499T>C mutations were detected ≥10^{-6} in 13/57 (22.8%) sperm samples, reaching a maximum of 1.44 x 10^{-5} in a sperm sample of a man aged 71.7 years, and an average mutation level of 0.97 x 10^{-6} across the entire cohort. In sperm, c.1499T>C mutation levels were significantly correlated with donor age (Spearman r_s = 0.301, p = 0.023). c.1499T>C was detected in 0/5 blood samples and measured in blood at an average level of 0.09 x 10^{-6}.

Three additional substitutions at p.I500 were detected at elevated levels in sperm. Two substitutions encoding the same p.I500L change, c.1498A>C and c.1498A>T, were detected in 8/57 (14%) and 4/57 (7%) of sperm samples, respectively, reaching maxima of 8.87 x 10^{-6} and 5.29 x 10^{-6}, respectively, in the same sperm sample (55.9 years), and neither substitution was

detected in blood (**Figure 5.7**). Whereas mutation levels in sperm for c.1498A>C were significantly correlated with donor age (Spearman r_s = 0.389, p = 0.003), this same relationship was not observed for c.1498A>T (Spearman r_s = 0.232, p = 0.082). Lastly, c.1499T>G, which encodes p.I500R, was detected in a single sperm sample at 3.13 x 10^{-6}, in a man aged 38.8 years.

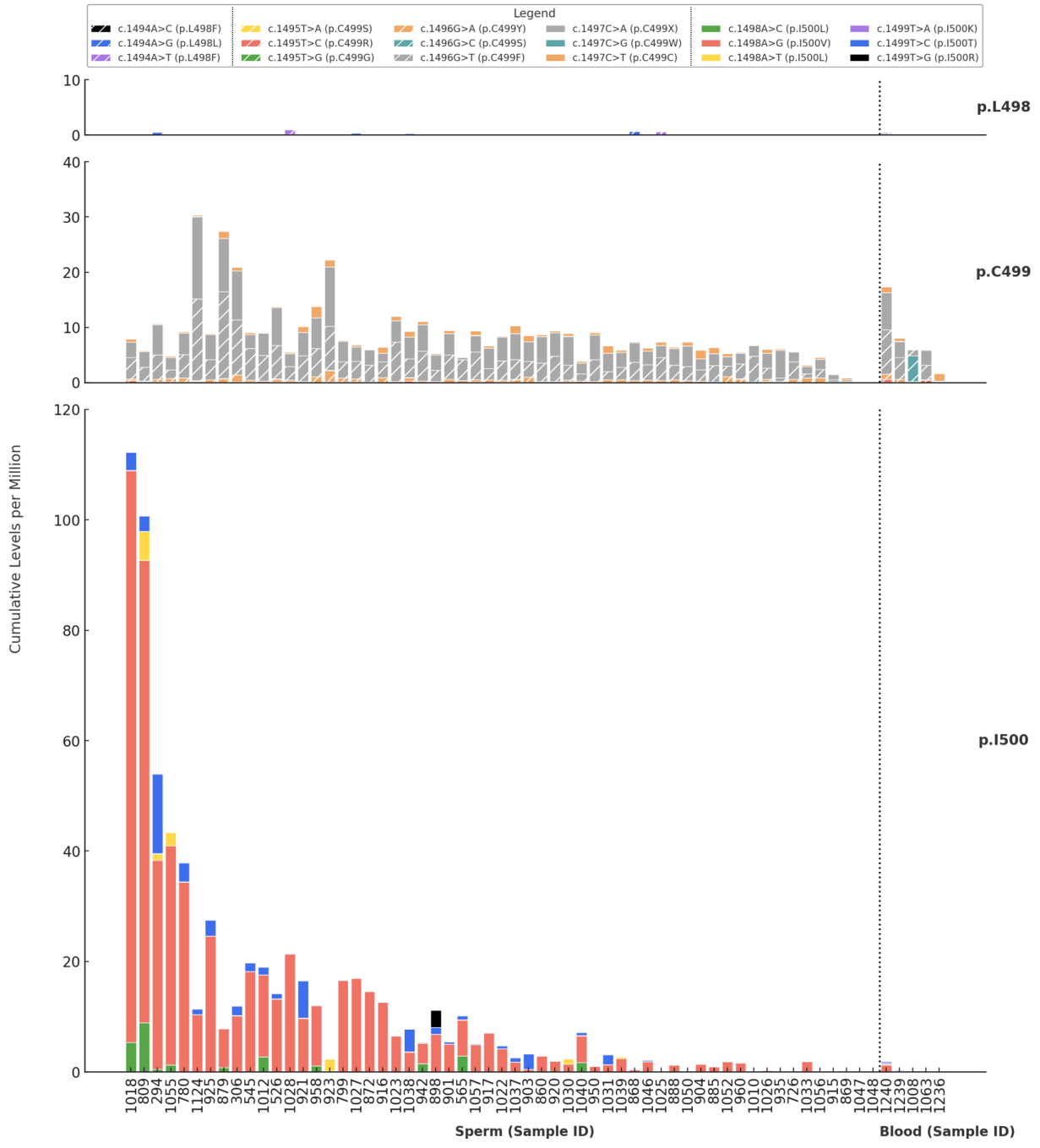


Figure 5.6: Cumulative mutation levels at *SMAD4* c.1494–1499 in a cohort of 57 sperm and 5 blood samples. All 18 possible substitutions within the *NsiI* restriction site are shown, with each bar representing the total mutational load for a given sample for each of the three amino acids encompassed within the *NsiI* site (p.L498, p.C499, and p.I500). The legend highlights how the substitutions are coloured and grouped according to these codons. A dotted line divides sperm (left) and blood (right) samples. Data is provided in **Supplementary Table S6**.

A key characteristic of the *NsiI* restriction site is its palindromic sequence, which can be used to contextualize enriched substitutions and distinguish true biological events from artefacts (**section 5.2.1**). For example, the c.1496G>T and c.1497C>A substitutions, which reciprocally occur at p.C499, were detected above background (~10^{-6}) in nearly all samples. Not only were mutation levels strongly correlated with one another across the entire cohort (Spearman $r_s = 0.69$, $p = 8.01 \times 10^{-10}$), mutation levels also did not differ within each sample ($p = 0.978$, Wilcoxon signed-rank test) (**Figure 5.8**). A similar pattern emerged for c.1497C>T and c.1496G>A, where mutation levels were also correlated (Spearman $r_s = 0.37$, $p = 0.0032$) and also did not differ in paired comparisons ($p = 0.183$, Wilcoxon signed-rank test).

By contrast, substitutions affecting the p.I500 codon presented a different profile. For instance, although c.1498A>G and its reciprocal c.1495T>C were correlated (Spearman $r_s = 0.43$, $p = 0.0004$) – as c.1495T>C was not detected ≥10^{-6} in any sample, this correlation was driven by values below the LOD (**Figure 5.7**) – mutation levels for c.1498A>G were higher across all samples ($p = 4.71 \times 10^{-11}$, Wilcoxon signed-rank test). Similar patterns were observed for the four other p.I500 substitutions that were detected at elevated levels in sperm: c.1498A>C, c.1498A>T, c.1499T>C and c.1499T>G (**Figure 5.8**).

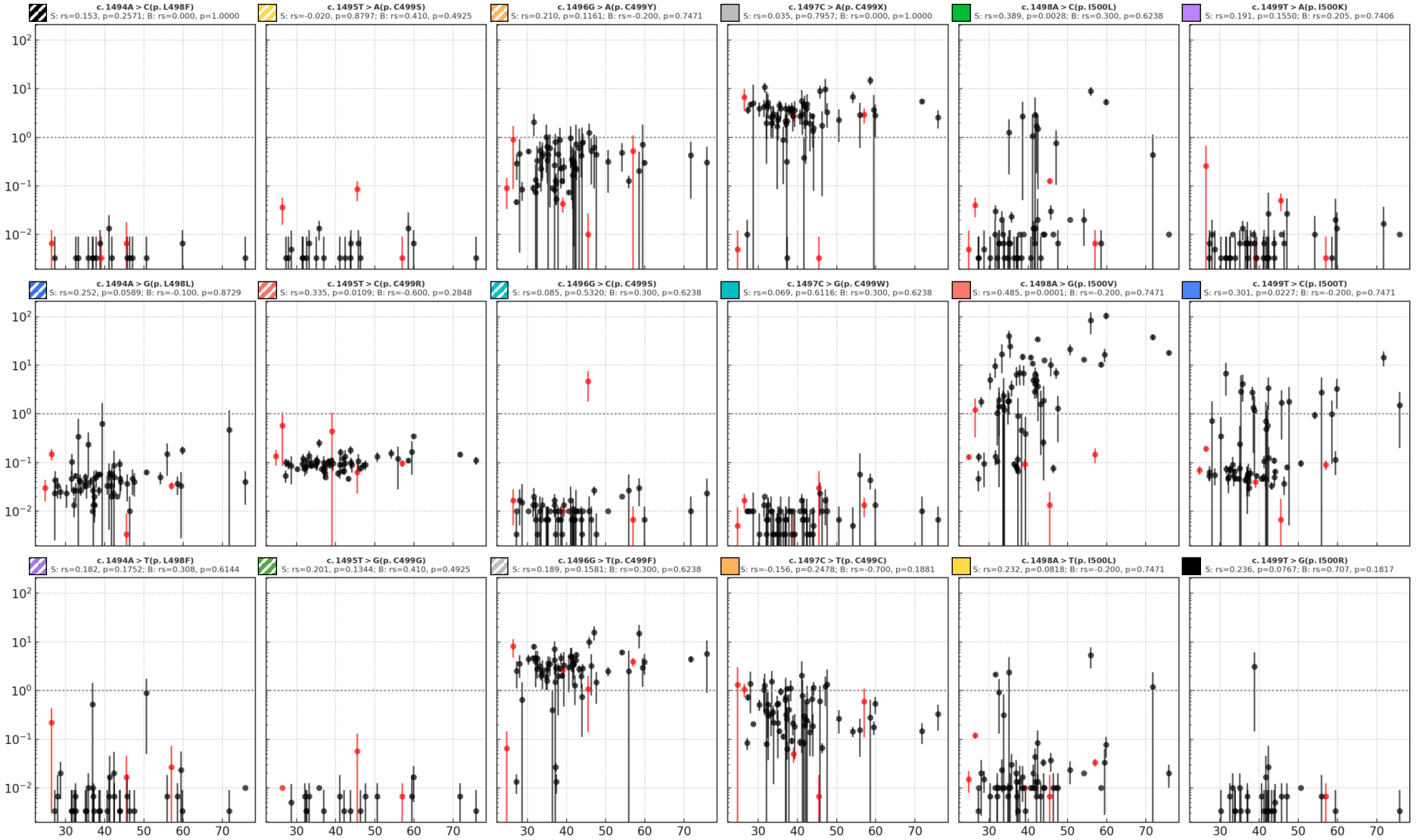


Figure 5.7: Quantification of all 18 substitutions at SMAD4 c.1494–1499 in a cohort of 57 sperm and 5 blood samples (24.8–75.9 years). Each column corresponds to one base pair within the *NsiI* restriction site, with reciprocal substitutions indicated by matching coloured and hatched boxes. For each sample, the mean (± SD) mutation levels across technical triplicates is plotted in black (sperm) or red (blood) against the age of the donor (x-axis, years). A horizontal dotted line at 1 marks the limit of detection for the *SMAD4* REDPCR assay. The Spearman rank correlation coefficient and p-value above each panel reflect how mutation levels and donor age relate for that particular substitution (S for sperm, B for blood). Data is provided in **Supplementary Table S6**.

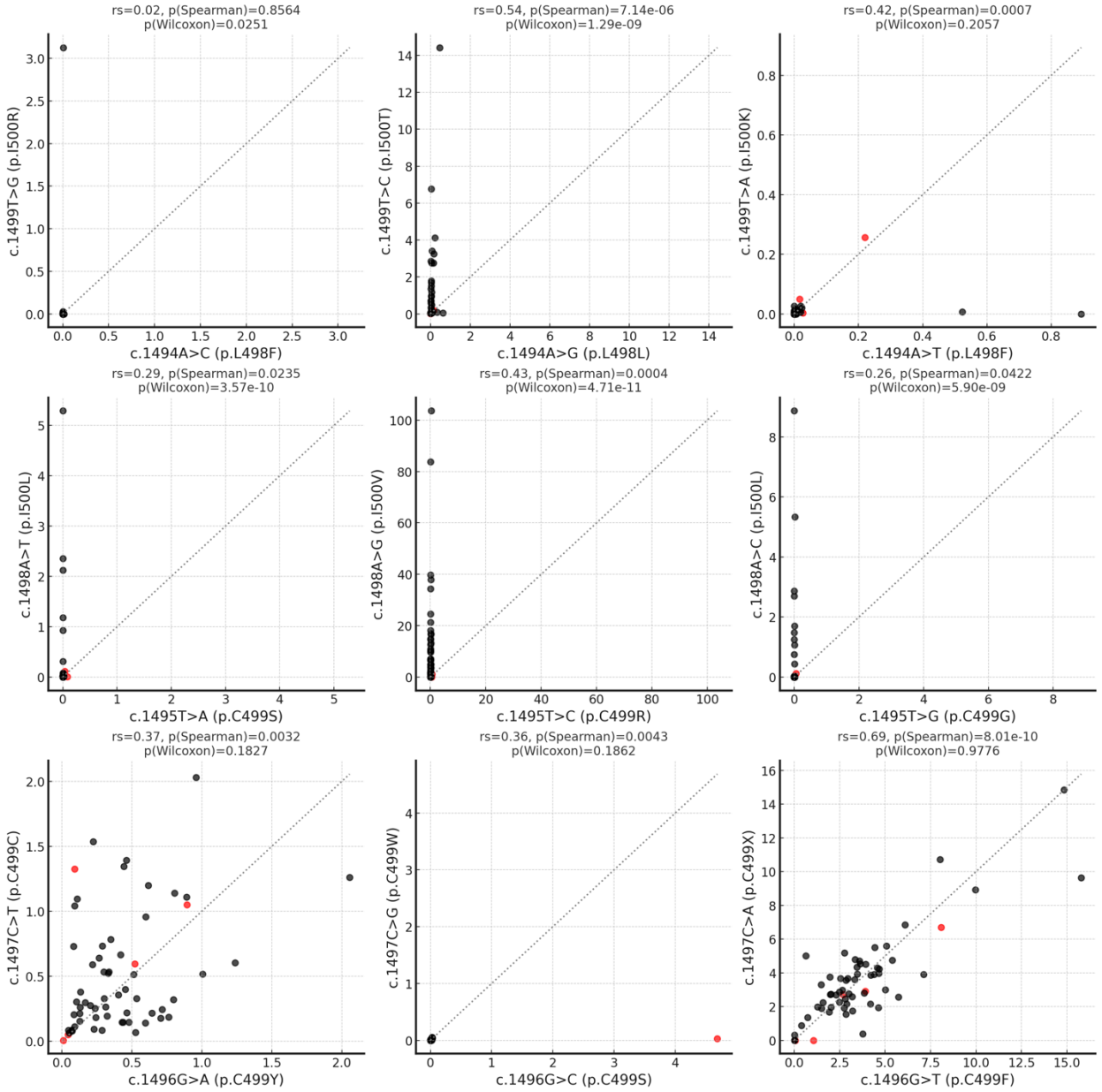


Figure 5.8: Correlation between mutation levels for the 9 reciprocal substitutions within the *NsiI* restriction site. Each panel corresponds to a reciprocal pair of substitutions, organized by the first base within the restriction site (e.g. c.1494) and its partner (e.g. c.1499). For each sample, the mean of three technical replicates is plotted in black (sperm) and red (blood). The Spearman rank correlation coefficient (r_s) and p-value are indicated at the top of each plot, along with the result of a paired analysis of mutation levels within each sample (Wilcoxon signed-rank test). Data is provided in **Supplementary Table S6**. Figure adapted from Wood et al., 2024.

5.3 – Discussion

In this chapter, an ultra-sensitive assay was adapted to exploit the *NsiI* restriction site (located at *SMAD4* c.1494–1499) and enable the comprehensive quantification of all 18 possible substitutions in this region. Through three separate titration-reconstruction experiments, the SMAD4 *REDPCR* assay demonstrated the ability to reliably capture and accurately quantify substitutions as low as ~10^{-6} (1 molecule per million). Next, by measuring mutation levels across multiple sperm samples of 4 individuals (with collections ranging from ~1 week to 3 years apart), results indicated that the assay was both precise among technical replicates (within ~3.1 fold) and consistent across biological samples from the same individual (within ~2.8-fold). Collectively these findings indicated that screening a single sperm sample was sufficient to capture an individual's mutational profile at *SMAD4* c.1494–1499 and underscored the reliability of the assay to detect ultra-rare variants.

Using this method, a cohort of 57 sperm (average age 40.7 years, range 27.2–75.9 years) and 5 blood samples (average age 39.5 years, range 24.8–57.0 years) was assayed in technical triplicates. Two substitutions, c.1496G>T (p.C499F) and c.1497C>A (p.C499X), were notably detected in virtually all sperm and blood samples, but mutation levels for neither c.1496G>T nor c.1497C>A differed between sperm nor blood, nor were they correlated with donor age. These substitutions are reciprocal within the *NsiI* site (**Figure 5.2**) and are equally enriched following restriction digest. Pairwise analyses further revealed that not only were mutation levels of c.1496G>T and c.1497C>A highly correlated within all samples (Spearman $r_s = 0.69$, $p = 8.01 \times 10^{-10}$), they also did not differ from each other ($p = 0.978$, Wilcoxon signed-rank test). Together, these findings suggest that these mutations are likely artefactual in nature. Supporting this notion, oxidative damage to guanine – producing 8-oxo-guanine, which can mispair with adenine – is known to arise under heat-intensive process (e.g. during PCR) or tissue processing (e.g. during

DNA extraction), leading to excessive G>T/C>A substitutions (Bruskov et al., 2002; Costello et al., 2013).

Substitutions encoding changes at p.I500 were, by contrast, enriched in sperm in absence of any reciprocal substitutions at c.1494–1495 detected $\geq 10^{-6}$ (**Figure 5.7**). The c.1498A>G and c.1499T>C mutations were collectively detected in the majority (~75.4%) of sperm samples at average levels of $\sim 10^{-5}$ (1 in 100,000) and $\sim 10^{-6}$ (1 in a million), respectively, and reached high maxima of $>10^{-4}$ and $>10^{-5}$, respectively, 3–4 orders of magnitude higher than the average germline mutation rate of $\sim 10^{-8}$ (Ségurel et al., 2014). As mutation levels in sperm for both c.1498A>G and c.1499T>C were also positively correlated with the age of the donor (Spearman r_s 0.485, p = 0.0001 and Spearman $r_s = 0.301$, p = 0.023, respectively), this suggests that pathogenic mutations in Myhre syndrome are positively selected in sperm and clonally enrich over time. These data also reinforce earlier findings of mutational foci for both c.1498A>G (p.I500V) and c.1486C>T (p.R496C) (located outside of the target space of the *SMAD4* REDPCR assay) in three testis biopsies of an aged man (Tes30G, 63 years) (**section 4**). In addition to the p.I500V and p.I500T mutations, two substitutions encoding p.I500L were also enriched in sperm, with mutation levels of c.1498A>C also positively correlated with age of the donor ($r_s = 0.389$; p = 0.0028), suggestive of positive selection.

Estimates of the birth prevalence of Myhre syndrome placed expectations of *SMAD4* mutations in sperm at $\sim 10^{-6}$. Although the average age of this cohort of 57 sperm samples (40.7 years, range 27.2–75.9 years) is far higher than current estimates of the average age of fatherhood (~31.4 years in the United States) (Greenberg et al., 2019), it was nevertheless surprising to find such high average levels of c.1498A>G (p.I500V) ($\sim 10^{-5}$) and c.1499T>C (p.I500T) ($\sim 10^{-6}$). By way of comparison, a previous survey of 99 sperm samples (average age 39.1 years, range 23.8–

73 years) reported a mean mutation level of ~2.4 x 10^{-5} for the *FGFR2* c.755C>G mutation in Apert syndrome – consistent with estimates of birth prevalence (~1 in 100,000), which is an order of magnitude higher than that of MS (Goriely et al., 2003). The discrepancy between the observed mutation prevalence in sperm and the purported birth prevalence of MS has several possible explanations. Although it is possible that p.I500V and p.I500T (and by extension, p.I500L) may not translate to functional spermatozoa or results in embryonic lethality, the mild phenotype of MS argues otherwise. Instead, challenges in clinical ascertainment likely contribute to MS being underdiagnosed. The cardinal features of MS can often develop later in life (Oldenburg et al., 2015; Garavelli et al., 2016; Lin et al., 2016; Meerschaut et al., 2019) and significant phenotypic overlap with other disorders can result in an initial misdiagnosis (Artemios et al., 2019; Jensen et al., 2020; Li et al., 2021). Such difficulties in establishing an MS diagnosis, along with incomplete penetrance, are further underlined by the occurrence of c.1498A>G and c.1486C>T in gnomAD v.4.1.0 (allele frequencies: 6.19 x 10^{-7} and 4.34 x 10^{-6}, respectively). Despite this apparent discrepancy, in this survey, c.1498A>G (p.I500V) and c.1499T>C (p.I500T) were detected $\geq 10^{-6}$ in 42/57 and 14/57 of sperm samples, respectively, consistent with earlier findings that p.I500V is reported ~3 times more than p.I500T in *de novo* cases of MS (**section 3.3.3; Supplemental Table S1**). In further support, it was noted in the subsequent publication that an individual presenting with mild MS bore the rare p.I500L mutation, highlighting the predictive value of identifying selfish mutations in their tissue of origin (Wood et al., 2024; Lin et al., 2024).

To date, all known selfish mutations occur in genes encoding components of the RTK-Ras pathway (Maher et al., 2018). The data in this chapter provide strong evidence that dominant mutations in *SMAD4* are under positive selection in the male germline. Given its role as the central regulator of BMP/TGF-β signalling, these findings substantiate the hypothesis that selfish

spermatogonial selection can operate outside of the canonical RTK-Ras pathway. Functional analyses of the MS SMAD4 mutations indicate a gain-of-function mechanism leading to constitutive BMP/TGF-β pathway activation (**section 3.3.3**). Although BMP signaling has been implicated in SSC maintenance (Guo et al., 2017), and SMAD4 co-expression with known spermatogonial markers was also validated using double immunofluorescence (**section 3.2.3**), the precise manner by which SMAD4 mutations at p.I500 confer a selective advantage to mutant spermatogonia remains unclear.

Following the completion of this work, three SMAD4 p.I500 mutations enriched in sperm – p.I500V, p.I500T, and p.I500L – were functionally characterized (Dr. Katherine Bush). In addition to increasing SMAD4 transcriptional activity, all three mutations also increased ERK phosphorylation upon stimulation with both FBS and TGF-β, suggesting that although MS SMAD4 mutation may activate the BMP/TGF-β pathways, their effects in spermatogonia could also involve cross-talk with RTK-Ras signalling (Wood et al., 2024). In the same functional assays, p.C499F and p.C499X (encoded by the reciprocal c.1496G>T and c.1497C>A substitutions) were also examined. Neither mutation resulted in any such increase in transcriptional activity, with the latter mutation causing a complete loss of function. These results further underscore that the elevated levels of c.1496G>T and c.1497C>A detected in all samples are likely attributable to technical artefact.

In addition to the findings presented in this chapter showing that MS SMAD4 mutations exhibit a high apparent germline mutation rate, two additional results were reported by collaborators in the associated publication (Wood et al., 2024). First, haplotypes of 22 individuals with MS were phased by identifying SNPs near the causative *SMAD4* mutation (up to 6.5 kb away) that differed between parents (i.e., were informative). In 16 MS parent-offspring trios with

informative SNPs, the *SMAD4* mutation was exclusively present on the paternal allele. Second, in two cohorts of 35 (USA) and 17 (EU) MS parent-offspring trios, parental age was compared against year-matched controls from their respective populations. On average, MS fathers were aged ~37 years – consistent with results of the literature review (**section 3.3.3; Supplemental Table S1**) – and were statistically older than the average age of fatherhood by ~4–6 years, thus establishing an epidemiological paternal age effect. Together, these results fulfilled the three recognized PAE criteria and established Myhre syndrome as a novel PAE disorder.

Chapter 6 - Conclusion and future directions

The progressive clonal enrichment of mutant spermatogonia is known to underlie a subset of dominant disorders whose incidence increases supralinearly with paternal age. To date, this mechanism, termed "selfish spermatogonial selection," has been associated exclusively with activating mutations occurring in genes within the RTK-Ras pathway. However, it is unlikely that positive selection is confined to a single signalling pathway since, in theory, any pathway or gene involved in SSC homeostasis may be vulnerable to selection (Goriely and Wilkie, 2012). Therefore, the primary goal of this thesis was to explore the broader landscape of positive selection in the male germline. To this end, the hypothesis that selection would be reflected by the recurrence of *de novo* mutations in clinically affected populations was tested.

6.1 – Identification of candidate selfish mutations

A systematic assessment of recurrently mutated amino acid residues in the largest-to-date study of >31,000 parent-offspring trios (Kaplanis et al., 2020) identified 22 DNMs across 20 genes that exceeded background rates – defined by the maximum observed recurrence of synonymous mutations (≥7 for non-CpG sites or ≥9 for CpG sites) – including 3 known selfish mutations in *FGFR3*, *MAP2K1*, and *PTPN11*. Each variant was evaluated through an integrated scoring framework that accounted for (i) expression in SSCs – including the direct evaluation of co-expression of select candidates with established spermatogonial markers (MAGEA4 and GFRA1) through double-immunofluorescent staining, (ii) known molecular function and mechanistic plausibility (e.g., involvement in known or implicated signalling pathways regulating SSC homeostasis), and (iii) alignment with established Paternal Age Effect criteria. As a result, 11 variants across 10 candidate genes were prioritized as strong candidates for selfish selection. In

contrast to established selfish mutations, which all act through a gain-of-function mechanism within the RTK-Ras pathway, the prioritized recurrent DNMs in this analysis were notably characterized by a broader range of functional changes and pathway involvement, including both loss-of-function and dominant-negative variants. For instance, recurrent nonsense mutations in *ADNP* involved in Helsmoortel-Van der Aa syndrome have been shown to attenuate Wnt signalling (D'Incal et al., 2024), a pathway that mediates SSC stemness in mice (Mäkelä and Hobbs, 2019), and may plausibly confer a fitness advantage to mutant SSCs.

In a recent preprint, Seplyarskiy applied a population genetics-driven approach to identify drivers of clonal expansions in spermatogonia (CES) from aggregated patient cohorts (autism spectrum and neurodevelopmental disorders (Kaplanis et al., 2020), developmental and epileptic encephalitis, and congenital heart disease) comprising ~55,000 trios, as well as ~6100 control trios and ~807,000 individuals from gnomAD v4, reasoning that CES causes an increase in the effective *de novo* mutation rate (Seplyarskiy et al., 2025). Single-base mutation rates were first estimated using the Roulette model, which accounts for extended sequence context, strand specificity, and regional variation (Seplyarskiy et al., 2023). Poisson and binomial statistics were then applied to detect significant deviations from expected *de novo* mutation rates. As the approach taken by Seplyarskiy was agnostic to the mechanism of mutation, it enabled the identification of (i) 15 candidate selfish genes exhibiting a significant excess of LOF mutations across the entire gene, and (ii) 20 residues across 17 genes acting in a GOF manner. These 32 genes included 4 known selfish genes and 8 of the 10 candidate genes prioritized in this thesis.

Whereas this thesis sought to identify candidates for selfish selection by focusing on the biologic plausibility of recurrent DNMs from patient cohorts, the approach by Seplyarskiy identified statistical deviations from baseline *de novo* mutation rates, regardless of disease

association. Indeed, Seplyarskiy identified elevated mutation rates in both disease and control cohorts (e.g., LOF mutations in *MIB1* with no overt disease association), indicating that clonal expansions in sperm do not always result in an established clinical phenotype but can still accumulate at unexpectedly high rates. Despite the different approaches, the convergent findings – such as overlapping candidate selfish genes and a wider range of mutational mechanisms – reaffirm the hypothesis that selfish spermatogonial selection is not restricted to activating mutations within the canonical RTK-Ras signalling pathway, and also highlight the value of patient cohorts as a rich resource to identify signatures of germline selection.

6.2 – Discovery of novel candidate mutations in aged testes

To facilitate a discovery screen of candidate selfish mutations, rhAmpSeq – a novel technique that leverages blocked PCR primers requiring RNase H2-mediated cleavage and has been shown to improve assay sensitivity over standard multiplexed PCR (Dobosy et al., 2021) – was adapted to detect low-frequency variants in the testes of aged men. Specifically, a custom rhAmpSeq panel targeting ~2.1 kb across 10 candidate genes (11 amplicons) and ~2 kb across 6 established genes (9 amplicons) was used to assay 186 testis biopsies dissected from three aged men (average age 70.7 years; range 63–76 years) in technical triplicates. In parallel, serially diluted DNA from 13 probands was screened to provide an empirical estimate of the limit of detection, aided by a custom statistical pipeline that estimated substitution-specific error rates at each nucleotide. This approach demonstrated a limit of detection of ~0.0005 VAF for regions achieving ≥5000x coverage. The sensitivity of the method was further substantiated in a pilot study, in which three known selfish mutations in *FGFR3, PTPN11,* and *RET* (range 0.0015–0.0047 VAF) were accurately recovered from previously-screened testis biopsies (Maher et al., 2018).

A total of 71 mutational events across 37 variants were identified, predominantly in established selfish genes with known associations to dominant disorders or cancer. Of these, 18 mutational events corresponding to 10 unique variants arose in 6 candidate selfish genes: *ADNP, CSNK2A1*, *PACS1*, *PPP1CB, PPP2R5D,* and *SMAD4*. Notably, three variants were detected in multiple neighbouring biopsies, consistent with clonal expansion along the length of a seminiferous tubule: *CSNK2A1* c. 593A>G (p.K198R) – the most common mutation in Okur-Chung neurodevelopmental syndrome (**section 3.3.10**), *PPP1CB* c.146C>G (p.P49R) – the most recurrent mutation in Noonan-like syndrome with loose anagen hair 2 (**section 3.3.8**), and *SMAD4* c.1486C>T (p.R496C) – a highly recurrent mutation in Myhre syndrome (**section 3.3.3**). By contrast, individual mutational foci (i.e., variants detected in only one testis biopsy) were found in *ADNP, PPP2R5D,* and *PACS1*. While detecting a candidate mutation in a single biopsy does not exclude the possibility of neutral drift or a passenger mutation, it may also reflect a nascent clone or a variant conferring a lower fitness effect. Consistent with this interpretation, mutations in candidate genes were detected at two-fold lower VAFs than those in established selfish genes (0.00078 VAF vs. 0.0016 VAF, $p = 0.00019$, Wilcoxon rank-sum test), suggesting milder fitness advantages conferred by mutations outside the RTK-Ras pathway. Indeed, all detected candidate variants (with the except of *PPP1CB* p.P49R) occurred in genes located outside of the RTK-Ras signalling pathway, implying a broader mechanistic basis for selfish spermatogonial selection than previously recognized.

One unanticipated outcome of this study was that the incorporation of UMIs into rhPCR primers did not improve assay sensitivity, potentially due to UMI overwriting arising from their degenerate design and short (5 bp) UMI length. Future rhAmpSeq-based discovery screens for selfish mutations may therefore benefit from employing longer UMIs (≥8 bp) to overcome these

challenges. Furthermore, as this study only screened one transverse slice from each aged testis, future screens may also consider analysing the entire organ to determine whether individual mutational foci (such as those detected in *ADNP, PPP2R5D,* and *PACS1*) are detectable in other slices. Such an approach could enable a three-dimensional reconstruction of clonal expansions in seminiferous tubules in the testis.

Following the completion of this work, a preprint was published describing the application of NanoSeq, an error-corrected duplex-sequencing method with an estimated sensitivity of ~5 x 10^{-9}, on 81 sperm from 63 men (average age 53 years, range 24–75 years) alongside 119 matched blood samples to investigate positive selection in the male germline (Neville et al., 2024). Their analysis identified 40 genes under significant positive selection in the male germline, including 8 of the 12 canonical selfish genes in the RTK-Ras pathway. Notably, 2 of the 32 novel genes identified in their dataset were *SMAD4* and *CSNK2B* – the former which contained two significant hotspots at p.R496 and p.I500 (loci at which variants were identified in this work), and the latter which interacts with *CSNK2A1* to form the α catalytic subunit of CK2 (**section 3.3.10**). Moreover, four positively selected genes (*SMAD6, MIB1, LZTR1,* and *SSX1*) showed LOF enrichment in excess of gnomAD, suggesting that germline selection can outpace purifying selection at the population level. The findings by Neville et al., both reinforce and expand upon the findings made in the current discovery screen.

6.3 – *SMAD4* mutations at p.I500 undergo positive selection in sperm

Mutational foci were detected at two loci in *SMAD4*: p.R496 and p.I500, both of which are established in the pathogenesis of Myhre syndrome, a rare dominant disorder. To investigate whether MS *SMAD4* mutations are detectable in sperm, an ultra-sensitive assay, REDPCR, was

adapted to target the *NsiI* restriction site at *SMAD4* c.1494–1499, a 6 bp region spanning 3 codons: p.L498, p.C499, and p.I500. This assay demonstrated reliable and accurate quantification of all 18 possible substitutions at *SMAD4* c.1494–1499 to a level of 10^{-6}. A survey of 57 sperm (average age 40.7 years, range 27.2–75.9 years) and 5 blood samples (average age 39.5 years; range 24.8–57.0 years) identified substitutions encoding changes at p.I500 to be enriched in the sperm of most men. In particular, mutation levels for the two MS mutations covered by this assay, c.1498A>G (p.I500V) and c.1499T>C (p.I500T), reached high levels of ~10^{-4} and ~10^{-5}, respectively, and were positively correlated with the age of the donor, consistent with positive selection. By sampling sperm across a range of donor ages, the results of this *SMAD4* REDPCR survey validated the findings of MS *SMAD4* mutational foci in aged testis biopsies. Together, with results from collaborators that demonstrated an exclusive paternal origin of mutation in MS parent-offspring trios, as well as an epidemiological paternal age effect, established Myhre syndrome as a novel PAE disorder (Wood et al., 2024).

6.4 – Future directions

Given the strong evidence that *SMAD4* p.I500 mutations are positively selected in the male germline, future studies could explore whether additional BMP/TGF-β genes are similarly subject to selection, as well as clarify the role this signalling cascade plays in mediating SSC homeostasis. Indeed, the related BMP/TGF-β genes, *SMAD6* and *BMPR2B*, were also identified as undergoing selection in the NanoSeq study (Neville et al., 2024). Of particular relevance, two of the originally described disorders shown to exhibit paternal age effects, fibrodysplasia ossificans progressiva and Marfan syndrome (Risch et al., 1987), are caused by mutations in ACVR1 and FBN1, respectively, both of which are BMP/TGF-β pathway components.

In the same NanoSeq study, pathway analysis implicated several Wnt-related genes (*CSNK2B, MIB1, CCAR2*) in positive selection, complementing evidence in this thesis that *ADNP* may also undergo selfish selection. Since both BMP/TGF-β and Wnt signalling are implicated in SSC homeostasis (Guo et al., 2017), these findings underscore the potential value of pathway-focused discovery screen for selfish mutations.

