

# IDENTIFICATION OF GERMLINE VARIANTS IN A COHORT OF SOUTH AFRICAN PAEDIATRIC CANCER PATIENTS

**Lana Naudé**

**Student number: 2740639**

**Supervisor: Dr Lindie Lamola**

**Co-supervisor: Dr Thandiswa Ngcungcu**



UNIVERSITY OF THE  
WITWATERSRAND,  
JOHANNESBURG

A research report (in the format of a “submissible” paper) submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the Degree of Master of Science in Medicine (Genomic Medicine) by Research and Coursework.

Johannesburg, 2025

# 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 9223 | [M] +27 78 080 8841 | [F] +27 11 489 9226 | [E] human.genetics@nhls.ac.za

**Re: Research report submitted to the Faculty of Health Sciences, University of the Witwatersrand, by Ms Lana Naudé in partial fulfilment of the requirements for the degree Master of Science in Medicine (Genomic Medicine)**

Dear examiner,

Thank you for agreeing to examine the research report submitted by Ms Lana Naudé (student number 2740639) entitled "Identification of germline variants in a cohort of South African paediatric cancer patients". This research report is being submitted by the candidate in the format of a "submissible" paper, as per the Style Guide for Thesis, Dissertation and Research Reports (updated March 2016) by the University of the Witwatersrand Faculty of Health Sciences. This report contributes 33% towards the final mark of the MSc (Med) Genomic Medicine degree for which the candidate is enrolled.

The candidate and her supervisors have chosen to submit her paper to the journal Frontiers. The submissible paper attached is therefore written in concordance with the author guidelines of the stipulated journal. These guidelines have been attached for your reference in the appendices section of this document. There are deviations purposively made from the journal submission guidelines, including that the tables and figures are placed in the text and not submitted separately as per the Frontiers guidelines. Additionally, the manuscript has been written in British English instead of American English, Calibri font was used, line numbers are not inserted and a line spacing of 1.5 was used instead of a line spacing of 1 to enhance readability.

The research protocol has also been attached as an appendix for the purposes of providing an extended literature review and further contextualise the study. The protocol has already been assessed and approved by the Faculty of Health Sciences Post-Graduate Assessors committee and is therefore not for examination.

Please refer to the Table of Contents for a complete list of the contents provided in this research report.

Yours sincerely,

Lana Naudé (Candidate)

Dr Lindie Lamola (Supervisor)

Dr Thandiswa Ngcungcu (Co-Supervisor)

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 CEO: Prof Koleka Mlisana  
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

## Declaration

I, Lana Naudé, declare that this Research Report (in the format of a “submissible” paper) is my own, unaided work. It is being submitted for the Degree of Master of Science in Medicine (Genomic Medicine) by Research and Coursework at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



---

Lana Naudé

Student number: 2740639

27<sup>th</sup> day of February 2025 in Johannesburg

## Contribution of Candidate

### Student's contribution to the article(s) and agreement of co-author(s)

I, Lana Naudé, student number 2740639, declare that this Research Report, entitled "Identification of germline variants in a cohort of South African paediatric cancer patients", is my own work and that I contributed significantly towards the research findings presented in the paper intended for publication below.

  
\_\_\_\_\_  
Signature of student

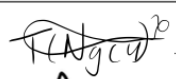
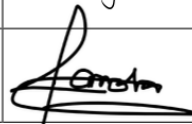
27/02/2025  
\_\_\_\_\_  
Date

  
\_\_\_\_\_  
Signature of Primary Supervisor

27/02/2025  
\_\_\_\_\_  
Date

Agreement by co-authors:

By signing this declaration, the co-authors listed below agree to the use of the research findings intended for a published article by the student as part of their Research Report.

| Authors         | Name                  | Signature                                                                            | Date                           |
|-----------------|-----------------------|--------------------------------------------------------------------------------------|--------------------------------|
| 1 <sup>st</sup> | Ms Lana Naudé         |   | 27 <sup>th</sup> February 2025 |
| 2 <sup>nd</sup> | Dr Thandiswa Ngcungcu |  | 27 <sup>th</sup> February 2025 |
| 3 <sup>rd</sup> | Dr Lindie Lamola      |  | 27 <sup>th</sup> February 2025 |

## **Dedication**

To the brave children and families who face the challenges of paediatric cancer with strength and resilience. Your courage inspires our pursuit of knowledge and innovation.

