

https://dspace.org/

Mutation Profiling of Paediatric Solid Tumours in a Cohort of South African Patients

Manolas, Erin

2023
University of the Witwatersrand, Johannesburg
https://hdl.handle.net/10539/42318
Manolas, Erin. (2024). Mutation Profiling of Paediatric Solid Tumours in a Cohort of South African Patients [Master's dissertation , University of the Witwatersrand, Johannesburg]. WireDSpace. https://hdl.handle.net/10539/42318

Downloaded from a DSpace repository. https://dspace.org/

Mutation Profiling of Paediatric Solid Tumours in a Cohort of South African Patients

Erin Manolas

Student Number: 733739

Supervisor: Dr Lindie Lamola

Co-supervisor: Prof Amanda Krause



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

A Dissertation submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Medicine (Human Genetics)

Johannesburg, 2023

Declaration

I, Erin Lee Manolas, declare that this Dissertation is my own, unaided work. It is being submitted for the Degree of Master of Science in Medicine (Human Genetics) at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



(Signature of the Candidate)

10th day of October 2023 in Johannesburg

Abstract

Childhood cancers are an emerging global health burden, with the highest increase in incidence and mortality rates occurring in low-middle income countries, such as South Africa (SA). Adding to this burden is the contribution of cancer-predisposing genes (CPGs), whose germline variants increase the risk of cancer development in childhood. These genes are largely associated with a set of disorders, known as cancer-predisposing syndromes (CPSs), which are characterised by an increased likelihood of cancer development and/or additional phenotypic malignant/non-malignant features. Next-Generation Sequencing (NGS) technologies have been key in determining the occurrence and contribution of such germline variants to paediatric cancer development across international research and diagnostic settings. However, these technologies have not been applied to paediatric cohorts in SA and thus there is a paucity of data regarding the contribution of germline variants to childhood cancer development in this setting. Through the design and evaluation of an NGS targeted-CPG panel, this study aimed to generate germline variant profiles of SA paediatric cancer patients, thereby gaining insight into their potential role in the pathogenesis of childhood cancers.

NGS was performed on the genomic DNA from 32 solid-tumour paediatric cancer patients using an Ion Ampliseq 50 CPG panel design and the Ion Torrent S5 sequencing instruments. Germline variants were called using the Ion Torrent Suite™ software (v.5.12.0) and annotated using the Ensembl Variant Effect Predictor. Variants were filtered using a bioinformatics pipeline assessing variant data from population and public databases, computational data, functional studies data, genetic pedigrees indicating family history and phenotypic data. Variant evidence was further interpreted for variant prioritisation and classification according to the ACMG-AMP guidelines. All putative pathogenic/likely-pathogenic (P/LP) variants identified were validated via Sanger Sequencing.

Seven pathogenic and/or likely pathogenic germline variants were identified and validated in seven patients. Three of these variants, identified in the *NF1, RET*, and *TP53* genes, were detected in patients who presented with phenotypes consistent with their genetic findings and are associated with well-known CPSs (diagnostic yield - 3/32, ~9.4%). The remaining four variants, identified in the *BRCA1, ERCC3, FAH*, and *RB1* genes, have not been previously associated with the patient's cancer phenotype and therefore require further investigation. At the time of this project's data generation, this is the first global report of the novel heterozygous, likely pathogenic *FAH* p.R162H variant. Additionally, although reported elsewhere, the majority of the variants identified in this study (6/7, ~86.7%) have been reported for the first time within the SA paediatric population. To our knowledge, this is the first study to utilise NGS technologies in the germline variant profiling of paediatric solid-tumour patients in SA and therefore has greatly added to filling the current knowledge gap. In addition, these findings have contributed towards the foundation for the development of a CPG sequencing panel suitable for implementation in a SA diagnostic setting.

Acknowledgements

I would like to acknowledge and thank the following people:

- This study's participants for consenting to participate and provide samples used in this study. This study would not have been possible without their participation, and I am grateful to be given the opportunity to contribute to the paediatric cancer research community.
- My supervisor Dr Lindie Lamola and co-supervisor Prof Amanda Krause, for firstly granting me the opportunity to undertake my MSc under the PeCan Project. Secondly, I would like to thank my supervisors for providing their continuous guidance, expertise, support, and understanding throughout my entire MSc journey.
- To my supervisor Dr Lindie Lamola and the National Research Foundation for the funding that was provided through the NRF Thuthuka grant awarded to her. This financial support made it possible for me to continue my post-graduate education, further allowing for my continued growth in my academic career.
- The PeCan Project collaborators, whose contributions have been key in making this study a possibility. To Dr Njabulo Mabaso, who was my partner during the patient recruitment process, and whose work provided crucial clinical information to this study. To the paediatric oncologists, Prof Jennifer Geel and Dr Gita Naidu, who facilitated the recruitment and sample collection process through the paediatric oncology units at Charlotte Maxeke and Chris Hani Baragwanath Academic Hospitals, respectively.
- To my family and David Stubbs for always providing me with unconditional love, support, and encouragement. Thank you for always believing in me, even when I didn't believe in myself. I love and appreciate you all immensely.
- To my fellow post-graduate students for their guidance and companionship throughout this journey, such as Heather Seymour, whose guidance and advice aided me in optimizing my Sanger sequencing validation technique.
- To colleagues who became life-long friends, Felicity Mhlongo and David Twesigomwe, I have thoroughly enjoyed growing and learning alongside you in this journey. I am immensely grateful for your friendship and truly appreciate all the support you have given me both in the academic and my personal life environment.
- To all my friends, thank you for your continued friendship, support and understanding over the past few years, through all the good times and the bad. I truly value your friendship.

Presentations Arising from This Research Project:

- Manolas E., Mabaso M., Geel J., Naidu G., Krause A., Lamola L. Oral presentation: "Mutation Profiling of Paediatric Solid Tumours in a Cohort of South African Patients". University of the Witwatersrand, Johannesburg, South Africa, Faculty of Health Sciences (FHS) Biennial Research Day (virtual), 15th October 2020.
- Manolas E., Mabaso M., Geel J., Naidu G., Krause A., Lamola L. Poster presentation: "Mutation Profiling of Paediatric Solid Tumours in a Cohort of South African Patients". 14th International Congress of Human Genetics (ICHG) – Cape Town, South Africa, 22nd – 26th February 2023.

Table of Contents

Declaration i

Abstract ii

Acknowledgements iii

Presentations Arising from This Research Project: iii

Table of Contents iv

List of Figures vii

List of Tables ix

Nomenclature (List of Abbreviations, Acronyms and Symbols) x

Preface xii

1. Introduction 1

1.1 Childhood vs Adult Cancers 1

1.2 Epidemiology of Childhood Cancers 2

1.3 Genetic Predisposition to Childhood Cancer 3

1.3.1 History & Identification of Cancer Predisposition Genes 4

1.3.2 Mechanisms and Functions of Cancer Predisposition Genes 5

1.3.3 Familial/Hereditary Predisposition Cancer Syndromes 8

1.3.4 Cancer Predisposition Genes in a Clinical Setting 11

1.3.5 The Complexities of Cancer Predisposition 14

1.4 Next-Generation Sequencing in a Cancer Research & Diagnostic Setting 15

1.4.1 Next-Generation Sequencing in Paediatric Cancer Genetics 17

1.5 Project Rationale 18

1.6 Aims & Objectives 19

1.6.1 Aims 19

1.6.2 Objectives 19

2. Methods & Materials 20

2.1 Study Participants 20

2.1.1 Ethics, Recruitment & Sample Collection 20

2.1.2 Patient Selection & Electronic Data Capturing 21

2.2 Experimental Techniques 22

2.2.1 DNA Extraction 22

2.2.2 DNA Quantification & Quality Control 22

2.2.3 Next Generation Sequencing 23

2.2.4 Bioinformatics – Analysis of NGS Data Output 30

2.2.5 Validation 38

2.2.6 Phenotype – Genotype Relationship 40

3. Results 41

3.1 Patient Demographics 41

3.2 Next-Generation Sequencing Data Quality Control 41

3.2.1 Pre-Alignment Metrics 41

3.2.2 Post-Alignment Metrics 45

3.3 Bioinformatics – Interpretation of NGS Data Output 47

3.3.1 Variant Filtering Output 47

3.3.2 Variant Prioritisation & Classification Output 48

3.4 NGS Prioritised Variants 49

3.4.1 Putative Disease-Causing/Contributing Variants 49

4. Discussion 77

4.1 Project Summary 77

4.2 Project Yield 78

4.3 Mutation-Positive Cases 79

4.3.1 Variants Identified in Patients with Strong Phenotypic Evidence 80

4.3.2 Variants Identified in Patients with Previously Unrelated Phenotypic Evidence 86

4.4 False Positives 96

4.5 Mutation-Negative Cases 97

4.6 Significance & Prognostic Implications of CPG Testing 99

4.7 Challenges, Limitations & Recommendations 101

4.8 Future Directions 104

4.9 Conclusion 106

Electronic Sources 108

References 108

Appendices 129

Appendix A: Plagiarism Declaration 129

Appendix B: Ethics & Recruitment Documentation 130

Appendix B1.1: PeCan Project Ethics Certificate 130

Appendix B1.2: Current Study Ethics Certificate 131

Appendix B2.1: Parent/Legal Guardian Information Sheets 132

Appendix B2.2: Child Participant Information Sheets 135

Appendix B3.1: Parent/Legal Guardian Consent Forms 137

Appendix B3.2: Child Participant Assent Forms 141

Appendix C: Methods Documentation 145

Appendix C1: Gene-Panel Table 145

Appendix C2: Ion Reporter™ & VEP Annotation Sets 150

Appendix C3: *In-Silico* Tools and Interpretations – utilised in variant assessment and filtration. 153

Appendix C4: ACMG-AMP Guidelines for Variant Interpretation & Classification (Tables taken from Richards et al (2015)) 155

Appendix C5: Validation Tables 157

Appendix C5.1: Websites/*In-silico* Tools used in Primer Design 157

Appendix C5.2: Primer Design Utilised in Validation of Seven Pathogenic/Likely Pathogenic Germline Variants: 157

Appendix C5.3: Standard Temperature Gradient PCR Thermal Cycling Conditions: 158

Appendix C5.4: Optimal Primer-Specific PCR Temperatures for the Seven Prioritised Variants: 158

Appendix C5.5: Cycle Sequencing Conditions used in Sanger Sequencing Validation: 159

Appendix D: Results Documentation 159

Appendix D1: Patient Demographics Table 159

Appendix D2: Post-Alignment Sequencing Quality Metrics per sample ...162

Appendix D3.1: Resultant Gel Electrophoresis Images for PCR Primer – Pair Optimisation & PCR Patient DNA Amplification 163

Appendix D3.1: Resultant Gel Electrophoresis Images for PCR Primer – Pair Optimisation & PCR Patient DNA 164

Appendix D3.2: Resultant Gel Electrophoresis Images for PCR Primer – Pair Optimisation & PCR Patient DNA Amplification 165

Appendix D3.2: Resultant Gel Electrophoresis Images for PCR Primer – Pair Optimisation & PCR Patient DNA Amplification 166

Appendix D4.1: IRGV Images for Mutation-Positive Cases 167

Appendix D4.1: IRGV Images for Mutation-Positive Cases 168

Appendix D4.2: IRGV Images for Mutation-Positive Cases 169

Appendix D4.2: IRGV Images for Mutation-Positive Cases 170

Appendix E: Turnitin Report 171

List of Figures

Figure 1.1: Graphical Illustration of the Two Broad Mechanisms of Oncogenesis (A & B) 6

Figure 2.1: Next-Generation Sequencing Workflow 24

Figure 2.2: Diagram depicting 50 Cancer-Predisposing Genes Selected for Targeted NGS Gene Panel (A & B) 25

Figure 2.3: Library Preparation Workflow 27

Figure 2.4: Template Preparation & Automated Chip Loading Workflow 30

Figure 2.5: Chip Sequencing & NGS Data Generation Workflow 31

Figure 2.6: Bioinformatics Workflow....32
Figure 2.7: Ion Torrent to VEP VCF Reformatting Example....36
Figure 2.8: VEP Annotation Command....36
Figure 2.9: Sanger Sequencing Workflow....40
Figure 3.1: Pie Charts and Column Graph Depicting the Summary of the 32 Patient Demographics (A-E) 45
Figure 3.2: ISP Loading, ISP Density & ISP Summary Generated for the First (A) and Second (B) Sequencing Runs....47
Figure 3.3: Read Length Histogram of the Trimmed Lengths of all the Reads Present in the First & Second Sequencing Runs' Output Files....48
Figure 3.4: First (A) and Second (B) Sequencing runs' Post-Alignment Sequencing Quality Metrics..... 49
Figure 3.5: Results of Variant Filtering....51
Figure 3.6: Patient PC.0023.01.1 Genetic Pedigree....58
Figure 3.7: Mutalyzer 2.0.35 Output for Variant NF1 (NM_001042492.3) c.1133_1136delTGAC (p.D378Afs*8) Identified in Patient PC.0023.01.1 (A & B) 59
Figure 3.8: Electropherogram Confirming Variant NF1 c.1133_1136delTGAC in Patient PC.0023.01.1 (A & B)....60
Figure 3.9: Patient PC.0032.01.1 Genetic Pedigree....62
Figure 3.10: Electropherogram Confirming Variant RET c.1901G>A in the Forward Strand of Patient PC.0032.01.1....66
Figure 3.11: Patient PC.0051.01.1 Genetic Pedigree....67
Figure 3.12: Electropherogram Confirming Variant TP53 c.817C>T in the Forward Strand of Patient PC.0051.01.1....71
Figure 3.13: Electropherogram Confirming Variant BRCA1 c.4199T>C in the Forward Strand of Patient PC.0011.01.1....74
Figure 3.14: Electropherogram Confirming Variant FAH c.485G>A in the Forward Strand of Patient PC.0009.01.1....76
Figure 3.15: Mutalyzer 2.0.35 Output for Variant ERCC3 (NM_000122.2): c.958C>T (p.Q320*) Identified in Patient PC.0003.01.1 (A & B)....79
Figure 3.16: Electropherogram Confirming Variant ERCC3 c.958C>T in the Forward Strand of Patient PC.0003.01.1....79

Figure 3.17: Electropherogram Confirming Variant RB1 c.1981C>T in the Forward Strand of Patient PC.0018.01.1 82

List of Tables

Table 1.1: General CPS Indicators 9

Table 2.1: Variant Filtering Criteria 38

Table 3.1: Pre-Alignment Statistics Generated for the First and Second Sequencing Runs 46

Table 3.2: Post-Alignment Sequencing Quality Metrics Generated by the Ion Torrent coverageAnalysis and variantCaller Plugins 50

Table 3.3: Sequencing Quality Data for the Seven Prioritised Variants 55

Table 3.4: In-silico Pathogenicity & Conservation Prediction Tool Results for the Prioritised Indel & MNV Variants 55

Table 3.5: In-silico Pathogenicity & Conservation Prediction Tool Results for the Prioritised SNV Variants 56

Table 3.6: ACMG-AMP Rules Applied to Variant NF1_p.D378Afs*8 61

Table 3.7: ACMG-AMP Rules Applied to Variant RET_p.C634Y 64

Table 3.8: ACMG-AMP Rules Applied to Variant TP53_p.R273C 69

Table 3.9: ACMG-AMP Rules Applied to Variant BRCA1_p.M1400T 73

Table 3.10: ACMG-AMP Rules Applied to Variant FAH_p.R162H 77

Table 3.11: ACMG-AMP Rules Applied to Variant ERCC3_p.Q320* 80

Table 3.12: ACMG-AMP Rules Applied to Variant RB1_p.R661W 83

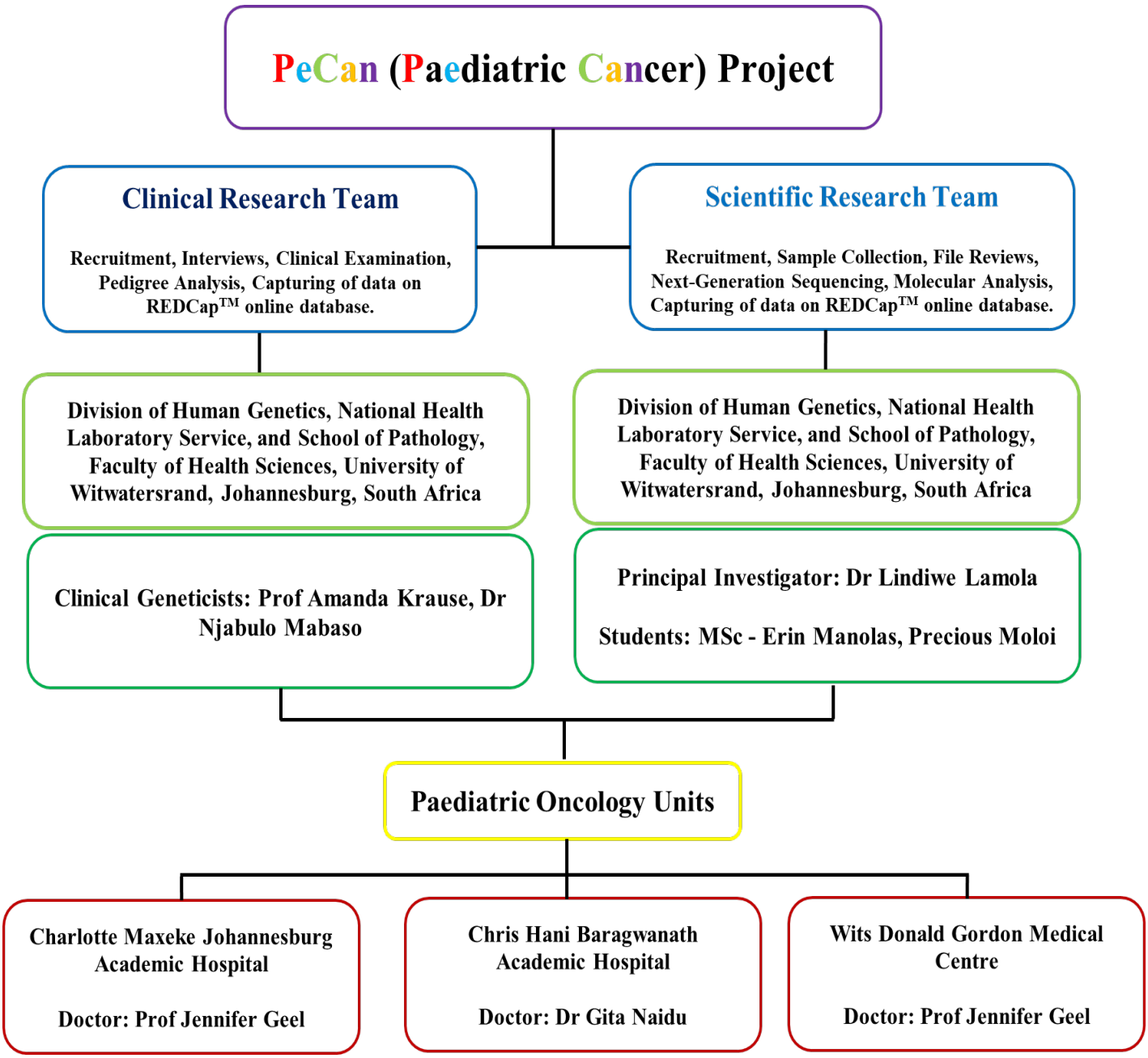
Nomenclature (List of Abbreviations, Acronyms and Symbols)

A, C, G, T	Adenine, Cytosine, Guanine, Thymine
ACMG – AMP	American College of Medical Genetics and Genomics and the Association for Molecular Pathology
BAM	Binary Alignment Map (file)
BED / .bed	Browser Extensible Data
bp	base pair
CADD	Combined Annotation Dependent Deletion
CCDS	Consensus Coding Sequence Set
CNS	Central Nervous System
CPGs	Cancer Predisposition Gene(s)
CPSs	Cancer Predisposition Syndrome(s)
dbNSFP	database for Non-Synonymous SNP's Functional Prediction
ddH_2O	double-distilled H_2O
DNA	Deoxyribonucleic Acid
dNTP	deoxyribonucleotide triphosphate
ddNTP	di-deoxynucleotide triphosphate
ds	double-stranded
EDTA	Ethylenediaminetetraacetic acid
ERMS	Embryonal Rhabdomyosarcoma
ExAC	Exome Aggregation Consortium
FMTC	Familial Medullary Thyroid Cancer
GERP RS	Genomic Evolutionary Rate Profiling Rejected Substitutions
gnomAD	Genome Aggregation Database
GRCh37	Genome Research Consortium Human Build 37
HBOC	Hereditary Breast & Ovarian Cancer (syndrome)
HGMD	Human Gene Mutation Database
HICs	High-Income Countries
HS	High Sensitivity
HT1	Hereditary Tyrosinemia Type 1
ID	Identification (number)
Indels	Insertions & Deletions
ISPs	Ion Sphere Particles
Kb	Kilobases
LFS	Li-Fraumeni Syndrome
LMICs	Low-Middle Income Countries

LOF	Loss-Of-Function
MAF	Minor Allele Frequency
MEN2A/B	Multiple Endocrine Neoplasia Type 2A/B
MLPA	Multiplex Ligation-Dependent Probe Amplification
MNV(s)	Multi-Nucleotide Variant(s)
MMR	Mismatch Repair
MTC	Medullary Thyroid Cancer
NCBI	National Center for Biotechnology Information
NCR(s)	National Cancer Registry/Registries
NGS	Next-Generation Sequencing
NHLBI GO ESP6500	National Heart, Lung and Blood Institute Grand Opportunity Exome Sequencing Project 6500
NHLS	National Health Laboratory Services
NF1	Neurofibromatosis Type 1
o/e	observed/expected (score)
PeCan	Paediatric Cancer Project
PhyloP	Phylogenetic P-value
P/LP	Pathogenic/Likely Pathogenic
PCR	Polymerase Chain Reaction
qPCR	quantitative Polymerase Chain Reaction
REDCap	Research Electronic Database Capture
SA	South Africa
SBIMB	Sydney Brenner Institute for Molecular Biosciences
SIFT-INDEL	Sorting Intolerant From Tolerant – Insertions & Deletions
SNV(s)	Single-Nucleotide Variant(s)
TF	Test Fragment
TMAP	Torrent Mapping Alignment Program
TSG	Tumour-Suppressor Gene
UniProt	Universal Protein Resource
UV	Ultra-Violet
VCF	Variant Call Format
VUS	Variant of Uncertain Significance
VEP	Variant Effect Predictor
WES	Whole-Exome Sequencing
Wits	University of the Witwatersrand
WGS	Whole-Genome Sequencing
XP	Xeroderma Pigmentosum

Preface

The Paediatric Cancer (PeCan) Project was formed in 2018 with the primary aim of investigating the genetic factors associated with paediatric cancers within a South African context. The PeCan project and all sub-studies which contribute to the aims and objectives of the larger project are based in the Division of Human Genetics, National Health Laboratory Services, and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand. Collaborators consist of a multi-disciplinary team of clinicians and scientists within the Division and three paediatric oncology units (see figure below). The figure below depicts the PeCan project team during the period of recruitment for the current project and has since been updated from 2022 onwards.



1. Introduction

Cancer can be defined as the evolution of healthy cells to a malignant neoplastic state (Hanahan & Weinberg, 2011). The role that our genome plays in cancer development cannot be understated - as cancer is essentially a genetic disease. Almost every step that is involved in the progression of a normal cell to a malignant one is the result of altered gene expression responsible for the proliferation, growth, differentiation and death of healthy cells. These genetic alterations, resulting in the formation of malignant cells are generally acquired, but, in rare cases, may be inherited (Cheetham et al., 2013).

1.1 Childhood vs Adult Cancers

Of the global cancer burden, childhood cancers account for less than 10% of all cancers reported (Ferlay et al., 2015). The frequent occurrence of cancer development in adulthood can be largely explained by the nature of cell biology and DNA replication. To maintain tissue homeostasis, cells undergo multiple divisions within a highly regulated cell cycle process, thereby maintaining cell growth and proliferation within normal limits, and allowing DNA synthesis to occur without error (Vassilev & DePamphilis, 2017). However, as the number of cell divisions increase exponentially with age so does the tendency for genetic aberrations to occur, and it is the accumulation of such genetic aberrations no longer repaired or tolerated by the genome which may result in the initiation of carcinogenesis (Vassilev & DePamphilis, 2017). This phenomenon was seen in a study by Gröbner *et al* (2018), which found that adult cancers had a mutation frequency 14 times higher than the investigated childhood tumours. The study also found that childhood cancers were more likely to have single putative driver gene mutations in comparison to adult cancers which were shown to be frequently driven by co-mutations (Gröbner et al., 2018).

In addition to the cellular ageing process, it is also important to consider the roles that environmental factors such as cumulative exposure to synthetic chemicals and pollutants, and lifestyle factors including cigarette smoking, alcohol consumption, obesity and poor nutrition may have in carcinogenesis (Madia et al., 2019; Jagai et al., 2017). Thus, it can be seen that the cellular ageing process and a cumulative exposure to extrinsic factors over time play a large role in the contribution towards cancer development in adulthood. However, without these cumulative and age-related risk factors at play in the development of childhood cancers, it leads to the question as to what role the genetic component, specifically the heritable genetic component, plays in the predisposition to cancer development in childhood. The need to investigate and fill this knowledge

gap is of utmost importance in countries such as South Africa where childhood cancer survival rates are low and may therefore benefit from implementing diagnostic measures which consider underlying genetic criteria (Stones et al., 2014).

1.2 Epidemiology of Childhood Cancers

Although only a small percentage of the global burden, the occurrence of childhood cancers has shown a gradual increase annually, with an expected 30% increase in new childhood cancer cases within this decade in low-middle income countries (LMICs), such as South Africa (Rodriguez-Galindo et al., 2015). A further increase in LMICs is of special concern as the under-5-year-mortality rate (U5MR) of 33 – 67 per million reported in LMICs is ~5x higher than the U5MR of 5 – 12 per million reported in high-income countries (HICs) (Ferlay et al., 2015; Howard et al., 2008). In South Africa, survival rates for paediatric cancer range from 46.4% - 63.9%, dependent on the type of cancer (Stones et al., 2014). This means that in South Africa ~50% of childhood cancer patients do not survive 5 years post-diagnosis, once again contrasting to survival rates above 80% in HICs, such as countries within Europe and the United States (Gatta et al., 2014; Stones et al., 2014; Ribeiro, 2012).

Although LMICs generally have higher relative proportions of children within their populations, childhood cancer incidence rates reported in HICs are generally higher (Fung et al., 2019). The incidence rate reported in HICs is ~130/million, in comparison to the incidence rate reported in LMICs of ~93/million, with specific rates of 33.4 – 47.2/million in Africa (Erdmann et al., 2015; Stones et al., 2014; Howard et al., 2008). The incidence rates within South Africa vary, with studies reporting rates between 45 – 100/million (Erdmann et al., 2015; Stones et al., 2014; Howard et al., 2008). There are several factors which may explain these statistics. The first factor to consider is the lack of any or adequate and efficient National Cancer Registries (NCRs) in LMICs, which may result in incomplete data capturing (Erdmann et al., 2015; Stefan, 2015). The South African NCR, as is the case with many NCRs in LMICs, has a pathology-based reporting system (Erdmann et al., 2015). This often results in the exclusion of cancer patients who are diagnosed through non-pathological based methods, such as patients presenting with brain tumours who are diagnosed through medical imaging, or patients who present at a late cancer stage and are not deemed necessary for pathological diagnosis. Another issue experienced by the South African NCR, further contributing to underreporting, was the voluntary nature of their submission system, whereby many private laboratories, mainly from 2006 onwards, decided to discontinue their voluntary reporting due to concerns about a lack of patient privacy (Erdmann et al., 2015).

The reporting of diagnosed cancer cases to the NCR by both private and public pathology services only became mandatory after the passing of new legislation in April 2011 (Erdmann et al., 2015).

Secondly, poor access to adequate healthcare systems is something that frequently occurs in LMICs (Erdmann et al., 2015; Stefan, 2015; Magrath et al., 2013). LMICs are largely dominated by rural, impoverished populations, where patients and their families have to travel long distances for treatment. This often increases the rate of abandonment of care as parents have to leave home for extended periods of time, affecting those they had to leave behind and losing essential income (Magrath et al., 2013). Thirdly, the standards of these healthcare systems are inadequate - having too few healthcare providers, at both the nursing and paediatric oncologist level and the lack of resources and technologies needed for effective diagnosis and treatment (Rodriguez-Galindo et al., 2015; Stefan, 2015; Erdmann et al., 2015). Lastly, due to a higher susceptibility to communicable diseases in LMICs, physicians may not consider a cancer diagnosis at first, as these other diseases hold a higher priority (Magrath et al., 2013). This often results in a delay in initial diagnosis and treatment and contributes to children presenting with an advanced stage of cancer where prognosis is poorer, once again contributing towards high mortality rates and the potential for underreporting (Gupta et al., 2015).

These factors which contribute to the high mortality rates together with the lack of knowledge on the actual incidence of childhood cancers in South Africa further emphasise the need to explore other diagnostic options. Investigating the inherited component of childhood cancers and the knowledge gained from this will contribute towards the improvement of diagnostic criteria and the development of research-specific treatment plans, ultimately improving current survival rates.

1.3 Genetic Predisposition to Childhood Cancer

The majority of childhood cancers develop as a result of an accumulation of genetic changes within somatic cells, i.e., somatic mutations (Carrasco Salas, Lapunzina & Pérez-Martínez, 2017). These somatic mutations are all *de novo*, meaning they occur for the first time in an individual, and are isolated to malignant cells (Stiller, 2004). However, cancer initiation may also be due to mutations acquired via germ cells, known as germline mutations. Germline mutations are acquired through genetic changes present in the parental DNA, which are then inherited and incorporated into the offspring DNA, or they occur as a result of a prezygotic *de novo* mutation event within the gametes and are present in every cell of the body (Carrasco Salas, Lapunzina & Pérez-Martínez, 2017; Davidoff, 2016).

A subset of genes are responsible for the maintenance of a healthy cell population through the highly regulated processes within the cell cycle. These processes include controlled cell proliferation through inhibited cell progression and apoptotic mechanisms, cell differentiation into specialised cell groups, and maintenance of genomic stability through effective DNA replication and repair (Carrasco Salas, Lapunzina & Pérez-Martínez, 2017; Davidoff, 2016). When genes involved in these processes are mutated, their altered functions may initiate carcinogenesis or promote the accumulation of carcinogenic genetic aberrations. Mutations in these genes increase the likelihood of cancer development and are thus known as Cancer Predisposition Genes (CPGs). CPGs can be defined as genes whose germline variants, whether inherited or *de novo,* increase an individual's risk of cancer development (Rasnic, Linial & Linial, 2020; Wang, 2016; Rahman, 2014b). A range of both monoallelic and biallelic CPG mutations have been reported to increase a person's likelihood of cancer development. This contrasts to the sporadic development of cancer, whereby two or more somatic mutations are needed to initiate carcinogenesis, thus carrying a lower risk than CPG mutation carriers (Wang, 2016). It is estimated that between 3-10% of all cancers worldwide are due to germline CPG mutations (Akhavanfard et al., 2020; Zhang et al., 2015; Rahman, 2014b).

1.3.1 History & Identification of Cancer Predisposition Genes

The discovery of a genetic predisposition to cancer development was initially reported over 100 years ago, through the detailing of the unusual clustering of cancers within families, by neuroanatomist Paul Broca (1866), pathologist Dr Alfred Warthin (1890 – 1910s) and zoologist and anatomist Theodor Boveri (1910s) (Wang, 2016; Rahman, 2014b). It would not be until nearly 50 years later that these findings would be echoed through the work of Dr H Lynch (1960 – 1990s) and Knudson's (1971) 'two-hit' theory based on the mathematical modelling of the epidemiology of retinoblastoma, which was later confirmed in 1987 with the discovery of the RB transcriptional corepressor 1 (*RB1*) predisposition gene (Wang, 2016; Rahman, 2014b; Lynch, Fusaro & Lynch, 1995; Knudson, 1971). Cancer predisposition research has greatly expanded in the last 30 years, largely due to the successful routine implementation of Genome-Wide Linkage Analysis studies (GWLAs) in the 1990s (Rahman, 2014b). Other strategies including candidate-based studies and more recently, targeted, Whole-Exome Sequencing (WES) and Whole-Genome Sequencing (WGS) Next-Generation Sequencing (NGS) techniques have significantly contributed towards the discovery of a new set of CPGs (Rahman, 2014b). In addition, Genome-Wide Association Studies (GWAS) have contributed to the discovery of many variants, mostly rare within unknown CPGs, which have been shown to have a small increase in the risk of cancer development (Rasnic, Linial & Linial, 2020; Chung & Chanock, 2011; Galvan, Ioannidis & Dragani, 2010). A list of 114 well-

known CPGs was compiled by Rahman (2014), who through an extensive literature and database review of over 30 years of research, identified CPGs with variants which confer a moderate or high risk of cancer development. This list was further extended to 152 CPGs in the largest variant predisposition study to date (Huang et al., 2018) through the addition of genes from St Jude's Paediatric Cancer Genome Project germline study (Zhang et al., 2015), the Catalogue of Somatic Mutations in Cancer (COSMIC) Cancer Gene Census (CGC) germline (https://cancer.sanger.ac.uk/census/) and other recent literature. Additionally, GWAS studies have identified 391 common variants which individually confer a small increase in the risk of cancer development (Hindorff et al., 2009).

The discovery of and further research into CPGs has provided invaluable knowledge to the scientific and medical communities. The investigation of cancer predisposition has resulted in gaining scientific insights into specific gene functions and mechanisms which subsequently result in cancer formation. The emergence of these known CPGs has paved the way for the identification of additional CPGs through their functional relationships to these well-known genes. This has expanded the knowledge of already established and now emerging functional networks involved in cancer predisposition and formation, many of which contain links to several CPGs (Rahman, 2014b; Sheppard et al., 2012). CPG research has also allowed an in-depth understanding of the mechanisms of the variants which they harbour and their association with specific phenotype development.

1.3.2 Mechanisms and Functions of Cancer Predisposition Genes

CPG variants that confer a moderate or high risk of cancer development are primarily inherited in an autosomal dominant pattern. Whilst some CPG variants only increase the risk of cancer in an autosomal recessive pattern, a smaller subset of CPG variants causes an increased cancer risk in cases of both monoallelic and biallelic inheritance, with the autosomal recessive condition often resulting in the development of a more severe phenotype than the dominant condition (Huang et al., 2018; Rahman, 2014b). Very few sex-linked CPGs with a moderate/large effect on cancer predisposition have been identified, with four X-linked and one Y-linked CPG reported (Huang et al., 2018; Rahman, 2014b).

The oncogenic mechanisms of CPGs can be broadly grouped into two categories – these are either the (1) inactivation or (2) activation of CPGs responsible for maintaining the balance between cell growth and proliferation and cell loss due to apoptosis (programmed cell death/senescence/differentiation) (Davidoff, 2016). – See Figure 1.1.

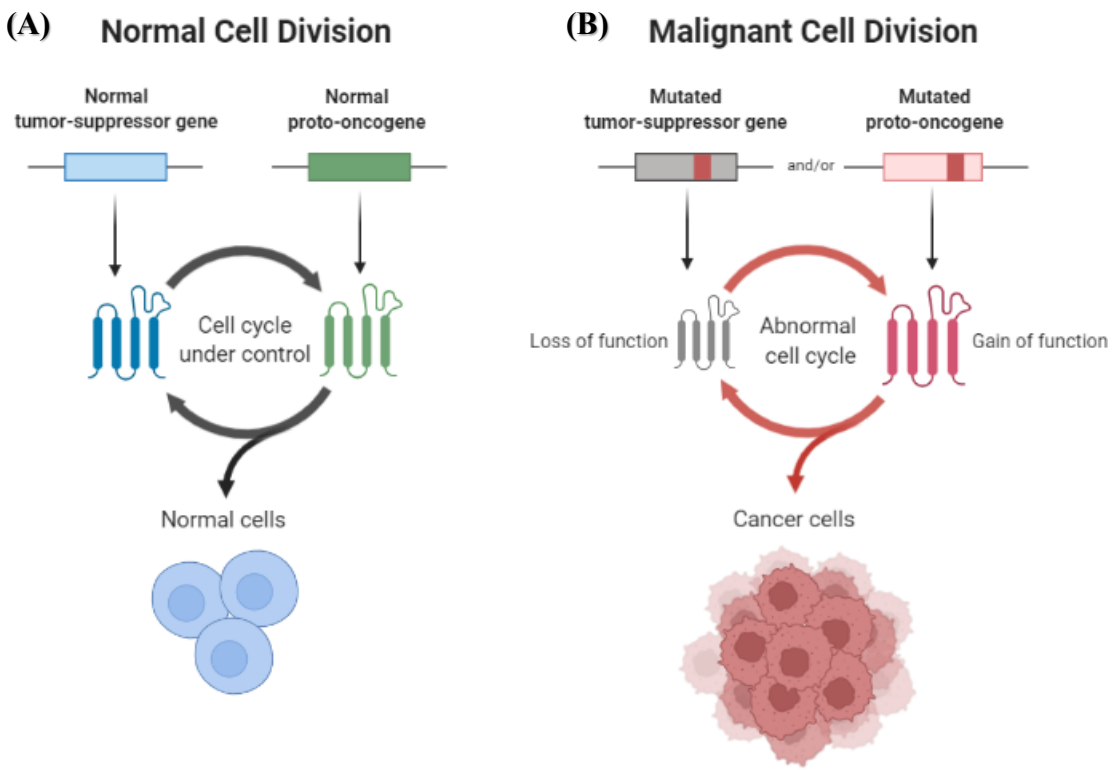


***Figure 1.1* – Graphical Illustration of the Two Broad Mechanisms of Oncogenesis, (1) Inactivation (Tumour – Suppressor Genes) & (2) Activation (Oncogenes): (A)** Image illustrates the role of tumour-suppressor (TSGs) and proto-oncogenes in normal cell division. TSGs produce negative growth signals, responsible for the restriction of cell growth and proliferation and the induction of apoptotic mechanisms. Proto-oncogenes produce positive growth signals when cell growth and proliferation are required. **(B)** Image illustrates the mechanisms by which TSGs and proto-oncogenes contribute to carcinogenesis. The oncogenic mechanism of TSGs was first described by Knudson's 'two-hit' hypothesis, which states that in order for an increased cancer risk and/or cancer development to occur, the first inherited germline mutation must be accompanied by a somatic 'hit'/mutation of the second allele. Loss of function of both alleles results in a complete loss in negative growth signals from that gene, allowing uncontrolled cell growth, proliferation, and evasion from apoptosis. Examples of paediatric-cancer TSGSs include the *RB1* & *TP53* genes. Proto-oncogenes when mutated by gain-of-function mutations become oncogenes and reside in a permanent state of positive signalling and are further unable to be inhibited by negative growth factors. This promotes uncontrolled cell growth and proliferation as well as an increased expression of genes which are necessary for the survival and replication of the malignant cell. Examples of common paediatric oncogenes include the *RET* & *ALK* genes. Legend references: (Carrasco Salas, Lapunzina & Pérez-Martínez, 2017; Davidoff, 2016; Ganjavi & Malkin, 2002; Knudson, 1971). Image designed by and obtained from "Tumour Suppressor and Proto-oncogenes", by BioRender.com (2023). Retrieved from https://app.biorender.com/biorender-templates.

Genes involved in the recognition of DNA damage and DNA repair pathways also play a key role in cancer predisposition and development. Like TSGs, the vast majority of DNA damage recognition and DNA repair genes require both alleles to be inactivated for a complete loss of function and subsequent increased risk of cancer development (Park & Vogelstein, 2003). However, unlike TSGs, DNA repair genes play a more indirect role in the regulation of cell growth

and proliferation (Carrasco Salas, Lapunzina & Pérez-Martínez, 2017). When DNA repair genes are inactivated, there is no direct increase in the cell's growth and proliferation, instead, there is an accumulation of genetic aberrations in the cell's DNA which is no longer repaired by the cell (Park & Vogelstein, 2003). This suggests that cancers which develop as a result of a DNA repair deficiency occur as a result of the accumulation of many errors in the genome and additional mutations in other key CPGs, rather than single point mutations within a DNA repair gene causing a cancer initiation event (Ganjavi & Malkin, 2002). There are over 100 DNA repair genes involved in several specialised DNA repair pathways which act to repair a multitude of genetic alterations which may occur (Knijnenburg et al., 2018; Friedberg et al., 2006). Thus, as there are many genes involved in several repair processes, there are many genes which may be inactivated due to mutations, which in turn may predispose individuals to a range of cancer types (Ganjavi & Malkin, 2002).

Although the activation of proto-oncogenes and the inactivation of TSGs and DNA repair genes account for the majority of cancer predisposition/development in childhood, there are other, rarer mechanisms of oncogenesis which occur. In some cases, the inactivation of a single allele results in haploinsufficiency, whereby the dosage of gene product produced by the other wild-type allele is approximately half of what is required for the gene to perform its complete function, thereby resulting in an abnormal phenotype (Cleary, 2001). Haploinsufficiency of the Nuclear receptor binding SET Domain protein 1 (*NSD1*) gene has been identified as the major cause of Sotos syndrome and more rarely Weaver syndrome (Tatton-Brown et al., 2005; Nagai et al., 2003). Mutations resulting in haploinsufficiency of the *NSD1* gene abrogate its epigenetic regulation of promoter regions involved in cell growth, thereby promoting excessive cell growth and malignant transformation (Weksberg et al., 2013). Sotos and Weaver's syndromes are classified as genetic childhood overgrowth disorders, whereby in addition to other phenotypic features, individuals are at an increased risk of developing several types of malignancies, the majority of which develop in childhood (Villani et al., 2017a; Douglas et al., 2003).

In another oncogenic mechanism, the mutation of a single allele results in the expression of a functionally altered gene product, one whose expression disrupts the functioning of the wild-type allele. This mechanism was first described as dominant-negative and occurs when the mutant gene product is capable of interacting with the same molecular complexes as the wild-type gene product (Veitia, 2007; Herskowitz, 1987). A dominant-negative mechanism of the *WT1* gene is seen in the development of Denys-Drash syndrome, characterised by abnormal kidney function, disorders of sexual development in majority males, and an increased risk of developing Wilms' tumour, a kidney cancer commonly developed in childhood (Achermann & Hughes, 2016). The WT1 gene

primarily acts as a TSG, by encoding a zinc finger transcription factor which acts to suppress the expression of growth-promoting genes and induce the expression of pro-apoptotic factors (Tobias, 2013). In cases of Denys-Drash syndrome, the mutant gene product exerts its dominant-negative effect by blocking the binding of the wild-type gene product to its target genomic regions, thereby promoting malignant transformation through increased expression of growth-promoting genes and a decreased expression of pro-apoptotic factors (Achermann & Hughes, 2016; Tobias, 2013).

Research has shown that some CPGs do not fit under a single functional label as a TSG or an oncogene, and thus the mutation of these genes may differ in their contribution towards oncogenesis. Key examples of this include the *TP53* and *WT1* genes, which are primarily known as TSGs for their role as a negative regulator of cell growth, and the subsequent loss of this function contributing towards cancer predisposition/development. However, mutations in these genes may not only result in the loss of their growth-inhibiting functions but may also exert a dominant-negative effect on the wild-type p53 protein and/or gain growth-promoting oncogenic functions, independent of the wild-type function (Achermann & Hughes, 2016; Tobias, 2013; Rivlin et al., 2011).

1.3.3 Familial/Hereditary Predisposition Cancer Syndromes

Familial cancer- and hereditary cancer-syndromes are inherited disorders characterised by an increased likelihood of cancer development (Rahner & Steinke, 2008; Stiller, 2004). The terms hereditary cancer- and familial cancer-syndromes are often used interchangeably but may also differ depending on the context. Familial cancer syndromes often refer to the development of specific and frequently common cancer types in several family members, more than is expected by chance. While familial cancers are indicative of an underlying inherited variant, the genetic causation is not always known and shared environmental and/or lifestyle factors may contribute to cancer onset and development (National Cancer Institute, 2019). Hereditary cancer syndromes often refer to the development of cancer where the causation is solely genetic and the clustering of cancer in a family is due to the inheritance of these genetic aberrations (National Cancer Institute, 2021). Familial cancer- and hereditary cancer-syndromes will collectively be referred to as Cancer Predisposing Syndromes (CPSs).

While there are numerous published guidelines and tools developed to assist clinicians in the identification and diagnosis of individuals and/or families with one of the >50 identified CPSs, there are also ‘general’ indicators which may suggest a hereditary susceptibility to cancer (Brown et al., 2020; Lindor et al., 2008). See Table 1.1 for a list of general CPS indicators in both the proband and close relatives of the proband, which may assist clinicians in identifying a hereditary

susceptibility to cancer. It is important to note that the lack of any of these general indicators does not necessarily suggest the lack of a CPS. For example, a negative family history of cancer does not necessarily suggest the lack of a CPS diagnosis. Family history may be inaccurate, non-informative or completely absent. Factors including small family size, modes of inheritance and penetrance of the underlying genetic aberration must be considered. There can be great variability in the degree of penetrance of the germline variants involved in CPSs, with very few showing 100% penetrance (Saletta, Dalla Pozza & Byrne, 2015; Sijmons, 2010). There may also be the possibility that the proband has a *de novo* germline variant, thus being the first family member to be diagnosed with a CPS. *De novo* germline variants associated with the onset of a CPS have been reported in cases of Familial Adenomatous Polyposis (FAP), Multiple Endocrine Neoplasia Type 2B (MEN2B), and between 7-25% of *TP53* variants causing Li-Fraumeni Syndrome (LFS) (Sijmons, 2010; Gonzalez et al., 2009a; Chompret et al., 2000).

Table 1.1 – General CPS Indicators: Table describing the general indicators in both the proband and close relatives which may indicate the presence of a Cancer-Predisposition Syndrome (CPS).

Individual(s) Being Assessed	**CPS Indicator**	**Example**	**Syndromes Associated with an example given**
Proband	Development of Rare Neoplasms	Development of Wilms' Tumour	BWS, Familial Wilms' Tumour, Frasier Syndrome, Perlman Syndrome, WAGR syndrome
	Unusually Early Age of Onset	Development of rare cancers in early childhood, such as Adrenocortical Carcinoma, MTC.	LFS (Adrenocortical carcinoma), MEN2A/MEN2B/FMTC (MTC)
	Presence of Bilateral/Multifocal Tumours	Development of Bilateral Retinoblastoma	Familial Retinoblastoma
	Metachronous and/or Synchronous Development of Multiple Tumours	Breast and Ovarian Cancer	HBOC
		Colon and Endometrial Cancer	Lynch Syndrome
		Colorectal Cancer and Brain Tumour (Generally Medulloblastomas/Glioblastomas)	Turcot Syndrome
Close Relatives of the Proband	Positive Family History of Cancer	• Same type of cancer in several family members, especially when occurring on the same side of the family • Several family members with different cancer types, occurring in combinations associated with a particular CPS. • Family members diagnosed with cancer at a relatively young age. • Diagnosis of cancer within a 1st – degree relationship, e.g., Parent/Child/Sibling	Examples of CPSs often associated with a positive family history of cancer include: Hereditary/Familial Retinoblastoma, Wilms', Neuroblastoma, GISTs. HBOC, Lynch, LFS, RTPS, NF1/NF2, MEN1/MEN2/FMTC

Abbreviations/Acronyms: BWS - Beckwith-Wiedemann Syndrome; FMTC – Familial Medullary Thyroid Cancer; GISTs - Gastrointestinal Stromal Tumours; HBOC - Hereditary Breast & Ovarian Cancer; LFS – Li-Fraumeni Syndrome; MEN – Multiple Endocrine Neoplasia; NF- Neurofibromatosis; RTPS – Rhabdoid Tumour Predisposition Syndrome; WAGR – Wilms' Tumour, Aniridia, Genital-urinary anomalies & Retardation syndrome. Table 1.1 References (Carrasco Salas, Lapunzina & Pérez-Martínez, 2017; Tobias, 2016; Sijmons, 2010; Lindor et al., 2008).

Some CPSs are defined by the occurrence of cancer alone and thus may be difficult to identify due to the lack of a distinctive non-malignant phenotype, which is often used as an indicator for further genetic investigation. These CPSs may be defined by the presence of the same cancer type in several family members and confirmed through molecular testing for mutations in known CPGs. Examples include Hereditary/Familial Retinoblastoma, Neuroblastoma, Wilms', and Gastro-Intestinal Stromal Tumours (GISTs) (Lindor et al., 2008). Other CPSs can be defined by the presence of two or more family members diagnosed with genetically related cancers or different cancer types occurring within a particular CPS tumour spectrum. Examples of CPSs associated with the predisposition to more than one cancer type include HBOC, Lynch Syndrome, LFS, RTPS and Turcot Syndrome (Lindor et al., 2008; Nagy, Sweet & Eng, 2004).

There are also CPSs which are characterised by an increased risk of the development of both benign and malignant neoplasms, such as FAP and BAP1-Tumour Predisposition Syndrome (BAP1-TPDS). FAP, which occurs as a result of mutations in the *APC* gene, is clinically diagnosed by the presence of approximately 100 or more adenomatous colorectal polyps, which are benign growths that develop on the mucosal membrane of the large intestine. If left untreated, nearly 100% of individuals with FAP will develop colon adenocarcinoma and also have an increased risk of developing other cancer types (Stec et al., 2010). Individuals with BAP1-TPDS as a result of pathogenic germline variants in the BRCA1-Associated Protein 1 (*BAP1*) gene, are at an increased risk of developing both benign and malignant tumours, most commonly in the kidneys, eyes, skin and of the mesothelium. The development of malignant tumours tends to occur at a younger age in comparison to the general population and also tends to be more aggressive and more likely to metastasize. Individuals may develop one or more types of tumours and affected family members may have different cancer types. The most common (generally) benign tumours are BAP1-inactivated melanocytic tumours and the most common cancer reported is uveal melanoma (Pilarski et al., 2016).

Other CPSs are characterised by distinct non-malignant phenotypes, and although they are associated with an increased risk of cancer development, they may also be independent of cancers. There is an abundance of non-malignant anomalies which are associated with CPSs, and thus the identification and interpretation of these anomalies are essential in making an accurate CPS diagnosis and cancer risk estimation. However, there is also significant phenotypic overlap between some of the CPSs and the presence of more common structural and/or functional abnormalities which may aid in the identification of an underlying CPS. Examples of more common non-malignant phenotypic features associated with several CPSs include: congenital anomalies, excessive somatic and/or visceral growth, growth deficiency, developmental delay,

anomalies of the integumentary system, characteristic facies, intellectual disability, and neurological/motor/co-ordination impairments (Manor & Lalani, 2020; Pierpont et al., 2014; Sijmons, 2010; Lindor et al., 2008).

1.3.4 Cancer Predisposition Genes in a Clinical Setting

Cancer Predisposition Genes in Diagnostics

The knowledge gained from CPG research has built the foundation for the implementation of CPG testing in the diagnosis of patients and their affected family members. Determining that cancer causation is due to an underlying germline variant within a CPG provides invaluable insights and has become standard for particular cancer types in some public and private genetic testing facilities. This is often the case in individuals who meet a certain set of criteria for hereditary *BRCA1* and *BRCA2* testing, whereby germline variants in these CPGs carry a high risk for breast and ovarian malignancy (Couch et al., 2017; Petrucelli, Daly & Pal, 2016). Previously, CPG testing was based on single-gene methods or hot-spots within these single CPGs. The development of these CPG tests was an expensive and time-consuming endeavour and was often developed in research-based settings with longer time-frames than those required in a diagnostic workflow (Rahman, 2014a). Due to the costly and laborious nature of previous methods, CPG testing often involved testing smaller groups who met a stricter set of criteria. These criteria often included testing individuals with evidence of the presence of a familial cancer syndrome only, whereby clinicians would refer for CPG testing with the presentation of certain tumour and other syndromic phenotypic features (Whitworth et al., 2018; Rahman, 2014a). The introduction of NGS technologies to CPG testing, through the development of targeted CPG panels, WES and WGS has facilitated massively parallel CPG testing. The implementation of NGS in genetic testing has shown to be a more time- and cost-efficient process, allowing more individuals to undergo multiple-gene testing under a more relaxed set of criteria (Biesecker, 2012; Green & Guyer, 2011). Multi-gene testing can be particularly useful if the predisposition and development of cancer is a heterogeneous process whereby it is more likely to identify causative variants through large gene set testing methods (Velázquez et al., 2020; Rahman, 2014a).

Cancer Predisposition Genes in Cancer Treatment & Prognosis

CPG research has provided insights into gene-specific mechanisms of action in cancer predisposition and development, some of which have been utilised in the development of research-specific treatment plans. This has provided clinicians with valuable information in cancer

management decisions such as whether to opt for conservative or radical surgery as well as whether to alter chemotherapy and/or radiotherapy regimens (Rahman, 2014a). For instance, individuals with germline *BRCA* variants, such as those with Hereditary Breast and Ovarian Cancer (HBOC) syndrome, may benefit from platinum-based and Poly-ADP Ribose Polymerase (PARP) inhibitor therapies (Sæther et al., 2018; Liu, Konstantinopoulos & Matulonis, 2014). This is because platinum-based neoadjuvant chemotherapy and PARP-inhibitor therapies utilise the faulty DNA repair mechanism of *BRCA* variant carriers. Although platinum-based and PARP-inhibitor therapies utilise the faulty repair mechanism of *BRCA* carriers differently, both therapies ultimately result in the trigger of apoptotic mechanisms due to unrepaired DNA damage within cells (Sæther et al., 2018; Liu, Konstantinopoulos & Matulonis, 2014).

The detection of underlying CPG variants has also been utilised as a valuable biomarker in predicting prognosis (Wang, 2016; Rahman, 2014b). An important prognostic factor which CPG detection can provide insight into is the likelihood of survival for CPG germline carriers who go on to develop CPG-related cancers. This is seen in *BRCA* carriers who are at risk of developing epithelial ovarian cancer (EOC). Both *BRCA1* and *BRCA2* genes are associated with the predisposition to and development of EOC. However, determining whether EOC development is due to either *BRCA1* or *BRCA2* germline variants may be of prognostic importance as overall survival and progression-free survival has reported to be significantly shorter in *BRCA1*- than in comparison to *BRCA2*-associated EOC patients (Vencken et al., 2013).

Another prognostic factor which CPG testing can provide valuable insights into is the possibility of developing a second primary malignancy. The development of multiple primary tumours at different sites is a strong indicator that an underlying germline variant within a CPG may be at play and the confirmation of such can provide insights into the range of cancer types the individual is predisposed to developing (Travis et al., 2006; Garber & Offit, 2005). This is especially true in cases of hereditary/familial cancer syndromes, such as LFS, hereditary breast and colon cancers and familial retinoblastoma, whereby patients and affected family members are at an increased risk of developing multiple primary and secondary cancers throughout life (Gonzalez et al., 2009b; Travis et al., 2006; Kleinerman et al., 2005). Some CPG carriers, such as *RB1* germline carriers, may be at a further increased risk of developing secondary malignancies, especially osteosarcomas, due to the receival of radiation therapy, often developing tumours in prior radiation fields (Ng et al., 2010; Marees et al., 2008; Kleinerman et al., 2005).

Cancer Predisposition Genes in Cancer Surveillance, Risk-Reduction & Prevention

Through CPG research, both malignant and non-malignant phenotypic, clinical, and family history criteria for the genetic testing of variants within CPGs and their associated hereditary cancer syndromes have been outlined and continue to be expanded on (Daly et al., 2021; LaDuca et al., 2020; National Comprehensive Cancer Network, 2019). These criteria allow for individuals and their close relatives with certain phenotypic characteristics or whose cancer type meets specific clinical criteria to be identified and screened for the presence of putative disease-causing/contributing germline CPG variants. The subsequent identification of germline CPG variants facilitates the implementation of tailored or targeted therapies, where available, for the affected cancer patient and appropriate surveillance measures for affected relatives (Rahman, 2014b). For mutation-positive CPG individuals, clinicians can decide on the appropriate screening, risk-reducing and cancer prevention measures to implement, dependent on the type of cancer and germline variant.

Screening measures often involve imaging techniques, biopsies and/or the detection of biochemical markers through blood analysis. These techniques aid in the identification of lesions and/or altered biochemical compounds which may indicate carcinogenesis before it presents clinically, the aim of which is for early cancer detection and improved prognosis (Greer, 2018; Wang, 2016; Rahman, 2014b). An example of effective surveillance is in individuals and family members diagnosed with Lynch syndrome who benefit from regular colonoscopies and transvaginal/transabdominal ultrasounds for the detection of colorectal cancers and endometrial hyperplasia/early-stage endometrial cancers, respectively (Wang, 2016; Vasen et al., 2013). Similarly, individuals and close relatives may also benefit from regular non-invasive surveillance, such as those diagnosed with MEN2 carrying variants in the *RET* gene, where annual blood and/or urine calcitonin and catecholamine tests are recommended for the detection of MTC and pheochromocytomas, respectively (Rahman, 2014b; Rich, Jimenez & Evans, 2010).

Patients found to have germline CPG variants may also benefit from risk-reducing measures, dependent on the cancer type they are predisposed to. Individuals with an increased risk of developing breast, pancreatic and colon cancers may reduce this risk by altering modifiable lifestyle factors, including levels of physical activity, tobacco smoking and excessive alcohol consumption (Basen-Engquist et al., 2020; Gapstur et al., 2018). Chemopreventative measures through the usage of certain drugs, vitamins and/or supplements have shown promising results in MMR germline variant carriers where daily aspirin intake significantly reduces the risk of colon cancer (Rahman, 2014b).

More drastic cancer prevention methods such as the partial or complete removal of organs are generally reserved for individuals whose CPG variant has full or near full penetrance and whose risk for developing cancer in that organ is greatly increased. Prophylactic surgery of this nature is also only considered in non-essential organs whose removal is associated with low morbidity and in organs where replacement therapy is effective for any functional loss which may occur (Rahman, 2014b; Wells et al., 2013). For example, individuals and close relatives diagnosed with MEN2A/2B and FMTC carrying high-risk *RET* variants meet these criteria for prophylactic thyroidectomy (Wells et al., 2013). Several guidelines have been published with recommendations for the age at which thyroidectomy should be performed and are largely developed according to the type and location of the *RET* variant (Tuttle et al., 2010; American Thyroid Association Guidelines Task Force et al., 2009).