6.5 – Concluding remarks

Paternal age at conception continues to rise, prompting heighted scrutiny of the ageing paternal germline and its contribution to public health. One such consequence is the anticipated increase in the *de novo* mutational load, and the corresponding rise in the incidence of selfishly driven diseases. As patient cohorts grow and molecular techniques offer both improved depth and breadth, the findings in this thesis provide a framework to identify and screen for novel genes and pathways that may be under selection in the male germline.

Appendix

Table S1 – Spectrum of *de novo SMAD4* mutations and parental age in Myhre syndrome

Reference	c.1486C>T (p.Arg496Cys)	c.1498A>G (p.Ile500Val)	c.1499T>C (p.Ile500Thr)	c.1500A>G (p.Ile500Met)	Paternal age(s) (years)	Maternal age(s) (years)
Myhre et al., 1981	-	-	-	-	37 38	26 34
Soljak et al., 1983	-	-	-	-	42	38
Garcia-Cruz et al., 1993	-	-	-	-	32 23	30 20
Rosser et al., 1995	-	-	-	-	-	-
Figuera, 1996.	-	-	-	-	-	-
Hopkin et al., 1998	-	-	-	-	-	-
Titomanlio et al., 2001	-	-	-	-	33	-
Whiteford et al., 2001	-	-	-	1	32	29
Burglen et al., 2003	-	-	-	-	40 28 38 43	34 - 32 35
Dávalos et al. 2003	-	-	-	-	49	39
Lopez-Cardona et al. 2004	-	-	-	-	39	28
van Steensel et al., 2005	-	-	-	-	-	-
Rulli et al., 2005	-	-	-	-	36	32
Becerra-Solano et al., 2008	-	-	-	-	43	42
Bachmann-Gagescu et al., 2011	-	-	-	-	-	-
McGowan *et al.,* 2011	-	-	-	-	46 37	-
Le Goff *et al* ., 2012	0	5	5	1	-	-
Caputo *et al* . 2012	0	5	3	0	35 34 38 36 31 33 28 36	-
Asakura et al., 2012	0	0	1	0	-	-
Lindor et al., 2012	0	1	1	0	-	-
Al Ageeli et al., 2012	0	0	0	0	35 31	32
Need et al., 2012	0	1	0	0	-	-
Picco et al., 2013	0	0	1	0	-	-
Piccolo et al., 2014	0	1	0	0	-	-

Reference	c.1486C>T (p.Arg496Cys)	c.1498A>G (p.Ile500Val)	c.1499T>C (p.Ile500Thr)	c.1500A>G (p.Ile500Met)	Paternal age(s) (years)	Maternal age(s) (years)
Michot et al., 2014	2	11	5	0	38 42 35 38 50 28 38 45 34 32 42 36	-
Caputo et al., 2014	1	0	0	0	22	22
Ishibashi et al., 2014	0	0	0	0	-	-
Kenis et al., 2014	1	0	0	0	-	-
Starr et al., 2015	0	2	3	0	29 37 30 26 28	29 29 28 26 25
Hawkes and Kini, 2015	0	0	1	0	38	27
Oldenburg et al., 2015	0	1	0	0	-	-
Lin et al., 2016	0	3	1	0	39 35 33 24	-
Garavelli *et al.* 2016	0	0	0	0	-	-
Bassett et al., 2016	0	1	0	0	-	-
Erdem et al., 2017	0	1	0	0	31	-
Nomura et al., 2017	0	1	0	0	-	-
Alagia e*t al.,* 2018	0	1	0	0	-	-
Vadeboncoeur and McHardy. 2018	0	0	1	0	-	-
Zheng *et al.,* 2019	0	0	1	0	38	35
Yu et al., 2019	0	3	3	0	33	28
Artemios et al., 2019	1	0	0	0	41	31
Gursoy et al., 2019	0	1	0	0	-	-
Lin et al., 2019	2	0	0	0	-	-
Alape *et al.,* 2020	0	1	0	0	-	-
Li *et al* ., 2020	0	2	0	0	-	-
Varenyiova et al., 2020	0	1	0	0	-	-
Kaplanis et al., 2020	9	21	1	0	-	-
Jensen *et al* ., 2020	0	0	1	0	-	-

Reference	c.1486C>T (p.Arg496Cys)	c.1498A>G (p.Ile500Val)	c.1499T>C (p.Ile500Thr)	c.1500A>G (p.Ile500Met)	Paternal age(s) (years)	Maternal age(s) (years)
Cappucio et al., 2021	0	1	0	0	-	-
Kandhaya-Pillai *et al* ., 2021	1	0	0	0	-	-
Lim et al., 2021	0	1	0	0	38	37
Li et al 2021	0	1	0	0	-	-
Jeon et al., 2021	0	0	1	0	32	34
Di Cesare et al., 2021	0	1	0	0	-	
Wu et al., 2021	0	1	0	0	-	
Inoeu et al., 2021	0	1	0	0	42	32
Kilci et al., 2022	0	1	0	0	26	23
Cappuccio et al., 2022	1	4	0	0	58 50	28 35
Catana et al., 2022	0	1	0	0	35	35
Yang et al., 2022	0	1	0	0	-	-
Alankarage et al., 2022	0	2	0	0	-	-
Starr et al., 2022	-	1	-	-	-	-
100000 Genomes Project (Genomics England)	2	6	-	-	-	-
gnomAD v3.0	2	1	-	-	-	-
Top MedBravo	1		-	-	-	-
denovoDB DDD., 2017	1	3	-	-	-	-
denovoDB Homsy et al., 2015.	-	1	-	-	-	-
denovoDB Krumm et al., 2015.	-	1	-	-	-	-
Summary data of *de novo* MS cases						
	c.1486C>T (p.Arg496Cys)	**c.1498A>G (p.Ile500Val)**	**c.1499T>C (p.Ile500Thr)**	**c.1500A>G (p.Ile500Met)**	**Mean paternal age (years ± SD)**	**Mean maternal age (years ± SD)**
# of cases	24	91	29	2	61	25
% of total cases	16.4%	62.3%	19.9%	1.4%	36.1 (2.5)	32.0 (0.6)

Table S2 – Non-reference allele fractions at SMAD4 c.1494–1499 from a dilution series of c.1498A>G (p.I500V)

Background	Sample ID	c.1498A>G spike-in level	c.1494A>C (p.L498F)	c.1494A>G (p.L498L)	c.1494A>T (p.L498F)	c.1495T>A (p.C499S)	c.1495T>C (p.C499R)	c.1495T>G (p.C499G)	c.1496G>A (p.C499Y)	c.1496G>C (p.C499S)	c.1496G>T (p.C499F)	c.1497C>A (p.C499X)	c.1497C>G (p.C499W)	c.1497C>T (p.C499C)	c.1498A>C (p.I500L)	c.1498A>G (p.I500V)	c.1498A>T (p.I500L)	c.1499T>A (p.I500K)	c.1499T>C (p.I500T)	c.1499T>G (p.I500R)
Blood	1063B	0	0.00%	0.02%	0.02%	0.00%	0.00%	0.00%	0.04%	0.00%	24.73%	60.34%	0.02%	0.10%	0.00%	0.02%	0.02%	0.02%	0.00%	14.68%
Blood	1063B	10^{-6}	0.02%	0.00%	0.02%	0.00%	0.06%	0.00%	0.00%	0.00%	55.77%	25.90%	0.00%	0.11%	0.02%	17.97%	0.02%	0.00%	0.06%	0.05%
Blood	1063B	3×10^{-6}	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%	34.44%	9.62%	0.02%	0.02%	0.00%	55.89%	0.00%	0.00%	0.00%	0.00%
Blood	1063B	10^{-5}	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	0.00%	12.18%	5.83%	0.01%	0.05%	0.00%	81.88%	0.01%	0.01%	0.00%	0.00%
Blood	1063B	3×10^{-5}	0.01%	0.00%	0.00%	0.01%	0.01%	0.00%	0.01%	0.00%	4.81%	8.01%	0.00%	1.37%	0.01%	85.71%	0.00%	0.00%	0.04%	0.01%
Blood	1063B	10^{-4}	0.00%	0.02%	0.00%	0.01%	0.02%	0.00%	0.00%	0.00%	1.28%	0.00%	0.00%	0.04%	0.00%	98.60%	0.00%	0.01%	0.00%	0.01%

Table S3 – Non-reference allele fractions at SMAD4 c.1494–1499 from a pilot screen of 12 sperm samples

Tissue	Sample ID	Age (Years)	c.1494A>C (p.L498F)	c.1494A>G (p.L498L)	c.1494A>T (p.L498F)	c.1495T>A (p.C499S)	c.1495T>C (p.C499R)	c.1495T>G (p.C499G)	c.1496G>A (p.C499Y)	c.1496G>C (p.C499S)	c.1496G>T (p.C499F)	c.1497C>A (p.C499X)	c.1497C>G (p.C499W)	c.1497C>T (p.C499C)	c.1498A>C (p.I500L)	c.1498A>G (p.I500V)	c.1498A>T (p.I500L)	c.1499T>A (p.I500K)	c.1499T>C (p.I500T)	c.1499T>G (p.I500R)
Sperm	716	29.94	0.00%	0.04%	0.02%	0.00%	0.06%	0.00%	1.68%	0.00%	36.03%	35.95%	0.02%	0.14%	0.00%	25.62%	0.00%	0.00%	0.44%	0.00%
Sperm	413	35.86	0.00%	0.00%	0.62%	0.93%	0.62%	0.31%	0.62%	0.31%	11.80%	18.94%	0.00%	2.17%	0.00%	47.83%	0.00%	0.31%	15.53%	0.00%
Sperm	406	36.38	0.00%	0.03%	0.05%	0.08%	0.05%	0.03%	7.23%	0.00%	52.15%	38.26%	0.03%	1.94%	0.00%	0.03%	0.03%	0.05%	0.03%	0.03%
Sperm	779	38.38	0.00%	0.03%	0.01%	0.01%	0.04%	0.00%	0.10%	0.01%	1.70%	6.05%	0.01%	2.84%	0.00%	71.93%	0.01%	0.01%	17.24%	0.00%
Sperm	409	42.09	0.00%	0.02%	0.00%	0.02%	0.00%	0.02%	2.93%	0.02%	10.70%	8.42%	0.00%	0.31%	0.02%	59.56%	0.02%	0.02%	17.95%	0.00%
Sperm	361	43.66	0.00%	0.07%	0.00%	0.02%	0.04%	0.00%	3.63%	0.00%	13.11%	23.28%	0.00%	3.37%	0.00%	56.45%	0.00%	0.00%	0.02%	0.00%
Sperm	307	45.78	0.00%	0.02%	0.00%	0.02%	0.00%	0.00%	0.55%	0.55%	31.25%	21.15%	0.00%	3.29%	0.00%	31.79%	0.00%	0.02%	11.36%	0.00%
Sperm	763	48.03	0.00%	0.02%	0.04%	0.00%	0.08%	0.00%	0.19%	0.00%	16.38%	24.12%	0.00%	0.21%	4.79%	49.82%	0.00%	0.00%	0.00%	4.33%
Sperm	766	50.27	0.00%	0.08%	0.04%	0.12%	0.04%	0.00%	6.02%	0.04%	37.13%	39.84%	0.00%	16.50%	0.00%	0.08%	0.04%	0.00%	0.08%	0.00%
Sperm	424	52.91	0.02%	0.02%	0.01%	0.00%	0.02%	0.00%	0.13%	0.00%	5.83%	3.15%	0.00%	0.12%	4.53%	66.71%	4.54%	0.00%	14.93%	0.00%
Sperm	788	56.21	0.00%	0.07%	0.02%	0.05%	0.05%	0.00%	0.05%	0.00%	25.10%	18.80%	0.00%	0.54%	0.00%	40.85%	0.02%	0.00%	14.45%	0.00%
Sperm	278	60.59	0.00%	0.00%	0.02%	0.00%	0.13%	0.00%	0.07%	0.02%	63.69%	22.02%	0.00%	0.62%	0.00%	0.04%	0.02%	0.02%	13.33%	0.00%
EBV	987	49	0.00%	0.07%	0.00%	0.00%	0.03%	0.00%	0.03%	0.00%	42.42%	56.99%	0.00%	0.38%	0.00%	0.00%	0.00%	0.03%	0.03%	0.00%

Table S4A – Estimated mutation levels (per million) at SMAD4 c.1494-1499 from a titration-reconstruction experiment of serially diluted MS DNA (c.1498A>G and c.1499T>C)

Background	Sample ID	MS DNA spike-in level	*NsiI* -integrant spike-in level	c.1494del # reads	Coverage	c.1494A>C (p.L498F)	c.1494A>G (p.L498L)	c.1494A>T (p.L498F)	c.1495T>A (p.C499S)	c.1495T>C (p.C499R)	c.1495T>G (p.C499G)	c.1496G>A (p.C499Y)	c.1496G>C (p.C499S)	c.1496G>T (p.C499F)	c.1497C>A (p.C499X)	c.1497C>G (p.C499W)	c.1497C>T (p.C499C)	c.1498A>C (p.I500L)	c.1498A>G (p.I500V)	c.1498A>T (p.I500L)	c.1499T>A (p.I500K)	c.1499T>C (p.I500T)	c.1499T>G (p.I500R)
Blood	1063C	0	3×10^{-5}	3006	4347	0.00	0.03	0.00	0.01	0.11	0.00	0.07	0.02	0.48	1.38	0.00	0.04	0.01	0.12	0.00	0.00	0.06	0.00
Blood	1197	0	3×10^{-5}	2049	2174	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.02	0.00	0.52	0.00	0.10	0.00	0.03	0.00	0.00	0.03	0.00
Blood	1063C	10^{-6}	3×10^{-5}	3370	4286	0.01	0.01	0.00	0.00	0.14	0.00	0.03	0.01	1.09	2.10	0.00	0.06	0.00	4.16	0.02	0.02	0.03	0.00
Blood	1197	10^{-6}	3×10^{-5}	1274	1720	0.00	0.00	0.00	0.00	0.13	0.00	0.03	0.03	1.86	0.34	0.00	0.00	0.00	0.05	0.05	0.00	3.46	0.00
Blood	1063C	3×10^{-6}	3×10^{-5}	2534	4173	0.00	0.00	0.00	0.00	0.12	0.01	0.04	0.00	2.51	4.98	0.00	0.03	0.00	5.84	0.00	0.01	5.49	0.00
Blood	1197	3×10^{-6}	3×10^{-5}	1392	1785	0.00	0.91	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.93	0.00	0.00	0.00	1.99	0.00	0.02	3.24	0.00
Blood	1063C	10^{-5}	3×10^{-5}	2213	3554	0.00	0.02	0.00	0.00	0.11	0.02	0.03	0.00	1.00	1.73	0.02	0.08	0.95	7.10	0.00	0.00	7.57	0.00
Blood	1197	10^{-5}	3×10^{-5}	1201	2104	0.00	0.00	0.00	0.00	0.08	0.00	0.06	0.00	0.47	1.03	0.03	0.03	0.03	11.08	0.00	0.00	10.17	0.00
Blood	1063C	3×10^{-5}	3×10^{-5}	1563	4839	0.00	0.02	0.02	0.02	0.15	0.04	0.06	0.02	0.00	3.82	0.00	0.06	0.04	27.70	0.00	0.00	35.87	0.00
Blood	1197	3×10^{-5}	3×10^{-5}	1494	2811	0.00	0.11	0.02	0.02	0.13	0.00	0.04	0.02	0.62	0.00	0.00	0.04	0.00	11.72	0.00	0.02	15.15	0.00
Blood	1063C	10^{-4}	3×10^{-5}	638	3272	0.05	0.16	0.00	0.00	0.21	0.00	0.00	0.05	2.50	2.09	0.00	0.10	0.10	59.77	0.00	0.00	71.36	0.00
Blood	1197	10^{-4}	3×10^{-5}	515	2316	0.00	1.10	0.00	0.00	0.13	0.00	0.19	0.00	0.13	0.45	0.06	0.13	0.13	63.28	0.00	0.00	49.45	0.00

Table S4B: Estimated mutation levels (per million) at SMAD4 c.1494-1499 from a titration-reconstruction experiment of serially diluted NsiI-resistant genomic integrant (c.1478-c.1494del)

Background	Sample ID	MS DNA spike-in level	*NsiI* - integrant spike-in level	c.1498A>G # reads	Coverage	c.1494A>C (p.L498F)	c.1494A>G (p.L498L)	c.1494A>T (p.L498F)	c.1495T>A (p.C499S)	c.1495T>C (p.C499R)	c.1495T>G (p.C499G)	c.1496G>A (p.C499Y)	c.1496G>C (p.C499S)	c.1496G>T (p.C499F)	c.1497C>A (p.C499X)	c.1497C>G (p.C499W)	c.1497C>T (p.C499C)	c.1498A>C (p.I500L)	c.1498A>T (p.I500L)	c.1499T>A (p.I500K)	c.1499T>G (p.I500R)	c.1494del *NsiI* - integrant
EBV	985	3×10^{-5}	0	4805	9669	0.01	0.07	0.01	0.03	0.13	0.10	0.09	0.01	0.57	2.54	0.01	0.11	0.01	0.03	0.01	0.02	0.08
EBV	985	3×10^{-5}	0	6054	12159	0.01	0.10	0.02	0.00	0.06	0.08	0.06	0.00	0.57	2.64	0.02	0.10	0.02	0.01	0.00	0.03	0.07
EBV	985	3×10^{-5}	10^{-6}	8244	15540	0.00	0.08	0.00	0.00	0.11	0.11	0.07	0.00	2.45	2.79	0.01	0.08	0.02	0.00	0.01	0.02	1.09
EBV	985	3×10^{-5}	10^{-6}	6285	11815	0.01	0.08	0.01	0.01	0.14	0.14	0.07	0.01	2.38	2.89	0.01	0.05	0.01	0.01	0.02	0.03	1.33
EBV	985	3×10^{-5}	3×10^{-6}	7125	14504	0.00	0.10	0.00	0.00	0.16	0.12	0.05	0.00	0.01	1.11	0.02	0.12	0.01	0.01	0.00	0.04	3.40
EBV	985	3×10^{-5}	3×10^{-6}	5933	11839	0.02	0.07	0.01	0.01	0.11	0.13	0.05	0.00	0.01	0.95	0.01	0.05	0.02	0.00	0.00	0.03	3.59
EBV	985	3×10^{-5}	10^{-5}	6704	13871	0.00	0.08	0.01	0.86	0.14	0.13	0.04	0.00	2.72	0.01	0.01	0.10	0.03	0.00	0.01	0.03	9.27
EBV	985	3×10^{-5}	10^{-5}	5996	12375	0.00	0.08	0.01	0.89	0.08	0.09	0.07	0.01	2.62	0.02	0.01	0.12	0.01	0.01	0.00	0.03	9.53
EBV	985	3×10^{-5}	3×10^{-5}	5955	14734	0.00	0.06	0.01	0.01	0.15	0.07	0.11	0.00	2.95	1.35	0.01	0.07	0.00	0.02	0.01	0.02	21.85
EBV	985	3×10^{-5}	3×10^{-5}	5145	12820	0.01	0.15	0.01	0.01	0.19	0.11	0.09	0.01	2.80	1.18	0.00	0.05	0.00	0.01	0.00	0.05	22.50
EBV	985	3×10^{-5}	10^{-4}	3826	17690	0.03	0.12	0.00	0.00	0.25	0.24	0.19	0.00	1.90	0.94	0.01	0.05	0.01	0.05	0.01	0.03	89.80
EBV	985	3×10^{-5}	10^{-4}	2543	11605	0.01	0.16	0.03	0.00	0.27	0.13	0.09	0.03	1.60	0.82	0.00	0.13	0.01	0.03	0.00	0.00	90.38

Table S5: Estimated mutation levels (per million) at SMAD4 c.1494-1499 from a survey of 2 – 4 sperm samples from 4 individuals

Individual	Sample ID	Replicate	Tissue	Age (Years)	c.1494del # reads	Coverage	c.1494A>C (p.L498F)	c.1494A>G (p.L498L)	c.1494A>T (p.L498F)	c.1495T>A (p.C499S)	c.1495T>C (p.C499R)	c.1495T>G (p.C499G)	c.1496G>A (p.C499Y)	c.1496G>C (p.C499S)	c.1496G>T (p.C499F)	c.1497C>A (p.C499X)	c.1497C>G (p.C499W)	c.1497C>T (p.C499C)	c.1498A>C (p.I500L)	c.1498A>G (p.I500V)	c.1498A>T (p.I500L)	c.1499T>A (p.I500K)	c.1499T>C (p.I500T)	c.1499T>G (p.I500R)
14	398	1	Sperm	39.35	480	2245	0.07	0.28	0.00	0.07	0.42	0.00	1.53	0.21	13.95	13.10	0.00	2.36	0.07	33.80	0.07	0.00	0.21	0.00
14	398	2	Sperm	39.35	307	1715	0.00	0.11	0.00	0.00	0.33	0.00	4.34	0.00	4.02	10.10	0.00	5.54	0.00	45.87	0.00	0.00	0.00	0.00
14	398	3	Sperm	39.35	798	3245	0.00	0.08	0.00	0.00	0.13	0.00	2.13	0.00	6.35	7.06	0.04	3.05	0.17	47.30	0.04	0.04	0.08	0.00
14	400	1	Sperm	39.37	609	2063	0.00	0.00	0.00	0.00	0.38	0.00	4.33	0.00	5.31	2.90	0.00	1.53	0.16	29.67	0.00	0.00	0.16	0.00
14	400	2	Sperm	39.37	615	1938	0.00	0.16	0.00	0.00	0.11	0.00	1.41	0.00	6.01	7.53	0.00	2.22	0.00	19.44	0.00	0.00	1.79	0.00
14	400	3	Sperm	39.37	1609	6677	0.00	0.08	0.00	0.04	0.25	0.00	3.09	0.04	7.92	10.07	0.04	2.03	0.10	37.10	0.04	0.00	2.26	0.00
14	778	1	Sperm	42.37	809	2325	0.00	0.00	0.00	0.04	0.21	0.00	0.16	0.00	1.89	2.68	0.00	0.16	0.04	30.43	0.00	0.04	1.52	0.04
14	778	2	Sperm	42.37	539	2148	0.00	0.06	0.00	0.06	0.56	0.00	1.24	0.00	5.44	6.18	0.00	0.31	0.00	52.09	0.06	0.00	1.67	0.00
14	778	3	Sperm	42.37	1848	4669	0.00	0.11	0.00	0.00	0.16	0.00	1.33	0.00	0.11	1.99	0.02	0.42	0.02	25.33	0.02	0.02	2.67	0.00
14	780	1	Sperm	42.39	416	1393	0.00	0.08	0.00	0.00	0.08	0.00	0.00	0.00	3.20	4.00	0.00	0.16	0.00	36.89	0.16	0.08	0.88	0.08
14	780	2	Sperm	42.39	594	1965	0.00	0.00	0.06	0.00	0.17	0.00	0.17	0.00	4.61	4.44	0.00	0.28	0.06	32.24	0.06	0.00	4.45	0.00
14	780	3	Sperm	42.39	2201	6969	0.00	0.18	0.00	0.00	0.15	0.00	1.98	0.03	4.80	3.19	0.02	0.29	0.00	33.91	0.03	0.00	4.92	0.00
17	415	1	Sperm	56.49	999	2184	0.00	0.03	0.00	0.00	0.10	0.00	0.20	0.00	6.34	3.07	0.00	0.03	0.07	13.67	0.00	0.03	0.17	0.00
17	415	2	Sperm	56.49	947	1827	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	2.92	0.95	0.04	0.25	0.04	9.52	0.00	0.00	0.07	0.00
17	415	3	Sperm	56.49	1067	2082	0.00	0.03	0.00	0.00	0.06	0.00	0.06	0.00	7.44	3.43	0.03	0.12	0.03	6.00	0.00	0.00	0.06	0.00
17	416	1	Sperm	56.51	913	2045	0.00	0.04	0.00	0.00	0.18	0.00	0.15	0.04	0.07	0.04	0.04	0.11	0.11	18.61	0.00	0.00	0.18	0.00
17	416	2	Sperm	56.51	1051	2029	0.00	0.00	0.00	0.00	0.13	0.00	0.10	0.00	1.39	1.05	0.00	0.06	2.51	9.02	0.00	0.00	0.06	0.00
17	416	3	Sperm	56.51	2008	5052	0.00	0.05	0.00	0.00	0.12	0.00	1.29	0.00	1.36	6.66	0.03	1.10	0.02	21.49	0.02	0.00	0.03	0.00
17	799	1	Sperm	59.48	2845	6220	0.00	0.04	0.01	0.00	0.04	0.00	0.07	0.00	1.03	3.41	0.00	0.15	0.00	21.54	0.00	0.00	0.08	0.01
17	799	2	Sperm	59.48	1681	3897	0.00	0.06	0.00	0.00	0.22	0.02	0.06	0.00	4.47	7.58	0.00	0.14	0.00	11.50	0.04	0.00	0.08	0.00
17	799	3	Sperm	59.48	553	1424	0.00	0.00	0.06	0.00	0.24	0.00	1.99	0.00	3.32	0.00	0.00	0.24	0.00	16.67	0.06	0.06	0.18	0.00
17	801	1	Sperm	59.53	463	3656	0.00	0.22	0.00	0.00	0.72	0.07	0.36	0.07	0.22	1.94	0.07	0.14	0.14	27.58	0.07	0.00	0.43	0.00
17	801	2	Sperm	59.53	1563	3373	0.00	0.06	0.02	0.00	0.09	0.09	0.04	0.02	0.06	4.18	0.00	1.81	0.02	15.79	0.00	0.02	2.37	0.00
17	801	3	Sperm	59.53	834	1557	0.00	0.04	0.00	0.00	0.04	0.00	0.40	0.00	4.83	0.04	0.00	0.08	0.04	8.15	0.08	0.00	2.60	0.00
27	565	1	Sperm	41.70	1674	3890	0.00	0.04	0.00	0.00	0.08	0.00	0.08	0.02	5.26	1.04	0.02	0.08	7.05	6.67	0.06	0.00	1.91	0.00
27	565	2	Sperm	41.70	1429	2825	0.00	0.02	0.00	0.00	0.09	0.00	0.16	0.00	3.45	0.00	0.00	0.07	0.02	3.97	0.02	0.00	0.07	0.00
27	565	3	Sperm	41.70	700	1280	0.00	0.00	0.00	0.00	0.10	0.00	0.62	0.00	2.57	0.10	0.00	0.10	1.53	8.96	0.05	0.00	0.14	0.05
27	566	1	Sperm	41.72	263	1766	0.00	0.25	0.00	0.00	0.25	0.00	0.13	0.00	0.13	0.51	0.00	1.27	0.13	0.89	0.00	0.00	31.81	0.13
27	566	2	Sperm	41.72	1371	2711	0.00	0.07	0.02	0.00	0.12	0.00	0.12	0.05	3.84	1.68	0.00	0.19	0.00	6.90	0.07	0.02	1.63	0.00
27	566	3	Sperm	41.72	660	1311	0.00	0.00	0.00	0.00	0.00	0.00	0.76	0.00	2.27	0.40	0.05	1.61	1.31	11.83	0.00	0.00	1.36	0.00
21	809	1	Sperm	55.92	1765	5071	0.00	0.04	0.02	0.00	0.04	0.00	0.11	0.02	0.04	1.47	0.00	0.08	9.17	37.36	7.72	0.00	0.02	0.02
21	809	2	Sperm	55.92	595	3438	0.00	0.22	0.00	0.00	0.22	0.00	0.17	0.06	7.22	5.48	0.17	0.28	6.95	108.46	2.86	0.00	5.76	0.00
21	809	3	Sperm	55.92	346	1778	0.00	0.19	0.00	0.00	0.10	0.00	0.10	0.00	0.19	1.64	0.00	0.10	10.48	105.60	5.29	0.00	2.50	0.00
21	810	1	Sperm	55.93	368	2159	0.00	0.27	0.00	0.00	0.27	0.00	0.00	0.09	2.71	6.24	0.00	0.18	3.72	94.34	23.38	0.00	15.90	0.00
21	810	2	Sperm	55.93	934	4643	0.00	0.07	0.00	0.00	0.25	0.00	0.11	0.00	1.67	3.82	0.00	0.25	4.00	90.11	16.05	0.00	2.50	0.00
21	810	3	Sperm	55.93	710	1918	0.00	0.09	0.00	0.00	0.09	0.00	0.05	0.00	2.07	1.69	0.00	0.05	3.25	39.10	4.66	0.00	0.05	0.00

Table S6: Estimated mutation levels (per million) at SMAD4 c.1494_1499 from a survey of 57 sperm and 5 blood samples