### **Presentations arising from this research project**

1. Identification of germline variants in a cohort of South African paediatric cancer patients. YEGG Symposium, University of the Witwatersrand: Poster presentation. 26 October 2024
2. Identification of germline variants in a cohort of South African paediatric cancer patients. SASHG Biennial Congress 2024, Sun City: Poster presentation. 27<sup>th</sup> – 29<sup>th</sup> October 2024

## Abstract

Paediatric cancers, which affect individuals under 18 years of age, are often attributed to inherited germline genetic variants due to the limited time for the accumulation of DNA alterations. Variants in cancer predisposition genes- such as tumour suppressor genes (TSGs), proto-oncogenes and DNA repair genes- are known to increase cancer risk. This study aimed to identify germline single nucleotide variants (SNVs) in paediatric cancer patients using Whole Exome Sequencing (WES), with a focus on four common childhood solid tumours; nephroblastoma, retinoblastoma, rhabdomyosarcoma and osteosarcoma. DNA was extracted from blood samples of 14 patients diagnosed with one of these cancers, and a virtual gene panel consisting of 64 cancer predisposition genes was used to analyse WES data. A total of 212,825 variants were processed, with the majority being intronic, followed by synonymous, missense, splice-site and indel variants. Variants were prioritised based on factors such as minor allele frequencies, variant types, locations and in silico prediction scores. These variants were further assessed using databases like ClinVar and AlphaMissense database (AMDB) and visualised using Integrative Genomics Viewer (IGV). Two heterozygous missense variants in TSGs were classified as variants of uncertain significance (VUSs) according to ACMG/ AMP guidelines: one in the *TP53* gene (NM\_000546.6:c.923T>C) in a retinoblastoma patient (PC79) and another in the *SDHB* gene (NM\_003000.3:c.455C>T) in a rhabdomyosarcoma patient (PC2), both of African ancestry. While the identification of these VUSs provide valuable insights into potential germline genetic causes of paediatric cancers, the absence of definitive pathogenic variants underscores the need for further investigation. This study also highlights the challenges of interpreting genetic data in underrepresented populations, such as those of African descent, where VUSs are more prevalent. Despite these challenges, WES proved to be valuable in identifying germline variants and advancing our understanding of paediatric cancer genetics in South Africa.

## **Acknowledgements**

I would like to express my sincere gratitude to my supervisor, Dr Lindie Lamola, for her guidance, unwavering support and insightful feedback throughout this research. Her mentorship and encouragement have been instrumental in shaping the direction and quality of this work. Dr Lamola's expertise has been invaluable in navigating the challenges of this project, and I am deeply appreciative of her continuous belief in my abilities.

I would like to equally extend my heartfelt thanks to my co-supervisor, Dr Thandiswa Ngcungcu, for her significant contributions to this research. Her thought-provoking ideas, constructive critiques, and exceptional knowledge have played a significant role in deepening my understanding of the subject. Dr Ngcungcu has not only provided academic guidance, but also personal encouragement, for which I am truly grateful.

I am genuinely thankful for the supportive environment fostered by everyone at the Division of Human Genetics. I would also like to extend my sincere thanks for the financial assistance that enabled me to attend the South African Society of Human Genetics Congress in Sun City. Your generosity has played a crucial role in my professional and personal development. To my fellow classmates, it has truly been a privilege to share this journey with you and to collaborate ideas and strengthen each other throughout this experience.

My deepest gratitude goes to my loved ones- Hannes, Gretha, Sumi, Eduan, and Thomas- whose unwavering love has been the foundation of this journey. Your understanding through the highs and lows have made all the difference. Together, you have all been my cornerstone, and I am eternally thankful for the belief you've shown in me.