1.3.5 The Complexities of Cancer Predisposition

As our knowledge of cancer predisposition continues to expand, so does the realisation of the previously unknown complexities of the disease. It is estimated that between 3-10% of all cancers worldwide are due to heritable CPG variants (Economopoulou, Dimitriadis & Psyrri, 2015; Grant et al., 2015; Lu et al., 2014). Although, this figure is likely an underestimate with more recent NGS datasets indicating that up to 17.5% of cancer patients may possess a clinically actionable CPG variant (Mandelker et al., 2017). One reason why this underestimation may exist is the poor characterisation of the contribution of known genes to the predisposition and development of specific cancer types. There lies a great complexity in the variability in which CPGs contribute to cancer predisposition (Rahman, 2014b). Some CPGs are well known for their contribution towards the development of individual cancer types. Such is the case with retinoblastomas and MTCs in childhood which can be highly attributed towards germline variants in the *RB1* and *RET* genes respectively (Paulson, Rudzinski & Hawkins, 2019; Stark, 2016). This contrasts with another childhood tumour, Wilms' tumour which has been associated with >50 clinical conditions and ~20 common CPGs. Further, Wilms' tumours are primarily sporadic and in many conditions are reported as a rare co-occurrence (Hohenstein, Pritchard-Jones & Charlton, 2015; Scott et al., 2006).

On the other side of the spectrum, there are also single CPGs which predispose individuals to many different cancer types, as well as different risks associated with each cancer type depending on the type of variant. This is observed in individuals with LFS, where different *TP53* variants are associated with various cancer types and risks. *TP53* variants which result in non-functional or truncated proteins are commonly associated with early-onset brain tumours, missense *TP53*

variants in the DNA binding domain are frequently seen in the development of breast cancers, and *TP53* variants outside the DNA-binding loop are specifically associated with adrenocortical tumours (Malkin, 2011; Figueiredo et al., 2006; Olivier et al., 2003). The variability between genotypes and cancer phenotypes may reflect the presence of additional modifying factors. Studies have identified both genetic and non-genetic modifying factors whose gene-gene interactions and regulation of gene expression may explain some of the phenotypic variability observed and act in regulating the penetrance of the mutant allele (Costa et al., 2019; Talseth-Palmer et al., 2013). The complexity of cancer predisposition has rendered cancer risk estimation challenging. This has highlighted the importance of the role genetic counsellors play in providing those with an increased risk of cancer development an expert heritable cancer risk assessment, as well as any risks, benefits and limitations associated with genetic testing (Druker et al., 2017; Wang, 2016).

1.4 Next-Generation Sequencing in a Cancer Research & Diagnostic Setting

Historically, the identification of CPGs and their underlying variants in a research setting was largely based on GWLAS and GWAS designs in large families with high penetrance (Rasnic, Linial & Linial, 2020; Rahman, 2014b; Chung & Chanock, 2011; Galvan, Ioannidis & Dragani, 2010). However, it was only within the last decade or so that Next-Generation Sequencing (NGS) technologies, a high-throughput, massively parallel and deep sequencing technique, has been implemented and made significant contributions to CPG research (Rahman, 2014b; Mardis, 2008).

Before the introduction of NGS techniques, molecular CPG variant detection in a diagnostic setting involved the development of single candidate gene-based tests (Rahman, 2014a). This involved using techniques such as Restriction-Fragment Length Polymorphism (RFLP), Multiplex Ligation-Dependent Probe Amplification (MLPA), Real-Time Quantitative PCR, and Sanger sequencing (Mahdieh & Rabbani, 2013; Sanger, Nicklen & Coulson, 1977). Sanger sequencing, known as the gold standard for variant detection in molecular diagnostics, proved to be an accurate method of examining smaller fragments in single-gene tests. However, the use of single gene-based approaches, such as Sanger sequencing, in investigating and obtaining a diagnosis for heterogeneous conditions can prove to be costly, time-consuming, and inconclusive. Thus, the implementation of NGS technologies, capable of sequencing and analysing multiple gene targets in several patients simultaneously, became a promising concept within the diagnostic setting.

The transition and successful implementation of NGS technologies from a research to a diagnostic environment in developing LMICs, like South Africa, is significantly behind that of well-resourced HICs (Helmy, Awad & Mosa, 2016; Yohe et al., 2015). However, the recent decrease

in the cost of next-generation technologies provides the opportunity to catalyse the implementation of molecular diagnostics in the SA setting. NGS technologies are well suited for allowing a comprehensive understanding of the genomic variants associated with cancer development (Kuhlen et al., 2018). The ability of NGS technologies to sequence multiple targets in a massively parallel manner makes it particularly useful for studying the heterogeneous and complex nature of cancer predisposition in childhood. It is capable of, according to the study design, screening multiple genes associated with several paediatric cancer subtypes simultaneously, and thus detecting multiple potential underlying variants. NGS bioinformatics tools facilitate the identification and analysis of variants' effects on gene function and thus provide insights into the variants' potential role in cancer pathogenesis. The variant evidence provided by these bioinformatic tools can be interpreted, weighted, and further classified, which aids in distinguishing between variants that are likely common polymorphisms and variants which are more likely to have a deleterious effect on gene function and thus play a larger role in cancer development. The classification and further investigation of Variants of Uncertain Significance (VUS) are of particular importance in examining genes for the presence of unique variants found in different populations, especially when examining genomic samples of understudied African origin.

Currently, in South Africa, there is limited to no routine diagnostic testing in place for the detection of germline variants in state paediatric cancer patients. Where testing does occur, the cases are generally private/referred and occur more often in adult patients and extended family members where a founder variant is suspected and is more routinely tested. There are also several genetic and epidemiologic factors on paediatric cancer in South Africa that are currently unknown. This includes updated statistics on the prevalence of different paediatric cancer types, the proportion of these cancers known to harbour disease-causing germline variants and the prevalence of CPSs within the paediatric South African population. These statistics will prove to be useful in determining which genes should be included in a NGS gene panel design, what cancer types should be tested, and what malignant and non-malignant testing criteria should be established.

It is also important to consider what type of diagnostic test would be most useful and effective. As mentioned, NGS technologies are well suited for investigating the complex and heterogeneous nature of cancer predisposition. Although the initial installation of NGS equipment and computational facilities may be expensive, once it is adapted to a diagnostic environment and used more routinely on a larger scale, it may prove to be more cost-effective compared to single-gene testing methods. NGS technologies can be utilised in many ways for the detection of pathogenic germline variants in paediatric cancer patients. This most commonly includes WGS, WES and

targeted sequencing (also known as a gene panel). As paediatric cancer-predisposition research is still within its infancy in SA, it may be better suited to focus on the well-known, moderate to high penetrance CPGs through a targeted gene panel approach before expanding into the whole exome and/or whole genome territory. Additionally, the higher costs, extensive analysis requirements, and longer turn-around times associated with WES/WGS may not be currently viable for implementation within a LMIC diagnostic setting. Thus, the use of a candidate CPG panel may be best suited for the current investigation of cancer predisposition in a sample of South African children. A small candidate CPG panel design and small sample size may go a long way to add knowledge that is currently lacking, guide future directions and ultimately lay the foundation for the implementation of a successful diagnostic system.

1.4.1 Next-Generation Sequencing in Paediatric Cancer Genetics

Several international (mainly United States and European Union populations) paediatric cancer genetics studies have utilised NGS technologies to determine the occurrence and contribution of germline variants in paediatric cancer development (Kim et al., 2021a, 2021b; von Stedingk et al., 2021; Wagener et al., 2021; Akhavanfard et al., 2020; Byrjalsen et al., 2020; Mirabello et al., 2020; Diets et al., 2018; Gröbner et al., 2018; Kline et al., 2017; Chang et al., 2016; Oberg et al., 2016; Parsons et al., 2016; Worst et al., 2016; Mody et al., 2015; Zhang et al., 2015). Collectively, these studies reported that 3.8% - 35.5% (within 2015-2021 time period) of their paediatric patients had pathogenic or likely pathogenic germline variants in CPGs. The majority of these studies have investigated a wide range of both haematological and solid tumour cancers. Parsons *et al* (2016) and Akhavanfard *et al* (2020) are the only studies to date that have used an NGS approach, specifically WES, to study the occurrence of germline variants in paediatric patients diagnosed with a wide range of solid tumour cancers only.

The results of these studies have shown the effectiveness of using NGS and bioinformatic techniques in the study of paediatric cancer genetics. For example, Parsons *et al* (2016), utilising a WES approach, studied a broad range of both central nervous system (CNS) and non-CNS malignancies and found that 10% of their cohort had germline variants that related to the patient's phenotype. The capability of using variant data produced by NGS and relating it to the patient's phenotype is of key importance in identifying clinical characteristics that point to underlying genetic causes. In fact, NGS studies have shown that the patient's phenotypic presentation is often more extensive than the established diagnostic criteria (Diets et al., 2018; Parsons et al., 2016). Thus, highlighting a need to improve the criteria for the identification of patients eligible for

genetic testing. This is especially important in LMICs where the screening of all patients is not possible (Diets et al., 2018).

These international cancer genetics studies have also shown the ability to develop a cost-effective diagnostic workflow using NGS technologies. For example, a study by Sahm *et al* (2016) developed a low genomic input, five-day NGS (targeted gene-panel) diagnostic workflow, capable of simultaneously analysing a large number of molecular markers within 130 genes for the diagnosis of 150 brain tumours. This efficient and cost-effective NGS diagnostic workflow was additionally capable of using a low genomic input (i.e., low genomic DNA concentration) successfully, thereby highlighting another advantage of using NGS in this field, as the ability to extract an adequate quantity of biological sample from a paediatric cancer patient may not be possible, due to disease progression or receipt of chemotherapy. In comparison to conventional variant detection techniques, the NGS workflows in these studies have shown the capability of NGS techniques in detecting previously uncharacterised variants, as well as variants that may be clinically relevant (Gröbner et al., 2018; Sahm et al., 2016; Worst et al., 2016).

There are very few African studies that have utilised NGS technologies in the research of paediatric cancer genetics. A recent systematic review by the PeCan Project (unpublished work) which looked at over 30 years (1980 – 2019) of paediatric cancer genetic studies within Africa, revealed that only ~5.6% (3/54) of studies used NGS in various ways, including WGS (for CNV evaluation), WES, targeted sequencing, and RNA-sequencing. These studies investigated Wilms' tumour (Lovvorn et al., 2015) and endemic Burkitt Lymphoma (eBL) in Kenya (Kaymaz et al., 2017), and Skin cancers in Xeroderma Pigmentosum (XP) patients in Tunisia (Rekaya et al., 2018). eBL and XP are prevalent conditions in Kenya and Tunisia, respectively (Rekaya et al., 2018; Kaymaz et al., 2017). To our knowledge, no research has been done on understanding the genetic basis of paediatric solid non-haematological tumours in a South African cohort. Thus, there is a need to increase knowledge of disease-causing variants within CPGs and to implement a diagnostic screening method in a South African clinical setting.

1.5 Project Rationale

In South Africa, little is known about the prevalence of paediatric cancer and its inherited component, which predisposes children to early cancer development. The majority of current literature reporting in a low-middle income or Sub-Saharan African context shows varying reports on the rates of paediatric cancer. This variation is indicative of a gap in the knowledge of the actual burden paediatric cancer poses. It is crucial to fill this knowledge gap so that evidence-based

treatment plans can be developed and so that diagnostic measures can be improved on. Clinical criteria of heritable cancers may not be obvious or extensive in patients and family history of cancer is often not known, both of which are heavily relied on in the diagnostics of heritable cancers. Thus, the generation of mutation profiles using NGS technology will prove to be invaluable for the patients' diagnosis and prognosis. The use of NGS technology in the generation of novel data will allow for a more in-depth understanding of population-specific cancer-predisposing genes and their underlying variants and will provide the foundation for the development of a diagnostic panel. The development and evaluation of a diagnostic panel will in turn provide invaluable information to the medical specialists, the patients and their families as it will allow early detection of at-risk family members and better clinical management of the patient. Ultimately the use of the diagnostic panel will contribute to better treatment outcomes and an increase in knowledge of cancer-predisposing genes in a South African setting.

1.6 Aims & Objectives

1.6.1 Aims

To generate germline mutation profiles of SA paediatric solid-tumour cancer patients via a targeted cancer-predisposing NGS gene panel and to further gain insight into their potential role in the pathogenesis of childhood cancers.

1.6.2 Objectives

1. To recruit, collect and extract genomic DNA from a minimum of 32 paediatric patients with solid tumour cancers through paediatric oncology clinics.
2. To design a NGS targeted gene panel, consisting of 50 cancer-predisposing genes.
3. To evaluate and utilise this targeted gene panel design to perform NGS of these candidate genes coding exons and intron-exon boundaries.
4. To identify the putative disease-causing/contributing germline mutations most prevalent in our South African paediatric patient subset via bioinformatics analysis on the sequencing output.
5. Validate any identified mutations using Sanger sequencing.

2. Methods & Materials

This chapter aims to describe the research methods & materials utilised in identifying putative disease-causing/contributing, pathogenic/likely pathogenic germline variants in a subset of South African paediatric cancer patients, diagnosed with solid, non-haematological tumours.

2.1 Study Participants

2.1.1 Ethics, Recruitment & Sample Collection

This project has ethics approval under both the larger PeCan Project's ethical clearance certificate (M180855 – See Appendix B1.1) and an additional ethics clearance certificate in my name (M190751 – See Appendix B1.2). Both ethics clearance certificates were issued by the Wits Human Research Ethics Committee (HREC) (Medical).

This research project took place in the Division of Human Genetics, National Health Laboratory Services (NHLS) and School of Pathology, The University of the Witwatersrand. Collaborators included a multi-disciplinary team within the Division and three paediatric Oncology Units (See Preface). Recruitment took place from two approved paediatric oncology units, namely, Chris Hani Baragwanath and Charlotte Maxeke Johannesburg Academic Hospitals. Patient files were reviewed and paediatric patients who were diagnosed with a non-haematological malignant solid tumour aged from birth to 18 years were recruited. Patients diagnosed with haematological cancers, such as leukaemias and lymphomas, were excluded during the recruitment process. It was decided to focus on solid-tumour cancer types in this study as haematological cancers are often more extensively studied and thus, we aimed to help fill this knowledge gap.

During recruitment, paediatric patients invited to participate in the study and their parent(s) and/or guardians received age-appropriate information sheets outlining, in detail, what the study aims to achieve and what their participation in the study entails – Appendix B2. After discussing information sheets, both parents/guardians and child participants were given age-appropriate consent forms – Appendix B3. Participation in the study required signed consent from both parents/guardians and the child participant where applicable. In cases where the child participant was unable to consent e.g., too young to read or write or in cases of mental/physical illness, just the parent(s)/guardians signed consent was considered. Child participants and their parent(s)/guardian were also asked to consent to the future use of their samples and data in future related projects and were given the choice of whether they would like to receive critical and actionable results from this study relating to the patient's cancer and any information which may impact the health of the immediate family members.

Where possible, 5- 15 ml of whole blood was collected from each participant. Whole blood was collected in Vacutainer® Ethylenediaminetetraacetic acid (EDTA) tubes (BD, New Jersey, USA), transferred to Nunc™ conical sterile polypropylene centrifuge tubes (Thermo Scientific™, MA, USA) and stored at ~ -20°C at the NHLS. We recruited paediatric patients who had been off chemotherapy for at least a month or more as there is some evidence indicating that chemotherapy can affect the integrity of the DNA in whole peripheral blood (Kværner et al., 2018).

2.1.2 Patient Selection & Electronic Data Capturing

A sample size of 32 patients was selected for this project on the basis that the Ion Ampliseq™ Kit for Chef DL8 (Thermo Fisher, MA, USA) contains 32 unique barcodes. From the larger PeCan Project sample set, the 32 patients required for this study were selected based on DNA quality. Pure, high-quality DNA was defined as the sample showing little to no protein, phenol or salt contamination which may interfere with downstream applications of the extracted DNA (Lucena-Aguilar et al., 2016). This criterion was chosen as we decided it was the best way to select patients without introducing bias by selecting via cancer type, age, gender, ethnicity, or phenotypic profile. Note, as stipulated in the consent forms, all samples were de-identified and kept anonymous during this procedure.

During recruitment, a review of the patient's file revealed we a 'blind' bioinformatics control patient, who had been both clinically (through a paediatric oncologist at Charlotte Maxeke Johannesburg Academic Hospital) and molecularly (at NHLS-Braamfontein) diagnosed. This patient was labelled a 'blind' bioinformatics control because I was kept blind to the type and location of this variant during NGS data analysis. This control patient was selected as part of the 32 patient samples included in this study. The purpose of including a control was to validate the bioinformatics pipeline.

For the purpose of data collection and storage for the larger PeCan project and all studies which fall under the larger umbrella project (including this research project), an online Research Electronic Data Capture (REDCap™) database was created. REDCap™ is a webtool which is widely used throughout WITS for the creation of research databases (https://redcap.core.wits.ac.za/). The Demographics data form was exported to Microsoft Excel and basic summary statistics (averages and frequencies) were calculated on the following patient characteristics: Age at diagnosis, Gender, Tumour type, Ethnolinguistic group(s) and Pedigree suggestive of family history of cancer.

2.2 Experimental Techniques

2.2.1 DNA Extraction

An adapted salting-out method was used to extract genomic DNA from whole blood samples (Miller, Dykes & Polesky, 1988). Briefly, red blood cells within the whole blood samples were lysed using a Triton-X lysing buffer. This solution was then treated with a concentrated salt solution as well as a series of centrifugation and wash steps to allow for the separation of the red blood cell debris from the pelleted white blood cells. Samples were then placed in an incubator at 42°C overnight in a mixed Proteinase-K, Tris-HCl, EDTA and 10% Sodium Dodecyl Sulphate (SDS) solution. DNA was then precipitated out through the addition of absolute ethanol and a highly concentrated salt solution. The precipitated DNA was re-suspended in a Tris-Buffer solution and stored in a fridge at ~ 4°C until required.

2.2.2 DNA Quantification & Quality Control

Quantitative and qualitative checks on the extracted genomic DNA were assessed using two different methods. Firstly, the DNA concentration and purity were determined using the Nanodrop Spectrophotometer - Nanodrop™ One (Thermo Fisher, MA, USA). The purity of the DNA sample was assessed using the A260/A280 and A260/A230 ratios, which indicate the presence of phenol, protein, RNA or salt contamination which may interfere with downstream DNA applications. Where patient DNA samples were not within an A260/A280 ratio range of ≤1.6-1.8 and/or an A260/230 ratio range of ≤ 2.0-2.2 and a concentration of > 50 ng/ul, a NucleoSpin Gel and PCR Clean-Up Kit was used (Macherey-Nagel, Düren, DE) to remove contaminants and any remaining chaotropic salts from the wash buffer, thus improving the purity of the samples. DNA concentration was measured again after NucleoSpin clean-up to assess improvement. Secondly, agarose gel electrophoresis was used to ensure the DNA is of high molecular weight and that it is of good integrity. The patients' samples were run on a 2% agarose gel. The band sizes were then determined relative to the 1 Kb molecular weight marker (New England Biolabs, MA, USA) loaded and run in parallel to the patient DNA samples. DNA integrity was then assessed by examining the shape and structure of the DNA bands produced in the agarose gel. Agarose gel results were visualised and captured under Ultra-Violet (UV) light in the Omega Fluor™ Gel Documentation System (Vacutec, Johannesburg, SA).

2.2.3 Next Generation Sequencing

Figure 2.1 illustrates the summarised NGS methodology followed in the sequencing of the extracted patient DNA samples using the designed 50 Cancer Predisposing Gene (CPG) targeted gene panel. This methodology will be expanded on in the coming subsections.

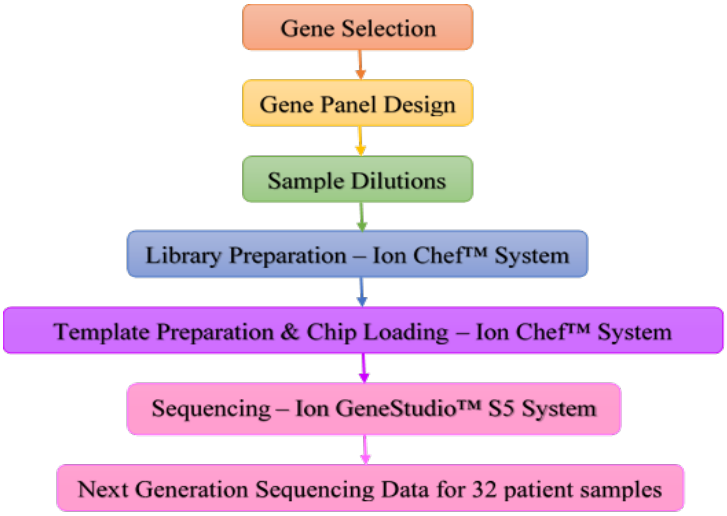


Figure 2.1 – Next-Generation Sequencing Workflow: Diagram illustrating Next-Generation Sequencing methodology overview.

Candidate Gene Selection

An extensive literature review and analysis was conducted to identify the genes most commonly associated with the development of paediatric cancers in order to select the 50 genes included in this project's gene panel design. The literature was searched for studies which utilised NGS technologies in investigating the role of germline variants within candidate CPGs in the formation of malignant solid non-haematological tumours in childhood. Papers which investigated the role of CPGs in haematological cancers only and those which researched CPGs in adults only were not further reviewed as they were not the focus of this study. At the time of this gene panel design, nine international studies and an extensive cancer-predisposition gene review by Rahman (2014) were identified - (Diets et al., 2018; Gröbner et al., 2018; Kline et al., 2017; Chang et al., 2016; Oberg et al., 2016; Parsons et al., 2016; Worst et al., 2016; Mody et al., 2015; Zhang et al., 2015). The most frequently reported CPGs harbouring disease-causing/contributing germline variants identified in these studies were included in the gene panel design. These genes are well-known in their association and role in solid tumour paediatric cancer development.

This CPG selection process particularly focused on genes which harbour heritable germline variants, responsible for the development of familial- and hereditary-cancer predisposing syndromes (CPSs) and thus increase the likelihood of childhood cancer development. From the

nine reviewed studies, 33 of the most frequently reported CPGs harbouring disease-causing/contributing germline variants were identified and included in the gene panel design – see Figure 2.2 (A). The remaining 17 genes were identified through a literature review for genes that harboured heritable germline variants, responsible for the development of familial/hereditary CPSs, which in turn increases the likelihood of developing cancer in childhood – see Figure 2.2 (B). For the full list of 50 CPGs, the full gene names, inheritance pattern, associated CPSs and childhood cancers included in this project's gene panel, see Appendix C1.

The majority of the CPGs included in this gene panel are inherited in an autosomal dominant pattern (70%), whilst ~20% of the included CPGs are seen to confer a moderate or high risk of cancer in an autosomal recessive manner. Nine of this panel's CPGs (28%) (*BRCA1, BRCA2, BUB1, LZTR1, MITF, MLH1, MSH2, MSH6, PMS2*) have been identified in both autosomal dominant and autosomal recessive conditions. Only one of this gene panel's CPGs is inherited in an X-linked manner (2%) (*GPC3*) – See Figure 2.2.

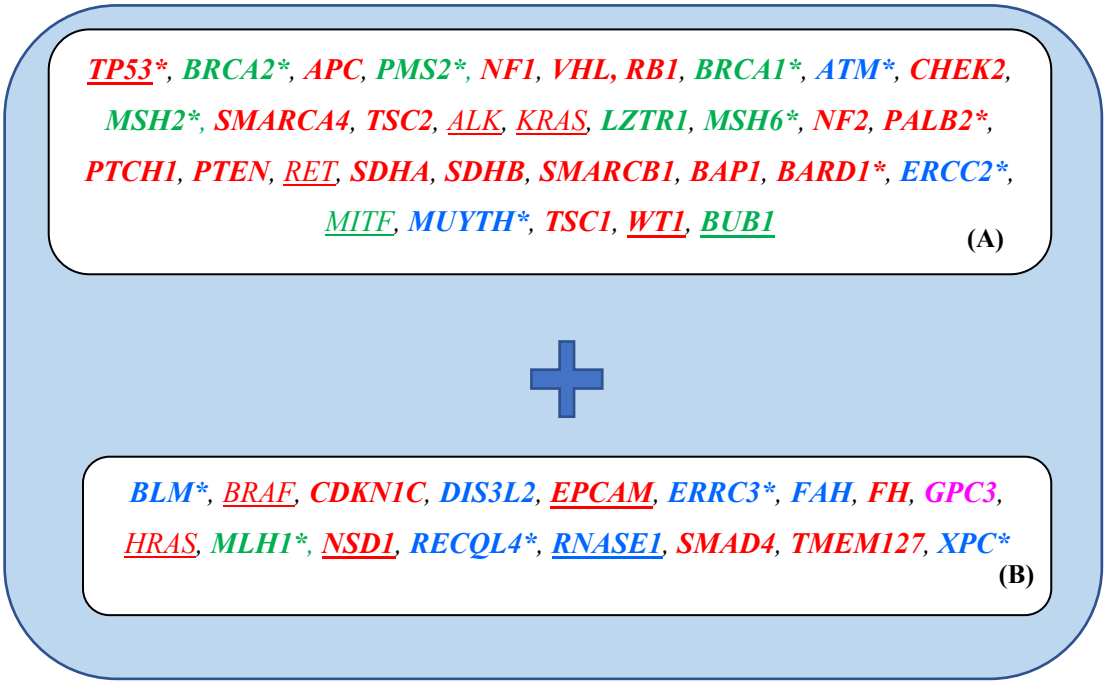


Figure 2.2 - Diagram depicting 50 Cancer-Predisposing Genes selected for targeted NGS gene panel (A & B): **(A)** The 33 CPGs selected through the nine international paediatric cancer NGS studies which are listed in order of frequency of variants detected in that gene (most-least). **(B)** The 17 genes selected and included in the gene panel design based off their associations with CPSs and increased predisposition to the development of cancer in childhood. **KEY** – Genes listed in Red: Autosomal Dominant Inheritance, Blue: Autosomal Recessive Inheritance, Green: Both Autosomal Dominant & Autosomal Recessive Inheritance, Pink/Purplish: X-linked Inheritance. Genes highlighted in **Bold:** Tumour-Suppressor Genes (TSGs), Underlined: Proto-oncogenes, **Bold & Underlined:** TSGs & Proto-oncogenes, Genes with *: DNA Repair genes. See Appendix C1– Gene Panel List, for full gene names.

As seen in the literature, only ~12% of this project's genes act as proto-oncogenes, conferring a predisposition to cancer through gain-of-function mutations, becoming oncogenes. Whilst the majority of the included CPGs are TSGs (~76%), inactivating their functions and contributing to oncogenesis through loss-of-function mutations. There are six genes (*TP53, WT1, BUB1, EPCAM, NSD1, RNASE1*) (~12%) that have been described to have both TSG and oncogenic mechanisms, depending on the cellular context in which the gene product is expressed. Notably, 16/50 CPGs included play vital roles in DNA repair, suggesting that mutations in these genes may increase the likelihood of cancer development through an accumulation of unrepaired errors in the genome, rather than a single cancer initiation event. There are other rarer mechanisms of oncogenesis of the 50 CPGs which may occur, such as Dominant-Negative Effect (*WT1*) and Haploinsufficiency (*NSD1*).

Gene Panel Design

After gene selection, the panel was designed using Thermo Fisher's Scientific Ion AmpliSeq online panel design tool (https://www.ampliseq.com/login/login.action). The Ion Ampliseq designer allows you to create a customisable targeted NGS panel for human disease research. Thermo Fisher's Scientific Ion AmpliSeq On-Demand Catalog contains over 5000 pretested genes, wet-lab validated for germline analysis, especially relevant in the research of inherited diseases, including hereditary cancers. The gene panel designer outputs pool(s) of multiplex oligonucleotide primer pairs, with each pair designed to amplify a targeted genomic region based on the gene selected in the On-Demand Catalog or an input gene list. 48 of the 50 candidate genes in this study were available in the On-Demand Catalog and their design provided high exon coverage with a high percentage of gene uniformity. This gene panel design was predicted to have an *in-silico* coverage of 100% and was designed using the *Homo Sapiens* (Human) reference genome - Genome Reference Consortium Human Build 37 (GRCh37). A total of 1316 amplicons were targeted, distributed over two primer pair pools (658 amplicons in each) and covered 248.773 kilobases (kb) of the human genome. The targeted regions covered all the exons within each of the 48 candidate genes as well as the exon-intron boundaries, allowing the detection of all coding variants as well as splice-region variants.

2/50 of our candidate genes, namely Cyclin-Dependent Kinase Inhibitor 1C (*CDKN1C*) Ribonuclease A family member 1 (*RNASE1*), were not available in the On-Demand Catalog and thus a Spike-In panel had to be created. A Spike-In panel is a made-to-order set of oligonucleotides synthesized for the additional genes which are not available in the On-Demand Catalog. This panel is then 'spiked-in' (pipetted) into the On-Demand panel during NGS through the manufacturer's protocol. This allows the sequencing of all 50 of the candidate genes as one combined panel and

to be recognised as such by the Ion Torrent Suite™ Software (version 5.12.0). Our Spike-In panel design covered 100% of the *RNASE1* gene's coding region and due to specificity issues covered only 43.7% of the *CDKN1C* gene's coding region.

Library Preparation

The 32 patient genomic DNA samples were sequenced using the 50-candidate gene panel design in two sequencing runs with 16 samples sequenced per run. During library preparation, each individual library run contained eight patient samples. Library runs one and two were combined to make a 16-patient pooled library used in the first sequencing run, and library runs three and four were combined to make the second 16-patient pooled library used in the second sequencing run – See Figure 2.3.

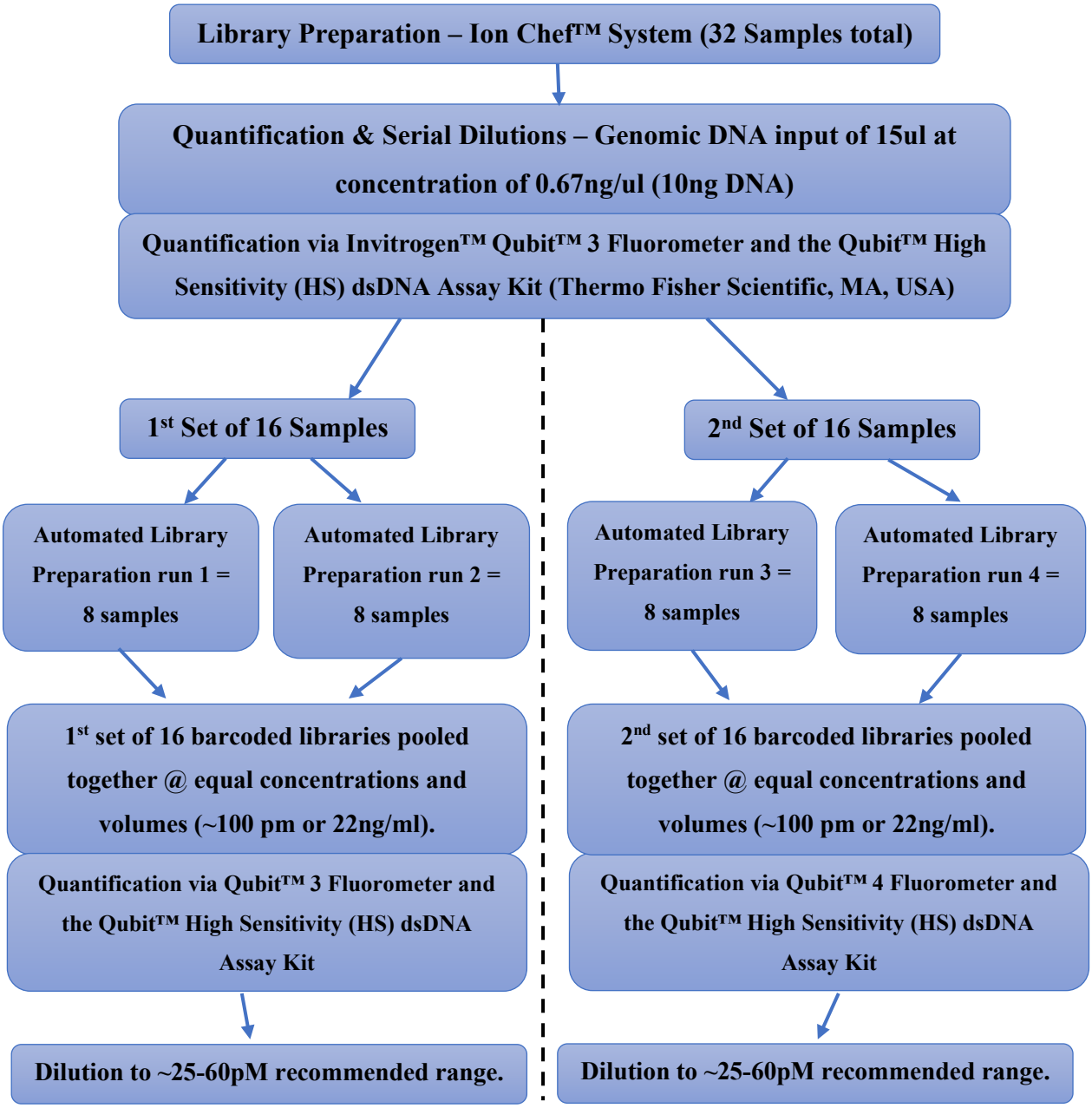


Figure 2.3 - Library Preparation Workflow: Diagram illustrating the distribution of the 32 patient samples during library preparation.

Automated library preparation was performed using the Ion Chef™ System of the Ion Torrent™ NGS Systems, using the Ion AmpliSeq™ Kit for Chef DL8 (Thermo Fisher, MA, USA). This process involves Polymerase Chain Reaction (PCR) amplification of the designed target regions; which produces DNA amplicons within the desired size range known as amplicon libraries. Specific DNA fragment sequences known as adapters were then annealed to the 3' and 5' ends of these amplicons. One of these adaptors, known as a sequencing adapter, is annealed so that the amplicon can bind and anchor the fragment to a complimentary sequence on the Ion Sphere Particles (ISPs) during template preparation. The second adapter is known as the barcode adapter, which binds to the primer annealing site and allows tracking of the sample during sequencing. Each library corresponds to a single sample and the Ion Chef™ system works to create 8 uniquely barcoded and combined libraries during one library preparation run, which will undergo library quantification and control before proceeding to templating.

15 ul of the first set of eight patient DNA samples were pipetted into the IonCode™ 96 Well PCR Plate which contains eight unique dried down IonCode™ barcodes. The two synthesised gene panel primer pools were then mixed with the two additional Spike-In primer pools before loading them into separate positions on the Ion AmpliSeq™ Chef DL8 reagents cartridge. The Ion Chef™ library preparation run plan was set up as per the manufacturer's instructions and the IonCode™ individual barcodes are linked up to their respective patient sample number. This feature allows the tracking of that particular sample throughout the preparation, templating and sequencing process. Cycling conditions were determined as instructed by the Ion AmpliSeq™ Preparation on the Ion Chef™ System user guide. This process was repeated for the second set of eight patient samples – see Figure 2.3.

The same library preparation protocol as described above was followed for the remaining 16 patient DNA samples when undergoing library preparation on the Ion Chef™ System for the second sequencing run. The only exception being the use of the Invitrogen™ Qubit™ 4 Fluorometer (Fisher Scientific part of Thermo Fisher Scientific, MA, USA) in all dsDNA quantification methods in place of the Invitrogen™ Qubit™ 3 Fluorometer as it as noted later that the machine was experiencing technical difficulties, an observation unfortunately only discovered after completion of the first sequencing run. The technical error experienced using the Qubit 3™ Fluorometer resulted in the over-dilution of combined library A (libraries one + two), ultimately resulting in the concentration of Library A being below ~25pM. The concentration of combined

library B (libraries three + four) was ~70 pM, which is slightly higher than the optimal range recommended (25 – 60 pM).

Template Preparation & Chip Loading

Automated template preparation was performed on the Ion Chef™ System (Thermo Fisher, MA, USA) using the Ion 510™ & Ion 520™ & Ion 530™ Kit – Chef (Thermo Fisher, MA, USA), according to the manufacturer's protocol. Automated template preparation and chip loading were set-up in the same run plan on the Ion Torrent Suite™ Software (version 5.12.0 Thermo Fisher, MA, USA). The barcodes ligated to the amplicons during library preparation were linked to the templating run, allowing the samples to continue being tracked throughout the templating and chip-loading process. See Figure 2.4.

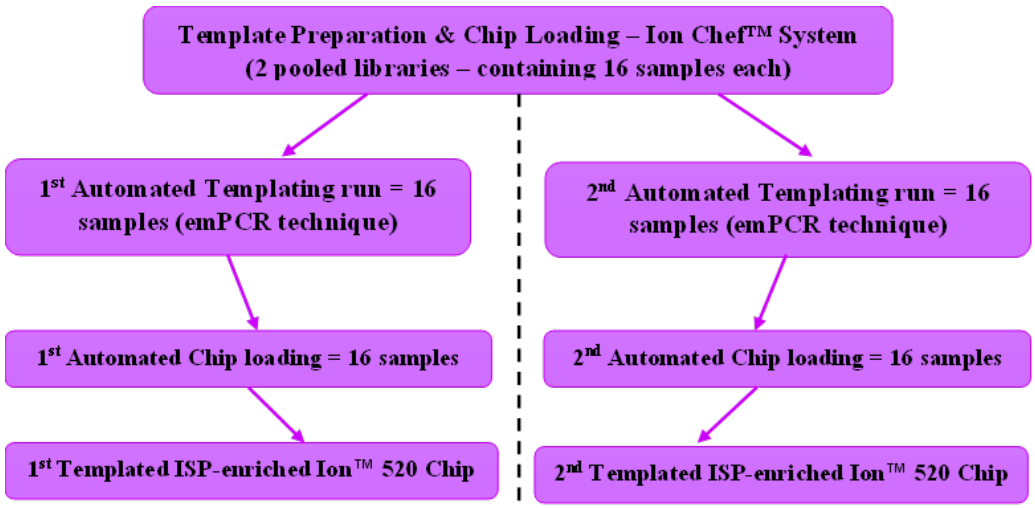


Figure 2.4 – Template Preparation & Automated Chip Loading Workflow: Diagram illustrating the distribution of the 32 patient samples during template preparation and automated chip loading.

The Ion Chef™ automated templating process uses an amplification technique known as emulsion PCR (emPCR) to prepare the DNA templates required for sequencing (Kanagal-Shamanna, 2016). During each templating run, the pooled and prepared library (containing 16 individual libraries), reagents, oils and ISPs are mixed and then amplified in the onboard thermal cycler (Thermo Fisher Scientific, 2019a). ISPs are essentially 'microscopic beads' with complementary adapter sequences attached to their surface (Kanagal-Shamanna, 2016). This allows for specific amplicons to attach to their complementary sequences on the ISPs during template preparation. This results in template-positive ISPs where one template-positive ISP contains multiple copies of the same amplicon, allowing a strong enough signal to be detected during the sequencing of those amplicons (Chai, 2019). The enriched templated ISPs were deposited into the Ion 520™ Chip (Thermo Fisher, MA, USA) by the Ion Chef's automated chip pipettor. The Ion semiconductor sequencing chip contains millions of microwells into which a single templated ISP is deposited during

automated chip loading. The same automated template preparation and chip loading protocol as described above was followed for the second set of 16 patient DNA samples. This resulted in a second templated ISP-enriched Ion 520™ Chip ready for the second sequencing run.

Ion Torrent S5™ - Sequencing of Candidate Genes

Upon completion of chip loading, the Ion 520™ Chip was loaded into the Ion GeneStudio™ S5 System (Thermo Fisher, MA, USA) of the Ion Torrent™ NGS Systems for sequencing. All necessary reagents and consumables used in sequencing were provided in the Ion S5™ Sequencing kit (Thermo Fisher Scientific, MA, USA). Genomic regions corresponding to our On-Demand Ion AmpliSeq™ panel design were uploaded to the Ion Torrent server in a Browser Extensible Data (BED) file format (Quinlan & Hall, 2010). The Ion Torrent Suite™ software (v.5.12.0) allows the expansion of sequencing and data analysis via preinstalled plugins. For our sequencing runs we selected the coverageAnalysis and variantCaller plugins. The coverageAnalysis plugin provides statistics on the level of sequence coverage achieved for the gene panel's targeted genomic regions and the variantCaller plugin calls a range of variants within a sample when compared to the reference genome GRCh37, including Single-Nucleotide Variants (SNVs), Multi-Nucleotide Variants (MNVs) and insertions and deletions (Indels). See Figure 2.5.

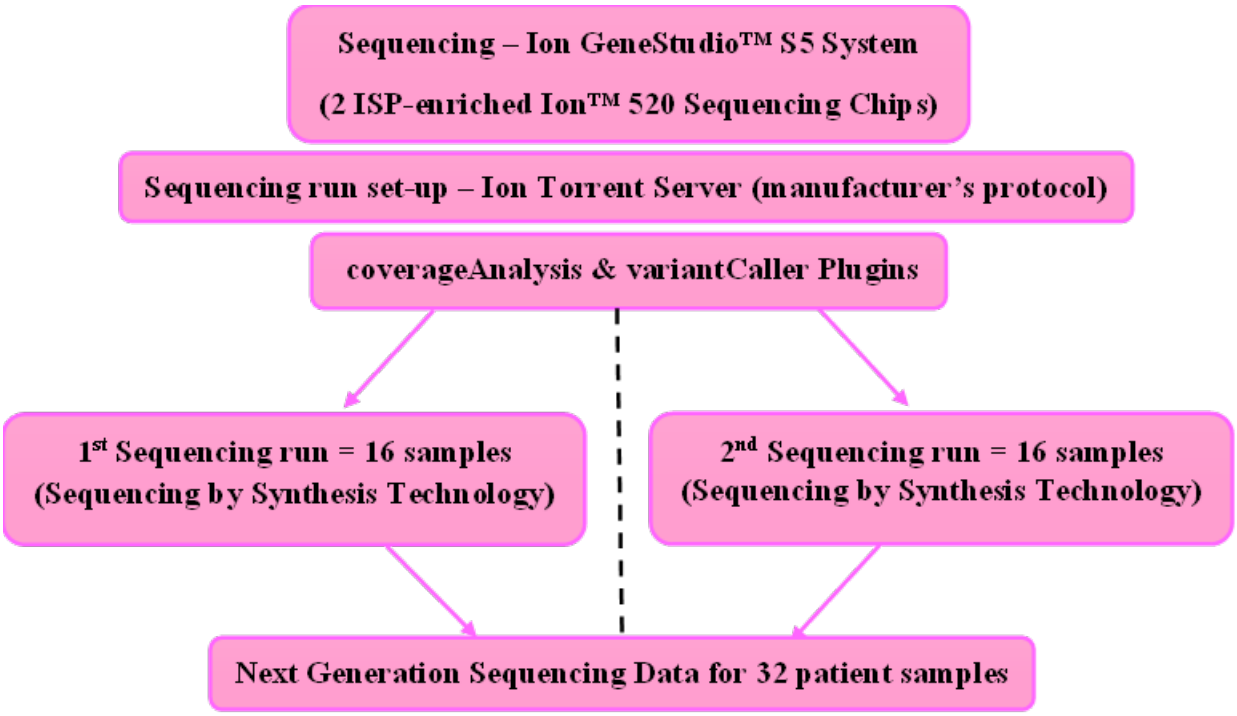


Figure 2.5 – Chip Sequencing & NGS Data Generation Workflow: Diagram illustrating the distribution of the 32 patient samples during sequencing on the Ion GeneStudio™ S5 system.

The Ion GeneStudio™ S5 System is a massively parallel technology which utilises a sequencing-by-synthesis technology. Whereby, the semi-conductor Ion 520™ Chip is capable of detecting a

single nucleotide incorporation during the DNA polymerisation of the ssDNA amplicons templated on the ISPs (Merriman & Rothberg, 2012). During sequencing, the chip's microwells were flooded with a solution containing buffers, DNA polymerase and one of the four deoxyribonucleotide triphosphates (dNTPs) at a time per sequencing flow over the sequencing chip (Merriman & Rothberg, 2012). As our gene panel was based on a ~200-base-read amplicon design, we selected 500 sequencing flows per sequencing run. The incorporation of a complementary dNTP into the growing DNA strand releases a pyrophosphate molecule and an H^+ ion thereby increasing the pH of the microwell solution (Merriman & Rothberg, 2012). This change in pH is detected by an ion-sensitive layer at the bottom of the microwell, which then triggers a series of electrical impulses to register that a specific dNTP was incorporated (Merriman & Rothberg, 2012). Signal processing and assembling of the DNA sequence are then performed by the Ion Torrent Suite™ software (v.5.12.0) through these electrical impulses. Each dNTP flow is followed by a wash step is performed to remove any remaining non-incorporated dNTP, thus ensuring greater accuracy when recording the sequence of dNTPs incorporated into the sequence.

Two sequencing runs were performed using the method described above on the Ion GeneStudio™ S5 system with each run sequencing 16 patient DNA samples. This resulted in a total of 32 patient DNA samples undergoing NGS sequencing for this research project.

2.2.4 Bioinformatics – Analysis of NGS Data Output

Figure 2.6 summarises the methodology followed during the bioinformatic analysis of the 32 patient NGS data. The quality of the sequencing runs were evaluated by assessing pre-alignment and post-alignment metrics generated by the Ion Torrent Suite™ software (version 5.12.0 Thermo Fisher, MA, USA). Mapping, alignment, and variant calling were performed by the Ion Torrent Suite™ software in an automated process (Thermo Fisher, MA, USA). Annotation of patient Variant Call Format (VCF) files was performed using the Ion Reporter™ software (Thermo Fisher, MA, USA) and Ensembl's Variant Effect Predictor (VEP). Annotated VCFs then underwent a filtering pipeline, including filtering according to variant effect, filtering according to predictions made by *in silico* pathogenicity and conservation prediction tools, a ClinVar filter and a variant frequency filter. Putative disease-causing/contributing variants which remained after the variant filtering pipeline underwent further variant prioritization. Variant prioritization included the classification of variants through the ACMG-AMP guidelines (assisted by VarSome). The various stages of bioinformatic data analysis will be expanded on in the coming sections.

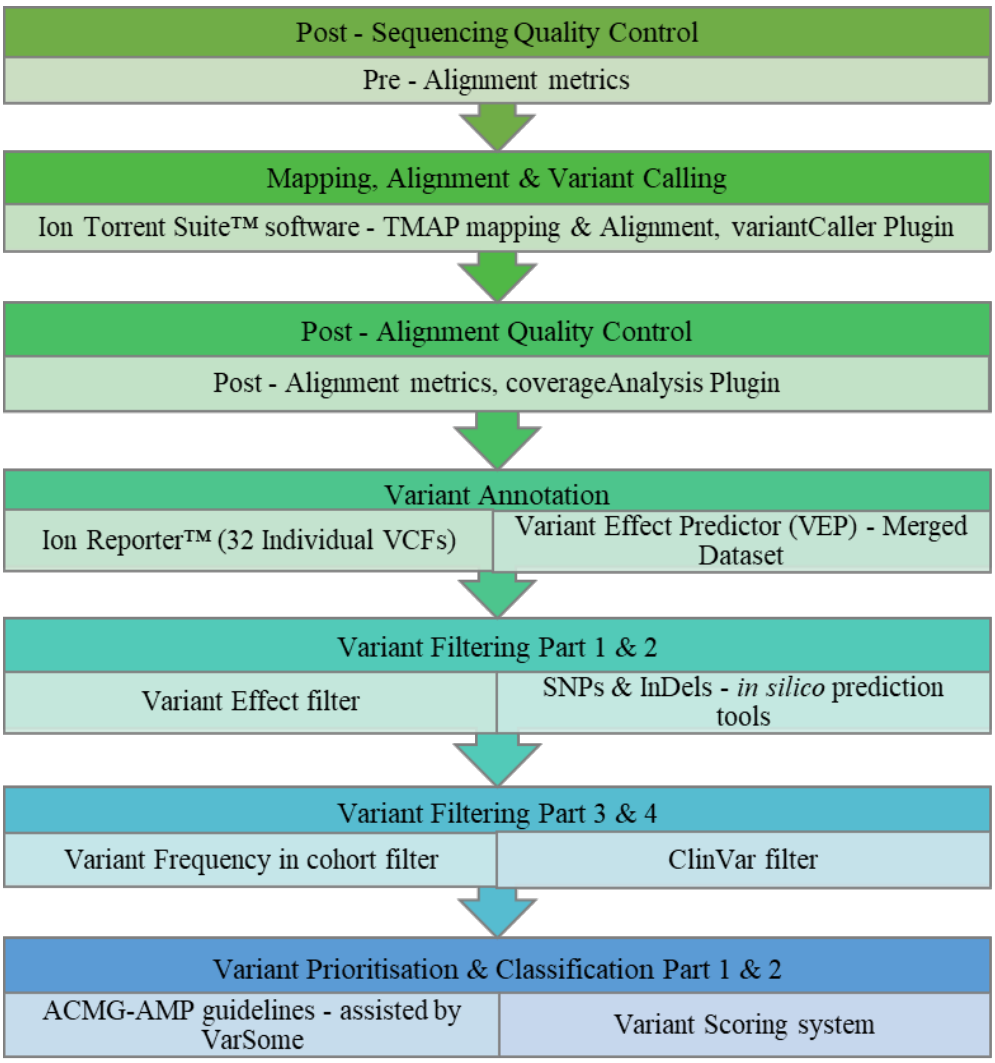


Figure 2.6 - Bioinformatics Workflow: A diagram illustrating the bioinformatics workflow followed during interpretation of Next-Generation Sequencing data. Acronyms – TMAP: Torrent Mapping Alignment Program, SNPs: Single Nucleotide Polymorphisms, ACMG-AMP: American College of Medical Genetics and Genomics – Association for Molecular Pathology

Post-Sequencing Quality Control

The quality of both sequencing runs was assessed using the sequencing metrics produced by the Ion Torrent Suite™ software (v.5.12.0) – including both pre-and post-alignment metrics. Pre-alignment metrics generated include statistics assessing the quality of templating and chip loading. The quality of templating was assessed through the run's generated ISP summary, which reports the % of wells containing an ISP (% of addressable wells), the number of those wells which contain a live ISP (i.e., the number of ISPs which have a key signal corresponding to the library key or test fragment (TF) key signal), the % clonal and polyclonal reads as well as the % of low quality, % of adapter primers and the 1% TF. Polyclonality refers to the attachment of two or more different DNA templates attached to a single ISP (de Leng et al., 2016). Polyclonal reads are removed before subsequent downstream data analysis as these ISPs contain different DNA

templates and thus different DNA sequences and will not respond in unison with each dNTP flow during sequencing, thus giving a mixed and inaccurate sequence. The ISP summary takes into account these pre-alignment metrics in order to calculate the number of usable reads from each sequencing run. This number is presented as a percentage of the total number of library ISPs detected as follows:

$$\%\ usable\ reads = \frac{library\ reads\ passing\ all\ filters\ x\ 100}{total\ number\ of\ library\ ISPs\ identified}$$

Before downstream analysis, these metrics were assessed according to the recommended thresholds; % polyclonal reads are < 35% and the % low-quality reads are < 20% (de Abreu et al., 2016).

The quality of chip loading was assessed through the generated run's % ISP loading and ISP density. The % ISP loading represents the % of the total wells which were loaded with ISPs during the automated chip loading process. The ISP density is a heat map of the sequencing chip's surface, allowing a visual representation of the distribution of the ISPs loaded across the wells. Areas closer to the red spectrum indicate a higher loading density whilst areas closer to the blue spectrum indicate a lower loading density.

All pre-alignment metrics facilitated the interpretation of the quality of the sequencing reads produced as a result of the automated library preparation, template preparation & chip loading. These metrics allowed us to assess if there was adequate quantity and quality NGS data after the sequencing run which can be relied on for accurate downstream data analysis.

Mapping, Alignment & Variant Calling

Sequencing reads were mapped and aligned to the reference genome GRCh37 by the Ion Torrent Suite™ software (v.5.12.0) (before the transition of software to GRCh38 build occurred) using the Torrent Mapping Alignment Program (TMAP) (Thermo Fisher, MA, USA), producing Binary Alignment Map (BAM) files, containing the sequence alignment data sorted by the corresponding reference location (Thermo Fisher Scientific, 2019c; Li et al., 2009). Germline variants were called by the Ion Torrent Suite™ software (v.5.12.0) via the inbuilt variantCaller plugin using the default germline parameters. The identified sequence variants and their locations were stored and output in Variant Call Format (VCF) files (Danecek et al., 2011).

Post-Alignment Quality Control

Post-alignment metrics assessed included the total number of reads, the number and percentage (of the total number of reads) of aligned and unaligned reads, and the alignment plot. The

coverageAnalysis plugin generated statistics on the sequence coverage achieved during the sequencing of the targeted genomic regions, these statistics were generated post mapping and alignment of the generated reads. The sequence coverage statistics were generated on a per barcode (i.e., per sample) basis. These statistics include the total number of reads mapped to the reference genome GRCh37, the percentage of reads mapped to any target region out of the total reads mapped, the mean base coverage depth and the percentage uniformity of the base coverage. The mean base coverage depth refers to the average number of reads for all the bases sequenced, mapped and aligned to the reference genome (Jennings et al., 2017). This statistic is of particular importance when determining the accuracy of base calling in a patient sample, regardless if the reference or alternate allele is called. For germline testing, we considered an average base coverage depth of >30x as sufficient for an accurate and confident base call (Jennings et al., 2017).

Variant Annotation & Data Sorting

The Ion Reporter™ version 5.12 software from Thermo Fisher Scientific was initially used to annotate the patient's VCF files generated by the variantCaller plugin on the Ion Torrent Suite™ software (v.5.12.0 Thermo Fisher Scientific, MA, USA). The software contains a set of bioinformatic tools, designed to simplify data annotation, filtering and interpreting of Ion Torrent™ sequencing data (Thermo Fisher Scientific, 2019b). Patient VCF files were uploaded from the Ion Torrent™ Server to the Ion Reporter™ software for an annotation-only workflow. The standard Ion Reporter™ annotation set was used to assign biological meaning to the variants called. See Appendix C2 for the list and description of the population and public databases and prediction tools in the standard Ion Reporter™ annotation set used to annotate patient VCF files.

In addition to Ion Reporter™, the Ensembl Variant Effect Predictor (VEP) annotation software was used to obtain more detailed variant impact annotations. VEP is a robust software suite which provides detailed genomic annotations in both coding and non-coding regions (McLaren et al., 2016). VEP accesses a large range of population and public databases and provides variant annotations from numerous prediction tools in order to provide a comprehensive understanding of the variant effect. VEP's variant annotation allows for the interpretation of a variant's effect on numerous transcript versions, protein isoforms and regulatory regions as well as providing disease and phenotype information (McLaren et al., 2016). Patient VCFs were supplied to VEP via the command line option, and this analysis was done on the Wits Core Computing Cluster (Hazelhurst, 2020) accessed through the Sydney Brenner Institute for Molecular Biosciences (SBIMB). We

opted for the command line option as it allows for greater flexibility in the variant annotations required for a particular dataset and is better suited for larger data sets.

However, before submission of the patient VCFs to VEP for annotation, all the patients' raw VCF files (downloaded from Ion Reporter™) were merged into one large dataset. Merging of the patients' VCFs into one large dataset was performed in order to reduce analysis time and to remove any filtering bias, such as filtering a patient's variants according to their cancer type or phenotype. This allowed for the same bioinformatics filtering and prioritisation pipeline to be applied to all patients simultaneously. Merging of the raw VCFs would also facilitate variant frequency determination within this study population.

The merged dataset then required reformatting to match the input required for VEP's annotation software. It was further noted that Indels and complex multi-allelic variants required a different command to separate and reformat compared to SNVs, thus, for reformatting purposes, the SNVs were separated from Indels and complex multi-allelic variants. This variant-type separation also facilitated more variant type specific downstream analysis – this will be expanded on under variant filtering. See Figure 2.7 for an example showing the format of the Ion Torrent™ VCFs compared to the VCF format produced after reformatting (and required for VEP input). The origin of each variant change after annotation could be found in the variant frequency file compiled, stating each position where a variant was detected, the change detected and in which patient(s) the variant change was detected in.

A

Ion Torrent VCF data format:			
Chromosome:Position	Reference	Allele1	Allele2
chr1:241682881	C	C	CG
chr1:241682881	C	CG	CGG
chr1:241682881	C	CG	G
chr1:45796256	GC	GC	AT
chr10:43606598	A	A	AG
chr10:89685245	TTTTG	TTTTG	T
chr10:89693022	TG	TG	GT
chr10:89693022	TG	TG	TGT
chr10:89693022	TG	TG	TT
chr10:89720578	TTTA	TTTA	T
chr10:89720579	TTA	TTA	T
chr10:89720619	T	T	TCT

B

Reformatted data input for VEP:			
Chromosome	Position	Reference Allele	Alternate Allele(s)
chr1	241682881	C	CG,CGG,G
chr1	45796256	GC	AT
chr10	43606598	A	AG
chr10	89685245	TTTTG	T
chr10	89693022	TG	GT,TGT,TT
chr10	89720578	TTTA	T
chr10	89720579	TTA	T
chr10	89720619	T	TCT

Figure 2.7 – Ion Torrent to VEP VCF reformatting example: (A) Screenshot of Ion Torrent VCF format before reformatting data. (B) Screenshot of VEP VCF format after reformatting data.