Sample ID	Replicate	Tissue	Age (Years)	c.1494del # reads	Coverage	c.1494A>C (p.L498F)	c.1494A>G (p.L498L)	c.1494A>T (p.L498F)	c.1495T>A (p.C499S)	c.1495T>C (p.C499R)	c.1495T>G (p.C499G)	c.1496G>A (p.C499Y)	c.1496G>C (p.C499S)	c.1496G>T (p.C499F)	c.1497C>A (p.C499X)	c.1497C>G (p.C499W)	c.1497C>T (p.C499C)	c.1498A>C (p.I500L)	c.1498A>G (p.I500V)	c.1498A>T (p.I500L)	c.1499T>A (p.I500K)	c.1499T>C (p.I500T)	c.1499T>G (p.I500R)
1047	1	Sperm	27.22	4385	4415	0.00	0.03	0.00	0.00	0.07	0.00	0.05	0.00	0.01	0.02	0.01	0.11	0.00	0.03	0.00	0.01	0.06	0.00
1047	2	Sperm	27.22	5544	5565	0.01	0.00	0.00	0.00	0.04	0.00	0.05	0.01	0.02	0.00	0.01	0.07	0.00	0.07	0.01	0.00	0.04	0.00
1047	3	Sperm	27.22	4611	4642	0.00	0.04	0.01	0.00	0.05	0.00	0.04	0.00	0.01	0.01	0.01	0.07	0.01	0.04	0.00	0.00	0.06	0.00
1050	1	Sperm	27.33	33836	41004	0.00	0.03	0.00	0.00	0.10	0.00	0.36	0.01	1.27	3.26	0.01	0.78	0.00	0.11	0.01	0.00	0.04	0.00
1050	2	Sperm	27.33	20818	27332	0.00	0.07	0.00	0.01	0.12	0.00	0.40	0.01	4.12	3.33	0.01	0.82	0.00	0.16	0.01	0.01	0.08	0.00
1050	3	Sperm	27.33	23502	29654	0.00	0.03	0.00	0.00	0.08	0.00	0.11	0.01	2.26	4.40	0.01	0.60	0.01	0.12	0.01	0.01	0.07	0.00
1037	1	Sperm	27.98	5663	8185	0.00	0.03	0.02	0.00	0.11	0.00	0.13	0.03	5.63	4.31	0.01	2.26	0.00	2.17	0.04	0.02	0.08	0.00
1037	2	Sperm	27.98	6979	9285	0.00	0.05	0.00	0.00	0.07	0.00	0.27	0.01	2.34	4.39	0.01	0.22	0.00	1.32	0.01	0.01	2.00	0.00
1037	3	Sperm	27.98	6298	8819	0.00	0.02	0.00	0.01	0.10	0.00	0.98	0.01	2.79	5.43	0.01	1.70	0.00	1.88	0.01	0.00	0.08	0.00
935	1	Sperm	28.70	3887	4095	0.00	0.03	0.03	0.01	0.12	0.00	0.06	0.03	1.24	0.08	0.00	0.21	0.00	0.05	0.01	0.00	0.04	0.00
935	2	Sperm	28.70	4929	6490	0.00	0.02	0.01	0.00	0.05	0.01	0.11	0.00	0.06	9.95	0.02	0.20	0.01	0.14	0.02	0.01	0.07	0.00
935	3	Sperm	28.70	126	134	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
901	1	Sperm	30.26	30292	44987	0.00	0.03	0.00	0.00	0.07	0.00	0.48	0.01	3.98	3.76	0.00	0.39	0.00	6.64	0.01	0.01	0.05	0.00
901	2	Sperm	30.26	33122	45325	0.00	0.01	0.00	0.00	0.07	0.00	0.57	0.01	3.63	2.72	0.00	0.65	0.00	2.97	0.00	0.00	0.95	0.00
901	3	Sperm	30.26	26908	41602	0.00	0.03	0.00	0.00	0.08	0.00	0.49	0.01	5.57	5.25	0.01	0.50	0.01	5.47	0.01	0.00	0.05	0.01
921	1	Sperm	31.49	5551	10541	0.00	0.04	0.00	0.01	0.09	0.00	0.10	0.01	5.22	2.40	0.01	1.25	0.02	12.80	0.01	0.00	8.23	0.00
921	2	Sperm	31.49	4621	9239	0.00	0.07	0.00	0.00	0.12	0.00	0.10	0.02	4.21	5.51	0.01	1.61	0.01	11.43	0.01	0.01	10.27	0.00
921	3	Sperm	31.49	6609	9859	0.00	0.03	0.01	0.00	0.07	0.00	0.07	0.01	4.59	4.76	0.02	0.27	0.00	4.89	0.01	0.00	1.82	0.00
923	1	Sperm	31.64	4749	8478	0.00	0.05	0.01	0.00	0.12	0.00	0.92	0.01	6.94	13.53	0.02	2.36	0.03	0.08	1.95	0.00	0.06	0.00
923	2	Sperm	31.64	5333	9272	0.00	0.14	0.00	0.01	0.15	0.00	2.67	0.02	8.81	9.70	0.02	0.72	0.04	0.16	2.15	0.01	0.09	0.00
923	3	Sperm	31.64	6774	11512	0.00	0.12	0.00	0.00	0.09	0.00	2.58	0.03	8.30	8.96	0.02	0.70	0.02	0.16	2.28	0.00	0.07	0.00
1010	1	Sperm	32.10	26536	32640	0.00	0.03	0.00	0.00	0.08	0.00	0.07	0.01	4.43	2.50	0.01	0.08	0.00	0.10	0.01	0.00	0.04	0.00
1010	2	Sperm	32.10	28324	34204	0.00	0.02	0.00	0.00	0.08	0.00	0.08	0.02	5.46	0.73	0.00	0.07	0.01	0.12	0.00	0.00	0.04	0.00
1010	3	Sperm	32.10	37155	45593	0.00	0.03	0.00	0.00	0.08	0.00	0.07	0.01	3.98	2.61	0.00	0.09	0.00	0.10	0.01	0.00	0.06	0.00
950	1	Sperm	32.12	7546	9502	0.00	0.02	0.00	0.00	0.05	0.00	0.07	0.00	3.72	2.32	0.01	1.01	0.00	1.55	0.02	0.00	0.07	0.00
950	2	Sperm	32.12	4831	7212	0.00	0.01	0.00	0.00	0.08	0.01	0.03	0.02	6.78	9.38	0.01	0.10	0.00	0.10	0.01	0.00	0.06	0.00
950	3	Sperm	32.12	9839	11257	0.00	0.01	0.01	0.00	0.09	0.01	0.30	0.00	1.28	1.83	0.00	0.03	0.00	1.45	0.00	0.00	0.02	0.00
1030	1	Sperm	32.49	17809	28040	0.00	0.06	0.01	0.00	0.13	0.00	0.13	0.00	3.76	7.97	0.02	0.98	0.00	2.91	1.46	0.01	0.08	0.01
1030	2	Sperm	32.49	27950	34026	0.00	0.05	0.00	0.00	0.10	0.00	0.14	0.00	1.77	2.40	0.00	0.43	0.01	1.23	0.01	0.00	0.07	0.01
1030	3	Sperm	32.49	19109	25884	0.00	0.05	0.01	0.00	0.10	0.01	0.72	0.00	2.74	5.19	0.00	0.16	0.01	0.14	1.31	0.00	0.07	0.00
920	1	Sperm	32.58	6140	8037	0.01	0.02	0.00	0.01	0.08	0.00	0.09	0.01	3.44	3.45	0.01	0.14	0.01	3.08	0.01	0.01	0.09	0.00
920	2	Sperm	32.58	5118	6912	0.00	0.02	0.00	0.00	0.11	0.01	0.10	0.00	6.77	4.38	0.02	0.16	0.00	0.12	0.01	0.00	0.06	0.00
920	3	Sperm	32.58	7841	10665	0.00	0.04	0.01	0.00	0.06	0.00	0.12	0.00	3.51	5.08	0.01	0.61	0.01	2.62	0.01	0.00	0.06	0.00
1027	1	Sperm	33.24	18421	27746	0.01	0.86	0.00	0.00	0.12	0.01	0.79	0.01	2.45	1.13	0.01	0.16	0.02	9.64	0.02	0.01	0.09	0.00
1027	2	Sperm	33.24	12096	26718	0.00	0.11	0.00	0.01	0.18	0.01	0.22	0.02	4.03	4.38	0.02	0.26	0.03	28.43	0.04	0.01	0.13	0.02
1027	3	Sperm	33.24	21962	35578	0.00	0.05	0.00	0.01	0.11	0.00	0.59	0.00	3.07	2.29	0.00	0.57	0.01	12.70	0.01	0.01	0.05	0.01
904	1	Sperm	33.51	25784	34356	0.00	0.04	0.00	0.00	0.09	0.00	0.20	0.01	2.33	2.41	0.01	1.62	0.01	3.15	0.01	0.00	0.07	0.00
904	2	Sperm	33.51	21985	27747	0.00	0.05	0.00	0.00	0.10	0.00	0.09	0.00	2.19	1.56	0.01	2.49	0.01	0.97	0.01	0.00	0.06	0.00
904	3	Sperm	33.51	23033	26855	0.00	0.04	0.00	0.00	0.08	0.00	0.38	0.01	1.56	1.92	0.01	0.50	0.01	0.11	0.01	0.00	0.03	0.00
1039	1	Sperm	33.75	33666	42375	0.00	0.03	0.00	0.00	0.08	0.00	0.51	0.00	2.66	2.66	0.00	0.38	0.00	1.06	0.01	0.00	0.06	0.00
1039	2	Sperm	33.75	19717	22785	0.00	0.06	0.00	0.00	0.13	0.00	0.61	0.00	1.75	0.96	0.00	0.20	0.00	0.16	0.01	0.00	0.07	0.00
1039	3	Sperm	33.75	23203	34167	0.00	0.03	0.00	0.00	0.09	0.00	0.09	0.00	2.48	4.45	0.01	0.49	0.02	5.87	0.92	0.00	0.07	0.01
888	1	Sperm	33.88	31201	40090	0.00	0.02	0.00	0.00	0.08	0.00	0.44	0.01	2.68	4.14	0.01	0.46	0.00	1.01	0.01	0.01	0.05	0.01
888	2	Sperm	33.88	26816	34163	0.00	0.03	0.00	0.00	0.10	0.00	0.79	0.02	4.21	3.13	0.00	0.09	0.01	0.11	0.01	0.00	0.05	0.00
888	3	Sperm	33.88	30589	36393	0.00	0.03	0.00	0.00	0.08	0.00	0.16	0.01	1.03	1.69	0.00	0.11	0.00	2.52	0.01	0.00	0.04	0.00
1052	1	Sperm	34.81	27402	32697	0.00	0.04	0.00	0.00	0.09	0.00	1.31	0.01	0.92	1.85	0.00	0.78	0.00	0.87	0.00	0.00	0.04	0.00
1052	2	Sperm	34.81	28879	36072	0.00	0.03	0.00	0.00	0.09	0.00	0.08	0.01	2.58	3.20	0.01	0.25	0.00	1.41	0.01	0.00	0.05	0.00
1052	3	Sperm	34.81	29720	37451	0.00	0.03	0.00	0.00	0.08	0.00	1.63	0.00	2.25	0.01	0.00	0.52	0.01	3.21	0.01	0.00	0.05	0.00
1055	1	Sperm	35.05	16960	41781	0.00	0.04	0.00	0.00	0.11	0.00	0.83	0.01	2.17	3.83	0.01	0.11	0.01	39.28	1.01	0.01	0.07	0.00
1055	2	Sperm	35.05	22802	63477	0.00	0.06	0.01	0.01	0.13	0.00	0.73	0.01	1.06	1.29	0.00	0.42	1.97	51.57	0.75	0.01	0.10	0.02
1055	3	Sperm	35.05	26734	58804	0.00	0.05	0.00	0.00	0.09	0.00	0.37	0.00	1.49	1.63	0.01	0.12	1.78	28.36	5.31	0.00	0.06	0.01
1046	1	Sperm	35.11	32891	41754	0.00	0.04	0.00	0.00	0.13	0.00	0.08	0.01	3.41	1.72	0.00	0.58	0.01	1.78	0.01	0.00	0.06	0.00
1046	2	Sperm	35.11	26003	33250	0.00	0.04	0.00	0.00	0.08	0.00	0.44	0.01	3.05	2.46	0.01	0.44	0.01	0.99	0.01	0.00	0.62	0.00
1046	3	Sperm	35.11	30586	40123	0.00	0.04	0.00	0.00	0.08	0.00	0.49	0.00	1.95	3.22	0.00	0.58	0.00	2.78	0.01	0.00	0.04	0.00
925	1	Sperm	35.35	6789	18081	0.00	0.05	0.01	0.00	0.14	0.00	0.12	0.01	5.06	3.48	0.00	0.16	0.01	35.00	0.02	0.00	6.59	0.00
925	2	Sperm	35.35	4865	10537	0.00	0.02	0.00	0.00	0.07	0.00	0.07	0.01	2.21	5.67	0.01	0.15	0.01	24.29	0.00	0.00	1.87	0.00
925	3	Sperm	35.35	10040	17999	0.00	0.03	0.00	0.00	0.11	0.00	1.12	0.00	3.61	4.48	0.01	0.13	0.00	14.50	0.01	0.00	0.09	0.01
1038	1	Sperm	35.76	15117	23884	0.00	0.45	0.01	0.01	0.20	0.01	0.30	0.01	4.63	3.17	0.01	1.02	0.02	4.93	0.03	0.01	2.80	0.00
1038	2	Sperm	35.76	14402	20840	0.00	0.13	0.00	0.01	0.26	0.01	0.25	0.01	1.82	2.97	0.00	1.08	0.02	3.33	0.01	0.01	2.76	0.00
1038	3	Sperm	35.76	8442	14635	0.01	0.13	0.02	0.02	0.30	0.01	1.25	0.01	3.96	5.63	0.02	0.77	0.03	2.39	0.05	0.02	6.82	0.00
915	1	Sperm	36.39	20336	22674	0.00	0.05	0.00	0.00	0.08	0.00	0.09	0.00	1.14	1.11	0.01	0.11	0.01	0.08	0.00	0.01	0.04	0.00
915	2	Sperm	36.39	31249	32787	0.00	0.04	0.00	0.00	0.07	0.00	0.09	0.00	0.02	0.02	0.00	0.13	0.00	0.10	0.00	0.00	0.05	0.00
915	3	Sperm	36.39	28310	30969	0.00	0.03	0.00	0.00	0.07	0.00	0.09	0.00	0.02	1.51	0.00	0.10	0.00	0.10	0.00	0.01	0.04	0.00

Sample ID	Replicate	Tissue	Age (Years)	c.1494del # reads	Coverage	c.1494A>C (p.L498F)	c.1494A>G (p.L498L)	c.1494A>T (p.L498F)	c.1495T>A (p.C499S)	c.1495T>C (p.C499R)	c.1495T>G (p.C499G)	c.1496G>A (p.C499Y)	c.1496G>C (p.C499S)	c.1496G>T (p.C499F)	c.1497C>A (p.C499X)	c.1497C>G (p.C499W)	c.1497C>T (p.C499C)	c.1498A>C (p.I500L)	c.1498A>G (p.I500V)	c.1498A>T (p.I500L)	c.1499T>A (p.I500K)	c.1499T>C (p.I500T)	c.1499T>G (p.I500R)
1025	1	Sperm	36.93	6500	7760	0.01	0.01	1.56	0.00	0.09	0.00	0.08	0.01	2.73	1.77	0.01	0.10	0.00	0.10	0.03	0.00	0.08	0.00
1025	2	Sperm	36.93	7768	9595	0.00	0.01	0.01	0.00	0.07	0.00	0.65	0.01	3.28	2.86	0.00	0.78	0.00	0.08	0.03	0.01	0.05	0.00
1025	3	Sperm	36.93	7850	10127	0.00	0.01	0.00	0.00	0.08	0.00	0.07	0.02	6.59	1.86	0.00	1.04	0.00	0.07	0.00	0.01	0.02	0.00
1023	1	Sperm	36.99	4499	7292	0.00	0.04	0.01	0.01	0.13	0.01	0.10	0.02	6.78	2.97	0.01	2.02	0.00	8.83	0.02	0.00	0.09	0.01
1023	2	Sperm	36.99	4699	7691	0.01	0.02	0.01	0.00	0.06	0.00	0.07	0.01	10.45	3.52	0.00	0.11	0.01	7.19	0.01	0.02	0.02	0.01
1023	3	Sperm	36.99	4824	6677	0.00	0.02	0.01	0.00	0.11	0.00	0.08	0.02	4.13	5.25	0.01	0.06	0.00	3.48	0.01	0.00	0.03	0.00
1056	1	Sperm	37.10	5753	6301	0.00	0.01	0.01	0.00	0.09	0.00	1.10	0.01	0.00	1.68	0.01	0.11	0.01	0.06	0.00	0.01	0.07	0.00
1056	2	Sperm	37.10	7062	8205	0.00	0.00	0.00	0.00	0.06	0.00	0.37	0.01	0.72	3.07	0.00	0.77	0.00	0.09	0.01	0.00	0.05	0.01
1056	3	Sperm	37.10	8050	9469	0.00	0.03	0.01	0.00	0.05	0.00	0.92	0.01	3.78	0.93	0.01	0.08	0.00	0.08	0.01	0.00	0.05	0.00
869	1	Sperm	37.25	25925	27264	0.00	0.03	0.00	0.00	0.06	0.00	0.10	0.00	0.03	0.02	0.00	0.47	0.00	0.09	0.00	0.00	0.04	0.00
869	2	Sperm	37.25	16753	17686	0.00	0.04	0.01	0.00	0.06	0.00	0.12	0.00	0.02	0.03	0.00	0.12	0.00	0.14	0.00	0.00	0.04	0.00
869	3	Sperm	37.25	13692	14989	0.00	0.07	0.00	0.00	0.11	0.00	0.17	0.01	0.03	0.90	0.00	0.19	0.00	0.12	0.00	0.00	0.10	0.00
885	1	Sperm	37.34	31396	37346	0.00	0.01	0.00	0.00	0.06	0.00	0.06	0.01	3.42	1.16	0.00	1.15	0.00	0.07	0.00	0.00	0.04	0.00
885	2	Sperm	37.34	3	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
885	3	Sperm	37.34	33508	42723	0.00	0.03	0.00	0.00	0.06	0.00	0.16	0.01	2.36	3.17	0.00	1.04	0.00	1.76	0.00	0.00	0.04	0.00
1048	1	Sperm	37.40	41568	42413	0.00	0.02	0.00	0.00	0.05	0.00	0.06	0.00	0.01	0.01	0.00	0.06	0.00	0.07	0.00	0.00	0.04	0.00
1048	2	Sperm	37.40	38014	38646	0.00	0.01	0.00	0.00	0.05	0.00	0.04	0.00	0.01	0.00	0.00	0.04	0.00	0.06	0.00	0.00	0.02	0.00
1048	3	Sperm	37.40	29732	30305	0.00	0.01	0.00	0.00	0.05	0.00	0.06	0.00	0.02	0.00	0.00	0.09	0.01	0.07	0.01	0.00	0.03	0.00
917	1	Sperm	37.84	21090	32375	0.00	0.05	0.00	0.00	0.10	0.00	0.13	0.01	3.03	3.40	0.00	0.89	0.01	8.91	0.01	0.01	0.07	0.00
917	2	Sperm	37.84	11982	18286	0.01	0.07	0.00	0.00	0.11	0.00	1.14	0.01	2.79	3.05	0.01	0.20	0.01	8.71	0.03	0.01	0.07	0.00
917	3	Sperm	37.84	20580	26494	0.00	0.03	0.00	0.00	0.08	0.00	0.09	0.00	0.06	4.82	0.02	0.11	0.01	3.41	0.00	0.00	0.04	0.00
903	1	Sperm	38.28	19560	27995	0.00	0.06	0.00	0.00	0.12	0.00	1.15	0.01	2.30	4.09	0.01	1.69	0.00	0.14	0.01	0.00	3.51	0.00
903	2	Sperm	38.28	21399	29171	0.00	0.06	0.00	0.00	0.09	0.00	0.11	0.01	3.01	3.81	0.01	0.81	0.00	0.11	0.01	0.00	2.92	0.00
903	3	Sperm	38.28	25648	35616	0.00	0.05	0.00	0.00	0.11	0.00	1.41	0.01	3.23	2.76	0.01	0.83	0.01	1.14	0.01	0.00	1.85	0.00
1012	1	Sperm	38.57	12368	23940	0.00	0.03	0.00	0.00	0.09	0.00	0.28	0.01	5.93	3.51	0.00	0.11	5.30	12.86	0.03	0.01	1.61	0.00
1012	2	Sperm	38.57	12424	23484	0.00	0.03	0.01	0.00	0.09	0.00	0.34	0.00	4.53	5.25	0.00	0.08	0.01	17.15	0.01	0.01	0.05	0.00
1012	3	Sperm	38.57	14128	25924	0.00	0.02	0.00	0.00	0.08	0.00	0.07	0.01	3.51	3.17	0.01	0.09	2.78	14.66	0.01	0.01	2.39	0.00
898	1	Sperm	38.84	26023	40453	0.01	0.05	0.00	0.00	0.09	0.00	0.14	0.00	2.10	4.04	0.01	0.14	0.01	3.75	0.01	0.01	1.12	5.94
898	2	Sperm	38.84	17241	26257	0.00	0.06	0.01	0.00	0.10	0.00	0.11	0.01	1.73	1.25	0.01	0.32	0.01	6.96	0.01	0.00	2.31	3.44
898	3	Sperm	38.84	18282	27535	0.01	0.06	0.00	0.00	0.13	0.00	0.13	0.00	2.26	2.92	0.01	0.18	0.01	9.75	0.01	0.01	0.08	0.00
868	1	Sperm	39.31	29545	36115	0.00	0.03	0.00	0.00	0.08	0.00	0.11	0.01	2.27	3.75	0.00	0.35	0.00	0.13	0.00	0.00	0.05	0.00
868	2	Sperm	39.31	22845	33298	0.00	1.82	0.00	0.00	0.13	0.00	0.38	0.02	6.55	5.50	0.01	0.11	0.00	0.09	0.01	0.01	0.05	0.00
868	3	Sperm	39.31	31297	35895	0.00	0.04	0.00	0.00	0.08	0.00	0.24	0.01	1.14	1.55	0.00	0.09	0.00	0.96	0.00	0.00	0.06	0.00
872	1	Sperm	40.63	30724	51358	0.00	0.04	0.01	0.00	0.06	0.00	0.07	0.01	5.36	2.48	0.01	0.09	0.00	13.72	0.01	0.00	0.04	0.01
872	2	Sperm	40.63	34322	54841	0.00	0.04	0.00	0.00	0.07	0.00	0.08	0.01	1.32	3.79	0.01	0.10	0.01	13.81	0.01	0.00	0.05	0.00
872	3	Sperm	40.63	34036	55714	0.00	0.02	0.00	0.00	0.05	0.00	0.07	0.01	2.37	2.01	0.01	0.07	0.00	16.23	0.01	0.00	0.05	0.01
958	1	Sperm	41.08	5970	10880	0.02	0.04	0.02	0.01	0.11	0.00	0.19	0.02	3.11	5.21	0.01	4.06	3.19	11.69	0.01	0.00	0.05	0.00
958	2	Sperm	41.08	4272	8304	0.02	0.10	0.00	0.00	0.12	0.02	1.63	0.03	7.68	9.57	0.02	1.92	0.00	10.48	0.01	0.01	0.04	0.00
958	3	Sperm	41.08	3774	5840	0.00	0.01	0.00	0.00	0.04	0.00	1.06	0.00	4.42	2.02	0.02	0.11	0.01	10.74	0.01	0.00	0.06	0.00
1057	1	Sperm	41.28	18255	25463	0.00	0.06	0.00	0.00	0.17	0.00	0.13	0.00	3.79	2.56	0.00	0.33	0.01	4.66	0.01	0.01	0.10	0.00
1057	2	Sperm	41.28	15937	27564	0.00	0.07	0.05	0.00	0.19	0.00	0.23	0.01	8.00	5.07	0.02	1.84	0.02	6.30	0.02	0.01	0.14	0.01
1057	3	Sperm	41.28	20112	26967	0.00	0.05	0.00	0.00	0.13	0.00	0.69	0.01	3.27	1.37	0.00	0.18	0.01	3.90	0.03	0.00	0.08	0.00
860	1	Sperm	41.62	38654	49832	0.00	0.02	0.00	0.00	0.07	0.00	0.10	0.00	1.56	5.36	0.01	0.06	0.01	1.99	0.01	0.00	0.03	0.00
860	2	Sperm	41.62	31783	43592	0.00	0.04	0.00	0.00	0.07	0.00	0.32	0.01	2.84	4.06	0.01	0.72	0.01	3.54	0.01	0.00	0.05	0.00
860	3	Sperm	41.62	31204	45114	0.00	0.04	0.00	0.00	0.06	0.00	0.08	0.01	5.63	4.95	0.01	0.11	0.00	3.02	0.01	0.00	0.06	0.00
565	1	Sperm	41.70	1674	3890	0.00	0.04	0.00	0.00	0.08	0.00	0.08	0.02	5.26	1.04	0.02	0.08	7.05	6.67	0.06	0.00	1.91	0.00
565	2	Sperm	41.70	1429	2825	0.00	0.02	0.00	0.00	0.09	0.00	0.16	0.00	3.45	0.00	0.00	0.07	0.02	3.97	0.02	0.00	0.07	0.00
565	3	Sperm	41.70	700	1280	0.00	0.00	0.00	0.00	0.10	0.00	0.62	0.00	2.57	0.10	0.00	0.10	1.53	8.96	0.05	0.00	0.14	0.05
1022	1	Sperm	41.85	21209	33770	0.01	0.07	0.00	0.00	0.12	0.00	0.15	0.01	4.60	7.64	0.01	0.18	0.01	5.30	0.02	0.01	0.10	0.00
1022	2	Sperm	41.85	26802	35140	0.00	0.03	0.00	0.00	0.11	0.00	0.13	0.00	1.41	2.95	0.01	0.19	0.01	2.67	0.01	0.00	1.28	0.00
1022	3	Sperm	41.85	18492	26236	0.00	0.07	0.00	0.00	0.13	0.00	0.69	0.01	4.35	2.46	0.01	0.21	0.02	4.56	0.01	0.00	0.08	0.01
1040	1	Sperm	42.14	28854	44342	0.00	0.03	0.00	0.00	0.08	0.01	0.08	0.01	2.07	3.87	0.00	0.07	2.97	6.27	0.01	0.00	1.64	0.01
1040	2	Sperm	42.14	34344	41084	0.00	0.02	0.00	0.00	0.06	0.00	0.57	0.00	0.52	0.55	0.00	0.61	2.14	1.70	0.01	0.00	0.05	0.00
1040	3	Sperm	42.14	30848	40279	0.00	0.02	0.00	0.00	0.06	0.00	0.06	0.01	1.24	1.55	0.00	0.08	0.01	6.52	0.00	0.01	0.04	0.00
780	1	Sperm	42.39	416	1393	0.00	0.08	0.00	0.00	0.08	0.00	0.00	0.00	3.20	4.00	0.00	0.16	0.00	36.89	0.16	0.08	0.88	0.08
780	2	Sperm	42.39	594	1965	0.00	0.00	0.06	0.00	0.17	0.00	0.17	0.00	4.61	4.44	0.00	0.28	0.06	32.24	0.06	0.00	4.45	0.00
780	3	Sperm	42.39	2201	6969	0.00	0.18	0.00	0.00	0.15	0.00	1.98	0.03	4.80	3.19	0.02	0.29	0.00	33.91	0.03	0.00	4.92	0.00
942	1	Sperm	42.47	19645	29758	0.00	0.07	0.01	0.01	0.13	0.00	0.26	0.01	6.02	7.94	0.01	0.30	0.01	0.85	0.01	0.00	0.09	0.00
942	2	Sperm	42.47	24404	38917	0.00	0.03	0.00	0.00	0.08	0.00	0.28	0.01	5.62	2.25	0.00	0.13	2.79	6.84	0.02	0.01	0.22	0.01
942	3	Sperm	42.47	27733	41515	0.00	0.05	0.01	0.00	0.08	0.00	0.12	0.01	4.56	4.10	0.00	1.34	1.65	3.44	0.01	0.00	0.06	0.00
960	1	Sperm	43.21	36305	40441	0.00	0.02	0.00	0.00	0.04	0.00	0.08	0.01	2.28	0.48	0.00	0.25	0.00	0.06	0.00	0.00	0.03	0.00
960	2	Sperm	43.21	37170	46718	0.00	0.02	0.00	0.00	0.05	0.00	0.06	0.01	2.67	3.27	0.01	0.09	0.00	1.77	0.01	0.00	0.03	0.00
960	3	Sperm	43.21	30422	40225	0.00	0.02	0.00	0.00	0.05	0.00	1.65	0.01	3.28	1.99	0.00	0.08	0.01	2.94	0.01	0.00	0.04	0.01
1026	1	Sperm	43.79	13790	18342	0.00	0.12	0.00	0.00	0.22	0.01	0.30	0.00	2.96	3.09	0.01	1.27	0.01	0.31	0.04	0.01	0.12	0.00
1026	2	Sperm	43.79	27460	34335	0.00	0.07	0.01	0.00	0.17	0.00	0.73	0.01	1.80	2.65	0.01	0.31	0.01	0.26	0.03	0.00	0.09	0.00
1026	3	Sperm	43.79	19314	23287	0.00	0.09	0.00	0.01	0.15	0.00	0.23	0.01	1.20	2.42	0.01	0.42	0.01	0.21	0.03	0.01	0.12	0.01

Sample ID	Replicate	Tissue	Age (Years)	c.1494del # reads	Coverage	c.1494A>C (p.L498F)	c.1494A>G (p.L498L)	c.1494A>T (p.L498F)	c.1495T>A (p.C499S)	c.1495T>C (p.C499R)	c.1495T>G (p.C499G)	c.1496G>A (p.C499Y)	c.1496G>C (p.C499S)	c.1496G>T (p.C499F)	c.1497C>A (p.C499X)	c.1497C>G (p.C499W)	c.1497C>T (p.C499C)	c.1498A>C (p.I500L)	c.1498A>G (p.I500V)	c.1498A>T (p.I500L)	c.1499T>A (p.I500K)	c.1499T>C (p.I500T)	c.1499T>G (p.I500R)
1033	1	Sperm	43.92	10366	13279	0.00	0.07	0.00	0.01	0.13	0.00	1.72	0.01	1.17	1.52	0.01	0.35	0.00	1.78	0.01	0.01	0.07	0.00
1033	2	Sperm	43.92	10291	12847	0.00	0.05	0.00	0.00	0.08	0.00	0.45	0.00	0.02	2.00	0.00	0.12	0.00	3.79	0.00	0.00	0.04	0.00
1033	3	Sperm	43.92	4167	5050	0.00	0.02	0.01	0.01	0.09	0.00	0.12	0.00	1.02	0.55	0.00	0.09	0.00	0.10	0.02	0.00	0.04	0.00
916	1	Sperm	44.12	18843	29774	0.00	0.05	0.00	0.00	0.11	0.00	1.08	0.01	2.38	0.51	0.00	1.40	0.01	12.61	0.01	0.00	0.07	0.00
916	2	Sperm	44.12	41	47	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
916	3	Sperm	44.12	31778	52015	0.00	0.03	0.00	0.00	0.07	0.00	0.53	0.01	3.33	2.59	0.01	0.88	0.01	12.59	0.01	0.00	0.06	0.01
306	1	Sperm	45.76	5297	12033	0.00	0.06	0.02	0.00	0.14	0.00	1.95	0.02	11.90	10.56	0.04	1.35	0.03	13.69	0.04	0.01	2.60	0.01
306	2	Sperm	45.76	6220	10269	0.01	0.02	0.00	0.01	0.12	0.00	0.61	0.01	6.97	5.86	0.01	0.30	0.04	5.46	0.02	0.01	2.38	0.01
306	3	Sperm	45.76	5667	11465	0.00	0.03	0.00	0.01	0.06	0.00	1.15	0.00	11.06	10.35	0.02	0.16	0.02	11.25	0.05	0.01	0.09	0.00
726	1	Sperm	46.34	8498	9343	0.01	0.00	0.00	0.00	0.07	0.00	0.69	0.00	0.42	1.87	0.01	0.08	0.01	0.09	0.01	0.01	0.05	0.00
726	2	Sperm	46.34	9944	12671	0.00	0.02	0.00	0.01	0.09	0.01	0.84	0.01	4.57	3.35	0.02	0.07	0.01	0.08	0.02	0.00	0.04	0.00
726	3	Sperm	46.34	5523	6309	0.00	0.01	0.01	0.00	0.07	0.00	0.05	0.01	4.63	0.00	0.00	0.05	0.01	0.06	0.00	0.01	0.02	0.00
879	1	Sperm	47.09	25594	48488	0.01	0.05	0.00	0.00	0.09	0.00	0.15	0.03	12.92	5.24	0.01	1.93	1.05	7.50	0.01	0.01	0.08	0.01
879	2	Sperm	47.09	21906	56211	0.00	0.07	0.00	0.00	0.10	0.00	1.19	0.03	21.64	16.73	0.03	1.19	1.20	8.58	0.01	0.01	0.09	0.01
879	3	Sperm	47.09	27345	49397	0.00	0.02	0.00	0.00	0.07	0.00	0.51	0.02	12.79	6.95	0.01	0.48	0.01	5.13	0.01	0.06	0.07	0.00
1031	1	Sperm	47.59	18406	24633	0.00	0.06	0.01	0.00	0.11	0.01	1.15	0.00	1.45	2.66	0.00	1.10	0.01	1.92	0.02	0.00	1.71	0.00
1031	2	Sperm	47.59	24958	37834	0.00	0.04	0.00	0.00	0.09	0.01	0.10	0.01	2.42	5.26	0.02	2.84	0.00	1.88	0.00	0.00	3.63	0.00
1031	3	Sperm	47.59	31843	35265	0.00	0.02	0.00	0.00	0.07	0.00	0.08	0.00	0.55	1.97	0.01	0.10	0.01	0.10	0.01	0.00	0.05	0.00
1028	1	Sperm	50.63	18336	32269	0.01	0.06	0.01	0.00	0.16	0.00	0.19	0.01	2.66	1.38	0.01	0.41	0.02	18.22	0.03	0.00	0.10	0.01
1028	2	Sperm	50.63	16056	34558	0.00	0.07	1.69	0.00	0.14	0.01	0.59	0.01	1.89	3.98	0.01	0.23	0.02	27.49	0.03	0.00	0.11	0.01
1028	3	Sperm	50.63	19006	33711	0.00	0.06	0.98	0.00	0.10	0.01	0.16	0.01	2.93	1.47	0.00	0.15	0.02	18.31	0.01	0.00	0.08	0.01
526	1	Sperm	54.17	0	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
526	3	Sperm	54.17	7205	13213	0.00	0.04	0.00	0.00	0.13	0.00	0.68	0.02	6.56	5.58	0.01	0.12	0.01	13.61	0.02	0.00	1.06	0.00
526	3	Sperm	54.17	7515	13907	0.00	0.06	0.00	0.00	0.18	0.00	0.28	0.02	5.65	8.12	0.00	0.17	0.03	12.74	0.02	0.02	0.83	0.00
809	1	Sperm	55.92	1765	5071	0.00	0.04	0.02	0.00	0.04	0.00	0.11	0.02	0.04	1.47	0.00	0.08	9.17	37.36	7.72	0.00	0.02	0.02
809	2	Sperm	55.92	595	3438	0.00	0.22	0.00	0.00	0.22	0.00	0.17	0.06	7.22	5.48	0.17	0.28	6.95	108.46	2.86	0.00	5.76	0.00
809	3	Sperm	55.92	346	1778	0.00	0.19	0.00	0.00	0.10	0.00	0.10	0.00	0.19	1.64	0.00	0.10	10.48	105.60	5.29	0.00	2.50	0.00
1124	1	Sperm	58.57	6087	16004	0.00	0.04	0.00	0.03	0.11	0.00	0.55	0.04	22.72	18.05	0.03	0.07	0.01	10.31	0.01	0.00	1.85	0.00
1124	2	Sperm	58.57	6643	13243	0.00	0.05	0.01	0.00	0.12	0.00	0.04	0.01	7.85	11.96	0.04	0.71	0.01	11.63	0.01	0.00	1.04	0.00
1124	3	Sperm	58.57	4530	9667	0.00	0.02	0.01	0.01	0.10	0.00	0.02	0.04	13.94	14.57	0.06	0.05	0.00	9.18	0.01	0.01	0.07	0.00
799	1	Sperm	59.48	2845	6220	0.00	0.04	0.01	0.00	0.04	0.00	0.07	0.00	1.03	3.41	0.00	0.15	0.00	21.54	0.00	0.00	0.08	0.01
799	2	Sperm	59.48	1681	3897	0.00	0.06	0.00	0.00	0.22	0.02	0.06	0.00	4.47	7.58	0.00	0.14	0.00	11.50	0.04	0.00	0.08	0.00
799	3	Sperm	59.48	553	1424	0.00	0.00	0.06	0.00	0.24	0.00	1.99	0.00	3.32	0.00	0.00	0.24	0.00	16.67	0.06	0.06	0.18	0.00
1018	1	Sperm	59.88	3731	19054	0.00	0.20	0.00	0.00	0.35	0.03	0.34	0.01	5.72	4.09	0.00	0.76	4.43	111.76	0.08	0.03	0.98	0.00
1018	2	Sperm	59.88	4434	18808	0.01	0.14	0.01	0.01	0.33	0.01	0.26	0.01	2.29	0.69	0.01	0.37	6.00	85.88	0.11	0.01	4.47	0.00
1018	3	Sperm	59.88	3708	19168	0.01	0.20	0.00	0.01	0.37	0.01	0.30	0.00	3.53	3.67	0.03	0.47	5.56	113.17	0.04	0.00	4.33	0.00
294	1	Sperm	71.77	4490	13754	0.00	0.05	0.01	0.00	0.15	0.01	0.07	0.01	4.48	6.11	0.01	0.18	0.03	35.65	2.43	0.01	20.12	0.00
294	2	Sperm	71.77	4224	11709	0.00	0.05	0.00	0.00	0.16	0.01	0.40	0.02	5.09	4.99	0.02	0.07	1.24	34.37	1.10	0.04	12.00	0.00
294	3	Sperm	71.77	4101	12293	0.00	1.31	0.00	0.00	0.13	0.00	0.82	0.00	3.75	5.41	0.00	0.19	0.03	43.70	0.01	0.00	11.14	0.00
545	1	Sperm	75.86	5003	9141	0.01	0.02	0.01	0.00	0.13	0.01	0.13	0.01	3.33	3.02	0.01	0.35	0.01	19.11	0.02	0.01	1.77	0.01
545	2	Sperm	75.86	5077	9666	0.00	0.07	0.01	0.00	0.11	0.00	0.69	0.05	11.30	1.37	0.01	0.50	0.01	16.15	0.03	0.01	0.09	0.00
545	3	Sperm	75.86	5419	9934	0.00	0.03	0.01	0.01	0.09	0.00	0.09	0.01	2.55	3.29	0.00	0.14	0.01	19.30	0.01	0.01	2.68	0.00
1236	1	Blood	24.76	16774	18619	0.00	0.02	0.00	0.00	0.10	0.00	0.05	0.00	0.01	0.01	0.01	2.52	0.00	0.12	0.01	0.00	0.06	0.00
1236	2	Blood	24.76	8402	8885	0.00	0.04	0.00	0.00	0.17	0.00	0.13	0.00	0.12	0.00	0.00	0.13	0.01	0.14	0.02	0.00	0.08	0.00
1240	1	Blood	26.41	3596	6627	0.01	0.19	0.47	0.06	1.15	0.01	1.81	0.03	9.85	9.85	0.02	1.41	0.06	1.28	0.13	0.74	0.17	0.00
1240	2	Blood	26.41	2376	3545	0.00	0.15	0.11	0.03	0.31	0.01	0.29	0.01	4.19	6.81	0.01	0.77	0.03	2.06	0.13	0.03	0.21	0.00
1240	3	Blood	26.41	4378	6628	0.01	0.11	0.08	0.02	0.28	0.01	0.58	0.01	10.18	3.42	0.02	0.97	0.03	0.30	0.10	0.00	0.20	0.00
1063	1	Blood	39.06	9413	10937	0.01	0.00	0.00	0.00	1.16	0.00	0.06	0.01	1.92	2.32	0.00	0.04	0.00	0.08	0.01	0.00	0.05	0.00
1063	2	Blood	39.06	8984	10931	0.00	0.00	0.00	0.00	0.09	0.00	0.04	0.01	3.73	3.41	0.00	0.07	0.00	0.12	0.01	0.01	0.04	0.00
1063	3	Blood	39.06	16818	19307	0.00	0.00	0.00	0.00	0.08	0.00	0.03	0.01	2.40	2.42	0.01	0.04	0.00	0.08	0.01	0.00	0.03	0.00
1008	1	Blood	45.51	4094	5211	0.02	0.00	0.05	0.07	0.07	0.14	0.00	8.02	0.00	0.00	0.02	0.00	0.14	0.02	0.02	0.05	0.00	0.00
1008	2	Blood	45.51	4158	5430	0.00	0.01	0.00	0.06	0.02	0.03	0.00	3.13	1.56	0.01	0.07	0.02	0.11	0.02	0.00	0.07	0.02	0.00
1008	3	Blood	45.51	1040	1320	0.00	0.00	0.00	0.13	0.10	0.00	0.03	2.91	1.65	0.00	0.00	0.00	0.13	0.00	0.00	0.03	0.00	0.00
1239	1	Blood	57.01	6102	8162	0.00	0.04	0.08	0.01	0.11	0.01	1.19	0.00	4.65	3.66	0.02	1.05	0.01	0.20	0.03	0.00	0.09	0.00
1239	2	Blood	57.01	5511	6675	0.00	0.03	0.00	0.00	0.10	0.01	0.09	0.01	3.99	1.79	0.01	0.66	0.01	0.10	0.04	0.01	0.11	0.01
1239	3	Blood	57.01	7000	8524	0.00	0.03	0.00	0.00	0.08	0.00	0.28	0.01	3.07	3.29	0.01	0.08	0.00	0.14	0.03	0.00	0.07	0.01

Bibliography

1. Abbassi, M., Bourmtane, A., Sayel, H., El Mouhi, H., Jalte, M., Elasri, Y. A., Askander, O., El Fahime, E., & Bouguenouch, L. (2023). Genetic characterization of Schuurs-Hoeijmakers syndrome in a moroccan individual with heterozygote PACS1 mutation. *Molecular Biology Reports*, *50*(11), 9121–9128. https://doi.org/10.1007/s11033-023-08803-9
2. Acuna-Hidalgo, R., Bo, T., Kwint, M. P., van de Vorst, M., Pinelli, M., Veltman, J. A., Hoischen, A., Vissers, L. E. L. M., & Gilissen, C. (2015). Post-zygotic Point Mutations Are an Underrecognized Source of De Novo Genomic Variation. *The American Journal of Human Genetics*, *97*(1), 67–74. https://doi.org/10.1016/j.ajhg.2015.05.008
3. Acuna-Hidalgo, R., Sengul, H., Steehouwer, M., Van De Vorst, M., Vermeulen, S. H., Kiemeney, L. A. L. M., Veltman, J. A., Gilissen, C., & Hoischen, A. (2017). Ultra-sensitive Sequencing Identifies High Prevalence of Clonal Hematopoiesis-Associated Mutations throughout Adult Life. *The American Journal of Human Genetics*, *101*(1), 50–64. https://doi.org/10.1016/j.ajhg.2017.05.013
4. Acuna-Hidalgo, R., Veltman, J. A., & Hoischen, A. (2016). New insights into the generation and role of de novo mutations in health and disease. *Genome Biology*, *17*(1), 241. https://doi.org/10.1186/s13059-016-1110-1
5. Al Ageeli, E., Mignot, C., Afenjar, A., Whalen, S., Dorison, N., Mayer, M., Esteva, B., Dubern, B., Momtchilova, M., Le Gargasson, J.-F., Bursztyn, J., & Héron, D. (2012). Retinal involvement in two unrelated patients with Myhre syndrome. *European Journal of Medical Genetics*, *55*(10), 541–547. https://doi.org/10.1016/j.ejmg.2012.05.006
6. Alagia, M., Cappuccio, G., Pinelli, M., Torella, A., Brunetti-Pierri, R., Simonelli, F., Limongelli, G., Oppido, G., Nigro, V., Brunetti-Pierri, N., & TUDP. (2018). A child with Myhre syndrome presenting with corectopia and tetralogy of Fallot. *American Journal of Medical Genetics Part A*, *176*(2), 426–430. https://doi.org/10.1002/ajmg.a.38560
7. Alankarage, D., Enriquez, A., Steiner, R. D., Raggio, C., Higgins, M., Milnes, D., Humphreys, D. T., Duncan, E. L., Sparrow, D. B., Giampietro, P. F., Chapman, G., & Dunwoodie, S. L. (2022a). Myhre syndrome is caused by dominant-negative dysregulation of SMAD4 and other co-factors. *Differentiation*, *128*, 1–12. https://doi.org/10.1016/j.diff.2022.09.002
8. Alankarage, D., Enriquez, A., Steiner, R. D., Raggio, C., Higgins, M., Milnes, D., Humphreys, D. T., Duncan, E. L., Sparrow, D. B., Giampietro, P. F., Chapman, G., & Dunwoodie, S. L. (2022b). Myhre syndrome is caused by dominant-negative dysregulation of SMAD4 and other co-factors. *Differentiation*, *128*, 1–12. https://doi.org/10.1016/j.diff.2022.09.002
9. Alape, D., Singh, R., Folch, E., Fernandez Bussy, S., Agnew, A., & Majid, A. (2020). Life-Threatening Multilevel Airway Stenosis Due to Myhre Syndrome. *American Journal of Respiratory and Critical Care Medicine*, *201*(6), 731–732. https://doi.org/10.1164/rccm.201905-0922IM
10. Alexandrov, L. B., Jones, P. H., Wedge, D. C., Sale, J. E., Campbell, P. J., Nik-Zainal, S., & Stratton, M. R. (2015). Clock-like mutational processes in human somatic cells. *Nature Genetics*, *47*(12), 1402–1407. https://doi.org/10.1038/ng.3441

11. Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Tian Ng, A. W., Wu, Y., Boot, A., Covington, K. R., Gordenin, D. A., Bergstrom, E. N., Islam, S. M. A., Lopez-Bigas, N., Klimczak, L. J., McPherson, J. R., Morganella, S., Sabarinathan, R., Wheeler, D. A., Mustonen, V., PCAWG Mutational Signatures Working Group, … Von Mering, C. (2020). The repertoire of mutational signatures in human cancer. *Nature*, *578*(7793), 94–101. https://doi.org/10.1038/s41586-020-1943-3
12. Alexandrov, L. B., Nik-Zainal, S., Aparicio, S. A. J. R., Wedge, D. C., Australian Pancreatic Cancer Genome Initiative, ICGC MMML-Seq Consortium, ICGC Breast Cancer Consortium, ICGC PedBrain, Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Børresen-Dale, A.-L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., … Stratton, M. R. (2013). Signatures of mutational processes in human cancer. *Nature*, *500*(7463), 415–421. https://doi.org/10.1038/nature12477
13. Anderson, J. (1998). Apert syndrome mutations in fibroblast growth factor receptor 2 exhibit increased affinity for FGF ligand. *Human Molecular Genetics*, *7*(9), 1475–1483. https://doi.org/10.1093/hmg/7.9.1475
14. Araki, T., Chan, G., Newbigging, S., Morikawa, L., Bronson, R. T., & Neel, B. G. (2009). Noonan syndrome cardiac defects are caused by *PTPN11* acting in endocardium to enhance endocardial-mesenchymal transformation. *Proceedings of the National Academy of Sciences*, *106*(12), 4736–4741. https://doi.org/10.1073/pnas.0810053106
15. Arighi, E., Borrello, M. G., & Sariola, H. (2005). RET tyrosine kinase signaling in development and cancer. *Cytokine & Growth Factor Reviews*, *16*(4–5), 441–467. https://doi.org/10.1016/j.cytogfr.2005.05.010
16. Arnesen, T., Van Damme, P., Polevoda, B., Helsens, K., Evjenth, R., Colaert, N., Varhaug, J. E., Vandekerckhove, J., Lillehaug, J. R., Sherman, F., & Gevaert, K. (2009). Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proceedings of the National Academy of Sciences*, *106*(20), 8157–8162. https://doi.org/10.1073/pnas.0901931106
17. Artemios, P., Areti, S., Katerina, P., Helen, F., Eirini, T., & Charalambos, P. (2019a). Autism Spectrum Disorder and Psychiatric Comorbidity in a Patient with Myhre Syndrome. *Journal of Autism and Developmental Disorders*, *49*(7), 3031–3035. https://doi.org/10.1007/s10803-019-04015-y
18. Artemios, P., Areti, S., Katerina, P., Helen, F., Eirini, T., & Charalambos, P. (2019b). Autism Spectrum Disorder and Psychiatric Comorbidity in a Patient with Myhre Syndrome. *Journal of Autism and Developmental Disorders*, *49*(7), 3031–3035. https://doi.org/10.1007/s10803-019-04015-y
19. Asakura, Y., Muroya, K., Sato, T., Kurosawa, K., Nishimura, G., & Adachi, M. (2012). First case of a Japanese girl with Myhre syndrome due to a heterozygous *SMAD4* mutation. *American Journal of Medical Genetics Part A*, *158A*(8), 1982–1986. https://doi.org/10.1002/ajmg.a.35440
20. Attisano, L., & Wrana, J. L. (2000). Smads as transcriptional co-modulators. *Current Opinion in Cell Biology*, *12*(2), 235–243. https://doi.org/10.1016/S0955-0674(99)00081-2
21. Aubry, F., Satie, A.-P., Rioux-Leclercq, N., Rajpert-De Meyts, E., Spagnoli, G. C., Chomez, P., De Backer, O., J□gou, B., & Samson, M. (2001). MAGE-A4, a germ cell specific marker, is expressed differentially in testicular tumors. *Cancer*, *92*(11), 2778–

2785. https://doi.org/10.1002/1097-0142(20011201)92:11<2778::AID-CNCR10125>3.0.CO;2-S
22. Averbeck, N., Gao, X.-D., Nishimura, S.-I., & Dean, N. (2008). Alg13p, the Catalytic Subunit of the Endoplasmic Reticulum UDP-GlcNAc Glycosyltransferase, Is a Target for Proteasomal Degradation. *Molecular Biology of the Cell*, *19*(5), 2169–2178. https://doi.org/10.1091/mbc.e07-10-1077
23. Bachmann-Gagescu, R., Hisama, F. M., & Yuen, A. L. (2011). Myhre syndrome with ataxia and cerebellar atrophy. *Clinical Dysmorphology*, *20*(3), 156–159. https://doi.org/10.1097/MCD.0b013e3283468043
24. Bassett, J. K., Douzgou, S., & Kerr, B. (2016). Severe constipation in a patient with Myhre syndrome: A case report. *Clinical Dysmorphology*, *25*(2), 54–57. https://doi.org/10.1097/MCD.0000000000000109
25. Becerra-Solano, L. E., Díaz-Rodriguez, M., Nastasi-Catanese, J. A., Toscano-Flores, J. J., Bañuelos-Robles, O., Figuera, L. E., Matute, E., & De Lourdes Ramírez-Dueñas, M. (2008). The fifth female patient with Myhre syndrome: Further delineation. *Clinical Dysmorphology*, *17*(2), 113–117. https://doi.org/10.1097/MCD.0b013e3282f52828
26. Beck, D. B., Cho, M. T., Millan, F., Yates, C., Hannibal, M., O'Connor, B., Shinawi, M., Connolly, A. M., Waggoner, D., Halbach, S., Angle, B., Sanders, V., Shen, Y., Retterer, K., Begtrup, A., Bai, R., & Chung, W. K. (2016). A recurrent de novo CTBP1 mutation is associated with developmental delay, hypotonia, ataxia, and tooth enamel defects. *Neurogenetics*, *17*(3), 173–178. https://doi.org/10.1007/s10048-016-0482-4
27. Beck, D. B., Subramanian, T., Vijayalingam, S., Ezekiel, U. R., Donkervoort, S., Yang, M. L., Dubbs, H. A., Ortiz-Gonzalez, X. R., Lakhani, S., Segal, D., Au, M., Graham, J. M., Verma, S., Waggoner, D., Shinawi, M., Bönnemann, C. G., Chung, W. K., & Chinnadurai, G. (2019). A pathogenic CtBP1 missense mutation causes altered cofactor binding and transcriptional activity. *Neurogenetics*, *20*(3), 129–143. https://doi.org/10.1007/s10048-019-00578-1
28. Bellus, G. A., Hefferon, T. W., de Luna, O., Hecht, J. T., Horton, W. A., & Machado, M. (1995). Achondroplasia Is Defined by Recurrent G380R Mutations of FGFR3. *Am. J. Hum. Genet.*
29. Bellus, G. A., Spector, E. B., Speiser, P. W., Weaver, C. A., Garber, A. T., Bryke, C. R., Israel, J., Rosengren, S. S., Webster, M. K., Donoghue, D. J., & Francomano, C. A. (2000). Distinct Missense Mutations of the FGFR3 Lys650 Codon Modulate Receptor Kinase Activation and the Severity of the Skeletal Dysplasia Phenotype. *The American Journal of Human Genetics*, *67*(6), 1411–1421. https://doi.org/10.1086/316892
30. Beyer, J., Albers, P., Altena, R., Aparicio, J., Bokemeyer, C., Busch, J., Cathomas, R., Cavallin-Stahl, E., Clarke, N. W., Claßen, J., Cohn-Cedermark, G., Dahl, A. A., Daugaard, G., De Giorgi, U., De Santis, M., De Wit, M., De Wit, R., Dieckmann, K. P., Fenner, M., … Wittekind, C. (2013). Maintaining success, reducing treatment burden, focusing on survivorship: Highlights from the third European consensus conference on diagnosis and treatment of germ-cell cancer. *Annals of Oncology*, *24*(4), 878–888. https://doi.org/10.1093/annonc/mds579
31. Blanc, A., Bonnet, C., Wandzel, M., Roth, V., Duffourd, Y., Safraou, H., Leheup, B., Muller, F., D Colne, J., Feillet, F., Schmitt, E., Castro, M., Savatt, J., Burcheri, A., Nemos, C., Philippe, C., & Lambert, L. (2024). Patient with a heterozygous pathogenic variant in *CSNK2A1* gene: A new case to update the Okur–Chung neurodevelopmental

syndrome. *American Journal of Medical Genetics Part A*, *194*(9), e63642. https://doi.org/10.1002/ajmg.a.63642
32. Blank, C. E. (1959). Apert's syndrome (a type of acrocephalosyndactyly)–observations on a British series of thirty-nine cases*. *Annals of Human Genetics*, *24*(2), 151–164. https://doi.org/10.1111/j.1469-1809.1959.tb01728.x
33. Bocciardi, R., Mograbi, B., Pasini, B., Borrello, M. G., Pierotti, M. A., Bourget, I., Fischer, S., Romeo, G., & Rossi, B. (1997). The multiple endocrine neoplasia type 2B point mutation switches the specificity of the Ret tyrosine kinase towards cellular substrates that are susceptible to interact with Crk and Nck. *Oncogene*, *15*(19), 2257–2265. https://doi.org/10.1038/sj.onc.1201413.
34. Bollen, M., Peti, W., Ragusa, M. J., & Beullens, M. (2010). The extended PP1 toolkit: Designed to create specificity. *Trends in Biochemical Sciences*, *35*(8), 450–458. https://doi.org/10.1016/j.tibs.2010.03.002
35. Borgo, C., D'Amore, C., Sarno, S., Salvi, M., & Ruzzene, M. (2021). Protein kinase CK2: A potential therapeutic target for diverse human diseases. *Signal Transduction and Targeted Therapy*, *6*(1), 183. https://doi.org/10.1038/s41392-021-00567-7
36. Braccioli, L., Vervoort, S. J., Adolfs, Y., Heijnen, C. J., Basak, O., Pasterkamp, R. J., Nijboer, C. H., & Coffer, P. J. (2017). FOXP1 Promotes Embryonic Neural Stem Cell Differentiation by Repressing Jagged1 Expression. *Stem Cell Reports*, *9*(5), 1530–1545. https://doi.org/10.1016/j.stemcr.2017.10.012
37. Brigman, J. L., Wright, T., Talani, G., Prasad-Mulcare, S., Jinde, S., Seabold, G. K., Mathur, P., Davis, M. I., Bock, R., Gustin, R. M., Colbran, R. J., Alvarez, V. A., Nakazawa, K., Delpire, E., Lovinger, D. M., & Holmes, A. (2010). Loss of GluN2B-Containing NMDA Receptors in CA1 Hippocampus and Cortex Impairs Long-Term Depression, Reduces Dendritic Spine Density, and Disrupts Learning. *The Journal of Neuroscience*, *30*(13), 4590–4600. https://doi.org/10.1523/JNEUROSCI.0640-10.2010
38. Brosens, L. A. (2011). Juvenile polyposis syndrome. *World Journal of Gastroenterology*, *17*(44), 4839. https://doi.org/10.3748/wjg.v17.i44.4839
39. Brown, J. E., Aldred, B., Boulter, T., Sullivan, R., Ver Hoeve, J., Pattnaik, B. R., & Schmitt, M. (2024). A case report of retinal dystrophy in patients with PACS1 syndrome. *Ophthalmic Genetics*, *45*(1), 103–107. https://doi.org/10.1080/13816810.2023.2216272
40. Brunner, S. F., Roberts, N. D., Wylie, L. A., Moore, L., Aitken, S. J., Davies, S. E., Sanders, M. A., Ellis, P., Alder, C., Hooks, Y., Abascal, F., Stratton, M. R., Martincorena, I., Hoare, M., & Campbell, P. J. (2019). Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. *Nature*, *574*(7779), 538–542. https://doi.org/10.1038/s41586-019-1670-9
41. Bruskov, V. I. (2002). Heat-induced formation of reactive oxygen species and 8-oxoguanine, a biomarker of damage to DNA. *Nucleic Acids Research*, *30*(6), 1354–1363. https://doi.org/10.1093/nar/30.6.1354
42. Burglen, L. (2003). Myhre syndrome: New reports, review, and differential diagnosis. *Journal of Medical Genetics*, *40*(7), 546–551. https://doi.org/10.1136/jmg.40.7.546
43. Bush, S. J., Nikola, R., Han, S., Suzuki, S., Yoshida, S., Simons, B. D., & Goriely, A. (2024). Adult Human, but Not Rodent, Spermatogonial Stem Cells Retain States with a Foetal-like Signature. *Cells*, *13*(9), 742. https://doi.org/10.3390/cells13090742
44. Caefer, D. M., Phan, N. Q., Liddle, J. C., Balsbaugh, J. L., O'Shea, J. P., Tzingounis, A. V., & Schwartz, D. (2022). The Okur-Chung Neurodevelopmental Syndrome Mutation

CK2K198R Leads to a Rewiring of Kinase Specificity. *Frontiers in Molecular Biosciences*, *9*, 850661. https://doi.org/10.3389/fmolb.2022.850661

45. Cappellen, D., De Oliveira, C., Ricol, D., De Medina, S., Bourdin, J., Sastre-Garau, X., Chopin, D., Thiery, J. P., & Radvanyi, F. (1999). Frequent activating mutations of FGFR3 in human bladder and cervix carcinomas. *Nature Genetics*, *23*(1), 18–20. https://doi.org/10.1038/12615
46. Cappuccio, G., Brunetti-Pierri, N., Clift, P., Learn, C., Dykes, J. C., Mercer, C. L., Callewaert, B., Meerschaut, I., Spinelli, A. M., Bruno, I., Gillespie, M. J., Dorfman, A. T., Grimberg, A., Lindsay, M. E., & Lin, A. E. (2022). Expanded cardiovascular phenotype of Myhre syndrome includes tetralogy of Fallot suggesting a role for *SMAD4* in human neural crest defects. *American Journal of Medical Genetics Part A*, *188*(5), 1384–1395. https://doi.org/10.1002/ajmg.a.62645
47. Cappuccio, G., Caiazza, M., Roca, A., Melis, D., Iuliano, A., Matyas, G., Rubino, M., Limongelli, G., & Brunetti-Pierri, N. (2021). A pilot clinical trial with losartan in Myhre syndrome. *American Journal of Medical Genetics Part A*, *185*(3), 702–709. https://doi.org/10.1002/ajmg.a.62019
48. Cappuyns, E., Huyghebaert, J., Vandeweyer, G., & Kooy, R. F. (2018). Mutations in *ADNP* affect expression and subcellular localization of the protein. *Cell Cycle*, *17*(9), 1068–1075. https://doi.org/10.1080/15384101.2018.1471313
49. Caputo, V., Bocchinfuso, G., Castori, M., Traversa, A., Pizzuti, A., Stella, L., Grammatico, P., & Tartaglia, M. (2014a). Novel *SMAD4* mutation causing Myhre syndrome. *American Journal of Medical Genetics Part A*, *164*(7), 1835–1840. https://doi.org/10.1002/ajmg.a.36544
50. Caputo, V., Bocchinfuso, G., Castori, M., Traversa, A., Pizzuti, A., Stella, L., Grammatico, P., & Tartaglia, M. (2014b). Novel *SMAD4* mutation causing Myhre syndrome. *American Journal of Medical Genetics Part A*, *164*(7), 1835–1840. https://doi.org/10.1002/ajmg.a.36544
51. Caputo, V., Cianetti, L., Niceta, M., Carta, C., Ciolfi, A., Bocchinfuso, G., Carrani, E., Dentici, M. L., Biamino, E., Belligni, E., Garavelli, L., Boccone, L., Melis, D., Andria, G., Gelb, B. D., Stella, L., Silengo, M., Dallapiccola, B., & Tartaglia, M. (2012a). A Restricted Spectrum of Mutations in the SMAD4 Tumor-Suppressor Gene Underlies Myhre Syndrome. *The American Journal of Human Genetics*, *90*(1), 161–169. https://doi.org/10.1016/j.ajhg.2011.12.011
52. Caputo, V., Cianetti, L., Niceta, M., Carta, C., Ciolfi, A., Bocchinfuso, G., Carrani, E., Dentici, M. L., Biamino, E., Belligni, E., Garavelli, L., Boccone, L., Melis, D., Andria, G., Gelb, B. D., Stella, L., Silengo, M., Dallapiccola, B., & Tartaglia, M. (2012b). A Restricted Spectrum of Mutations in the SMAD4 Tumor-Suppressor Gene Underlies Myhre Syndrome. *The American Journal of Human Genetics*, *90*(1), 161–169. https://doi.org/10.1016/j.ajhg.2011.12.011
53. Carpentieri, G., Leoni, C., Pietraforte, D., Cecchetti, S., Iorio, E., Belardo, A., Pietrucci, D., Di Nottia, M., Pajalunga, D., Megiorni, F., Mercurio, L., Tatti, M., Camero, S., Marchese, C., Rizza, T., Tirelli, V., Onesimo, R., Carrozzo, R., Rinalducci, S., … Flex, E. (2022). Hyperactive HRAS dysregulates energetic metabolism in fibroblasts from patients with Costello syndrome via enhanced production of reactive oxidizing species. *Human Molecular Genetics*, *31*(4), 561–575. https://doi.org/10.1093/hmg/ddab270

54. Castellano, E., & Downward, J. (2011). RAS Interaction with PI3K: More Than Just Another Effector Pathway. *Genes & Cancer*, *2*(3), 261–274. https://doi.org/10.1177/1947601911408079
55. Cătană, A., Simonescu-Colan, R., Cuzmici-Barabaș, Z., Militaru, D., Iordănescu, I., & Militaru, M. (2022). First documented case of Myhre syndrome in Romania: A case report. *Experimental and Therapeutic Medicine*, *23*(5), 323. https://doi.org/10.3892/etm.2022.11252
56. Cerritelli, S. M., & Crouch, R. J. (2009). Ribonuclease H: The enzymes in eukaryotes. *The FEBS Journal*, *276*(6), 1494–1505. https://doi.org/10.1111/j.1742-4658.2009.06908.x
57. Chang, Y., Wang, X., Yang, J., Tien, J. C.-Y., Mannan, R., Cruz, G., Zhang, Y., Vo, J. N., Magnuson, B., Mahapatra, S., Cho, H., Dhanasekaran, S. M., Wang, C., Wang, Z., Zhou, L., Zhou, K., Zhou, Y., Zhang, P., Huang, W., … Chinnaiyan, A. M. (2024). Development of an orally bioavailable CDK12/13 degrader and induction of synthetic lethality with AKT pathway inhibition. *Cell Reports Medicine*, *5*(10), 101752. https://doi.org/10.1016/j.xcrm.2024.101752
58. Chen, M., & Zhou, X. (2018). VIPER: Variability-preserving imputation for accurate gene expression recovery in single-cell RNA sequencing studies. *Genome Biology*, *19*(1), 196. https://doi.org/10.1186/s13059-018-1575-1
59. Chen, P.-C., Wakimoto, H., Conner, D., Araki, T., Yuan, T., Roberts, A., Seidman, C. E., Bronson, R., Neel, B. G., Seidman, J. G., & Kucherlapati, R. (2010). Activation of multiple signaling pathways causes developmental defects in mice with a Noonan syndrome–associated Sos1 mutation. *Journal of Clinical Investigation*, *120*(12), 4353–4365. https://doi.org/10.1172/JCI43910
60. Cheng, H., Gottlieb, L., Marchi, E., Kleyner, R., Bhardwaj, P., Rope, A. F., Rosenheck, S., Moutton, S., Philippe, C., Eyaid, W., Alkuraya, F. S., Toribio, J., Mena, R., Prada, C. E., Stessman, H., Bernier, R., Wermuth, M., Kauffmann, B., Blaumeiser, B., … Lyon, G. J. (2019). Phenotypic and biochemical analysis of an international cohort of individuals with variants in NAA10 and NAA15. *Human Molecular Genetics*, *28*(17), 2900–2919. https://doi.org/10.1093/hmg/ddz111
61. Cheng, T.-L., Wang, Z., Liao, Q., Zhu, Y., Zhou, W.-H., Xu, W., & Qiu, Z. (2014). MeCP2 Suppresses Nuclear MicroRNA Processing and Dendritic Growth by Regulating the DGCR8/Drosha Complex. *Developmental Cell*, *28*(5), 547–560. https://doi.org/10.1016/j.devcel.2014.01.032
62. Chesi, M., Nardini, E., Brents, L. A., Schröck, E., Ried, T., Kuehl, W. M., & Bergsagel, P. L. (1997). Frequent translocation t(4;14)(p16.3;q32.3) in multiple myeloma is associated with increased expression and activating mutations of fibroblast growth factor receptor 3. *Nature Genetics*, *16*(3), 260–264. https://doi.org/10.1038/ng0797-260
63. Chevalier, N., Hinault, C., Clavel, S., Paul-Bellon, R., & Fenichel, P. (2021). GPER and Testicular Germ Cell Cancer. *Frontiers in Endocrinology*, *11*, 600404. https://doi.org/10.3389/fendo.2020.600404
64. Chinnadurai, G. (2003). CtBP family proteins: More than transcriptional corepressors. *BioEssays*, *25*(1), 9–12. https://doi.org/10.1002/bies.10212
65. Choi, S.-K., Yoon, S.-R., Calabrese, P., & Arnheim, N. (2008). A germ-line-selective advantage rather than an increased mutation rate can explain some unexpectedly common

human disease mutations. *Proceedings of the National Academy of Sciences*, *105*(29), 10143–10148. https://doi.org/10.1073/pnas.0801267105
66. Choi, S.-K., Yoon, S.-R., Calabrese, P., & Arnheim, N. (2012). Positive Selection for New Disease Mutations in the Human Germline: Evidence from the Heritable Cancer Syndrome Multiple Endocrine Neoplasia Type 2B. *PLoS Genetics*, *8*(2), e1002420. https://doi.org/10.1371/journal.pgen.1002420
67. Chou, I.-J., Hou, J.-Y., Fan, W.-L., Tsai, M.-H., & Lin, K.-L. (2023). Long-Term Outcome of Neonatal Seizure with PACS2 Mutation: Case Series and Literature Review. *Children*, *10*(4), 621. https://doi.org/10.3390/children10040621
68. Chow, E., & Macrae, F. (2005). A review of Juvenile Polyposis Syndrome. *Journal of Gastroenterology and Hepatology*, *20*(11), 1634–1640. https://doi.org/10.1111/j.1440-1746.2005.03865.x
69. Cintrón-Colón, A. F., Almeida-Alves, G., Boynton, A. M., & Spitsbergen, J. M. (2020). GDNF synthesis, signaling, and retrograde transport in motor neurons. *Cell and Tissue Research*, *382*(1), 47–56. https://doi.org/10.1007/s00441-020-03287-6
70. Clapham, D. E., & Neer, E. J. (1997). G PROTEIN βγ SUBUNITS. *Annual Review of Pharmacology and Toxicology*, *37*(1), 167–203. https://doi.org/10.1146/annurev.pharmtox.37.1.167
71. Clermont, Y. (1966). Spermatogenesis in Man. *Fertility and Sterility*, *17*(6), 705–721. https://doi.org/10.1016/S0015-0282(16)36120-9
72. Cohen, M. M., Kreiborg, S., Lammer, E. J., Cordero, J. F., Mastroiacovo, P., Erickson, J. D., Roeper, P., & Martínez-Frías, M. L. (1992). Birth prevalence study of the apert syndrome. *American Journal of Medical Genetics*, *42*(5), 655–659. https://doi.org/10.1002/ajmg.1320420505
73. Colino-Sanguino, Y., Clark, S. J., & Valdes-Mora, F. (2022). The H2A.Z-nucleosome code in mammals: Emerging functions. *Trends in Genetics*, *38*(3), 273–289. https://doi.org/10.1016/j.tig.2021.10.003
74. Cooper, D. N., & Krawczak, M. (1989). Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Human Genetics*, *83*(2), 181–188. https://doi.org/10.1007/BF00286715
75. Coorens, T. H. H., Collord, G., Jung, H., Wang, Y., Moore, L., Hooks, Y., Mahbubani, K., Saeb-Parsy, K., Campbell, P. J., Martincorena, I., Leung, S. Y., & Stratton, M. R. (2024). *The somatic mutation landscape of normal gastric epithelium*. https://doi.org/10.1101/2024.03.17.585238
76. Cordeddu, V., Di Schiavi, E., Pennacchio, L. A., Ma'ayan, A., Sarkozy, A., Fodale, V., Cecchetti, S., Cardinale, A., Martin, J., Schackwitz, W., Lipzen, A., Zampino, G., Mazzanti, L., Digilio, M. C., Martinelli, S., Flex, E., Lepri, F., Bartholdi, D., Kutsche, K., … Tartaglia, M. (2009). Mutation of SHOC2 promotes aberrant protein N-myristoylation and causes Noonan-like syndrome with loose anagen hair. *Nature Genetics*, *41*(9), 1022–1026. https://doi.org/10.1038/ng.425
77. Corona-Rivera, J. R., Zenteno, J. C., López-Pérez, L. G., Yokoyama-Rebollar, E., Villarroel, C. E., Barragán-Arévalo, T., Montes-Almanza, L. Á., Zepeda-Romero, L. C., Morales-Domínguez, G. E., Peña-Padilla, C., Bobadilla-Morales, L., & Corona-Rivera, A. (2023). First Report of Mexican Patients with PACS1-Related Neurodevelopmental Disorder and Review of the PACS1-, PACS2-, and WDR37-Related Ophthalmological

Manifestations. *Molecular Syndromology*, *14*(2), 143–151. https://doi.org/10.1159/000526975

78. Costello, M., Pugh, T. J., Fennell, T. J., Stewart, C., Lichtenstein, L., Meldrim, J. C., Fostel, J. L., Friedrich, D. C., Perrin, D., Dionne, D., Kim, S., Gabriel, S. B., Lander, E. S., Fisher, S., & Getz, G. (2013). Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Research*, *41*(6), e67–e67. https://doi.org/10.1093/nar/gks1443
79. Cranston, A. N., Carniti, C., Oakhill, K., Radzio-Andzelm, E., Stone, E. A., McCallion, A. S., Hodgson, S., Clarke, S., Mondellini, P., Leyland, J., Pierotti, M. A., Whittaker, J., Taylor, S. S., Bongarzone, I., & Ponder, B. A. J. (2006). RET Is Constitutively Activated by Novel Tandem Mutations that Alter the Active Site Resulting in Multiple Endocrine Neoplasia Type 2B. *Cancer Research*, *66*(20), 10179–10187. https://doi.org/10.1158/0008-5472.CAN-06-0884
80. Crow, J. F. (2000). The origins, patterns and implications of human spontaneous mutation. *Nature Reviews Genetics*, *1*(1), 40–47. https://doi.org/10.1038/35049558
81. Cruz-Gamero, J. M., Ballardin, D., Lecis, B., Zhang, C.-L., Cobret, L., Gast, A., Morisset-Lopez, S., Piskorowski, R., Langui, D., Jose, J., Chevreux, G., & Rebholz, H. (2024). Missense mutation in the activation segment of the kinase CK2 models Okur-Chung neurodevelopmental disorder and alters the hippocampal glutamatergic synapse. *Molecular Psychiatry*. https://doi.org/10.1038/s41380-024-02762-8
82. Cui, Y., Hackenmiller, R., Berg, L., Jean, F., Nakayama, T., Thomas, G., & Christian, J. L. (2001). The activity and signaling range of mature BMP-4 is regulated by sequential cleavage at two sites within the prodomain of the precursor. *Genes & Development*, *15*(21), 2797–2802. https://doi.org/10.1101/gad.940001
83. Czeizel, A. E., Elek, C., & Susánszky, É. (1993). Birth prevalence study of the Apert syndrome. *American Journal of Medical Genetics*, *45*(3), 392–392. https://doi.org/10.1002/ajmg.1320450322
84. Da Silva, J. D., Costa, M. D., Almeida, B., Lopes, F., Maciel, P., & Teixeira-Castro, A. (2021). Case Report: A Novel GNB1 Mutation Causes Global Developmental Delay With Intellectual Disability and Behavioral Disorders. *Frontiers in Neurology*, *12*, 735549. https://doi.org/10.3389/fneur.2021.735549
85. Dancy, B. M., & Cole, P. A. (2015). Protein Lysine Acetylation by p300/CBP. *Chemical Reviews*, *115*(6), 2419–2452. https://doi.org/10.1021/cr500452k
86. Davidoff, M. S., Breucker, H., Holstein, A. F., & Seidl, K. (1990). Cellular architecture of the lamina propria of human seminiferous tubules. *Cell and Tissue Research*, *262*(2), 253–261. https://doi.org/10.1007/BF00309880
87. Deciphering Developmental Disorders Study. (2017). Prevalence and architecture of de novo mutations in developmental disorders. *Nature*, *542*(7642), 433–438. https://doi.org/10.1038/nature21062
88. Der, C. J., Finkel, T., & Cooper, G. M. (1986). Biological and biochemical properties of human rasH genes mutated at codon 61. *Cell*, *44*(1), 167–176. https://doi.org/10.1016/0092-8674(86)90495-2
89. Di Cesare, T., Rossi, G., Girotto, G., & Di Nardo, W. (2021). Benefit of cochlear implantation in a patient with Myhre syndrome. *BMJ Case Reports*, *14*(8), e243164. https://doi.org/10.1136/bcr-2021-243164

90. Di Persio, S., Tekath, T., Siebert-Kuss, L. M., Cremers, J.-F., Wistuba, J., Li, X., Meyer Zu Hörste, G., Drexler, H. C. A., Wyrwoll, M. J., Tüttelmann, F., Dugas, M., Kliesch, S., Schlatt, S., Laurentino, S., & Neuhaus, N. (2021). Single-cell RNA-seq unravels alterations of the human spermatogonial stem cell compartment in patients with impaired spermatogenesis. *Cell Reports Medicine*, *2*(9), 100395. https://doi.org/10.1016/j.xcrm.2021.100395
91. Di Rocco, F., Biosse Duplan, M., Heuzé, Y., Kaci, N., Komla-Ebri, D., Munnich, A., Mugniery, E., Benoist-Lasselin, C., & Legeai-Mallet, L. (2014). FGFR3 mutation causes abnormal membranous ossification in achondroplasia. *Human Molecular Genetics*, *23*(11), 2914–2925. https://doi.org/10.1093/hmg/ddu004
92. Dillon, M., Lopez, A., Lin, E., Sales, D., Perets, R., & Jain, P. (2021). Progress on Ras/MAPK Signaling Research and Targeting in Blood and Solid Cancers. *Cancers*, *13*(20), 5059. https://doi.org/10.3390/cancers13205059
93. D'Incal, C. P., Van Rossem, K. E., De Man, K., Konings, A., Van Dijck, A., Rizzuti, L., Vitriolo, A., Testa, G., Gozes, I., Vanden Berghe, W., & Kooy, R. F. (2023). Chromatin remodeler Activity-Dependent Neuroprotective Protein (ADNP) contributes to syndromic autism. *Clinical Epigenetics*, *15*(1), 45. https://doi.org/10.1186/s13148-023-01450-8
94. Ding, Z., Wu, C.-J., Chu, G. C., Xiao, Y., Ho, D., Zhang, J., Perry, S. R., Labrot, E. S., Wu, X., Lis, R., Hoshida, Y., Hiller, D., Hu, B., Jiang, S., Zheng, H., Stegh, A. H., Scott, K. L., Signoretti, S., Bardeesy, N., … DePinho, R. A. (2011). SMAD4-dependent barrier constrains prostate cancer growth and metastatic progression. *Nature*, *470*(7333), 269–273. https://doi.org/10.1038/nature09677
95. Dobosy, J. R., Rose, S. D., Beltz, K. R., Rupp, S. M., Powers, K. M., Behlke, M. A., & Walder, J. A. (2011). RNase H-dependent PCR (rhPCR): Improved specificity and single nucleotide polymorphism detection using blocked cleavable primers. *BMC Biotechnology*, *11*(1), 80. https://doi.org/10.1186/1472-6750-11-80
96. Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., & Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, *34*(2), 184–191. https://doi.org/10.1038/nbt.3437
97. Dominguez, I., Cruz-Gamero, J. M., Corasolla, V., Dacher, N., Rangasamy, S., Urbani, A., Narayanan, V., & Rebholz, H. (2021). Okur-Chung neurodevelopmental syndrome-linked CK2α variants have reduced kinase activity. *Human Genetics*, *140*(7), 1077–1096. https://doi.org/10.1007/s00439-021-02280-5
98. Dvořáková, Š., Václavíková, E., Sýkorová, V., Dušková, J., Vlček, P., Ryška, A., Novák, Z., & Bendlová, B. (2006). New Multiple Somatic Mutations in the *RET* Proto-oncogene Associated with a Sporadic Medullary Thyroid Carcinoma. *Thyroid*, *16*(3), 311–316. https://doi.org/10.1089/thy.2006.16.311
99. Eble, J. N. (1994). Spermatocytic seminoma. *Human Pathology*, *25*(10), 1035–1042. https://doi.org/10.1016/0046-8177(94)90062-0
100. Eboreime, J., Choi, S., Yoon, S., Sadybekov, A., Katritch, V., Calabrese, P., & Arnheim, N. (2022). Germline selection of *PTPN11* (HGNC:9644) variants make a major contribution to both Noonan syndrome's high birth rate and the transmission of sporadic cancer variants resulting in fetal abnormality. *Human Mutation*, *43*(12), Article 12. https://doi.org/10.1002/humu.24493