After merging, reformatting, and separating the SNV, Indel and MNV data, files were individually submitted to VEP for annotation. See Figure 2.8 for the command used to annotate the merged SNVs and Indels + complex multi-allelic variants through Ensembl's VEP software. See

Appendix C3 for the list and description of the population and public databases and prediction tools selected for annotation of the patients' VCF files through Ensembl's VEP.

```
### Run VEP for snps:

/opt/exp_soft/bioinf/ensembl-vep/vep --ASSEMBLY GRCh37 -i all_snps.txt -o snps_vep.txt --dir_cache
/dataB/aux/37/cadd/vep_caches/ --cache --port 3337 --symbol --canonical --ccds --tab --check_existing --sift s --
dir_plugins /opt/exp_soft/bioinf/ensembl-vep/plugins/ --plugin CADD,/dataB/aux/37/cadd/whole_genome_SNVs.tsv.gz --sift b
--polyphen b --plugin Condel,/dataB/aux/37/cadd/config/Condel/,b,2 --plugin
dbNSFP,/dataB/aux/37/cadd/dbNSFP_hg19.gz,MetaSVM_pred,MetaSVM_score,MetaSVM_rankscore,MetaLR_pred,MetaLR_score,MetaLR_ranks
core,FATHMM_pred,FATHMM_score,FATHMM_converted_rankscore,CADD_phred,CADD_raw,CADD_raw_rankscore,DANN_rankscore,DANN_score,L
RT_converted_rankscore,LRT_pred,LRT_score,fathmm-MKL_coding_group,fathmm-MKL_coding_pred,fathmm-
MKL_coding_rankscore,fathmm-
MKL_coding_score,PROVEAN_converted_rankscore,PROVEAN_pred,PROVEAN_score,MutationAssessor_UniprotID,MutationAssessor_pred,Mu
tationAssessor_score,MutationAssessor_score_rankscore,MutationAssessor_variant,MutationTaster_score,MutationTaster_pred,Mut
ationTaster_converted_rankscore --af_1kg --af_gnomad

## The same command above was used to run the merged indels file on VEP
```

Figure 2.8 – VEP annotation command: screenshot of the command used to submit merged SNVs and Indels VCF files to VEP via Wits Core Computing Cluster

Variant annotation by both the Ion Reporter™ and VEP software was based on the GRCh37 human reference genome. The merged SNVs and Indels annotated VCFs from VEP provided greater variant evidence provided by additional *in silico* prediction tools and were output on the Linux terminal, allowing more flexibility in the application of downstream filters and prioritisation of the variants. Thus, it was decided to proceed with the annotated VCF files from VEP only for the further filtering and prioritisation of variants.

Variant Filtering

The final result of variant annotation using VEP was two merged annotated VCF files, one with all the annotated patient SNV data and the other with all the annotated patient Indel + complex multi-allelic data. However, it was noted that some of the variants (predominantly Indels & MNVs) in the annotated VEP output had blank or empty fields where there should be scores from pathogenicity and conservation prediction tools. The variants with no scores or pathogenicity predictions were resubmitted to selected tool plugins via their individual web platforms (outside of the VEP environment) as a way to debug issues with importing scores from the dbNSFP (database for Non-Synonymous SNPs' functional prediction) database (Liu, Jian & Boerwinkle, 2011) as relied on by VEP. The plugins that were prioritised for this troubleshooting included CADD and SIFT-INDEL.

To ensure a standard uniform approach for transcript choice, expression of the variant in Ensembl's VEP canonical transcript was selected. Ensembl's canonical transcript is determined by the transcript with the longest coding sequence in the Consensus Coding Sequence Set (CCDS) with no stop codon (Ensembl, 2020). The CCDS represents the coding sequence which is consistently annotated between Ensembl, The National Center for Biotechnology Information

(NCBI), the HUGO Genome Nomenclature Committee (HGNC) and the Mouse Genome Informatics (MGI) (Ensembl, 2020). The Ensembl canonical transcript and protein length are most often in agreement with the UniProtKB/Swiss-Prot's canonical sequence, which is defined to be the most prevalent and due to its length or amino-acid composition, facilitates the clearest description of characteristics such as isoforms and domains (UniProtKB/Swiss-Prot, 2018).

The annotated variants were not filtered using MAF to avoid filtering out a potentially rare or novel disease-causing variant. This is because numerous annotated variants in both variant-type files did not report any MAFs under the population-database fields. However, it should be noted that the MAF of variants were checked after prioritisation to ensure they were not frequently occurring polymorphisms. All variant filters were applied to both the merged SNVs, Indels and complex multiallelic variant files, with the exception of prediction tool filters which were specific for SNVs or Indels and complex multi-allelic variants only. Variants were filtered based on variant effect, *in-silico* tools (see Appendix C3 for the list of prediction tools and their interpretation used for the filtering of SNV and Indels and multi-allelic variants), variant frequency in our patient cohort, ClinVar classification, and variant sequencing quality. Table 2.1 describes the use of these criteria in detail.

Table 2.1 Variant Filtering Criteria: Table describing the different variant filtering-out criteria employed in achieving the list of putative disease-causing/contributing variants.

Variant Filtering Criteria	Exclusion Criteria	Reason
Variant Type (1st variant filter)	Intronic, synonymous, and in-frame variants	• No effect on protein structure and function • Aim to identify germline variants which are likely to be disease-causing/contributing and would therefore have an effect on protein structure and/or function.
***In-silico* Tools (2nd variant filter)**	Variants predicted to be non-damaging	• To identify variants predicted to have a deleterious effect on protein function and structure, as predicted by various *in-silico* pathogenicity and conservation prediction tools - See Appendix C3 for the list of prediction tools and their interpretation used for the filtering of SNVs, and Indels and multi-allelic variants.
Variant Frequency (3rd variant filter)	Variants observed in ≥ 5 patients (~20%) patients in our cohort	• A threshold of 20% was selected as it takes into account the project's smaller sample size without the filter being too lenient or stringent. A variant frequency filter facilitates the identification of disease-causing/contributing variants occurring at lower frequencies while filtering out higher frequency polymorphisms within our patient subset.
ClinVar Report (4th variant filter)	Benign & Likely Benign Variants	• To filter out variants reported by ClinVar to have a benign/likely benign clinical significance. Thereby leaving variants which may be necessary for normal biological functioning or variants implicated in disease states and thus may play a role in cancer development.

Sequencing Quality (5th variant filter)	• Variants within a homopolymer region • Variant base coverage depth ≤ 30x • p-value ≥ 0.01 • Phred score ≥ 20	• Homopolymer region criterion was considered as NGS technologies, including Ion Torrent sequencing technologies, have been reported to produce false positives, especially Indel variants within or near homopolymer regions (Poulet et al., 2016; Yeo et al., 2014). • Three sequencing metrics which assessed the quality of the region sequenced (base coverage depth) and the accuracy of a base called (p-value and Phred score).

Variant Prioritisation & Classification

In order to ascertain the role of the filtered variants on disease onset and progression, variant prioritisation and classification were performed using the following two steps. The first step involved the interpretation and classification of filtered variants according to the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) guidelines (Richards et al., 2015). The guidelines act as a set of rules to guide you in the interpretation and weighting of available variant evidence. The interpretation of variant evidence using the ACMG-AMP criteria facilitated the assignment of variant class of the filtered variant set in this study based on the following five-tier system: Benign, Likely Benign, Uncertain Significance or Variant of Uncertain Significance (VUS), Likely Pathogenic and Pathogenic (Richards et al., 2015). Variant evidence incorporated into these classifications included data from population and public databases, computational data from *in silico* pathogenicity and conservation prediction tools, functional studies data, genetic pedigrees indicating family history, and phenotypic data (Richards et al., 2015).

Each variant which remained after filtering was assessed using the ACMG-AMP guidelines with the assistance of the bioinformatics tool VarSome.com (Saphetor, Lausanne, Switzerland). VarSome.com (which will be referred to as VarSome from here onwards) acts as a search engine for human genomic variants, whereby users can search for evidence regarding their input genomic variant (Kopanos et al., 2018). VarSome imports data from multiple databases into one central location, making the variant data easily available and ready to interpret (Kopanos et al., 2018). Of interest here VarSome applies the ACMG-AMP criteria for the automated classification of sequence variants according to the recommended five-tier classification system - See Appendix C4 for the ACMG-AMP guidelines used for variant classification. Automated classification of variants using the ACMG-AMP guidelines is possible through VarSome's integration of population data, computational data, the use of specific terms such as 'hot-spot' and 'well-known', the use of statistically significant thresholds, and guidance from clinicians and the VarSome user community (Kopanos et al., 2018). We utilised VarSome's recommendations regarding the

ACMG-AMP code strengths, and where variant evidence justified, upgraded or downgraded code strengths of the rules applied. After each variant received a classification based on this five-tier classification system, any variant that was classified as benign or likely benign was excluded from further analysis. Variants which were classified as VUSs were assessed for further evidence which may reclassify them as likely pathogenic variants.

2.2.5 Validation

Validation of the prioritised putative disease-causing/contributing variants was performed using Sanger sequencing technology. Although NGS technologies have been well implemented in diagnostic labs around the world, Sanger sequencing is still required to maintain the highest level of specificity and sensitivity in the process of variant confirmation (Mu et al., 2016). Validation of the prioritised variants via the Sanger sequencing technology described consisted of five main steps, namely, primer design, PCR optimisation, PCR amplification of the patient DNA, Sanger sequencing on the 3130 xL Genetic Analyzer (Applied Biosystems by Thermo Fisher Scientific, CA, USA) and analysis of the Sanger output on the BioEdit software (Hall, 1999). See Figure 2.9 for the validation via the Sanger sequencing workflow. It is important to note that the patient DNA samples used in validation were extracted from the same whole blood patient samples, but were not the same ones used in NGS, thereby ensuring no sample swopping occurred.

Primer design involved designing short oligonucleotide sequences capable of annealing to the 5'-3' end and 3'-5' end around the variant of interest – see Figure 2.9. See Appendix C5.1 for the list and URLs of websites and *in-silico* tools mentioned in designing primer pairs and Appendix C5.2 for the final primer designs utilised in validating the prioritised variants. Completed primer designs were submitted, ordered, purchased, and manufactured by Inqaba Biotec™ (Inqaba Biotechnical Industries (Pty) Ltd, PTA, SA).

PCR optimisation was performed via a temperature gradient PCR in order to determine the optimal annealing temperature of the specific primer-pair –See Appendix C5.3 for the standard thermal cycling conditions followed when performing a temperature gradient PCR. These PCR products were then migrated through a 2% Agarose Gel Electrophoresis and visualised and captured under UV light in the Omega Fluor™ Gel Documentation System (Vacutec, JHB, South Africa). The region of interest was then PCR amplified in the patients' DNA using specific primer pairs at the optimal annealing temperature – see Appendix C5.4 for optimal primer-specific PCR temperatures utilised in validation. The successful amplification of the region of interest in the patient's DNA

was confirmed through Gel Electrophoresis and the Omega Fluor™ Gel Documentation System (Vacutec, JHB, South Africa).

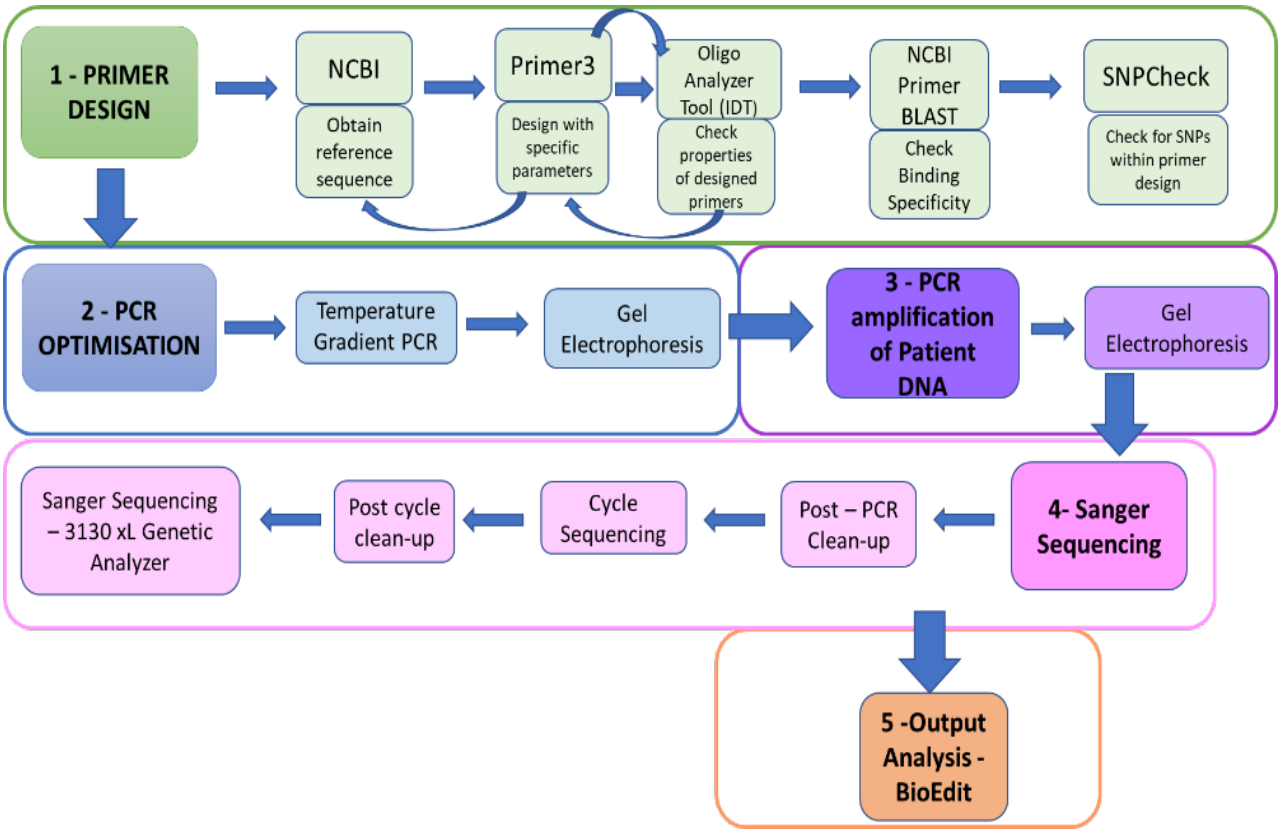


Figure 2.9 – Sanger Sequencing Workflow: Diagram illustrating the methodology followed for validation via Sanger sequencing. The five main stages of the Sanger sequencing workflow are numbered 1-5 and flow according to the larger arrows. The smaller steps followed within each of the five main stages are illustrated within the larger boxes and flow according to the smaller horizontal arrows.

The resultant PCR products and double-distilled H_2O (ddH_2O) were then transferred into MultiScreen-PCRμ96 Filter plates (MilliporeSigma, MA, USA) and used together with the Millivac-Maxi Vacuum pump, 230V (MilliporeSigma, MA, USA) for purification – see Figure .0

2.9. The purified PCR products were then cycle sequenced via the ddNTP chain termination method using the Big Dye™ v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, MA, USA). See Appendix C5.5 for the thermal cycling conditions followed during Cycle Sequencing. Prior to Sanger sequencing, the amplified PCR products were purified once again using the MultiScreen-PCRμ96 Filter plates, the Millivac-Maxi Vacuum pump 230V and the Millipore injection solution. The cycle sequencing products were heat-denatured to single-stranded amplicons before loading the samples into the 3130 xL Genetic Analyzer for capillary electrophoresis. The 3130 xL Genetic Analyzer uses capillary electrophoresis to separate the fluorescently tagged amplicons by size. These fluorescently tagged ddNTPs are detected through laser excitation and translated into a nucleotide sequence. The resultant electropherograms produced by the 3130 xL Genetic Analyzer were further analysed for the presence of the variant

of interest using BioEdit (Hall, 1999), a freely available software for the alignment and analysis of biological sequences.

2.2.6 Phenotype – Genotype Relationship

The relationship between a patient's clinical phenotype and genotype was further assessed in all cases where a pathogenic or likely pathogenic variant was identified, especially patients whose phenotype indicated the presence of a CPS. Phenotypic information was obtained by the PeCan's Medical Geneticists during the recruitment process. This included the type of cancer found within the proband, a full clinical examination of the patient, medical history of the patient, and a genetic pedigree indicating history and penetrance of specific cancers within the family. The likely role of the variant in the patient's presenting phenotype and cancer development was determined through VarSome (via UniProt and ClinVar entries), Mastermind®, and a literature search and review of any studies or medical reports documenting the association of the variant/gene that the variant lies in and the development of a particular phenotype and/or development of that patient's particular cancer.

Mastermind® is a genomic search engine created by Genomenon®, containing the most comprehensive database of genomic literature (https://www.genomenon.com/mastermind). Mastermind® facilitates a comprehensive understanding of the relationships between disease, genes, variants, phenotypes, and therapies through a thorough genomic literature search and subsequent identification of literature documenting these relationships (Chunn et al., 2020). Articles catalogued in Mastermind® relevant to the mutation-positive CPGs in this study were reviewed for evidence regarding the clinical phenotypes and cancer types associated with the variant of interest. Where VarSome's UniProt, ClinVar and Mastermind® entries did not identify any clinical evidence supporting the relationship between the patient's particular phenotype/cancer type and identified variant, the scientific and medical literature was searched and reviewed further using Google Scholar (scholar.google.com), PubMed (www.ncbi.nlm.nih.gov/pubmed) and ScienceDirect (http://www.sciencedirect.com/science/search).

3. Results

This chapter provides the NGS results generated, and their interpretation in identifying any pathogenic/likely pathogenic germline variants in this project's cohort of South African paediatric cancer patients.

3.1 Patient Demographics

Thirty-two South African paediatric patients diagnosed with solid non-haematological malignancies were recruited from the paediatric oncology units at Charlotte Maxeke Johannesburg Academic Hospital and Chris Hani Baragwanath Academic Hospital. Figure 3.1 (A-E), provides a graphic summary of the patient demographics, including the proportions of sex, ethnicity, cancer types, individuals with/without a family history of cancer, and age at diagnosis. Details of the patient demographics for all 32 participants can be seen in Appendix D1.

3.2 Next-Generation Sequencing Data Quality Control

Two sequencing runs were performed using the Ion GeneStudio™ S5 system (Thermo Fisher, MA, USA) resulting in sequencing data generation for a total of 32 individual patient DNA samples.

3.2.1 Pre-Alignment Metrics

The quality of library preparation, template preparation, and chip loading was determined through the evaluation of specific pre-alignment metrics. Library and template preparation was assessed using the pre-alignment ISP and read statistics seen in Table 3.1 and Figures 3.2 A (ii) & B (ii). The quality of chip-loading was assessed by evaluating the % ISP loading and the ISP Density heat map. The ISP Density heat map seen in Figure 3.2 A (i) & B (i) shows a visual representation of the automated well loading distribution on the first and second sequencing runs chips' surface, respectively. Figure 3.3 (A & B) illustrates the read length histogram of all the trimmed read lengths present in the output files of the first and second sequencing runs, respectively. The mean, median and modal read lengths of the first and second sequencing runs are extremely similar and occur within the read length range expected (~200 – 225 bp) based on our gene panel design.

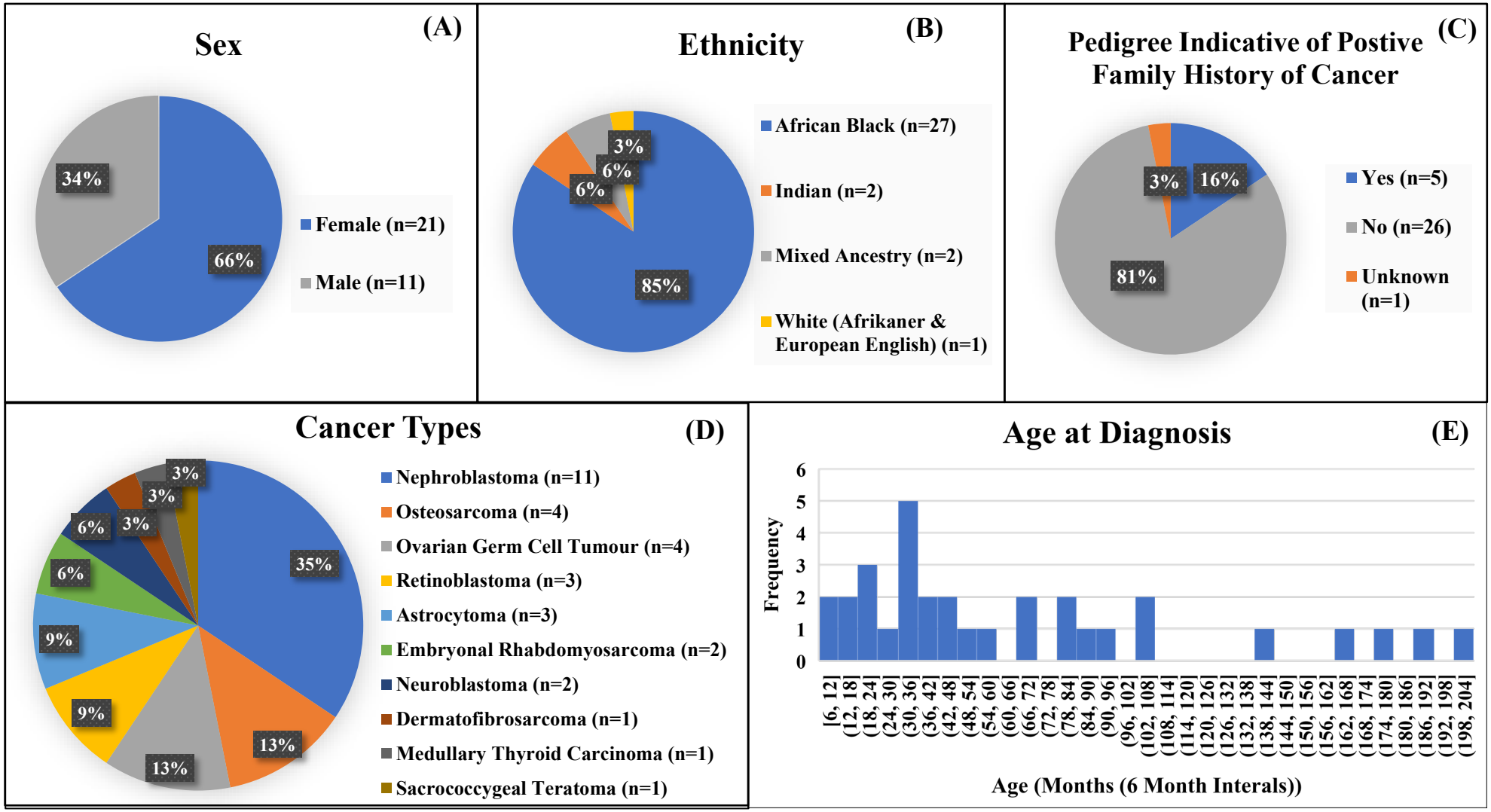


Figure 3.1 – Pie Charts and Column Graph Depicting the Summary of the 32 Patient Demographics (A-E): **(A)** Pie chart depicting the Sex distribution of the patient subset. **(B)** Pie chart depicting the Ethnicity distribution of the patient subset. **(C)** Pie chart depicting which patients have pedigrees indicating a positive family history of cancer**. (D)** The proportions of Cancer Types within this patient subset. **(E)** Age at Diagnosis (in Months) of this patient subset.

In summary of the pre-alignment metrics for the first sequencing run, although the combined Library A was over-diluted and led to a lower number of wells with ISPs, the run still generated a sufficient % of usable reads, a low proportion of polyclonal reads (27%), and the low-quality reads being <20% (19%). This deemed the pre-alignment sequencing data for the first sequencing run to be of sufficient quality and quantity data to move forward to alignment and further downstream analysis.

In summary of the pre-alignment metrics for the second sequencing run, although the combined Library B concentration was too high thus leading to a high number of polyclonal reads, the sequencing run still generated a sufficient % of usable reads, and the low-quality reads were <20% (%11). Thus, deeming the sequencing data generated in the second sequencing run of sufficient quality and quantity for further alignment and downstream analysis.

Table 3.1 – Pre-Alignment Statistics Generated for the First and Second Sequencing Runs: Table summarising the pre-alignment statistics used in assessing the quality and quantity of data generated for the first and second sequencing runs.

	First Sequencing Run	**Second Sequencing Run**
Addressable Wells	12, 530, 194	12, 530, 194
Number of Wells with ISPs (% of total addressable wells)	7, 508, 021 (59.9%)	11, 777, 544 (94%)
Number of Wells with Live ISPs (% of the number of wells with ISPs)	7, 496, 795 (99.9%)	11, 777, 544 (100%)
Total Reads	4, 076, 218	6, 082, 089
% Usable Reads	57% (57% of 4, 076, 218 = ~ 2, 323, 444 reads)	52% (52% of 6,082, 089 = ~3, 162, 686 reads)
% Polyclonal Reads	27%	41%
% Low-Quality Reads	19%	11%

Acronyms/Abbreviations/Symbols – ISPs: Ion Sphere Particles, %: Percentage

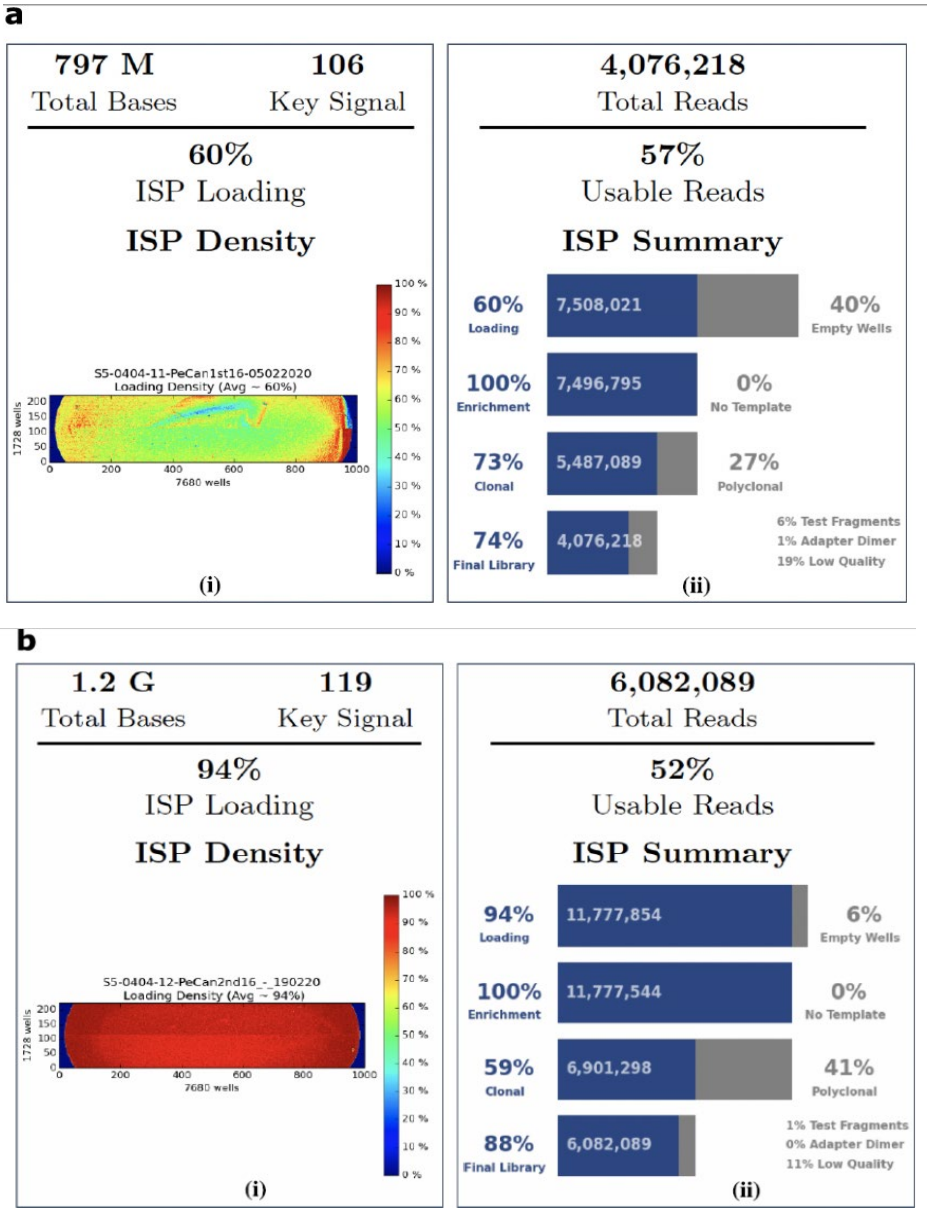


Figure 3.2 - ISP Loading, ISP Density & ISP Summary Generated for the First (A) and Second (B) Sequencing Runs: **(A) (i)** ISP Loading of 60%, indicating that ~60% of the sequencing chip's total well capacity was addressable. **(ii)** This figure provides an ISP summary. After filtering out the polyclonal and low-quality reads, 57% of the total reads were deemed as usable. **(B) (i)** ISP Loading of 94%, indicating that ~94% of the sequencing chip's total well capacity was addressable **(ii)** This figure provides an ISP summary. After filtering out the polyclonal and low-quality reads, 52% of the total reads were deemed as usable. Acronyms/Symbols – ISP: Ion Sphere Particles, %: Percentage

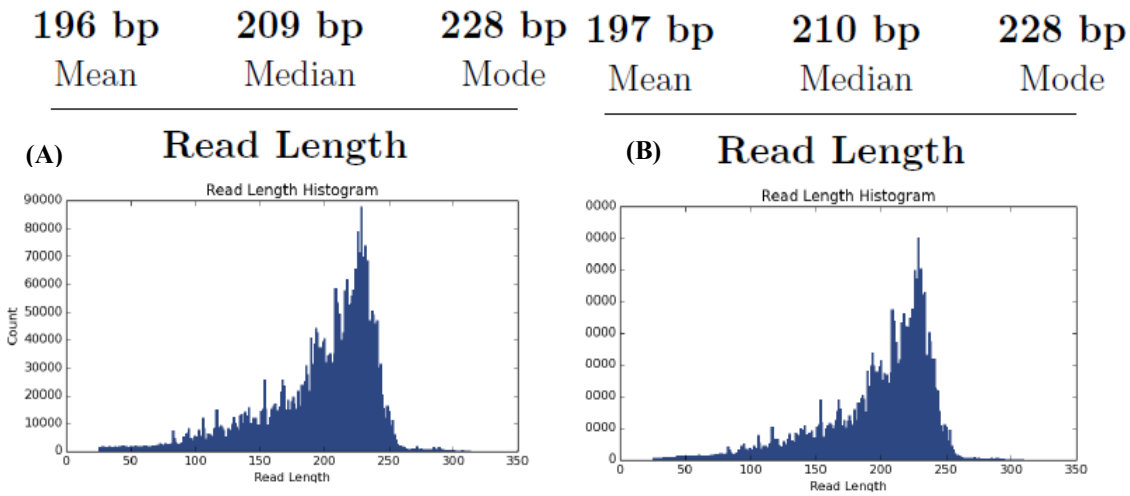
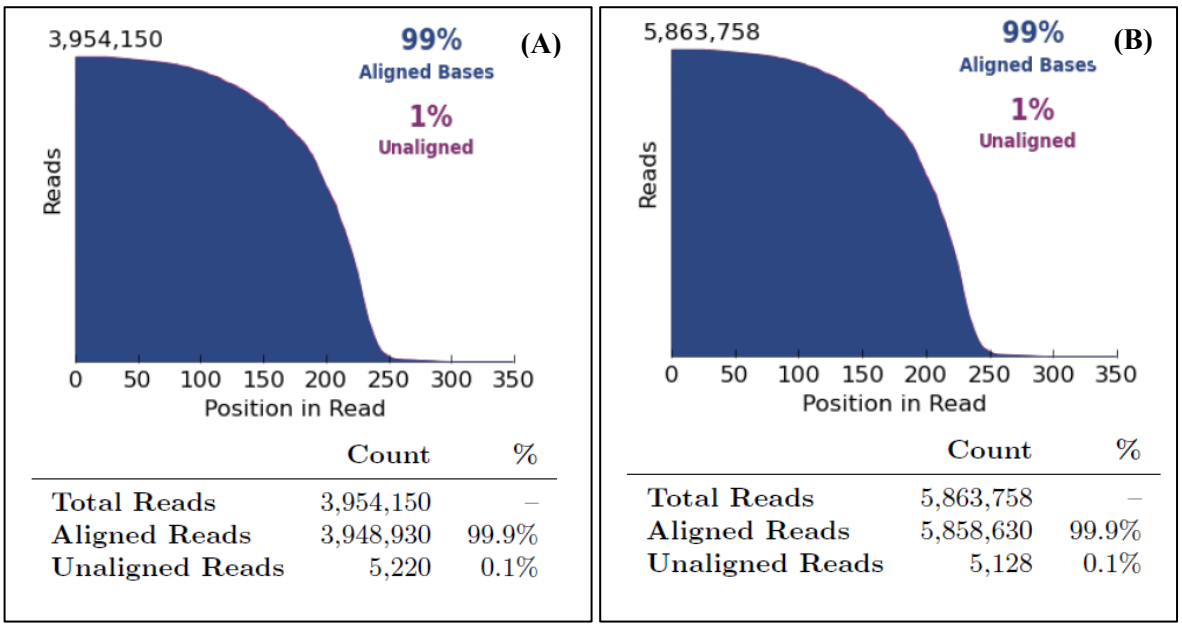


Figure 3.3 – Read Length Histogram of the Trimmed Lengths of all the Reads Present in the First (A) & Second (B) Sequencing Runs' Output Files: **(A & B)** The X-axis represents the read lengths (in base pairs) and the Y-axis represents the count number. The mean, median, and modal trimmed read lengths (in base pairs) for the first and second sequencing runs are shown above the read length histograms and fall within the expected read length range for the gene panels design (~200 – 225 bp). Acronyms – bp: Base Pairs

3.2.2 Post-Alignment Metrics

The quantity and quality of the sequencing data were further assessed using post-alignment metrics, as well as the results of the coverageAnalysis and variantCaller Ion Torrent plugins. The first (Figure 3.4 (A)) and second (Figure 3.4 (B)) sequencing run reports generated quality and quantity metrics of the generated sequencing data after the reads had been mapped and aligned to the designed gene panel's targeted reference sequences. These metrics included the total number of reads generated after pre-alignment filtering, the number and % of aligned and unaligned reads (out of the total number of reads after pre-alignment filtering), and an alignment plot – showing the number of aligned and unaligned reads by position in an aligned sequence.



	Count	%
Total Reads	3,954,150	–
Aligned Reads	3,948,930	99.9%
Unaligned Reads	5,220	0.1%

	Count	%
Total Reads	5,863,758	–
Aligned Reads	5,858,630	99.9%
Unaligned Reads	5,128	0.1%

Figure 3.4 - First (A) and Second (B) Sequencing runs' Post-Alignment Sequencing Quality Metrics:
(A & B) The alignment plots show the number of total aligned reads (~99% - shown in blue) and unaligned reads (<1% - shown by a small purple line) by position in an aligned sequence. Below the alignment plot is the total number of reads generated after filtering, and the count and % of aligned and unaligned reads (of the total reads).

The Ion Torrent coverageAnalysis and variantCaller plugins generate sequencing metrics on a per barcode basis (i.e., per sample). The variantCaller plugin was run to get a preliminary view of the variants identified in the patient samples, this plugin was run in the Ion Torrent Suite™ software (v.5.12.0) before patient VCFs were exported to the Ion Reporter and VEP annotation software and does not necessarily reflect the number of variants called by either annotation software. The variantCaller plugin was run using the Ion Torrent Ampliseq library and the germline variant frequency settings. See Table 3.2 for the averages and ranges of the first and second sequencing run quality metrics generated by the coverageAnalysis and variantCaller plugins. The post-run quality metrics generated by the coverageAnalysis and variantCaller plugins. The post-alignment sequencing quality metrics per sample can be seen in Appendix D2.

Although the second sequencing run generated a greater quantity and quality of sequencing data, the alignment and coverage quality metrics for both sequencing runs showed that both runs generated sufficient reads (both overall and per sample) and that each sample had a high enough mean base coverage depth. The overall alignment and coverage quality metrics indicated that the sequencing data generated for both sequencing runs were of good enough quality to commence to further bioinformatic analysis.

Table 3.2 - Post-Alignment Sequencing Quality Metrics Generated by the Ion Torrent coverageAnalysis and variantCaller Plugins: Table summarising the averages and ranges of the first and second sequencing runs quality metrics generated by the coverageAnalysis and variantCaller plugins. These include the total number of reads mapped to the reference sequence (per sample), the % reads mapped to the targeted regions, the mean base depth coverage, the % coverage uniformity (coverageAnalysis plugin), and the number of variants (SNVs & Indels) called by the variantCaller plugin.

	First Sequencing Run		Second Sequencing Run	
	Average	**Range**	**Average**	**Range**
coverageAnalysis Plugin				
Mapped Reads	246, 808	207, 977 - 279, 065	366, 039	211, 811 - 418, 540
On Target (%)	92,73	91,62 - 93,31	94,55	93,83 - 94,87
Mean Base Coverage Depth	182,6	153,9 - 205,9	266,2	153,7 - 306,0
Coverage Uniformity (%)	95,94	95,10 - 96,70	93,55	88,20 - 95,20
VariantCaller Plugin				
Number of Variants Detected by Ion Torrent Suite™ Software*	90	71-104	215	178 - 250

*: Indicates the number of variants called by the variantCaller in the Ion Torrent Suite™ software (v.5.12.0), before patient VCFs were exported to the Ion Reporter and VEP annotation software programs and does not necessarily reflect the number of variants called per sample by these annotation software programs. Acronyms/Abbreviations/Symbols - ID: Identification (number), PeCan: Paediatric Cancer (Project).

3.3 Bioinformatics – Interpretation of NGS Data Output

3.3.1 Variant Filtering Output

The merged SNVs, Indels and Complex MNV VEP data files underwent five data filtering steps, each interpreting multiple lines of variant evidence individually. This allowed for a more in-depth interpretation of the variants remaining after filtering, allowing for the identification of putative disease-causing/contributing variants– See Figure 3.5.

Together the merged SNVs, and Indels and Complex MNV files contained ~1942 variants (all canonical CCDS transcript). Approximately 83% of these variants were excluded after applying the variant effect filter, which filtered out any variants predicted to have no effect on protein structure and function (This left a total of ~ 313 variants – Figure 3.5). The second variant filter applied was based on the scores and interpretations provided by *in-silico* pathogenicity and conservation tools. See Appendix C3 for the list of *in-silico* tools and their interpretation used for the filtering of SNVs, Indels and complex MNVs. The third and fourth variant filters removed any variants which occurred in more than 20% of the patient subset and any variants that were reported

to have a benign or likely benign clinical significance as predicted by ClinVar. These three filters excluded a further ~82% of the variants which remained after the first filter. The fifth variant filter assessed the sequencing quality of the remaining 57 variants, excluding variants which were close to or within homopolymer regions, as well as those variants which failed other sequencing quality criteria. A total of 26 variants remained after all variant filters were applied – Figure 3.5

3.3.2 Variant Prioritisation & Classification Output

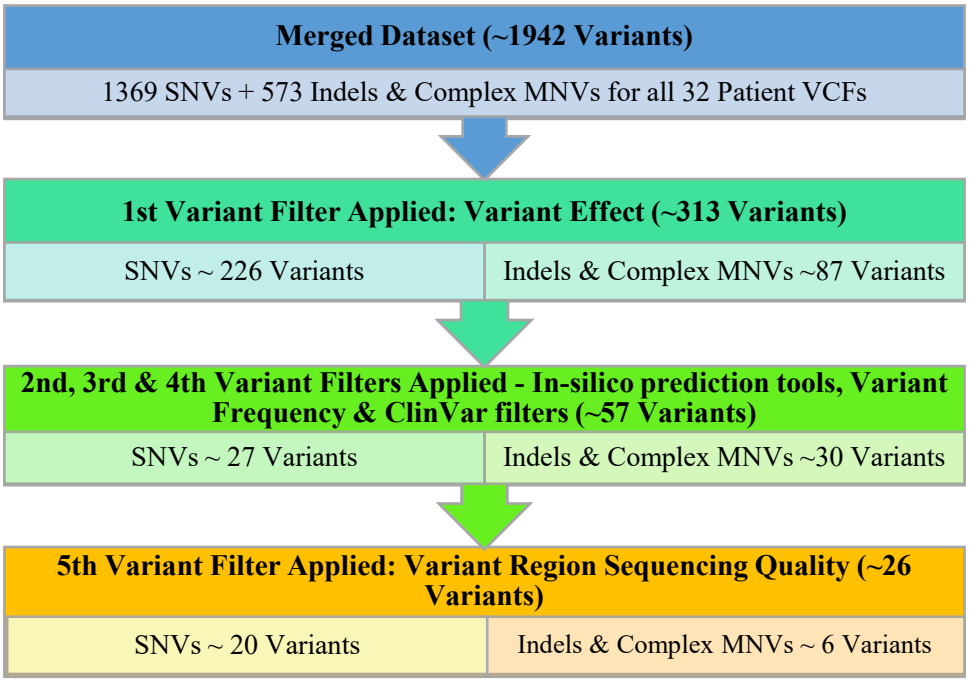


Figure 3.5 – Results of Variant Filtering: Variant filtering pipeline followed in the identification of putative disease-causing/contributing variants in this patient subset. Each level indicates the application of the five variant filters and the number of variants remaining after applying that filter (in brackets). Acronyms/Abbreviations/Symbols - CCDS: Consensus Coding Sequence Set, Indels: Insertions & Deletions, MNVs: Multi-Nucleotide Variants, SNVs: Single Nucleotide Variants, VEP: Variant Effect Predictor

Variant prioritisation and classification of the variants remaining after filtering assessed multiple lines of variant evidence simultaneously, allowing a holistic view of the variant's potential role in disease onset and progression. Variant prioritisation and classification involved the interpretation and classification of the remaining 26 variants according to the ACMG-AMP guidelines (Richards et al., 2015), with the assistance of the bioinformatics tool VarSome.com (Saphetor, Lausanne, Switzerland) (Kopanos et al., 2018). The 26 variants which remained after filtering were classified according to the five-tier system: Benign, Likely Benign, Uncertain Significance or Variant of Uncertain Significance (VUS), Likely Pathogenic and Pathogenic (Richards et al., 2015) – See Appendix C4 for the ACMG-AMP guidelines used for variant classification. Five of these variants

were classified as Pathogenic, two variants were classified as Likely Pathogenic, 14 were classified as VUSs, one was classified as Likely Benign, and four variants were classified as Benign. All Likely Benign and Benign variants were excluded from further analysis. The 14 variants classified as VUSs were further assessed for any *in-silico* and/or clinical criteria which may reclassify them as Likely Pathogenic. However, no additional criteria were identified that supported this reclassification, and thus the 14 VUSs were excluded from further assessment. With the exclusion of all Benign, Likely Benign and VUSs, a total of seven variants were prioritised for validation.

3.4 NGS Prioritised Variants

3.4.1 Putative Disease-Causing/Contributing Variants

This section addresses the seven prioritised variants, providing all the lines of variant evidence which led to their prioritisation and classification as pathogenic or likely pathogenic. Each variant will be addressed individually and the evidence includes one/more of the following criteria: adequate sequencing quality, the variant's predicted effect on protein function and/or structure as predicted by *in-silico* pathogenicity and conservation tools, the variant's allele frequency reported in population databases as well as the variant frequency in this patient subset, any clinical and/or family history data which may suggest the presence of a heritable/familial CPS, any clinical evidence previously reported (as reported by ClinVar/HGMD/UniProt), variant evidence which supported an ACMG-AMP Pathogenic or Likely Pathogenic prediction, and any evidence supporting the gene to patient's cancer relationship and /or CPS association. Results of variant validation, including primer design, PCR optimisation of primer pairs, PCR amplification of patient DNA (Appendix D3), and electropherograms are also shown, where applicable. Some of the variant evidence has been grouped and described in a single table/diagram. This includes the sequencing quality (Table 3.3) and the *in-silico* pathogenicity and conservation tools results (Table 3.4 & 3.5). Ion Reporter™ Genomic Viewer (IRGV) images depicting the coverage of the seven prioritised variants can be seen in Appendix D4. IRGV is a genomics viewer built into the Ion Reporter software, based on the Integrative Genomics Viewer (IGV) developed by the Broad Institute (Robinson et al., 2017).

Table 3.3 Sequencing Quality Data for the Seven Prioritised Variants: Table describing sequencing quality data assessed for the second set of seven prioritised variants. As seen all seven variants have allele ratios close to a true heterozygous call (~0.5:0.5).

PeCan Project Patient ID	Gene	Variant Coding Change	Variant Protein Change	Allele Coverage	Allele Count	Allele Ratio	Phred Score	p-value
PC.0023.01.1	*NF1*	c.1133_1136delTGAC	p.D378Afs*8	150x	TTGAC=75x T=75x	TTGAC=0.50 T=0.50	346.70	0.00001
PC.0032.01.1	*RET*	c.1901G>A	p.C634Y	290x	G=134x A=156x	G=0.46 A=0.54	768.13	0.00001
PC.0051.01.1	*TP53*	c.817C>T	p.R273C	95x	C=44 T=51	C=0.46 T=0.54	254.66	0.00001
PC.0011.01.1	*BRCA1*	c.4199T>C	p.M1400T	257x	T=108x C=149x	T=0.42 C=0.58	794.56	0.00001
PC.0009.01.1	*FAH*	c.485G>A	p.R162H	514x	G=258x A=256x	G=0.51 A=0.49	1144.45	0.00001
PC.0003.01.1	*ERCC3*	c.958C>T	p.Q320*	243x	C=121x T=122x	C=0.50 T=0.50	556.01	0.00001
PC.0018.01.1	*RB1*	c.1981C>T	p.R661W	102x	C=54x T=48x	C=0.53 T=0.47	210.55	0.00001

Abbreviations/Acronyms/Symbols – ID: Identification (number), PeCan: Paediatric Cancer (Project).

Table 3.4 - In-silico Pathogenicity & Conservation Prediction Tool Results for the Prioritised Indel & MNV Variants: Table showing the *in-silico* pathogenicity and conservation tool results for one/seven prioritised variants, categorised under the Indels & MNVs category. This variant passed the pathogenicity and conservation thresholds set when assessing these variant types, therefore indicating the variant change occurs in a conserved region and likely has a damaging effect on protein structure and/or function.

PeCan Project Patient ID	Gene	Variant Coding Change	Variant Protein Change	CADD-PHRED Scaled Score	PhyloP100Way	GERP RS	SIFT-INDEL Confidence Score	SIFT-INDEL Prediction
PC.0023.01.1	*NF1*	c.1133_1136delTGAC	p.D378Afs*8	32	7.66	5.46	0.858	Damaging

Abbreviations/Acronyms/Symbols – CADD: Combined Annotation Dependent Depletion, GERP RS: Genomic Evolutionary Rate Profiling Rejected Substitutions, ID: Identification (Number), PeCan: Paediatric Cancer (Project), PhyloP: Phylogenetic P-values, SIFT-INDEL: Sorting Intolerant From Intolerant – Insertions & Deletions.

Table 3.5 - In-silico Pathogenicity & Conservation Prediction Tool Results for the Prioritised SNV Variants: Table showing the *in-silico* pathogenicity and conservation tool results for six/seven of the second set of prioritised variants, categorised under the SNVs category. Shown below are all *in-silico* tools which were selected to assess missense and nonsense SNVs. All six variants passed the pathogenicity and conservation thresholds set when assessing these variants, thus indicating the variant change occurs in a conserved region and likely has a damaging effect on protein structure and/or function.

PeCan Project Patient ID						
	PC.0032.01.1	**PC.0051.01.1**	**PC.0011.01.1**	**PC.0009.01.1**	**PC.0003.01.1**	**PC.0018.01.1**
Gene	*RET*	*TP53*	*BRCA1*	*FAH*	*ERCC3*	*RB1*
Variant Coding Change	c.1901G>A	c.817C>T	c.4199T>C	c.485G>A	c.958C>T	c.1981C>T
Variant Protein Change	p.C634Y	p.R273C	p.M1400T	p.R162H	p.Q320*	p.R661W
Variant Type	Missense	Missense	Missense	Missense	Stop-Gain (Nonsense)	Missense
CADD - PHRED Scaled Score	23	32	28	34	38	35
***FATHMM / FATHMM-MKL Score & Prediction**	-5.79 (Damaging)	-7.34 (Damaging)	-3.15 (Damaging)	-3.61 (Damaging)	0.99 (Damaging)	-3.72 (Damaging)
GERP RS	4.52	4.92	5.35	5.41	5.50	5.70
LRT Score	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
LRT Prediction	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious
MetaLR Score	0.976	0.993	0.820	0.953	NA	0.949
MetaLR Prediction	Damaging	Damaging	Damaging	Damaging	NA	Damaging

MutationAssessor Rankscore & Prediction	0.801 (M = Medium Functional Impact	0.883 (M = Medium Functional Impact)	0.719 (M = Medium Functional Impact)	0.991 (H = High Functional Impact)	NA	0.913 (M = Medium Functional Impact)
MutationTaster Converted Rankscore	0.803	0.810	0.810	0.810	0.813	0.813
MutationTaster Prediction	D = Disease-Causing	D = Disease-Causing	D = Disease-Causing	D = Disease-Causing	A = Automatic Disease-Causing	A = Automatic Disease-Causing
phyloP100way	9.62	4.54	5.83	8.70	9.79	4.60
PolyPhen Score	0.999	0.999	0.999	0.999	NA	1.000
PolyPhen Prediction	Probably Damaging	Probably Damaging	Probably Damaging	Probably Damaging	NA	Probably Damaging
PROVEAN Score	-7.34	-7.38	-2.91	-4.97	NA	-7.17
PROVEAN Prediction	Deleterious	Deleterious	Deleterious	Deleterious	NA	Deleterious
SIFT Score	0.001	0.002	0.002	0.000	NA	0.000
SIFT Prediction	Deleterious	Deleterious	Deleterious	Deleterious	NA	Deleterious

NA = Not Applicable, shown in cases where the tool produced a blank output as it was not capable of assessing that specific variant type. In this case, MetaLR, MutationAssessor, PolyPhen, PROVEAN and SIFT were unable to provide a prediction for the *ERCC3* c.958C>T (p.Q320*) stop-gain (nonsense) variant. In addition, FATHMM-MKL was used instead of FATHMM when assessing the *ERCC3* c.958C>T (p.Q320*) stop-gain (nonsense) variant. Abbreviations/Acronyms/Symbols – CADD: Combined Annotation Dependent Depletion, FATHMM-MKL: Functional Analysis Through Hidden Markov Models – Multiple Kernel Learning, GERP RS: Genomic Evolutionary Rate Profiling Rejected Substitutions, ID: Identification (Number), LRT: Likelihood Ratio Test, PeCan: Paediatric Cancer (Project), phyloP: phylogenetic P-values, PolyPhen: Polymorphism Phenotyping, PROVEAN: Protein Variation Effect Analyser, SIFT: Sorting Intolerant From Tolerant.

PC.0023.01.1 – *NF1* (NM_001042492.3): c.1133_1136delTGAC (p.D378Afs*8)

Patient PC.0023.01.1 is a White South African male who was diagnosed with a Pilocytic Astrocytoma at 9 years old. This patient previously received a clinical diagnosis of Neurofibromatosis Type 1 (NF1) as he presented with phenotypic features, a positive family history suggesting NF1 inheritance, and cancer type all characteristic of the syndrome. The reported family history of the patient indicated that the proband's father had been previously diagnosed with cancer and presented with a similar NF1-related phenotype to the proband (see Figure 3.6).

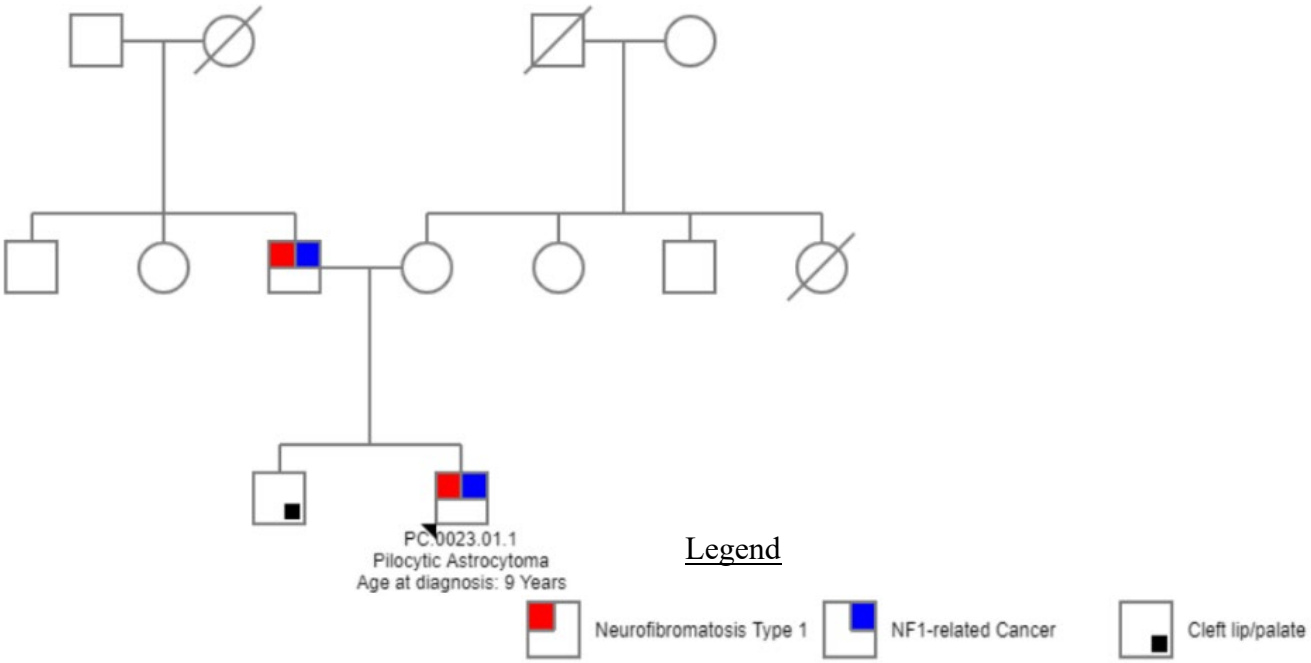


Figure 3.6 – Patient PC.0023.01.1 Genetic Pedigree: Pedigree of patient PC.0023.01.1 depicting the family history of Neurofibromatosis Type 1 and NF1-related cancers. The heterozygous *NF1* c.1133_1136delTGAC (p.Asp378AlafsTer8) variant was identified in the proband (PC.0023.01.1) who was diagnosed with a pilocytic astrocytoma at age 9 years.

A variant in the Neurofibromin 1 gene (*NF1)* (NM_001042492.3): c.1133_1136delTGAC (p.D378Afs*8) was identified as the likely disease-causing variant related to the patient's cancer type. The variant showed adequate sequencing quality (see Table 3.3). The variant is heterozygous with an allele ratio of exactly 0.5:0.5. The coverage of this variant can be seen in the IRGV image in Appendix D4.

Variant *NF1* c.1133_1136delTGAC was classified as pathogenic according to the ACMG-AMP codes: PVS1, PM2, PP3 and PP4. See Table 3.6 for the variant evidence and reasoning used to apply these ACMG-AMP rules. The Mutalyzer software version 2.0.35 (https://mutalyzer.nl/) was used to predict and visualise the variant's effect on the *NF1* reference protein (Lefter et al., 2021). The four base-pair deletion is predicted to result in a frameshift whereby the first amino acid residue affected Aspartate 378 (Asp378) is replaced by an Alanine residue (Asp378Ala). Subsequently the frameshift results in a truncated protein due to the position of a translation termination codon in the new reading frame (Ter8/*8) – see Mutalyzer Figure 3.7 (A & B).



Figure 3.7 – Mutalyzer 2.0.35 Output for Variant NF1 (NM_001042492.3): c.1133_1136delTGAC (p.D378Afs*8) Identified in Patient PC.0023.01.1 (A & B): **(A)** Mutalyzer image depicting the effect of the variant's four bp deletion c.1133_1136delTGAC on the reference protein. Black amino acid residues occur before the four bp deletion and are unaffected by the frameshift. Red amino acid residues are those which have shifted and are affected by the four bp deletion. **(B)** Mutalyzer image depicting the predicted protein as a result of the four bp deletion. A truncated protein is produced due to a translation termination codon in the new reading frame (fsTer8/*8) indicated by an * in the figure.

The *NF1* c.1133_1136delTGAC variant has not been reported in ClinVar, HGMD, and UniProt. Mastermind® identified five articles containing the *NF1*_p.D378Afs*8 variant (Pasmant et al., 2015; Sabbagh et al., 2013; Laycock-Van Spyk et al., 2011; Pros et al., 2008; Upadhyaya et al., 2008). Although the literature evidence identified for this variant further supports the genotype-phenotype relationship of patient PC.00023.01.1, additional clinical evidence from ClinVar/HGMD/UniProt is needed to apply the ACMG-AMP rule PP5.

All five stages of validation via Sanger sequencing were successful for this variant. Figure 3.8 depicts the electropherogram output of the forward strand of the patient DNA sequenced via Sanger sequencing, confirming the presence of the heterozygous pathogenic variant c.1133_1136delTGAC in patient PC.0023.01.1. As the variant is heterozygous, the reference sequence (TGAC) is present in the one strand, while deleted from the other strand, resulting in a frameshift in the variant strand only. The heterozygous nature of this variant is depicted by reduced peak sizes and multiple base calls beyond the point of deletion.