101. Eboreime, J., Choi, S.-K., Yoon, S.-R., Arnheim, N., & Calabrese, P. (2016). Estimating Exceptionally Rare Germline and Somatic Mutation Frequencies via Next Generation Sequencing. *PLOS ONE*, *11*(6), e0158340. https://doi.org/10.1371/journal.pone.0158340
102. Erdem, H. B., Sahin, I., & Tatar, A. (2018). Myhre syndrome with novel findings: Bilateral congenital cortical cataract, bilateral papilledema, accessory nipple, and adenoid hypertrophy. *Clinical Dysmorphology*, *27*(1), 12–14. https://doi.org/10.1097/MCD.0000000000000188
103. Erickson, J. D., & Michael Cohen, M. (1974). A study of parental age effects on the occurrence of fresh mutations for the Apert syndrome. *Annals of Human Genetics*, *38*(1), 89–96. https://doi.org/10.1111/j.1469-1809.1974.tb01996.x
104. Escalier, D., Silvius, D., & Xu, X. (2003). Spermatogenesis of mice lacking CK2α′: Failure of germ cell survival and characteristic modifications of the spermatid nucleus. *Molecular Reproduction and Development*, *66*(2), 190–201. https://doi.org/10.1002/mrd.10346
105. Fan, Z., Devlin, J. R., Hogg, S. J., Doyle, M. A., Harrison, P. F., Todorovski, I., Cluse, L. A., Knight, D. A., Sandow, J. J., Gregory, G., Fox, A., Beilharz, T. H., Kwiatkowski, N., Scott, N. E., Vidakovic, A. T., Kelly, G. P., Svejstrup, J. Q., Geyer, M., Gray, N. S., … Johnstone, R. W. (2020). CDK13 cooperates with CDK12 to control global RNA polymerase II processivity. *Science Advances*, *6*(18), eaaz5041. https://doi.org/10.1126/sciadv.aaz5041
106. Fang, X., Baggett, L. M., Caylor, R. C., Percy, A. K., Neul, J. L., Lane, J. B., Glaze, D. G., Benke, T. A., Marsh, E. D., Motil, K. J., Barrish, J. O., Annese, F. E., & Skinner, S. A. (2024). Parental age effects and Rett syndrome. *American Journal of Medical Genetics Part A*, *194*(2), 160–173. https://doi.org/10.1002/ajmg.a.63396
107. Ferreira, M., Beullens, M., Bollen, M., & Van Eynde, A. (2019). Functions and therapeutic potential of protein phosphatase 1: Insights from mouse genetics. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1866*(1), 16–30. https://doi.org/10.1016/j.bbamcr.2018.07.019
108. Ferreira, M., Boens, S., Winkler, C., Szekér, K., Verbinnen, I., Van Eynde, A., Fardilha, M., & Bollen, M. (2017). The protein phosphatase 1 regulator NIPP1 is essential for mammalian spermatogenesis. *Scientific Reports*, *7*(1), 13364. https://doi.org/10.1038/s41598-017-13809-y
109. Figuera, L. E. (1996). Geleophysic dysplasia vs. Myhre syndrome. *American Journal of Medical Genetics*, *65*(4), 361–361. https://doi.org/10.1002/(SICI)1096-8628(19961111)65:4<361::AID-AJMG25>3.0.CO;2-R
110. Ford, C. E., Skiba, N. P., Bae, H., Daaka, Y., Reuveny, E., Shekter, L. R., Rosal, R., Weng, G., Yang, C.-S., Lefkowitz, R. J., & Hammt, H. E. (1998). *Molecular Basis for Interactions of G Protein By Subunits with Effectors*. *280*.
111. França, L. R., Hess, R. A., Dufour, J. M., Hofmann, M. C., & Griswold, M. D. (2016). The Sertoli cell: One hundred fifty years of beauty and plasticity. *Andrology*, *4*(2), 189–212. https://doi.org/10.1111/andr.12165
112. Francioli, L. C., Genome of the Netherlands Consortium, Polak, P. P., Koren, A., Menelaou, A., Chun, S., Renkens, I., Van Duijn, C. M., Swertz, M., Wijmenga, C., Van Ommen, G., Slagboom, P. E., Boomsma, D. I., Ye, K., Guryev, V., Arndt, P. F., Kloosterman, W. P., De Bakker, P. I. W., & Sunyaev, S. R. (2015). Genome-wide

patterns and properties of de novo mutations in humans. *Nature Genetics*, *47*(7), 822–826. https://doi.org/10.1038/ng.3292

113. Fuster, J. J., MacLauchlan, S., Zuriaga, M. A., Polackal, M. N., Ostriker, A. C., Chakraborty, R., Wu, C.-L., Sano, S., Muralidharan, S., Rius, C., Vuong, J., Jacob, S., Muralidhar, V., Robertson, A. A. B., Cooper, M. A., Andrés, V., Hirschi, K. K., Martin, K. A., & Walsh, K. (2017). Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science*, *355*(6327), 842–847. https://doi.org/10.1126/science.aag1381
114. Gadzicki, D., Döcker, D., Schubach, M., Menzel, M., Schmorl, B., Stellmer, F., Biskup, S., & Bartholdi, D. (2015). Expanding the phenotype of a recurrent *de novo* variant in *PACS1* causing intellectual disability. *Clinical Genetics*, *88*(3), 300–302. https://doi.org/10.1111/cge.12544
115. Galama, W. H., Verhaagen – Van Den Akker, S. L. J., Lefeber, D. J., Feenstra, I., & Verrips, A. (2017). ALG13-CDG with Infantile Spasms in a Male Patient Due to a De Novo ALG13 Gene Mutation. In E. Morava, M. Baumgartner, M. Patterson, S. Rahman, J. Zschocke, & V. Peters (Eds.), *JIMD Reports, Volume 40* (Vol. 40, pp. 11–16). Springer Berlin Heidelberg. https://doi.org/10.1007/8904_2017_53
116. Gallione, C., Aylsworth, A. S., Beis, J., Berk, T., Bernhardt, B., Clark, R. D., Clericuzio, C., Danesino, C., Drautz, J., Fahl, J., Fan, Z., Faughnan, M. E., Ganguly, A., Garvie, J., Henderson, K., Kini, U., Leedom, T., Ludman, M., Lux, A., … Marchuk, D. A. (2010). Overlapping spectra of *SMAD4* mutations in juvenile polyposis (JP) and JP-HHT syndrome. *American Journal of Medical Genetics Part A*, *152A*(2), 333–339. https://doi.org/10.1002/ajmg.a.33206
117. Gao, X.-D., Tachikawa, H., Sato, T., Jigami, Y., & Dean, N. (2005). Alg14 Recruits Alg13 to the Cytoplasmic Face of the Endoplasmic Reticulum to Form a Novel Bipartite UDP-N-acetylglucosamine Transferase Required for the Second Step of N-Linked Glycosylation. *Journal of Biological Chemistry*, *280*(43), 36254–36262. https://doi.org/10.1074/jbc.M507569200
118. Garavelli, L., Maini, I., Baccilieri, F., Ivanovski, I., Pollazzon, M., Rosato, S., Iughetti, L., Unger, S., Superti-Furga, A., & Tartaglia, M. (2016a). Natural history and life-threatening complications in Myhre syndrome and review of the literature. *European Journal of Pediatrics*, *175*(10), 1307–1315. https://doi.org/10.1007/s00431-016-2761-3
119. Garavelli, L., Maini, I., Baccilieri, F., Ivanovski, I., Pollazzon, M., Rosato, S., Iughetti, L., Unger, S., Superti-Furga, A., & Tartaglia, M. (2016b). Natural history and life-threatening complications in Myhre syndrome and review of the literature. *European Journal of Pediatrics*, *175*(10), 1307–1315. https://doi.org/10.1007/s00431-016-2761-3
120. Garcia-Cruz, D., Figuera, L. E., Feria-Velazco, A., Sánchez-Corona, J., Garcia-Cruz, M. O., Ramirez-Duenãs, R. M., Hernandez-Córdova, A., Ruiz, M. X., Bitar-Alatorre, W. E., Ramirez-Dueñas, M. L., & Cantú, J. M. (1993). The Myhre syndrome: Report of two cases. *Clinical Genetics*, *44*(4), 203–207. https://doi.org/10.1111/j.1399-0004.1993.tb03880.x
121. Genovese, G., Kähler, A. K., Handsaker, R. E., Lindberg, J., Rose, S. A., Bakhoum, S. F., Chambert, K., Mick, E., Neale, B. M., Fromer, M., Purcell, S. M., Svantesson, O., Landén, M., Höglund, M., Lehmann, S., Gabriel, S. B., Moran, J. L., Lander, E. S., Sullivan, P. F., … McCarroll, S. A. (2014). Clonal Hematopoiesis and Blood-Cancer

Risk Inferred from Blood DNA Sequence. *New England Journal of Medicine*, *371*(26), 2477–2487. https://doi.org/10.1056/NEJMoa1409405

122. Giaimo, B. D., Ferrante, F., Herchenröther, A., Hake, S. B., & Borggrefe, T. (2019). The histone variant H2A.Z in gene regulation. *Epigenetics & Chromatin*, *12*(1), 37. https://doi.org/10.1186/s13072-019-0274-9
123. Giaimo, B. D., Ferrante, F., Vallejo, D. M., Hein, K., Gutierrez-Perez, I., Nist, A., Stiewe, T., Mittler, G., Herold, S., Zimmermann, T., Bartkuhn, M., Schwarz, P., Oswald, F., Dominguez, M., & Borggrefe, T. (2018). Histone variant H2A.Z deposition and acetylation directs the canonical Notch signaling response. *Nucleic Acids Research*, *46*(16), 8197–8215. https://doi.org/10.1093/nar/gky551
124. Giannoulatou, E., Maher, G. J., Ding, Z., Gillis, A. J. M., Dorssers, L. C. J., Hoischen, A., Rajpert-De Meyts, E., WGS500 Consortium, McVean, G., Wilkie, A. O. M., Looijenga, L. H. J., & Goriely, A. (2017). Whole-genome sequencing of spermatocytic tumors provides insights into the mutational processes operating in the male germline. *PLOS ONE*, *12*(5), e0178169. https://doi.org/10.1371/journal.pone.0178169
125. Giannoulatou, E., McVean, G., Taylor, I. B., McGowan, S. J., Maher, G. J., Iqbal, Z., Pfeifer, S. P., Turner, I., Burkitt Wright, E. M. M., Shorto, J., Itani, A., Turner, K., Gregory, L., Buck, D., Rajpert-De Meyts, E., Looijenga, L. H. J., Kerr, B., Wilkie, A. O. M., & Goriely, A. (2013). Contributions of intrinsic mutation rate and selfish selection to levels of de novo *HRAS* mutations in the paternal germline. *Proceedings of the National Academy of Sciences*, *110*(50), 20152–20157. https://doi.org/10.1073/pnas.1311381110
126. Giudicelli, M. D., Serazin, V., Le Sciellour, C. R., Albert, M., Selva, J., & Giudicelli, Y. (2008). Increased achondroplasia mutation frequency with advanced age and evidence for G1138A mosaicism in human testis biopsies. *Fertility and Sterility*, *89*(6), 1651–1656. https://doi.org/10.1016/j.fertnstert.2007.04.037
127. Glaser, R. L., Broman, K. W., Schulman, R. L., Eskenazi, B., Wyrobek, A. J., & Jabs, E. W. (2003). The Paternal-Age Effect in Apert Syndrome Is Due, in Part, to the Increased Frequency of Mutations in Sperm. *The American Journal of Human Genetics*, *73*(4), 939–947. https://doi.org/10.1086/378419
128. Goldmann, J. M., Seplyarskiy, V. B., Wong, W. S. W., Vilboux, T., Neerincx, P. B., Bodian, D. L., Solomon, B. D., Veltman, J. A., Deeken, J. F., Gilissen, C., & Niederhuber, J. E. (2018). Germline de novo mutation clusters arise during oocyte aging in genomic regions with high double-strand-break incidence. *Nature Genetics*, *50*(4), 487–492. https://doi.org/10.1038/s41588-018-0071-6
129. Goldmann, J. M., Veltman, J. A., & Gilissen, C. (2019). De Novo Mutations Reflect Development and Aging of the Human Germline. *Trends in Genetics*, *35*(11), 828–839. https://doi.org/10.1016/j.tig.2019.08.005
130. Goldmann, J. M., Wong, W. S. W., Pinelli, M., Farrah, T., Bodian, D., Stittrich, A. B., Glusman, G., Vissers, L. E. L. M., Hoischen, A., Roach, J. C., Vockley, J. G., Veltman, J. A., Solomon, B. D., Gilissen, C., & Niederhuber, J. E. (2016). Parent-of-origin-specific signatures of de novo mutations. *Nature Genetics*, *48*(8), 935–939. https://doi.org/10.1038/ng.3597
131. Goriely, A., Hansen, R. M. S., Taylor, I. B., Olesen, I. A., Jacobsen, G. K., McGowan, S. J., Pfeifer, S. P., McVean, G. A. T., Meyts, E. R.-D., & Wilkie, A. O. M. (2009). Activating mutations in FGFR3 and HRAS reveal a shared genetic origin for congenital

disorders and testicular tumors. *Nature Genetics*, *41*(11), 1247–1252. https://doi.org/10.1038/ng.470

132. Goriely, A., McGrath, J. J., Hultman, C. M., Wilkie, A. O. M., & Malaspina, D. (2013). "Selfish Spermatogonial Selection": A Novel Mechanism for the Association Between Advanced Paternal Age and Neurodevelopmental Disorders. *American Journal of Psychiatry*, *170*(6), 599–608. https://doi.org/10.1176/appi.ajp.2013.12101352
133. Goriely, A., McVean, G. A. T., Röjmyr, M., Ingemarsson, B., & Wilkie, A. O. M. (2003). Evidence for Selective Advantage of Pathogenic FGFR2 Mutations in the Male Germ Line. *Science*, *301*(5633), 643–646. https://doi.org/10.1126/science.1085710
134. Goriely, A., McVean, G. A. T., Van Pelt, A. M. M., O'Rourke, A. W., Wall, S. A., De Rooij, D. G., & Wilkie, A. O. M. (2005). Gain-of-function amino acid substitutions drive positive selection of *FGFR2* mutations in human spermatogonia. *Proceedings of the National Academy of Sciences*, *102*(17), 6051–6056. https://doi.org/10.1073/pnas.0500267102
135. Goriely, A., & Wilkie, A. O. M. (2012). Paternal Age Effect Mutations and Selfish Spermatogonial Selection: Causes and Consequences for Human Disease. *The American Journal of Human Genetics*, *90*(2), 175–200. https://doi.org/10.1016/j.ajhg.2011.12.017
136. Greenberg, D. R., Khandwala, Y. S., Lu, Y., Stevenson, D. K., Shaw, G. M., & Eisenberg, M. L. (2020). Disease burden in offspring is associated with changing paternal demographics in the United States. *Andrology*, *8*(2), 342–347. https://doi.org/10.1111/andr.12700
137. Greifenberg, A. K., Hönig, D., Pilarova, K., Düster, R., Bartholomeeusen, K., Bösken, C. A., Anand, K., Blazek, D., & Geyer, M. (2016). Structural and Functional Analysis of the Cdk13/Cyclin K Complex. *Cell Reports*, *14*(2), 320–331. https://doi.org/10.1016/j.celrep.2015.12.025
138. Gripp, K. W., Aldinger, K. A., Bennett, J. T., Baker, L., Tusi, J., Powell-Hamilton, N., Stabley, D., Sol-Church, K., Timms, A. E., & Dobyns, W. B. (2016). A novel rasopathy caused by recurrent de novo missense mutations in *PPP1CB* closely resembles Noonan syndrome with loose anagen hair. *American Journal of Medical Genetics Part A*, *170*(9), 2237–2247. https://doi.org/10.1002/ajmg.a.37781
139. Gripp, K. W., Morse, L. A., Axelrad, M., Chatfield, K. C., Chidekel, A., Dobyns, W., Doyle, D., Kerr, B., Lin, A. E., Schwartz, D. D., Sibbles, B. J., Siegel, D., Shankar, S. P., Stevenson, D. A., Thacker, M. M., Weaver, K. N., White, S. M., & Rauen, K. A. (2019a). Costello syndrome: Clinical phenotype, genotype, and management guidelines. *American Journal of Medical Genetics Part A*, *179*(9), 1725–1744. https://doi.org/10.1002/ajmg.a.61270
140. Gripp, K. W., Morse, L. A., Axelrad, M., Chatfield, K. C., Chidekel, A., Dobyns, W., Doyle, D., Kerr, B., Lin, A. E., Schwartz, D. D., Sibbles, B. J., Siegel, D., Shankar, S. P., Stevenson, D. A., Thacker, M. M., Weaver, K. N., White, S. M., & Rauen, K. A. (2019b). Costello syndrome: Clinical phenotype, genotype, and management guidelines. *American Journal of Medical Genetics Part A*, *179*(9), 1725–1744. https://doi.org/10.1002/ajmg.a.61270
141. Gripp, K. W., Zand, D. J., Demmer, L., Anderson, C. E., Dobyns, W. B., Zackai, E. H., Denenberg, E., Jenny, K., Stabley, D. L., & Sol-Church, K. (2013). Expanding the SHOC2 mutation associated phenotype of noonan syndrome with loose anagen hair:

Structural brain anomalies and myelofibrosis. *American Journal of Medical Genetics Part A*, *161*(10), 2420–2430. https://doi.org/10.1002/ajmg.a.36098
142. Guo, J., Grow, E. J., Mlcochova, H., Maher, G. J., Lindskog, C., Nie, X., Guo, Y., Takei, Y., Yun, J., Cai, L., Kim, R., Carrell, D. T., Goriely, A., Hotaling, J. M., & Cairns, B. R. (2018). The adult human testis transcriptional cell atlas. *Cell Research*, *28*(12), 1141–1157. https://doi.org/10.1038/s41422-018-0099-2
143. Guo, J., Grow, E. J., Yi, C., Mlcochova, H., Maher, G. J., Lindskog, C., Murphy, P. J., Wike, C. L., Carrell, D. T., Goriely, A., Hotaling, J. M., & Cairns, B. R. (2017). Chromatin and Single-Cell RNA-Seq Profiling Reveal Dynamic Signaling and Metabolic Transitions during Human Spermatogonial Stem Cell Development. *Cell Stem Cell*, *21*(4), 533-546.e6. https://doi.org/10.1016/j.stem.2017.09.003
144. Gürsoy, S., Hazan, F., Öztürk, T., & Ateş, H. (2019). Novel Ocular and Inner Ear Anomalies in a Patient with Myhre Syndrome. *Molecular Syndromology*, *10*(6), 339–343. https://doi.org/10.1159/000504829
145. Hafner, C., Hartmann, A., Van Oers, J. M. M., Stoehr, R., Zwarthoff, E. C., Hofstaedter, F., Landthaler, M., & Vogt, T. (2007). FGFR3 mutations in seborrheic keratoses are already present in flat lesions and associated with age and localization. *Modern Pathology*, *20*(8), 895–903. https://doi.org/10.1038/modpathol.3800837
146. Hafner, C., Toll, A., Fernández-Casado, A., Earl, J., Marqués, M., Acquadro, F., Méndez-Pertuz, M., Urioste, M., Malats, N., Burns, J. E., Knowles, M. A., Cigudosa, J. C., Hartmann, A., Vogt, T., Landthaler, M., Pujol, R. M., & Real, F. X. (2010). Multiple oncogenic mutations and clonal relationship in spatially distinct benign human epidermal tumors. *Proceedings of the National Academy of Sciences*, *107*(48), 20780–20785. https://doi.org/10.1073/pnas.1008365107
147. Hafner, C., Vogt, T., & Hartmann, A. (2006). FGFR3 Mutations in Benign Skin Tumors. *Cell Cycle*, *5*(23), 2723–2728. https://doi.org/10.4161/cc.5.23.3509
148. Hahn, S. A., Schutte, M., Moskaluk, C. A., Weinstein, C. L., Fischer, A., Yeo, C. J., Hruban, R. H., & Kern, S. E. (n.d.). *DPC4, A Candidate Tumor Suppressor Gene at Human Chromosome 18q21.I*.
149. Haldane, J. B. S. (1946). THE MUTATION RATE OF THE GENE FOR HAEMOPHILIA, AND ITS SEGREGATION RATIOS IN MALES AND FEMALES. *Annals of Eugenics*, *13*(1), 262–271. https://doi.org/10.1111/j.1469-1809.1946.tb02367.x
150. Halloran, D., Pandit, V., & Nohe, A. (2022). The Role of Protein Kinase CK2 in Development and Disease Progression: A Critical Review. *Journal of Developmental Biology*, *10*(3), 31. https://doi.org/10.3390/jdb10030031
151. Hamici, S., Bastaki, F., & Khalifa, M. (2017). Exome sequence identified a c.320A > G ALG13 variant in a female with infantile epileptic encephalopathy with normal glycosylation and random X inactivation: Review of the literature. *European Journal of Medical Genetics*, *60*(10), 541–547. https://doi.org/10.1016/j.ejmg.2017.07.014
152. Hamilton, M. J., Caswell, R. C., Canham, N., Cole, T., Firth, H. V., Foulds, N., Heimdal, K., Hobson, E., Houge, G., Joss, S., Kumar, D., Lampe, A. K., Maystadt, I., McKay, V., Metcalfe, K., Newbury-Ecob, R., Park, S.-M., Robert, L., Rustad, C. F., … Suri, M. (2018). Heterozygous mutations affecting the protein kinase domain of *CDK13* cause a syndromic form of developmental delay and intellectual disability. *Journal of Medical Genetics*, *55*(1), 28–38. https://doi.org/10.1136/jmedgenet-2017-104620

153. Hamilton, M. J., & Suri, M. (2019). CDK13-related disorder. In *Advances in Genetics* (Vol. 103, pp. 163–182). Elsevier. https://doi.org/10.1016/bs.adgen.2018.11.001
154. Hansen, K. B., Yi, F., Perszyk, R. E., Menniti, F. S., & Traynelis, S. F. (2017). NMDA Receptors in the Central Nervous System. In N. Burnashev & P. Szepetowski (Eds.), *NMDA Receptors* (Vol. 1677, pp. 1–80). Springer New York. https://doi.org/10.1007/978-1-4939-7321-7_1
155. Hansen, R. M., Goriely, A., Wall, S. A., Roberts, I. S., & Wilkie, A. O. (2005). Fibroblast growth factor receptor 2, gain-of-function mutations, and tumourigenesis: Investigating a potential link. *The Journal of Pathology*, *207*(1), 27–31. https://doi.org/10.1002/path.1816
156. Hauseman, Z. J., Fodor, M., Dhembi, A., Viscomi, J., Egli, D., Bleu, M., Katz, S., Park, E., Jang, D. M., Porter, K. A., Meili, F., Guo, H., Kerr, G., Mollé, S., Velez-Vega, C., Beyer, K. S., Galli, G. G., Maira, S.-M., Stams, T., … King, D. A. (2022a). Structure of the MRAS–SHOC2–PP1C phosphatase complex. *Nature*, *609*(7926), 416–423. https://doi.org/10.1038/s41586-022-05086-1
157. Hauseman, Z. J., Fodor, M., Dhembi, A., Viscomi, J., Egli, D., Bleu, M., Katz, S., Park, E., Jang, D. M., Porter, K. A., Meili, F., Guo, H., Kerr, G., Mollé, S., Velez-Vega, C., Beyer, K. S., Galli, G. G., Maira, S.-M., Stams, T., … King, D. A. (2022b). Structure of the MRAS–SHOC2–PP1C phosphatase complex. *Nature*, *609*(7926), 416–423. https://doi.org/10.1038/s41586-022-05086-1
158. Hawkes, L., & Kini, U. (2015). Myhre syndrome with facial paralysis and branch pulmonary stenosis. *Clinical Dysmorphology*, *24*(2), 84–85. https://doi.org/10.1097/MCD.0000000000000068
159. He, Y., Sun, M. M., Zhang, G. G., Yang, J., Chen, K. S., Xu, W. W., & Li, B. (2021). Targeting PI3K/Akt signal transduction for cancer therapy. *Signal Transduction and Targeted Therapy*, *6*(1), 425. https://doi.org/10.1038/s41392-021-00828-5
160. Heller, C. G., & Clermont, Y. (1963). Spermatogenesis in Man: An Estimate of Its Duration. *Science*, *140*(3563), 184–186. https://doi.org/10.1126/science.140.3563.184
161. Helsmoortel, C., Vulto-van Silfhout, A. T., Coe, B. P., Vandeweyer, G., Rooms, L., Van Den Ende, J., Schuurs-Hoeijmakers, J. H. M., Marcelis, C. L., Willemsen, M. H., Vissers, L. E. L. M., Yntema, H. G., Bakshi, M., Wilson, M., Witherspoon, K. T., Malmgren, H., Nordgren, A., Annerén, G., Fichera, M., Bosco, P., … Van Der Aa, N. (2014). A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. *Nature Genetics*, *46*(4), 380–384. https://doi.org/10.1038/ng.2899
162. Hertz, E. P. T., Kruse, T., Davey, N. E., López-Méndez, B., Sigurðsson, J. O., Montoya, G., Olsen, J. V., & Nilsson, J. (2016). A Conserved Motif Provides Binding Specificity to the PP2A-B56 Phosphatase. *Molecular Cell*, *63*(4), 686–695. https://doi.org/10.1016/j.molcel.2016.06.024
163. Hildebrand, J. D., & Soriano, P. (2002). Overlapping and Unique Roles for C-Terminal Binding Protein 1 (CtBP1) and CtBP2 during Mouse Development. *Molecular and Cellular Biology*, *22*(15), 5296–5307. https://doi.org/10.1128/MCB.22.15.5296-5307.2002
164. Hoban, R., Roberts, A. E., Demmer, L., Jethva, R., & Shephard, B. (2012). Noonan syndrome due to a *SHOC2* mutation presenting with fetal distress and fatal hypertrophic cardiomyopathy in a premature infant. *American Journal of Medical Genetics Part A*, *158A*(6), 1411–1413. https://doi.org/10.1002/ajmg.a.35318

165. Holmes, W. M., Mannakee, B. K., Gutenkunst, R. N., & Serio, T. R. (2014). Loss of amino-terminal acetylation suppresses a prion phenotype by modulating global protein folding. *Nature Communications*, *5*(1), 4383. https://doi.org/10.1038/ncomms5383
166. Holstein, A.-F., Schulze, W., & Davidoff, M. (2003). Understanding spermatogenesis is a prerequisite for treatment. *Reproductive Biology and Endocrinology*, *1*(1), 107. https://doi.org/10.1186/1477-7827-1-107
167. Hood, R. L., Lines, M. A., Nikkel, S. M., Schwartzentruber, J., Beaulieu, C., Nowaczyk, M. J. M., Allanson, J., Kim, C. A., Wieczorek, D., Moilanen, J. S., Lacombe, D., Gillessen-Kaesbach, G., Whiteford, M. L., Quaio, C. R. D. C., Gomy, I., Bertola, D. R., Albrecht, B., Platzer, K., McGillivray, G., … Boycott, K. M. (2012). Mutations in SRCAP, Encoding SNF2-Related CREBBP Activator Protein, Cause Floating-Harbor Syndrome. *The American Journal of Human Genetics*, *90*(2), 308–313. https://doi.org/10.1016/j.ajhg.2011.12.001
168. Hopkin, R. J., Cotton, R., Langer, L. O., & Saal, H. M. (1998). Progressive laryngotracheal stenosis with short stature and arthropathy. *American Journal of Medical Genetics*, *80*(3), 241–246. https://doi.org/10.1002/(SICI)1096-8628(19981116)80:3<241::AID-AJMG12>3.0.CO;2-#
169. Houge, G., Haesen, D., Vissers, L. E. L. M., Mehta, S., Parker, M. J., Wright, M., Vogt, J., McKee, S., Tolmie, J. L., Cordeiro, N., Kleefstra, T., Willemsen, M. H., Reijnders, M. R. F., Berland, S., Hayman, E., Lahat, E., Brilstra, E. H., Van Gassen, K. L. I., Zonneveld-Huijssoon, E., … Janssens, V. (2015). B56δ-related protein phosphatase 2A dysfunction identified in patients with intellectual disability. *Journal of Clinical Investigation*, *125*(8), 3051–3062. https://doi.org/10.1172/JCI79860
170. Hu, C., Chen, W., Myers, S. J., Yuan, H., & Traynelis, S. F. (2016). Human GRIN2B variants in neurodevelopmental disorders. *Journal of Pharmacological Sciences*, *132*(2), 115–121. https://doi.org/10.1016/j.jphs.2016.10.002
171. Hu, H., Wang, B., Borde, M., Nardone, J., Maika, S., Allred, L., Tucker, P. W., & Rao, A. (2006). Foxp1 is an essential transcriptional regulator of B cell development. *Nature Immunology*, *7*(8), 819–826. https://doi.org/10.1038/ni1358
172. Hurst, L. D., & Ellegren, H. (1998). Sex biases in the mutation rate. *Trends in Genetics*, *14*(11), 446–452. https://doi.org/10.1016/S0168-9525(98)01577-7
173. Ibrahimi, O. A., Eliseenkova, A. V., Plotnikov, A. N., Yu, K., Ornitz, D. M., & Mohammadi, M. (2001). Structural basis for fibroblast growth factor receptor 2 activation in Apert syndrome. *Proceedings of the National Academy of Sciences*, *98*(13), 7182–7187. https://doi.org/10.1073/pnas.121183798
174. Im, K., Mareninov, S., Diaz, M. F. P., & Yong, W. H. (2019). An Introduction to Performing Immunofluorescence Staining. In W. H. Yong (Ed.), *Biobanking* (Vol. 1897, pp. 299–311). Springer New York. https://doi.org/10.1007/978-1-4939-8935-5_26
175. Inoue, K., Eiro, T., Semoto, M., Roppongi, T., Nomoto, M., Takahashi, Y., & Hishimoto, A. (2021). First case of Myhre syndrome with schizophrenia. *Clinical Dysmorphology*, *30*(4), 207–208. https://doi.org/10.1097/MCD.0000000000000386
176. Ip, J. P. K., Mellios, N., & Sur, M. (2018). Rett syndrome: Insights into genetic, molecular and circuit mechanisms. *Nature Reviews Neuroscience*, *19*(6), 368–382. https://doi.org/10.1038/s41583-018-0006-3

177. Ishibashi, N., Sasaki, Y., & Asakura, Y. (2014). Myhre syndrome: A rare craniofacial disorder. *CRANIO®*, *32*(4), 300–306. https://doi.org/10.1179/0886963413Z.00000000024
178. Ishii, K., Kanatsu-Shinohara, M., Toyokuni, S., & Shinohara, T. (2012). FGF2 mediates mouse spermatogonial stem cell self-renewal via upregulation of *Etv5* and *Bcl6b* through MAP2K1 activation. *Development*, *139*(10), 1734–1743. https://doi.org/10.1242/dev.076539
179. Iyer, N. K., Burke, C. A., Leach, B. H., & Parambil, J. G. (2010). SMAD4 mutation and the combined syndrome of juvenile polyposis syndrome and hereditary haemorrhagic telangiectasia. *Thorax*, *65*(8), 745–746. https://doi.org/10.1136/thx.2009.129932
180. Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P. V., Mar, B. G., Lindsley, R. C., Mermel, C. H., Burtt, N., Chavez, A., Higgins, J. M., Moltchanov, V., Kuo, F. C., Kluk, M. J., Henderson, B., Kinnunen, L., Koistinen, H. A., Ladenvall, C., Getz, G., … Ebert, B. L. (2014). Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine*, *371*(26), 2488–2498. https://doi.org/10.1056/NEJMoa1408617
181. Jan, S. Z., Vormer, T. L., Jongejan, A., Röling, M. D., Silber, S. J., De Rooij, D. G., Hamer, G., Repping, S., & Van Pelt, A. M. M. (2017). Unraveling transcriptome dynamics in human spermatogenesis. *Development*, *144*(20), 3659–3673. https://doi.org/10.1242/dev.152413
182. Jay, K., Mitra, A., Harding, T., Matthes, D., & Van Ness, B. (2019). Identification of a de novo *FOXP1* mutation and incidental discovery of inherited genetic variants contributing to a case of autism spectrum disorder and epilepsy. *Molecular Genetics & Genomic Medicine*, *7*(7), e00751. https://doi.org/10.1002/mgg3.751
183. Jensen, B., James, R., Hong, Y., Omoyinmi, E., Pilkington, C., Sebire, N. J., Howell, K. J., Brogan, P. A., & Eleftheriou, D. (2020a). A case of Myhre syndrome mimicking juvenile scleroderma. *Pediatric Rheumatology*, *18*(1), 72. https://doi.org/10.1186/s12969-020-00466-1
184. Jensen, B., James, R., Hong, Y., Omoyinmi, E., Pilkington, C., Sebire, N. J., Howell, K. J., Brogan, P. A., & Eleftheriou, D. (2020b). A case of Myhre syndrome mimicking juvenile scleroderma. *Pediatric Rheumatology*, *18*(1), 72. https://doi.org/10.1186/s12969-020-00466-1
185. Jensen, B., James, R., Hong, Y., Omoyinmi, E., Pilkington, C., Sebire, N. J., Howell, K. J., Brogan, P. A., & Eleftheriou, D. (2020c). A case of Myhre syndrome mimicking juvenile scleroderma. *Pediatric Rheumatology*, *18*(1), 72. https://doi.org/10.1186/s12969-020-00466-1
186. Jeon, M. J., Kim, M. J., Kim, J. H., Park, J. S., Yim, J., Kim, M., Kwon, S. K., Lee, S., Ko, J. M., Chae, J.-H., & Suh, D. I. (2021). Multilevel Airway Stenosis Being Bypassed by a Customized Tracheostomy Tube in an Infant with Myhre Syndrome. *Pediatric Allergy, Immunology, and Pulmonology*, *34*(2), 83–87. https://doi.org/10.1089/ped.2021.0029
187. Jónsson, H., Sulem, P., Kehr, B., Kristmundsdottir, S., Zink, F., Hjartarson, E., Hardarson, M. T., Hjorleifsson, K. E., Eggertsson, H. P., Gudjonsson, S. A., Ward, L. D., Arnadottir, G. A., Helgason, E. A., Helgason, H., Gylfason, A., Jonasdottir, A., Jonasdottir, A., Rafnar, T., Frigge, M., … Stefansson, K. (2017). Parental influence on

human germline de novo mutations in 1,548 trios from Iceland. *Nature*, *549*(7673), 519–522. https://doi.org/10.1038/nature24018