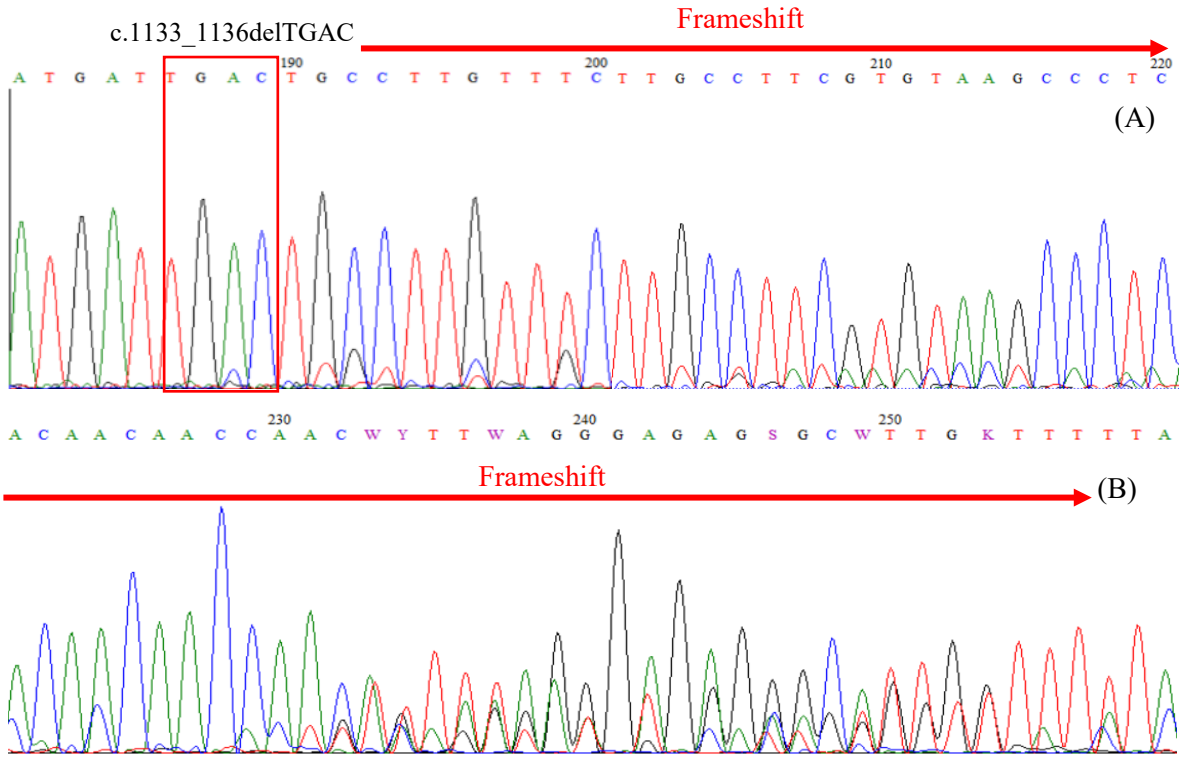


Figure 3.8 – Electropherogram Confirming Variant NF1 c.1133_1136delTGAC in Patient PC.0023.01.1 (A & B): **(A)** Forward strand of patient PC.0023.01.1 confirming the presence of the *NF1* c.1133_1136 deletion (TGAC highlighted in red box) in one strand and the presence of the reference alleles in the other strand, confirming the heterozygous nature of this variant. **(B)** Electropherogram result of sequence produced after the point of deletion. The heterozygous deletion results in a shift in the reading frame of the remaining bases in one strand, whilst the strand without the variant is read like the reference strand. This is depicted through multiple base calls at various points throughout the remaining sequence and reduced peaks after the point of deletion.

Table 3.6 – ACMG-AMP Rules Applied to Variant NF1_p.D378Afs*8: This table summarises the ACMG-AMP rules applied to variant *NF1*_p.D378Afs*8 identified in patient PC.0023.01.1. These rules and their strengths together result in an ACMG-AMP pathogenic classification.

PeCan Patient ID	Cancer Type	Confirmed/Suspected Clinical CPS Diagnosis	Variant Change	ACMG-AMP Classification	ACMG Codes	Reasoning & Evidence
PC.0023.01.1	Pilocytic Astrocytoma	Confirmed Clinical NF1 Diagnosis	*NF1*_p.D378Afs*8	Pathogenic	PVS1 (Very Strong)	A frameshift variant in the *NF1* gene, where LOF is a known mechanism of disease. The NF1 gene has 2213 pathogenic LOF variants and the gnomAD LOF o/e score of 0,215 which is < than the gnomAD recommended threshold of 0,35 and the VarSome recommended threshold of 0,763.
					PM2 (Moderate)	Variant allele frequency is not reported in the control population databases gnomAD, the 1000 Genomes Project, ExAc and the NHLBI GO ESP6500. The variant region is reported to be well covered in gnomAD exomes (80.1x) and gnomAD genomes (31.1x).
					PP3 (Supporting)	Pathogenic computational verdict based on deleterious/damaging predictions from four different *in-silico* tools - CADD, SIFT-INDEL, GERP and phyloP, with no other benign *in-silico* predictions. See Table 3.4
					PP4 (Supporting)	Patient PC.0023.01.1 phenotype: Delayed speech and language development, inguinal freckling, Lisch nodules, multiple cafe-au-lait spots, upslanting palpebral fissures, and the NF1-related cancer Pilocytic Astrocytoma. This phenotype and positive family history of the proband is highly specific to NF1 disease, known to be caused by mutations in the *NF1* gene.

PC.0032.01.1 – *RET* (NM_020975.6): c.1901G>A (p.C634Y)

Patient PC.0032.01.1 is an Indian South African female who was diagnosed with a Medullary Thyroid Carcinoma (MTC) at age 14 years. This patient had been both clinically (through a paediatric oncologist at CMJAH), and molecularly (at NHLS-Braamfontein) diagnosed with Multiple Endocrine Neoplasia Type 2A (MEN2A) caused by a mutation in the *RET* gene. It was decided to recruit and include this patient in this project's patient subset to be used as a 'blind' bioinformatics control patient. The research team was blinded to the exact location and type of variant this patient had during NGS data analysis. This ensured that the same bioinformatics pipeline applied to all patients allowed for the detection of a case with a known variant.

Patient PC.0032.01.1 presented with a positive family history of cancer, which was further suggestive of the presence of a heritable CPS – see Figure 3.9. The proband's paternal grandfather passed away from leukaemia and the paternal grandmother was said to be affected and passed away from a disease relating to the breast and uterine, although an exact cancer diagnosis was not known. The proband's father, although not present at the time of recruitment, is said to have a similar phenotype to that of the patient and was also diagnosed with MEN2A-related cancer. This suggests that the disease-causing variant identified in the proband was likely inherited from the father.

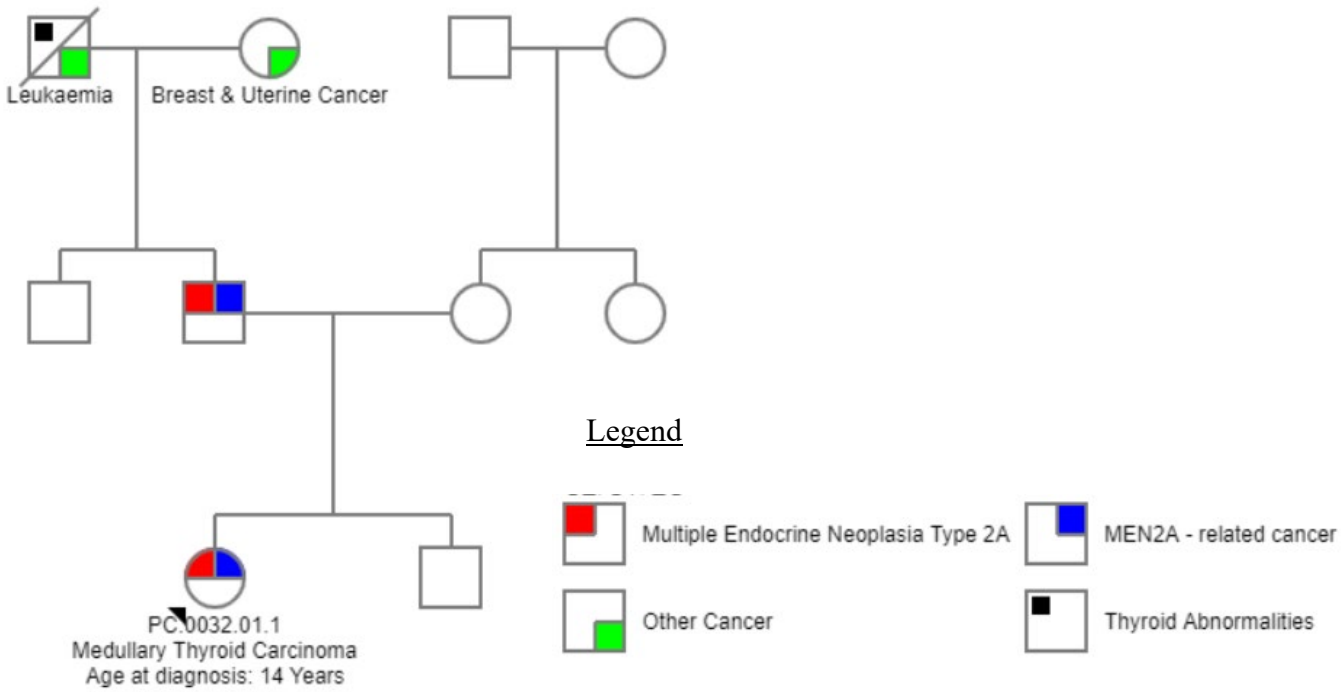


Figure 3.9 – Patient PC.0032.01.1 Genetic Pedigree: Pedigree of patient PC.0032.01.1 depicting the family history of MEN2A-related and other cancer types. The heterozygous *RET* c.1901G>A (p.C634Y) variant was identified in the proband (PC.0032.01.1) who was diagnosed with a medullary thyroid carcinoma at age 14 years.

The *RET* (NM_020975.6): c.1901G>A (p.C634Y) variant was identified as the likely disease-causing variant related to the patient's cancer type. The variant showed adequate sequencing quality (see Table 3.3). The variant is heterozygous with a G:A allele ratio of 0.46:0.54. Coverage of this variant is depicted in the IRGV image in Appendix D4.

The *RET* c.1901G>A is a missense variant resulting in the replacement of the Cysteine 634 amino acid residue with a Tyrosine amino acid residue. The single nucleotide substitution c.1901G>A occurs in exon 11 of the NM_020975.6 *RET* transcript and ultimately results in the expression of a *RET* gene product which differs from that of the reference protein. The *RET* c.1901G>A variant was classified as pathogenic according to the ACMG-AMP codes: PM1, PM2, PM5, PP5, PP3, and PP4. Variant evidence analysed justified increasing the default strength of PM1 and PM2 from moderate to strong, as well as the default strength of PP5 from supporting to strong. See Table 3.7 for the variant evidence and reasoning used to apply these ACMG-AMP rules.

The ACMG-AMP rule PP2 could not be applied as only 289/1981 clinically reported variants in the *RET* gene are pathogenic (289/1981 = 14.6% is lower than the VarSome recommended threshold of 27.5%). The *RET* gene also has an ExAC missense z-score of 1.77 and a gnomAD missense z-score of 0.96, both of which are lower than the z > 3.09 threshold score recommended by the ClinGen Sequence Variation Interpretation Working Group (SVI WG).

All five stages of validation via Sanger sequencing were successful for this variant. The resultant electropherogram of the forward strand for patient PC.0032.01.1 confirms the presence of the heterozygous RET c.1901G>A variant – Figure 3.10.

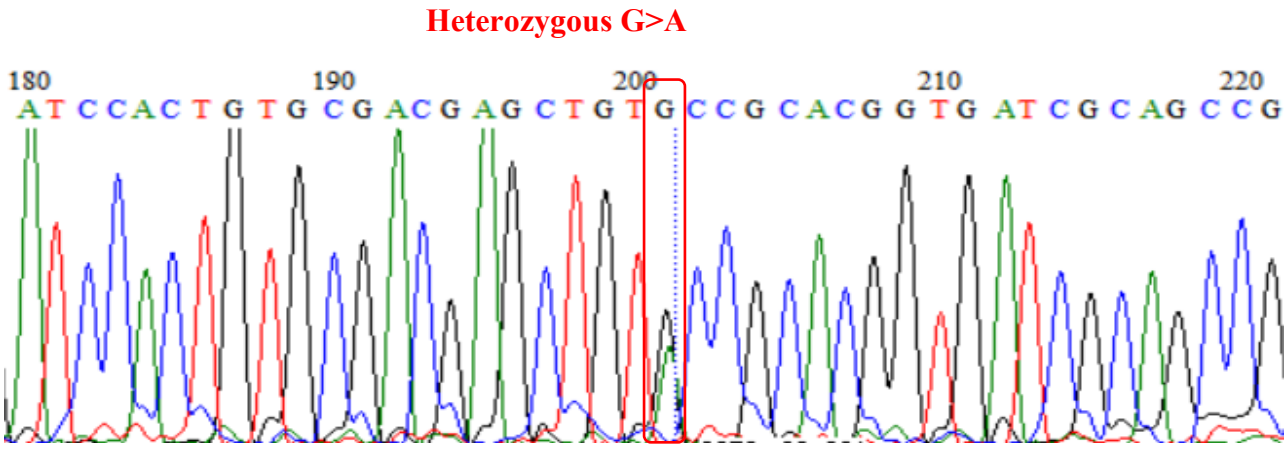


Figure 3.10 – Electropherogram Confirming Variant RET c.1901G>A in the Forward Strand of Patient PC.0032.01.1: Forward strand of patient PC.0032.01.1 confirming the presence of the heterozygous *RET* c.1901G>A variant (highlighted in red box). This is shown by the reduced double peak, where a black peak represents the reference G allele (G) and the green peak represents the alternate A allele.

Table 3.7 – ACMG-AMP Rules Applied to Variant RET_p.C634Y: This table summarises the ACMG-AMP rules applied to variant *RET*_p.C634Y identified in patient PC.0032.01.1. These rules and their strengths together result in an ACMG-AMP pathogenic classification.

PeCan Patient ID	Cancer Type	Confirmed/Suspected Clinical CPS Diagnosis	Variant Change	ACMG-AMP Classification	ACMG Codes	Reasoning & Evidence
PC.0032.01.1	Medullary Thyroid Carcinoma (MTC)	Confirmed Clinical MEN2A Diagnosis	*RET*_p.C634Y	Pathogenic	PM1 (Strong)	The variant is located in a mutational hotspot without any benign variation. The location of the variant is in a hotspot of length 17 amino acids and has 47 missense/in-frame/non-synonymous variants (23 pathogenic, 24 uncertain significance, and 0 benign), which qualifies this as a particularly dense hotspot and thus adds to the default strength of PM1, increasing it from moderate to strong.
					PM2 (Strong)	Variant reported in gnomAD exomes (good gnomAD exomes coverage = 85.5x) at an extremely low frequency of f = 0.00000404. The gnomAD homozygous allele count for the *RET* gene is 0. The variant has not been reported in gnomAD genomes (good gnomAD genomes coverage = 28.2x), ExAC, 1000 Genomes Project, and the NHLBI GO ESP6500. As recommended by VarSome, the strength of PM2 was increased from the default moderate to strong as the variant does not alter protein length and is located in a region which is predicted to be highly conserved by phyloP100way (9.62).
					PM5 (Moderate)	Several ClinVar, UniProt, and HGMD pathogenic variants affecting the same codon (Cysteine 634) as the c.1901G>A variant, with a different nucleotide change have been previously reported.

					PP5 (Strong)	Variant *RET* c.1901G>A (p.C634Y) has been classified as pathogenic by ClinVar (accession number: VCV000013909.23), UniProt (annotation ID: VAR_006325), and HGMD (citing 1 article, phenotype report (relating to the MEN2A phenotype), and functional characterisation report, accession number: CM941237). Variant *RET* p.C634Y was linked to 371 articles by Mastermind® (Link) and two VarSome users have classified this variant as pathogenic, citing two articles ((Liu et al., 2017; Uchino et al., 1999). Due to the cumulative pathogenic evidence provided by several reputable sources, the default PP5 strength of moderate has been increased to very strong.
					PP3 (Supporting)	Pathogenic computational verdict based on deleterious/damaging predictions from 11 different *in-silico* tools, with no other benign *in-silico* predictions. See Table 3.5
					PP4 (Supporting)	Patient PC.0032.01.1 presented with the following MEN2A-related phenotype: Clinodactyly of the fifth finger on both hands, a high-arched palate, hypopigmentation, and the MEN2A-related MTC. This phenotype and positive family history of the proband is highly specific to MEN2A disease, known to be caused by mutations in the *RET* gene.

PC.0051.01.1 – *TP53* (NM_000546.6): c.817C>T (p.R273C)

Patient PC.0051.01.1 is a Black South African female who was diagnosed with an embryonal rhabdomyosarcoma (ERMS) of the right buttock at age 1 year, 6 months. This patient's pedigree indicated a positive family history of cancer, furthermore indicative of a hereditary CPS – see Figure 3.11.

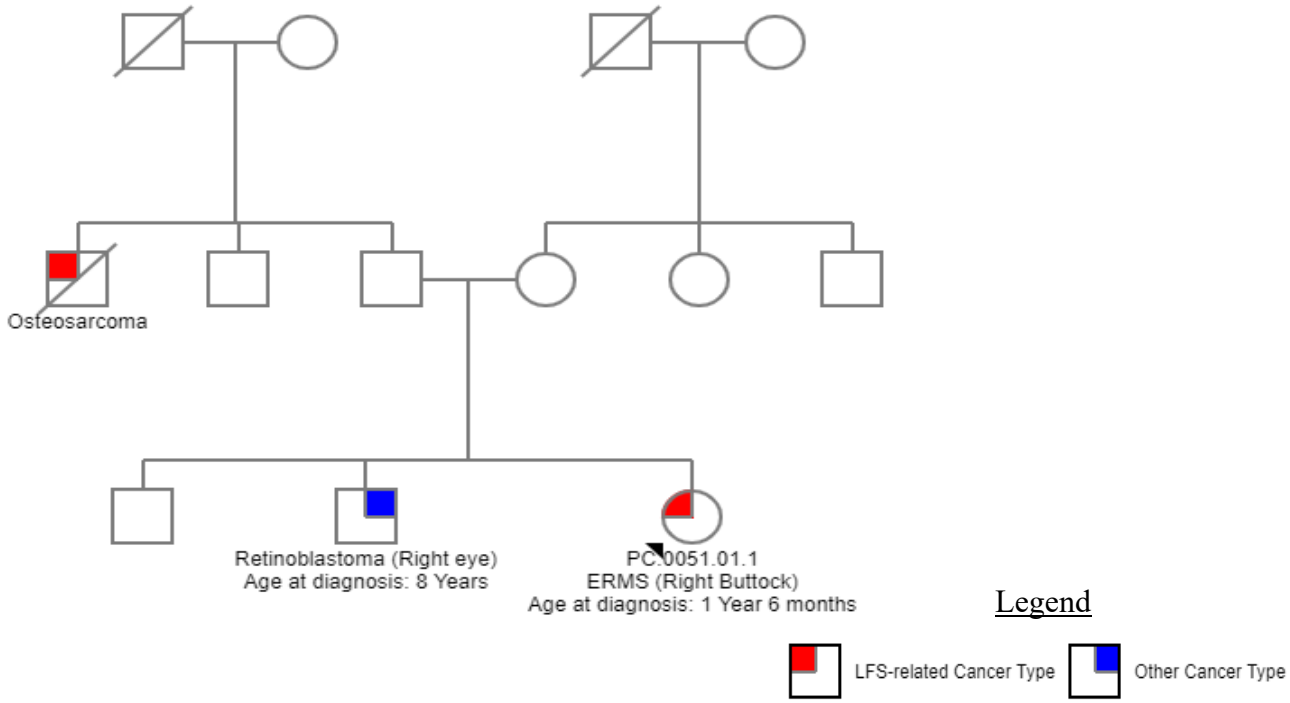


Figure 3.11 Patient PC.0051.01.1 Genetic Pedigree: Patient PC.0051.01.1 pedigree showing a positive family history of cancer within first- and second-degree relatives of the proband. This pattern of positive family history of cancer and detection of the pathogenic heterozygous *TP53* c.817C>T (p.R273C) variant in the proband, diagnosed with an ERMS of the right buttock at age 1 year, 6 months, is suggestive of LFS.

The *TP53* (NM_000546.6): c.817C>T (p.R273C) variant was identified as the likely disease-causing variant related to the patient's cancer type. The variant showed adequate sequencing quality (see Table 3.3). The variant is heterozygous with a C:T allele ratio of 0.46:0.54. Coverage of this variant can be seen in the IRGV image in Appendix D4 and depicts the missense variant in the complementary strand, thus a G>A substitution is shown.

The *TP53* c.817C>T is a missense variant resulting in the replacement of the Arginine 273 amino acid residue with a Cysteine amino acid residue. The single nucleotide substitution c.817C>T occurs in exon 8 (of 11) of the NM_000546.6 *TP53* transcript. The *TP53* c.817C>T variant was classified as pathogenic according to the ACMG-AMP codes: PM1, PM2, PM5, PS3, PP5, PP3,

and PP4. Variant evidence analysed justified increasing the default strength of PM1 from moderate to strong, as well as the default strength of PP5 from supporting to strong. See Table 3.8 for the variant evidence and reasoning used to apply these ACMG-AMP rules.

Although the *TP53* gene has a high proportion of pathogenic missense variants reported, the low gnomAD/ExAC missense z-score (0.98 and 1.38, respectively, which are both < the recommended ClinGen pathogenic missense z-score of > 3.09) and a high gnomAD o/e score (0.82, which is > gnomAD recommended score of 0.35) do not provide enough evidence to apply the ACMG-AMP rule PP2.

All five stages of validation via Sanger sequencing were successful for this variant. The heterozygous variant *TP53* c.817C>T is confirmed in the forward strand of patient PC.0051.01.1. This is indicated by the double peak (base call) shown in Figure 3.12 of both the reference and alternate allele.

Patient PC.0051.01.1 genetic pedigree and validation of germline TP53 variant is indicative of LFS type 1, meeting both the classical and 2015 revised Chompret LFS criteria: With the proband having a soft-tissue sarcoma (ERMS) before the age of 45/46 (diagnosed at age 1 year & 6 months), a first-degree relative (sibling) diagnosed with any cancer (retinoblastoma) before the age of 45 years (diagnosed at age 8 years), a second-degree relative (paternal uncle) diagnosed with an LFS related cancer (osteosarcoma) before the age of 56 years, and a germline *TP53* variant identified by a multi-gene NGS panel (Schneider et al., 2019; Bougeard et al., 2015; Mai et al., 2012; Chompret et al., 2000).

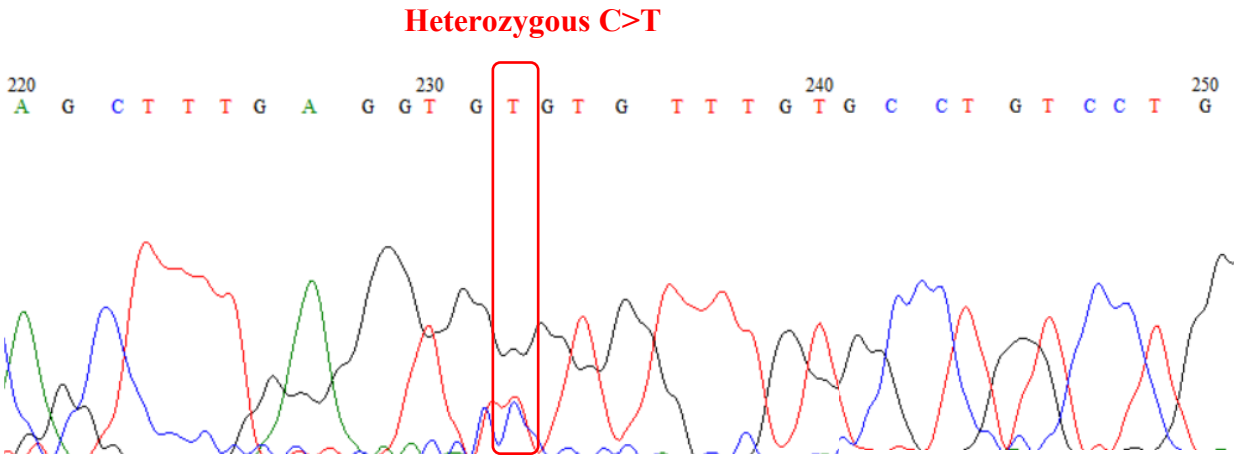


Figure 3.12 – Electropherogram Confirming Variant TP53 c.817C>T in the Forward Strand of Patient PC.0051.01.1: Forward strand of patient PC.0051.01.1 confirming the presence of the heterozygous *TP53* c.817C>T variant (highlighted in red box). This is shown by the reduced double peak, where the blue peak represents the reference C allele, and the red peak represents the alternate T allele. There is also evidence of a black (G allele) peak, however this is likely 'carry over' from the previous G homopolymer.

Table 3.8 – ACMG-AMP Rules Applied to Variant TP53_p.R273C: This table summarises the ACMG-AMP rules applied to variant *TP53*_p.R273C identified in patient PC.0051.01.1. These rules and their strengths together result in an ACMG-AMP pathogenic classification.

PeCan Patient ID	Cancer Type	Confirmed/ Suspected Clinical CPS Diagnosis	Variant Change	ACMG-AMP Classification	ACMG Codes	Reasoning & Evidence
PC.0051.01.1	Embryonal Rhabdomyosarcoma of Right Buttock (ERMS)	Pedigree Suggestive of Li-Fraumeni Syndrome	*TP53*_p.R273C	Pathogenic	PM1 (Strong)	The variant is located in a crucial DNA-binding domain and is also within a mutational hotspot without any benign variation. This mutational hotspot of length 17 amino acids and has 118 missense/in-frame/non-synonymous variants (47 pathogenic, 71 uncertain significance, and 0 benign), which qualifies this as a particularly dense hotspot, and thus adds to the default strength of PM1, increasing it from moderate to strong.
					PM2 (Moderate)	The variant is reported at an extremely low frequency in gnomAD exomes (f = 0.0000106) and ExAC (f = 0.00000888). The *TP53* c.817C>T has not been reported in gnomAD genomes, however, gnomAD has reported that the data quality for the *TP53* gene in gnomAD genomes is suspect, with codes AC0 (allele count equals zero and is not a high-confidence genotype) and RF (data has failed the random forest filters). This variant is absent from the 1000 Genomes Project and NHLBI GO ESP6500 population databases.
					PM5 (Moderate)	Several previously reported ClinVar, HGMD, and UniProt pathogenic missense variants affecting the same codon (Arginine 273), with a different nucleotide change.
					PS3 (Strong)	Several functional studies support a damaging effect on gene function/gene product for variants which occur in the same crucial *TP53* DNA-binding domain as the *TP53* c.817C>T mutation identified in patient PC.0051.01.1 (Li et al., 2014; Malcikova et al., 2010; Végran et al., 2007).

					PP5 (Strong)	The *TP53* c.817C>T (p.R273C) variant has been classified as pathogenic by ClinVar (with 14 submissions and citing 32 articles, accession number: VCV000043594.24), UniProt (citing 254 articles, annotation ID: VAR_005993), and HGMD (citing 1 article, phenotype report (relating to the LFS phenotype), and functional characterisation report, accession number: CM951233). In addition to these pathogenicity classifications, variant *TP53* p.R273C was linked to 1331 articles by Mastermind® (Link) and three VarSome users have classified this variant as pathogenic (Muller & Vousden, 2014; Cancer Genome Atlas Research Network et al., 2013), and one as likely pathogenic (Wada et al., 1993). Due to the cumulative pathogenic evidence provided by several reputable sources, the default PP5 strength of moderate has been increased to strong.
					PP3 (Supporting)	Pathogenic computational verdict based on deleterious/damaging predictions from 11 different *in-silico* tools, with no other benign *in-silico* predictions. See Table 3.5
					PP4 (Supporting)	Patient PC.0051.01.1 presented with the following phenotype: ERMS, and a positive history of heritable cancers, all specific to the LFS disease, known to be caused by mutations in the *TP53* gene.

PC.0011.01.1 – *BRCA1* (NM_007294.4): c.4199T>C (p.M1400T)

Patient PC.0051.01.1 is a South African Indian female who was diagnosed with an Embryonal Rhabdomyosarcoma (ERMS) of the bladder at age 3 years. The patient's genetic pedigree did not indicate any first- or second-degree relatives affected by cancer of any kind and therefore was not suggestive of a heritable/familial CPS.

The likely pathogenic *BRCA1* (NM_007294.4): c.4199T>C (p.M1400T) was identified in this patient. The variant showed adequate sequencing quality (see Table 3.3). The variant is heterozygous with a T:C allele ratio of 0.42:0.58. Coverage of this variant can be seen in the IRGV image in Appendix D4, which depicts the missense variant in the complementary strand, thus an A>G substitution is shown.

The *BRCA1* c.4199T>C is a missense variant resulting in the substitution of the Methionine 4199 amino acid residue with a Threonine amino acid residue. The single nucleotide substitution c.4199T>C occurs in exon 12 (of 23) of the NM_007294.4 transcript and ultimately results in the expression of a *BRCA1* gene product which differs from that of the reference protein. The *BRCA1* c.4199T>C variant was classified as likely pathogenic according to the following ACMG-AMP codes: PM1, PM2 and PP3. Variant evidence analysed justified increasing the default strength of PM1 from moderate to strong. See Table 3.9 for the variant evidence and reasoning used to apply these ACMG-AMP rules.

Although the *BRCA1* c.4199T>C variant is located in a mutational hotspot which is also a UniProt region of interest for its interaction with the key homologous repair gene, *PALB2,* the ACMG-AMP rule PP2 cannot be applied as the *BRCA1* gene missense z-score and o/e score does not meet the recommended pathogenicity thresholds. *BRCA1* gene has a gnomAD missense z-score of 0.58 and an ExAC missense z-score of -1.26, both of which are lower than the ClinGen SVI WG pathogenic threshold of >3.09. The gnomAD missense o/e score is 0.95. which is greater than the gnomAD recommended score of 0.35 and the VarSome recommended score of 0.76.

HGMD (CM041698), UniProt (VAR_070488) and ClinVar (VCV000055137.3) report an alternate variant affecting the same codon (Met1400), *BRCA1* c.4198A>G (p.M1400V) as a variant of uncertain significance (VUS), and thus the ACMG-AMP rule PM5 could not be applied. Likewise, ClinVar (VCV000055138.6) reports the *BRCA1* c.4199T>C (p.M1400T) variant detected here as a VUS, and thus the ACMG-AMP rule PP5 could also not be applied.

All five stages of validation via Sanger sequencing were successful for this variant. The heterozygous *BRCA1* c.4199T>C variant is confirmed in the forward strand of patient PC.0011.01.1 and the complementary heterozygous A>G change is confirmed in the negative strand of patient PC.0011.01.1. This is indicated by the double peak (base call) shown in Figure 3.13 of both the reference and alternate alleles.

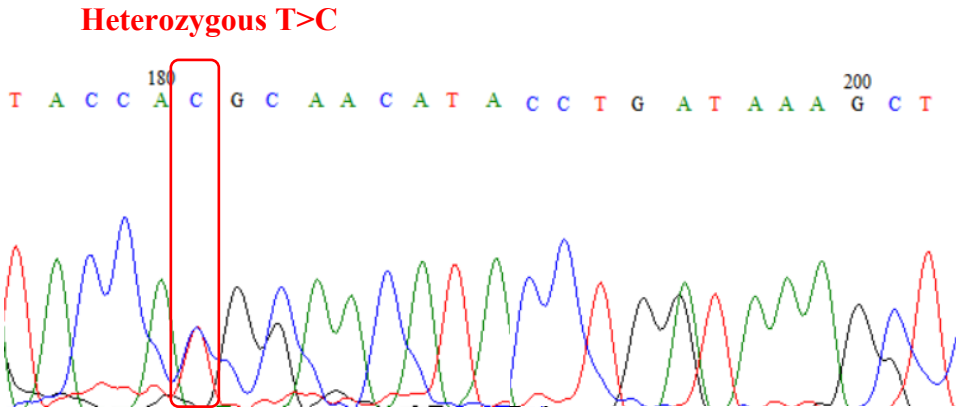


Figure 3.13 – Electropherogram Confirming Variant BRCA1 c.4199T>C in the Forward Strand of Patient PC.0011.01.1: Forward strand of patient PC.0011.01.1 generated via Sanger sequencing and analysed on BioEdit. Electropherogram image confirms the presence of the heterozygous *BRCA1* c.4199T>C variant (highlighted in red box). This is shown by the reduced double peak, where the red peak represents the reference T allele and the blue peak represents the alternate C allele.

Table 3.9 – ACMG-AMP Rules Applied to Variant BRCA1_p.M1400T: This table summarises the ACMG-AMP rules applied to variant *BRCA1*_p.M1400T identified in patient PC.0011.01.1. These rules and their strengths together result in an ACMG-AMP likely pathogenic classification.

PeCan Patient ID	Cancer Type	Confirmed/Suspected Clinical CPS Diagnosis	Variant Change	ACMG-AMP Classification	ACMG Codes	Reasoning & Evidence
PC.0011.01.1	Embryonal Rhabdomyosarcoma (ERMS) Bladder	None	*BRCA1*_p.M1400T	Likely Pathogenic	PM1 (Strong)	UniProt reports as a region of interest, due to the interaction of *BRCA1* with key Homologous Repair gene *PALB2*. Additionally, the variant is located in a mutational hotspot. The mutational hotspot of length 61 amino acids has 23 missense/in-frame/non-synonymous variants (22 pathogenic and 1 benign), which qualifies this as a particularly dense hotspot, and thus adds to the default strength of PM1, increasing it from moderate to strong.
					PM2 (Moderate)	Variant allele frequency is not reported in the control population databases gnomAD, the 1000 Genomes Project, ExAc and the NHLBI GO ESP6500. The variant region is reported to be well covered in gnomAD exomes (70.7x) and gnomAD genomes (32.4x).
					PP3 (Supporting)	Pathogenic computational verdict based on deleterious/damaging predictions from 11 different *in-silico* tools, with no other benign *in-silico* predictions. See Table 3.5

PC.0009.01.1 – *FAH* (NM_000137.4): c.485G>A (p.R162H)

Patient PC.0009.01.1 is a Black South African female who was diagnosed with a juvenile dermatofibrosarcoma protuberans (giant cell fibroblastoma) at the age of 5 years. The patient's genetic pedigree did not indicate any first- or second-degree relatives affected by cancer and thus was not suggestive of a familial/heritable CPS.

A likely pathogenic variant was identified in the Fumarylacetoacetate Hydrolase (*FAH)* gene. Variant *FAH* c.485G>A showed adequate sequencing quality (see Table 3.3). The variant is heterozygous with a G:A allele ratio of 0.51:0.49. Coverage of this variant can be seen in the IRGV image in Appendix D4.

The *FAH* c.485G>A is a missense variant resulting in the replacement of the Arginine 485 amino acid residue with a Histidine amino acid residue. The single nucleotide substitution c.485G>A occurs in exon 6 (of 14) of the NM_000137.4 *FAH* transcript and ultimately results in the expression of a *FAH* gene product which differs from that of the reference protein. The *FAH* c.485G>A variant was classified as likely pathogenic according to the ACMG-AMP codes: PM1, PM2 and PP3. Variant evidence analysed justified increasing the default strength of PM2 from moderate to strong. See Table 3.10 for the variant evidence and reasoning used to apply these ACMG-AMP rules.

Of the 54 non-VUS missense variants in the *FAH* gene, 48 are pathogenic (88.9%), which is greater than the VarSome recommended threshold of 38.6%. In addition, 139/361 (38.5%) clinically reported *FAH* variants are pathogenic, which is greater than the VarSome recommended threshold of 27.5%. Although the *FAH* gene has a high percentage of pathogenic missense variants and a high number of clinically reported variants which are pathogenic, the ACMG-AMP rule PP2 could not be applied due to the *FAH* gene's low missense z-score and high missense o/e score. The *FAH* gene has a gnomAD missense z-score of -0.03 and an ExAC missense z-score of 0.38, which is less than the recommended ClinGen SVI WG threshold of >3.09. The *FAH* gene missense o/e score is 1.01 which is greater than the gnomAD recommended threshold of 0.35 and the VarSome recommended threshold of 0.76.

A single alternate amino acid change at the *FAH* codon 162 has been reported in ClinVar. However, the alternate variant reported, c.484C>T (p.R162C) in ClinVar (VCV000887285.1), has been classified as a VUS and thus the ACMG-AMP rule PM5 could not be applied. Likewise, no equivalent amino acid changes nor the actual variant identified here (*FAH* c.485G>A p.R162H)

has been reported in ClinVar, UniProt, and HGMD, therefore the ACMG-AMP rules PS1 and PP5 could not be applied, respectively.

No articles were linked to the *FAH* p.R162H variant by Mastermind®. Additionally, both an external literature search and brief pathway analysis of the *FAH* gene and the patient's cancer type (dermatofibrosarcoma), showed no evidence directly linking the *FAH* gene to the pathogenesis of dermatofibrosarcoma. However, the identification of a pathogenic or likely pathogenic variant in a patient where no previous evidence supporting the patient's genotype-phenotype relationship has been reported is not an isolated occurrence within cancer genetics NGS studies. This will be expanded on in the Discussion chapter.

All five stages of validation via Sanger sequencing were successful for this variant. The heterozygous *FAH* c.485G>A variant is confirmed in the forward strand of patient PC.0009.01.1. This is indicated by the double peak (base call) shown in Figure 3.14 of both the reference and alternate alleles.

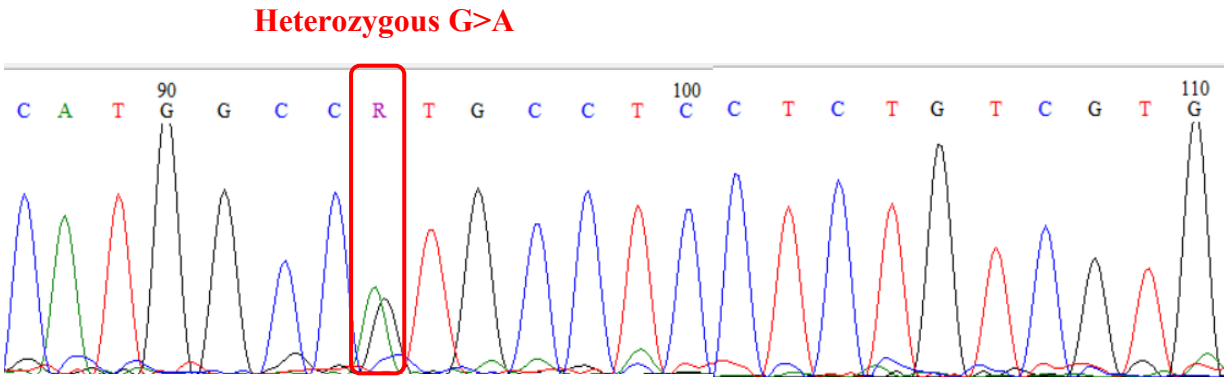


Figure 3.14 – Electropherogram Confirming Variant FAH c.485G>A in the Forward Strand of Patient PC.0009.01.1: Forward strand of patient PC.0009.01.1 generated via Sanger sequencing and analysed on BioEdit. Electropherogram image confirms the presence of the heterozygous *FAH* c.485G>A variant (highlighted in red box). This is shown by the reduced double peak, where the black peak represents the reference G allele, and the green peak represents the alternate A allele.

Table 3.10 – ACMG-AMP Rules Applied to Variant FAH_p.R162H: This table summarises the ACMG-AMP rules applied to variant *FAH*_p.R162H identified in patient PC.0009.01.1. These rules and their strengths together result in an ACMG-AMP likely pathogenic classification.

PeCan Patient ID	Cancer Type	Confirmed/Suspected Clinical CPS Diagnosis	Variant Change	ACMG-AMP Classification	ACMG Codes	Reasoning & Evidence
PC.0009.01.01	Juvenile Dermatofibrosarcoma Protuberans	None	*FAH*_p.R162H	Likely Pathogenic	PM2 (Strong)	The variant is reported in gnomAD exomes (good gnomAD exomes coverage = 81.4x) at an extremely low frequency of f = 0.0000159 and ExAC f = 0,0000165 The gnomAD homozygous allele count for the recessive *FAH* gene is 0. The variant has not been reported in gnomAD genomes (good gnomAD genomes coverage = 29.2x), 1000 Genomes Project, and the NHLBI GO ESP6500. As recommended by VarSome, the strength of PM2 was increased from the default moderate to strong as the variant does not alter protein length and is located in a region which is predicted to be highly conserved by phyloP100way (8.7).
					PM1 (Moderate)	The variant is located in a crucial CA^{2+} metal ion binding domain. This region has 29 missense/in-frame/non-synonymous variants (9 pathogenic, 20 uncertain, 0 benign).
					PP3 (Supporting)	Pathogenic computational verdict based on deleterious/damaging predictions from 11 different *in-silico* tools, with no other benign *in-silico* predictions. See Table 3.5

PC.0003.01.1 – *ERCC3* (NM_000122.2): c.958C>T (p.Q320*)

Patient PC.0003.01.1 is a Black South African female who was diagnosed with a nephroblastoma (Wilms' tumour) at the age of 7 years. The patient's genetic pedigree did not indicate any first- or second-degree relatives affected with cancer of any type and is therefore not suggestive of any heritable/familial CPS.

A pathogenic variant was identified in the *ERCC3* gene. Variant *ERCC3* (NM_000122.2): c.958C>T (p.Q320*) showed an adequate sequencing quality (see Table 3.3). The variant is heterozygous with a C:T allele ratio of exactly 0.50:0.50. Coverage of this variant can be seen in the IRGV image in Appendix D4, which depicts the variant in the complementary strand, thus a G>A substitution is shown.

Variant *ERCC3* c.958C>T is a nonsense variant whereby the single nucleotide change results in the substitution of the 320 Glutamine amino acid with a stop codon and thus premature termination in the translation of the *ERCC3* gene product. See Mutalyzer 2.0.35 for the prediction and visualisation of the reference (Figure 3.15 A) and resultant truncated *ERCC3* protein (Figure 3.15 B). The *ERCC3* c.958C>T variant identified in patient PC.0003.01.1 was classified as pathogenic according to the following ACMG-AMP criteria: PVS1, PM2, and PP3. See Table 3.11 for the variant evidence and reasoning used to apply these ACMG-AMP rules.

The *ERCC3* c.958C>T variant has not been reported in ClinVar, UniProt or HGMD. Without additional pathogenic criteria from these reputable sources, the ACMG-AMP rule PP5 could not be applied. Mastermind® identified one paper linked to the *ERCC3* p.Q320* variant. However, this variant was identified in a study profiling circulating tumour DNA in metastatic breast cancer (Cresswell et al., 2020). A literature review revealed two studies whereby a truncating *ERCC3* germline variant was identified in a patient with a Wilms' tumour (Kim et al., 2021a; Byrjalsen et al., 2020). No other literature evidence nor a brief pathway analysis further supported the patient's genotype–phenotype (*ERCC3* – Nephroblastoma) relationship. The identification of pathogenic/likely pathogenic variants in genes not previously known to be involved in the pathogenesis of the patient's cancer type is not an isolated event in cancer genetics NGS studies. This will be expanded on in the Discussion chapter.

All five stages of validation via Sanger sequencing were successful for this variant. The heterozygous *ERCC3* c.958C>T variant is confirmed in the forward strand of patient

PC.0003.01.1. This is indicated by the double peak (base call) shown in Figure 3.16 of both the reference and alternate alleles.

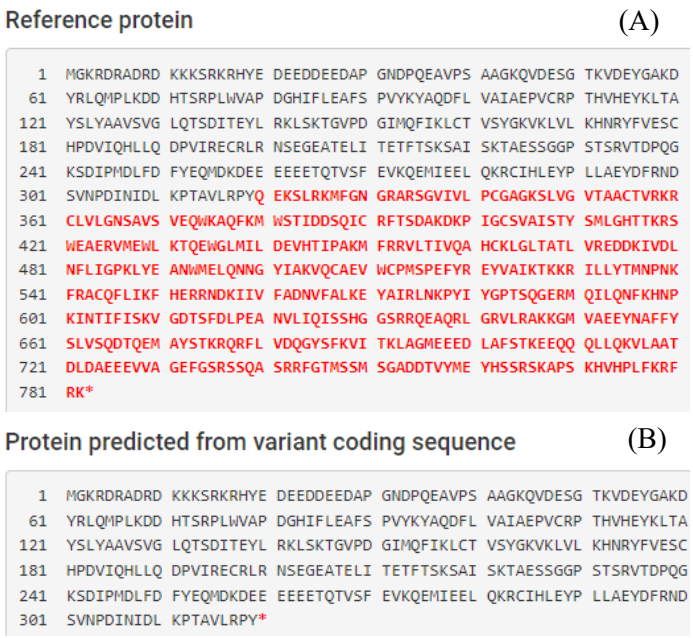


Figure 3.15 - Mutalyzer 2.0.35 Output for Variant ERCC3 (NM_000122.2): c.958C>T (p.Q320*) Identified in Patient PC.0003.01.1 (A & B): **(A)** Mutalyzer 2.0.35 image depicting the effect of the variant's single base substitution c.958C>T on the reference protein. Amino acid residues shown in black occur before the single base pair substitution and are unaffected by the variant change. Amino acid residues shown in red are those which are no longer translated due to the gain of a premature stop codon at the amino acid 320 Glutamine residue **(B)** Mutalyzer 2.0.35 image depicting the predicted protein as a result of the nonsense *ERCC3* variant. The result is a truncated ERCC3 protein, whereby amino acid 320 Glutamine and those which occur after are no longer translated.

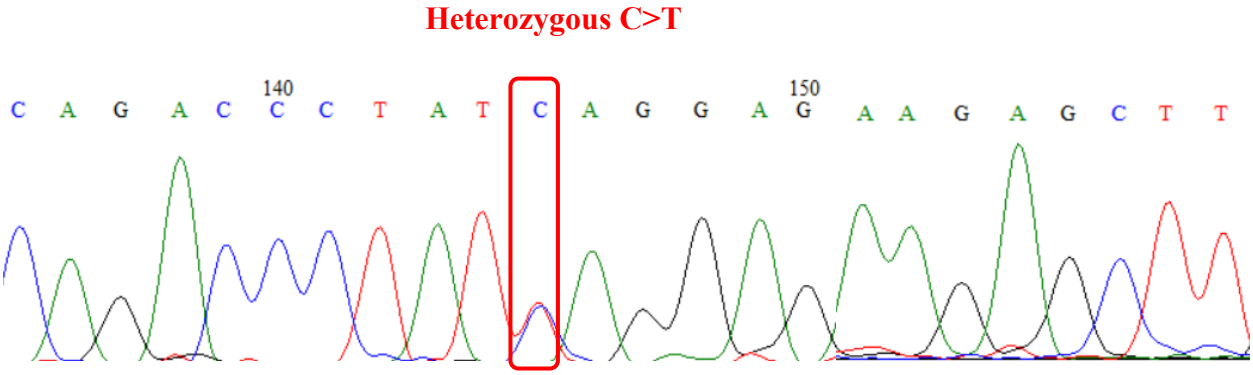


Figure 3.16 – Electropherogram Confirming Variant ERCC3 c.958C>T in the Forward Strand of Patient PC.0003.01.1: Forward strand of patient PC.0003.01.1 generated via Sanger sequencing and analysed on BioEdit. Electropherogram image confirms the presence of the heterozygous *ERCC3* c.958C>T variant (highlighted in red box). This is shown by the reduced double peak, where the blue peak represents the reference C allele, and the red peak represents the alternate T allele.

Table 3.11 – ACMG-AMP Rules Applied to Variant ERCC3_p.Q320*: This table summarises the ACMG-AMP rules applied to variant ERCC3_p.Q320* identified in patient PC.0003.01.1. These rules and their strengths together result in an ACMG-AMP pathogenic classification

PeCan Patient ID	Cancer Type	Confirmed/Suspected Clinical CPS Diagnosis	Variant Change	ACMG-AMP Classification	ACMG Codes	Reasoning & Evidence
PC.0003.01.1	Nephroblastoma (Right sided)	None	*ERCC3*_p.Q320*	Pathogenic	PVS1 (Very Strong)	A nonsense variant in the *ERCC3* gene, where LOF is a known mechanism of disease. The ERCC3 gene's gnomAD LOF o/e score is 0,559, which is less than the VarSome recommended threshold of 0,763.
					PM2 (Moderate)	Variant allele frequency is not reported in the control population databases gnomAD, the 1000 Genomes Project, ExAc and the NHLBI GO ESP6500. The variant region is reported to be well covered in gnomAD exomes (99.2x) and gnomAD genomes (31.4x).
					PP3 (Supporting)	Pathogenic computational verdict based on deleterious/damaging predictions from 7 different *in-silico* tools, with no other benign *in-silico* predictions. See Table 3.5

PC.0018.01.1 – *RB1* (NM_000321.3): c.1981C>T (p.R661W)

Patient PC.0018.01 is a Black South African Male who was diagnosed with a right-sided nephroblastoma (Wilms' tumour) at age 6 months. The patient's genetic pedigree did not indicate any first- or second-degree relatives diagnosed with any type of cancer and was therefore not suggestive of a familial/heritable CPS.

A pathogenic variant was identified in the *RB1* gene. Variant *RB1* (NM_000321.3): c.1981C>T (p.R661W) showed an adequate sequencing quality (see Table 3.3). The variant is heterozygous with a C:T allele ratio of 0.53:0.47. Coverage of this variant can be seen in the IRGV image in Appendix D4. The *RB1* c.1981C>T is a missense variant resulting in the replacement of the Arginine 320 amino acid residue with a Tryptophan amino acid residue. The single nucleotide substitution c.1981C>T occurs in exon 20 (of 27) of the NM_000321.3 *RB1* transcript and ultimately results in the expression of an *RB1* gene product which differs from that of the reference protein. The *RB1* c.1981C>T variant was classified as pathogenic according to the ACMG-AMP codes: PS3, PM1, PM2, PP5, and PP3. See Table 3.12 for the variant evidence and reasoning used to apply these ACMG-AMP rules.

Although the *RB1* gene has a high percentage of pathogenic non-VUS missense variants as well as a large number of pathogenic clinically reported variants, ACMG-AMP rule PP2 could not be applied due to the *RB1* gene's low missense z-score and high missense o/e score. The *RB1* gene has a gnomAD missense z-score of 2.67 and an ExAC missense z-score of 1.98, both of which are lower than the recommended ClinGen SVI WG pathogenic threshold of >3.09. The *RB1* gnomAD missense o/e score is 0.65, which is less than the VarSome recommended threshold of 0.763 but is still higher than the gnomAD recommended threshold of 0.35. A single alternative amino acid change affecting the *RB1* 661 codon has been reported in ClinVar. However, this alternative *RB1* c.1982G>T (p.R661Q) (accession: VCV000820461.4) variant has been classified as a VUS, and therefore the ACMG-AMP rule PM5 could not be applied. Likewise, no equivalent amino acid changes for this variant have been reported in ClinVar, HGMD and UniProt and thus the ACMG-AMP rule PS1 could not be applied.

The large majority of studies linked to this variant via Mastermind® (link) report the variant in cases of familial retinoblastoma, where *RB1* is a known causal gene. Pathogenic and likely pathogenic variants in the *RB1* gene have been reported in cancer types other than retinoblastoma, especially in cases of secondary cancer formation in familial retinoblastoma patients. However, a literature review and brief *in-silico* pathway analysis did not provide any evidence to support the

patient's genotype-phenotype relationship (*RB1* germline variants in cases of Wilms' tumour). Variants identified in this study which have no previously reported evidence to support their role in the pathogenesis of the patient's cancer type will be expanded on in the Discussion chapter.

All five stages of validation via Sanger sequencing were successful for this variant. The heterozygous *RB1* c.1981C>T variant is confirmed in the forward strand of patient PC.0018.01.1 and is indicated by the double peak (base call) shown in Figure 3.17 of both the reference and alternate alleles.

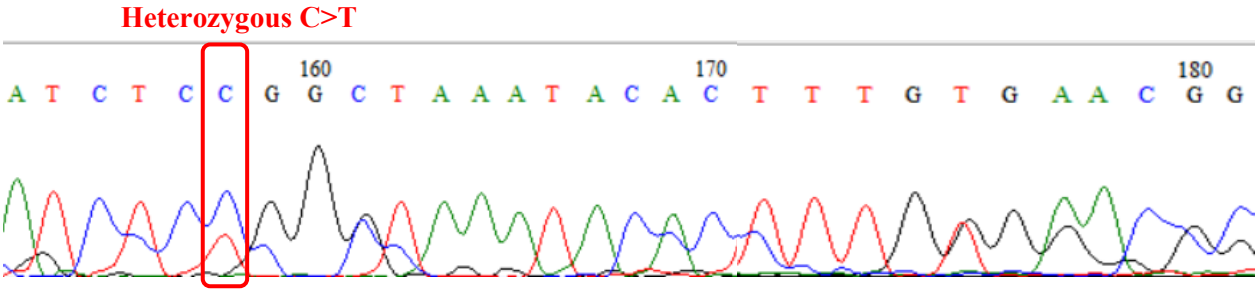


Figure 3.17 – Electropherogram Confirming Variant RB1 c.1981C>T in the Forward Strand of Patient PC.0018.01.1: Forward strand of patient PC.0018.01.1 generated via Sanger sequencing and analysed on BioEdit. Electropherogram image confirms the presence of the heterozygous *RB1* c.1981C>T variant (highlighted in red box). This is shown by the reduced double peak, where the blue peak represents the reference C allele and the red peak represents the alternate T allele.

Table 3.12 – ACMG-AMP Rules Applied to Variant RB1_p.R661W: This table summarises the ACMG-AMP rules applied to variant *RB1*_p.P661W identified in patient PC.0018.01.1. These rules and their strengths together result in an ACMG-AMP pathogenic classification.

PeCan Patient ID	Cancer Type	Confirmed/Suspected Clinical CPS Diagnosis	Variant Change	ACMG-AMP Classification	ACMG Codes	Reasoning & Evidence
PC.0018.01.1	Nephroblastoma (Right-Sided)	None	*RB1*_p.R661W	Pathogenic	PS3 (Strong)	Functional studies supporting a damaging effect on gene product/gene function for variants which occur in the same *RB1* pocket B domain as the R661W variant identified in patient PC.0018.01.1. Variants in this domain, including R661W, show defective protein binding activity and inducing cell cycle arrest *in vitro*. *In vivo,* the pocket domain shows a temperature-sensitive binding threshold, which may explain the variable frequency of tumour development seen in families carrying this mutation (Park et al., 2008; Otterson et al., 1999).
					PM1 (Moderate)	The variant is located in the *RB1* pocket B domain, crucial for the binding of the E2F transcription factor which plays a key role in the regulation of cell proliferation and thus mitigates *RB1*'s tumour suppressor function. The variant is also within a mutational hotspot without any benign variation. The mutational hotspot of length 17 amino acids has 21 missense/ in-frame/ synonymous variants (6 pathogenic, 15 uncertain, and 0 benign).
					PM2 (Moderate)	The variant is reported at a very low frequency in gnomAD exomes (good gnomAD exomes coverage = 83.2x) at (f = 0.0000159), 1000 Genomes Project (f = 0.00001997), and ExAC (f = 0.00000824). The variant has not been reported in gnomAD genomes (good gnomAD genomes coverage = 28,9x) and the NHLBI GO ESP6500 population databases.
					PP3 (Supporting)	Pathogenic computational verdict based on deleterious/damaging predictions from 11 different *in-silico* tools, with no other benign i*n-silico* predictions. See Table 3.5.
					PP5 (Supporting)	Variant RB1 c.1981C>T (p.R661W) has been classified as pathogenic by ClinVar (with 8 submissions and 20 articles cited, accession number: VCV000013087.18), UniProt (citing 9 articles, annotation ID: VAR_005582) and HGMD (citing 1 article and an additional phenotype report, accession number: CM920603). Additionally, the R661W variant was linked to 111 articles by Mastermind® (Link).

4. Discussion

4.1 Project Summary

The increasing incidence of new childhood cancer cases together with the high mortality rates in LMICs emphasises the need to further explore the genetic mechanisms in the pathogenesis of childhood cancer, thereby contributing to the development of other diagnostic testing options, evidence-based treatment plans, and ultimately improving current survival rates.

The pathogenesis of childhood cancers has been shown to be distinct from that of adult cancers, with the latter being more frequently driven by co-driver mutations and showing a higher mutation frequency overall (Gröbner et al., 2018). This may be reflective of the large role that cumulative exposure to extrinsic carcinogenic factors and the cellular ageing process plays in the development of adult cancers (Madia et al., 2019; Jagai et al., 2017; Vassilev & DePamphilis, 2017). Without exposure to these cumulative factors in childhood, it leads to the question as to what role the genetic component, specifically the germline component, plays in the development of cancer in childhood.

In South Africa, there is no current routine diagnostic testing in place for the detection of germline variants in paediatric cancer patients. Testing only occurs in private/referred cases which rely heavily on defined clinical CPS and/or family history criteria, thereby restricting referral criteria of patients who may be eligible for genetic testing. This testing also involves the use of single gene-based molecular techniques, which in comparison to multi-gene sequencing techniques, can be laborious, costly, and often inconclusive (Sahm et al., 2016; Nikiforova et al., 2015; Rahman, 2014a). Therefore, the implementation of NGS technologies is a promising concept within the SA research and diagnostic settings. The use of these NGS technologies in research and diagnostics has been well-established in HICs, with several paediatric cancer studies identifying P/LP germline variants in CPGs in 3.8 – 35.5% of their patients (von Stedingk et al., 2021; Kline et al., 2017; Chang et al., 2016; Oberg et al., 2016; Parsons et al., 2016). To the best of our knowledge, there are no studies which have investigated the germline profiles of solid-tumour cancer patients within the SA paediatric population.

In the present study, we aimed to gain insight into the occurrence and contribution of germline variants to the development of paediatric cancer in a SA cohort. We utilised the increasing availability of targeted NGS panel design software and short-read sequencing to generate high throughput genomic data for these patients, thus facilitating germline variant identification and

subsequent analysis. Given the considerable genetic heterogeneity in SA, our study approach and findings are highly relevant for informing genomic medicine efforts aimed at addressing current paediatric cancer diagnostic testing limitations in the country and potentially across other LMICs in Africa.

4.2 Project Yield

Seven P/LP germline variants were identified and validated in seven of the 32 paediatric solid-tumour cancer patients recruited and investigated in this study, resulting in an overall variant detection rate of ~22%. Three of these P/LP variants were identified in patients whose phenotypes are consistent with their genetic findings and are associated with well-known CPSs, resulting in a diagnostic yield of ~9.4% (3/32). The remaining four P/LP variants were identified in CPGs with little to no previous relation to the patient's cancer phenotype and require further investigation. A diagnostic yield of ~9.4% was anticipated for this project's research question as it falls within the 3.8 – 35.5% range as determined by international paediatric cancer NGS studies (von Stedingk et al., 2021; Kline et al., 2017; Chang et al., 2016; Oberg et al., 2016; Parsons et al., 2016). This project's yield of ~9.4% was also expected according to the global estimation that 3-10% of all paediatric cancer patients are expected to harbour disease-causing/contributing germline variants (Akhavanfard et al., 2020; Zhang et al., 2015).