188. Kandhaya-Pillai, R., Hou, D., Zhang, J., Yang, X., Compoginis, G., Mori, T., Tchkonia, T., Martin, G. M., Hisama, F. M., Kirkland, J. L., & Oshima, J. (2021). SMAD4 mutations and cross-talk between TGF-β/IFNγ signaling accelerate rates of DNA damage and cellular senescence, resulting in a segmental progeroid syndrome—The Myhre syndrome. *GeroScience*, *43*(3), 1481–1496. https://doi.org/10.1007/s11357-020-00318-6
189. Kaplanis, J., Samocha, K. E., Wiel, L., Zhang, Z., Arvai, K. J., Eberhardt, R. Y., Gallone, G., Lelieveld, S. H., Martin, H. C., McRae, J. F., Short, P. J., Torene, R. I., De Boer, E., Danecek, P., Gardner, E. J., Huang, N., Lord, J., Martincorena, I., Pfundt, R., … Retterer, K. (2020). Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature*, *586*(7831), 757–762. https://doi.org/10.1038/s41586-020-2832-5
190. Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J., Dusart, P., Edfors, F., Oksvold, P., Ozcan, M., Mardonoglu, A., Fagerberg, L., Mulder, J., Luo, Y., Ponten, F., Uhlén, M., & Lindskog, C. (2021). A single–cell type transcriptomics map of human tissues. *SCIENCE ADVANCES*.
191. Kaye, E. G., Basavaraju, K., Nelson, G. M., Zomer, H. D., Roy, D., Joseph, I. I., Rajabi-Toustani, R., Qiao, H., Adelman, K., & Reddi, P. P. (2024). RNA polymerase II pausing is essential during spermatogenesis for appropriate gene expression and completion of meiosis. *Nature Communications*, *15*(1), 848. https://doi.org/10.1038/s41467-024-45177-3
192. Kenis, C., Verstreken, M., Gieraerts, K., De Foer, B., Van Der Aa, N., Offeciers, E. F., & Casselman, J. W. (2014). Bilateral Otospongiosis and a Unilateral Vestibular Schwannoma in a Patient With Myhre Syndrome. *Otology & Neurotology*, *35*(9), e253–e255. https://doi.org/10.1097/MAO.0000000000000314
193. Kennedy, S. R., Schmitt, M. W., Fox, E. J., Kohrn, B. F., Salk, J. J., Ahn, E. H., Prindle, M. J., Kuong, K. J., Shen, J.-C., Risques, R.-A., & Loeb, L. A. (2014). Detecting ultralow-frequency mutations by Duplex Sequencing. *Nature Protocols*, *9*(11), 2586–2606. https://doi.org/10.1038/nprot.2014.170
194. Khater, M., Bryant, C. N., & Wu, G. (2021). Gβγ translocation to the Golgi apparatus activates ARF1 to spatiotemporally regulate G protein–coupled receptor signaling to MAPK. *Journal of Biological Chemistry*, *296*, 100805. https://doi.org/10.1016/j.jbc.2021.100805
195. Kilci, F., Hürmüzlü-Kozler, S., Jones, J., Doğan, K., Cerrah Güneş, M., & Çizmecioğlu-Jones, F. M. (2022). Myhre syndrome associated with hyperinsulinism and impaired glucose tolerance: A novel finding. *Clinical Dysmorphology*, *31*(1), 42–44. https://doi.org/10.1097/MCD.0000000000000396
196. Kitadate, Y., Jörg, D. J., Tokue, M., Maruyama, A., Ichikawa, R., Tsuchiya, S., Segi-Nishida, E., Nakagawa, T., Uchida, A., Kimura-Yoshida, C., Mizuno, S., Sugiyama, F., Azami, T., Ema, M., Noda, C., Kobayashi, S., Matsuo, I., Kanai, Y., Nagasawa, T., … Yoshida, S. (2019). Competition for Mitogens Regulates Spermatogenic Stem Cell Homeostasis in an Open Niche. *Cell Stem Cell*, *24*(1), 79-92.e6. https://doi.org/10.1016/j.stem.2018.11.013
197. Klein, A. M., Nakagawa, T., Ichikawa, R., Yoshida, S., & Simons, B. D. (2010). Mouse Germ Line Stem Cells Undergo Rapid and Stochastic Turnover. *Cell Stem Cell*, *7*(2), 214–224. https://doi.org/10.1016/j.stem.2010.05.017

198. Knudson, A. G. (1971). Mutation and Cancer: Statistical Study of Retinoblastoma. *Proceedings of the National Academy of Sciences*, *68*(4), 820–823. https://doi.org/10.1073/pnas.68.4.820
199. Koelling, N., Bernkopf, M., Calpena, E., Maher, G. J., Miller, K. A., Ralph, H. K., Goriely, A., & Wilkie, A. O. M. (2019). amplimap: A versatile tool to process and analyze targeted NGS data. *Bioinformatics*, *35*(24), 5349–5350. https://doi.org/10.1093/bioinformatics/btz582
200. Kondrashov, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature*, *336*(6198), 435–440. https://doi.org/10.1038/336435a0
201. Kondrashov, A. S. (2003). Direct estimates of human per nucleotide mutation rates at 20 loci causing mendelian diseases. *Human Mutation*, *21*(1), 12–27. https://doi.org/10.1002/humu.10147
202. Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S. A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., Wong, W. S. W., Sigurdsson, G., Walters, G. B., Steinberg, S., Helgason, H., Thorleifsson, G., Gudbjartsson, D. F., Helgason, A., Magnusson, O. Th., … Stefansson, K. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature*, *488*(7412), 471–475. https://doi.org/10.1038/nature11396
203. Krishnaraj, R., Ho, G., & Christodoulou, J. (2017). RettBASE: Rett syndrome database update. *Human Mutation*, *38*(8), 922–931. https://doi.org/10.1002/humu.23263
204. Krygowska, A. A., & Castellano, E. (2018). PI3K: A Crucial Piece in the RAS Signaling Puzzle. *Cold Spring Harbor Perspectives in Medicine*, *8*(6), a031450. https://doi.org/10.1101/cshperspect.a031450
205. Kuo, Y.-C., Huang, K.-Y., Yang, C.-H., Yang, Y.-S., Lee, W.-Y., & Chiang, C.-W. (2008). Regulation of Phosphorylation of Thr-308 of Akt, Cell Proliferation, and Survival by the B55α Regulatory Subunit Targeting of the Protein Phosphatase 2A Holoenzyme to Akt. *Journal of Biological Chemistry*, *283*(4), 1882–1892. https://doi.org/10.1074/jbc.M709585200
206. Kuppuswamy, M., Vijayalingam, S., Zhao, L.-J., Zhou, Y., Subramanian, T., Ryerse, J., & Chinnadurai, G. (2008). Role of the PLDLS-Binding Cleft Region of CtBP1 in Recruitment of Core and Auxiliary Components of the Corepressor Complex. *Molecular and Cellular Biology*, *28*(1), 269–281. https://doi.org/10.1128/MCB.01077-07
207. Kweon, H. Y., Lee, M.-N., Dorfel, M., Seo, S., Gottlieb, L., PaPazyan, T., McTiernan, N., Ree, R., Bolton, D., Garcia, A., Flory, M., Crain, J., Sebold, A., Lyons, S., Ismail, A., Marchi, E., Sonn, S., Jeong, S.-J., Jeon, S., … Lyon, G. J. (2021). Naa12 compensates for Naa10 in mice in the amino-terminal acetylation pathway. *eLife*, *10*, e65952. https://doi.org/10.7554/eLife.65952
208. Lambrecht, C., Libbrecht, L., Sagaert, X., Pauwels, P., Hoorne, Y., Crowther, J., Louis, J. V., Sents, W., Sablina, A., & Janssens, V. (2018). Loss of protein phosphatase 2A regulatory subunit B56δ promotes spontaneous tumorigenesis in vivo. *Oncogene*, *37*(4), 544–552. https://doi.org/10.1038/onc.2017.350
209. Lanauze, C. B., Sehgal, P., Hayer, K., Torres-Diz, M., Pippin, J. A., Grant, S. F. A., & Thomas-Tikhonenko, A. (2021). Colorectal Cancer-Associated Smad4 R361 Hotspot Mutations Boost Wnt/β-Catenin Signaling through Enhanced Smad4–LEF1 Binding. *Molecular Cancer Research*, *19*(5), 823–833. https://doi.org/10.1158/1541-7786.MCR-20-0721

210. Lawrence, M. S., Stojanov, P., Mermel, C. H., Robinson, J. T., Garraway, L. A., Golub, T. R., Meyerson, M., Gabriel, S. B., Lander, E. S., & Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*, *505*(7484), 495–501. https://doi.org/10.1038/nature12912
211. Lawson, A. R. J., Abascal, F., Coorens, T. H. H., Hooks, Y., O'Neill, L., Latimer, C., Raine, K., Sanders, M. A., Warren, A. Y., Mahbubani, K. T. A., Bareham, B., Butler, T. M., Harvey, L. M. R., Cagan, A., Menzies, A., Moore, L., Colquhoun, A. J., Turner, W., Thomas, B., … Martincorena, I. (2020). Extensive heterogeneity in somatic mutation and selection in the human bladder. *Science*, *370*(6512), 75–82. https://doi.org/10.1126/science.aba8347
212. Le Fevre, A. K., Taylor, S., Malek, N. H., Horn, D., Carr, C. W., Abdul-Rahman, O. A., O'Donnell, S., Burgess, T., Shaw, M., Gecz, J., Bain, N., Fagan, K., & Hunter, M. F. (2013). *FOXP1* mutations cause intellectual disability and a recognizable phenotype. *American Journal of Medical Genetics Part A*, *161*(12), 3166–3175. https://doi.org/10.1002/ajmg.a.36174
213. Le Goff, C., Mahaut, C., Abhyankar, A., Le Goff, W., Serre, V., Afenjar, A., Destrée, A., Di Rocco, M., Héron, D., Jacquemont, S., Marlin, S., Simon, M., Tolmie, J., Verloes, A., Casanova, J.-L., Munnich, A., & Cormier-Daire, V. (2012a). Mutations at a single codon in Mad homology 2 domain of SMAD4 cause Myhre syndrome. *Nature Genetics*, *44*(1), 85–88. https://doi.org/10.1038/ng.1016
214. Le Goff, C., Mahaut, C., Abhyankar, A., Le Goff, W., Serre, V., Afenjar, A., Destrée, A., Di Rocco, M., Héron, D., Jacquemont, S., Marlin, S., Simon, M., Tolmie, J., Verloes, A., Casanova, J.-L., Munnich, A., & Cormier-Daire, V. (2012b). Mutations at a single codon in Mad homology 2 domain of SMAD4 cause Myhre syndrome. *Nature Genetics*, *44*(1), 85–88. https://doi.org/10.1038/ng.1016
215. Lee, C.-C., Peng, S.-H., Shen, L., Lee, C.-F., Du, T.-H., Kang, M.-L., Xu, G.-L., Upadhyay, A. K., Cheng, X., Yan, Y.-T., Zhang, Y., & Juan, L.-J. (2017). The Role of N-α-acetyltransferase 10 Protein in DNA Methylation and Genomic Imprinting. *Molecular Cell*, *68*(1), 89-103.e7. https://doi.org/10.1016/j.molcel.2017.08.025
216. Lee, J., Kanatsu-Shinohara, M., Inoue, K., Ogonuki, N., Miki, H., Toyokuni, S., Kimura, T., Nakano, T., Ogura, A., & Shinohara, T. (2007). Akt mediates self-renewal division of mouse spermatogonial stem cells. *Development*, *134*(10), 1853–1859. https://doi.org/10.1242/dev.003004
217. Lee, J., Kanatsu-Shinohara, M., Morimoto, H., Kazuki, Y., Takashima, S., Oshimura, M., Toyokuni, S., & Shinohara, T. (2009). Genetic Reconstruction of Mouse Spermatogonial Stem Cell Self-Renewal In Vitro by Ras-Cyclin D2 Activation. *Cell Stem Cell*, *5*(1), 76–86. https://doi.org/10.1016/j.stem.2009.04.020
218. Lee, M.-N., Kweon, H. Y., & Oh, G. T. (2018). N-α-acetyltransferase 10 (NAA10) in development: The role of NAA10. *Experimental & Molecular Medicine*, *50*(7), 1–11. https://doi.org/10.1038/s12276-018-0105-2
219. Lee, Y.-S., Ehninger, D., Zhou, M., Oh, J.-Y., Kang, M., Kwak, C., Ryu, H.-H., Butz, D., Araki, T., Cai, Y., Balaji, J., Sano, Y., Nam, C. I., Kim, H. K., Kaang, B.-K., Burger, C., Neel, B. G., & Silva, A. J. (2014). Mechanism and treatment for learning and memory deficits in mouse models of Noonan syndrome. *Nature Neuroscience*, *17*(12), 1736–1743. https://doi.org/10.1038/nn.3863

220. Lee-Six, H., Olafsson, S., Ellis, P., Osborne, R. J., Sanders, M. A., Moore, L., Georgakopoulos, N., Torrente, F., Noorani, A., Goddard, M., Robinson, P., Coorens, T. H. H., O'Neill, L., Alder, C., Wang, J., Fitzgerald, R. C., Zilbauer, M., Coleman, N., Saeb-Parsy, K., … Stratton, M. R. (2019). The landscape of somatic mutation in normal colorectal epithelial cells. *Nature*, *574*(7779), 532–537. https://doi.org/10.1038/s41586-019-1672-7
221. Lelieveld, S. H., Wiel, L., Venselaar, H., Pfundt, R., Vriend, G., Veltman, J. A., Brunner, H. G., Vissers, L. E. L. M., & Gilissen, C. (2017). Spatial Clustering of de Novo Missense Mutations Identifies Candidate Neurodevelopmental Disorder-Associated Genes. *The American Journal of Human Genetics*, *101*(3), 478–484. https://doi.org/10.1016/j.ajhg.2017.08.004
222. Levine, A. D., & Chung, W. K. (2023a). Clinical features of PPP2 syndrome type R5D (Jordan's syndrome) to support standardization of care. *Molecular Case Studies*, *9*(3), a006285. https://doi.org/10.1101/mcs.a006285
223. Levine, A. D., & Chung, W. K. (2023b). Clinical features of PPP2 syndrome type R5D (Jordan's syndrome) to support standardization of care. *Molecular Case Studies*, *9*(3), a006285. https://doi.org/10.1101/mcs.a006285
224. Li, C., Li, L., Yang, M., Zeng, L., & Sun, L. (2020). PACS-2: A key regulator of mitochondria-associated membranes (MAMs). *Pharmacological Research*, *160*, 105080. https://doi.org/10.1016/j.phrs.2020.105080
225. Li, H., Cheng, B., Hu, X., Li, C., Su, J., Zhang, S., Li, L., Li, M., Yang, K., He, S., Chen, S., Wang, H., Liu, G., & Shen, Y. (2020). The first two Chinese Myhre syndrome patients with the recurrent SMAD4 pathogenic variants: Functional consequences and clinical diversity. *Clinica Chimica Acta*, *500*, 128–134. https://doi.org/10.1016/j.cca.2019.10.006
226. Li, J., Zhu, T., Yang, S., Yang, F., Wu, J., & Xiong, F. (2021a). Myhre Syndrome Misdiagnosed as Marfan Syndrome: An Educational Presentation. *Brazilian Journal of Cardiovascular Surgery*, *36*(5). https://doi.org/10.21470/1678-9741-2020-0592
227. Li, J., Zhu, T., Yang, S., Yang, F., Wu, J., & Xiong, F. (2021b). Myhre Syndrome Misdiagnosed as Marfan Syndrome: An Educational Presentation. *Brazilian Journal of Cardiovascular Surgery*, *36*(5). https://doi.org/10.21470/1678-9741-2020-0592
228. Li, S., Weidenfeld, J., & Morrisey, E. E. (2004). Transcriptional and DNA Binding Activity of the Foxp1/2/4 Family Is Modulated by Heterotypic and Homotypic Protein Interactions. *Molecular and Cellular Biology*, *24*(2), 809–822. https://doi.org/10.1128/MCB.24.2.809-822.2004
229. Li, Y., Roberts, N. D., Wala, J. A., Shapira, O., Schumacher, S. E., Kumar, K., Khurana, E., Waszak, S., Korbel, J. O., Haber, J. E., Imielinski, M., PCAWG Structural Variation Working Group, Akdemir, K. C., Alvarez, E. G., Baez-Ortega, A., Beroukhim, R., Boutros, P. C., Bowtell, D. D. L., Brors, B., … Von Mering, C. (2020). Patterns of somatic structural variation in human cancer genomes. *Nature*, *578*(7793), 112–121. https://doi.org/10.1038/s41586-019-1913-9
230. Liau, N. P. D., Johnson, M. C., Izadi, S., Gerosa, L., Hammel, M., Bruning, J. M., Wendorff, T. J., Phung, W., Hymowitz, S. G., & Sudhamsu, J. (2022). Structural basis for SHOC2 modulation of RAS signalling. *Nature*, *609*(7926), 400–407. https://doi.org/10.1038/s41586-022-04838-3

231. Lim, D., Kim, J. H., & Lee, J. (2021). Myhre syndrome: The first case in Korea. *Annals of Pediatric Endocrinology & Metabolism*, *26*(3), 210–214. https://doi.org/10.6065/apem.2040214.107
232. Lim, J., Maher, G. J., Turner, G. D. H., Dudka-Ruszkowska, W., Taylor, S., Meyts, E. R.-D., Goriely, A., & Wilkie, A. O. M. (2012). Selfish Spermatogonial Selection: Evidence from an Immunohistochemical Screen in Testes of Elderly Men. *PLoS ONE*, *7*(8), e42382. https://doi.org/10.1371/journal.pone.0042382
233. Lin, A. E., Alali, A., Starr, L. J., Shah, N., Beavis, A., Pereira, E. M., Lindsay, M. E., & Klugman, S. (2020). Gain-of-function pathogenic variants in *SMAD4* are associated with neoplasia in Myhre syndrome. *American Journal of Medical Genetics Part A*, *182*(2), 328–337. https://doi.org/10.1002/ajmg.a.61430
234. Lin, A. E., Michot, C., Cormier-Daire, V., L'Ecuyer, T. J., Matherne, G. P., Barnes, B. H., Humberson, J. B., Edmondson, A. C., Zackai, E., O'Connor, M. J., Kaplan, J. D., Ebeid, M. R., Krier, J., Krieg, E., Ghoshhajra, B., & Lindsay, M. E. (2016). Gain-of-function mutations in *SMAD4* cause a distinctive repertoire of cardiovascular phenotypes in patients with Myhre syndrome. *American Journal of Medical Genetics Part A*, *170*(10), 2617–2631. https://doi.org/10.1002/ajmg.a.37739
235. Lin, A. E., Michot, C., Cormier-Daire, V., L'Ecuyer, T. J., Matherne, G. P., Barnes, B. H., Humberson, J. B., Edmondson, A. C., Zackai, E., O'Connor, M. J., Kaplan, J. D., Ebeid, M. R., Krier, J., Krieg, E., Ghoshhajra, B., & Lindsay, M. E. (2016). Gain-of-function mutations in *SMAD4* cause a distinctive repertoire of cardiovascular phenotypes in patients with Myhre syndrome. *American Journal of Medical Genetics Part A*, *170*(10), 2617–2631. https://doi.org/10.1002/ajmg.a.37739
236. Lin, A. E., Scimone, E. R., Thom, R. P., Balaguru, D., Kinane, T. B., Moschovis, P. P., Cohen, M. S., Tan, W., Hague, C. D., Dannheim, K., Levitsky, L. L., Lilly, E., DiGiacomo, D. V., Masse, K. M., Kadzielski, S. M., Zar-Kessler, C. A., Ginns, L. C., Neumeyer, A. M., Colvin, M. K., … MGH Myhre Syndrome Study Group. (2024a). Emergence of the natural history of Myhre syndrome: 47 patients evaluated in the Massachusetts General Hospital Myhre Syndrome Clinic (2016–2023). *American Journal of Medical Genetics Part A*, *194*(10), e63638. https://doi.org/10.1002/ajmg.a.63638
237. Lin, A. E., Scimone, E. R., Thom, R. P., Balaguru, D., Kinane, T. B., Moschovis, P. P., Cohen, M. S., Tan, W., Hague, C. D., Dannheim, K., Levitsky, L. L., Lilly, E., DiGiacomo, D. V., Masse, K. M., Kadzielski, S. M., Zar-Kessler, C. A., Ginns, L. C., Neumeyer, A. M., Colvin, M. K., … MGH Myhre Syndrome Study Group. (2024b). Emergence of the natural history of Myhre syndrome: 47 patients evaluated in the Massachusetts General Hospital Myhre Syndrome Clinic (2016–2023). *American Journal of Medical Genetics Part A*, *194*(10), e63638. https://doi.org/10.1002/ajmg.a.63638
238. Lindor, N. M., Gunawardena, S. R., & Thibodeau, S. N. (2012). Mutations of SMAD4 account for both LAPS and Myhre syndromes. *American Journal of Medical Genetics Part A*, *158A*(6), 1520–1521. https://doi.org/10.1002/ajmg.a.35374
239. Liu, J., Zhuang, T., Pi, J., Chen, X., Zhang, Q., Li, Y., Wang, H., Shen, Y., Tomlinson, B., Chan, P., Yu, Z., Cheng, Y., Zheng, X., Reilly, M., Morrisey, E., Zhang, L., Liu, Z., & Zhang, Y. (2019). Endothelial Forkhead Box Transcription Factor P1 Regulates Pathological Cardiac Remodeling Through Transforming Growth Factor-β1–Endothelin-1 Signal Pathway. *Circulation*, *140*(8), 665–680. https://doi.org/10.1161/CIRCULATIONAHA.119.039767

240. Lopez-Cardona, M. G., Garcia-Cruz, D., Garcia-Ortiz, J. E., Davalos, N. O., Feria-Velasco, A., Rodriguez-Rojas, L. X., Garcia-Cruz, M. O., Figuera-Villanueva, L. E., Stephens, A., Larios-Arceo, F., & Sanchez-Corona, J. (2004). Second female case of Myhre syndrome: *Clinical Dysmorphology*, *13*(2), 91–94. https://doi.org/10.1097/00019605-200404000-00007
241. Louis, J. V., Martens, E., Borghgraef, P., Lambrecht, C., Sents, W., Longin, S., Zwaenepoel, K., Pijnenborg, R., Landrieu, I., Lippens, G., Ledermann, B., Götz, J., Van Leuven, F., Goris, J., & Janssens, V. (2011). Mice lacking phosphatase PP2A subunit PR61/B'δ (*Ppp2r5d*) develop spatially restricted tauopathy by deregulation of CDK5 and GSK3β. *Proceedings of the National Academy of Sciences*, *108*(17), 6957–6962. https://doi.org/10.1073/pnas.1018777108
242. Lynch, M. (2010). Rate, molecular spectrum, and consequences of human mutation. *Proceedings of the National Academy of Sciences*, *107*(3), 961–968. https://doi.org/10.1073/pnas.0912629107
243. Ma, L., Bayram, Y., McLaughlin, H. M., Cho, M. T., Krokosky, A., Turner, C. E., Lindstrom, K., Bupp, C. P., Mayberry, K., Mu, W., Bodurtha, J., Weinstein, V., Zadeh, N., Alcaraz, W., Powis, Z., Shao, Y., Scott, D. A., Lewis, A. M., White, J. J., … Chung, W. K. (2016). De novo missense variants in PPP1CB are associated with intellectual disability and congenital heart disease. *Human Genetics*, *135*(12), 1399–1409. https://doi.org/10.1007/s00439-016-1731-1
244. Ma, X., Shao, Y., Tian, L., Flasch, D. A., Mulder, H. L., Edmonson, M. N., Liu, Y., Chen, X., Newman, S., Nakitandwe, J., Li, Y., Li, B., Shen, S., Wang, Z., Shurtleff, S., Robison, L. L., Levy, S., Easton, J., & Zhang, J. (2019). Analysis of error profiles in deep next-generation sequencing data. *Genome Biology*, *20*(1), 50. https://doi.org/10.1186/s13059-019-1659-6
245. Maher, G. J., Goriely, A., & Wilkie, A. O. M. (2014). Cellular evidence for selfish spermatogonial selection in aged human testes. *Andrology*, *2*(3), 304–314. https://doi.org/10.1111/j.2047-2927.2013.00175.x
246. Maher, G. J., McGowan, S. J., Giannoulatou, E., Verrill, C., Goriely, A., & Wilkie, A. O. M. (2016). Visualizing the origins of selfish de novo mutations in individual seminiferous tubules of human testes. *Proceedings of the National Academy of Sciences*, *113*(9), 2454–2459. https://doi.org/10.1073/pnas.1521325113
247. Maher, G. J., Ralph, H. K., Ding, Z., Koelling, N., Mlcochova, H., Giannoulatou, E., Dhami, P., Paul, D. S., Stricker, S. H., Beck, S., McVean, G., Wilkie, A. O. M., & Goriely, A. (2018). Selfish mutations dysregulating RAS-MAPK signaling are pervasive in aged human testes. *Genome Research*, *28*(12), 1779–1790. https://doi.org/10.1101/gr.239186.118
248. Mäkelä, J.-A., & Hobbs, R. M. (2019). Molecular regulation of spermatogonial stem cell renewal and differentiation. *Reproduction*, *158*(5), R169–R187. https://doi.org/10.1530/REP-18-0476
249. Makova, K. D., & Li, W.-H. (2002). Strong male-driven evolution of DNA sequences in humans and apes. *Nature*, *416*(6881), 624–626. https://doi.org/10.1038/416624a
250. Malpe, M. S., McSwain, L. F., Kudyba, K., Ng, C. L., Nicholson, J., Brady, M., Qian, Y., Choksi, V., Hudson, A. G., Parrott, B. B., & Schulz, C. (2020). G-protein signaling is required for increasing germline stem cell division frequency in response to mating in

Drosophila males. *Scientific Reports*, *10*(1), 3888. https://doi.org/10.1038/s41598-020-60807-8

251. Mandel, S., Spivak-Pohis, I., & Gozes, I. (2008). ADNP Differential Nucleus/Cytoplasm Localization in Neurons Suggests Multiple Roles in Neuronal Differentiation and Maintenance. *Journal of Molecular Neuroscience*, *35*(2), 127–141. https://doi.org/10.1007/s12031-007-9013-y
252. Martin, H. C., Jones, W. D., McIntyre, R., Sanchez-Andrade, G., Sanderson, M., Stephenson, J. D., Jones, C. P., Handsaker, J., Gallone, G., Bruntraeger, M., McRae, J. F., Prigmore, E., Short, P., Niemi, M., Kaplanis, J., Radford, E. J., Akawi, N., Balasubramanian, M., Dean, J., … on behalf of the Deciphering Developmental Disorders Study. (2018). Quantifying the contribution of recessive coding variation to developmental disorders. *Science*, *362*(6419), 1161–1164. https://doi.org/10.1126/science.aar6731
253. Martin, L. A., Assif, N., Gilbert, M., Wijewarnasuriya, D., & Seandel, M. (2014). Enhanced Fitness of Adult Spermatogonial Stem Cells Bearing a Paternal Age-Associated FGFR2 Mutation. *Stem Cell Reports*, *3*(2), 219–226. https://doi.org/10.1016/j.stemcr.2014.06.007
254. Martincorena, I., & Campbell, P. J. (2015). Somatic mutation in cancer and normal cells. *Science*, *349*(6255), 1483–1489. https://doi.org/10.1126/science.aab4082
255. Martincorena, I., Fowler, J. C., Wabik, A., Lawson, A. R. J., Abascal, F., Hall, M. W. J., Cagan, A., Murai, K., Mahbubani, K., Stratton, M. R., Fitzgerald, R. C., Handford, P. A., Campbell, P. J., Saeb-Parsy, K., & Jones, P. H. (2018). Somatic mutant clones colonize the human esophagus with age. *Science*, *362*(6417), 911–917. https://doi.org/10.1126/science.aau3879
256. Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D. C., Fullam, A., Alexandrov, L. B., Tubio, J. M., Stebbings, L., Menzies, A., Widaa, S., Stratton, M. R., Jones, P. H., & Campbell, P. J. (2015). High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*, *348*(6237), 880–886. https://doi.org/10.1126/science.aaa6806
257. Martínez-Ruiz, C., Black, J. R. M., Puttick, C., Hill, M. S., Demeulemeester, J., Larose Cadieux, E., Thol, K., Jones, T. P., Veeriah, S., Naceur-Lombardelli, C., Toncheva, A., Prymas, P., Rowan, A., Ward, S., Cubitt, L., Athanasopoulou, F., Pich, O., Karasaki, T., Moore, D. A., … McGranahan, N. (2023). Genomic–transcriptomic evolution in lung cancer and metastasis. *Nature*, *616*(7957), 543–552. https://doi.org/10.1038/s41586-023-05706-4
258. Massagué, J., & Wotton, D. (n.d.). *Transcriptional control by the TGF-b/Smad signaling system*.
259. Matijevic, T., Knezevic, J., Slavica, M., & Pavelic, J. (2009). Rett Syndrome: From the Gene to the Disease. *European Neurology*, *61*(1), 3–10. https://doi.org/10.1159/000165342
260. Mayerhofer, A. (2013). Human testicular peritubular cells: More than meets the eye. *REPRODUCTION*, *145*(5), R107–R116. https://doi.org/10.1530/REP-12-0497
261. Mazzanti, L., Cacciari, E., Cicognani, A., Bergamaschi, R., Scarano, E., & Forabosco, A. (2003). Noonan-like syndrome with loose anagen hair: A new syndrome? *American Journal of Medical Genetics Part A*, *118A*(3), 279–286. https://doi.org/10.1002/ajmg.a.10923

262. McCright, B., Rivers, A. M., Audlin, S., & Virshup, D. M. (1996). The B56 Family of Protein Phosphatase 2A (PP2A) Regulatory Subunits Encodes Differentiation-induced Phosphoproteins That Target PP2A to Both Nucleus and Cytoplasm. *Journal of Biological Chemistry*, *271*(36), 22081–22089. https://doi.org/10.1074/jbc.271.36.22081
263. McGowan, R., Gulati, R., McHenry, P., Cooke, A., Butler, S., Keng, W. T., Murday, V., Whiteford, M., Dikkers, F. G., Sikkema-Raddatz, B., Van Essen, T., & Tolmie, J. (2011). Clinical features and respiratory complications in Myhre syndrome. *European Journal of Medical Genetics*, S1769721211000838. https://doi.org/10.1016/j.ejmg.2011.07.001
264. McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, *17*(1), 122. https://doi.org/10.1186/s13059-016-0974-4
265. Medrano, J. V., Rombaut, C., Simon, C., Pellicer, A., & Goossens, E. (2016). Human spermatogonial stem cells display limited proliferation in vitro under mouse spermatogonial stem cell culture conditions. *Fertility and Sterility*, *106*(6), 1539-1549.e8. https://doi.org/10.1016/j.fertnstert.2016.07.1065
266. Meerschaut, I., Beyens, A., Steyaert, W., De Rycke, R., Bonte, K., De Backer, T., Janssens, S., Panzer, J., Plasschaert, F., De Wolf, D., & Callewaert, B. (2019). Myhre syndrome: A first familial recurrence and broadening of the phenotypic spectrum. *American Journal of Medical Genetics Part A*, *179*(12), 2494–2499. https://doi.org/10.1002/ajmg.a.61377
267. Meerschaut, I., Rochefort, D., Revençu, N., Pètre, J., Corsello, C., Rouleau, G. A., Hamdan, F. F., Michaud, J. L., Morton, J., Radley, J., Ragge, N., García-Miñaúr, S., Lapunzina, P., Bralo, M. P., Mori, M. Á., Moortgat, S., Benoit, V., Mary, S., Bockaert, N., … Callewaert, B. (2017). *FOXP1* -related intellectual disability syndrome: A recognisable entity. *Journal of Medical Genetics*, *54*(9), 613–623. https://doi.org/10.1136/jmedgenet-2017-104579
268. Mellios, N., Feldman, D. A., Sheridan, S. D., Ip, J. P. K., Kwok, S., Amoah, S. K., Rosen, B., Rodriguez, B. A., Crawford, B., Swaminathan, R., Chou, S., Li, Y., Ziats, M., Ernst, C., Jaenisch, R., Haggarty, S. J., & Sur, M. (2018). MeCP2-regulated miRNAs control early human neurogenesis through differential effects on ERK and AKT signaling. *Molecular Psychiatry*, *23*(4), 1051–1065. https://doi.org/10.1038/mp.2017.86
269. Meng, X., Lindahl, M., Hyvönen, M. E., Parvinen, M., De Rooij, D. G., Hess, M. W., Raatikainen-Ahokas, A., Sainio, K., Rauvala, H., Lakso, M., Pichel, J. G., Westphal, H., Saarma, M., & Sariola, H. (2000). Regulation of Cell Fate Decision of Undifferentiated Spermatogonia by GDNF. *Science*, *287*(5457), 1489–1493. https://doi.org/10.1126/science.287.5457.1489
270. Messina, G., Atterrato, M. T., & Dimitri, P. (2016). When chromatin organisation floats astray: The *Srcap* gene and Floating–Harbor syndrome. *Journal of Medical Genetics*, *53*(12), 793–797. https://doi.org/10.1136/jmedgenet-2016-103842
271. Meyers, G. A., Orlow, S. J., Munro, I. R., Przylepa, K. A., & Jabs, E. W. (1995). Fibroblast growth factor receptor 3 (FGFR3) transmembrane mutation in Crouzon syndrome with acanthosis nigricans. *Nature Genetics*, *11*(4), 462–464. https://doi.org/10.1038/ng1295-462
272. Michaelson, J. J., Shi, Y., Gujral, M., Zheng, H., Malhotra, D., Jin, X., Jian, M., Liu, G., Greer, D., Bhandari, A., Wu, W., Corominas, R., Peoples, Á., Koren, A., Gore, A., Kang, S., Lin, G. N., Estabillo, J., Gadomski, T., … Sebat, J. (2012). Whole-Genome

Sequencing in Autism Identifies Hot Spots for De Novo Germline Mutation. *Cell*, *151*(7), 1431–1442. https://doi.org/10.1016/j.cell.2012.11.019

273. Michot, C., Le Goff, C., Mahaut, C., Afenjar, A., Brooks, A. S., Campeau, P. M., Destree, A., Di Rocco, M., Donnai, D., Hennekam, R., Heron, D., Jacquemont, S., Kannu, P., Lin, A. E., Manouvrier-Hanu, S., Mansour, S., Marlin, S., McGowan, R., Murphy, H., … Cormier-Daire, V. (2014). Myhre and LAPS syndromes: Clinical and molecular review of 32 patients. *European Journal of Human Genetics*, *22*(11), 1272–1277. https://doi.org/10.1038/ejhg.2013.288
274. Milholland, B., Dong, X., Zhang, L., Hao, X., Suh, Y., & Vijg, J. (2017). Differences between germline and somatic mutation rates in humans and mice. *Nature Communications*, *8*(1), 15183. https://doi.org/10.1038/ncomms15183
275. Miyake, N., Ozasa, S., Mabe, H., Kimura, S., Shiina, M., Imagawa, E., Miyatake, S., Nakashima, M., Mizuguchi, T., Takata, A., Ogata, K., & Matsumoto, N. (2018). A novel missense mutation affecting the same amino acid as the recurrent *PACS1* mutation in Schuurs-Hoeijmakers syndrome. *Clinical Genetics*, *93*(4), 929–930. https://doi.org/10.1111/cge.13105
276. Miyazawa, K., Shinozaki, M., Hara, T., Furuya, T., & Miyazono, K. (2002). Two major Smad pathways in TGF-β superfamily signalling. *Genes to Cells*, *7*(12), 1191–1204. https://doi.org/10.1046/j.1365-2443.2002.00599.x
277. Moloney, D. M., Slaney, S. R., Oldridge, M., Wall, S. A., Sahlin, P., Stenman, G., & Wilkie, A. O. M. (1996). Exclusive paternal origin of new mutations in Apert syndrome. *Nature Genetics*, *13*(1), 48–53. https://doi.org/10.1038/ng0596-48
278. Moore, L., Cagan, A., Coorens, T. H. H., Neville, M. D. C., Sanghvi, R., Sanders, M. A., Oliver, T. R. W., Leongamornlert, D., Ellis, P., Noorani, A., Mitchell, T. J., Butler, T. M., Hooks, Y., Warren, A. Y., Jorgensen, M., Dawson, K. J., Menzies, A., O'Neill, L., Latimer, C., … Rahbari, R. (2021). The mutational landscape of human somatic and germline cells. *Nature*, *597*(7876), 381–386. https://doi.org/10.1038/s41586-021-03822-7
279. Moore, L., Leongamornlert, D., Coorens, T. H. H., Sanders, M. A., Ellis, P., Dentro, S. C., Dawson, K. J., Butler, T., Rahbari, R., Mitchell, T. J., Maura, F., Nangalia, J., Tarpey, P. S., Brunner, S. F., Lee-Six, H., Hooks, Y., Moody, S., Mahbubani, K. T., Jimenez-Linan, M., … Stratton, M. R. (2020). The mutational landscape of normal human endometrial epithelium. *Nature*, *580*(7805), 640–646. https://doi.org/10.1038/s41586-020-2214-z
280. Morén, A., Hellman, U., Inada, Y., Imamura, T., Heldin, C.-H., & Moustakas, A. (2003). Differential Ubiquitination Defines the Functional Status of the Tumor Suppressor Smad4. *Journal of Biological Chemistry*, *278*(35), 33571–33582. https://doi.org/10.1074/jbc.M300159200
281. Motta, M., Chillemi, G., Fodale, V., Cecchetti, S., Coppola, S., Stipo, S., Cordeddu, V., Macioce, P., Gelb, B. D., & Tartaglia, M. (2016). SHOC2 subcellular shuttling requires the KEKE motif-rich region and *N* -terminal leucine-rich repeat domain and impacts on ERK signalling. *Human Molecular Genetics*, *25*(17), 3824–3835. https://doi.org/10.1093/hmg/ddw229
282. Myhre, S. A., Ruvalcaba, R. H. A., & Graham, C. B. (1981). A new growth deficiency syndrome. *Clinical Genetics*, *20*(1), 1–5. https://doi.org/10.1111/j.1399-0004.1981.tb01798.x

283. Nachman, M. W., & Crowell, S. L. (2000). Estimate of the Mutation Rate per Nucleotide in Humans. *Genetics*, *156*(1), 297–304. https://doi.org/10.1093/genetics/156.1.297
284. Nakagawa, T., Nabeshima, Y., & Yoshida, S. (2007). Functional Identification of the Actual and Potential Stem Cell Compartments in Mouse Spermatogenesis. *Developmental Cell*, *12*(2), 195–206. https://doi.org/10.1016/j.devcel.2007.01.002
285. Nakata, H., Wakayama, T., Sonomura, T., Honma, S., Hatta, T., & Iseki, S. (2015). Three-dimensional structure of seminiferous tubules in the adult mouse. *Journal of Anatomy*, *227*(5), 686–694. https://doi.org/10.1111/joa.12375
286. Naski, M. C., Wang, Q., Xu, J., & Ornitz, D. M. (1996). Graded activation of fibroblast growth factor receptor 3 by mutations causing achondroplasia and thanatophoric dysplasia. *Nature Genetics*, *13*(2), 233–237. https://doi.org/10.1038/ng0696-233
287. Naudin, C., Hattabi, A., Michelet, F., Miri-Nezhad, A., Benyoucef, A., Pflumio, F., Guillonneau, F., Fichelson, S., Vigon, I., Dusanter-Fourt, I., & Lauret, E. (2017). PUMILIO/FOXP1 signaling drives expansion of hematopoietic stem/progenitor and leukemia cells. *Blood*, *129*(18), 2493–2506. https://doi.org/10.1182/blood-2016-10-747436
288. Need, A. C., Shashi, V., Hitomi, Y., Schoch, K., Shianna, K. V., McDonald, M. T., Meisler, M. H., & Goldstein, D. B. (2012). Clinical application of exome sequencing in undiagnosed genetic conditions. *Journal of Medical Genetics*, *49*(6), 353–361. https://doi.org/10.1136/jmedgenet-2012-100819
289. Neto, F. T. L., Bach, P. V., Najari, B. B., Li, P. S., & Goldstein, M. (2016). Spermatogenesis in humans and its affecting factors. *Seminars in Cell & Developmental Biology*, *59*, 10–26. https://doi.org/10.1016/j.semcdb.2016.04.009
290. Neville, M. D., Lawson, A. R., Sanghvi, R., Abascal, F., Pham, M. H., Cagan, A., Nicola, P. A., Bayzetinova, T., Baez-Ortega, A., Roberts, K., Lensing, S. V., Widaa, S., Alcantara, R. E., García, M. P., Wadge, S., Stratton, M. R., Campbell, P. J., Small, K., Martincorena, I., … Rahbari, R. (2024). *Sperm sequencing reveals extensive positive selection in the male germline*. https://doi.org/10.1101/2024.10.30.24316414
291. Ng, B. G., Eklund, E. A., Shiryaev, S. A., Dong, Y. Y., Abbott, M., Asteggiano, C., Bamshad, M. J., Barr, E., Bernstein, J. A., Chelakkadan, S., Christodoulou, J., Chung, W. K., Ciliberto, M. A., Cousin, J., Gardiner, F., Ghosh, S., Graf, W. D., Grunewald, S., Hammond, K., … Freeze, H. H. (2020). Predominant and novel de novo variants in 29 individuals with *ALG13* deficiency: Clinical description, biomarker status, biochemical analysis, and treatment suggestions. *Journal of Inherited Metabolic Disease*, *43*(6), 1333–1348. https://doi.org/10.1002/jimd.12290
292. Ng, S. W. K., Rouhani, F. J., Brunner, S. F., Brzozowska, N., Aitken, S. J., Yang, M., Abascal, F., Moore, L., Nikitopoulou, E., Chappell, L., Leongamornlert, D., Ivovic, A., Robinson, P., Butler, T., Sanders, M. A., Williams, N., Coorens, T. H. H., Teague, J., Raine, K., … Campbell, P. J. (2021). Convergent somatic mutations in metabolism genes in chronic liver disease. *Nature*, *598*(7881), 473–478. https://doi.org/10.1038/s41586-021-03974-6
293. Nikkel, S. M., Dauber, A., De Munnik, S., Connolly, M., Hood, R. L., Caluseriu, O., Hurst, J., Kini, U., Nowaczyk, M. J. M., Afenjar, A., Albrecht, B., Allanson, J. E., Balestri, P., Ben-Omran, T., Brancati, F., Cordeiro, I., Da Cunha, B. S., Delaney, L. A., Destrée, A., … Boycott, K. M. (2013). The phenotype of Floating-Harbor syndrome:

Clinical characterization of 52 individuals with mutations in exon 34 of SRCAP. *Orphanet Journal of Rare Diseases*, *8*(1), 63. https://doi.org/10.1186/1750-1172-8-63
294. Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L. A., Menzies, A., Martin, S., Leung, K., Chen, L., Leroy, C., Ramakrishna, M., Rance, R., Lau, K. W., Mudie, L. J., … Stratton, M. R. (2012). Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell*, *149*(5), 979–993. https://doi.org/10.1016/j.cell.2012.04.024
295. Nomura, R., Miyai, K., Nishimura, G., Kashimada, K., & Morio, T. (2017). Myhre syndrome: Age-dependent progressive phenotype. *Pediatrics International*, *59*(11), 1205–1206. https://doi.org/10.1111/ped.13413
296. Nowell, P. C. (1976). The Clonal Evolution of Tumor Cell Populations: Acquired genetic lability permits stepwise selection of variant sublines and underlies tumor progression. *Science*, *194*(4260), 23–28. https://doi.org/10.1126/science.959840
297. Nürnberg, B., & Beer-Hammer, S. (2019). Function, Regulation and Biological Roles of PI3Kγ Variants. *Biomolecules*, *9*(9), 427. https://doi.org/10.3390/biom9090427
298. Oatley, J. M., Avarbock, M. R., Telaranta, A. I., Fearon, D. T., & Brinster, R. L. (2006). Identifying genes important for spermatogonial stem cell self-renewal and survival. *Proceedings of the National Academy of Sciences*, *103*(25), 9524–9529. https://doi.org/10.1073/pnas.0603332103
299. Okur, V., Cho, M. T., Henderson, L., Retterer, K., Schneider, M., Sattler, S., Niyazov, D., Azage, M., Smith, S., Picker, J., Lincoln, S., Tarnopolsky, M., Brady, L., Bjornsson, H. T., Applegate, C., Dameron, A., Willaert, R., Baskin, B., Juusola, J., & Chung, W. K. (2016). De novo mutations in CSNK2A1 are associated with neurodevelopmental abnormalities and dysmorphic features. *Human Genetics*, *135*(7), 699–705. https://doi.org/10.1007/s00439-016-1661-y
300. Olafsson, S., McIntyre, R. E., Coorens, T., Butler, T., Jung, H., Robinson, P. S., Lee-Six, H., Sanders, M. A., Arestang, K., Dawson, C., Tripathi, M., Strongili, K., Hooks, Y., Stratton, M. R., Parkes, M., Martincorena, I., Raine, T., Campbell, P. J., & Anderson, C. A. (2020). Somatic Evolution in Non-neoplastic IBD-Affected Colon. *Cell*, *182*(3), 672-684.e11. https://doi.org/10.1016/j.cell.2020.06.036
301. Oldenburg, M. S., Frisch, C. D., Lindor, N. M., Edell, E. S., Kasperbauer, J. L., & O'Brien, E. K. (2015a). Myhre-LAPs syndrome and intubation related airway stenosis: Keys to diagnosis and critical therapeutic interventions. *American Journal of Otolaryngology*, *36*(5), 636–641. https://doi.org/10.1016/j.amjoto.2015.04.005
302. Oldenburg, M. S., Frisch, C. D., Lindor, N. M., Edell, E. S., Kasperbauer, J. L., & O'Brien, E. K. (2015b). Myhre-LAPs syndrome and intubation related airway stenosis: Keys to diagnosis and critical therapeutic interventions. *American Journal of Otolaryngology*, *36*(5), 636–641. https://doi.org/10.1016/j.amjoto.2015.04.005
303. Oldridge, M., Zackai, E. H., McDonald-McGinn, D. M., Iseki, S., Morriss-Kay, G. M., Twigg, S. R. F., Johnson, D., Wall, S. A., Jiang, W., Theda, C., Jabs, E. W., & Wilkie, A. O. M. (1999). De Novo Alu-Element Insertions in FGFR2 Identify a Distinct Pathological Basis for Apert Syndrome. *The American Journal of Human Genetics*, *64*(2), 446–461. https://doi.org/10.1086/302245
304. Olson, H. E., Jean-Marçais, N., Yang, E., Heron, D., Tatton-Brown, K., Van Der Zwaag, P. A., Bijlsma, E. K., Krock, B. L., Backer, E., Kamsteeg, E.-J., Sinnema, M., Reijnders, M. R. F., Bearden, D., Begtrup, A., Telegrafi, A., Lunsing, R. J., Burglen, L., Lesca, G.,

Cho, M. T., … Thauvin-Robinet, C. (2018). A Recurrent De Novo PACS2 Heterozygous Missense Variant Causes Neonatal-Onset Developmental Epileptic Encephalopathy, Facial Dysmorphism, and Cerebellar Dysgenesis. *The American Journal of Human Genetics*, *102*(5), 995–1007. https://doi.org/10.1016/j.ajhg.2018.03.005

305. O'Roak, B. J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J. J., Girirajan, S., Karakoc, E., MacKenzie, A. P., Ng, S. B., Baker, C., Rieder, M. J., Nickerson, D. A., Bernier, R., Fisher, S. E., Shendure, J., & Eichler, E. E. (2011). Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nature Genetics*, *43*(6), 585–589. https://doi.org/10.1038/ng.835
306. Owen, C. I., Bowden, R., Parker, M. J., Patterson, J., Patterson, J., Price, S., Sarkar, A., Castle, B., Deshpande, C., Splitt, M., Ghali, N., Dean, J., Green, A. J., Crosby, C., Deciphering Developmental Disorders Study, & Tatton-Brown, K. (2018a). Extending the phenotype associated with the *CSNK2A1*- related Okur–Chung syndrome—A clinical study of 11 individuals. *American Journal of Medical Genetics Part A*, *176*(5), 1108–1114. https://doi.org/10.1002/ajmg.a.38610
307. Owen, C. I., Bowden, R., Parker, M. J., Patterson, J., Patterson, J., Price, S., Sarkar, A., Castle, B., Deshpande, C., Splitt, M., Ghali, N., Dean, J., Green, A. J., Crosby, C., Deciphering Developmental Disorders Study, & Tatton-Brown, K. (2018b). Extending the phenotype associated with the *CSNK2A1*- related Okur–Chung syndrome—A clinical study of 11 individuals. *American Journal of Medical Genetics Part A*, *176*(5), 1108–1114. https://doi.org/10.1002/ajmg.a.38610
308. Pan, X., Chen, X., Tong, X., Tang, C., & Li, J. (2015). Ppp2ca knockout in mice spermatogenesis. *REPRODUCTION*, *149*(4), 385–391. https://doi.org/10.1530/REP-14-0231
309. Paoletti, P., Bellone, C., & Zhou, Q. (2013). NMDA receptor subunit diversity: Impact on receptor properties, synaptic plasticity and disease. *Nature Reviews Neuroscience*, *14*(6), 383–400. https://doi.org/10.1038/nrn3504
310. Papke, C. M., Smolen, K. A., Swingle, M. R., Cressey, L., Heng, R. A., Toporsian, M., Deng, L., Hagen, J., Shen, Y., Chung, W. K., Kettenbach, A. N., & Honkanen, R. E. (2021). A disorder-related variant (E420K) of a PP2A-regulatory subunit (PPP2R5D) causes constitutively active AKT-mTOR signaling and uncoordinated cell growth. *Journal of Biological Chemistry*, *296*, 100313. https://doi.org/10.1016/j.jbc.2021.100313
311. Park, W.-J., Theda, C., Maestri, N. E., Meyers, G. A., Fryburg, J. S., Dufresne, C., Cohen, M. M., & Jabs', E. W. (1995). Analysis of Phenotypic Features and FGFR2 Mutations in Apert Syndrome. *Am. J. Hum. Genet.*
312. Pawliczek, P., Patel, R. Y., Ashmore, L. R., Jackson, A. R., Bizon, C., Nelson, T., Powell, B., Freimuth, R. R., Strande, N., Shah, N., Paithankar, S., Wright, M. W., Dwight, S., Zhen, J., Landrum, M., McGarvey, P., Babb, L., Plon, S. E., Milosavljevic, A., & on behalf of the Clinical Genome (ClinGen) Resource. (2018a). ClinGen Allele Registry links information about genetic variants. *Human Mutation*, *39*(11), 1690–1701. https://doi.org/10.1002/humu.23637
313. Pawliczek, P., Patel, R. Y., Ashmore, L. R., Jackson, A. R., Bizon, C., Nelson, T., Powell, B., Freimuth, R. R., Strande, N., Shah, N., Paithankar, S., Wright, M. W., Dwight, S., Zhen, J., Landrum, M., McGarvey, P., Babb, L., Plon, S. E., Milosavljevic, A., & on behalf of the Clinical Genome (ClinGen) Resource. (2018b). ClinGen Allele

Registry links information about genetic variants. *Human Mutation*, *39*(11), 1690–1701. https://doi.org/10.1002/humu.23637
314. Pearson, C. A., Moore, D. M., Tucker, H. O., Dekker, J. D., Hu, H., Miquelajáuregui, A., & Novitch, B. G. (2020). Foxp1 Regulates Neural Stem Cell Self-Renewal and Bias Toward Deep Layer Cortical Fates. *Cell Reports*, *30*(6), 1964-1981.e3. https://doi.org/10.1016/j.celrep.2020.01.034
315. Pefkianaki, M., Schneider, A., Capasso, J. E., Wasserman, B. N., Bardakjian, T., & Levin, A. V. (2018). Ocular manifestations of PACS1 mutation. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, *22*(4), 323–325. https://doi.org/10.1016/j.jaapos.2017.12.008
316. Penrose, L. (1955). Parental age and mutation. *The Lancet*, *266*(6885), 312–313. https://doi.org/10.1016/S0140-6736(55)92305-9
317. Penrose, L. S. (1957). *Parental Age in Achondroplasia and Mongolism*.
318. Petrovski, S., Küry, S., Myers, C. T., Anyane-Yeboa, K., Cogné, B., Bialer, M., Xia, F., Hemati, P., Riviello, J., Mehaffey, M., Besnard, T., Becraft, E., Wadley, A., Politi, A. R., Colombo, S., Zhu, X., Ren, Z., Andrews, I., Dudding-Byth, T., … Goldstein, D. B. (2016). Germline De Novo Mutations in GNB1 Cause Severe Neurodevelopmental Disability, Hypotonia, and Seizures. *The American Journal of Human Genetics*, *98*(5), 1001–1010. https://doi.org/10.1016/j.ajhg.2016.03.011
319. Pfeifer, G. P., Denissenko, M. F., Olivier, M., Tretyakova, N., Hecht, S. S., & Hainaut, P. (2002). Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene*, *21*(48), 7435–7451. https://doi.org/10.1038/sj.onc.1205803
320. Piccinini, E., Kalkkinen, N., Saarma, M., & Runeberg-Roos, P. (2013). Glial cell line-derived neurotrophic factor: Characterization of mammalian posttranslational modifications. *Annals of Medicine*, *45*(1), 66–73. https://doi.org/10.3109/07853890.2012.663927
321. Picco, P., Naselli, A., Pala, G., Marsciani, A., Buoncompagni, A., & Martini, A. (2013). Recurrent Pericarditis in Myhre Syndrome. *American Journal of Medical Genetics Part A*, *161*(5), 1164–1166. https://doi.org/10.1002/ajmg.a.35892
322. Piccolo, P., Mithbaokar, P., Sabatino, V., Tolmie, J., Melis, D., Schiaffino, M. C., Filocamo, M., Andria, G., & Brunetti-Pierri, N. (2014a). SMAD4 mutations causing Myhre syndrome result in disorganization of extracellular matrix improved by losartan. *European Journal of Human Genetics*, *22*(8), 988–994. https://doi.org/10.1038/ejhg.2013.283
323. Piccolo, P., Mithbaokar, P., Sabatino, V., Tolmie, J., Melis, D., Schiaffino, M. C., Filocamo, M., Andria, G., & Brunetti-Pierri, N. (2014b). SMAD4 mutations causing Myhre syndrome result in disorganization of extracellular matrix improved by losartan. *European Journal of Human Genetics*, *22*(8), 988–994. https://doi.org/10.1038/ejhg.2013.283
324. Pierreux, C. E., As, F. J. N., & Hill, C. S. (n.d.). *Transforming Growth Factor b-Independent Shuttling of Smad4 between the Cytoplasm and Nucleus*.
325. Pinhasov, A., Mandel, S., Torchinsky, A., Giladi, E., Pittel, Z., Goldsweig, A. M., Servoss, S. J., Brenneman, D. E., & Gozes, I. (2003). Activity-dependent neuroprotective protein: A novel gene essential for brain formation. *Developmental Brain Research*, *144*(1), 83–90. https://doi.org/10.1016/S0165-3806(03)00162-7

326. Platzer, K., Yuan, H., Schütz, H., Winschel, A., Chen, W., Hu, C., Kusumoto, H., Heyne, H. O., Helbig, K. L., Tang, S., Willing, M. C., Tinkle, B. T., Adams, D. J., Depienne, C., Keren, B., Mignot, C., Frengen, E., Strømme, P., Biskup, S., … Lemke, J. R. (2017). *GRIN2B* encephalopathy: Novel findings on phenotype, variant clustering, functional consequences and treatment aspects. *Journal of Medical Genetics*, *54*(7), 460–470. https://doi.org/10.1136/jmedgenet-2016-104509
327. Plotnikov, A., Cuderland, D., Karamansha, Y., Llivnah, O., & Seger, R. (2019). Nuclear ERK Translocation is Mediated by Protein Kinase CK2 and Accelerated by Autophosphorylation. *Cellular Physiology and Biochemistry*, *53*(2), 366–387. https://doi.org/10.33594/000000144
328. Poon, G. Y. P., Watson, C. J., Fisher, D. S., & Blundell, J. R. (2021). Synonymous mutations reveal genome-wide levels of positive selection in healthy tissues. *Nature Genetics*, *53*(11), 1597–1605. https://doi.org/10.1038/s41588-021-00957-1
329. Qin, J., Calabrese, P., Tiemann-Boege, I., Shinde, D. N., Yoon, S.-R., Gelfand, D., Bauer, K., & Arnheim, N. (2007). The Molecular Anatomy of Spontaneous Germline Mutations in Human Testes. *PLoS Biology*, *5*(9), e224. https://doi.org/10.1371/journal.pbio.0050224
330. Rahbari, R., Wuster, A., Lindsay, S. J., Hardwick, R. J., UK10K Consortium, Alexandrov, L. B., Al Turki, S., Dominiczak, A., Morris, A., Porteous, D., Smith, B., Stratton, M. R., & Hurles, M. E. (2016). Timing, rates and spectra of human germline mutation. *Nature Genetics*, *48*(2), 126–133. https://doi.org/10.1038/ng.3469
331. Rajpert-De Meyts, E., Jacobsen, G. K., Bartkova, J., Aubry, F., Samson, M., Bartek, J., & Skakkebæk, N. E. (2003). The immunohistochemical expression pattern of Chk2, p53, p19 ^{INK4d} , MAGE-A4 and other selected antigens provides new evidence for the premeiotic origin of spermatocytic seminoma. *Histopathology*, *42*(3), 217–226. https://doi.org/10.1046/j.1365-2559.2003.01587.x
332. Ralph, H. K. (n.d.). *Evaluating selfish spermatogonial selection and its role in human disease*.
333. Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, *8*(11), 2281–2308. https://doi.org/10.1038/nprot.2013.143
334. Ree, R., Geithus, A. S., Tørring, P. M., Sørensen, K. P., Damkjær, M., Lynch, S. A., & Arnesen, T. (2019). A novel NAA10 p.(R83H) variant with impaired acetyltransferase activity identified in two boys with ID and microcephaly. *BMC Medical Genetics*, *20*(1), 101. https://doi.org/10.1186/s12881-019-0803-1
335. Risch, N., Reich, E. W., Wishnick, M. M., & McCarthy, J. G. (n.d.). *Spontaneous Mutation and Parental Age in Humans*.
336. Roach, J. C., Glusman, G., Smit, A. F. A., Huff, C. D., Hubley, R., Shannon, P. T., Rowen, L., Pant, K. P., Goodman, N., Bamshad, M., Shendure, J., Drmanac, R., Jorde, L. B., Hood, L., & Galas, D. J. (2010). Analysis of Genetic Inheritance in a Family Quartet by Whole-Genome Sequencing. *Science*, *328*(5978), 636–639. https://doi.org/10.1126/science.1186802
337. Rocca, D. L., Wilkinson, K. A., & Henley, J. M. (2017). SUMOylation of FOXP1 regulates transcriptional repression via CtBP1 to drive dendritic morphogenesis. *Scientific Reports*, *7*(1), 877. https://doi.org/10.1038/s41598-017-00707-6
338. Rodriguez-Viciana, P., Oses-Prieto, J., Burlingame, A., Fried, M., & McCormick, F. (2006). A Phosphatase Holoenzyme Comprised of Shoc2/Sur8 and the Catalytic Subunit

of PP1 Functions as an M-Ras Effector to Modulate Raf Activity. *Molecular Cell*, *22*(2), 217–230. https://doi.org/10.1016/j.molcel.2006.03.027

339. Rope, A. F., Wang, K., Evjenth, R., Xing, J., Johnston, J. J., Swensen, J. J., Johnson, W. E., Moore, B., Huff, C. D., Bird, L. M., Carey, J. C., Opitz, J. M., Stevens, C. A., Jiang, T., Schank, C., Fain, H. D., Robison, R., Dalley, B., Chin, S., … Lyon, G. J. (2011). Using VAAST to Identify an X-Linked Disorder Resulting in Lethality in Male Infants Due to N-Terminal Acetyltransferase Deficiency. *The American Journal of Human Genetics*, *89*(1), 28–43. https://doi.org/10.1016/j.ajhg.2011.05.017
340. Rosenberger, G., Meien, S., & Kutsche, K. (2009). Oncogenic *HRAS* mutations cause prolonged PI3K signaling in response to epidermal growth factor in fibroblasts of patients with Costello syndrome. *Human Mutation*, *30*(3), 352–362. https://doi.org/10.1002/humu.20855
341. Ross, P. D., Guy, J., Selfridge, J., Kamal, B., Bahey, N., Tanner, K. E., Gillingwater, T. H., Jones, R. A., Loughrey, C. M., McCarroll, C. S., Bailey, M. E. S., Bird, A., & Cobb, S. (2016). Exclusive expression of MeCP2 in the nervous system distinguishes between brain and peripheral Rett syndrome-like phenotypes. *Human Molecular Genetics*, ddw269. https://doi.org/10.1093/hmg/ddw269
342. Ross, S., & Hill, C. S. (2008). How the Smads regulate transcription. *The International Journal of Biochemistry & Cell Biology*, *40*(3), 383–408. https://doi.org/10.1016/j.biocel.2007.09.006
343. Rosser, E. M., Wilkinson, A. R., Hurst, J. A., McGaughran, J. M., & Donnai, D. (1995). Geleophysic dysplasia: A report of three affected boys—prenatal ultrasound does not detect recurrence. *American Journal of Medical Genetics*, *58*(3), 217–221. https://doi.org/10.1002/ajmg.1320580304
344. Rots, D., Chater-Diehl, E., Dingemans, A. J. M., Goodman, S. J., Siu, M. T., Cytrynbaum, C., Choufani, S., Hoang, N., Walker, S., Awamleh, Z., Charkow, J., Meyn, S., Pfundt, R., Rinne, T., Gardeitchik, T., De Vries, B. B. A., Deden, A. C., Leenders, E., Kwint, M., … Weksberg, R. (2021). Truncating SRCAP variants outside the Floating-Harbor syndrome locus cause a distinct neurodevelopmental disorder with a specific DNA methylation signature. *The American Journal of Human Genetics*, *108*(6), 1053–1068. https://doi.org/10.1016/j.ajhg.2021.04.008
345. Rousso, D. L., Gaber, Z. B., Wellik, D., Morrisey, E. E., & Novitch, B. G. (2008). Coordinated Actions of the Forkhead Protein Foxp1 and Hox Proteins in the Columnar Organization of Spinal Motor Neurons. *Neuron*, *59*(2), 226–240. https://doi.org/10.1016/j.neuron.2008.06.025
346. Rulli, I., Ferrero, G. B., Belligni, E., Delmonaco, A. G., Defilippi, C., & Silengo, M. (2005). Myhre's syndrome in a girl with normal intelligence. *American Journal of Medical Genetics Part A*, *134A*(1), 100–102. https://doi.org/10.1002/ajmg.a.30444
347. Rylaarsdam, L., Rakotomamonjy, J., Pope, E., & Guemez-Gamboa, A. (2024). iPSC-derived models of PACS1 syndrome reveal transcriptional and functional deficits in neuron activity. *Nature Communications*, *15*(1), 827. https://doi.org/10.1038/s41467-024-44989-7
348. Ryu, B.-Y., Orwig, K. E., Oatley, J. M., Avarbock, M. R., & Brinster, R. L. (2006). Effects of Aging and Niche Microenvironment on Spermatogonial Stem Cell Self-Renewal. *Stem Cells*, *24*(6), 1505–1511. https://doi.org/10.1634/stemcells.2005-0580

349. Sada, A., Suzuki, A., Suzuki, H., & Saga, Y. (2009). The RNA-Binding Protein NANOS2 Is Required to Maintain Murine Spermatogonial Stem Cells. *Science*, *325*(5946), 1394–1398. https://doi.org/10.1126/science.1172645
350. Salazar, R., Arbeithuber, B., Ivankovic, M., Heinzl, M., Moura, S., Hartl, I., Mair, T., Lahnsteiner, A., Ebner, T., Shebl, O., Pröll, J., & Tiemann-Boege, I. (2022). Discovery of an unusually high number of de novo mutations in sperm of older men using duplex sequencing. *Genome Research*, *32*(3), 499–511. https://doi.org/10.1101/gr.275695.121
351. Sandal, P., Jong, C. J., Merrill, R. A., Song, J., & Strack, S. (2021). Protein phosphatase 2A – structure, function and role in neurodevelopmental disorders. *Journal of Cell Science*, *134*(13), jcs248187. https://doi.org/10.1242/jcs.248187
352. Santillo, A., Falvo, S., Chieffi, P., Burrone, L., Chieffi Baccari, G., Longobardi, S., & Di Fiore, M. M. (2014). D-aspartate affects NMDA receptor-extracellular signal–regulated kinase pathway and upregulates androgen receptor expression in the rat testis. *Theriogenology*, *81*(5), 744–751. https://doi.org/10.1016/j.theriogenology.2013.12.009
353. Saunders, K., Bert, A. G., Dredge, B. K., Toubia, J., Gregory, P. A., Pillman, K. A., Goodall, G. J., & Bracken, C. P. (2020). Insufficiently complex unique-molecular identifiers (UMIs) distort small RNA sequencing. *Scientific Reports*, *10*(1), Article 1. https://doi.org/10.1038/s41598-020-71323-0
354. Saunier, C., Støve, S. I., Popp, B., Gérard, B., Blenski, M., AhMew, N., Bie, C., Goldenberg, P., Isidor, B., Keren, B., Leheup, B., Lampert, L., Mignot, C., Tezcan, K., Mancini, G. M. S., Nava, C., Wasserstein, M., Bruel, A., Thevenon, J., … Zweier, C. (2016). Expanding the Phenotype Associated with NAA10-Related N-Terminal Acetylation Deficiency. *Human Mutation*, *37*(8), 755–764. https://doi.org/10.1002/humu.23001
355. Schmitt, M. W., Kennedy, S. R., Salk, J. J., Fox, E. J., Hiatt, J. B., & Loeb, L. A. (2012). Detection of ultra-rare mutations by next-generation sequencing. *Proceedings of the National Academy of Sciences*, *109*(36), 14508–14513. https://doi.org/10.1073/pnas.1208715109
356. Schuurs-Hoeijmakers, J. H. M., Oh, E. C., Vissers, L. E. L. M., Swinkels, M. E. M., Gilissen, C., Willemsen, M. A., Holvoet, M., Steehouwer, M., Veltman, J. A., de Vries, B. B. A., van Bokhoven, H., de Brouwer, A. P. M., Katsanis, N., Devriendt, K., & Brunner, H. G. (2012). Recurrent De Novo Mutations in PACS1 Cause Defective Cranial-Neural-Crest Migration and Define a Recognizable Intellectual-Disability Syndrome. *The American Journal of Human Genetics*, *91*(6), 1122–1127. https://doi.org/10.1016/j.ajhg.2012.10.013
357. Scott, G. K. (2003). The phosphorylation state of an autoregulatory domain controls PACS-1-directed protein traffic. *The EMBO Journal*, *22*(23), 6234–6244. https://doi.org/10.1093/emboj/cdg596
358. Ségurel, L., Wyman, M. J., & Przeworski, M. (2014). Determinants of Mutation Rate Variation in the Human Germline. *Annual Review of Genomics and Human Genetics*, *15*(1), 47–70. https://doi.org/10.1146/annurev-genom-031714-125740
359. Seo, J. H., Cha, J.-H., Park, J.-H., Jeong, C.-H., Park, Z.-Y., Lee, H.-S., Oh, S. H., Kang, J.-H., Suh, S. W., Kim, K. H., Ha, J. Y., Han, S. H., Kim, S.-H., Lee, J.-W., Park, J. A., Jeong, J.-W., Lee, K.-J., Oh, G. T., Lee, M.-N., … Kim, K.-W. (2010). Arrest Defective 1 Autoacetylation Is a Critical Step in Its Ability to Stimulate Cancer Cell Proliferation. *Cancer Research*, *70*(11), 4422–4432. https://doi.org/10.1158/0008-5472.CAN-09-3258

360. Seplyarskiy, V., Moldovan, M. A., Koch, E., Kar, P., Neville, M. D., Rahbari, R., & Sunyaev, S. (2025). Cohort-level analysis of human de novo mutations points to drivers of clonal expansion in spermatogonia. *medRxiv*. https://doi.org/10.1101/2025.01.03.25319979
361. Seplyarskiy, V., Koch, E. M., Lee, D. J., Lichtman, J. S., Luan, H. H., & Sunyaev, S. R. (2023). A mutation rate model at the basepair resolution identifies the mutagenic effect of polymerase III transcription. *Nature Genetics*, *55*(12), 2235–2242. https://doi.org/10.1038/s41588-023-01562-0
362. Seshacharyulu, P., Pandey, P., Datta, K., & Batra, S. K. (2013). Phosphatase: PP2A structural importance, regulation and its aberrant expression in cancer. *Cancer Letters*, *335*(1), 9–18. https://doi.org/10.1016/j.canlet.2013.02.036
363. Shah, A. A., Amjad, M., Hassan, J.-U., Ullah, A., Mahmood, A., Deng, H., Ali, Y., Gul, F., & Xia, K. (2022). Molecular Insights into the Role of Pathogenic nsSNPs in GRIN2B Gene Provoking Neurodevelopmental Disorders. *Genes*, *13*(8), 1332. https://doi.org/10.3390/genes13081332
364. Shang, L., Henderson, L. B., Cho, M. T., Petrey, D. S., Fong, C.-T., Haude, K. M., Shur, N., Lundberg, J., Hauser, N., Carmichael, J., Innis, J., Schuette, J., Wu, Y. W., Asaikar, S., Pearson, M., Folk, L., Retterer, K., Monaghan, K. G., & Chung, W. K. (2016). De novo missense variants in PPP2R5D are associated with intellectual disability, macrocephaly, hypotonia, and autism. *Neurogenetics*, *17*(1), 43–49. https://doi.org/10.1007/s10048-015-0466-9
365. Shinde, D. N., Elmer, D. P., Calabrese, P., Boulanger, J., Arnheim, N., & Tiemann-Boege, I. (2013). New evidence for positive selection helps explain the paternal age effect observed in achondroplasia. *Human Molecular Genetics*, *22*(20), 4117–4126. https://doi.org/10.1093/hmg/ddt260
366. Shu, W., Yang, H., Zhang, L., Lu, M. M., & Morrisey, E. E. (2001). Characterization of a New Subfamily of Winged-helix/Forkhead (Fox) Genes That Are Expressed in the Lung and Act as Transcriptional Repressors. *Journal of Biological Chemistry*, *276*(29), 27488–27497. https://doi.org/10.1074/jbc.M100636200
367. Shukla, V., Coumoul, X., Wang, R.-H., Kim, H.-S., & Deng, C.-X. (2007). RNA interference and inhibition of MEK-ERK signaling prevent abnormal skeletal phenotypes in a mouse model of craniosynostosis. *Nature Genetics*, *39*(9), 1145–1150. https://doi.org/10.1038/ng2096
368. Sieburth, D. S., Sun, Q., & Han, M. (1998). SUR-8, a Conserved Ras-Binding Protein with Leucine-Rich Repeats, Positively Regulates Ras-Mediated Signaling in C. elegans. *Cell*, *94*(1), 119–130. https://doi.org/10.1016/S0092-8674(00)81227-1
369. Silverstein, A. M., Barrow, C. A., Davis, A. J., & Mumby, M. C. (2002). Actions of PP2A on the MAP kinase pathway and apoptosis are mediated by distinct regulatory subunits. *Proceedings of the National Academy of Sciences*, *99*(7), 4221–4226. https://doi.org/10.1073/pnas.072071699
370. Sisakhtnezhad, S., & Heshmati, P. (2018). Comparative analysis of single-cell RNA sequencing data from mouse spermatogonial and mesenchymal stem cells to identify differentially expressed genes and transcriptional regulators of germline cells. *Journal of Cellular Physiology*, *233*(7), 5231–5242. https://doi.org/10.1002/jcp.26303