Notably, there is a large variation in the diagnostic yield reported by these international paediatric cancer NGS studies. This variation may be attributed to one of the many different methodologies observed between studies, including differences in sample size, NGS techniques utilised, paediatric cancer types, age of participants, different population groups, and clinical inclusion criteria. Interestingly, studies with the highest variant detection rates (16.8 – 35.5%) (Byrjalsen et al., 2020; Mirabello et al., 2020; Diets et al., 2018; Kline et al., 2017; Mody et al., 2015) tend to be associated with study designs which analysed a high number of CPGs (≥ 146 CPGs) and/or utilised WES in their NGS approach. This suggests that the expansion of the current project's targeted CPG panel or selecting a WES study design may aid in achieving a higher variant detection rate. Regardless, this project's diagnostic variant detection rate of ~9.4% still falls within the international yield range (3.8-35.5%) and thus speaks to the efficacy of this project's NGS design.

This project's NGS design is capable of simultaneously screening 32 patients for germline variants within the exonic regions and exon-intron boundaries of 50 CPGs. This allowed for the identification of both well-known variants consistent with patient cancer phenotypes and rare or

novel variants which would not have been suspected according to the patient's cancer phenotype. Achieving the same results using single-gene testing approaches would require the development of numerous individual diagnostic tests, which would be a laborious and impractical approach. Additionally, performing multiple single CPG tests sequentially would prove to be costly and time-consuming, and therefore not suitable for implementation in a diagnostic environment. Therefore, this project's findings indicate that the NGS platform utilised here would be suitable as a primary screening method for the identification of disease-causing/contributing germline variants in SA paediatric cancer patients.

4.3 Mutation-positive Cases

This project's targeted gene panel and bioinformatic analysis pipeline facilitated the identification, and further validation, of five pathogenic (P) and two likely pathogenic (LP) germline variants in a total of seven different probands. The large majority of these variants are missense (n=5, 71%), with one frameshift and one nonsense variant also identified. The seven germline variants were identified in the following genes: (P) - *NF1, RET, TP53, ERCC3, RB1*, (LP) - *BRCA1* and *FAH*. As is consistent with the literature, the majority of the variants identified here occur in CPGs which primarily act as TSGs (6/7 genes), with one pathogenic variant identified in the *RET* proto-oncogene (Reviewed by (Saletta, Dalla Pozza & Byrne, 2015; Rahman, 2014)). Additionally, the majority of the variants identified here occur in CPGs which play a key role in the recognition and repair of DNA damage, acting in both a direct (*BRCA1*, *ERCC3*) and indirect manner (*TP53, RB1*), as is reported in several paediatric cancer studies (Kim et al., 2021b; Mirabello et al., 2020; Byrjalsen et al., 2020; Gröbner et al., 2018; Zhang et al., 2015). Of the 50 CPGs analysed in this study, the *TP53* gene is the most prevalent in literature, with several studies reporting it as the CPG with the highest burden of variants (Kim et al., 2021a; Akhavanfard et al., 2020; Mirabello et al., 2020; Gröbner et al., 2018; Zhang et al., 2015).

Three of the pathogenic variants identified, namely in the *NF1*, *RET*, and *TP53* genes, have been previously reported and classified as pathogenic in ClinVar, HGMD, and UniProt. Additionally, all three variants have literature evidence supporting the patient's genotype-phenotype relationship. The P/LP variants identified in *BRCA1*, *ERRC3*, *RB1,* and *FAH* have been reported in cancer literature, however, the context of these variants differs from that of the patients' cancer phenotype. To our knowledge, the *FAH* c.485G>A (p.R162H) variant identified in patient PC.0009.01.1 is novel, as it has not been reported in ClinVar, UniProt, HGMD, and has no published literature to date. Each identified and validated putative disease-causing/contributing

variant will be further expanded on here, in terms of the strength of the patient's genotypic-phenotypic evidence which supports the variant.

4.3.1 Variants Identified in Patients with Strong Phenotypic Evidence

NF1 c.1133_1136delTGAC (p.D378Afs*8) Frameshift Deletion Variant

The heterozygous *NF1* c.1133_1136delTGAC (p.D378Afs*8) frameshift variant was detected and validated in patient PC.0023.01.1, who was diagnosed at age 9 years with a malignant pilocytic astrocytoma. The patient had been previously clinically diagnosed with Neurofibromatosis Type 1 (NF1), a known CPS characterised by changes in skin pigmentation, growths of tumours along the nerves, and a predisposition to cancer formation from early childhood, including pilocytic astrocytomas (Karajannis & Ferner, 2015). During recruitment, the patient presented with an NF1-associated phenotype and a family history which revealed that the proband's father was previously clinically diagnosed with NF1 and NF1-associated cancer.

The pathogenic *NF1* p.D378Afs*8 variant is a four base-pair deletion resulting in the introduction of a premature stop codon, predicted by SIFT-INDEL to cause nonsense-mediated decay (NMD) of the *NF1* mRNA transcript. Although unlikely, if the mutated transcript were to escape the NMD mechanism, it would result in the translation of a shortened *NF1* protein, lacking major protein domains and motifs which are essential for the interaction of the *NF1* protein with several signalling pathways and ultimately for *NF1*'s role as a TSG (Gurung & Bhattacharjee, 2015). According to gnomAD, the *NF1* gene is highly likely to be loss-of-function intolerant (pLI), indicating that LOF is a potential mechanism for disease within this gene—where very few benign LOF variants are reported, adding to the pathogenicity of this variant.

The *NF1* gene encodes Neurofibromin, a large multifunctional regulatory protein expressed mainly in the neural cells of the peripheral and central nervous systems (Bergoug et al., 2020). A non-functional Neurofibromin protein loses its ability to interact and regulate key signal transduction pathways, such as the *RAS* and cAMP signalling pathways, thus rendering them constitutively active, ultimately resulting in uncontrolled cell growth and proliferation signalling. It is this LOF mechanism which results in the formation of the NF1 phenotype, where individuals tend to develop both benign (neurofibromas, Lisch nodules) and cancerous (malignant peripheral nerve sheath tumours (MPNSTs), optic nerve gliomas, low-grade gliomas) tumours (Goldblum, Folpe & Weiss, 2020; Gurung & Bhattacharjee, 2015). Patient PC.0023.01.1 developed a pilocytic astrocytoma, which is a type of low-grade glioma, seen to occur more frequently in the NF1 paediatric population. Heterozygous *NF1* germline variants are seen to be involved in the

pathogenesis of pilocytic astrocytomas through the dysregulated interaction of the mutated neurofibromin protein with the *RAS* CNS and cAMP signalling pathways. These pathways are involved in cell growth, development, and plasticity in the brain. (Goldblum, Folpe & Weiss, 2020; Molosh & Shekhar, 2018).

The presence of a heterozygous *NF1* variant is sufficient for the development of a typical non-malignant NF1 phenotype, as well as the predisposition to the development of NF1-associated malignancies. However, like other TSGs, the development of NF1-associated malignancies requires the somatic inactivation of the other allele. In the case of pilocytic astrocytomas, studies have shown that those affected by NF1 require an acquired germline *NF1* variant, a second somatic *NF1* variant, and the cooperation of the right tumour microenvironment for the full penetrance of the malignant brain tumour phenotype (Gutmann et al., 2013; Bajenaru et al., 2003). Therefore, the likely mechanism of pathogenesis in the development of the pilocytic astrocytoma of patient PC.0023.01.1 is the LOF of the *NF1*-gene due to the acquired pathogenic *NF1* p.D378Afs*8 germline variant and a probable second somatic *NF1* variant within the tumour genome. Although, NGS of the patient's tumour biopsy is needed for further confirmation of this mechanism of pathogenesis.

Although the mutational spectrum within NF1 patients continues to be expanded, with the identification of groups of recurrent variants, the large majority of NF1 patients/families harbour private variants (Bianchessi et al., 2020; Sabbagh et al., 2013). Thus, it is not unusual that the *NF1* p.D378Afs*8 variant identified in patient PC.0023.01.1 has not been reported in ClinVar, UniProt, and HGMD. This variant has been previously reported in NF1 patients, with one study (Upadhyaya et al., 2008) mentioning an MPNST with LOH diagnosis in addition to their NF1 phenotype (Sabbagh et al., 2013; Pros et al., 2008). To our knowledge, this is the first report of a South African paediatric NF1 patient harbouring the pathogenic putative disease-causing *NF1* p.D378Afs*8 variant. The patient's family history indicates that the pathogenic *NF1* variant is likely inherited from the proband's father, although genetic testing of both parents and segregation analysis would need to be performed in order to confirm inheritance.

RET c.1901G>A (p.C634Y) Missense Variant

The heterozygous pathogenic *RET* c.1901G>A (p.C634Y) missense variant was detected and validated in patient PC.0032.01.1, who was diagnosed with a Medullary Thyroid Carcinoma (MTC) at the age of 14 years. The patient previously received a clinical diagnosis of Multiple

Endocrine Neoplasia Type 2A (MEN2A) which is a well-known CPS that affects the endocrine glands and greatly increases the likelihood of the development of a MTC (>90%), pheochromocytoma (~50%), parathyroid adenoma/hyperplasia (20-30%), and cutaneous lichen amyloidosis (CLA) (30-35%) (Neumann, Young & Eng, 2019; Wells et al., 2015; American Thyroid Association Guidelines Task Force et al., 2009). The patient's family history indicates that the pathogenic *RET* variant is likely inherited from the proband's father, although genetic testing of both parents and segregation analysis would need to be performed in order to confirm inheritance. This patient was recruited as a blind bioinformatics control patient, and it has been further confirmed that the *RET* c.1901G>A variant detected in patient PC.0032.01.1 is the same variant previously detected in a diagnostic setting.

The *RET* gene is a proto-oncogene that encodes a transmembrane cell surface receptor known as a receptor tyrosine kinase (RTK). The RTK encoded by *RET* acts as a receptor for a family of extracellular signalling molecules, which then activate signalling pathways involved in the growth and differentiation of neurons and other peripheral cells (De Falco et al., 2017; Davidoff, 2016). Gain-of-function (GOF) mutations within the *RET* RTK domain result in the constitutive activation of the receptor, in the absence of a ligand. This results in the loss of the gene's regulatory capacity and constant activation of signalling pathways, promoting malignant transformation (Du & Lovly, 2018; Davidoff, 2016). GOF mutations in the *RET* gene have been established as the mechanism of pathogenesis in the development of the MEN2A disease and subsequently the development of MTC in >90% of MEN2A cases. MTC is a solid-tumour malignancy that arises from parafollicular C-cells in the thyroid gland, which are derived from embryonic neural crest cells where *RET* is known to be highly expressed (Lodish & Stratakis, 2008; Pachnis, Mankoo & Costantini, 1993).

The *RET* c.1901G>A variant occurs within a mutational hotspot, where a high proportion of missense variants reported are pathogenic, with no benign missense variants reported. This variant lies in the extracellular juxtamembrane region of the *RET* encoded RTK domain, which spans amino acids 29-635. Variants that affect cysteine residues in the extracellular domain which also lie in close proximity to the transmembrane domain (spanning amino acids 636-657), such as the variant detected here affecting the cysteine 634 amino acid residue, are said to gain protein homodimerization activity (neXtProt human protein database entry: NX_P07949-1). Heterozygous variants affecting cysteine residues in this region result in a single wild-type functional cysteine residue left unpaired. This unpaired cysteine residue is now free to form disulfide bonds within the same domain, thereby generating receptor homodimers. It is this strong

self-association that results in constitutive activation of the *RET* kinase, thereby promoting malignant transformation (Kjær et al., 2006).

The *RET* p.C634Y variant has been previously reported and classified as pathogenic in several reputable sources (ClinVar (Accession VCV000013909.23), UniProt, HGMD) and has been linked to over 350 articles (Mastermind® Link) mentioning the variant in MEN2A and less commonly FMTC families. This variant has been reported in populations all over the world including South Africa (SA). Cases of MEN2A and its associated endocrinopathies have been documented and reported to occur in SA families for over three decades (Zorgania, Piriea & Motalaa, 2018; Moore & Zaahl, 2010; Hopley & Huddle, 1997). However, the exact p.C634Y variant has only been reported in SA families over the last decade or so, mostly occurring in White, followed by Black South Africans. The most prevalent MEN2A variant in the SA population to date is the C634S variant (Moore & Zaahl, 2010; Moore & Zaahl, 2008; Moore, Appfelstaedt & Zaahl, 2007; Hopley & Huddle, 1997). Like ~95% of all MEN2A cases, the heterozygous p.C634Y variant in patient PC.0032.01 is likely inherited (Al-Salameh, Baudry & Cohen, 2018; Moline & Eng, 2011).

Adding to the pathogenicity of this variant are several reports (ClinVar, UniProt, HGMD) of alternate pathogenic variants affecting the same amino acid residue (Cysteine 634). These variants have been reported within different populations of MEN2A individuals with MTC, and/or pheochromocytomas, and/or parathyroid adenoma/hyperplasia, and/or CLA, thus highlighting the clinical importance of this amino acid residue in the pathogenesis of MEN2A and MTC. Of the pathogenic variants affecting the C634 codon in MEN2A patients, the majority of individuals are reported to harbour the C634R variant, followed by the C634Y variant (Melmed & Young, 2020; Moline & Eng, 2011; Puñales et al., 2003). Research has shown that in comparison to other variant sites, MEN2A individuals with C634 variants have a very high likelihood of developing MTC with an earlier onset and may follow a more aggressive course (Valdés et al., 2015; Toledo et al., 2006; Puñales et al., 2003). In fact, the 2015 revised American Thyroid Association (ATA) guidelines recommend that MEN2A individuals with C634 variants undergo complete thyroidectomy at/before the age of 5 years, and also begin screening for the presence of pheochromocytomas from 11 years onwards (Wells et al., 2015).

Like other oncogenes, the presence of a single GOF variant in the *RET* gene is generally sufficient for the development of the MEN2A disease phenotype and associated benign/malignant tumours. In fact, a large percentage of individuals with MEN2A and MTC are seen to carry heterozygous

variants affecting the C634 codon (Young, 2013; Puñales et al., 2003). Therefore, the mechanism of pathogenesis in the development of MEN2A and subsequently MTC in patient PC.0032.01 is highly likely due to the heterozygous pathogenic GOF p.C634Y variant in the *RET* proto-oncogene.

TP53 c.817C>T (p.R273C) Missense Variant

A heterozygous *TP53* c.817C>T (p.R273C) missense variant was detected and validated in patient PC.0051.01.1, who was diagnosed with an embryonal rhabdomyosarcoma (ERMS) of the right buttock at age 1 year, 6 months. The patient's family history showing a specific pattern of familial cancer, together with the patient's specific cancer type and detection of the pathogenic heterozygous *TP53* variant is suggestive of a hereditary CPS known as Li-Fraumeni Syndrome (LFS). LFS is an inherited syndrome whereby individuals have a greatly increased likelihood of developing a spectrum of both childhood- and adulthood-onset malignancies. Patients with LFS do not present with any specific non-malignant phenotypic features but are instead characterised by the occurrence of specific cancer types within the proband and 1st- and 2nd-degree relatives. These cancer types include breast cancer, soft-tissue sarcomas, osteosarcomas, brain tumours, leukaemia, and adrenocortical carcinoma. Patient PC.0051.01 meets both the classical and 2015 revised Chompret LFS diagnostic criteria.

The *TP53* gene encodes the nuclear phosphoprotein p53 which acts within the nucleus of a cell to regulate cell proliferation (Davidoff, 2016; Ganjavi & Malkin, 2002). *TP53*'s primary function as a TSG plays a key role in the cell's response to DNA damage. This includes the induction of cell arrest when DNA damage occurs, the interaction with DNA repair transcription factors and activation of transcription of the appropriate DNA repair genes, and the initiation of apoptosis where DNA damage is too extensive (Tobias, 2016; Hanahan & Weinberg, 2011). LOF mutations in one of the *TP53* domains result in the loss of p53's ability to interact and/or bind to its targets, thereby resulting in unrepaired DNA damage, the uncontrolled proliferation of cells, and evasion from apoptosis, all contributing to carcinogenesis (Carrasco Salas, Lapunzina & Pérez-Martínez, 2017; Rivlin et al., 2011). It has long been established that germline *TP53* mutations are the major underlying cause of the familial clustering of cancers which characterises LFS, including the soft-tissue sarcoma (ERMS) diagnosed in patient PC.0051.01.1 (Olivier, Hollstein & Hainaut, 2010).

The *TP53* p.R273C variant is located within a crucial DNA-binding domain (DBD), spanning amino acids 102-292. This variant region is also a dense mutational hotspot where a high proportion of missense variants reported are pathogenic with no benign missense variants

reported. Several *in vitro* and/or *in vivo* functional studies have shown that variants within the *TP53* DBD, including those affecting codon R273, have a damaging effect on gene function/gene product (Li et al., 2014; Malcikova et al., 2010; Végran et al., 2007). It is these LOF *TP53* variants that are the major contributor to the development of cancer types seen within the LFS tumour spectrum. However, like other TSGs, a LOH of the wild-type allele in germline *TP53* variant carriers within the tumour genome has been reported to occur in LFS families (Donehower et al., 2019; Dearth et al., 2007). Variants within the *TP53* DBD may not only exhibit a LOF of its negative growth signals but may also exert a dominant-negative effect (DNE) on the wild-type p53. Variants in the *TP53* binding domain often result in p53 proteins with no alteration to the tetramerization domain. Thus, the mutant p53 proteins are capable of forming inactive complexes with the wild-type p53 protein. This interferes with the binding of the wild-type p53 to its targets and the regulation of their transcriptional activity (Willis et al., 2004; Milner & Medcalf, 1991).

The International Agency for Research on Cancer (IARC) *TP53* database (link) reports the p.R273C variant to exhibit no functional transcriptional activity and a DNE-LOF mechanism of pathogenicity, as predicted by the (Kato et al., 2003) and (Giacomelli et al., 2018) methodology, respectively. Thus, the mechanism of pathogenesis for the development of the ERMS in patient PC.0051.01.1 is likely due to the LOF of the mutant p53 protein as a result of the heterozygous *TP53* c.1901C>T variant, in addition to the DNE of the mutant allele on the wild type p53 protein, with a possible LOH occurring within the tumour genome. Further *in vitro* and/or functional studies of the *TP53* p.R273C variant in patient PC.0051.01.1 would need to be performed in order to confirm the LOF-DNE mechanism of pathogenesis and to determine if the variant gains any additional oncogenic properties.

There have been several pathogenic variants with alternate amino acid changes affecting the same *TP53* R273 codon reported in ClinVar, HGMD, and UniProt. Variants affecting the R273 codon, along with five other DBD variants, are considered hotspot residues, which collectively account for 30% of all *TP53* missense variants. Variants within the DBD, including R273, have been associated with a more penetrant cancer phenotype and earlier age of onset, which may explain the early onset of ERMS in patient PC.0051.01.1 at age 1 year, 6 months (Rivlin et al., 2011). Although all variants affecting the R273 codon show some level of wild-type p53 LOF, studies have shown that different amino acid changes at this site may result in differing effects on p53 function and/or structure, and subsequently likely clinical implications (Garg et al., 2020; Li et al., 2020, 2014). In particular, the p.R273C and p.R273H variants show a decreased DNA-binding capacity, less active apoptotic pathways, as well as unique stability and aggregation profiles *in*

vitro in comparison to the p.R273G and p.R273L *TP53* missense variants (Garg et al., 2020; Li et al., 2014). This again emphasises the possible gain of oncogenic functions of particular *TP53* missense variants where DNA-contact inhibition alone may not explain the differing effects of different variant changes affecting the same codon. These unique R273 variant changes seen *in vitro* may form the molecular basis for differing cancer phenotypes and therapeutic responses *in vivo*, highlighting the importance of genetic testing and determining the exact variant change (Garg et al., 2020).

Several reputable sources (ClinVar (Accession: VCV000043594.24), UniProt, HGMD, VarSome) have previously reported and classified the *TP53* p.R273C variant as pathogenic and the variant has been linked to over 1331 articles (Mastermind® Link), with the majority mentioning the variant within LFS families and a wide range of cancer types. The IARC *TP53* database reports (61 entries in total) the p.R273C variant in populations around the world, most commonly the UK, USA and India, with no reports of the variant in the SA population. The variant is most commonly found in breast cancer patients (27.87%), with soft-tissue sarcomas being the 4th most common cancer type reported (6.56%). The majority of these entries have missing age of onset information. A literature search and review, external to the IARC *TP53* database, revealed that the *TP53* p.R273C variant has only been reported in the SA once before. However, the variant is reported as somatic, occurring in a patient with laryngeal carcinoma at age 65 years (Barnard et al., 2003). To our knowledge, this is the first report of the germline *TP53* p.R273C variant in a SA paediatric cancer patient. The patient's family history suggests that the variant is likely inherited from the proband's father (affected paternal uncle). However, genetic testing and segregation analysis of both parents would need to be performed to confirm the inherited/*de novo* status of the variant, and whether the father is an unaffected carrier of the variant.

4.3.2 Variants Identified in Patients with Previously Unrelated Phenotypic Evidence

The implementation of NGS technologies in paediatric cancer research has facilitated the screening of a large number of CPGs simultaneously – some of which were not previously considered to play a role in childhood cancer predisposition. This has led to an increase in the number of studies reporting P/LP variants in genes with no previously reported association with the patient's cancer phenotype (Mirabello et al., 2020; Gröbner et al., 2018; Brohl et al., 2017; Franceschi et al., 2017; Parsons et al., 2016; Zhang et al., 2015). These variants often have little to no literature evidence/case studies supporting this relationship and may also have no

obvious/direct role in the pathogenesis of that particular cancer type. This has prompted the discussion as to whether these types of results should be reported as incidental/secondary findings or if these variants are in fact disease-causing/contributing and therefore play a crucial role in paediatric cancer development. Until *in vitro*/*in vivo* functional studies utilising these variants are performed as well as further validation in independent cohorts, their exact contribution to carcinogenesis cannot be elucidated. However, caution should be taken when reporting certain P/LP variants as incidental/secondary findings or not reporting them at all, as an accumulating body of evidence on these variants and subsequent knowledge on new mechanisms of pathogenesis continue to expand.

Recent NGS paediatric cancer studies support the reporting of these variants, with a large proportion of children carrying P/LP variants in 'unknown' CPGs or within genes in pathways not previously implicated in the particular cancer's pathogenesis (Akhavanfard et al., 2020; Byrjalsen et al., 2020). There is also a recent body of literature which supports the association of specific P/LP germline variants to the somatic events that follow, suggesting that these mechanisms of pathogenesis extend beyond that of Knudson's 'two-hit' hypothesis (Capasso et al., 2020; Musa & Grünewald, 2020; Waszak et al., 2020). This research indicates that uncommon germline variants may play a key role in determining the sequential somatic events which occur in genes which are more common/well-known in the cancer's pathogenesis, thus emphasising the importance of reporting these uncommon germline variants. The following sections discuss the P/LP germline variants in this study which have not been previously implicated in the pathogenesis of the patient's cancer phenotype.

BRCA1 c.4199C>T (p.M1400T) Missense Variant

A heterozygous *BRCA1* c.4199C>T (p.M1400T) missense variant was detected and further validated in patient PC.0011.01.1, who was diagnosed with an ERMS of the bladder at age 3 years. The patient's phenotype and genetic pedigree were not suggestive of a heritable/familial CPS. Genetic testing and segregation analysis of the proband's parents would need to be performed to confirm the *de novo*/inheritance status of the variant.

BRCA1 exerts its tumour suppressor activities through its involvement in ubiquitination, DNA damage repair through the homologous recombination repair (HRR) mechanism, and transcriptional regulation of key genes (Clark et al., 2012). Variants which result in a LOF of BRCA1's DNA repair and other tumour suppressor mechanisms result in an accumulation of

unrepaired genetic aberrations contributing to carcinogenesis and subsequently tumour formation (Sæther et al., 2018). *BRCA1* is a well-known causal gene in HBOC syndrome, a hereditary CPS where germline carriers are at a greatly increased risk of developing breast and/or ovarian cancer from early adulthood (Lindor et al., 2008; Wilbur, Rojan & Legare, 2007). Although most associated with breast/ovarian cancers, pathogenic germline *BRCA1* variants have been detected and reported in other cancer types, such as prostate, pancreatic, melanoma, leukaemia, CNS, and soft-tissue sarcomas (Akhavanfard et al., 2020; Petrucelli, Daly & Pal, 2016; Zhang et al., 2015).

The *BRCA1* p.M1400T variant was classified as likely pathogenic according to the ACMG-AMP guidelines with the assistance of VarSome in August 2020. The variant was viewed as likely pathogenic due to the application of the PM1 rule, which was further upgraded from a default strength of moderate to strong. The PM1 rule strength was upgraded as the variant region was reported as a dense mutational hotspot and a UniProt region of interest in VarSome. The variant lies in the coiled-coil (CC) domain of BRCA1 (spanning amino acids 1397 – 1424) which directly binds to PALB2, another key protein which forms an integral part of the BRCA complex required for successful HRR to occur (Sy, Huen & Chen, 2009; Zhang et al., 2009). However, since the original classification, several benign missense variants have been reported in this region (Dines et al., 2020). This indicates that the region may no longer be considered a dense mutational hotspot and may flag the likely pathogenic classification of the variant for future review and possible reclassification. A study using a transcriptional activation assay (Woods et al., 2016), and another study utilising a functionally validated computational prediction model (Hart et al., 2019), in the analysis of the functional impact of *BRCA1* germline variants may provide further evidence to support this reclassification. The results of these studies indicate that the variant may have some effect on transcriptional activation but to a lesser extent than variants within the *BRCA1* C-Terminus (BRCT) domain, and that the variant's overall effect on protein function is likely neutral to intermediary (Hart et al., 2019; Woods et al., 2016).

The interpretation of the functional impact of *BRCA1* missense germline variants remains a challenge and this is reflected in the large proportion (~89%) of *BRCA1* missense variants (including p.M1400T) classified as VUSs in ClinVar (Dines et al., 2020). There is conflicting evidence regarding the VUS classification of *BRCA1* exon 12 (CC domain) variants (Lyra et al., 2021; Dines et al., 2020; Bouwman et al., 2013), with some functional studies supporting the pathogenicity of this variant by reporting a disrupted interaction between these variants and the *PALB2* gene. This disrupted interaction results in below basal wild-type transcriptional activity and defective HRR. This provides evidence that *BRCA1* CC domain variants may

contribute to carcinogenesis through an impaired HR DNA repair mechanism, resulting in genomic instability and subsequent tumorigenesis (Woods et al., 2016; Sy, Huen & Chen, 2009).

Another challenge for the interpretation of *BRCA1* germline variants is the lack of literature regarding the role of the *BRCA* genes in paediatric cancers. *BRCA1/2* testing generally only occurs in adulthood, as the genes are most commonly associated with early-onset adult cancer types. *BRCA1/2* testing likely only occurs in a paediatric setting in rare cases of suspected Fanconi Anaemia (Walsh et al., 2017). However, since the introduction and increased use of NGS multi-CPG panels in the paediatric cancer setting, there has been an increasing number of pathogenic germline *BRCA1/2* variants reported in paediatric patients with apparently sporadic and non-syndromic cancer phenotypes (Akhavanfard et al., 2020; Qin et al., 2020; Brohl et al., 2017; Parsons et al., 2016; Zhang et al., 2015). Including two studies reporting a pathogenic germline *BRCA1* variant in a paediatric patient with an ERMS (Kim et al., 2021b; Li et al., 2021).

This suggests that *BRCA1* germline variants may play a critical role in the development of both paediatric and adult cancers (Kratz et al., 2022; Walsh et al., 2017). Supporting this hypothesis is the evidence regarding the embryonic lethality of a complete LOF of the BRCA1 protein due to biallelic *BRCA1* germline variants, as seen in mouse models (as reviewed by (Evers & Jonkers, 2006)). Thereby indicating that the role of the BRCA1 protein is crucial for normal early life development and that even a slight diminish of the BRCA1 protein function due to monoallelic variants may promote oncogenesis in childhood (Woodward & Meyer, 2021). This evidence encourages the further inclusion and analysis of the *BRCA* genes in future paediatric cancer studies so that their exact role and contribution to the pathogenesis of childhood cancer can be determined. However, it should be noted that the literature reports a higher prevalence of pathogenic *BRCA2* germline variants in comparison to *BRCA1* (Kim et al., 2021b; Li et al., 2021; Walsh et al., 2017). This suggests that although reports of pathogenic *BRCA1* germline variants are increasing, their contribution to carcinogenesis in childhood is still a rare occurrence, with *BRCA2* variants more likely to be involved. While the *BRCA1* p.M1400T variant was classified as likely pathogenic in our study, there is not enough evidence to determine the variant's potential role in the carcinogenesis of the ERMS tumour in patient PC.0011.01.1. Future functional *in vivo/in vitro* studies in the context of the patient's cancer phenotype will aid in determining if the variant contributes to a partial (heterozygous) *BRCA1* DNA repair deficiency, thereby resulting in the accumulation of many unrepaired errors in the genome, including variants in other key CPGs.

In a South African setting, *BRCA1* has almost exclusively been studied in cases of adult-onset familial breast and/or ovarian cancers, with a few founder variants identified (Combrink et al., 2021; Francies et al., 2015; Schoeman et al., 2013). However, *BRCA1* has not been studied within the SA paediatric cancer context. Thus, to our knowledge, this is the first report of the likely pathogenic heterozygous c.4199C>T (p.M1400T) *BRCA1* variant in a South African paediatric patient diagnosed with an ERMS of the bladder.

FAH c.485G>A (p.R162H) Missense Variant

The heterozygous *FAH* c.485G>A (p.R162H) missense variant was detected and validated in patient PC.0009.01.1, who was diagnosed with a juvenile dermatofibrosarcoma protuberans (giant cell fibroblastoma) at age 5 years. The patient's genetic pedigree and phenotype were not suggestive of any particular familial/heritable CPS. Heterozygous carriers of *FAH* variants are often unaffected, thus there is a possibility that the proband inherited the variant from an unaffected parent. Genetic testing and segregation analysis of the proband's parents would need to be performed in order to confirm the inherited/*de novo* status of the variant.

Dermatofibrosarcoma protuberans (DFSP) is a rare cutaneous soft-tissue sarcoma, arising from the dermis layer of the skin, often spreading to the epidermal and subcutaneous layers of the skin, but rarely beyond this (Llombart et al., 2011). The giant cell fibroblastoma histological subtype is referred to as juvenile DFSP, as it commonly affects children and adolescents, with a typical age of onset within the first decade of life (Hassan et al., 2012). Approximately 90% of DFSP cases occur as a result of a somatic chromosomal translocation between a part of the *COL1A1* (Collagen Type 1 Alpha 1) gene on chromosome 17 and the *PDGFB* (Platelet Derived Growth Factor subunit B) gene on chromosome 22 (t(17;22)) (Llombart et al., 2013; Patel et al., 2008). Genetic testing of the probands tumour biopsy would need to be performed to confirm the presence/absence of the t(17;22) translocation. Following the discovery of the t(17;22) chromosomal translocation in the majority of DFSP patients, molecular genetics studies have focused on the sequencing and identification of variation within the tumour genome. Thus, the potential contribution of germline variants to the development of the remaining 10% of DFSP cases is unknown.

Variants within the *FAH* gene have not been previously implicated in the pathogenesis of DFSP. The *FAH* gene encodes the enzyme fumarylacetoacetate hydrolase (FAH) which is the last in a series of enzymes responsible for tyrosine catabolism and is most abundantly expressed in the liver and kidneys (Thompson, Portmann & Roberts, 2012). Germline *FAH* variants result in the expression of an unstable or completely inactive FAH enzyme. Without sufficient FAH enzyme

activity, the by-products of tyrosine are not broken down completely and begin to accumulate in the liver and kidneys. This tyrosine by-product accumulation is toxic to cells through its generation of reactive oxygen species and is responsible for the subsequent development of the characteristic liver and kidney disease seen in patients with Hereditary Tyrosinemia type 1 (HT1). These patients are at an increased risk of developing hepatocellular carcinoma, or more rarely hepatoblastoma, from childhood onwards (Macias et al., 2019; Geppert et al., 2017). HT1 is an autosomal recessive disease and therefore develops due to homozygous, or very rarely compound heterozygous variants in the *FAH* gene (Sniderman King, Trahms & Scott, 2017; Hahn et al., 1995). Heterozygous carriers of P/LP *FAH* variants, such as patient PC.0009.01.1, have sufficient levels of FAH enzyme activity and thus are unaffected carriers of the HT1 disease.

Adding to the likely pathogenic classification of the *FAH* p.R162H variant is its location within a crucial metal ion binding domain (~ spanning amino acids 126-233) containing CA^{2+} ion binding sites. These sites are key in the function of the FAH enzyme in binding the tyrosine by-product substrate for its subsequent degradation (Bateman et al., 2001). The FAH enzyme is active once in its homodimer state, and variants which affect the formation of this homodimer, including those within this CA^{2+} metal ion binding domain, ultimately reduce the overall catalytic activity of the enzyme. Variants which affect the dimer interface and the binding pocket (such as p.R162) loosen the connection between these residues, weakening the stability of the dimer formation, leading to an accumulation of undegraded tyrosine by-products (Dweikat et al., 2021; Macias et al., 2019).

Although variants in the *FAH* gene are primarily associated with HT1-related hepatocellular carcinoma, studies have reported P/LP germline *FAH* variants in other paediatric cancer types, including sarcomas (Kim et al., 2021a, 2021b; Mirabello et al., 2020). Interestingly, a study has shown that expression of the *FAH* gene played a critical role in reprogramming the metabolism of melanoma cells by stimulating anaplerotic reactions, thus allowing the survival and growth of the cancerous skin cells in a hypoxic tumour microenvironment. This suggests that metabolic reprogramming by the *FAH* gene may be essential in the pathogenesis of skin cancers (Gu et al., 2017). Functional *in vitro* and/or *in vivo* studies of the *FAH* p.R162H variant would need to be performed to determine its exact role in carcinogenesis, and if germline variants in the *FAH* gene play a role in the development and/or survival of DFSP. Until such evidence is provided, the heterozygous *FAH* p.R162H cannot be named a causal/driver mutation in the pathogenesis of the DFSP in patient PC.0009.01.1. The *FAH* p.R162H variant may however play a role as a passenger mutation, through an indirect contribution to malignant transformation by an accumulation of oxidative damage to DNA and other cellular components, altered gene expression and DNA repair

mechanisms, and the potential metabolic reprogramming of cancer cells promoting their survival (Squires & Balisteri, 2021; van Ginkel, Pennings & van Spronsen, 2017; Koren & Palladino, 2016). A literature search and review of the reported *FAH* germline variants show no reports of the *FAH* p.R162H variant (Ibarra-González et al., 2019; Ijaz et al., 2016; Angileri et al., 2015). However, this may be due to a lack of reporting of heterozygous carriers of germline variants in the recessive *FAH* gene. To our knowledge this is the first reported case of a South African paediatric patient diagnosed with a juvenile DFSP carrying the likely pathogenic *FAH* c.485G>A (p.R162H) variant.

ERCC3 c.958C>T (p.Q320*) Nonsense Variant

A pathogenic heterozygous *ERCC3* c.958C>T (p.Q320*) nonsense variant was detected and further validated in patient PC.0003.01.1, who was diagnosed with a Wilms' tumour at age 7 years. The patient's phenotype and familial history were not specific to a particular heritable/familial CPS. Individuals carrying heterozygous variants in the recessive *ERCC3* gene may be unaffected, and thus there is a possibility that the variant was inherited from an unaffected parent. Genetic testing and segregation analysis of the proband's parents would need to be performed in order to confirm the inherited/*de novo* status of the variant.

Wilms' tumour is an embryonal tumour of the kidney which primarily develops in childhood. This cancer type is genetically heterogeneous and complex, with an association with>50 CPSs and ~40 (somatic and/or germline variants) CPGs identified in its tumorigenesis (Reviewed by (Treger et al., 2019)). Patient PC.0003.01.1 presented with what would be characterised as a non-syndromic Wilms' tumour, of which causal germline variants have been identified in ~10% of non-syndromic Wilms' tumour cases (Gadd et al., 2017). Heterozygous germline variants in the *ERCC3* gene, as identified in patient PC.0003.01.1, have not been previously confirmed as causal variants for the development of Wilms' tumours.

The *ERCC3* gene, a recessive TSG, encodes the XPB protein, which is an essential subunit of the Transcription Factor II Human (TFIIH) protein complex. The TFIIH complex has two essential roles, one of which is to repair DNA damage using the nucleotide excision repair (NER) mechanism, and the other is in the initiation and regulation of the transcription of genes transcribed by RNA polymerase II (Rimel & Taatjes, 2018; Compe & Egly, 2012; Drapkin et al., 1994). The helicase ATP-binding domain (spanning amino acids 327-488) of *ERCC3* is required for both transcription initiation and NER of DNA damage carried out by the TFIIH complex (Lin, Choi &

Gralla, 2005; Drapkin et al., 1994). The *ERCC3* nonsense variant identified in patient PC.0003.01.1 results in the substitution of the 320 Glutamine amino acid with a premature stop codon (PSC) and thus the translation of a truncated XPB protein lacking the major domains and motifs, including the helicase ATP-binding domain, the ATP-binding site, and the helicase C-terminal (UniProt Entry). Without these domains, the resultant XPB protein would be non-functional during its role in transcription initiation and NER of DNA damage (Popp & Maquat, 2013).

LOF variants in *ERCC3* greatly impact the ability of the TFIIH complex to repair DNA damage, thereby resulting in the accumulation of genetic aberrations, particularly single-stranded breaks (SSBs) (Coin, Oksenych & Egly, 2007). Unrepaired SSBs induce replication blocks which may further progress to lethal double-stranded DNA breaks (Sharma, Lewis & Wlodarski, 2020). Further promoting carcinogenesis are unrepaired DNA lesions occurring in CPGs directly involved in the growth and proliferation of cells, thereby allowing the uncontrolled cell growth and division of mutagenic cells (Ganjavi & Malkin, 2002). Variants within the *ERCC3* subunit of TFIIH also affect transcription initiation, the recruitment of NER factors during DNA repair, and result in the disruption of the post-translational modification of proteins – all of which disrupt normal cellular functioning and promote carcinogenesis (Singh et al., 2015; Oh et al., 2007). Biallelic germline *ERCC3* variants (and other NER genes) are most associated with the autosomal recessive disorder Xeroderma Pigmentosum (XP). Individuals with XP experience extreme photosensitivity and an increased risk of skin cancer at a young age as a result of a diminished ability of the NER pathway to repair UV-induced DNA damage (Sharma, Lewis & Wlodarski, 2020). Individuals with heterozygous *ERCC3* germline variants, like patient PC.0003.01.1, are generally asymptomatic and have not been seen to have an increased risk of skin cancer (Kraemer, DiGiovanna & Tamura, 2003).

Although not previously considered a direct causal gene in the pathogenesis of tumours outside the XP spectrum, several studies have identified *ERCC3* as a putative pathogenic gene associated with an increased risk of overall cancer development (Dumbrava et al., 2019; Feliubadaló et al., 2017; Franceschi et al., 2017). Most of the patients in these studies carrying truncating *ERCC3* variants were reported to have a more severe cancer phenotype, including earlier age of onset, more aggressive clinical course, and higher cancer incidence/genetic anticipation events in cases where multiple family members carried the variant. Individuals in these studies were reported both with and without the presence of additional P/LP variants in other high/moderate-penetrance CPGs. To date, only two paediatric cancer studies have reported the presence of a P/LP

heterozygous germline *ERCC3* variant in a Wilms' tumour patient (Kim et al., 2021a; Byrjalsen et al., 2020). Like patient PC.0003.01.1, the study by Byrjalsen *et al* (2020) identified a pathogenic truncating *ERCC3* mutation in a paediatric non-syndromic Wilms' tumour patient with no history of familial cancer. However, no information regarding the patient's age of onset, disease severity/progression, and proposed mechanism of pathogenicity was provided. Additionally, it should be noted that the study identified a second pathogenic variant in the patient, a large deletion of multiple exons of the *BARD1* gene (Byrjalsen et al., 2020).

The heterozygous *ERCC3* p.Q320* variant is likely not a causal mutation in the pathogenesis of the Wilms' tumour in patient PC.0003.01.1. However, literature evidence suggests that truncating E*RCC3* variants may contribute to carcinogenesis by acting as a modifying factor to another underlying P/LP variant in a high/moderate penetrance CPG. It may additionally accelerate tumorigenesis by contributing to an accumulation of unrepaired DNA damage and disruption in the transcription of other key CPGs. To our knowledge, this is the first report of the pathogenic heterozygous *ERCC3* p.Q320* germline variant in a SA paediatric cancer patient as a literature review revealed the variant has only been reported once globally in a study profiling the circulating tumour DNA in metastatic breast cancer (Cresswell et al., 2020).

RB1 c.1981C>T (p.Arg661Trp) Missense Variant

The pathogenic heterozygous *RB1* c.1981C>T (p.R661W) missense variant was detected and validated in patient PC.0018.01.1, who was diagnosed with a non-syndromic Wilms' tumour at age 6 months. The patient's family history and phenotype did not suggest the presence of a familial/heritable CPS. It is possible that the variant was inherited from an unaffected parent, as not all individuals carrying heterozygous *RB1* variants develop a cancer phenotype. Inheritance/*de novo* status of the variant would need to be confirmed through genetic testing and segregation analysis of the proband's parents.

To date, there have been no literature published nor functional studies implicating heterozygous *RB1* germline mutations as causal variants in the onset and development of Wilms' tumours. Germline *RB1* mutations are almost exclusively known as causal variants in the development of familial retinoblastoma (Stark, 2016; Westman, 2006). However, *RB1* germline variants have also been associated with the development of secondary malignancies outside the retina, including tumours of the brain and nasal cavity, osteosarcomas, soft-tissue sarcomas, melanomas, breast, and lung cancers (Ng et al., 2010; Marees et al., 2008; Kleinerman et al., 2005). The *RB1* gene encodes the pRB nuclear phosphoprotein, known as the 'gate-keeper' of the cell cycle due to its

regulatory activity of the major G1 checkpoint, halting or allowing the cell to progress into DNA synthesis (Davidoff, 2016; Ganjavi & Malkin, 2002). *RB1*-inactivating variants result in a loss of the gene's ability to regulate the cell cycle and to repair DNA damage effectively, thereby promoting oncogenesis (Marshall et al., 2019; di Fiore et al., 2013)

The *RB1* p.R661W variant lies within domain B (spanning amino acids 646 – 771) of the highly conserved pRB pocket. This region is also considered a mutational hotspot without any benign variation, adding to the pathogenic nature of this variant. This region plays a crucial role in the binding of the E2F transcription factor, which when bound to pRB inhibits E2F's transcriptional activity, thereby blocking the cell cycle from progressing and subsequently resulting in the arrest of the cell cycle (Tobias, 2016; Macaluso et al., 2005). Defective binding of pRB-E2F results in the dysregulation of the cell cycle, allowing uncontrolled growth and division of the cell, as well as defective DNA repair mechanisms. The *RB1* gene, like other TSGs, generally requires the somatic loss of the second wild-type allele for the complete loss of the negative growth signals of the gene and subsequent initiation of tumorigenesis. Functional studies have supported a damaging effect on gene function for variants which occur within the B domain of the pRB pocket, including the p.R661W variant identified in patient PC.0018.01.1. These studies show the defective ability of the p.R661W mutant to bind the E2F1 transcription factor as well as induce cell cycle arrest *in vitro*, thereby promoting oncogenesis (Park et al., 2008; Otterson et al., 1999).

The *RB1* p.R661W variant has been previously reported and classified as pathogenic by UniProt, HGMD and ClinVar (Accession: VCV000013087.18), and has also been linked to over 100 articles by Mastermind® (link). However, the majority of these studies report the variant as a causal variant in the development of familial retinoblastoma. Although reported as a pathogenic causal variant, older studies report the variant as low penetrance, with unrelated families varying in retinoblastoma development, including bilateral, unilateral, benign retinomas, and unaffected cases (Otterson et al., 1997; Lohmann et al., 1994; Onadim et al., 1992). The incomplete penetrance of this variant was investigated and described by Otterson *et al* (1999), which demonstrated a temperature-sensitive binding threshold of the pocket domain *in vivo*, whereby variable frequency in tumour development may be explained by reversible fluctuations in this threshold affecting the level of the *RB1* pocket binding activity. Another theory explaining the low-penetrance and variable expressivity of this variant was proposed by Eloy *et al* (2016), which found a parent-of-origin DNA methylation effect. This study demonstrated that the likelihood of being unaffected when the variant is of maternal origin is ~90%, due to the overexpression and retention of sufficient tumour suppressor activity of the mutant maternal allele (Eloy et al., 2016).

The temperature-sensitive pocket domain and parent-of-origin (if maternally inherited) theories may explain why patient PC.0018.01.1 has not developed a retinoblastoma, despite the presence of their pathogenic *RB1* germline variant associated with the autosomal dominant retinoblastoma trait, with a high penetrance of ~90% (Onadim et al., 1992).

Until functional *in vitro* and/or *in vivo* studies of the *RB1* p.R661W variant within the context of Wilms' tumour development are performed, its potential causal role in the development of patient PC.0018.01.1 Wilms' tumour cannot be fully elucidated. However, evidence from functional studies indicates that the *RB1* p.R661W variant may contribute to carcinogenesis indirectly through the dysregulation of the cell cycle, subsequently resulting in uncontrolled cell growth and division, as well as the accumulation of unrepaired genetic aberrations. It is important to note that the heterozygous *RB1* p.R661W variant alone may not be deleterious, and that the variant often corresponds to partial inactivation only. Thus, the variant requires a destabilisation event and a LOH within the tumour genome in order to contribute/initiate a carcinogenic event (Eloy et al., 2016; Park et al., 2008; Otterson et al., 1999, 1997). Sequencing of the tumour genome in patient PC.0018.01.1 would determine if there is a LOH of the remaining R661 wild-type allele and aid in determining the contribution of the p.R661W variant in the development of the patient's Wilms' tumour. A literature search and review revealed no reports of the *RB1* p.R661W variant in the SA population. Thus, to our knowledge, this is the first report of the pathogenic, germline *RB1* c.1981C>T (p.R661W) variant in a SA patient diagnosed with a non-syndromic Wilms' tumour.

4.4 False Positives

The bioinformatics pipeline generated and utilised (as expanded on in the Methods chapter, section 2.2.3) in the analysis of the patient NGS data successfully identified seven P/LP germline variants in 7/32 patients recruited. This bioinformatics pipeline identified and prioritised an additional nine P/LP variants (16 total variants initially prioritised), which through Sanger sequencing were revealed to be false positive variants. These variants were prioritised as they passed all variant filters including the effect of the variant, their *in-silico* pathogenicity/conservation prediction scores, their low frequency within the patient cohort, and no reports of benign/likely benign in ClinVar. The variants were also initially classified as pathogenic/likely pathogenic using the ACMG-AMP guidelines and VarSome. These variants did however have allele depth ratios which deviated from a true homozygous (~0.0:1.0) or heterozygous state (~0.5:0.5) state.

One of the nine additional variants which was detected in the *BLM* gene, had good sequencing quality and an allele ratio close to a true heterozygous state (~0.45:0.55), but was also revealed as

a false positive through validation of the variant via Sanger sequencing. A closer inspection of the IRGV image of this variant revealed a possible alignment error within that region of that patient's NGS data. This alignment error likely resulted in the 'alternate' allele (which was an insertion) being called at a high frequency as the variant detection algorithm tried to compensate for the misalignment. Therefore, for future analysis using this bioinformatics pipeline, it is recommended that in addition to the current variant filtering workflow, the variants which are not close to a true heterozygous and homozygous state be filtered out. Furthermore, it is recommended that manual inspection of variant IRGV plots and/or BAM files is performed during the variant prioritisation stage before the validation of the variant is considered. Manual inspection of all variant IRGV plots/BAM files before the variant prioritisation stage will likely prove to be impractical and unnecessary.

As the additional nine P/LP variants had allele ratios which deviated from a true heterozygous/homozygous state which would indicate the presence of a mutation, we considered the possibility of these variants being somatic mutations. However, there are a few factors which indicate that the variants detected in our study were likely not somatic. Firstly, as the patient DNA was extracted from whole blood samples in patients with solid tumour cancers, the only somatic DNA detectable would be within the patients' circulating cell-free DNA (cfDNA) containing tumour-derived DNA (ctDNA) fragments. The NGS sequencing and somatic analysis workflow used here is not optimised for such analysis and thus may not be sensitive enough to accurately detect somatic variants within ctDNA. For the analysis of somatic variants in solid-tumour cancer patients, it may be better to consider using Thermo Fisher's Scientific Liquid Biopsy Analysis or the Paired Tumour-Normal sequencing workflows (Kleftogiannis et al., 2019). Secondly, it is unlikely that the 9 additional P/LP variants were of somatic origin, as the large majority of our patient cohort were in remission at the time of recruitment and whole blood sampling (remission 6 months/more). The lack of ctDNA can be used as a biomarker of treatment success and remission status (Reinert et al., 2016; Diehl et al., 2008), and thus it is unlikely that ctDNA fragments were still present within the cfDNA of our patients in remission.

4.5 Mutation-Negative Cases

No P/LP germline variants were identified in 25 of the 32 (25/32 ~78%) paediatric solid-tumour cancer patients recruited and investigated in this study. Although high, 78% is not unusual, as this study focused on the germline profiling of paediatric solid-tumour cancer patients, and the

literature estimates that ~90% of all childhood cancers are attributable to somatic variants (Akhavanfard et al., 2020; Zhang et al., 2015). It is important to note that this study aimed to include patients with and without criteria suggestive of a heritable CPS and thus the patient samples analysed in this study were selected based on the extracted DNA quality and purity. However, the diagnostic yield (3/32, 9.4%) of putative disease-causing/contributing P/LP germline variants identified may have been higher if the selection criteria were solely based on confirmed/suspected CPS diagnoses and/or other general criteria suggestive of heritable/familial cancers.

Some of the patients which did not receive a molecular diagnosis were thought to have a higher likelihood of a P/LP germline variant due to the presence of general CPS indicators. This included patients with a bilateral tumour, with a positive family history of cancer, and those with early cancer onset for their particular paediatric cancer type. We suspected the presence of P/LP germline variants in these patients as the occurrence of these factors in childhood is often indicative of a hereditary susceptibility to cancer (Lindor et al., 2008). The cancer types of the patients who did not receive a molecular diagnosis include nephroblastoma (n=9), osteosarcoma (n=4), ovarian germ cell tumour (OGCT, n=4), retinoblastoma (n=3), neuroblastoma (n=2), astrocytoma (n=2), and sacrococcygeal teratoma (n=1).

Over 30 years of cancer genetics research has identified and compiled a list of 152 well-known CPGs seen to confer a moderate-high risk of cancer development (Huang et al., 2018; Zhang et al., 2015; Rahman, 2014b) and an additional >300 candidate CPGs seen to confer a small increase in the risk of cancer development (Rasnic, Linial & Linial, 2020; Hindorff et al., 2009). As this is the first study to investigate the germline variant profiles of South African paediatric solid-tumour patients, it is explorative in its nature and focused on 50 of the well-known childhood CPGs. Thus, it is possible that the mutation-negative patients harbour a P/LP germline variant in one of the other >100 well-known CPGs not tested for here. This includes CPGs associated with the specific cancer types seen in the mutation-negative cases here (Reviewed by (Phi, 2021; Capasso et al., 2020; Czarnecka et al., 2020; Muskens et al., 2019; Treger et al., 2019; Barr & Applebaum, 2018; Pierce, Frazier & Amatruda, 2018)), with the exception of retinoblastoma, whereby hereditary retinoblastoma is predominantly associated with germline variants in the *RB1* gene.

There are also types of genetic aberrations, such as Copy Number Variants (CNVs), implicated in the pathogenesis of paediatric cancers other than the SNVs and small indels (<50bp) focused on in this current study. A key example is seen in the pathogenesis of Wilms' tumours, which includes

cases of uniparental disomy (UPD) of the 11p15.5 imprinted gene cluster and/or epimutation of the same imprinted region in cases of BWS, and the chromosomal microdeletion of 11p13 seen in cases of WAGR syndrome (Romanelli et al., 2011; Breslow et al., 2003). Thus, it is plausible that the development of cancers in some of the mutation-negative patients may be due to variant types not investigated in the present study. The targeted gene panel design utilised in this study sequenced the exonic regions as well as the exon-intron boundaries only, and thus deep intronic and regulatory regions were not sequenced. Although increasing, the knowledge of the role that non-coding variants play in cancer predisposition and/or development is still severely lacking (Yang & Adli, 2019). The majority of non-coding variants found to be associated with disease are regulatory elements (Rojano et al., 2019). This includes perhaps the most well-studied example of non-coding variants in cancer development, which are variants in the promoter region of the Telomerase Reverse Transcriptase (*TERT*) gene (Horn et al., 2013; Huang et al., 2013). These causal variants have been identified in cases of melanoma and brain cancers and are associated with earlier age of onset, aggressive course, and shorter survival time (Schwaederle et al., 2018; Horn et al., 2013). Therefore, a proportion of the mutation-negative cases, although likely marginal, may harbour P/LP causal variants in non-coding regions of the genes not sequenced here.

4.6 Significance & Prognostic Implications of CPG Testing

The detection of cancer-predisposing variants can provide invaluable information to clinicians, benefitting both the proband and their close affected relatives. This project's targeted-CPG NGS workflow facilitated the molecular confirmation of the clinical CPS diagnoses of NF1 and MEN2A in patients PC.0023.01.1 and PC.0032.01.1, respectively. Molecular confirmation of clinical CPS diagnoses, such as NF1 and MEN2A, is beneficial for altering current treatment regimens where applicable, for the implementation of appropriate surveillance measures, as well as the proper management of any future malignant and non-malignant disease manifestations. Both NF1 and MEN2A CPSs may be characterised by non-malignant disease manifestations and thus these patients would both benefit from annual physical examinations and the implementation of appropriate surveillance measures (Li et al., 2020b; Dunning-Davies & Parker, 2016; Rahman, 2014b). Additionally, the molecular CPG testing of this patient cohort, together with obtaining a detailed genetic pedigree of the proband's family history, allowed for the identification of another previously unsuspected CPS, namely LFS in patient PC.0051.01.1. This finding highlights the importance of CPG testing in paediatric cancer patients with a positive history of familial cancer, as the presence of family history and probands cancer type alone is often not sufficient in

identifying a genetic causation related to a familial CPS. This is especially true in cases of LFS, which is not characterised by any specific non-malignant phenotypic features and thus the diagnosis of the CPS may not be considered.

A molecular CPG diagnosis further allows clinicians to implement appropriate screening protocols in cases where children are at an increased risk of developing a secondary malignancy (Travis et al., 2006; Garber & Offit, 2005). This can be seen in the current study through the identification of pathogenic variants in the *NF1* and *TP53* CPGs. The identification of their cancer-predisposing variants aids clinicians in determining the appropriate screening protocols, such as when to begin screening, the suitable screening techniques, and the optimal frequency of screening. This will ensure that patients receive risk-reducing screening measures while avoiding overexposure to radiation and subsequent increased risk of radiation-induced cancers, where possible (Kratz et al., 2017; Sharif et al., 2006). Furthermore, patients with such germline cancer-predisposing variants may be at risk for developing a secondary malignancy from childhood into adulthood and thus it is recommended that screening commence in childhood and continue throughout life (Kratz et al., 2017; Dunning-Davies & Parker, 2016; Villani et al., 2016).

This project's multi-gene NGS panel has also identified four additional P/LP germline variants in CPGs which have not been previously implicated in the direct pathogenesis of the patient's cancer type. The possible role that these variants may play in these patients' cancer development is discussed under section 4.3.2. Although these variants may contribute toward the patient's cancer development in an indirect manner and the patients do not present with any particular CPS, the identification of these P/LP germline variants may still provide clinicians with important prognostic information regarding patient treatment, future surveillance protocols, and future cancer risks. A key example of this is the identification of P/LP germline variants in the DNA repair genes *BRCA1* and *ERCC3* in patients PC.0011.01.1 and PC.0003.01.1 respectively. The identification of germline DNA repair variants provides clinicians with insights into the natural DNA repair potential of the patient, thereby aiding in determining the most beneficial and least cytotoxic therapies accordingly (As reviewed by (Li et al., 2021b; Kiwerska & Szyfter, 2019; Hosoya & Miyagawa, 2014)). Individuals with variants resulting in DNA repair deficiency may result in hypersensitive tumour cells but may also be subject to the formation of secondary tumours due to the genotoxic effects of the treatment on healthy non-tumour cells, as demonstrated in a project by Qin et al (2020). Although patients PC.0011.01.1 and PC.0003.01.1 have already received cancer treatment and are currently in remission, the knowledge gained from these

patients' molecular findings will prove beneficial in the treatment of possible secondary cancer formation.

The detection of cancer-predisposing variants in this study has additionally provided clinicians with information regarding possible inheritance patterns, thereby aiding in determining where sequencing and segregation analysis of the proband's parents is necessary. Confirming inheritance is highly beneficial for unaffected parents whose child carries a germline variant associated with a lifetime risk of cancer development, such as the patients in this study with cancer-predisposing variants identified in the *BRCA1*, *NF1,* and *TP53* genes (Daly et al., 2021; Dullens et al., 2020; Evans et al., 2017; Bougeard et al., 2015). This facilitates the employment of early surveillance measures in mutation-positive parents, with the aim of early detection and treatment where prognostic outcomes are more favourable. This information further facilitates in determining the necessity of testing close relatives, such as the proband's siblings, who have a 50% chance of inheriting the P/LP germline variant in cases where the parents are mutation-positive. This is especially important in cases where germline variants may carry a high risk of cancer onset in childhood, such as the patients in this study with cancer-predisposing variants identified in the *NF1, RET, TP53*, and *RB1* genes (Skalet et al., 2018; Kratz et al., 2017; Wells et al., 2013; Lindor et al., 2008; Ferner et al., 2007).

4.7 Challenges, Limitations & Recommendations

There are a few challenges and limitations recognised in the current study. **(i)** Due to the lack of research on the epidemiology and genetic basis of paediatric cancer in South Africa, this project's 50-candidate gene panel selection process was limited to evidence provided by international NGS paediatric cancer studies. This posed a challenge when designing a targeted CPG panel which would be the most applicable to the SA paediatric cancer population and the most feasible for future implementation in a diagnostic environment. Studies similar to this one with larger sample sizes will greatly contribute to determining which CPGs are the most prevalent within the South African paediatric cancer population and ultimately the most relevant CPGs to include and/or exclude in a diagnostic gene panel design. Future epidemiology studies may also contribute to a CPG panel design by providing updated statistics on the prevalence of different paediatric cancer types seen in SA. These statistics will aid in determining the most useful CPGs to include according to the various cancer types and may also contribute to defining the malignant and non-malignant phenotypic criteria to include for CPG testing.