371. Smith, T., Heger, A., & Sudbery, I. (2017). UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Research*, *27*(3), 491–499. https://doi.org/10.1101/gr.209601.116
372. Smolen, K. A., Papke, C. M., Swingle, M. R., Musiyenko, A., Li, C., Salter, E. A., Camp, A. D., Honkanen, R. E., & Kettenbach, A. N. (2023). Quantitative proteomics and phosphoproteomics of PP2A-PPP2R5D variants reveal deregulation of RPS6 phosphorylation via converging signaling cascades. *Journal of Biological Chemistry*, *299*(9), 105154. https://doi.org/10.1016/j.jbc.2023.105154
373. Sohni, A., Tan, K., Song, H.-W., Burow, D., De Rooij, D. G., Laurent, L., Hsieh, T.-C., Rabah, R., Hammoud, S. S., Vicini, E., & Wilkinson, M. F. (2019). The Neonatal and Adult Human Testis Defined at the Single-Cell Level. *Cell Reports*, *26*(6), 1501-1517.e4. https://doi.org/10.1016/j.celrep.2019.01.045
374. Soljak, M. A., Aftimos, S., & Gluckman, P. D. (1983). A new syndrome of short stature, joint limitation and muscle hypertrophy. *Clinical Genetics*, *23*(6), 441–446. https://doi.org/10.1111/j.1399-0004.1983.tb01979.x
375. Sollis, E., Graham, S. A., Vino, A., Froehlich, H., Vreeburg, M., Dimitropoulou, D., Gilissen, C., Pfundt, R., Rappold, G. A., Brunner, H. G., Deriziotis, P., & Fisher, S. E. (2016). Identification and functional characterization of *de novo FOXP1* variants provides novel insights into the etiology of neurodevelopmental disorder. *Human Molecular Genetics*, *25*(3), 546–557. https://doi.org/10.1093/hmg/ddv495
376. Song, D. H., Dominguez, I., Mizuno, J., Kaut, M., Mohr, S. C., & Seldin, D. C. (2003). CK2 Phosphorylation of the Armadillo Repeat Region of β-Catenin Potentiates Wnt Signaling. *Journal of Biological Chemistry*, *278*(26), 24018–24025. https://doi.org/10.1074/jbc.M212260200
377. Sonne, S. B., Almstrup, K., Dalgaard, M., Juncker, A. S., Edsgard, D., Ruban, L., Harrison, N. J., Schwager, C., Abdollahi, A., Huber, P. E., Brunak, S., Gjerdrum, L. M., Moore, H. D., Andrews, P. W., Skakkebaek, N. E., Meyts, E. R.-D., & Leffers, H. (2009). Analysis of Gene Expression Profiles of Microdissected Cell Populations Indicates that Testicular Carcinoma *In situ* Is an Arrested Gonocyte. *Cancer Research*, *69*(12), 5241–5250. https://doi.org/10.1158/0008-5472.CAN-08-4554
378. Spencer, D. H., Tyagi, M., Vallania, F., Bredemeyer, A. J., Pfeifer, J. D., Mitra, R. D., & Duncavage, E. J. (2014). Performance of Common Analysis Methods for Detecting Low-Frequency Single Nucleotide Variants in Targeted Next-Generation Sequence Data. *The Journal of Molecular Diagnostics*, *16*(1), 75–88. https://doi.org/10.1016/j.jmoldx.2013.09.003
379. Splendore, A., Jabs, E. W., Félix, T. M., & Passos-Bueno, M. R. (2003). Parental origin of mutations in sporadic cases of Treacher Collins syndrome. *European Journal of Human Genetics*, *11*(9), 718–722. https://doi.org/10.1038/sj.ejhg.5201029
380. Starr, L. J., Grange, D. K., Delaney, J. W., Yetman, A. T., Hammel, J. M., Sanmann, J. N., Perry, D. A., Schaefer, G. B., & Olney, A. H. (2015). Myhre syndrome: Clinical features and restrictive cardiopulmonary complications. *American Journal of Medical Genetics Part A*, *167*(12), 2893–2901. https://doi.org/10.1002/ajmg.a.37273
381. Starr, L. J., Lindsay, M. E., Perry, D., Gheewalla, G., VanderLaan, P. A., Majid, A., Strange, C., Costea, G.-C., Lungu, A., & Lin, A. E. (2022). Review of the Pathologic Characteristics in Myhre Syndrome: Gain-of-Function Pathogenic Variants in *SMAD4*

cause a Multisystem Fibroproliferative Response. *Pediatric and Developmental Pathology*, *25*(6), 611–623. https://doi.org/10.1177/10935266221079569

382. Steele, C. D., Abbasi, A., Islam, S. M. A., Bowes, A. L., Khandekar, A., Haase, K., Hames-Fathi, S., Ajayi, D., Verfaillie, A., Dhami, P., McLatchie, A., Lechner, M., Light, N., Shlien, A., Malkin, D., Feber, A., Proszek, P., Lesluyes, T., Mertens, F., … Pillay, N. (2022). Signatures of copy number alterations in human cancer. *Nature*, *606*(7916), 984–991. https://doi.org/10.1038/s41586-022-04738-6
383. Sun, X., Peng, X., Cao, Y., Zhou, Y., & Sun, Y. (2020). ADNP promotes neural differentiation by modulating Wnt/β-catenin signaling. *Nature Communications*, *11*(1), 2984. https://doi.org/10.1038/s41467-020-16799-0
384. Swanger, S. A., Chen, W., Wells, G., Burger, P. B., Tankovic, A., Bhattacharya, S., Strong, K. L., Hu, C., Kusumoto, H., Zhang, J., Adams, D. R., Millichap, J. J., Petrovski, S., Traynelis, S. F., & Yuan, H. (2016). Mechanistic Insight into NMDA Receptor Dysregulation by Rare Variants in the GluN2A and GluN2B Agonist Binding Domains. *The American Journal of Human Genetics*, *99*(6), 1261–1280. https://doi.org/10.1016/j.ajhg.2016.10.002
385. Szczałuba, K., Biernacka, A., Szymańska, K., Gasperowicz, P., Kosińska, J., Rydzanicz, M., & Płoski, R. (2018). Novel GNB1 de novo mutation in a patient with neurodevelopmental disorder and cutaneous mastocytosis: Clinical report and literature review. *European Journal of Medical Genetics*, *61*(3), 157–160. https://doi.org/10.1016/j.ejmg.2017.11.010
386. Takahashi, M. (2022). RET receptor signaling: Function in development, metabolic disease, and cancer. *Proceedings of the Japan Academy, Series B*, *98*(3), 112–125. https://doi.org/10.2183/pjab.98.008
387. Takashima, S., Kanatsu-Shinohara, M., Tanaka, T., Morimoto, H., Inoue, K., Ogonuki, N., Jijiwa, M., Takahashi, M., Ogura, A., & Shinohara, T. (2015). Functional Differences between GDNF-Dependent and FGF2-Dependent Mouse Spermatogonial Stem Cell Self-Renewal. *Stem Cell Reports*, *4*(3), 489–502. https://doi.org/10.1016/j.stemcr.2015.01.010
388. Tartaglia, M., Cordeddu, V., Chang, H., Shaw, A., Kalidas, K., Crosby, A., Patton, M. A., Sorcini, M., Van Der Burgt, I., Jeffery, S., & Gelb, B. D. (2004). Paternal Germline Origin and Sex-Ratio Distortion in Transmission of PTPN11 Mutations in Noonan Syndrome. *The American Journal of Human Genetics*, *75*(3), 492–497. https://doi.org/10.1086/423493
389. Tartaglia, M., & Gelb, B. D. (2005). Germ-line and somatic PTPN11 mutations in human disease. *European Journal of Medical Genetics*, *48*(2), 81–96. https://doi.org/10.1016/j.ejmg.2005.03.001
390. Tartaglia, M., Martinelli, S., Stella, L., Bocchinfuso, G., Flex, E., Cordeddu, V., Zampino, G., Van Der Burgt, I., Palleschi, A., Petrucci, T. C., Sorcini, M., Schoch, C., Foà, R., Emanuel, P. D., & Gelb, B. D. (2006a). Diversity and Functional Consequences of Germline and Somatic PTPN11 Mutations in Human Disease. *The American Journal of Human Genetics*, *78*(2), 279–290. https://doi.org/10.1086/499925
391. Tartaglia, M., Martinelli, S., Stella, L., Bocchinfuso, G., Flex, E., Cordeddu, V., Zampino, G., Van Der Burgt, I., Palleschi, A., Petrucci, T. C., Sorcini, M., Schoch, C., Foà, R., Emanuel, P. D., & Gelb, B. D. (2006b). Diversity and Functional Consequences of Germline and Somatic PTPN11 Mutations in Human Disease. *The American Journal of Human Genetics*, *78*(2), 279–290. https://doi.org/10.1086/499925

392. Tartaglia, M., Martinelli, S., Stella, L., Bocchinfuso, G., Flex, E., Cordeddu, V., Zampino, G., Van Der Burgt, I., Palleschi, A., Petrucci, T. C., Sorcini, M., Schoch, C., Foà, R., Emanuel, P. D., & Gelb, B. D. (2006c). Diversity and Functional Consequences of Germline and Somatic PTPN11 Mutations in Human Disease. *The American Journal of Human Genetics*, *78*(2), 279–290. https://doi.org/10.1086/499925
393. Terrell, E. M., & Morrison, D. K. (2019). Ras-Mediated Activation of the Raf Family Kinases. *Cold Spring Harbor Perspectives in Medicine*, *9*(1), a033746. https://doi.org/10.1101/cshperspect.a033746
394. Teven, C. M., Farina, E. M., Rivas, J., & Reid, R. R. (2014). Fibroblast growth factor (FGF) signaling in development and skeletal diseases. *Genes & Diseases*, *1*(2), 199–213. https://doi.org/10.1016/j.gendis.2014.09.005
395. The Treacher Collins Syndrome Collaborative Group, Dixon, J., Edwards, S. J., Gladwin, A. J., Dixon, M. J., Loftus, S. K., Bonner, C. A., Koprivnikar, K., & Wasmuth, J. J. (1996). Positional cloning of a gene involved in the pathogenesis of Treacher Collins syndrome. *Nature Genetics*, *12*(2), 130–136. https://doi.org/10.1038/ng0296-130
396. Thi My Nhung, T., Phuoc Long, N., Diem Nghi, T., Suh, Y., Hoang Anh, N., Jung, C. W., Minh Triet, H., Jung, M., Woo, Y., Yoo, J., Noh, S., Kim, S. J., Lee, S. B., Park, S., Thomas, G., Simmen, T., Mun, J., Rhee, H.-W., Kwon, S. W., & Park, S. K. (2023). Genome-wide kinase-MAM interactome screening reveals the role of CK2A1 in MAM Ca^{2+} dynamics linked to DEE66. *Proceedings of the National Academy of Sciences*, *120*(32), e2303402120. https://doi.org/10.1073/pnas.2303402120
397. Thomas, G., Aslan, J. E., Thomas, L., Shinde, P., Shinde, U., & Simmen, T. (2017). Caught in the act – protein adaptation and the expanding roles of the PACS proteins in tissue homeostasis and disease. *Journal of Cell Science*, *130*(11), 1865–1876. https://doi.org/10.1242/jcs.199463
398. Thompson, J., & Williams, C. (2018). Protein Phosphatase 2A in the Regulation of Wnt Signaling, Stem Cells, and Cancer. *Genes*, *9*(3), 121. https://doi.org/10.3390/genes9030121
399. Thul, P. J., & Lindskog, C. (2018). The human protein atlas: A spatial map of the human proteome. *Protein Science*, *27*(1), 233–244. https://doi.org/10.1002/pro.3307
400. Tiemann-Boege, I., Navidi, W., Grewal, R., Cohn, D., Eskenazi, B., Wyrobek, A. J., & Arnheim, N. (2002). The observed human sperm mutation frequency cannot explain the achondroplasia paternal age effect. *Proceedings of the National Academy of Sciences*, *99*(23), 14952–14957. https://doi.org/10.1073/pnas.232568699
401. Tillotson, R., & Bird, A. (2020). The Molecular Basis of MeCP2 Function in the Brain. *Journal of Molecular Biology*, *432*(6), 1602–1623. https://doi.org/10.1016/j.jmb.2019.10.004
402. Tillotson, R., Cholewa-Waclaw, J., Chhatbar, K., Connelly, J. C., Kirschner, S. A., Webb, S., Koerner, M. V., Selfridge, J., Kelly, D. A., De Sousa, D., Brown, K., Lyst, M. J., Kriaucionis, S., & Bird, A. (2021). Neuronal non-CG methylation is an essential target for MeCP2 function. *Molecular Cell*, *81*(6), 1260-1275.e12. https://doi.org/10.1016/j.molcel.2021.01.011
403. Titomanlio, L., Marzano, M. G., Rossi, E., D'Armiento, M., De Brasi, D., Vega, G. R., Andreucci, M. V., Orsini, A. V. M., Santoro, L., & Sebastio, G. (2001). Case of Myhre syndrome with autism and peculiar skin histological findings. *American Journal of Medical Genetics*, *103*(2), 163–165. https://doi.org/10.1002/ajmg.1517

404. Tolarova, M. M., Harris, J. A., Ordway, D. E., & Vargervik, K. (1997). Birth prevalence, mutation rate, sex ratio, parents' age, and ethnicity in Apert syndrome. *American Journal of Medical Genetics*, *72*(4), 394–398. https://doi.org/10.1002/(SICI)1096-8628(19971112)72:4<394::AID-AJMG4>3.0.CO;2-R
405. Trappe, R., Laccone, F., Cobilanschi, J., Meins, M., Huppke, P., Hanefeld, F., & Engel, W. (2001). MECP2 Mutations in Sporadic Cases of Rett Syndrome Are Almost Exclusively of Paternal Origin. *The American Journal of Human Genetics*, *68*(5), 1093–1101. https://doi.org/10.1086/320109
406. Trinh, J., Hüning, I., Budler, N., Hingst, V., Lohmann, K., & Gillessen-Kaesbach, G. (2017). A novel de novo mutation in CSNK2A1: Reinforcing the link to neurodevelopmental abnormalities and dysmorphic features. *Journal of Human Genetics*, *62*(11), 1005–1006. https://doi.org/10.1038/jhg.2017.73
407. United States Government Accountability Office. (2023). *Medicaid: Characteristics of and Expenditures for Adults with Intellectual or Developmental Disabilities* (No. GAO-23-105457).
408. Vadeboncoeur, C., & McHardy, M. (2018). Benefits of Early Referral to Pediatric Palliative Care for a Child With a Rare Disease. *Pediatrics*, *141*(6), e20173417. https://doi.org/10.1542/peds.2017-3417
409. Van Dijck, A., Vulto-van Silfhout, A. T., Cappuyns, E., Van Der Werf, I. M., Mancini, G. M., Tzschach, A., Bernier, R., Gozes, I., Eichler, E. E., Romano, C., Lindstrand, A., Nordgren, A., Kvarnung, M., Kleefstra, T., De Vries, B. B. A., Küry, S., Rosenfeld, J. A., Meuwissen, M. E., Vandeweyer, G., … Slattery, L. (2019). Clinical Presentation of a Complex Neurodevelopmental Disorder Caused by Mutations in ADNP. *Biological Psychiatry*, *85*(4), 287–297. https://doi.org/10.1016/j.biopsych.2018.02.1173
410. Varenyiova, Z., Hrckova, G., Ilencikova, D., & Podracka, L. (2020). Myhre Syndrome Associated With Dunbar Syndrome and Urinary Tract Abnormalities: A Case Report. *Frontiers in Pediatrics*, *8*, 72. https://doi.org/10.3389/fped.2020.00072
411. Veltman, J. A., & Brunner, H. G. (2012). De novo mutations in human genetic disease. *Nature Reviews Genetics*, *13*(8), 565–575. https://doi.org/10.1038/nrg3241
412. Villar-Pazos, S., Thomas, L., Yang, Y., Chen, K., Lyles, J. B., Deitch, B. J., Ochaba, J., Ling, K., Powers, B., Gingras, S., Kordasiewicz, H. B., Grubisha, M. J., Huang, Y. H., & Thomas, G. (2023). Neural deficits in a mouse model of PACS1 syndrome are corrected with PACS1- or HDAC6-targeting therapy. *Nature Communications*, *14*(1), 6547. https://doi.org/10.1038/s41467-023-42176-8
413. Vogel, F., & Rathenberg, R. (n.d.). *Spontaneous Mutation in Man*.
414. Von Kopylow, K., Staege, H., Spiess, A.-N., Schulze, W., Will, H., Primig, M., & Kirchhoff, C. (2012). Differential marker protein expression specifies rarefaction zone-containing human Adark spermatogonia. *REPRODUCTION*, *143*(1), 45–57. https://doi.org/10.1530/REP-11-0290
415. Vulih-Shultzman, I., Pinhasov, A., Mandel, S., Grigoriadis, N., Touloumi, O., Pittel, Z., & Gozes, I. (2007). Activity-Dependent Neuroprotective Protein Snippet NAP Reduces Tau Hyperphosphorylation and Enhances Learning in a Novel Transgenic Mouse Model. *Journal of Pharmacology and Experimental Therapeutics*, *323*(2), 438–449. https://doi.org/10.1124/jpet.107.129551

416. Walker, B. A., Murdoch, J. L., McKusick, V., Langer, L., & Beals, R. (1971). Hypochondroplasia. *Archives of Pediatrics & Adolescent Medicine*, *122*(2), 95. https://doi.org/10.1001/archpedi.1971.02110020029001
417. Wan, L., Molloy, S. S., Thomas, L., Liu, G., Xiang, Y., Rybak, S. L., & Thomas, G. (1998). PACS-1 Defines a Novel Gene Family of Cytosolic Sorting Proteins Required for trans-Golgi Network Localization. *Cell*, *94*(2), 205–216. https://doi.org/10.1016/S0092-8674(00)81420-8
418. Wang, J., Chandrasekhar, V., Abbadessa, G., Yu, Y., Schwartz, B., & Kontaridis, M. I. (2017). In vivo efficacy of the AKT inhibitor ARQ 092 in Noonan Syndrome with multiple lentigines-associated hypertrophic cardiomyopathy. *PLOS ONE*, *12*(6), e0178905. https://doi.org/10.1371/journal.pone.0178905
419. Wang, Y., Spatz, M. K., Kannan, K., Hayk, H., Avivi, A., Gorivodsky, M., Pines, M., Yayon, A., Lonai, P., & Givol, D. (1999). A mouse model for achondroplasia produced by targeting fibroblast growth factor receptor 3. *Proceedings of the National Academy of Sciences*, *96*(8), 4455–4460. https://doi.org/10.1073/pnas.96.8.4455
420. Wang, Y., Xiao, R., Yang, F., Karim, B. O., Iacovelli, A. J., Cai, J., Lerner, C. P., Richtsmeier, J. T., Leszl, J. M., Hill, C. A., Yu, K., Ornitz, D. M., Elisseeff, J., Huso, D. L., & Jabs, E. W. (2005). Abnormalities in cartilage and bone development in the Apert syndrome FGFR2+/S252W mouse. *Development*, *132*(15), 3537–3548. https://doi.org/10.1242/dev.01914
421. Weinbauer, G. F., Luetjens, C. M., Simoni, M., & Nieschlag, E. (2010). Physiology of Testicular Function. In E. Nieschlag, H. M. Behre, & S. Nieschlag (Eds.), *Andrology* (pp. 11–59). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-78355-8_2
422. Weinberg, W. (1912). Zur Vererbung des Zwergwuchses. *Arch. Rassen-u Gesell. Biol.*, *9*, 710–718.
423. Wen, Y., Wang, J., Zhang, Q., Yang, X., Wei, L., & Bao, X. (2023). MECP2 germline mosaicism plays an important part in the inheritance of Rett syndrome: A study of MECP2 germline mosaicism in males. *BMC Medicine*, *21*(1), 155. https://doi.org/10.1186/s12916-023-02846-2
424. Werner, C., Gast, A., Lindenblatt, D., Nickelsen, A., Niefind, K., Jose, J., & Hochscherf, J. (2022). Structural and Enzymological Evidence for an Altered Substrate Specificity in Okur-Chung Neurodevelopmental Syndrome Mutant CK2αLys198Arg. *Frontiers in Molecular Biosciences*, *9*, 831693. https://doi.org/10.3389/fmolb.2022.831693
425. Whiteford, M. L., Doig, W. B., Raine, P. A. M., Hollman, A. S., & Tolmie, J. L. (2001). A new case of Myhre syndrome: *Clinical Dysmorphology*, *10*(2), 135–140. https://doi.org/10.1097/00019605-200104000-00011
426. Wilkie, A. O. M. (2005). Bad bones, absent smell, selfish testes: The pleiotropic consequences of human FGF receptor mutations. *Cytokine & Growth Factor Reviews*, *16*(2), 187–203. https://doi.org/10.1016/j.cytogfr.2005.03.001
427. Wilkie, A. O. M., Slaney, S. F., Oldridge, M., Poole, M. D., Ashworth, G. J., Hockley, A. D., Hayward, R. D., David, D. J., Pulleyn, L. J., Rutland, P., Malcolm, S., Winter, R. M., & Reardon, W. (1995). Apert syndrome results from localized mutations of FGFR2 and is allelic with Crouzon syndrome. *Nature Genetics*, *9*(2), 165–172. https://doi.org/10.1038/ng0295-165
428. Williams, J.-C. B., Hamilton, J. K., Shiller, M., Fischer, L., Deprisco, G., & Boland, C. R. (2012). Combined Juvenile Polyposis and Hereditary Hemorrhagic Telangiectasia.

Baylor University Medical Center Proceedings, *25*(4), 360–364. https://doi.org/10.1080/08998280.2012.11928877

429. Wöhrle, F. U., Daly, R. J., & Brummer, T. (2009). Function, regulation and pathological roles of the Gab/DOS docking proteins. *Cell Communication and Signaling*, *7*(1), 22. https://doi.org/10.1186/1478-811X-7-22
430. Wong, M. M., Cox, L. K., & Chrivia, J. C. (2007). The Chromatin Remodeling Protein, SRCAP, Is Critical for Deposition of the Histone Variant H2A.Z at Promoters. *Journal of Biological Chemistry*, *282*(36), 26132–26139. https://doi.org/10.1074/jbc.M703418200
431. Wood, K. A., Tong, R. S., Motta, M., Cordeddu, V., Scimone, E. R., Bush, S. J., Maxwell, D. W., Giannoulatou, E., Caputo, V., Traversa, A., Mancini, C., Ferrero, G. B., Benedicenti, F., Grammatico, P., Melis, D., Steindl, K., Brunetti-Pierri, N., Trevisson, E., Wilkie, A. Om., … Goriely, A. (2024). SMAD4 mutations causing Myhre syndrome are under positive selection in the male germline. *The American Journal of Human Genetics*, *111*(9), 1953–1969. https://doi.org/10.1016/j.ajhg.2024.07.006
432. Wright, D. E., Wagers, A. J., Gulati, A. P., Johnson, F. L., & Weissman, I. L. (2001). Physiological Migration of Hematopoietic Stem and Progenitor Cells. *Science*, *294*(5548), 1933–1936. https://doi.org/10.1126/science.1064081
433. Wu, C., Xie, T., Guo, Y., Wang, D., Qiu, M., Han, R., Qing, G., Liang, K., & Liu, H. (2023). CDK13 phosphorylates the translation machinery and promotes tumorigenic protein synthesis. *Oncogene*, *42*(16), 1321–1330. https://doi.org/10.1038/s41388-023-02653-2
434. Wu, H., Wang, X., Cui, Y., & Wang, X. (2021). Case Report: Improved Height in a Patient With Myhre Syndrome Using a Combination of Growth Hormone and Letrozole. *Frontiers in Pediatrics*, *9*, 675934. https://doi.org/10.3389/fped.2021.675934
435. Xiang, X., Deng, L., Zhang, J., Zhang, X., Lei, T., Luan, G., Yang, C., Xiao, Z.-X., Li, Q., & Li, Q. (2014). A Distinct Expression Pattern of Cyclin K in Mammalian Testes Suggests a Functional Role in Spermatogenesis. *PLoS ONE*, *9*(7), e101539. https://doi.org/10.1371/journal.pone.0101539
436. Xiao, L., Gong, L.-L., Yuan, D., Deng, M., Zeng, X.-M., Chen, L.-L., Zhang, L., Yan, Q., Liu, J.-P., Hu, X.-H., Sun, S.-M., Liu, J., Ma, H.-L., Zheng, C.-B., Fu, H., Chen, P.-C., Zhao, J.-Q., Xie, S.-S., Zou, L.-J., … Li, D. W.-C. (2010). Protein phosphatase-1 regulates Akt1 signal transduction pathway to control gene expression, cell survival and differentiation. *Cell Death & Differentiation*, *17*(9), 1448–1462. https://doi.org/10.1038/cdd.2010.16
437. Xie, Y., Su, N., Yang, J., Tan, Q., Huang, S., Jin, M., Ni, Z., Zhang, B., Zhang, D., Luo, F., Chen, H., Sun, X., Feng, J. Q., Qi, H., & Chen, L. (2020). FGF/FGFR signaling in health and disease. *Signal Transduction and Targeted Therapy*, *5*(1), 181. https://doi.org/10.1038/s41392-020-00222-7
438. Xu, R., Wang, X., Safi, S., Braunegger, N., Hipgrave Ederveen, A., Rottmann, M., Wittbrodt, J., Wuhrer, M., Wesslowski, J., & Davidson, G. (2023). N-Glycosylation of LRP6 by B3GnT2 Promotes Wnt/β-Catenin Signalling. *Cells*, *12*(6), 863. https://doi.org/10.3390/cells12060863
439. Xu, X., Brodie, S. G., Yang, X., Im, Y.-H., Parks, W. T., Chen, L., Zhou, Y.-X., Weinstein, M., Kim, S.-J., & Deng, C.-X. (2000). Haploid loss of the tumor suppressor Smad4/Dpc4 initiates gastric polyposis and cancer in mice. *Oncogene*, *19*(15), 1868–1874. https://doi.org/10.1038/sj.onc.1203504

440. Xu, X., Toselli, P. A., Russell, L. D., & Seldin, D. C. (1999). Globozoospermia in mice lacking the casein kinase II α′ catalytic subunit. *Nature Genetics*, *23*(1), 118–121. https://doi.org/10.1038/12729
441. Yamada, M., Cai, W., Martin, L. A., N'Tumba-Byn, T., & Seandel, M. (2019). Functional robustness of adult spermatogonial stem cells after induction of hyperactive Hras. *PLOS Genetics*, *15*(5), e1008139. https://doi.org/10.1371/journal.pgen.1008139
442. Yang, K., Wang, X., Wang, W., Han, M., Hu, L., Kang, D., Yang, J., Liu, M., Gao, X., Yuan, Y., & Xu, J. (2023). A newborn male with Myhre syndrome, hearing loss, and complete syndactyly of fingers 3–4. *Molecular Genetics & Genomic Medicine*, *11*(3), e2103. https://doi.org/10.1002/mgg3.2103
443. Yang, Z., Le, J. T., Hutter, D., Bradley, K. M., Overton, B. R., McLendon, C., & Benner, S. A. (2020). Eliminating primer dimers and improving SNP detection using self-avoiding molecular recognition systems. *Biology Methods and Protocols*, *5*(1), bpaa004. https://doi.org/10.1093/biomethods/bpaa004
444. Yasuda, T., Nah, H. D., Laurita, J., Kinumatsu, T., Shibukawa, Y., Shibutani, T., Minugh-Purvis, N., Pacifici, M., & Koyama, E. (2012). Muenke Syndrome Mutation, *FgfR3* ^{P244R} , Causes TMJ Defects. *Journal of Dental Research*, *91*(7), 683–689. https://doi.org/10.1177/0022034512449170
445. Yin, L., Du, X., Li, C., Xu, X., Chen, Z., Su, N., Zhao, L., Qi, H., Li, F., Xue, J., Yang, J., Jin, M., Deng, C., & Chen, L. (2008). A Pro253Arg mutation in fibroblast growth factor receptor 2 (Fgfr2) causes skeleton malformation mimicking human Apert syndrome by affecting both chondrogenesis and osteogenesis. *Bone*, *42*(4), 631–643. https://doi.org/10.1016/j.bone.2007.11.019
446. Yoda, A., Adelmant, G., Tamburini, J., Chapuy, B., Shindoh, N., Yoda, Y., Weigert, O., Kopp, N., Wu, S.-C., Kim, S. S., Liu, H., Tivey, T., Christie, A. L., Elpek, K. G., Card, J., Gritsman, K., Gotlib, J., Deininger, M. W., Makishima, H., … Lane, A. A. (2015). Mutations in G protein β subunits promote transformation and kinase inhibitor resistance. *Nature Medicine*, *21*(1), 71–75. https://doi.org/10.1038/nm.3751
447. Yoon, S.-R., Choi, S.-K., Eboreime, J., Gelb, B. D., Calabrese, P., & Arnheim, N. (2013). Age-Dependent Germline Mosaicism of the Most Common Noonan Syndrome Mutation Shows the Signature of Germline Selection. *The American Journal of Human Genetics*, *92*(6), 917–926. https://doi.org/10.1016/j.ajhg.2013.05.001
448. Yoshida, S., Sukeno, M., & Nabeshima, Y. (2007). A Vasculature-Associated Niche for Undifferentiated Spermatogonia in the Mouse Testis. *Science*, *317*(5845), 1722–1726. https://doi.org/10.1126/science.1144885
449. Youker, R. T., Shinde, U., Day, R., & Thomas, G. (2009). At the crossroads of homoeostasis and disease: Roles of the PACS proteins in membrane traffic and apoptosis. *Biochemical Journal*, *421*(1), 1–15. https://doi.org/10.1042/BJ20081016
450. Young, L. C., Hartig, N., Boned Del Río, I., Sari, S., Ringham-Terry, B., Wainwright, J. R., Jones, G. G., McCormick, F., & Rodriguez-Viciana, P. (2018). SHOC2–MRAS–PP1 complex positively regulates RAF activity and contributes to Noonan syndrome pathogenesis. *Proceedings of the National Academy of Sciences*, *115*(45). https://doi.org/10.1073/pnas.1720352115
451. Yu, J., Sui, F., Gu, F., Li, W., Yu, Z., Wang, Q., He, S., Wang, L., & Xu, Y. (2024). Structural insights into histone exchange by human SRCAP complex. *Cell Discovery*, *10*(1), 15. https://doi.org/10.1038/s41421-023-00640-1

452. Yu, K. P.-T., Luk, H.-M., Chung, B. H.-Y., & Lo, I. F.-M. (2019). Myhre syndrome: A report of six Chinese patients and literature review. *Clinical Dysmorphology*, *28*(3), 143–148. https://doi.org/10.1097/MCD.0000000000000271
453. Zambrano, R. M., Marble, M., Chalew, S. A., Lilje, C., Vargas, A., & Lacassie, Y. (2017). Further evidence that variants in *PPP1CB* cause a rasopathy similar to Noonan syndrome with loose anagen hair. *American Journal of Medical Genetics Part A*, *173*(2), 565–567. https://doi.org/10.1002/ajmg.a.38056
454. Zang, R. X., Mumby, M. J., & Dikeakos, J. D. (2022). The Phosphofurin Acidic Cluster Sorting Protein 2 (PACS-2) E209K Mutation Responsible for PACS-2 Syndrome Increases Susceptibility to Apoptosis. *ACS Omega*, *7*(38), 34378–34388. https://doi.org/10.1021/acsomega.2c04014
455. Zhang, M., Ma, T., Hu, B., & Xiang, W. (2022). FOXP1 promotes osteoblast differentiation via regulation of TGF-β/ALK-5 pathway. *ScienceAsia*, *48*(4), 423. https://doi.org/10.2306/scienceasia1513-1874.2022.071
456. Zhang, Q., Yang, X., Wang, J., Li, J., Wu, Q., Wen, Y., Zhao, Y., Zhang, X., Yao, H., Wu, X., Yu, S., Wei, L., & Bao, X. (2019). Genomic mosaicism in the pathogenesis and inheritance of a Rett syndrome cohort. *Genetics in Medicine*, *21*(6), 1330–1338. https://doi.org/10.1038/s41436-018-0348-2
457. Zheng, J., Wang, A., Huang, P., Xiong, J., Liu, L., Xiao, Y., Li, J., Mao, D., & Liu, L. (2019). *A child with Myhre syndrome presenting with moyamoya disease: A case report and literature review*. https://doi.org/10.21203/rs.2.1868/v1
458. Zheng, Y., Thomas, A., Schmidt, C. M., & Dann, C. T. (2014). Quantitative detection of human spermatogonia for optimization of spermatogonial stem cell culture. *Human Reproduction*, *29*(11), 2497–2511. https://doi.org/10.1093/humrep/deu232
459. Zhou, P., Zhu, L., Fan, Q., Liu, Y., Zhang, T., Yang, T., Chen, J., Cheng, Q., Li, T., & Chen, L. (2020). A case report of Noonan syndrome-like disorder with loose anagen hair 2 treated with recombinant human growth hormone. *American Journal of Medical Genetics Part A*, *182*(8), 1967–1971. https://doi.org/10.1002/ajmg.a.61638
460. Zmolikova, M., Puchmajerova, A., Hecht, P., Lebl, J., Trkova, M., & Krepelova, A. (2014). Coarctation of the aorta in Noonan-like syndrome with loose anagen hair. *American Journal of Medical Genetics Part A*, *164*(5), 1218–1221. https://doi.org/10.1002/ajmg.a.36404
461. Zuo, Q., Jin, K., Zhang, Y., Song, J., & Li, B. (2017). Dynamic expression and regulatory mechanism of TGF-β signaling in chicken embryonic stem cells differentiating into spermatogonial stem cells. *Bioscience Reports*, *37*(4), BSR20170179. https://doi.org/10.1042/BSR20170179
462. Zhu, B., Dong, Q., Huang, X., Ji, G., Che, Y., Wang, W., Jiang, H., Gao, J. (2003). Mutation analysis of fibroblast growth factor receptor 3 gene in an achondroplasia family. *Chinese Journal of Medical Genetics, 20*(5), 373–375.