This project's targeted NGS gene panel was designed using Thermo Fisher's Scientific Ion AmpliSeq online panel design tool (https://www.ampliseq.com/login/login.action), which resulted in a gene panel design with an *in-silico* exon coverage of 100% and a high % of gene uniformity (as determined through wet-lab validation) for 49/50 candidate CPGs selected. **(ii)** Due to specificity issues, the NGS gene panel design covered only 43.7% of the *CDKN1C* gene's coding region, thus possibly limiting variant detection in these missing regions. When designing a multi-gene NGS panel, factors such as primer specificity and primer cross-hybridization need to be considered. Thus, to minimize an *in-silico* predicted increase in off-target amplification, we opted for a design which had high specificity for the *CDKN1C* gene but unfortunately missed 547 bp within exon 1 of the gene. However, CPVs within the *CDKN1C* gene are very rare and are generally attributed to cases of BWS only. Therefore, it is recommended that patients with a suspected BWS diagnosis who are mutation-negative after sequencing with this gene panel undergo additional Sanger sequencing of the missing region.

(iii) One of the biggest challenges encountered in this project was the annotation and analysis of Indels and multi-nucleotide variants (MNVs). The VCF generated from the NGS data was annotated using VEP, which uses the dbNSFP framework to import data on variants using dbNSFP's prediction tools' database. The dbNSFP by definition imports functional data on non-synonymous SNVs i.e., missense variants only, even in cases where the particular tool's algorithm is capable of assessing other variant types -thus leaving variants such as Indels and other complex MNVs unannotated. Therefore, in order to gain sufficient *in-silico* evidence for the subsequent analysis of these variant types, all Indels and MNVs had to be separated from all SNVs and manually submitted to each *in-silico* tool capable of assessing these variant types. This also required manually reformatting the variant data to meet each tool's specific data input requirements. Having to manually separate, reformat and then analyse Indels and MNVs from SNVs greatly added to the data analysis time, a factor which needs to be considered for implementation in a diagnostic timeframe. Perhaps a way to alleviate this analysis burden is for annotation software programs, such as VEP, to improve on their method of annotation, by utilising each tool's individual algorithm when annotating data, instead of importing data for specific variant types only. Additionally, there are very few *in-silico* tools capable of predicting the functional and/or structural effects of Indels (frameshifts in particular) and other MNVs. The lack of *in-silico* pathogenicity and conservation data limits the extent to which these variant types can be analysed and interpreted, especially in comparison to other synonymous, non-synonymous and splice-site SNVs. As the field of bioinformatics continues to expand so will the development of

tools capable of assessing the functional effect of variant classes beyond synonymous and missense variants. Alternatively, the challenges of data formatting and analysis may be alleviated through future improvements to the Ion Reporter software. These future advancements would need to provide a more comprehensive annotation of non-synonymous SNVs, greater flexibility in the downstream filtering of variants, and allow for the batch filtering of patient VCFs.

(iv) Another challenge encountered during the recruitment phase of this study was in obtaining a complete, accurate and informative report of the history of cancers within the patient's family. The patient's clinical family history (FH) is self-reported by their biological parent(s)/legal guardian during recruitment and is thus subjective to the reporter's knowledge, or lack thereof. In our recruitment experience, the reporting of FH often lacked important details, such as the type of cancer and age at diagnosis of the affected family members. Several paediatric patients recruited also came from single-parent households, where the medical FH of the absent parent was not known. The lack of complete or accurate history of familial cancers limited the extent to which the patient's genetic pedigrees could be interpreted for signs of a heritable/familial CPS. Although a positive FH of cancers does not always predict the presence of a heritable CPS, it can be used among other factors when identifying heritable cancers. Some CPS diagnostic criteria are also reliant on a detailed FH, such as the LFS diagnostic criteria, which requires details on a specific spectrum of LFS cancers and the age of diagnosis of the affected relatives. Although studies show that FH should not be used as a sole indicator when identifying heritable cancers, this criterion is often used by clinicians when identifying patients to refer for CPG testing. Thus, the lack or inaccurate record of FH may exclude patients who could benefit from CPG testing.

(v) We acknowledge that although some of the electropherograms produced here were deemed adequate for a 'student-research' environment, that they may not be of the high quality required for validation within a diagnostic environment. Although NGS technologies are constantly improving in accuracy, they are not error-free – and therefore it is recommended that NGS variants undergo validation via a secondary sequencing technique (Matthijs et al., 2016). However, it should be noted that the so-called gold-standard Sanger technique utilised in the validation of these NGS variants, among being costly and time-consuming, is also not error-free. This was experienced in the current project whereby the validation of 16 NGS variants (seven P/LP variants and additional nine false positive variants) using Sanger sequencing proved to be an extensive and laborious process. With the multi-stage process of Sanger sequencing, we experienced many difficulties in optimising PCR conditions for each variant and producing electropherograms with clear shoulder peaks and little background noise, especially for frameshift variants and regions

with homopolymers. The electropherograms produced here were the best of our ability given the validation time frame.

It may also be worth considering the need to utilise Sanger sequencing in the validation of future NGS variants. Several NGS studies have assessed the necessity for the validation of NGS variants using Sanger sequencing in a clinical environment (De Cario et al., 2020; Zheng et al., 2019; Beck et al., 2016; Mu et al., 2016; Baudhuin et al., 2015; Strom et al., 2014; Sikkema-Raddatz et al., 2013). These studies collectively report NGS data to be as accurate as Sanger sequencing results (validation rate range 98.7 – 100%), with one study by De Cario *et al* (2020) reporting their Sanger results (99.7%, n= 942/945) to be less accurate than their NGS results due to Sanger errors including allelic dropout and variants within the primer binding regions. Studies reporting validation ranges <100% (98.7-99.7%) stated that the discrepancy occurred for variants with 'poorer' NGS data quality, as well as for variants within complex genomic regions (Beck et al., 2016; Mu et al., 2016; Strom et al., 2014). These findings indicate that while the validation of complex NGS variants is still recommended, that the validation of high-quality SNVs may be redundant (Baudhuin et al., 2015; Strom et al., 2014). As proposed by De Cario *et al* (2020), the implementation of international clinical NGS guidelines with standardised NGS quality parameters will aid in reducing the Sanger sequencing validation burden. Limiting the need of Sanger validation for all NGS variants will greatly reduce the turnaround times and costs, factors which are important to consider in a LMIC diagnostic environment. This approach to validation may have proven beneficial to the current study as well, as the variants which showed skewed allele ratios proved to be false positives.

4.8 Future Directions

This project's study design included the targeted NGS of 32 patients using a 50 CPG panel, which is suitable for the explorative nature of this study. Future directions of this research should focus on optimising P/LP CPG variant detection (i.e., increasing the diagnostic yield) and on investigating a larger set of moderate-high penetrance CPGs within a larger sample size of SA paediatric cancer patients.

There are several ways this project's design could be expanded on, including WGS, WES, and/or a larger targeted-CPG panel design. The option of using an expanded gene panel may be more suitable as it is more cost-efficient compared to WGS and WES in the SA diagnostic setting. Another challenge with WES/WGS is that they result in more potential CPGs to investigate.

However, if a WES/WGS workflow is used, it may be beneficial to consider a 'tiered' NGS data analysis approach (Kim et al., 2021b, 2021a; Akhavanfard et al., 2020; Byrjalsen et al., 2020), by firstly analysing the well-known moderate to high penetrance CPGs before expanding the analysis to lower penetrance CPGs as well as candidate or novel CPGs. Utilising a 'tiered' WES/WGS data analysis approach focuses on identifying P/LP germline variants in well-known CPGs first, which are often clinically actionable and/or are associated with AD/AR CPSs - while still allowing the expansion of data analysis to other regions of the exome/genome in patients who are mutation-negative after first tier CPG analysis.

Future studies may increase the yield of P/LP CPG variants through the detection and analysis of variant types in addition to the SNVs and small Indels (<50 bp) focused on in the current study. This includes the detection and analysis of larger structural variants (SVs - inversions, translocations), CNVs (duplications, deletions), epimutations (DNA methylation, genomic imprinting, histone modifications), UPD, and chromosomal abnormalities (aneuploidies). With the exception of a few well-known, recurring pathogenic SVs identified in particular cancer types, such as osteosarcomas, medulloblastomas, and Wilms' tumours, the exact contribution of SVs to the development of childhood cancers is unknown (Capasso et al., 2020; Czarnecka et al., 2020; Muskens et al., 2019). However, several NGS paediatric cancer studies have shown the ability to increase the detection of disease-causing/contributing CPG variants through the expansion of their NGS study design to include the analysis of other larger genetic aberrations (Kim et al., 2021b; Newman et al., 2021; Byrjalsen et al., 2020).

Targeted, whole-exome and whole-genome sequencing data can be used for the identification of CNVs and SVs by utilising one or more of the many bioinformatic algorithms, tools, and software developed for CNV/SV discovery from short-read NGS data (Singh et al., 2021; Yao et al., 2019; Zare et al., 2017; Li, 2013; Alkan, Coe & Eichler, 2011). Although CNV/SV detection is possible through targeted panels and WES, the fragmented nature of these techniques may affect the accuracy of CNV discovery and is limiting in cases where SVs or the breakpoints of SVs occur in intronic regions, which are only covered by WGS (Newman et al., 2021; Yao et al., 2019). Perhaps one of the biggest challenges remains the identification of SVs in repetitive regions covered by NGS technologies. Therefore, to allow for a higher degree of accuracy and confidence in SV discovery, future studies should ensure that NGS data is high coverage, and that multiple methods of quality control are implemented to limit background noise and identify potential SV artefacts (As reviewed by (Yang, 2020)).

This project focused on identifying P/LP germline variants in the paediatric probands only. Therefore, the inherited/*de novo* status of the proband's variant can only be hypothesised based on the current phenotype of the parents and close relatives. Prospective work of the current study would include sequencing and segregation analysis of the proband's parents to determine the inheritance pattern, gain information regarding familial cancer risk predictions, and guide the necessity of testing other close relatives, i.e., siblings, aunts/uncles, and grandparents. This would also aid in determining the contribution and pathogenicity of the germline variants in some cases.

As the mechanisms of pathogenicity of the variants identified in this study were deduced using various sources of computational data and available published literature, future functional *in vivo/in vitro* studies are still needed to confirm the variant's mechanism of pathogenicity (*NF1, RET, TP53* variants) and/or the variant's potential role in the development of the patient's cancer phenotype (*BRCA1, ERCC3, FAH, RB1* variants).

During variant classification, using the ACMG-AMP guidelines, we utilised VarSome's statistically significant thresholds to upgrade the strength of certain ACMG-AMP rules where recommended. VarSome generates these statistically significant thresholds during their testing and calibration process, where over a million variants having >= 1 star on ClinVar are used for benchmarking. However, upon later review of variant evidence, we acknowledge that the use of these thresholds may have resulted in the possible 'over classification' of some variants. Therefore, all mutation-positive cases will be further reviewed and discussed by a team of medical geneticists and genetic counsellors to assess variant evidence, finalise variant classification and decide what medical information will be relayed back to the patients in an appropriate clinical setting. This includes reviewing the four P/LP variants identified in the *BRCA1, RB1, ERRC3*, and *FAH* genes, as they may fall under the classification of secondary findings as they were purposely tested as a part of the 50 CPG NGS panel design but have little to no evidence directly relating them to the patient's cancer type. Although all variants have been validated here, they require additional confirmation under diagnostic conditions before they can be fed back to the patients. Mutation-negative results will also be communicated, with emphasis on the possibility that they may harbour a P/LP variant in a CPG not tested for here.

4.9 Conclusion

To our knowledge, this is the first study of its kind to utilise NGS in the germline mutation profiling of paediatric solid-tumour patients in SA. The design and evaluation of this project's targeted 50-CPG panel and our molecular findings have built a foundation for genomic-based

cancer diagnostics in a SA setting. This is exemplified by the panel's diagnostic yield of ~9.4%. The data generated adds to our knowledge and understanding of the contribution of CPGs such as *NF1, RET, TP53,* and their underlying P/LP variants towards paediatric cancer pathogenesis in SA. Additionally, confirmation of a potential underlying cancer genetic causation will provide clinicians with a molecular diagnosis which informs on patient prognosis as well as the implementation of appropriate treatment and screening measures, where applicable. Identification of these P/LP germline variants will also aid in determining where genetic testing of close relatives is necessary, as well as the subsequent prediction of familial cancer risks.

Electronic Sources

Ensembl (2020) *Help - Glossary - Homo_sapiens - Ensembl genome browser 102*. 2020. http://www.ensembl.org/Help/Glossary.

ClinVar Database: https://www.ncbi.nlm.nih.gov/clinvar/

Hazelhurst, S. (2020) *User guide for the ZA-Wits-Core Cluster grid.core.wits.ac.za*. 2020. http://ip-lookup.net/.

Human Gene Mutation Database (HGMD): https://www.hgmd.cf.ac.uk/ac/index.php

Mutalyzer: https://mutalyzer.nl/

REDCap™ Database: https://redcap.core.wits.ac.za/

Thermo Fisher's Scientific Ion AmpliSeq online panel design tool (https://www.ampliseq.com/login/login.action).

Thermo Fisher Scientific (2019a) Ion 510 & Ion 520 & Ion 530 Kit - Chef User Guide. 25 June 2019. Thermo Fisher Scientific. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0016854_510_520_530_Kit_Chef_UG.pdf.

Thermo Fisher Scientific (2018) Ion Ampliseq Library Kit Plus User Guide. 5 November 2018. Thermo Fisher Scientific. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0017003_IonAmpliSeqLibaryKitPlus_UG.pdf.

Thermo Fisher Scientific (2019b) Ion Reporter Software 5.12 User Guide. 12 November 2019. Thermo Fisher Scientific. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0018032_IonReporterSoftware_5_12_UG.pdf.

Thermo Fisher Scientific (2019c) Torrent Suite Software 5.12 User Guide. 5 November 2019. Thermo Fisher Scientific. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0017972_031419_TorrentSuite_5_12_UG_.pdf.

UniProtKB/Swiss-Prot (2018) *What is the canonical sequence? Are all isoforms described in one entry?* 2018. https://www.uniprot.org/help/canonical_and_isoforms.

VarSome: https://varsome.com/

References

Achermann, J.C. & Hughes, I.A. (2016) Pediatric Disorders of Sex Development. In: S. Melmed, K.S. Polonsky, P.R. Larsen, & H.M. Kronenberg (eds.). *Williams Textbook of Endocrinology (Thirteenth Edition)*. 13th edition. Elsevier. pp. 893–963.

Akhavanfard, S., Padmanabhan, R., Yehia, L., Cheng, F. & Eng, C. (2020) Comprehensive germline genomic profiles of children, adolescents and young adults with solid tumors. *Nature Communications*. 11 (1). 2206-2219.

Alkan, C., Coe, B.P. & Eichler, E.E. (2011) Genome structural variation discovery and genotyping. Nature reviews. Genetics. 12 (5), 363-376.

Al-Salameh, A., Baudry, C. & Cohen, R. (2018) Update on multiple endocrine neoplasia Type 1 and 2. *Presse medicale*. 47 (9), 722–731.

American Thyroid Association Guidelines Task Force, Kloos, R.T., Eng, C., Evans, D.B., Francis, G.L., Gagel, R.F., et al. (2009) Medullary thyroid cancer: management guidelines of the American Thyroid Association. *Thyroid*. 19 (6), 565–612.

Angileri, F., Bergeron, A., Morrow, G., Lettre, F., Gray, G., Hutchin, T., Ball, S. & Tanguay, R.M. (2015) Geographical and Ethnic Distribution of Mutations of the Fumarylacetoacetate Hydrolase Gene in Hereditary Tyrosinemia Type 1. *JIMD Reports*. 19, 43-58.

Bajenaru, M.L., Garbow, J.R., Perry, A., Hernandez, M.R. & Gutmann, D.H. (2005) Natural history of neurofibromatosis 1–associated optic nerve glioma in mice. *Annals of Neurology*. 57 (1), 119–127.

Bajenaru, M.L., Hernandez, M.R., Perry, A., Zhu, Y., Parada, L.F., Garbow, J.R. & Gutmann, D.H. (2003) Optic Nerve Glioma in Mice Requires Astrocyte Nf1 Gene Inactivation and Nf1 Brain Heterozygosity. *Cancer Research*. 63 (24), 8573–8577.

Barnard, D., Lehmann, K., Hoal, E.G., van Helden, P.D. & Victor, T.C. (2003) The spectrum of mutations in TP53 in laryngeal cancer patients from a high-incidence population shows similarities to many of the known mutational hotspots. *Cancer genetics and cytogenetics*. 145 (2), 126–132.

Barr, E.K. & Applebaum, M.A. (2018) Genetic Predisposition to Neuroblastoma. *Children (Basel)*. 5 (9), 119.

Basen-Engquist, K., Brown, P., Coletta, A.M., Savage, M., Colbert Maresso, K. & Hawk, E. (2020) Lifestyle and cancer prevention. In: *Abeloff's Clinical Oncology*. 6th edition. Elsevier. pp. 337–374.

Bateman, R.L., Bhanumoorthy, P., Witte, J.F., McClard, R.W., Grompe, M. & Timm, D.E. (2001) Mechanistic Inferences from the Crystal Structure of Fumarylacetoacetate Hydrolase with a Bound Phosphorus-based Inhibitor. *Journal of Biological Chemistry*. 276 (18), 15284–15291.

Baudhuin, L.M., Lagerstedt, S.A., Klee, E.W., Fadra, N., Oglesbee, D. & Ferber, M.J. (2015) Confirming Variants in Next-Generation Sequencing Panel Testing by Sanger Sequencing. *The Journal of Molecular Diagnostics*. 17 (4), 456–461.

Beck, T.F., Mullikin, J.C., NISC Comparative Sequencing Program & Biesecker, L.G. (2016) Systematic evaluation of sanger validation of next-generation sequencing variants. *Clinical Chemistry*. 62 (4), 647–654.

Bergoug, M., Doudeau, M., Godin, F., Mosrin, C., Vallée, B. & Bénédetti, H. (2020) Neurofibromin Structure, Functions and Regulation. *Cells*. 9 (11), 2365-2397

Bianchessi, D., Ibba, M.C., Saletti, V., Blasa, S., Langella, T., Paterra, R., et al. (2020) Simultaneous Detection of NF1, SPRED1, LZTR1, and NF2 Gene Mutations by Targeted NGS in an Italian Cohort of Suspected NF1 Patients. *Genes*. 11 (6), 671-689.

Biesecker, L.G. (2012) Opportunities and challenges for the integration of massively parallel genomic sequencing into clinical practice: Lessons from the ClinSeq project. *Genetics in Medicine*. 14 (4), 393–398.

Bougeard, G., Renaux-Petel, M., Flaman, J.M., Charbonnier, C., Fermey, P., Belotti, M., et al. (2015) Revisiting Li-Fraumeni Syndrome From TP53 Mutation Carriers. *JCO*. 33 (21), 2345–2352.

Breslow, N.E., Norris, R., Norkool, P.A., Kang, T., Beckwith, J.B., Perlman, E.J., et al. (2003) Characteristics and outcomes of children with the Wilms tumor-Aniridia syndrome: a report from the National Wilms Tumor Study Group. *JCO.* 21 (24), 4579–4585.

Brohl, A.S., Patidar, R., Turner, C.E., Wen, X., Song, Y.K., Wei, J.S., et al. (2017) Frequent inactivating germline mutations in DNA repair genes in patients with Ewing sarcoma Germline mutations in Ewing sarcoma. *Genetics in medicine*. 19 (8), 955-958.

Brown, G.R., Simon, M., Wentling, C., Spencer, D.M., Parker, A.N. & Rogers, C.A. (2020) A review of inherited cancer susceptibility syndromes. *JAAPA.* 33 (12), 10–16.

Byrjalsen, A., Hansen, T.V.O., Stoltze, U.K., Mehrjouy, M.M., Barnkob, N.M., et al. (2020) Nationwide germline whole genome sequencing of 198 consecutive pediatric cancer patients reveals a high incidence of cancer prone syndromes. *PLOS Genetics*. 16 (12), e1009231.

Capasso, M., Montella, A., Tirelli, M., Maiorino, T., Cantalupo, S. & Iolascon, A. (2020) Genetic Predisposition to Solid Pediatric Cancers. *Frontiers in Oncology*. 10, 2083. e590033.

Carlo, M.I., Ravichandran, V., Srinavasan, P., Bandlamudi, C., Kemel, Y., et al. (2020) Cancer Susceptibility Mutations in Patients With Urothelial Malignancies. *JCO*. 38 (5), 406–414.

Carrasco Salas, P., Lapunzina, P. & Pérez-Martínez, A. (2017) Genetic predisposition to childhood cancer. *Anales de Pediatría (English Edition)*. 87 (3), 125–127.

Chai, C. (2019) Principle of Emulsion PCR and Its Applications in Biotechnology. *Journal of Animal Reproduction and Biotechnology*. 34 (4), 259–266.

Chang, W., Brohl, A.S., Patidar, R., Sindiri, S., Shern, J.F., et al. (2016) MultiDimensional ClinOmics for Precision Therapy of Children and Adolescent Young Adults with Relapsed and Refractory Cancer: A Report from the Center for Cancer Research. *Clinical Cancer Research*. 22 (15), 3810–3820.

Cheetham, S.W., Gruhl, F., Mattick, J.S. & Dinger, M.E. (2013) Long noncoding RNAs and the genetics of cancer. *British Journal of Cancer*. 108 (12), 2419–2425.

Chompret, A., Brugières, L., Ronsin, M., Gardes, M., Dessarps-Freichey, F., Abel, A., Hua, D., et al. (2000) P53 germline mutations in childhood cancers and cancer risk for carrier individuals. *British Journal of Cancer*. 82 (12), 1932–1937.

Chung, C.C. & Chanock, S.J. (2011) Current status of genome-wide association studies in cancer. *Human Genetics*. 130 (1), 59–78.

Chunn, L.M., Nefcy, D.C., Scouten, R.W., Tarpey, R.P., Chauhan, G., Lim, M.S., et al. (2020) Mastermind: A Comprehensive Genomic Association Search Engine for Empirical Evidence Curation and Genetic Variant Interpretation. *Frontiers in Genetics*. 11, e577152.

Cleary, M.A. (2001) Haploinsufficiency. In: S. Brenner & J.H. Miller (eds.). *Encyclopedia of Genetics*. Elsevier. p. 911.

Combrink, H.M., Oosthuizen, J., Visser, B., Chabilal, N., Buccimazza, I., Foulkes, W.D. & van der Merwe, N.C. (2021) Mutations in BRCA-related breast and ovarian cancer in the South African Indian population: A descriptive study. *Cancer genetics*. 258–259, 1–6.

Compe, E. & Egly, J.M. (2012) TFIIH: when transcription met DNA repair. *Nature reviews. Molecular cell biology*. 13 (6), 343–354.

Costa, Gerber, Ibañez, Melanda, Parise, et al. (2019) Penetrance of the TP53 R337H Mutation and Pediatric Adrenocortical Carcinoma Incidence Associated with Environmental Influences in a 12-Year Observational Cohort in Southern Brazil. *Cancers*. 11 (11), 1804-1820.

Couch, F.J., Shimelis, H., Hu, C., Hart, S.N., Polley, E.C., Na, J., et al. (2017) Associations between cancer predisposition testing panel genes and breast cancer. *JAMA Oncology*. 3 (9), 1190–1196.

Cresswell, G.D., Nichol, D., Spiteri, I., Tari, H., Zapata, L., Heide, T., Maley, C.C., Magnani, L., Schiavon, G., Ashworth, A., Barry, P. & Sottoriva, A. (2020) Mapping the breast cancer metastatic cascade onto ctDNA using genetic and epigenetic clonal tracking. *Nature Communications*. 11 (1), 1446-1458.

Czarnecka, A.M., Synoradzki, K., Firlej, W., Bartnik, E., Sobczuk, P., Fiedorowicz, M., Grieb, P. & Rutkowski, P. (2020) Molecular Biology of Osteosarcoma. *Cancers*. 12 (8), 1–27.

Daly, M.B., Pal, T., Berry, M.P., Buys, S.S., Dickson, P., et al. (2021) Genetic/familial high-risk assessment: Breast, ovarian, and pancreatic, version 2.2021, NCCN Clinical Practice Guidelines in Oncology. *JNCCN*. 19 (1), 77–102.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G. & Durbin, R. (2011) The variant call format and VCFtools. *Bioinformatics*. 27 (15), 2156-2158.

Davidoff, A.M. (2016) Tumor Biology and Environmental Carcinogenesis. In: R. Carachi & J.L. Grosfeld (eds.). *The Surgery of Childhood Tumors*. Third edit. Springer-Verlag Berlin Heidelberg. pp. 19–51.

de Abreu, F.B., Peterson, J.D., Amos, C.I., Wells, W.A. & Tsongalis, G.J. (2016) Effective quality management practices in routine clinical next-generation sequencing. *Clinical Chemistry and Laboratory Medicine*. 54 (5), 761–771.

Dearth, L.R., Qian, H., Wang, T., Baroni, T.E., Zeng, J., Chen, S.W., et al. (2007) Inactive full-length p53 mutants lacking dominant wild-type p53 inhibition highlight loss of heterozygosity as an important aspect of p53 status in human cancers. *Carcinogenesis*. 28 (2), 289–298.

De Cario, R., Kura, A., Suraci, S., Magi, A., Volta, A., Marcucci, R., et al. (2020) Sanger Validation of High-Throughput Sequencing in Genetic Diagnosis: Still the Best Practice? *Frontiers in Genetics*. 11, e592588.

de Leng, W.W.J., Gadellaa-Van Hooijdonk, C.G., Barendregt-Smouter, F.A.S., Koudijs, M.J., Nijman, I., Hinrichs, J.W.J., et al. (2016) Targeted next generation sequencing as a reliable diagnostic assay for the detection of somatic mutations in tumours using minimal DNA amounts from formalin fixed paraffin embedded material. *PLoS ONE*. 11 (2), e0149405.

Diehl, F., Schmidt, K., Choti, M.A., Romans, K., Goodman, S., Li, M., et al. (2008) Circulating mutant DNA to assess tumor dynamics. *Nature medicine*. 14 (9), 985-990.

Diets, I.J., Waanders, E., Ligtenberg, M.J., van Bladel, D.A.G., Kamping, E.J., et al. (2018) High Yield of Pathogenic Germline Mutations Causative or Likely Causative of the Cancer Phenotype in Selected Children with Cancer. *Clinical Cancer Research*. 24 (7), 1594–1603.

di Fiore, R., D'Anneo, A., Tesoriere, G. & Vento, R. (2013) RB1 in cancer: different mechanisms of RB1 inactivation and alterations of pRb pathway in tumorigenesis. *Journal of cellular physiology*. 228 (8), 1676–1687.

Donehower, L.A., Soussi, T., Korkut, A., Weinstein, J.N., Akbani, R. & Correspondence, D.A.W. (2019) Integrated Analysis of TP53 Gene and Pathway Alterations in The Cancer Genome Atlas. *Cell Reports*. 1370–1384.

Douglas, J., Hanks, S., Temple, I.K., Davies, S., Murray, A., Upadhyaya, M., et al. (2003) NSD1 mutations are the major cause of sotos syndrome and occur in some cases of weaver syndrome but are rare in other overgrowth phenotypes. *American Journal of Human Genetics*. 72 (1), 132–143.

Drapkin, R., Reardon, J.T., Ansari, A., Huang, J.C., Zawel, L., Ahn, K., et al. (1994) Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. *Nature*. 368 (6473), 769–772.

Druker, H., Zelley, K., McGee, R.B., Scollon, S.R., Kohlmann, W.K., Schneider, K.A. & Schneider, K.W. (2017) Genetic counselor recommendations for cancer predisposition evaluation and surveillance in the pediatric oncology patient. *Clinical Cancer Research*. 23 (13), e91–e97.

Du, Z. & Lovly, C.M. (2018) Mechanisms of receptor tyrosine kinase activation in cancer. *Molecular Cancer*. 17 (1), 1–13.

Dullens, B., de Putter, R., Lambertini, M., Toss, A., Han, S., et al. (2020) Cancer Surveillance in Healthy Carriers of Germline Pathogenic Variants in BRCA1/2: A Review of Secondary Prevention Guidelines. *Journal of Oncology*. 2020, 13. e9873954.

Dumbrava, E.I., Brusco, L., Daniels, M.S., Wathoo, C., Shaw, K.R., et al. (2019) Expanded Analysis of Secondary Germline Findings From Matched Tumor/Normal Sequencing Identifies Additional Clinically Significant Mutations. *JCO Precision Oncology*. 3, 1–11.

Dunning-Davies, B.M. & Parker, A.P.J. (2016) Annual review of children with neurofibromatosis type 1. *Archives of disease in childhood. Education and practice edition*. 101 (2), 102–111.

Economopoulou, P., Dimitriadis, G. & Psyrri, A. (2015) Beyond BRCA: New hereditary breast cancer susceptibility genes. *Cancer Treatment Reviews*. 41 (1), 1–8.

Eloy, P., Dehainault, C., Sefta, M., Aerts, I., Doz, F., Cassoux, N., Lumbroso le Rouic, L., Stoppa-Lyonnet, D., Radvanyi, F., Millot, G.A., Gauthier-Villars, M. & Houdayer, C. (2016) A Parent-of-Origin Effect Impacts the Phenotype in Low Penetrance Retinoblastoma Families Segregating the c.1981C>T/p.Arg661Trp Mutation of RB1. *PLoS Genetics*. 12 (2), e1005888.

Erdmann, F., Kielkowski, D., Schonfeld, S.J., Kellett, P., Stanulla, M., Dickens, C., et al. (2015) Childhood cancer incidence patterns by race, sex and age for 2000-2006: A report from the South African National Cancer Registry. *International Journal of Cancer*. 136 (11), 2628–2639.

Espenschied, C.R., LaDuca, H., Li, S., McFarland, R., Gau, C.L. & Hampel, H. (2017) Multigene panel testing provides a new perspective on Lynch syndrome. *Journal of Clinical Oncology*. 35 (22), 2568–2575.

Evans, D.G.R., Salvador, H., Chang, V.Y., Erez, A., Voss, S.D., Schneider, K.W., et al. (2017) Cancer and Central Nervous System Tumor Surveillance in Pediatric Neurofibromatosis 1. *Clinical Cancer Research*. 23 (12), e46–e53.

Feliubadaló, L., Tonda, R., Gausachs, M., Trotta, J.R., Castellanos, E., et al. (2017) Benchmarking of Whole Exome Sequencing and Ad Hoc Designed Panels for Genetic Testing of Hereditary Cancer. *Scientific reports*. 7, e37984.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., et al. (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer*. 136 (5), e359-386.

Ferner, R.E., Huson, S.M., Thomas, N., Moss, C., Willshaw, H., Evans, D.G., et al. (2007) Guidelines for the diagnosis and management of individuals with neurofibromatosis 1. *JMG*. 44 (2), 81-88.

Figueiredo, B.C., Sandrini, R., Zambetti, G.P., Pereira, R.M., Cheng, C., Liu, W., et al. (2006) Penetrance of adrenocortical tumours associated with the germline TP53 R337H mutation. *Jou*. 43 (1), 91–96.

Franceschi, S., Spugnesi, L., Aretini, P., Lessi, F., Scarpitta, R., Galli, A., et al. (2017) Whole-exome analysis of a Li-Fraumeni family trio with a novel TP53 PRD mutation and anticipation profile. *Carcinogenesis*. 38 (9), 938–943.

Francies, F.Z., Wainstein, T., de Leeneer, K., Cairns, A., Murdoch, M., Nietz, S., et al. (2015) BRCA1, BRCA2 and PALB2 mutations and CHEK2 c.1100delC in different South African ethnic groups diagnosed with premenopausal and/or triple negative breast cancer. *BMC Cancer*. 15 (1), 1–10.

Friedberg, E.C., Walker, G.C., Siede, W., Wood, R.D., Schultz, R.A. & Ellenberger, T. (2006) *DNA Repair and Mutagenesis, 2nd edition*. 2nd edition. Washington DC, ASM Press.

Fung, A., Horton, S., Zabih, V., Denburg, A. & Gupta, S. (2019) Cost and cost-effectiveness of childhood cancer treatment in low-income and middle-income countries: a systematic review. BMJ Global Health. 4 (5), e001825.

Gadd, S., Huff, V., Walz, A.L., Ooms, A.H.A.G., Armstrong, A.E., et al. (2017) A Children's Oncology Group and TARGET initiative exploring the genetic landscape of Wilms tumor. *Nature genetics*. 49 (10), 1487–1494.

Galvan, A., Ioannidis, J.P.A. & Dragani, T.A. (2010) Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer. *Trends in Genetics*. 26 (3), 132–141.

Ganjavi, H. & Malkin, D. (2002) Genetics of childhood cancer. *CORR*. (401), 75–87.

Gapstur, S.M., Drope, J.M., Jacobs, E.J., Teras, L.R., McCullough, M.L., Douglas, C.E., et al. (2018) A blueprint for the primary prevention of cancer: Targeting established, modifiable risk factors. *CA: A Cancer Journal for Clinicians*. 68 (6), 446–470.

Garber, J.E. & Offit, K. (2005) Hereditary cancer predisposition syndromes. *JCO*.23 (2) pp.276–292.

Garg, A., Hazra, J.P., Sannigrahi, M.K., Rakshit, S. & Sinha, S. (2020) Variable Mutations at the p53-R273 Oncogenic Hotspot Position Leads to Altered Properties. *Biophysical Journal*. 118 (3), 720-728.

Geppert, J., Stinton, C., Freeman, K., Fraser, H., Clarke, A., Johnson, S., et al. (2017) Evaluation of pre-symptomatic nitisinone treatment on long-term outcomes in Tyrosinemia type 1 patients: a systematic review. *Orphanet Journal of Rare Diseases*. 12 (1), 154.

Giacomelli, A.O., Yang, X., Lintner, R.E., McFarland, J.M., Duby, M., et al. (2018) Mutational processes shape the landscape of TP53 mutations in human cancer. *Nature genetics*. 50 (10), 1381–1387.

Gonzalez, K.D., Buzin, C.H., Noltner, K.A., Gu, D., Li, W., Malkin, D. & Sommer, S.S. (2009a) High frequency of de novo mutations in Li-Fraumeni syndrome. *Journal of Medical Genetics*. 46 (10), 689–693.

Gonzalez, K.D., Noltner, K.A., Buzin, C.H., Gu, D., Wen-Fong, C.Y., Nguyen, V.Q., et al. (2009b) Beyond li fraumeni syndrome: Clinical characteristics of families with p53 germline mutations. *Journal of Clinical Oncology*. 27 (8), 1250–1256.

Grant, R.C., Selander, I., Connor, A.A., Selvarajah, S., Borgida, A., Briollais, L., et al. (2015) Prevalence of germline mutations in cancer predisposition genes in patients with pancreatic cancer. *Gastroenterology*. 148 (3), 556–564.

Green, E.D. & Guyer, M.S. (2011) Charting a course for genomic medicine from base pairs to bedside. *Nature*. 470 (7333), 204–213.

Greer, M.L.C. (2018) Imaging of cancer predisposition syndromes. *Pediatric Radiology*. 48 (9), 1364–1375.

Gröbner, S.N., Worst, B.C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., et al. (2018a) The landscape of genomic alterations across childhood cancers. *Nature*. 555 (7696), 321–327. doi:10.1038/nature25480.

Gu, Z., Zhang, H., Li, Y., Shen, S., Yin, X., Zhang, W., et al. (2017) CDC5L drives FAH expression to promote metabolic reprogramming in melanoma. *Oncotarget*. 8 (69), 114328-114343.

Gupta, S., Howard, S.C., Hunger, S.P., Antillon, F.G., Metzger, M.L., Israels, T., Harif, M. & Rodriguez-Galindo, C. (2015) Treating Childhood Cancer in Low- and Middle-Income Countries. In: *Cancer: Disease Control Priorities, Third Edition (Volume 3)*. 3rd edition. The International Bank for Reconstruction and Development / The World Bank. pp. 121–146.

Gurung, A.B. & Bhattacharjee, A. (2015) Significance of Ras Signaling in Cancer and Strategies for its Control. *Oncology & Hematology Review (US)*. 11 (02), 147-152.

Gutmann, D.H., McLellan, M.D., Hussain, I., Wallis, J.W., Fulton, L.L., Fulton, R.S., et al. (2013) Somatic neurofibromatosis type 1 (NF1) inactivation characterizes NF1-associated pilocytic astrocytoma. *Genome research*. 23 (3), 431–439.

Hahn, S.H., Krasnewich, D., Brantly, M., Kvittingen, E.A. & Gahl, W.A. (1995) Heterozygosity for an exon 12 splicing mutation and a W234G missense mutation in an American child with chronic tyrosinemia type 1. *Human mutation*. 6 (1), 66–73.

Hall, M.J., Obeid, E.I., Schwartz, S.C., Mantia-Smaldone, G., Forman, A.D. & Daly, M.B. (2016) Genetic testing for hereditary cancer predisposition: BRCA1/2, Lynch syndrome, and beyond. *Gynecologic Oncology*. 140 (3), 565–574.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*.

Hanahan, D. & Weinberg, R.A. (2011) Hallmarks of Cancer: The Next Generation. *Cell*. 144 (5), 646–674.

Hart, S.N., Hoskin, T., Shimelis, H., Moore, R.M., Feng, B., Thomas, A., et al. (2019) Comprehensive annotation of BRCA1 andBRCA2 missense variants by functionally validated sequence-based computational prediction models. *Genetics in medicine*. 21 (1), 71-80.

Hassan, M.J., Ahmad, S.S., Hasan, M., Basha, M. & Khurram, M.F. (2012) Giant Cell Fibroblastoma: A Case Report. *Oman Medical Journal*. 27 (3), e037.

Helmy, M., Awad, M. & Mosa, K.A. (2016) Limited resources of genome sequencing in developing countries: Challenges and solutions. *Applied and Translational Genomics*. 9, 15–19.

Herskowitz, I. (1987) Functional inactivation of genes by dominant negative mutations. *Nature*. 329 (6136), 219–222.

Hindorff, L., Sethupathy, P., Junkins, H., Ramos, E., JP, Mehta., Collins, F. & Manolio, T. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *PNAS.* 106 (23), 9362–9367.

Hohenstein, P., Pritchard-Jones, K. & Charlton, J. (2015) The yin and yang of kidney development and Wilms' tumors. *Genes and Development*. 29 (5), 467–482.

Hopley, M. & Huddle, K.R. (1997) Multiple endocrine neoplasia type 2A in a black South African family. *South African Medical Journal*. 87 (3), 371–372.

Horn, S., Figl, A., Rachakonda, P.S., Fischer, C., Sucker, A., Gast, A., et al. (2013) TERT promoter mutations in familial and sporadic melanoma. *Science*. 339 (6122), 959–961.

Howard, S.C., Metzger, M.L., Wilimas, J.A., Quintana, Y., Pui, C.-H., Robison, L.L. & Ribeiro, R.C. (2008) Childhood cancer epidemiology in low-income countries. *Cancer*. 112 (3), 461–472.

Huang, F.W., Hodis, E., Xu, M.J., Kryukov, G. v., Chin, L. & Garraway, L.A. (2013) Highly recurrent TERT promoter mutations in human melanoma. *Science*. 339 (6122), 957–959.

Huang, K. lin, Mashl, R.J., Wu, Y., Ritter, D.I., Wang, J., et al. (2018) Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell*. 173 (2), 355-370.

Ibarra-González, I., Fernández-Lainez, C., Alcántara-Ortigoza, M.A., González-Del Angel, A., Fernández-Henández, L., Guillén-López, S., et al. (2019) Mutational spectrum of Mexican patients with tyrosinemia type 1: In silico modeling and predicted pathogenic effect of a novel missense FAH variant. *Molecular Genetics & Genomic Medicine*. 7 (12), e937.

Ijaz, S., Zahoor, M.Y., Imran, M., Afzal, S., Bhinder, M.A., Ullah, I., Cheema, H.A., Ramzan, K. & Shehzad, W. (2016) Direct sequencing of FAH gene in Pakistani tyrosinemia type 1 families reveals a novel mutation. *JPEM*. 29 (3), 327–332.

Jagai, J.S., Messer, L.C., Rappazzo, K.M., Gray, C.L., Grabich, S.C. & Lobdell, D.T. (2017) County-level cumulative environmental quality associated with cancer incidence. *Cancer*. 123 (15), 2901–2908.

Jansen, S., Matthew, C.G., Vermaak, W.J., Kruger, A.J., Ponder, B.A., Calitz, M.S., et al. (1991) Multiple endocrine neoplasia type 2A (MEN 2A) syndrome in a South African family. Biochemistry and molecular genetics. *South African Medical Journal.* 20 (80), 83–87.

Jennings, L.J., Arcila, M.E., Corless, C., Kamel-Reid, S., Lubin, I.M., Pfeifer, J., et al. (2017) Guidelines for Validation of Next-Generation Sequencing–Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *Journal of Molecular Diagnostics*. 19 (3), 341–365.

Kanagal-Shamanna, R. (2016) Emulsion PCR: Techniques and applications. In: *Methods in Molecular Biology*. Humana Press Inc. pp. 33–42.

Karajannis, M.A. & Ferner, R.E. (2015) NEUROFIBROMATOSIS-RELATED TUMORS: EMERGING BIOLOGY AND THERAPIES. *Current opinion in pediatrics*. 27 (1), 33-36.

Kato, S., Han, S.Y., Liu, W., Otsuka, K., Shibata, H., Kanamaru, R. & Ishioka, C. (2003) Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *PNAS.* 100 (14), 8424–8429.

Kaymaz, Y., Odour, C.I., Yu, H., Otieno, J.A., Ong'echa, J.M., Moormann, A.M. & Bailey, J.A. (2017) Comprehensive Transcriptome and Mutational Profiling of Endemic Burkitt Lymphoma Reveals EBV Type-Specific Differences. *Molecular Cancer Research*. 15 (5), 563–576.

Kim, J., Gianferante, M., Karyadi, D.M., Hartley, S.W., Frone, M.N., et al. (2021a) Frequency of Pathogenic Germline Variants in Cancer-Susceptibility Genes in the Childhood Cancer Survivor Study. *JNCI Cancer Spectrum*. 5 (2), pkab007.

Kim, J., Light, N., Subasri, V., Young, E.L., Wegman-Ostrosky, T., et al. (2021b) Pathogenic Germline Variants in Cancer Susceptibility Genes in Children and Young Adults With Rhabdomyosarcoma. *JCO Precision Oncology*. (5), 75–87.

Kjær, S., Kurokawa, K., Perrinjaquet, M., Abrescia, C. & Ibáñez, C.F. (2006) Self-association of the transmembrane domain of RET underlies oncogenic activation by MEN2A mutations. *Oncogene*. 25 (53), 7086–7095.

Kleftogiannis, D., Punta, M., Jayaram, A., Sandhu, S., Wong, S.Q., Gasi Tandefelt, D., et al. (2019) Identification of single nucleotide variants using position-specific error estimation in deep sequencing data. *BMC Medical Genomics*. 12 (1), 115.

Kleinerman, R.A., Tucker, M.A., Tarone, R.E., Abramson, D.H., Seddon, J.M., Stovall, M., et al. (2005) Risk of new cancers after radiotherapy in long-term survivors of retinoblastoma: An extended follow-up. *Journal of Clinical Oncology*. 23 (10), 2272–2279.

Kline, C.N., Joseph, N.M., Grenert, J.P., van Ziffle, J., Talevich, E., et al. (2017) Targeted next-generation sequencing of pediatric neuro-oncology patients improves diagnosis, identifies pathogenic germline mutations, and directs targeted therapy. *Neuro-Oncology*. 19 (5), 699–709.

Knijnenburg, T.A., Wang, L., Zimmermann, M.T., Chambwe, N., Gao, G.F., et al. (2018) Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. *Cell Reports*. 23 (1), 239-254.

Knudson, A.G. (1971) Mutation and cancer: statistical study of retinoblastoma. *PNAS*. 68 (4), 820–823. http://www.ncbi.nlm.nih.gov/pubmed/5279523.

Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C.E., Albarca Aguilera, M., Meyer, R. & Massouras, A. (2018) VarSome: the human genomic variant search engine. *Bioinformatics*. 35 (11), 1978–1980.

Koren, D. & Palladino, A. (2016) Hypoglycemia (Tyrosinemia). In: R.E. Weiss & S. Refetoff (eds.). *Genetic Diagnosis of Endocrine Disorders*. 2nd edition. Academic Press. pp. 31–75.

Kraemer, K.H., DiGiovanna, J.J. & Tamura, D. (2003) Xeroderma Pigmentosum. In: M.P. Adam, G.M. Mirzaa, R.A. Pargon, & S.E. Wallace (eds.). *GeneReviews® [Internet]*. Seattle (WA), University of Washington, Seattle. p. https://www.ncbi.nlm.nih.gov/books/NBK1397/.

Kratz, C.P., Achatz, M.I., Brugi Eres, L., Frebourg, T., Garber, J.E., et al. (2017) Cancer Screening Recommendations for Individuals with Li-Fraumeni Syndrome. *Clinical Cancer Research*. 23 (11), 38–45.

Kuhlen, M., Taeubner, J., Brozou, T., Wieczorek, D., Siebert, R. & Borkhardt, A. (2018) Family-based germline sequencing in children with cancer. *Oncogene*. 1367–1380.

Kværner, A.S., Minaguchi, J., Yamani, N. el., Henriksen, C., Ræder, H., Paur, I., et al. (2018) DNA damage in blood cells in relation to chemotherapy and nutritional status in colorectal cancer patients—A pilot study. *DNA Repair*. 63, 16–24.

LaDuca, H., Polley, E.C., Yussuf, A., Hoang, L., Gutierrez, S., et al. (2020) A clinical guide to hereditary cancer panel testing: evaluation of gene-specific cancer associations and sensitivity of genetic testing criteria in a cohort of 165,000 high-risk patients. *Genetics in Medicine*. 22 (2), 407–415.

Laycock-Van Spyk, S., Thomas, N., Cooper, D.N. & Upadhyaya, M. (2011) Neurofibromatosis type 1-associated tumours: Their somatic mutational spectrum and pathogenesis. *Human Genomics*. 5 (6), 623-690.

Lefter, M., Vis, J.K., Vermaat, M., den Dunnen, J.T., Taschner, P.E.M. & Laros, J.F.J. (2021) Mutalyzer 2: next generation HGVS nomenclature checker. *Bioinformatics*. 37 (18), 2811–2817.

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Oxford University Press, Preprint, viewed 7 September 2022.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25 (16), 2078–2079.

Li, H., Sisoudiya, S.D., Martin-Giacalone, B.A., Khayat, M.M., Dugan-Perez, S., et al. (2021) Germline Cancer Predisposition Variants in Pediatric Rhabdomyosarcoma: A Report From the Children's Oncology Group. *JNCI*. 113 (7), 875–883.

Li, J., Yang, L., Gaur, S., Zhang, K., Wu, X., Yuan, Y.C., et al. (2014) Mutants TP53 p.R273H and p.R273C but not p.R273G enhance cancer cell malignancy. *Human Mutation*. 35 (5), 575–584.

Li, L., Li, X., Tang, Y., Lao, Z., Lei, J. & Wei, G. (2020a) Common cancer mutations R175H and R273H drive the p53 DNA-binding domain towards aggregation-prone conformations. *Physical Chemistry Chemical Physics*. 22 (17), 9225–9232.

Li, S.Y., Ding, Y.Q., Si, Y.L., Ye, M.J., Xu, C.M. & Qi, X.P. (2020b) 5P Strategies for Management of Multiple Endocrine Neoplasia Type 2: A Paradigm of Precision Medicine. *Frontiers in Endocrinology*. 11, e543246.

Lin, Y.C., Choi, W.S. & Gralla, J.D. (2005) TFIIH XPB mutants suggest a unified bacterial-like mechanism for promoter opening but not escape. *Nature Structural and Molecular Biology*. 12 (7), 603–607.

Lindor, N.M., McMaster, M.L., Lindor, C.J. & Greene, M.H. (2008) Concise handbook of familial cancer susceptibility syndromes: Second edition. *JNCI - Monographs*. 2008 (38), 3–93.

Liu, J.F., Konstantinopoulos, P.A. & Matulonis, U.A. (2014) PARP inhibitors in ovarian cancer: Current status and future promise. *Gynecologic Oncology*. 133 (2), 362–369.

Liu, X., Jian, X. & Boerwinkle, E. (2011) dbNSFP: A lightweight database of human nonsynonymous SNPs and their functional predictions. *Human Mutation*. 32 (8), 894–899.

Llombart, B., Monteagudo, C., Sanmartín, O., López-Guerrero, J.A., Serra-Guillén, C., Poveda, A., et al. (2011) Dermatofibrosarcoma protuberans: A clinicopathological, immunohistochemical, genetic

(COL1A1-PDGFB), and therapeutic study of low-grade versus high-grade (fibrosarcomatous) tumors. *Journal of the American Academy of Dermatology*. 65 (3), 564–575.

Llombart, B., Serra-Guillén, C., Monteagudo, C., López Guerrero, J.A. & Sanmartín, O. (2013) Dermatofibrosarcoma protuberans: a comprehensive review and update on diagnosis and management. *Seminars in Diagnostic Pathology*. 30 (1), 13–28.

Lodish, M.B. & Stratakis, C.A. (2008) RET oncogene in MEN2, MEN2B, MTC, and other forms of thyroid cancer: molecular genetics and therapeutic advances. *Expert review of anticancer therapy*. 8 (4), 625-632.

Lohmann, D.R., Brandt, B., Höpping, W., Passarge, E. & Horsthemke, B. (1994) Distinct RB1 gene mutations with low penetrance in hereditary retinoblastoma. *Human genetics*. 94 (4), 349–354.

Loman, N.J., Misra, R. V., Dallman, T.J., Constantinidou, C., Gharbia, S.E., Wain, J. & Pallen, M.J. (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology*. 30 (5), 434–439.

Lovvorn, H.N.I., Pierce, J., Libes, J., Li, B., Wei, Q., Correa, H., et al (2015) Genetic and chromosomal alterations in Kenyan Wilms Tumor. *Genes, Chromosomes & Cancer*. 54 (11), 702-715.

Lu, K.H., Wood, M.E., Daniels, M., Burke, C., Ford, J., Kauff, N.D., et al. (2014) American society of clinical oncology expert statement: Collection and use of a cancer family history for oncology providers. *Journal of Clinical Oncology*. 32 (8), 833–840.

Lucena-Aguilar, G., Sánchez-López, A.M., Barberán-Aceituno, C., Carrillo-Ávila, J.A., López-Guerrero, J.A. & Aguilar-Quesada, R. (2016) DNA Source Selection for Downstream Applications Based on DNA Quality Indicators Analysis. *Biopreservation and Biobanking*. 14 (4), 264–270.

Lynch, H.T., Fusaro, R.M. & Lynch, J. (1995) Hereditary cancer in adults. *Cancer Detection and Prevention*. 19 (3), 219–233. https://pubmed.ncbi.nlm.nih.gov/7750110/.

Macaluso, M., Montanari, M., Cinti, C. & Giordano, A. (2005) Modulation of cell cycle components by epigenetic and genetic events. *Seminars in Oncology*. 32 (5), 452–457.

Macias, I., Laín, A., Bernardo-Seisdedos, G., Gil, D., Gonzalez, E., Falcon-Perez, J.M. & Millet, O. (2019) Hereditary tyrosinemia type I–associated mutations in fumarylacetoacetate hydrolase reduce the enzyme stability and increase its aggregation rate. *The Journal of Biological Chemistry*. 294 (35), 13051-13060.

Madia, F., Worth, A., Whelan, M. & Corvi, R. (2019) Carcinogenicity assessment: Addressing the challenges of cancer and chemicals in the environment. *Environment International*. 128, 417–429.

Magrath, I., Steliarova-Foucher, E., Epelman, S., Ribeiro, R.C., Harif, M., Li, C.K., et al. (2013) Paediatric cancer in low-income and middle-income countries. *The Lancet Oncology*. 14 (3), e104–e116.

Mahdieh, N. & Rabbani, B. (2013) An overview of mutation detection methods in genetic disorders. *Iranian Journal of Pediatrics*. 23 (4), 375–388.

Mai, P.L., Malkin, D., Garber, J.E., Schiffman, J.D., Weitzel, J.N., et al. (2012) Li-Fraumeni syndrome: report of a clinical research workshop and creation of a research consortium. *Cancer Genetics*. 205 (10), 479-487.

Malcikova, J., Tichy, B., Damborsky, J., Kabathova, J., Trbusek, M., Mayer, J. & Pospisilova, S. (2010) Analysis of the DNA-binding activity of p53 mutants using functional protein microarrays and its relationship to transcriptional activation. *Biological Chemistry*. 391 (2–3), 197–205.

Malkin, D. (2011) Li-fraumeni syndrome. *Genes and Cancer*. 2 (4), 475–484.

Mandelker, D., Zhang, L., Kemel, Y., Stadler, Z.K., Joseph, V., et al. (2017) Mutation detection in patients with advanced cancer by universal sequencing of cancer-related genes in tumor and normal DNA vs guideline-based germline testing. *JAMA*. 318 (9), 825–835.

Manor, J. & Lalani, S.R. (2020) Overgrowth Syndromes—Evaluation, Diagnosis, and Management. *Frontiers in Pediatrics*. 8, 574857.

Mardis, E.R. (2008) Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics*. 9, 387–402.

Marees, T., Moll, A.C., Imhof, S.M., de Boer, M.R., Ringens, P.J. & van Leeuwen, F.E. (2008) Risk of Second Malignancies in Survivors of Retinoblastoma: More Than 40 Years of Follow-up. *JNCI*. 100 (24), 1771–1779.

Marshall, A.E., Roes, M. v., Passos, D.T., DeWeerd, M.C., Chaikovsky, A.C., Sage, J., Howlett, C.J. & Dick, F.A. (2019) RB1 Deletion in Retinoblastoma Protein Pathway-Disrupted Cells Results in DNA Damage and Cancer Progression. *Molecular and Cellular Biology*. 39 (16), 105–119.

Matthijs, G., Souche, E., Alders, M., Corveleyn, A., Eck, S., Feenstra, I., Race, V., Sistermans, E., Sturm, M., Weiss, M., Yntema, H., Bakker, E., Scheffer, H. & Bauer, P. (2016) Guidelines for diagnostic next-generation sequencing. *European Journal of Human Genetics*. 24 (1), 2–5.

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.

Melmed, S. & Young, W.F. (2020) Endocrine Hypertension. In: *Williams Textbook of Endocrinology, Fourteenth Edition*. 14th edition. Philadelphia, Elsevier. pp. 542–572.

Merriman, B. & Rothberg, J.M. (2012) Progress in Ion Torrent semiconductor chip based sequencing. *Electrophoresis*. 33 (23), 3397–3417..

Miller, S.A., Dykes, D.D. & Polesky, H.F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*. 16 (3), 1215–1215. doi:10.1093/nar/16.3.1215.

Milner, J. & Medcalf, E.A. (1991) Cotranslation of activated mutant p53 with wild type drives the wild-type p53 protein into the mutant conformation. *Cell*. 65 (5), 765–774.

Mirabello, L., Zhu, B., Koster, R., Karlins, E., Dean, M., et al. (2020) Frequency of Pathogenic Germline Variants in Cancer-Susceptibility Genes in Patients With Osteosarcoma. *JAMA Oncology*. 6 (5), 724–734.

Mody, R.J., Wu, Y.-M., Lonigro, R.J., Cao, X., Roychowdhury, S., et al. (2015) Integrative Clinical Sequencing in the Management of Refractory or Relapsed Cancer in Youth. *JAMA*. 314 (9), 913–925.

Moline, J. & Eng, C. (2011) Multiple endocrine neoplasia type 2: An overview. *Genetics in Medicine*. 13 (9), 755–764.

Moore, S.W., Appfelstaedt, J. & Zaahl, M.G. (2007) Familial medullary carcinoma prevention, risk evaluation, and RET in children of families with MEN2. *Journal of Pediatric Surgery*. 42 (2), 326–332.

Moore, S.W. & Zaahl. M.G. (2010) Chasing the ubiquitous RET proto-oncogene in South African MEN2 families - Implications for the surgeon. *South African Journal of Surgery*. 48 (4), 127–131.

Moore, S.W. & Zaahl, M.G. (2008) Multiple endocrine neoplasia syndromes, children, Hirschsprung's disease and RET. *Pediatric Surgery International*. 24 (5), 521–530.

Mu, W., Lu, H.M., Chen, J., Li, S. & Elliott, A.M. (2016) Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing. *Journal of Molecular Diagnostics*. 18 (6), 923–932.

Musa, J. & Grünewald, T.G.P. (2020) Interaction between somatic mutations and germline variants contributes to clinical heterogeneity in cancer. *Molecular & Cellular Oncology*. 7 (1) e1682924.

Muskens, I.S., Zhang, C., de Smith, A.J., Biegel, J.A., Walsh, K.M. & Wiemels, J.L. (2019) Germline genetic landscape of pediatric central nervous system tumors. *Neuro-Oncology*. 21 (11), 1376-1388.

Nagai, T., Matsumoto, N., Kurotaki, N., Harada, N., Niikawa, N., et al. (2003) Sotos syndrome and haploinsufficiency of NSD 1: Clinical features of intragenic mutations and submicroscopic deletions. *Journal of Medical Genetics*. 40 (4), 285–289.

Nagy, R., Sweet, K. & Eng, C. (2004) Highly penetrant hereditary cancer syndromes. *Oncogene*. 23 (38), 6445–6470.

National Cancer Institute (2019) *Definition of familial cancer - NCI Dictionary of Cancer Terms - National Cancer Institute*. 2019. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/familial-cancer

National Cancer Institute (2021) *Definition of hereditary cancer syndrome - NCI Dictionary of Cancer Terms - National Cancer Institute*. 2021. NCI Dictionaries. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/hereditary-cancer-syndrome

National Comprehensive Cancer Network (2019) *NCCN Clinical Practice Guidelines in Oncology for Genetic/Familial High-Risk Assessment: Colorectal Version 3.2019*.

Neumann, H.P.H., Young, W.F. & Eng, C. (2019) Pheochromocytoma and Paraganglioma. *The New England journal of medicine*. 381 (6), 552–565.

Newman, S., Nakitandwe, J., Kesserwan, C.A., Azzato, E.M., Wheeler, D.A., et al. (2021) Genomes for Kids: The Scope of Pathogenic Mutations in Pediatric Cancer Revealed by Comprehensive DNA and RNA Sequencing. *Cancer discovery*. 11 (12), 3008–3027.

Ng, A.K., Kenney, L.B., Gilbert, E.S. & Travis, L.B. (2010) Secondary Malignancies Across the Age Spectrum. *Seminars in Radiation Oncology*.20 (1) pp.67–78.

Nikiforova, M.N., Wald, A.I., Melan, M.A., Roy, S., Zhong, S., Hamilton, R.L., Lieberman, F.S., Drappatz, J., Amankulor, N.M., Pollack, I.F., Nikiforov, Y.E. & Horbinski, C. (2016) Targeted next-generation sequencing panel (GlioSeq) provides comprehensive genetic profiling of central nervous system tumors. *Neuro-Oncology*. 18 (3), 379–387.

Oberg, J.A., Bender, J.L.G., Sulis, M.L., Pendrick, D., Sireci, A.N., et al. (2016) Implementation of next generation sequencing into pediatric hematology-oncology practice: moving beyond actionable alterations. *Genome Medicine*. 8 (1), 1–19.

Oh, K.S., Imoto, K., Boyle, J., Khan, S.G. & Kraemer, K.H. (2007) Influence of XPB helicase on recruitment and redistribution of nucleotide excision repair proteins at sites of UV-induced DNA damage. *DNA repair*. 6 (9), 1359-1370.

Olivier, M., Goldgar, D.E., Sodha, N., Ohgaki, H., Kleihues, P., Hainaut, P. & Eeles, R.A. (2003) Li-Fraumeni and related syndromes: correlation between tumor type, family structure, and TP53 genotype. *Cancer Research*. 63 (20), 6643–6650.

Olivier, M., Hollstein, M. & Hainaut, P. (2010) TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use. *Cold Spring Harbor Perspectives in Biology*. 2 (1), a001008.

Onadim, Z., Hogg, A., Baird, P.N. & Cowell, J.K. (1992) Oncogenic point mutations in exon 20 of the RB1 gene in families showing incomplete penetrance and mild expression of the retinoblastoma phenotype. *PNAS*. 89 (13), 6177–6181.

Otterson, G.A., Chen, W.D., Coxon, A.B., Khleif, S.N. & Kaye, F.J. (1997) Incomplete penetrance of familial retinoblastoma linked to germ-line mutations that result in partial loss of RB function. *PNAS*. 94 (22), 12036–12040.

Otterson, G.A., Modi, S., Nguyen, K., Coxon, A.B. & Kaye, F.J. (1999) Temperature-Sensitive RB Mutations Linked to Incomplete Penetrance of Familial Retinoblastoma in 12 Families. *American Journal of Human Genetics*. 65 (4), 1040-1046.

Pachnis, V., Mankoo, B. & Costantini, F. (1993) Expression of the c-ret proto-oncogene during mouse embryogenesis. *Development*. 119 (4), 1005–1017.

Park, B.H. & Vogelstein, B. (2003) DNA Repair Pathway Genes. In: D. Kufe, R. Pollock, & R. Weichselbaum (eds.). *Holland-Frei Cancer Medicine*. 6th edition. BC Decker. p. https://www.ncbi.nlm.nih.gov/books/NBK12469/.

Park, Y., Kubo, A., Komiya, T., Coxon, A., Beebe, K., Neckers, L., Meltzer, P.S. & Kaye, F.J. (2008) Low-penetrant RB allele in small-cell cancer shows geldanamycin instability and discordant expression with mutant ras. *Cell cycle (Georgetown, Tex.)*. 7 (15), 2384-2391.

Parsons, W.D., Roy, A., Yang, Y., Wang, T., Scollon, S., et al. (2016) Diagnostic Yield of Clinical Tumor and Germline Whole-Exome Sequencing for Children With Solid Tumors. *JAMA oncology*. 2 (5), 616–624.

Pasmant, E., Parfait, B., Luscan, A., Goussard, P., Briand-Suleau, A., Laurendeau, I., et al. (2015) Neurofibromatosis type 1 molecular diagnosis: what can NGS do for you when you have a large gene with loss of function mutations? *European Journal of Human Genetics*. 23 (5), 596-601.

Patel, K.U., Szabo, S.S., Hernandez, V.S., Prieto, V.G., Abruzzo, L. v., Lazar, A.J.F. & López-Terrada, D. (2008) Dermatofibrosarcoma protuberans COL1A1-PDGFB fusion is identified in virtually all dermatofibrosarcoma protuberans cases when investigated by newly developed multiplex reverse transcription polymerase chain reaction and fluorescence in situ hybridization assays. *Human Pathology*. 39 (2), 184–193.

Paulson, V.A., Rudzinski, E.R. & Hawkins, D.S. (2019) Thyroid cancer in the pediatric population. *Genes*. 10 (9), 723-743.

Petrucelli, N., Daly, M.B. & Pal, T. (2016) BRCA1- and BRCA2-Associated Hereditary Breast and Ovarian Cancer. In: *GeneReviews®*. University of Washington, Seattle. p. http://www.ncbi.nlm.nih.gov/pubmed/20301425.

Phi, J.H. (2021) Sacrococcygeal Teratoma: A Tumor at the Center of Embryogenesis. *Journal of Korean Neurosurgical Society*. 64 (3), 406-413.

Pierce, J.L., Frazier, A.L. & Amatruda, J.F. (2018) Pediatric Germ Cell Tumors: A Developmental Perspective. *Advances in Urology*. 2018, e9059382.

Pierpont, M.E.M., Magoulas, P.L., Adi, S., Kavamura, M.I., Neri, G., Noonan, J., et al. (2014) Cardio-Facio-Cutaneous Syndrome: Clinical Features, Diagnosis, and Management Guidelines. *Pediatrics*. 134 (4), e1162.

Pilarski, R., Carlo, M., Cebulla, C. & Abel-Rahman, M. (2016) BAP1 Tumor Predisposition Syndrome. In: M. Adam, H. Ardinger, & R. Pagon (eds.). *GeneReviews® [Internet]*. Seattle, University of Washington. p. https://www.ncbi.nlm.nih.gov/books/NBK390611/.

Poulet, A., Privat, M., Ponelle, F., Viala, S., Decousus, S., Perin, A., et al. (2016) Improved Efficiency and Reliability of NGS Amplicon Sequencing Data Analysis for Genetic Diagnostic Procedures Using AGSA Software. *BioMed Research International*. 2016, 1–11.

Pros, E., Gómez, C., Martín, T., Fábregas, P., Serra, E. & Lázaro, C. (2008) Nature and mRNA effect of 282 different NF1 point mutations: focus on splicing alterations. *Human Mutation*. 29 (9), e173–e193.

Puñales, M.K., Graf, H., Gross, J.L. & Maia, A.L. (2003) RET codon 634 mutations in multiple endocrine neoplasia type 2: variable clinical features and clinical outcome. *JCEM*. 88 (6), 2644–2649.

Qin, N., Wang, Z., Liu, Q., Song, N., Wilson, C.L., Ehrhardt, M.J., et al. (2020) Pathogenic Germline Mutations in DNA Repair Genes in Combination With Cancer Treatment Exposures and Risk of Subsequent Neoplasms Among Long-Term Survivors of Childhood Cancer. *JCO*. 38 (24), 2728–2740.

Quinlan, A.R. & Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 26 (6), 841–842.

Rahman, N. (2014a) Mainstreaming genetic testing of cancer predisposition genes. *Clinical Medicine, Journal of the Royal College of Physicians of London*. 14 (4), 436–439.

Rahman, N. (2014b) Realizing the promise of cancer predisposition genes. *Nature*. 505 (7483), 302–308.

Rahner, N. & Steinke, V. (2008) Hereditary Cancer Syndromes. *Deutsches Arzteblatt International*. 105 (41), 706–714.

Rasnic, R., Linial, N. & Linial, M. (2020) Expanding cancer predisposition genes with ultra-rare cancer-exclusive human variations. *Scientific Reports*. 10 (1), 1–9.

Reinert, T., Schøler, L. v., Thomsen, R., Tobiasen, H., Vang, S., Nordentoft, I., et al. (2016) Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery. *Gut*. 65 (4), 625–634.

Rekaya, M. ben, Naouali, C., Messaoud, O., Jones, M., Bouyacoub, Y., et al. (2018) Whole Exome Sequencing allows the identification of two novel groups of Xeroderma pigmentosum in Tunisia, XP-D and XP-E: Impact on molecular diagnosis. *Journal of dermatological science*. 89 (2), 172–180.

Rich, T.A., Jimenez, C. & Evans, D.B. (2010) Hereditary Pheochromocytoma and Multiple Endocrine Neoplasia Type 2 (MEN2). In: *Genetic Diagnosis of Endocrine Disorders*. Elsevier Inc. pp. 181–191.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Grody, W.W., et al. (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*. 17 (5), 405–423.

Rimel, J.K. & Taatjes, D.J. (2018) The essential and multifunctional TFIIH complex. *Protein Science: A Publication of the Protein Society*. 27 (6), 1018-1037.

Rivlin, N., Brosh, R., Oren, M. & Rotter, V. (2011) Mutations in the p53 tumor suppressor gene: Important milestones at the various steps of tumorigenesis. *Genes and Cancer*. 2 (4), 466–474.

Robinson, J.T., Thorvaldsdóttir, H., Wenger, A.M., Zehir, A. & Mesirov, J.P. (2017) Variant review with the integrative genomics viewer. *Cancer Research*. 77 (21), e31–e34.

Rodriguez-Galindo, C., Friedrich, P., Alcasabas, P., Antillon, F., Banavali, S., Castillo, L., et al. (2015) Toward the cure of all children with cancer through collaborative efforts: Pediatric oncology as a global challenge. *Journal of Clinical Oncology*. 33 (27), 3065–3073.

Rojano, E., Seoane, P., Ranea, J.A.G. & Perkins, J.R. (2019) Regulatory variants: from detection to predicting impact. *Briefings in Bioinformatics*. 20 (5), 1639–1654.

Romanelli, V., Meneses, H.N.M., Fernández, L., Martínez-Glez, V., Gracia-Bouthelier, R., F Fraga, M., et al. (2011) Beckwith-Wiedemann syndrome and uniparental disomy 11p: Fine mapping of the recombination breakpoints and evaluation of several techniques. *European Journal of Human Genetics*. 19 (4), 416–421.

Sabbagh, A., Pasmant, E., Imbard, A., Luscan, A., Soares, M., Blanché, H., et al. (2013) NF1 Molecular Characterization and Neurofibromatosis Type I Genotype–Phenotype Correlation: The French Experience. *Human Mutation*. 34 (11), 1510–1518.

Sæther, N.H., Skuja, E., Irmejs, A., Maksimenko, J., Miklasevics, E., Purkalne, G. & Gardovskis, J. (2018) Platinum-based neoadjuvant chemotherapy in BRCA1-positive breast cancer: A retrospective cohort analysis and literature review. *Hereditary Cancer in Clinical Practice*. 16 (1), 1-5.

Sahm, F., Schrimpf, D., Jones, D.T.W., Meyer, J., Kratz, A., et al. (2016) Next-generation sequencing in routine brain tumor diagnostics enables an integrated diagnosis and identifies actionable targets. *Acta Neuropathologica*. 131 (6), 903–910.

Saletta, F., Dalla Pozza, L. & Byrne, J.A. (2015) Genetic causes of cancer predisposition in children and adolescents. *Translational pediatrics*. 4 (2), 67–75.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *PNAS.* 74 (12), 5463–5467.

Schneider, K., Zelley, K., Nichols, K.E. & Garber, J. (2019) Li-Fraumeni Syndrome. In: M. Adam, H. Ardinger, R. Pagon, S. Wallace, L. Bean, K. Stephens, & A. Amemiya (eds.). *GeneReviews(®)*. Seattle, University of Washington, Seattle. p. https://www.ncbi.nlm.nih.gov/books/NBK1311/.

Schoeman, M., Apffelstaedt, J.P., Baatjes, K. & Urban, M. (2013) Implementation of a breast cancer genetic service in South Africa - lessons learned. *South African medical journal*. 103 (8), 529–533.

Schwaederle, M., Krishnamurthy, N., Daniels, G.A., Piccioni, D.E., Kesari, S., Fanta, P.T., et al. (2018) Telomerase reverse transcriptase promoter alterations across cancer types as detected by next-generation sequencing: A clinical and molecular analysis of 423 patients. *Cancer*. 124 (6), 1288–1296.

Scott, R.H., Stiller, C.A., Walker, L. & Rahman, N. (2006) Syndromes and constitutional chromosomal abnormalities associated with Wilms tumour. *Journal of Medical Genetics*. 43 (9), 705–715.

Sharif, S., Ferner, R., Birch, J.M., Gillespie, J.E., Gattamaneni, H.R., Baser, M.E. & Evans, D.G.R. (2006) Second primary tumors in neurofibromatosis 1 patients treated for optic glioma: Substantial risks after radiotherapy. *Journal of Clinical Oncology*. 24 (16), 2570–2575.

Sharma, R., Lewis, S. & Wlodarski, M.W. (2020) DNA Repair Syndromes and Cancer: Insights Into Genetics and Phenotype Patterns. *Frontiers in Pediatrics*. 8, e570084.

Sheppard, K.E., Kinross, K.M., Solomon, B., Pearson, R.B. & Phillips, W.A. (2012) Targeting PI3 kinase/AKT/mTOR signaling in cancer. *Critical Reviews in Oncogenesis*. 17 (1), 69–95.

Sikkema-Raddatz, B., Johansson, L.F., de Boer, E.N., Almomani, R., Boven, L.G., van den Berg, M.P., et al. (2013) Targeted Next-Generation Sequencing can Replace Sanger Sequencing in Clinical Diagnostics. Human Mutation. 34 (7), 1035–1042.

Sijmons, R.H. (2010) Identifying Patients with Familial Cancer Syndromes. In: D. Riegert-Johnson, L. Boardman, & T. Hefferon (eds.). *Cancer Syndromes*. National Center for Biotechnology Information (US). p. https://www.ncbi.nlm.nih.gov/books/NBK45295/.

Singh, A., Compe, E., le May, N. & Egly, J.M. (2015) TFIIH Subunit Alterations Causing Xeroderma Pigmentosum and Trichothiodystrophy Specifically Disturb Several Steps during Transcription. *The American Journal of Human Genetics*. 96 (2), 194–207.

Singh, A.K., Olsen, M.F., Lavik, L.A.S., Vold, T., Drabløs, F. & Sjursen, W. (2021) Detecting copy number variation in next generation sequencing data from diagnostic gene panels. BMC Medical Genomics. 14 (1), 1–12.

Skalet, A.H., Gombos, D.S., Gallie, B.L., Kim, J.W., Shields, C.L., Marr, B.P., Plon, S.E. & Chévez-Barrios, P. (2018) Screening Children at Risk for Retinoblastoma: Consensus Report from the American Association of Ophthalmic Oncologists and Pathologists. *Ophthalmology*. 125 (3), 453–458.

Sniderman King, L., Trahms, C. & Scott, R.C. (2017) Tyrosinemia Type I. In: M.P. Adam, H.H. Ardinger, R.A. Pagon, & S.E. Wallace (eds.). *Gene Reviews*. Seattle (WA), University of Washington, Seattle. pp. 2132–2133.

Squires, J.E. & Balisteri, W. (2021) Other Inherited Metabolic Disorders of the Liver. In: M. Feldman, L.S. Friedman, L.J. Brandt, R.T. Chung, D.T. Rubin, & C.M. Wilcox (eds.). *Sleisenger and Fordtran's Gastrointestinal and Liver Disease.* 11th edition. Elsevier. pp. 1189–1209. https://www.clinicalkey.com/#!/content/book/3-s2.0-B9780323609623000771.

Stark, A.E. (2016) Determining the incidence of the hereditary form of retinoblastoma. *Annals of Translational Medicine*. 4 (9).

Stefan, D.C. (2015) Cancer Care in Africa: An Overview of Resources. *Journal of Global Oncology*. 1 (1), 30–36.

Stiller, C.A. (2004) Epidemiology and genetics of childhood cancer. *Oncogene*. 23 (38), 6429–6444.

Stones, D.K., De Bruin, G.P., Esterhuizen, T.M. & Stefan, D.C. (2014) Childhood cancer survival rates in two South African units. *South African Medical Journal*. 104 (7), 501–504.

Strom, S.P., Lee, H., Das, K., Vilain, E., Nelson, S.F., Grody, W.W. & Deignan, J.L. (2014) Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory. *Genetics in Medicine*. 16 (7), 510–515.

Sy, S.M.H., Huen, M.S.Y. & Chen, J. (2009) PALB2 is an integral component of the BRCA complex required for homologous recombination repair. *PNAS.* 106 (17), 7155-7160.

Talseth-Palmer, B.A., Wijnen, J.T., Grice, D.M. & Scott, R.J. (2013) Genetic modifiers of cancer risk in Lynch syndrome: A review. *Familial Cancer*. 12 (2), 207–216.

Tatton-Brown, K., Douglas, J., Coleman, K., Baujat, G., Cole, T.R.P., Das, S., et al. (2005) Genotype-phenotype associations in Sotos syndrome: An analysis of 266 individuals with NSD1 aberrations. *American Journal of Human Genetics*. 77 (2), 193–204.

Thompson, D. & Easton, D.F. (2002) Cancer incidence in BRCA1 mutation carriers. *Journal of the National Cancer Institute*. 94 (18), 1358–1365.

Thompson, R.J., Portmann, B.C. & Roberts, E.A. (2012) Genetic and metabolic liver disease. In: A.D. Burt, B.C. Portmann, & L.D. Ferrell (eds.). *MacSween's Pathology of the Liver*. 6th edition. Elsevier. pp. 157–259.

Tobias, E.S. (2016) Genetic counseling for childhood tumors and inherited cancer-predisposing syndromes. In: R. Carachi & J.L. Grosfeld (eds.). *The Surgery of Childhood Tumors*. 3rd edition. Berlin, Heidelberg, Springer Berlin Heidelberg. pp. 51–66.

Tobias, E.S. (2013) The Molecular Biology of Cancer. In: D. Rimoin, R. Pyeritz, & B. Korf (eds.). *Emery and Rimoin's Principles and Practice of Medical Genetics*. 6th edition. Elsevier. pp. 1–44.

Toledo, S.P. de A., dos Santos, M.A.C.G., de Almeida, R. & Lourenço Junior, D.M. (2006) Impact of RET proto-oncogene analysis on the clinical management of multiple endocrine neoplasia type 2. *Clinics*. 61 (1), 59–70.

Travis, L.B., Rabkin, C.S., Brown, L.M., Allan, J.M., Alter, B.P., et al. (2006) Cancer Survivorship—Genetic Susceptibility and Second Primary Cancers: Research Strategies and Recommendations. *JNCI: Journal of the National Cancer Institute*. 98 (1), 15–25.

Treger, T.D., Chowdhury, T., Pritchard-Jones, K. & Behjati, S. (2019) The genetic changes of Wilms tumour. *Nature Reviews Nephrology*. 15 (4), 240–251.

Tuttle, R.M., Ball, D.W., Byrd, D., Daniels, G.H., Dilawari, R.A., et al. (2010) Medullary carcinoma: Clinical practice guidelines in oncology™. *JNCCN*. 8 (5), 512–530.

Upadhyaya, M., Kluwe, L., Spurlock, G., Monem, B., Majounie, E., Mantripragada, K., et al. (2008) Germline and somatic NF1 gene mutation spectrum in NF1-associated malignant peripheral nerve sheath tumors (MPNSTs). *Human mutation*. 29 (1), 74–82.

Valdés, N., Navarro, E., Mesa, J., Casterás, A., Alcázar, V., Lamas, C., Tébar, J., Castaño, L., Gaztambide, S. & Forga, L. (2015) RET Cys634Arg mutation confers a more aggressive multiple endocrine neoplasia type 2A phenotype than Cys634Tyr mutation. *European journal of endocrinology*. 172 (3), 301–307.

van Ginkel, W.G., Pennings, J.P. & van Spronsen, F.J. (2017) Liver Cancer in Tyrosinemia Type 1. *Advances in experimental medicine and biology*. 959, 101–109. doi:10.1007/978-3-319-55780-9_9.

Vasen, H.F.A., Blanco, I., Aktan-Collan, K., Gopie, J.P., Alonso, A., et al. (2013) Revised guidelines for the clinical management of Lynch syndrome (HNPCC): Recommendations by a group of European experts. *Gut*. 62 (6), 812–823.

Vassilev, A. & DePamphilis, M.L. (2017) Links between DNA replication, stem cells and cancer. *Genes*. 8 (2), 45-78.

Végran, F., Boidot, R., Oudin, C., Defrain, C., Rebucci, M. & Lizard-Nacol, S. (2007) Association of p53 gene alterations with the expression of antiapoptotic survivin splice variants in breast cancer. *Oncogene*. 26 (2), 290–297.

Veitia, R.A. (2007) Exploring the molecular etiology of dominant-negative mutations. *Plant Cell*. 19 (12), 3843–3851.

Velázquez, C., Lastra, E., Avila Cobos, F., Abella, L., De La Cruz, V., Hernando, B.A., Hernández, L., et al. (2020) A comprehensive custom panel evaluation for routine hereditary cancer testing: Improving the yield of germline mutation detection. *Journal of Translational Medicine*. 18 (1), 232-242.

Vencken, P.M.L.H., Reitsma, W., Kriege, M., Mourits, M.J.E., de Bock, G.H., de Hullu, J.A., et al. (2013) Outcome of BRCA1- compared with BRCA2-associated ovarian cancer: A nationwide study in the Netherlands. *Annals of Oncology*. 24 (8), 2036–2042.

Villani, A., Greer, M.L.C., Kalish, J.M., Nakagawara, A., Nathanson, K.L., Pajtler, K.W., et al. (2017) Recommendations for cancer surveillance in individuals with RASopathies and other rare genetic conditions with increased cancer risk. *Clinical Cancer Research*. 23 (12), e83–e90.

Villani, A., Shore, A., Wasserman, J.D., Stephens, D., Kim, R.H., Druker, H., et al. (2016) Biochemical and imaging surveillance in germline TP53 mutation carriers with Li-Fraumeni syndrome: 11 year follow-up of a prospective observational study. *The Lancet. Oncology*. 17 (9), 1295–1305.

von Stedingk, K., Stjernfelt, K.J., Kvist, A., Wahlström, C., Kristoffersson, U., Stenmark-Askmalm, M., et al. (2021) Prevalence of germline pathogenic variants in 22 cancer susceptibility genes in Swedish pediatric cancer patients. *Scientific Reports*. 11 (1), e5307.

Wagener, R., Taeubner, J., Walter, C., Yasin, L., Alzoubi, D., Bartenhagen, C., et al. (2021) Comprehensive germline-genomic and clinical profiling in 160 unselected children and adolescents with cancer. *European Journal of Human Genetics*. 29 (8), 1301-1311.

Walsh, M.F., Kennedy, J., Harlan, M., Kentsis, A., Shukla, N., Musinsky, J., Roberts, S., Kung, A.L., Robson, M., Kushner, B.H., Meyers, P. & Offit, K. (2017) Germline BRCA2 mutations detected in pediatric sequencing studies impact parents' evaluation and care. *Cold Spring Harbor Molecular Case Studies*. 3 (6), a001925.

Wang, Q. (2016) Cancer predisposition genes: Molecular mechanisms and clinical impact on personalized cancer care: Examples of Lynch and HBOC syndromes. *Acta Pharmacologica Sinica*. 37 (2), 143–149.

Waszak, S.M., Robinson, G.W., Gudenas, B.L., Smith, K.S., Forget, A., et al. (2020) Germline Elongator mutations in Sonic Hedgehog medulloblastoma. *Nature*. 580 (7803), 396–401. doi:10.1038/s41586-020-2164-5.

Weksberg, R., Butcher, D.T., Grafodatskaya, D., Choufani, S. & Tycko, B. (2013) Epigenetics. In: D. Rimoin, R. Pyeritz, & B. Korf (eds.). *Emery and Rimoin's Principles and Practice of Medical Genetics*. 6th edition. Elsevier Ltd. pp. 1–31.

Wells, S.A., Asa, S.L., Dralle, H., Elisei, R., Evans, D.B., Gagel, R.F., et al. (2015) Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid*. 25 (6), 567–610.

Wells, S.A.J., Pacini, F., Robinson, B.G. & Santoro, M. (2013) Multiple Endocrine Neoplasia Type 2 and Familial Medullary Thyroid Carcinoma: An Update. *JCEM.* 98 (8), 3149-3164.

Westman, J.A. (2006) Hereditary Syndromes of Tumor Suppressor Genes - Retinoblastoma. In: *Medical Genetics For the Modern Clinician*. Lippincott Williams & Wilkins. pp. 110–111.

Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., et al. (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 452 (7189), 872–876. doi:10.1038/nature06884.

Whitworth, J., Smith, P.S., Martin, J.E., West, H., Luchetti, A., et al. (2018) Comprehensive Cancer-Predisposition Gene Testing in an Adult Multiple Primary Tumor Series Shows a Broad Range of Deleterious Variants and Atypical Tumor Phenotypes. *American Journal of Human Genetics*. 103 (1), 3–18.

Wilbur, J.S., Rojan, A.A. & Legare, R.D. (2007) The Genetics of Breast and Gynecologic Cancers. In: A.I. Sokol & E.R. Sokol (eds.). *General Gynecology*. Elsevier. pp. 43–71.

Willis, A., Jung, E.J., Wakefield, T. & Chen, X. (2004) Mutant p53 exerts a dominant negative effect by preventing wild-type p53 from binding to the promoter of its target genes. *Oncogene*. 23 (13), 2330–2338.

Woods, N.T., Baskin, R., Golubeva, V., Jhuraney, A., De-Gregoriis, G., Vaclova, T., et al. (2016) Functional assays provide a robust tool for the clinical annotation of genetic variants of uncertain significance. *NPJ Genomic Medicine*. 1, e16001.

Worst, B.C., Tilburg, C.M. van, Balasubramanian, G.P., Fiesel, P., Witt, R., et al. (2016) Next-generation personalised medicine for high-risk paediatric cancer patients – The INFORM pilot study. *European Journal of Cancer*. 65, 91–101.

Yang, J. & Adli, M. (2019) Mapping and making sense of noncoding mutations in the genome. *American Association for Cancer Research*. 79 (17), 4309–4314.

Yang, L. (2020) A practical guide for structural variation detection in human genome. *Current Protocols in Human Genetics*. 107 (1), e103.

Yao, R., Yu, T., Qing, Y., Wang, J. & Shen, Y. (2019) Evaluation of copy number variant detection from panel-based next-generation sequencing data. *Molecular Genetics & Genomic Medicine*. 7 (1), e00513.

Yeo, Z.X., Wong, J.C.L., Rozen, S.G. & Lee, A.S.G. (2014) Evaluation and optimisation of indel detection workflows for ion torrent sequencing of the BRCA1 and BRCA2 genes. *BMC Genomics*. 15 (1), e516.

Yohe, S., Hauge, A., Bunjer, K., Kemmer, T., Bower, M., Schomaker, M., et al. (2015) Clinical validation of targeted next-generation sequencing for inherited disorders. *Archives of Pathology and Laboratory Medicine*. 139 (2), 204–210.

Young, W.F. (2013) Secondary Hypertension: Pheochromocytoma. In: H.R. Black & W.J. Elliott (eds.). *Hypertension: A Companion to Braunwald's Heart Disease: Second Edition*. 2nd edition. W.B. Saunders. pp. 89–98.

Zare, F., Dow, M., Monteleone, N., Hosny, A. & Nabavi, S. (2017) An evaluation of copy number variation detection tools for cancer using whole exome sequencing data. BMC Bioinformatics. 18 (1), 286.

Zhang, F., Ma, J., Wu, J., Ye, L., Cai, H., Xia, B. & Yu, X. (2009) PALB2 Links BRCA1 and BRCA2 in the DNA-Damage Response. *Current Biology*. 19 (6), 524-529.

Zhang, J., Walsh, M.F., Wu, G., Edmonson, M.N., Gruber, T.A., et al. (2015) Germline Mutations in Predisposition Genes in Pediatric Cancer. *New England Journal of Medicine*. 373 (24), 2336–2346.

Zheng, J., Zhang, H., Banerjee, S., Li, Y., Zhou, J., et al. (2019) A comprehensive assessment of Next-Generation Sequencing variants validation using a secondary technology. *Molecular Genetics & Genomic Medicine*. 7 (7), e00748.

Zorgania, A.E., Piriea, F.J. & Motalaa, A.A. (2018) Characteristics and outcome of patients with pheochromocytoma at a tertiary endocrinology clinic in Durban, South Africa over 14 years. *Journal of Endocrinology, Metabolism and Diabetes of South Africa*. 23 (2), 52–58.

Appendices

Appendix A: Plagiarism Declaration



PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I _____Erin Manolas_____ (Student number: _733739_____) am a student registered for the degree of ___Master of Science in Medicine – Human Genetics_(MRA05)_____ in the academic year _2023___.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.
- I have included as an appendix a report from "Turnitin" (or other approved plagiarism detection) software indicating the level of plagiarism in my research document.

Signature: E Manolas Date: _____10/10/2023_____

Appendix B: Ethics & Recruitment Documentation

Appendix B1.1: PeCan Project Ethics Certificate

R14/49 Dr L Lamola, et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M180855

NAME: (Principal Investigator)	Dr L Lamola, et al
DEPARTMENT:	School of Pathology Department of Human Genetics National Health Laboratory Service
PROJECT TITLE:	Inherited cancer predisposing mutations in a cohort of childhood cancer patients
DATE CONSIDERED:	31/08/2018
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Not applicable
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	26/02/2019

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on 3rd floor, Phillip V Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.
I/We fully understand the conditions under which I am/we are authorised to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated from the research protocol as approved, I/we undertake to resubmit to the Committee. I agree to submit a yearly progress report. When a funder requires annual re-certification, the application date will be one year after the date of the meeting when the study was initially reviewed. In this case, the study was initially reviewed in August and will therefore reports and re-certification will be due early in the month of August each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature — Date: 26/02/2019

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix B1.2: Current Study Ethics Certificate

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R14/49 Ms E Manolas

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M190751

NAME: (Principal Investigator)	Ms E Manolas
DEPARTMENT:	School of Pathology Department of Human Genetics National Health Laboratory Service
PROJECT TITLE:	Mutation profiling of pediatric solid tumours in a cohort of South African patients
DATE CONSIDERED:	2019/07/26
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Dr L Lamola
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	2019/10/23

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on the 3rd Floor, Phillip Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to submit details to the Committee. I **agree to submit a yearly progress report**. When a funder requires annual re-certification, the application date will be one year after the date when the study was initially reviewed. In this case, the study was initially reviewed in July and will therefore reports and re-certification will be due early in the month of July each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

E Manolas — Principal Investigator Signature

23/10/2019 — Date

PLEASE QUOTE THE CLEARANCE CERTIFICATE NUMBER IN ALL ENQUIRIES

Appendix B2.1: Parent/Legal Guardian Information Sheets

NATIONAL HEALTH LABORATORY SERVICE
School of Pathology, University of the Witwatersrand
DIVISION OF HUMAN GENETICS
Hospital Street, Johannesburg, 2001 | PO Box 1038, Johannesburg, 2000
[T]: +27 11 489 9211 | [F]: +27 11 489 9226 | [E]: human.genetics@nhls.ac.za

Protocol: M180855
Protocol: M190751

Inherited cancer-predisposing mutations in a cohort of childhood cancer patients

Dear Potential Participant,

Hello, we are Dr Lindiwe Lamola, Dr Candice Feben, Prof. Amanda Krause, Dr Njabulo Mabaso and Ms Erin Manolas from the Division of Human Genetics at the University of the Witwatersrand and National Health Laboratory Service. We are doing research on childhood cancers. We work with other members of the Division of Human Genetics to study the genetic factors that cause childhood cancers in South Africans.

As we are trying to understand the genetic causes of cancer, we would like to invite your child to take part in this study, because your child has cancer. Participation in the study will mean answering some questions about the health and background of your child and giving a blood and/or saliva sample for genetic studies. We will also examine your affected child, review their clinical file and investigations, and take a blood and/ or saliva sample from them.

What are childhood cancers?

Childhood cancers are the cancers that develop in children from birth to 18 years of age. Childhood cancer is not just one disease, but a group of several major types of disease. There are many types of childhood cancers; these may include; neuroblastoma, retinoblastoma, Wilms tumour, and cancers of the brain, bone, and soft tissue. Childhood cancers are not common and differ from adult cancers in the way they grow, spread and how they are treated.

What is cancer?

Cancer is the abnormal growth of cells. Cells are small basic building blocks of all living things. All parts of the body are made up of cells. Cancer develops when damaged cells grow and spread faster than normal healthy cells.

What causes childhood cancers?

Childhood cancers are complex diseases and doctors sometimes aren't sure why some children get cancer and others don't. However, about 10% (1 in 10) of all cancers in children are caused by inherited changes in your genes – inherited means these changes can be passed from parents to children. For this study, we are interested in studying the genetic causes of childhood cancers in order to understand their role in the development of cancer.

What will the study involve?

In this study, we would like to understand the factors which may have played a part in your child developing cancer. To do this, we want to learn more about your family, specifically whether other family members have had cancer, and to learn more about your child, specifically related to his/her type of cancer, how old the child was when the cancer was diagnosed and whether there are any other problems which your child may have been born with or may have developed over time. We would like to take a full family history from you and then perform a clinical examination of your child, looking at his/her weight and growth, skin, chest, abdomen, limbs and spine. The family history and examination will take approximately half an hour. We would like to look at the records in your child's hospital file to make sure that we collect all the relevant information

In order to study the genes, we will take a sample of blood and saliva/ buccal swabs from your child. Our blood and saliva/ buccal swab contains cells. All the cells in our bodies contain a chemical called DNA that we inherit from our parents. DNA carries the instructions that your body must follow in order to exist and function. It contains all your genes. Your cells also contain a chemical called RNA. RNA is formed from DNA and helps DNA to do its job.

We can take the DNA out of the cells that are in the blood or the saliva/ buccal swab sample that we take from your child. Once we have isolated the DNA we can study the genes. In this study, we may need to examine all of the genes to try and find the gene that may be the cause of your child's cancer.

What can you expect from the study?

Once you understand what the study is about and are happy to take part, we will ask you to sign a consent form on behalf of your child to join the study. If your child is old enough and able to understand the study, they may sign an assent form. A copy of these forms will be given to you to keep.

Once you have signed the consent form and your child has signed the assent form we will ask you some questions about your child's background and health. A complete clinical examination will be performed on your affected child. We may ask to access data relevant to your child's diagnosis from his/her clinical file or test results. This information will help us to understand your child's condition better and to interpret the data generated by our research.

A nursing sister or doctor will take up to three teaspoons (15ml) of blood from your child. Venipuncture (drawing blood) is normally done as part of routine medical care and presents a slight risk of discomfort. Drawing blood may result in faintness, inflammation of the vein, pain, bruising or bleeding at the puncture site. Your protection is that experienced staff perform the procedures under sterile conditions.

We will also ask for a saliva/ buccal swab sample from your affected child. A nursing sister or doctor will use a small, soft sponge to collect the saliva sample. The sponge is placed inside your child's mouth against the cheek and gently moved along the gums and cheek to soak up as much saliva as possible. For a buccal sample, a nursing sister or doctor will use a small soft brush to collect a buccal swab. These procedures are painless and harmless.

Long-term storing of samples and data

Once we have done the research that we are planning for this project, we would like to store your child's DNA, genomic information and data on their health history and laboratory test results for future research on childhood cancer, and we will ask you to sign a separate consent form to give us permission to do this. The sample will be stored together with other participants' samples. The samples will be used in future childhood cancer studies.

Confidentiality

Your child's samples will be stored in a locked fridge and the personal information will be stored on a secure computer in a locked office. Only the investigators of this study will have access to your child's name and personal details. Each participant will be given a code and number for the study so that no one knows who you are. The blood, saliva, RNA and DNA samples and the associated data will not be attached to your names – they will only have their code and number on them.

What are my rights as a participant?

You and your child do not have to join this study. Whether he/she decides to join this study or not will not affect your treatment in our clinic. You can also decide to leave the study at any time. You may choose to have your child's biological samples (blood, DNA, RNA and saliva) destroyed at any time. If you do not want your child to be part of this study or decide to leave the study, it will not affect your child's regular treatments or medical care in any way. We invite your child to approach the researchers to be re-consented when he/she turns 18 if he/she wishes. He/she can also decide to leave the study at any time.

You may not get any benefit directly from this study, but we hope it may help others who are affected by childhood cancers in the future. We may contact you in future if we believe we have identified the cause of your child's condition if you give your permission for this. Please understand that what we are trying to do is very difficult and could take a long time.

Contact details

If you have any problems you can contact us at any time at this telephone number: Dr Lindiwe Lamola, Dr Candice Feben or Prof Amanda Krause – 011 489 9223.

PLEASE NOTE!

Some patients may experience psychological distress should a complication be detected. You can contact the Genetic Counselling Unit at the Division of Human Genetics, NHLS Braamfontein to arrange a session with an HPCSA-registered genetic counsellor to discuss these issues. They are experienced and qualified counsellors who are able to offer distress counselling, decision support, address concerns associated with genetic disorders and testing and explain medical terminology. Appointments can be arranged by contacting Ms Janet Robbie at (011) 489 9223 or Ms Shelley Macaulay at (011) 489 9243.

If you have any complaints or problems with this research you may report them to the Chair of the Human Research Ethics Committee at the University of the Witwatersrand at this telephone number – (011) 717 1234.

Appendix B2.2: Child Participant Information Sheets

Information sheet – Child Participant

Inherited cancer-predisposing mutations in a cohort of childhood cancer patients

Protocol: M180855
Protocol: M190751

Division of Human Genetics

University of the Witwatersrand & National Health Laboratory Service

P.O. Box 1038

Braamfontein

Johannesburg

2000

Tel: +2711 489 9211

Fax : +2711 489 9226

Dear Potential Participant,

Hello, we are Dr Lindie Lamola, Dr Candice Feben, Prof. Amanda Krause, Dr Njabulo Mabaso and Ms Erin Manolas from the Division of Human Genetics at the University of the Witwatersrand and National Health Laboratory Service. We are doing research on childhood cancers. We work with other members of the Division of Human Genetics to study the genetic factors that cause childhood cancers in South Africans.



What is cancer?

Cancer is the abnormal growth of cells. Cells are small structures which our bodies are made up of. Sometimes our cells get damaged and they start to grow differently and become big and abnormal – when they grow like this, they become what we call cancer.

What causes childhood cancers?

Cancer in children is difficult and sometimes doctors are not sure why some children get cancer and others don't.

For this study, we would like to look inside your DNA and find what is making your cells grow so much to cause cancer. By doing so, we will be able to understand how and why children like you have cancer.

What do I have to do?

Once you understand what the study is about and are happy to take part, we will ask you to sign a form to join the study. A copy of these forms will be given to you to keep.





Once you have become a participant in the study the doctor will do some checks like how heavy you are and how tall you are



The doctor will need to take some of your blood and saliva/ buccal swabs which we will use to get your DNA.



Our work in this study is to test your DNA to find what caused your cancer.



We will ask your Dr for permission to access your hospital file. This information will help us to understand your condition better.

Long-term storing of samples and data



Your sample will be stored safely together with other participants' samples. The samples will only be used in the childhood cancer study and not in any other research.

Confidentiality

You will be given a code and number for the study so that no one knows your real name and who you are. All samples and the data will not be attached to your name.

You do not have to join this study. Whether you decide to join this study or not will not affect your treatment in our clinic.

.

What will I benefit from being part of this study?

You may not get any benefit directly from this study, but we hope it may help others who are affected by childhood cancers in the future. As it might help us to understand why you got cancer. We may contact you in future if we believe we have identified the cause of your condition.

Contact details



If you have any problems you can contact us at any time on this telephone number: Dr Lindie Lamola, Dr Candice Feben, Dr Njabulo Mabaso or Prof Amanda Krause – 011 489 9223.



Appendix B3.1: Parent/Legal Guardian Consent Forms

NATIONAL HEALTH LABORATORY SERVICE
School of Pathology, University of the Witwatersrand
DIVISION OF HUMAN GENETICS
Hospital Street, Johannesburg, 2001 | PO Box 1038, Johannesburg, 2000
[T]: +27 11 489 9211 | [F]: +27 11 489 9226 | [E]: human.genetics@nhls.ac.za

Protocol: M180855

Protocol: M190751

Inherited cancer-predisposing mutations in a cohort of childhood cancer patients

CONSENT FORM – PARENTAL CONSENT FOR CHILD PARTICIPANTS
PARTICIPANT STUDY ID:

The purpose of this study, procedures to be followed, risks and benefits, have been explained to me and my child. I have been allowed to ask questions, and my questions have been answered to my satisfaction. I have been told whom to contact if I have questions, to discuss problems, concerns, or suggestions related to the research, or to obtain information or offer input about the research. I have read the attached information sheet and agree to allow my child to participate in this research study with the understanding that I may withdraw my child at any time. To the best of my ability, I have discussed the study with my child, who agrees to be in the study. I have explained to my child that he/she may approach the researchers to be re-consented when he/she turns 18 if he/she wishes and that he/she may withdraw from the study at any time. I have been told that I will be given a signed and dated copy of this consent form if I would like one.

	Please initial
I understand that the purpose of this consent form is for me to inform the study what they can or cannot do with my child's samples, clinical information and photos.	
I understand that all procedures/tests on the samples have been or will be approved by the Human Research Ethics Committee of the University of the Witwatersrand. I understand that every time a new study is done on my child's sample, permission will be obtained from this Ethics Committee for the study to make sure that it is used only for the purposes stated above.	
I understand that if I have any questions at any time, they will be answered.	
I understand that confidentiality will be maintained at all times through the use of a code and numbering system.	
I understand that I may withdraw my child from the study at any time.	
I have discussed the study with my child.	

I am in agreement that my child's samples may be stored and used for the purposes described in the information sheet.	
I understand that I can choose to have my child's sample destroyed at any time.	
I understand that my child may not benefit directly from the research done on their sample.	
If this research does identify information that is relevant to my child's cancer, I would like to be contacted to receive the information.	Yes ☐ No ☐
If this research does identify genetic information that would be important or of health benefit to my family, or me, I would like to be contacted to receive the information.	Yes ☐ No ☐

Researcher	☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐ Printed Name & Surname _____ ☐☐/☐☐/☐☐☐☐ Signature/Thumbprint Date (dd/mm/yyyy)
Parent Participant	☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐ Printed Name & Surname _____ ☐☐/☐☐/☐☐☐☐ Signature/Thumbprint Date (dd/mm/yyyy)

NATIONAL HEALTH LABORATORY SERVICE
School of Pathology, University of the Witwatersrand
DIVISION OF HUMAN GENETICS
Hospital Street, Johannesburg, 2001 | PO Box 1038, Johannesburg, 2000
[T]: +27 11 489 9211 | [F]: +27 11 489 9226 | [E]: human.genetics@nhls.ac.za

Protocol: M180855
Protocol: M190751

Inherited cancer-predisposing mutations in a cohort of childhood cancer patients

CONSENT FORM – PARENTAL CONSENT FOR CHILD PARTICIPANTS PARTICIPANT STUDY ID:

PERMISSION FOR STORAGE	
I agree to have my samples stored in the research facility for future childhood cancer studies only without any identifiers.	Yes ☐ No ☐
I agree to have the data generated from my samples stored without any identifiers and be reanalysed if need be.	Yes ☐ No ☐

PERMISSION FOR REUSE		
	Samples	**Data**
I agree that my child's samples & data (including clinical data) can be shared with other investigators for research *related to childhood cancers* once the ethics committee has approved the new study. My child's samples & data will be made available without any identifiers.	Yes ☐ No ☐	Yes ☐ No ☐

Researcher	☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐ Printed Name & Surname _____ ☐☐/☐☐/☐☐☐☐ Signature/Thumbprint Date (dd/mm/yyyy)
Parent Participant	☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐ Printed Name & Surname _____ ☐☐/☐☐/☐☐☐☐ Signature/Thumbprint Date (dd/mm/yyyy)

This report is intended solely to record the observations and/or opinion of the writer. It does not constitute a medico-legal report.

Chairperson: Prof Eric Buch Acting CEO: Dr Karmani Chetty
Physical Address: 1 Modderfontein Road, Sandringham, Johannesburg, South Africa Postal Address: Private Bag X8, Sandringham, 2131, South Africa
Tel: +27 (0) 11 386 6000/ 0860 00 NHLS(6457) www.nhls.ac.za
Practice number: 5200296

NATIONAL HEALTH LABORATORY SERVICE
School of Pathology, University of the Witwatersrand
DIVISION OF HUMAN GENETICS
Hospital Street, Johannesburg, 2001 | PO Box 1038, Johannesburg, 2000
[T]: +27 11 489 9211 | [F]: +27 11 489 9226 | [E]: human.genetics@nhls.ac.za

Protocol: M180855
Protocol: M190751

Inherited cancer-predisposing mutations in a cohort of childhood cancer patients

CONSENT FORM – PARENTAL CONSENT FOR CHILD PARTICIPANTS

PARTICIPANT STUDY ID:

PERMISSION FOR PHOTOGRAPHS	
I agree to the use of my child's photographs for medical records ONLY.	Yes ☐ No ☐
I agree for my child's photographs to be used for teaching purposes AND to be used for medical records but NOT for medical publication.	Yes ☐ No ☐
I agree for my child's photographs to be used in medical publications, including medical journals, textbooks and electronic publications AND to be used for teaching purposes AND to be used for my child's medical records. I understand that the images may be seen by members of the general public, in addition to scientists and medical researchers that use these publications in their professional education. Although these photographs will be used without identifying information, such as my child's name, I understand that it is possible that someone may recognize my child.	Yes ☐ No ☐ ☐Understood

Researcher	☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐ Printed Name & Surname _____ ☐☐/☐☐/☐☐☐☐ Signature/Thumbprint Date (dd/mm/yyyy)
Parent Participant	☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐ Printed Name & Surname _____ ☐☐/☐☐/☐☐☐☐ Signature/Thumbprint Date (dd/mm/yyyy)

This report is intended solely to record the observations and/or opinion of the writer. It does not constitute a medico-legal report.

Chairperson: Prof Eric Buch Acting CEO: Dr Karmani Chetty
Physical Address: 1 Modderfontein Road, Sandringham, Johannesburg, South Africa Postal Address: Private Bag X8, Sandringham, 2131, South Africa
Tel: +27 (0) 11 386 6000/ 0860 00 NHLS(6457) www.nhls.ac.za
Practice number: 5200296

Appendix B3.2: Child Participant Assent Forms

NATIONAL HEALTH LABORATORY SERVICE
School of Pathology, University of the Witwatersrand
DIVISION OF HUMAN GENETICS
Hospital Street, Johannesburg, 2001 | PO Box 1038, Johannesburg, 2000
[T]: +27 11 489 9211 | [F]: +27 11 489 9226 | [E]: human.genetics@nhls.ac.za

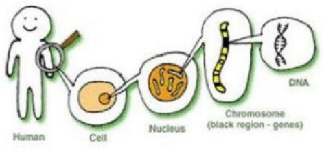
Protocol: M180855

Protocol: M190751

Inherited cancer-predisposing mutations in a cohort of childhood cancer patients

ASSENT FORM FOR CHILD PARTICIPANTS (7-13 Years)
PARTICIPANT STUDY ID:

Do you understand… **(please mark an 'X' in the boxes below)**

 that this study is about looking into the cells (DNA) to try better understand why children, like you, get cancer? ☐	  that this study is important to help doctors understand and treat children with cancer in the future? ☐	 that we need some blood and cheek cells from you, so we can look into your cells to find the DNA? ☐
 that the doctor will be doing a check-up and taking photos to record in your study file? ☐	 that this study may not help you directly, but if we do find something that will help you or your family, we will contact your parents? ☐	 that future scientists and doctors may want to look at your DNA & records (if you & your parents allow)? ☐
 that if someone picks up your sample or looks at your information, they won't see your name, only a number? ☐	 that you don't have to participate in this study if you don't want to and you can stop being in this study at anytime and it will not affect your treatment? ☐	 that if you have any questions in the future, you can ask your parents to contact the doctor? ☐

NATIONAL HEALTH LABORATORY SERVICE
School of Pathology, University of the Witwatersrand
DIVISION OF HUMAN GENETICS
Hospital Street, Johannesburg, 2001 | PO Box 1038, Johannesburg, 2000
[T]: +27 11 489 9211 | [F]: +27 11 489 9226 | [E]: human.genetics@nhls.ac.za

I am happy to participate in this study, if true please sign below:

Child Participant	Date (dd/mm/yyyy)	Researcher
Parent/legal guardian	Date (dd/mm/yyyy)	Date (dd/mm/yyyy)
Child participant's name in BLOCK CAPITALS		Researcher's name in BLOCK CAPITALS

I, _____ the parent/legal guardian of _____ acknowledge that he/she is unable to provide assent due to _____ and am providing consent to participate in this study on their behalf.	**Please initial**

This report is intended solely to record the observations and/or opinion of the writer. It does not constitute a medico-legal report.

Inherited cancer-predisposing mutations in a cohort of childhood cancer patients

ASSENT FORM FOR CHILD PARTICIPANTS (14 – 18 Years)
PARTICIPANT STUDY ID:

Do you understand… (please mark an 'X' in the boxes below)

That this study is about looking into the cells (DNA) of children to try better understand why children, like you, get cancer?	☐
That this study is important to help doctors understand and treat children with cancer in the future?	☐
That we need some blood and cheek cells from you, so we can look into your cells to find the DNA?	☐
That the doctor will be doing a check-up and taking photos to record in your study file?	☐
That this study may not help you directly, but if we do find something that will help you or your family, we will contact your parents?	☐
That future scientists and doctors may want to look at your DNA and records (with your + your parents' permission)?	☐
That if someone picks up your sample or looks at your information, they won't see your name – only a number?	☐
That you don't have to participate in this study if you don't want to and you can stop being in this study at any time and it will not affect your treatment?	☐
That if you have any questions in the future, you can ask your parents to contact the doctor?	☐

I am happy to participate in this study, if true please sign below:

Child Participant	Date (dd/mm/yyyy)	Researcher
Parent/legal (dd/mm/yyyy) guardian	Date (dd/mm/yyyy)	Date

Child participant's name in BLOCK CAPITALS	Researcher's name in BLOCK CAPITALS

I, _____ the parent/legal guardian of _____ acknowledge that he/she is unable to due to _____ and am providing consent to participate in this study on their behalf.	**Please initial**

This report is intended solely to record the observations and/or opinion of the writer. It does not constitute a medico-legal report.

Chairperson: Prof Eric Buch Acting CEO: Dr Karmani Chetty
Physical Address: 1 Modderfontein Road, Sandringham, Johannesburg, South Africa Postal Address: Private Bag X8, Sandringham, 2131, South Africa
Tel: +27 (0) 11 386 6000/ 0860 00 NHLS(6457) www.nhls.ac.za
Practice number: 5200296

Appendix C: Methods Documentation

Appendix C1: Gene-Panel Table

Gene Symbol (HGNC)	Gene Name	Inheritance	Familial Syndrome	Associated Childhood Cancers
ALK (427)	Anaplastic Lymphoma receptor tyrosine Kinase	Autosomal Dominant	Familial Neuroblastoma	Neuroblastoma, Anaplastic Large Cell Lymphoma, Inflammatory Myofibroblastic tumour, Lung Cancer
APC (583)	Adenomatosis Polyposis Coli, WNT signalling pathway regulator	Autosomal Dominant	Familial Adenomatous Polyposis, Turcot syndrome	Medulloblastoma, Hepatoblastoma, Osteosarcoma
ATM (795)	Ataxia Telangiectasia Mutated serine/threonine kinase	Autosomal Recessive	Ataxia-telangiectasia	Melanoma, Stomach cancer, Pancreatic cancer, Leukaemia, Lymphoma, Bladder cancer
BAP1 (950)	BRAC1 Associated Protein 1	Autosomal Dominant	BAP1 - Tumour predisposition syndrome	Cutaneous melanoma, Basal cell carcinoma, Uveal melanoma, Clear Cell Renal Carcinoma, Malignant Mesothelioma
BARD1 (952)	BRCA1 Associated RING Domain 1	Autosomal Dominant	n/a	Neuroblastoma, Nephroblastoma, Ovarian cancer
BLM (1058) ***Aka - RECQL2/3***	Bloom Syndrome RecQ-like helicase	Autosomal Recessive	Bloom syndrome	Non-Hodgkin's Lymphoma, Wilms' tumour, Osteosarcoma, Gastrointestinal Stromal Tumours
BRAF (1097)	B-Raf proto-oncogene, serine/threonine kinase (RASopathy gene)	Autosomal Dominant	Cardiofaciocutaneous syndrome, Erdheim-Chester Disease, Giant Congenital Melanocytic Nevus, Noonan syndrome, Familial Gastrointestinal Stromal Tumours, Costello syndrome	Melanoma, Leukaemia, Cholangiocarcinoma, Gastrointestinal Stromal Tumours, Multiple Myeloma, Non-Hodgkin Lymphoma, Colorectal cancer, Thyroid carcinoma, Lung Adenocarcinoma
BRCA1 (1100) *Aka FANC**S***	Breast Cancer 1 DNA repair associated	Autosomal Dominant	Fanconi Anaemia, Hereditary Breast-Ovarian Cancer	Pancreatic cancer, Colon cancer, Leukaemia, Central Nervous System tumours, Acute Myeloid Leukaemia
*BRCA2 (1101) Aka FANCD**1***	Breast Cancer 2 DNA repair associated	Autosomal Dominant	Fanconi Anaemia, Hereditary Breast-Ovarian Cancer	Acute Myeloid Leukaemia, Central Nervous System tumours, Wilms' tumour, Osteosarcoma
BUB1 (1148)	Budding Uninhibited by Benzimidazoles 1 homolog, mitotic checkpoint serine/threonine kinase	Autosomal Dominant / Autosomal Recessive	Lynch Syndrome, Primary Microcephaly-30	Colorectal cancer, Oesophageal cancer, Gastric cancer, Melanoma, Lung cancer, Hepatoblastoma, Glioblastoma

CDKN1C (1786)	Cyclin Dependent Kinase Inhibitor 1C	Autosomal Dominant	Beckwith-Wiedemann syndrome	Wilms' tumour, Neuroblastoma, Adrenal carcinoma, Hepatoblastoma, Rhabdomyosarcoma
CHEK2 (16627)	Checkpoint Kinase 2	Autosomal Dominant	Li-Fraumeni syndrome	Soft-tissue carcinoma, Colorectal cancer, Thyroid cancer, Kidney cancer
DIS3L2 (28648)	DIS3 Like 3'-5' exoribonuclease 2	Autosomal Recessive	Perlman syndrome	Wilms tumour (nephroblastomatosis – precancerous lesion of Wilms tumour)
EPCAM (11529) ***AKA HNPCC8***	Epithelial Cell Adhesion Molecule	Autosomal Dominant	Lynch Syndrome	Hereditary Non-Polyposis Colorectal Cancer, Glioblastoma, Pancreatic cancer
ERCC2 (3434) AKA XPD/XPCD	Excision Repair Cross-Complementation group 2, TFIIH core complex helicase subunit (Nucleotide Excision Repair gene).	Autosomal Recessive	Xeroderma Pigmentosum	Skin carcinoma, Melanoma
ERCC3 (3435) AKA XPB/XPCB	Excision Repair Cross-Complementation group 3, TFIIH core complex helicase subunit (Nucleotide Excision Repair gene)	Autosomal Recessive	Xeroderma Pigmentosum (XP), XP-Cockayne Syndrome (XP-CS)	Basal cell carcinomas, Melanoma
FAH (3579)	Fumarylacetoacetate Hydrolase	Autosomal Recessive	Tyrosinaemia	Hepatocellular carcinoma
FH (3700)	Fumarate Hydratase	Autosomal Dominant	Hereditary Leiomyomatosis and renal cell cancer	Renal cell carcinoma
GPC3 (4451)	Glypican Proteoglycan 3	X-linked	Simpson-Golabi-Behmel syndrome	Wilms' tumour
HRAS (5173)	HRas proto-oncogene, GTPase (RASopathy gene)	Autosomal Dominant	Costello syndrome	Rhabdomyosarcoma, Neuroblastoma, Bladder cancer, Head and Neck Squamous cell carcinoma, Kidney cancer
KRAS (6407)	KRas proto-oncogene, GTPase (RASopathy gene)	Autosomal Dominant	Cardiofaciocutaneous syndrome, Noonan syndrome, Autoimmune Lymphoproliferative syndrome, KRAS Mutation-Associated Phenotype, Hereditary Diffuse Gastric Cancer	Leukaemia, Cholangiocarcinoma, Lung Adenocarcinoma, Pancreatic cancer, Colorectal carcinoma, Acute Myeloid Leukaemia
LZTR1 (6742)	Leucine Zipper like Transcription Regulator 1 (RASopathy gene)	Autosomal Dominant /Autosomal Recessive	Noonan syndrome, Schwannomatosis	Neuroblastoma, Embryonal Rhabdomyosarcoma, Leukaemia

MITF ***(7105)***	Melanocyte Inducing Transcription Factor	Autosomal Dominant /Autosomal Recessive	Tietz syndrome, Waardenburg syndrome	Melanoma
MLH1 ***(7127) AKA HNPCC2***	MutL Homolog 1 (Mismatch Repair gene)	Autosomal Dominant/ Autosomal Recessive (CMMRD)	Constitutional Mismatch Repair Deficiency syndrome, Lynch syndrome, Turcot syndrome	Hereditary Non-Polyposis Colorectal Cancer, Glioblastoma, Medulloblastoma, Leukaemia, Lymphoma
MSH2 ***(7325)*** *AKA HNPCC**1***	MutS Homolog 2 (Mismatch Repair gene)	Autosomal Dominant/ Autosomal Recessive (CMMRD)	Constitutional Mismatch Repair Deficiency syndrome, Lynch syndrome, Muir-Torre syndrome, Turcot syndrome	Hereditary Non-Polyposis Colorectal Cancer, Glioblastoma, Medulloblastoma, Nephroblastoma, Leukaemia, Lymphoma
MSH6 ***(7329)*** *AKA HNPCC**5***	MutS Homolog 6 (Mismatch Repair gene)	Autosomal Dominant / Autosomal Recessive (CMMRD)	Constitutional Mismatch Repair Deficiency syndrome, Lynch Syndrome, Turcot syndrome	Hereditary Non-Polyposis Colorectal Cancer, Leukaemia, Lymphoma, Neurofibromas, Stomach cancer, Pancreatic cancer, Glioblastoma, Medulloblastoma
MUTYH ***(7527)***	MutY Homolog DNA glycosylase	Autosomal Recessive	Familial Adenomatous Polyposis	Colorectal cancer
*NF1 (**7765**)*	Neurofibromin 1 (RASopathy gene)	Autosomal Dominant	Neurofibromatosis type 1	Astrocytoma, Juvenile Myelomonocytic Leukaemia, Acute Lymphocytic Leukaemia, Rhabdomyosarcoma, Malignant Peripheral Nerve Sheath Tumour (MPNST), Central Nervous System Gliomas
*NF2 (**7773**)*	Neurofibromin 2, moesin-ezrin-radixin-like (MERLIN) tumour suppressor	Autosomal Dominant	Neurofibromatosis Type 2	Meningioma, Gliomas, Ependymomas
NSD1 ***(14234)***	Nuclear receptor binding SET Domain protein 1	Autosomal Dominant	Sotos syndrome, Weaver syndrome	Acute Myeloid Leukaemia, Neuroblastoma, Gliomas
PALB2 ***(26144)*** *AKA FANC**N***	Partner And Localizer of BRCA2	Autosomal Dominant	Fanconi Anaemia, Hereditary Breast-Ovarian Cancer	Acute Myeloid Leukaemia, Pancreatic cancer, Wilms' Tumour
PMS2 ***(9122)*** *AKA HNPCC**4**/* ***MLH4***	PMS1 homolog 2, mismatch repair system component (Mismatch Repair gene)	Autosomal Dominant / Autosomal Recessive (CMMRD)	Constitutional Mismatch Repair Deficiency syndrome, Lynch syndrome, Turcot syndrome	Hereditary Non-Polyposis Colorectal Cancer, Stomach cancer, Gliomas, Leukaemia, Lymphoma
PTCH1 ***(9585)***	Patched 1	Autosomal Dominant	Gorlin Syndrome	Medulloblastoma, Basal Cell Carcinoma, Osteosarcoma

*PTEN (**9588**)*	Phosphatase and Tensin homolog	Autosomal Dominant	Bannayan-Riley-Ruvalcaba syndrome aka PTEN hamartoma tumour syndrome, Cowden syndrome	Thyroid cancer, Head and Neck Squamous Cell Carcinoma, Lung cancer, Astrocytomas, Gliomas, Melanoma
*RB1 (**9884**)*	RB transcriptional corepressor 1	Autosomal Dominant	Familial Retinoblastoma	Retinoblastoma, Osteosarcoma
*RECQL4 (**9949**)*	RecQ Like helicase 4	Autosomal Recessive	Rothmund-Thompson syndrome	Osteosarcoma
*RET (**9967**) AKA MEN2A/B*	Rearranged during Transfection proto-oncogene	Autosomal Dominant	Multiple Endocrine Neoplasia Type 2	Medullary Thyroid Carcinoma, Malignant Pheochromocytomas
*RNASE1 (**10044**)*	Ribonuclease A family member 1, pancreatic	Autosomal Recessive	Noonan Syndrome	Myelomonocytic Leukaemia, Acute Lymphocytic Leukaemia, Neuroblastoma, Rhabdomyosarcoma
SDHA (10680)	Succinate Dehydrogenase complex flavoprotein subunit A	Autosomal Dominant	Non-Syndromic Paraganglioma, Leigh syndrome, Friedreighs' Ataxia	Gastrointestinal Stromal Tumours, Stomach cancer, Non-Small Cell Lung carcinoma, Endometrial cancer, Colorectal Adenocarcinoma, Melanoma, Sarcoma, Renal Cell Carcinoma
*SDHB (**10681**)*	Succinate Dehydrogenase complex iron sulphur subunit B	Autosomal Dominant	Cowden or Cowden-like syndrome, Carney-Stratakis syndrome	Thyroid cancer, Gastrointestinal Stromal Tumours, Renal Cell Carcinoma, Malignant Paragangliomas
SMAD4 (6770)	SMAD family member 4	Autosomal Dominant	Hereditary Haemorrhagic Telangiectasia, Juvenile Polyposis syndrome, Myhre syndrome	Malignant Gastrointestinal Polyps, Cholangiocarcinoma, Colon cancer, Pancreatic cancer, Endometrial cancer
SMARCA4 (11100)	SWI/SNF related, Matrix associated, Actin dependent Regulator of Chromatin, subfamily A, member 4	Autosomal Dominant	Coffin-Siris syndrome, Rhabdoid Tumour Predisposition Syndrome	Rhabdoid tumours – including Central Nervous System tumours, Kidney tumours, Lung cancer
*SMARCB1 (**11103**)*	SWI/SNF related, Matrix associated, Actin dependent Regulator of Chromatin, subfamily B, member 1	Autosomal Dominant	Li-Fraumeni syndrome, Schwannomatosis, Rhabdoid Tumour Predisposition Syndrome	Osteosarcoma, Rhabdoid tumours – including Central Nervous System tumours, Meningiomas
*TMEM127 (**26038**)*	Transmembrane protein 127	Autosomal Dominant	Hereditary Pheochromocytoma	Malignant Paragangliomas, Malignant Pheochromocytomas
*TP53 (**11998**)*	Tumour Protein p53	Autosomal Dominant	Li-Fraumeni Syndrome	Adrenocortical carcinoma, Brain cancers, Osteosarcoma, Sarcoma, Leukaemia, Wilms' Tumour
*TSC1 (**12362**)*	Tuberous Sclerosis Complex subunit 1	Autosomal Dominant	Tuberous Sclerosis	Subependymal Giant Cell Astrocytoma, Kidney cancer, Cholangiocarcinoma

TSC2 (***12363***)	Tuberous Sclerosis Complex subunit 2	Autosomal Dominant	Tuberous Sclerosis	Subependymal giant cell Astrocytoma, Kidney cancer
VHL (***12687***)	Von Hippel-Lindau tumour suppressor	Autosomal Dominant	Von-Hippel Lindau syndrome, Non-syndromic Paragangliomas	Renal Cell Carcinoma, Hemangioblastomas, Pancreatic tumours, Malignant Paragangliomas, Malignant Pheochromocytomas
WT1 (***12796***)	Wilms' Tumour 1 transcription factor	Autosomal Dominant	Denys-Drash syndrome, Frasier syndrome, Familial Wilms' Tumour, Cytogenetically Normal Acute Myeloid Leukaemia, WAGR (Wilms' tumour, Aniridia, Genitourinary anomalies, and intellectual disability (formerly referred to as mental Retardation)) syndrome	Wilms' tumour, Cytogenetically Normal-Acute Myeloid Leukaemia
XPC (***12816***)	Xeroderma Pigmentosum, complementation group C, DNA damage recognition and repair factor (Nucleotide Excision Repair gene)	Autosomal Recessive	Xeroderma Pigmentosum	Squamous Cell carcinoma, Basal Cell carcinoma, Melanomas

Appendix C2: Ion Reporter™ & VEP Annotation Sets

	Description	Ion Reporter	Variant Effect Predictor (VEP)	Source
Population Databases				
1000 Genomes	A freely accessible public database providing whole-genome sequencing data from a total of 2504 individuals from 26 different populations.	X	X	(The 1000 Genomes Project Consortium, 2015)
5000 Exomes	5000 Exome Project as a part of the NHLBI - ESP provides freely accessible exome data from diverse populations	X		(Fu *et al.*, 2013)
Genome Aggregation Database (gnomAD)	Publicly available resource of NGS exome and genome data. The v2 data set (reference genome - GRCh37) contains data from unrelated individuals as a result of various population and disease-specific studies.		X	(Karczewski *et al.*, 2020)
Exome Aggregation Consortium (ExAC)	Publicly and freely available aggregation of exome data from large population datasets (60 706 individuals). ExAC browser is no longer available, and its data can be located in the gnomAD browser.	X		(Karczewski *et al.*, 2017)
Public Databases				
ClinVar	A freely accessible public database containing reports detailing a variant's relationship to the patient's phenotype. These reports provide interpretations of the variant's clinical significance to disease.	X	X	(Landrum *et al.*, 2014)
Catalogue of Somatic Mutations in Cancer (COSMIC)	Largest publicly available database of somatic mutations in human cancers. Contains both high-precision data manually curated by experts and peer-reviewed genome-wide screening data.	X		(Bamford *et al.*, 2004)
Single Nucleotide Polymorphism Database (dbSNP)	dbSNP, established by the NCBI, is the largest nucleotide variation database. It contains > 12 million non-redundant, species-specific genetic variation data	X	X	(Sherry *et al.*, 2001)
Online Mendelian Inheritance in Man (OMIM)	Database of human genes and genetic disorders. OMIM focuses on the relationship between phenotypic expression and genetic variation and is based on peer-reviewed biomedical literature	X		(Hamosh *et al.*, 2005)

Database of Genomic Variants (DGV)	Publicly accessible database of structural variation identified in control individuals within populations around the world. The control data is used for the correlation of genetic variation with phenotypic data	X		(MacDonald *et al.*, 2014)
DrugBank	Knowledgebase containing clinical and molecular information on drug and drug target (protein) interactions	X		(Wishart *et al.*, 2006)
Pfam	Database of annotated protein families which uses Hidden-Markov Models (HMMs) to generate multiple sequence alignments of the protein sequences	X		(Finn *et al.*, 2008; Sonnhammer, Eddy & Durbin, 1997)
Gene ontology (GO)	GO aims to unify biological vocabulary regarding genes and proteins. GO annotates genes and/or gene products with cellular location, biological processes, and molecular function.	X	X	(The Gene Ontology Consortium, 2021; Ashburner *et al.*, 2000)
Pathogenicity & Conservation Prediction Tools				
Combined Annotation Dependent Depletion (CADD)	A scoring tool which integrates multiple variant annotations into a single C-score metric, calculated by contrasting simulated mutations to those that survived natural selection. C-scores aid in the interpretation of the deleteriousness of SNVs and InDels		X	(Kircher *et al.*, 2014)
CONsensus DELeteriousness (CONDEL)	Scores deleteriousness of missense SNVs by taking the weighted average of the normalised scores from multiple bioinformatic tools and integrates them into one unified classification		X	(González-Pérez & López-Bigas, 2011)
Deleterious Annotation of genetic variants using Neural Networks (DANN)	DANN is a deep neural network trained with the same training data and feature set as CADD. Like CADD, DANN can assess both coding and non-coding variants, with the addition of DANN's capability to assess non-linear relationships between variables		X	(Quang, Chen & Xie, 2015)
Likelihood Ratio Test (LRT)	The LRT compares two data probability models to one another. One is a neutral model, whereby each codon is assumed to evolve naturally and there is no difference between the synonymous and non-synonymous substitution rates. This is compared to a constrained model where it is assumed that each codon has evolved under negative selection		X	(Chun & Fay, 2009)
MetaLR	MetaLR is an Ensembl score which assesses the pathogenicity of missense variants by integrating nine independent deleteriousness scores through a logistic regression model. The dbNSFP project is used to calculate MetaLR scores		X	(Dong *et al.*, 2015)

MutationAssessor	MutationAssessor predicts the functional impact of an amino-acid substitution in proteins. It generates conservation scores by assessing the impact of amino-acid substitutions within a protein family multiple sequence alignment.		X	(Reva, Antipin & Sander, 2011)
MutationTaster	MutationTaster is capable of assessing the disease-causing potential of all variant types as it investigates variants according to its resultant gene alteration. MutationTaster also uses integrated databases to discard known polymorphisms		X	(Schwarz *et al.*, 2010)
Polymorphism Phenotyping (PolyPhen)	PolyPhen is a tool used to predict the structural and functional impact on human proteins due to amino-acid substitutions. It specifically assesses the impact of non-synonymous SNVs by investigating multiple physical and comparative variables	X	X	(Adzhubei, Jordan & Sunyaev, 2013)
Protein Variant Effect Analyzer (PROVEAN)	PROVEAN assess the functional effect of single/multiple amino acid substitutions/indels using an alignment score approach. PROVEAN's web server is capable of processing multiple queries at once and stores a large database of pre-computed PROVEAN scores. PROVEAN generates scores by comparing pairwise alignments of protein sequences with a number of single-locus variants		X	(Choi & Chan, 2015)
Sorting Intolerant From Tolerant (SIFT)	SIFT predicts the effect of non-synonymous SNVs on protein function. It uses measures of the conservation and physical properties of amino acids to determine this effect	X	X	(Ng & Henikoff, 2003)
Functional Analysis through Hidden Markov Models (FATHMM)	FATHMM's web server uses Hidden Markov Models combined with sequence conservation to predict the functional impacts of both coding and non-coding variants	X	X	(Shihab *et al.*, 2013)
Grantham score	Ion Reporter software calculates the Grantham score based on the work of R. Grantham., 1974. Grantham score is a measure of the evolutionary distance between two amino acids, with a greater evolutionary distance indicating that the two amino acids in question are less likely to be substituted with one another, hence, the occurrence of their substitution is predicted to be more damaging.	X		(Grantham, 1974)
phyloP	A phyloP score represents a measure of the evolutionary conservation at an individual alignment site. phyloP integrates score tests, distributions of the number of nucleotide substitutions, the LRT and the Genomic Evolutionary Rate Profiling (GERP) test to measure if the substitution has drifted from the neutral rate of substitution, in both a positive (conservation) and negative (acceleration) direction, in a clade-specific manner	X	X	(Pollard *et al.*, 2010)

Appendix C3: *In-Silico* Tools and Interpretations – utilised in variant assessment and filtration.

Prediction Tool	Tool's scores and interpretations				Used for SNVs	Used for Indels & complex variants
CADD PHRED-scaled score	Likely Benign = score < 10	score ≥ 10 = Top 10% of deleterious variants in the human genome	score ≥ 20 = Top 1% of deleterious variants in the human genome	score ≥ 30 = Top 0.1% of the deleterious variants in the human genome	X	X
	The larger the score the more likely the variant has a damaging effect					
FATHMM score	T: score > -1.5	D: score ≤ -.15			X	
	The smaller the score the more likely the variant is deleterious. Scores range from -18.9 to 11.0 with a threshold between Tolerated (T) and Damaging (D) of ≤ -1.5					
GERP	Less conserved region and likely less deleterious effect = score < 4	More conserved region and more likely deleterious effect = score ≥ 4			X	X
	The greater the score the greater the evolutionary constraint predicted and the more deleterious the variant effect predicted.					
LRT	U: score 1 - 0.050	N: score 0.050 - 0.0005	D: 0.0005 - 0		X	
	A lower score indicates that the codon is more likely to be constrained, therefore a variant affecting this codon is more likely to be deleterious. D = Deleterious, N = Neutral, U = Unknown. Scores interpreted, as indicated above, according to the relationship between LRT prediction and LRT score output by VEP					
MetaLR	Tolerated (T) = score ≤ 0.8	Damaging (D) = score > 0.8			X	
	Scores ranging from 0-1, with higher scores indicating a variant with a more damaging effect. The prediction threshold between Damaging (D) and Tolerated (T) output by VEP were scores ≤ 0.8 = Tolerated					
MutationAssessor	N: score 0.0 - 0.2	L: score 0.2 - 0.5	M: 0.5 - 0.9	H: 0.9 - 1.0	X	
	High Functional Impact - H; Medium Functional Impact - M; Low Functional Impact (L); Neutral/No Functional Impact - N. Scores interpreted, as indicated above, according to MutationAssessor Rankscore and MutationAssessor prediction relationship output by VEP					

MutationTaster	P: score ≤ 0.2	N: score 0.2 - 0.3	D: score 0.3 - 0.8	A: score ≥ 0.8	X	
	Scores range from 0 - 1, with higher scores indicating a higher probability of a variant having a damaging effect. 1 = highest probability. A = Automatic Disease-Causing, D = Disease-Causing, N = Polymorphism, likely harmless, P = Polymorphism, known to be harmless. Scores interpreted, as indicated above, according to MutationTaster converted Rankscore and MutationTaster prediction relationship output by VEP					
phyloP	Less conserved region and likely less deleterious effect = score < 0.00		More conserved region and more likely deleterious effect = score > 0.00		X	X
	Higher, positive scores indicate conservation and therefore a more deleterious effect of a variant at this site. Lower, negative scores indicate fast-evolving sites and therefore a less deleterious effect of a variant at this site					
PolyPhen	B: score 0.00 - 0.15		P: 0.15 - 0.85	D: 0.85 - 1.00	X	
	Scores ranging from 0-1, with higher scores indicating a variant with a more damaging effect. 1 = highest probability. D = Probably Damaging, P = Possibly Damaging, B = Benign					
PROVEAN	N: score > -2.5		D: score ≤ -2.5		X	
	The lower the score the more damaging the effect predicted. Scores equal to or less than a threshold of -2.5 are predicted to have a deleterious effect, and scores above this threshold are predicted to be neutral. D = Deleterious, N = Neutral					
SIFT	T: score > 0.05		D: score ≤ 0.05		X	
	Scores range from 0-1, with lower scores indicating a more deleterious variant. The threshold between Deleterious (D) and Tolerated (T) is 0.05					
SIFT-INDEL	Neutral Prediction + confidence score close to 1 = confident that this insertion/deletion is not damaging		Damaging Prediction + confident score close to 1 = confident that this insertion/deletion is likely to have a damaging effect on gene function			X
	Uses a decision tree algorithm to predict if an insertion/deletion is 'gene damaging' = Damaging or has a 'neutral' effect = Neutral. A damaging or neutral prediction is made along with a confidence score in that prediction. The confidence scores range from 0 - 1 and the closer the score is to 1 the more confident is the prediction.					

Appendix C4: ACMG-AMP Guidelines for Variant Interpretation & Classification (Tables taken from Richards et al (2015))

Criteria for classifying pathogenic variants:

Evidence of pathogenicity	Category
Very strong	PVS1 null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease Caveats: • Beware of genes where LOF is not a known disease mechanism (e.g., *GFAP*, *MYH7*) • Use caution interpreting LOF variants at the extreme 3′ end of a gene • Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact • Use caution in the presence of multiple transcripts
Strong	PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change Example: Val→Leu caused by either G>C or G>T in the same codon Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level PS2 De novo (both maternity and paternity confirmed) in a patient with the disease and no family history Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity. PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established. PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls Note 1: Relative risk or OR, as obtained from case–control studies, is >5.0, and the confidence interval around the estimate of relative risk or OR does not include 1.0. See the article for detailed guidance. Note 2: In instances of very rare variants where case–control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.
Moderate	PM1 Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation PM2 Absent from controls (or at extremely low frequency if recessive) (**Table 6**) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium Caveat: Population data for insertions/deletions may be poorly called by next-generation sequencing. PM3 For recessive disorders, detected in *trans* with a pathogenic variant Note: This requires testing of parents (or offspring) to determine phase. PM4 Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants PM5 Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before Example: Arg156His is pathogenic; now you observe Arg156Cys Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level. PM6 Assumed de novo, but without confirmation of paternity and maternity
Supporting	PP1 Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease Note: May be used as stronger evidence with increasing segregation data PP2 Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease PP3 Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.) Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant. PP4 Patient's phenotype or family history is highly specific for a disease with a single genetic etiology PP5 Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation

LOF, loss of function; OR, odds ratio.

Criteria for classifying benign variants:

Evidence of benign impact	Category
Stand-alone	BA1 Allele frequency is >5% in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium
Strong	BS1 Allele frequency is greater than expected for disorder (see **Table 6**)
	BS2 Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder, with full penetrance expected at an early age
	BS3 Well-established in vitro or in vivo functional studies show no damaging effect on protein function or splicing
	BS4 Lack of segregation in affected members of a family
	Caveat: The presence of phenocopies for common phenotypes (i.e., cancer, epilepsy) can mimic lack of segregation among affected individuals. Also, families may have more than one pathogenic variant contributing to an autosomal dominant disorder, further confounding an apparent lack of segregation.
Supporting	BP1 Missense variant in a gene for which primarily truncating variants are known to cause disease
	BP2 Observed in *trans* with a pathogenic variant for a fully penetrant dominant gene/disorder or observed in *cis* with a pathogenic variant in any inheritance pattern
	BP3 In-frame deletions/insertions in a repetitive region without a known function
	BP4 Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc.)
	Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm cannot be counted as an independent criterion. BP4 can be used only once in any evaluation of a variant.
	BP5 Variant found in a case with an alternate molecular basis for disease
	BP6 Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation
	BP7 A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved

Rules for combining criteria to classify sequence variants:

Pathogenic	(i) 1 Very strong (PVS1) *AND* (a) ≥1 Strong (PS1–PS4) *OR* (b) ≥2 Moderate (PM1–PM6) *OR* (c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) *OR* (d) ≥2 Supporting (PP1–PP5) (ii) ≥2 Strong (PS1–PS4) *OR* (iii) 1 Strong (PS1–PS4) *AND* (a)≥3 Moderate (PM1–PM6) *OR* (b)2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR* (c)1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Likely pathogenic	(i) 1 Very strong (PVS1) *AND* 1 moderate (PM1–PM6) *OR* (ii) 1 Strong (PS1–PS4) *AND* 1–2 moderate (PM1–PM6) *OR* (iii) 1 Strong (PS1–PS4) *AND* ≥2 supporting (PP1–PP5) *OR* (iv) ≥3 Moderate (PM1–PM6) *OR* (v) 2 Moderate (PM1–PM6) *AND* ≥2 supporting (PP1–PP5) *OR* (vi) 1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Benign	(i) 1 Stand-alone (BA1) *OR* (ii) ≥2 Strong (BS1–BS4)
Likely benign	(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) *OR* (ii) ≥2 Supporting (BP1–BP7)
Uncertain significance	(i) Other criteria shown above are not met OR (ii) the criteria for benign and pathogenic are contradictory

Appendix C5: Validation Tables

Appendix C5.1: Websites/*In-silico* Tools used in Primer Design

Name of Website/*in-silico* Tool	Purpose of Website/*in-silico* Tool	Website/*in-silico* Tool URL
NCBI (Gene Database)	Extract reference sequences for the region of interest – GRCh37 build	https://www.ncbi.nlm.nih.gov/gene/?term=
Primer3 version 0.4.0	Design primers using reference sequences within a set of parameters	https://bioinfo.ut.ee/primer3-0.4.0/
OligoAnalyzer™	Assess for the formation of secondary structures (heterodimers, homodimers, hairpins) which may interfere with primer binding	https://eu.idtdna.com/calc/analyzer
NCBI Primer-BLAST	Check primer specificity to ensure they are unique to the region of interest	https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?ORGANISM=9606&INPUT_SEQUENCE=NM_001618.3
SNPCheck version 3	Check for the presence of SNPs reported in primer design regions	https://genetools.org/SNPCheck/snpcheck.htm

Appendix C5.2: Primer Design Utilised in Validation of Seven Pathogenic/Likely Pathogenic Germline Variants:

PeCan Project Patient ID	Variant	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	Product Size (bp)
PC.0023.01.1	*NF1*_p.D378Afs*8	GAACCACAGTATGGGTGCTT	GGACTTACATTGGTGATGATTCG	604
PC.0032.01.1	*RET*_p.C634Y	GGGCAATAGTGGTCTAGGAG	CTTGTGGTAGCAGTGGATGC	312
PC.0051.01.1	*TP53*_p.R273C	CCACCTCTTACCGATTTCT	TGTTGTTGGGCAGTGCTA	516
PC.0011.01.1	*BRCA1*_p.M1400T	CCCTCACTAACATCATTTGG	GCTTTGTTCTGGATTTCGC	364
PC.0009.01.1	*FAH*_p.R162H	GTGTATGTGGCGACTCTTTG	AAAACGGTGCTGTGAGTG	290
PC.0003.01.1	*ERCC3*_p.Q320*	TTGACCATTGGACCTCTTAG	GCAGATCCAGACACAACAG	340
PC.0018.01.1	*RBI*_p.R661W	GGGGGAAAGAAAAGAGTGGTA	GGAGAGAAGGTGAAGTGC	324

Appendix C5.3: Standard Temperature Gradient PCR Thermal Cycling Conditions:

Step	Temperature (°C)	Time	Cycles
Initial Denaturation	95	10 minutes	1
Denature	95	15 seconds	35
Anneal	Range (45 – 65) *	30 seconds	
Extend	72	30 seconds	
Final Extension	72	5 minutes	1
Hold	4	Indefinitely	1

*Exact range followed is dependent on the primer pair.

Appendix C5.4: Optimal Primer-Specific PCR Temperatures for the Seven Prioritised Variants:

PeCan Project Patient ID	Variant	Optimal Temperature (°C) for PCR Steps					
		Initial Denaturation	**Denaturation**	**Annealing**	**Extension**	**Final Extension**	**Hold**
PC.0023.01.1	*NF1*_p.D378Afs*8	95	95	56.1	72	72	4
PC.0032.01.1	*RET*_p.C634Y	95	95	59.2	72	72	4
PC.0051.01.1	*TP53*_p.R273C	95	95	45.7	72	72	4
PC.0011.01.1	*BRCA1*_p.M1400T	95	95	55.2	72	72	4
PC.0009.01.1	*FAH*_p.R162H	95	95	56.0	72	72	4
PC.0003.01.1	ERCC3_p.Q320*	95	95	56.5	72	72	4
PC.0018.01.1	RBI_p.R661W	95	95	56.5	72	72	4

Appendix C5.5: Cycle Sequencing Conditions used in Sanger Sequencing Validation:

Temperature (°C)	Time	Cycles
96	1 minute	1
96	30 seconds	25
50	15 seconds	
60	4 minutes	
4-12	Indefinite Hold	1

Appendix D: Results Documentation

Appendix D1: Patient Demographics Table

PeCan Project Patient ID	Age at Diagnosis (Years, Months)	Age at Recruitment (Years, Months)	Sex	Ethnicity	Cancer Type	Relatives Affected	Pedigree Indicative of Positive Family History of Cancer	Suspected/Confirmed CPS Diagnosis
PC.0003.01.1	7	7, 2	Female	African Black	Nephroblastoma (Right-sided)	No	No	None
PC.0006.01.1	15	15, 9	Female	Mixed Ancestry	Ovarian Germ Cell Tumour (Dysgerminoma)	No	No	None
PC.0007.01.1	8	8, 8	Male	African Black	Pilocytic Astrocytoma	No	No	None
PC.0008.01.1	9	10, 3	Female	African Black	Osteosarcoma (Left Thigh)	No	No	None
PC.0009.01.1	5	11, 10	Female	African Black	Juvenile Dermatofibrosarcoma Protuberans	No	No	None

PC.0011.01.1	3	6, 2	Female	Indian	Embryonal Rhabdomyosarcoma (Bladder)	No	No	None
PC.0012.01.1	3	15, 1	Female	African Black	Nephroblastoma (Left-sided)	No	No	None
PC.0014.01.1	4	7, 2	Female	African Black	Neuroblastoma (MYCN-amplified)	No	No	None
PC.0016.01.1	6	10	Female	African Black	Ovarian Germ Cell Tumour (Choriocarcinoma, non-gestational)	No	No	None
PC.0018.01.1	0, 6	11, 1	Male	African Black	Nephroblastoma (Right-sided)	No	No	None
PC.0019.01.1	3, 6	9, 1	Male	African Black	Nephroblastoma (Right-sided)	No	No	None
PC.0021.01.1	2	13, 9	Female	African Black	Nephroblastoma (Right-sided)	No	No	None
PC.0023.01.1	9	9, 4	Male	White (Afrikaner & European English)	Pilocytic Astrocytoma	Yes	Yes	NF1
PC.0024.01.1	12	14, 7	Female	African Black	Ovarian Germ Cell Tumour (Yolk Sac Tumour)	No	No	None
PC.0026.01.1	1, 8	4, 8	Female	African Black	Retinoblastoma (Left-sided)	No	No	None
PC.0028.01.1	16	17, 9	Male	Mixed Ancestry	Osteosarcoma (Right Femur)	Yes	Yes	None
PC.0031.01.1	2	3, 1	Male	African Black	Astrocytoma (Brain, Grade III, Unbiopsied)	No	No	None
PC.0032.01.1	14	16, 1	Female	Indian	Medullary Thyroid Carcinoma	Yes	Yes	MEN2A
PC.0033.01.1	7, 6	11, 7	Male	African Black	Nephroblastoma (Right-sided)	No	No	None
PC.0034.01.1	1	12, 8	Female	African Black	Sacrococcygeal Teratoma	No	No	None
PC.0037.01.1	7	12, 9	Female	African Black	Osteosarcoma (Right Tibia)	No	No	None

PC.0040.01.1	3	7, 6	Female	African Black	Ovarian Germ Cell Tumour (Mixed, Right Ovary)	No	No	None
PC.0041.01.1	6	8, 8	Female	African Black	Nephroblastoma (Right-sided)	No	No	None
PC.0042.01.1	2, 5	5, 7	Male	African Black	Nephroblastoma (Right-sided)	No	No	None
PC.0043.01.1	1, 6	2, 10	Female	African Black	Retinoblastoma (Left-sided)	No	No	None
PC.0044.01.1	3, 3	8, 6	Female	African Black	Nephroblastoma (Right-sided)	No	No	None
PC.0046.01.1	2, 11	2, 11	Male	African Black	Neuroblastoma	Yes	No	None
PC.0047.01.1	3	9	Female	African Black	Nephroblastoma (Bilateral)	No	No	None
PC.0048.01.1	17	17, 10	Male	African Black	Osteosarcoma (Right Femur)	Yes	Yes	None
PC.0049.01.1	4	7, 4	Male	African Black	Nephroblastoma (Left-sided)	Unknown	Unknown	None
PC.0050.01.1	4, 2	7, 9	Female	African Black	Retinoblastoma (Right-sided)	No	No	None
PC.0051.01.1	1, 6	4, 2	Female	African Black	Embryonal Rhabdomyosarcoma (Right buttock)	Yes	Yes	None

Abbreviations/Acronyms – CPS: Cancer Predisposing Syndrome, MEN2: Multiple Endocrine Neoplasia Type 2, NF1: Neurofibromatosis type 1

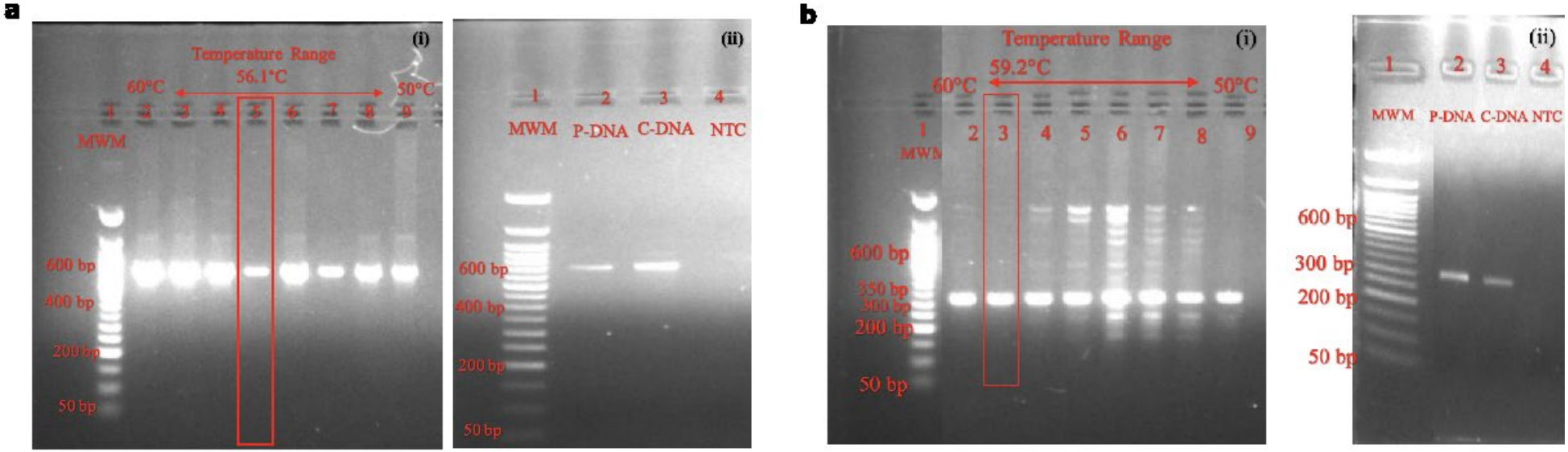
Appendix D2: Post-Alignment Sequencing Quality Metrics per sample

PeCan Project Patient ID	coverageAnalysis Plugin				VariantCaller Plugin
	Mapped Reads	On Target (%)	Mean Base Coverage Depth	Coverage Uniformity (%)	Number of Variants Detected by Ion Torrent Suite™ Software
First Sequencing Run					
PC.0011.01.1	270,852	92.61	199.7	96.14	71
PC.0014.01.1	257,653	93.03	190.9	96.22	92
PC.0016.01.1	245,646	92.43	181.0	95.62	99
PC.0018.01.1	279,065	92.34	205.9	96.14	97
PC.0019.01.1	245,634	93.03	182.5	95.68	87
PC.0028.01.1	241,480	92.65	178.2	95.93	90
PC.0033.01.1	239,029	93.02	177.6	95.85	90
PC.0034.01.1	216,674	93.03	162.3	95.57	94
PC.0040.01.1	250,075	91.62	181.6	95.87	79
PC.0041.01.1	255,727	92.68	189.3	95.84	88
PC.0043.01.1	266,068	92.92	197.1	96.09	95
PC.0044.01.1	244,656	93.03	182.6	96.70	81
PC.0046.01.1	250,205	91.72	181.2	96.13	99
PC.0049.01.1	234,974	93.31	175.0	95.51	104
PC.0050.01.1	207,977	92.86	153.9	95.76	79
PC.0051.01.1	243,215	93.29	182.5	95.92	96
Second Sequencing Run					
PC.0003.01.1	367,586	93.83	261.9	95.19	250
PC.0006.01.1	393,081	94.61	284.1	94.23	209
PC.0007.01.1	399,049	94.18	284.5	93.61	235
PC.0008.01.1	394,734	94.79	306.0	94.20	216
PC.0009.01.1	346,375	94.23	248.8	92.28	233
PC.0012.01.1	373,733	94.28	270.3	94.21	222
PC.0021.01.1	399,529	94.38	289.9	94.59	205
PC.0023.01.1	418,540	94.79	306.0	94.20	178
PC.0024.01.1	404,330	94.02	290.9	94.43	191
PC.0026.01.1	377,177	94.83	275.4	93.48	225
PC.0031.01.1	344,898	94.84	250.3	93.37	235
PC.0032.01.1	349,326	94.83	253.0	93.34	183
PC.0037.01.1	344,619	94.69	249.0	92.70	204
PC.0042.01.1	359,480	94.75	261.5	94.50	223
PC.0047.01.1	372,362	94.87	273.3	94.34	230
PC.0048.01.1	211,811	94.84	153.7	88.19	198

Appendix D3.1: Resultant Gel Electrophoresis Images for PCR Primer – Pair Optimisation & PCR Patient DNA Amplification

Variant *NF1* c.1133_1136delTGAC in Patient PC.0023.01.1

Variant *RET* c.1901G>A in Patient PC.0032.01.1

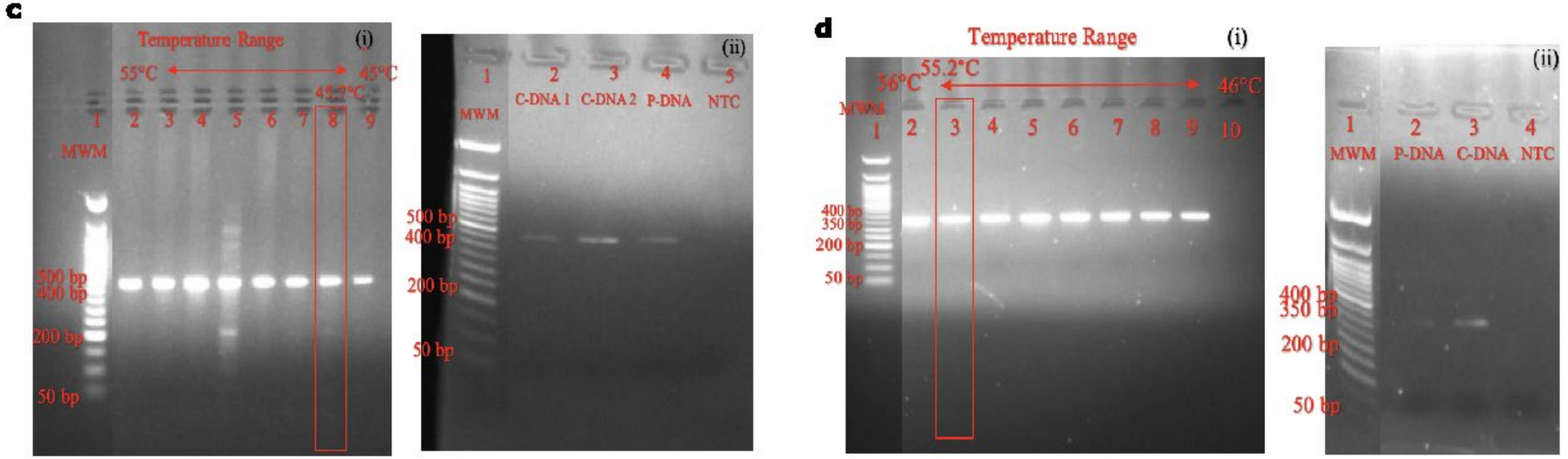


A (i) Optimal annealing temperature determined as 56.1°C (lane 5). MWM (lane 1) indicates the amplified region is around 600 bp, as expected according to our primer design (604 bp). **(ii) Patient PC.0023.01.1 DNA** (P-DNA – lane 2) and control DNA (C-DNA – lane 3) successfully amplified, as shown by clear single bands. NTC (lane 4) shows no band present. MWM (lane 1) indicates the amplified region is around 600 bp, as expected according to our primer design (604 bp). **B (i) Optimal annealing temperature determined as 59.2°C (lane 3).** MWM (lane 1) indicates the amplified region is around 300 bp, as expected according to our primer design (312 bp). **(ii) Patient PC.0032.01.1 DNA** (P-DNA – lane 2) and control DNA (C-DNA – lane 3) successfully amplified, as shown by clear single bands. NTC (lane 4) shows no bands present. MWM (lane 1) indicates the amplified region is around 300 bp, as expected according to our primer design (312 bp).

Appendix D3.1: Resultant Gel Electrophoresis Images for PCR Primer – Pair Optimisation & PCR Patient DNA

Variant *TP53* c.817C>T in Patient PC.0051.01.1

Variant *BRCA1* c.4199T>C in Patient PC.0011.01.1

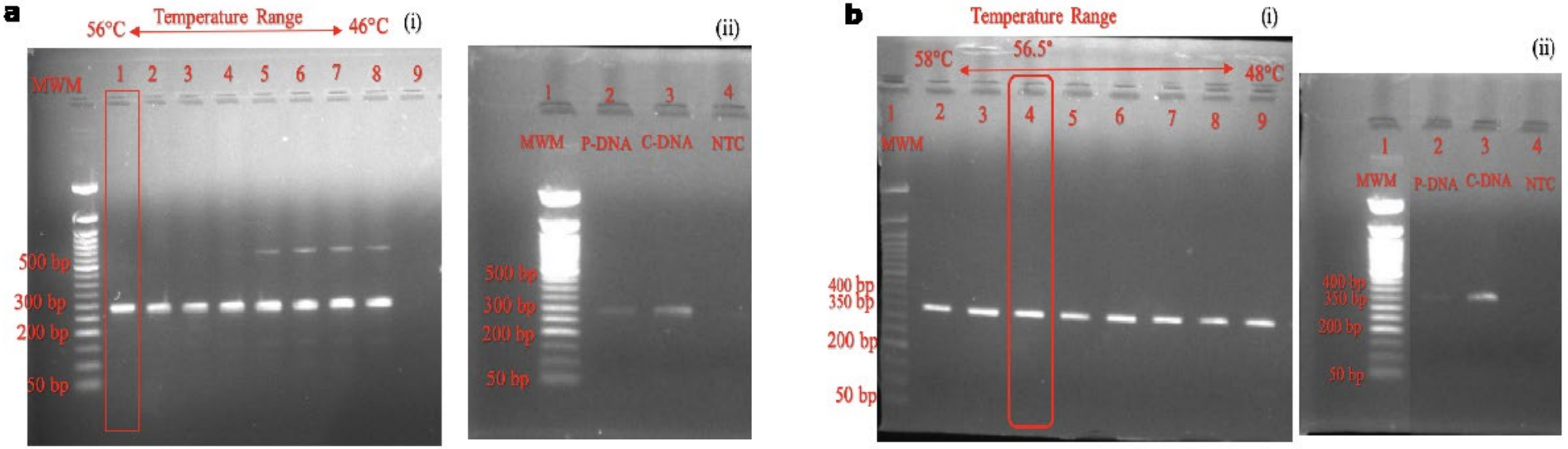


C (i) Optimal annealing temperature determined as 45.7°C (lane 8). MWM (lane 1) indicates the amplified region is around 500 bp, as expected according to our primer design (516 bp). **(ii) Patient PC.0051.01.1 DNA** (P-DNA – lane 4) and control DNA (C-DNA – lane 2 & 3) successfully amplified, as shown by clear single bands. NTC (lane 5) shows no band present. MWM (lane 1) indicates the amplified region is around 500 bp, as expected according to our primer design (516 bp). **D (i) Optimal annealing temperature determined as 55.2°C (lane 3).** MWM (lane 1) indicates the amplified region is around 350 bp, as expected according to our primer design (364 bp). **(ii) Patient PC.0011.01.1 DNA** (P-DNA – lane 2) and control DNA (C-DNA – lane 3) successfully amplified, as shown by clear single bands. NTC (lane 4) shows no band present. MWM (lane 1) (although slightly skew) indicates the amplified region is around 350 bp, as expected according to our primer design (364 bp).

Appendix D3.2: Resultant Gel Electrophoresis Images for PCR Primer – Pair Optimisation & PCR Patient DNA Amplification

Variant *FAH* c.485G>A in Patient PC. 0009.01.1

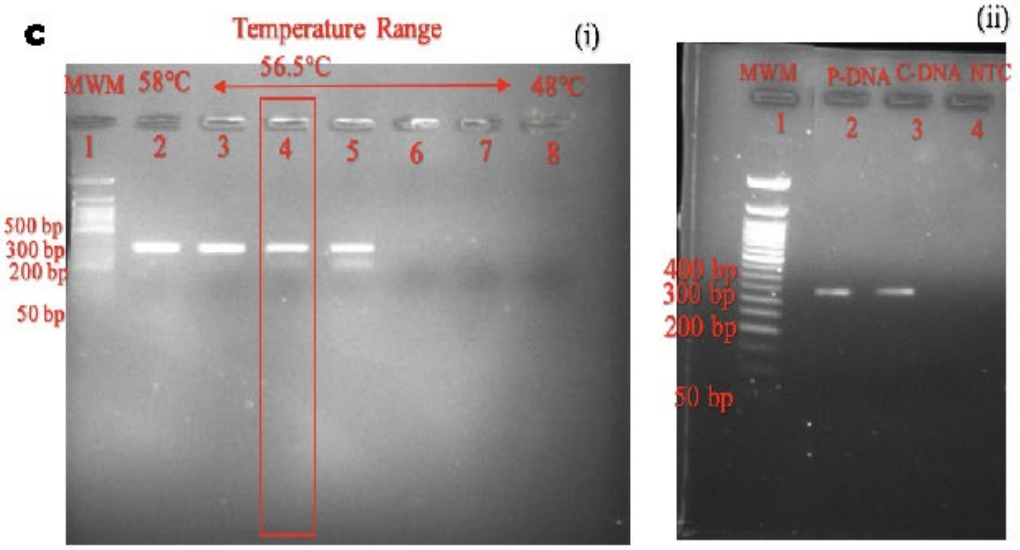
Variant *ERCC3* c.958C>T in Patient PC.0003.01.1



(A) (i) Optimal annealing temperature determined as 56°C (lane 1). MWM indicates the amplified region is around 300 bp, as expected according to our primer design (290 bp). **(ii) Patient PC.0009.01.1 DNA** (P-DNA – lane 2) and control DNA (C-DNA – lane 3) successfully amplified, as shown by clear single bands. NTC (lane 4) shows no band present. MWM (lane 1) (although slightly skew) indicates the amplified region is around 300 bp, as expected according to our primer design (290 bp). **B (i) Optimal annealing temperature determined as 56.5°C (lane 4)**. MWM indicates the amplified region is around 350 bp, which is expected according to our primer design (340 bp). **(ii)**. **Patient PC.0003.01.1 DNA** (P-DNA – lane 2) and control DNA (C-DNA – lane 3) successfully amplified, as shown by clear single bands. NTC (lane 4) shows no band present. MWM (lane 1) indicates the amplified region is around 350 bp, which is expected according to our primer design (340 bp).

Appendix D3.2: Resultant Gel Electrophoresis Images for PCR Primer – Pair Optimisation & PCR Patient DNA Amplification

Variant RB1 c.1981C>T in Patient PC.0018.01.1

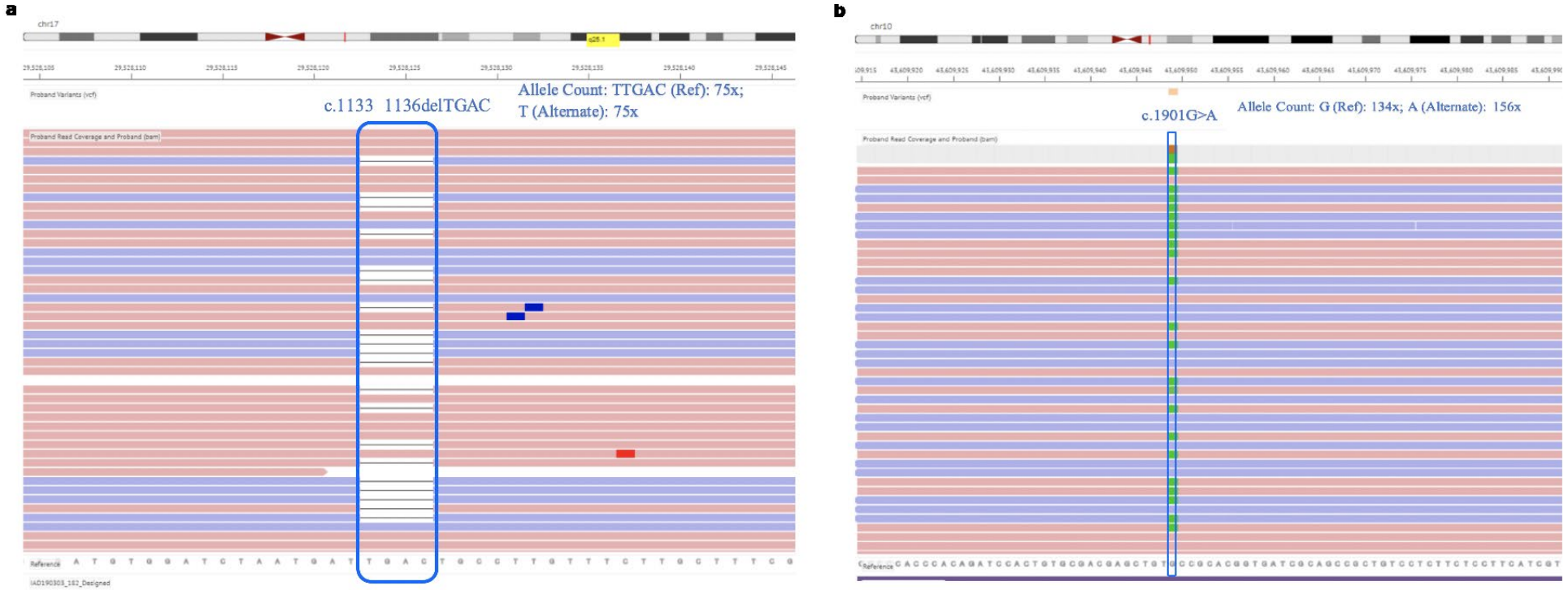


C (i) Optimal annealing temperature determined as 56.5°C (lane 4), the primer pair did not amplify at the lower temperatures. MWM indicates the amplified region is around 300 bp, as expected according to our primer design (324 bp). **C (ii) Patient PC.0018.01.1 DNA** (P-DNA – lane 2) and control DNA (C-DNA – lane 3) successfully amplified, as shown by clear single bands. NTC (lane 4) shows no band present. MWM (lane 1) indicates the amplified region is around 300 bp, as expected according to our primer design (324 bp).

Appendix D4.1: IRGV Images for Mutation-Positive Cases

IRGV Image for Heterozygous Variant *NF1* c.1133_1136delTGAC Identified in Patient PC.0023.01.1

IRGV Image for Heterozygous Variant *RET* c.1901G>A Identified in Patient PC.0032.01.1



(A) The sequencing region which contains the NF1 four base-pair deletion (TGAC) identified in patient PC.0023.01.1 is highlighted in the blue box. Reads with and without the deletion can be seen, depicting the heterozygous nature of this variant. The reference sequence can be seen at the bottom of the image. Pink reads indicate positive strand reads and purple reads indicated negative strand reads. **(B)** The RET single base pair substitution (G>A) identified in patient PC.0032.01.1. is highlighted in the blue box. Reads with the green blocks are those where the alternate A allele was detected. Reads with and without the substitution can be seen, depicting the heterozygous nature of this variant.

Appendix D4.1: IRGV Images for Mutation-Positive Cases

IRGV Image for Heterozygous Variant *TP53* c.817G>A Identified in Patient PC.0051.01.1

c



IRGV Image for Heterozygous Variant BRCA1 c.4199A>G Identified in Patient PC.0011.01.1

d

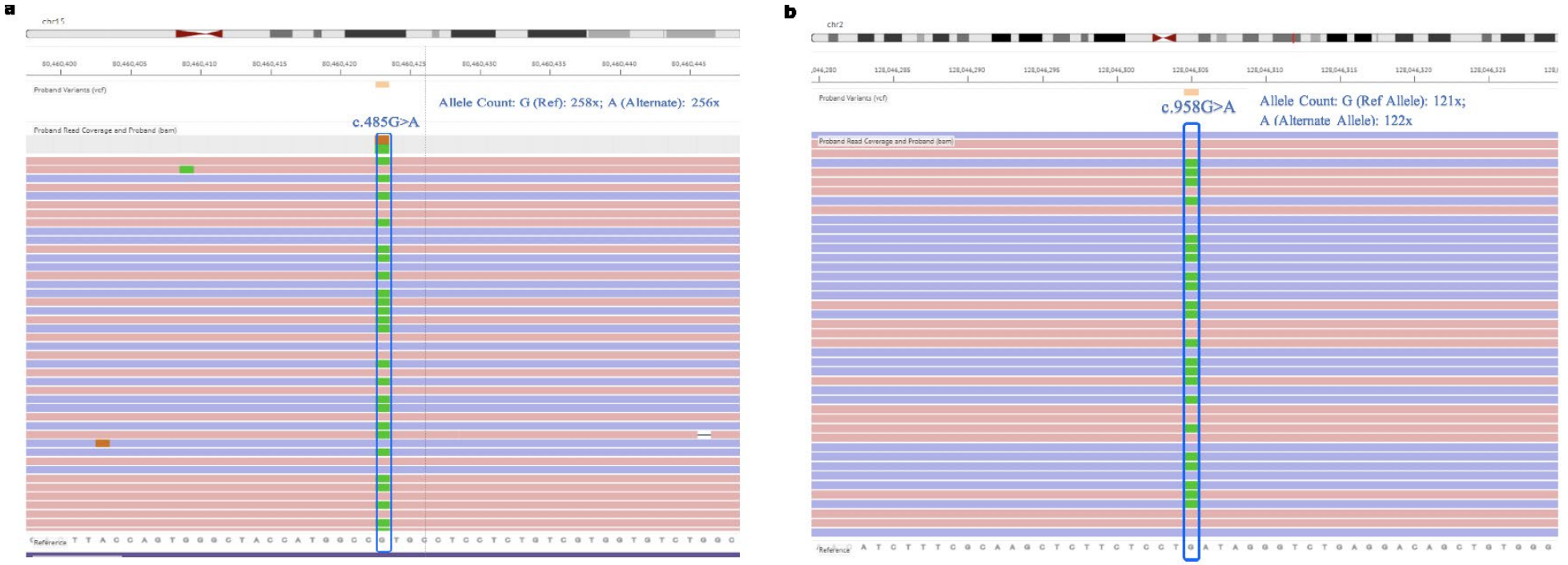


(C) The variant was reported in the strand complementary to the c.817C>T variant. Sequencing region which contains the *TP53* single base pair substitution (G>A) is highlighted in the blue box. Reads with the green blocks are those where the alternate A allele was detected. Reads with and without the substitution can be seen, depicting the heterozygous nature of this variant. The reference sequence can be seen at the bottom of the image. Pink reads indicate positive strand reads and purple reads indicate negative strand reads. **(D)** The variant was reported in the strand complementary to the c.4199T>C variant. Sequencing region which contains the *BRCA1* c.4199A>G substitution identified in patient PC.0011.01.1 is highlighted in the blue box. Reads with the dark orange blocks are those where the alternate G allele was detected. Reads with and without the substitution can be seen, depicting the heterozygous nature of this variant. The reference sequence can be seen at the bottom of the image. Pink reads indicate positive strand reads and purple reads indicate negative strand reads.

Appendix D4.2: IRGV Images for Mutation-Positive Cases

IRGV Image for Heterozygous Variant *FAH* c.485G>A Identified in Patient PC.0009.01.1

IRGV Image for Heterozygous Variant ERCC3 c.958G>A Identified in Patient PC.0003.01.1

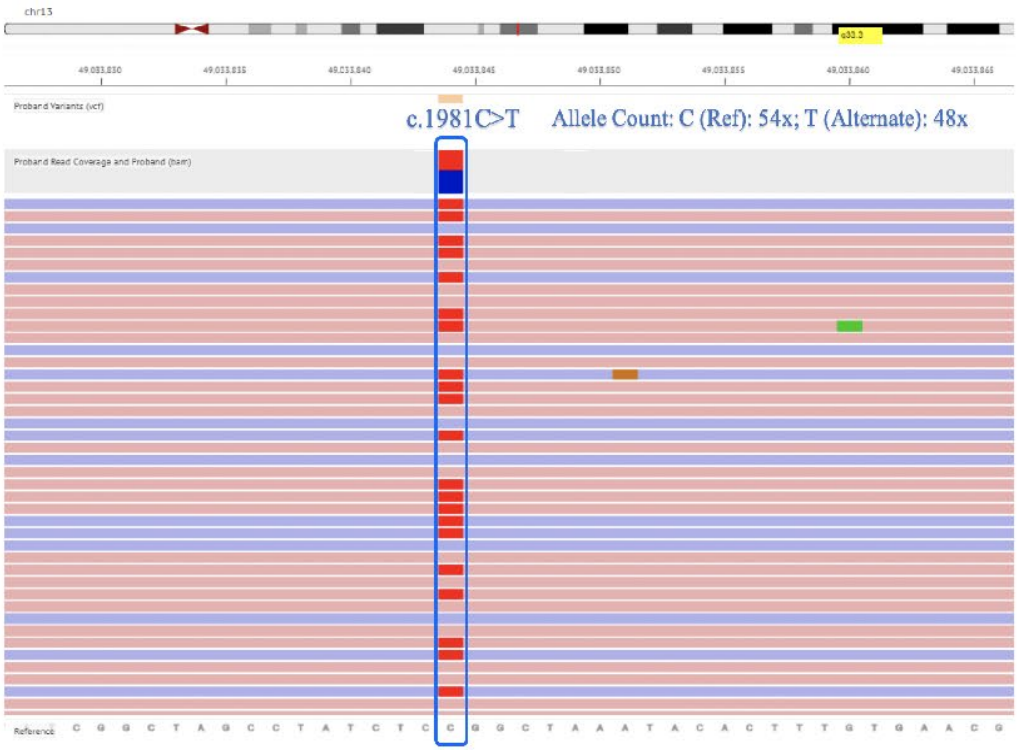


(A) Sequencing region which contains the *FAH* single base pair substitution (G>A) identified in patient PC.0009.01.1 is highlighted in the blue box. Reads with the green blocks are those where the alternate A allele was detected. Reads with and without the substitution can be seen, depicting the heterozygous nature of this variant. The reference sequence can be seen at the bottom of the image. Pink reads indicate positive strand reads and purple reads indicate negative strand reads. **(B)** The variant was reported in the strand complementary to the c.958C>T variant. Sequencing region which contains the *ERCC3* single base pair substitution (G>A) identified in patient PC.0003.01.1 is highlighted in the blue box. Reads with the green blocks are those where the alternate A allele was detected. Reads with and without the substitution can be seen, depicting the heterozygous nature of this variant. The reference sequence can be seen at the bottom of the image. Pink reads indicate positive strand reads and purple reads indicate negative strand reads.

Appendix D4.2: IRGV Images for Mutation-Positive Cases

IRGV Image for Heterozygous Variant *RB1* c.1981C>T Identified in Patient PC.0018.01.1

C



(C) Sequencing region which contains the *RB1* single base pair substitution (C>T) identified in patient PC.0018.01.1 is highlighted in the blue box. Reads with the red blocks are those where the alternate T allele was detected. Reads with and without the substitution can be seen, depicting the heterozygous nature of this variant. The reference sequence can be seen at the bottom of the image. Pink reads indicate positive strand reads and purple reads indicate negative strand reads.

Appendix E: Turnitin Report

Erin Manolas MSc Dissertation 733739 Final.pdf

ORIGINALITY REPORT

12%	8%	10%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	<1%
2	Julia Würtemberger, Tim Ripperger, Christian Vokuhl, Sebastian Bauer et al. "Genetic susceptibility in children, adolescents, and young adults diagnosed with soft-tissue sarcomas", European Journal of Medical Genetics, 2023 Publication	<1%
3	Neurofibromatosis Type 1, 2012. Publication	<1%
4	www.ncbi.nlm.nih.gov Internet Source	<1%
5	dokumen.pub Internet Source	<1%
6	link.springer.com Internet Source	<1%
7	www.frontiersin.org Internet Source	<1%