## Table of Contents

|                                                                                                        |    |
|--------------------------------------------------------------------------------------------------------|----|
| Letter of approval.....                                                                                | 2  |
| Declaration.....                                                                                       | 3  |
| Contribution of Candidate .....                                                                        | 4  |
| Dedication .....                                                                                       | 5  |
| Presentations arising from this research project.....                                                  | 6  |
| Abstract.....                                                                                          | 7  |
| Acknowledgements .....                                                                                 | 8  |
| Table of Contents .....                                                                                | 9  |
| List of Figures .....                                                                                  | 11 |
| List of Tables.....                                                                                    | 12 |
| List of Abbreviations .....                                                                            | 13 |
| Research report in the format of a “submissible paper” .....                                           | 15 |
| Abstract for Journal Submission .....                                                                  | 16 |
| 1. Introduction .....                                                                                  | 17 |
| 2. Materials and Methods.....                                                                          | 21 |
| 2.1 Study cohort .....                                                                                 | 21 |
| 2.2 Virtual gene panel generation .....                                                                | 22 |
| 2.3 Variant annotation and prioritisation.....                                                         | 22 |
| 2.4 Variant visualisation.....                                                                         | 23 |
| 2.5 Variant classification .....                                                                       | 24 |
| 3. Results .....                                                                                       | 24 |
| 3.1 Frequency and types of variants across different paediatric cancers .....                          | 25 |
| 3.2 Variants of uncertain significance identified in retinoblastoma and rhabdomyosarcoma patients..... | 25 |
| 3.2.1 The NM_000546.6:c.923T>C variant in the <i>TP53</i> gene of a retinoblastoma patient .....       | 26 |
| 3.2.2 The NM_003000.3:c.455C>T variant in the <i>SDHB</i> gene of a rhabdomyosarcoma patient .....     | 27 |

|                                                                                        |    |
|----------------------------------------------------------------------------------------|----|
| 4. Discussion .....                                                                    | 27 |
| 4.1 The <i>TP53</i> NM_000546.6:c.923T>C variant in retinoblastoma patient PC79 .....  | 28 |
| 4.2 The <i>SDHB</i> NM_003000.3:c.455C>T variant in rhabdomyosarcoma patient PC2 ..... | 29 |
| 4.3 Variants of Uncertain Significance .....                                           | 30 |
| 4.4 Limitations .....                                                                  | 31 |
| 4.5 Future work .....                                                                  | 31 |
| 5. Conclusion .....                                                                    | 32 |
| Acknowledgement .....                                                                  | 33 |
| Statement of Ethics .....                                                              | 33 |
| Conflict of Interest Statement .....                                                   | 33 |
| Funding Sources .....                                                                  | 33 |
| Author contributions .....                                                             | 32 |
| Data availability statement .....                                                      | 33 |
| References .....                                                                       | 34 |
| Appendices .....                                                                       | 38 |
| Appendix A: Supplementary Tables and Figures .....                                     | 39 |
| Appendix B: Ethics Clearance Certificate .....                                         | 42 |
| Appendix C: Approved Research Protocol .....                                           | 43 |
| Appendix D: Turnitin Report and Plagiarism Declaration .....                           | 66 |
| Appendix E: Author Guidelines for <i>Frontiers</i> .....                               | 69 |

## List of Figures

|                                                                                                |    |
|------------------------------------------------------------------------------------------------|----|
| Figure 1: Proportional distribution of paediatric cancer types worldwide .....                 | 17 |
| Figure 2: Schematic overview of the methodology employed in this study .....                   | 21 |
| Figure 3: SNV Prioritisation Workflow .....                                                    | 23 |
| Figure 4: Distribution of variant types prior to variant prioritisation .....                  | 25 |
| Figure 5: Genomic location of the <i>TP53</i> NM_000546.6:c.923T>C (p.Leu308Pro) variant ..... | 29 |
| Figure 6: Genomic location of the <i>SDHB</i> NM_003000.3:c.455C>T (p.Ser152Phe) variant. .... | 30 |

## List of Tables

|                                                                                                   |    |
|---------------------------------------------------------------------------------------------------|----|
| Table 1: Genes most commonly associated with four cancer types and their related CPSs .....       | 19 |
| Table 2: Patient characteristics .....                                                            | 22 |
| Table 3: Demographic characteristics of the study cohort (n=14).....                              | 24 |
| Table 4: Details of the variants identified in retinoblastoma and rhabdomyosarcoma patients ..... | 26 |

## List of Abbreviations

|             |                                                                                      |
|-------------|--------------------------------------------------------------------------------------|
| ACMG/AMP    | American College of Medical Genetics and Genomics / Association of Medical Pathology |
| AMDB        | AlphaMissense Database                                                               |
| BAM         | Binary Alignment Map                                                                 |
| BWS         | Beckwith-Wiedemann syndrome                                                          |
| CADD        | Combined Annotation Depletion                                                        |
| CNV         | Copy number variant                                                                  |
| CPS         | Cancer predisposition syndrome                                                       |
| DDS         | Denys-Drash syndrome                                                                 |
| GIST        | Gastrointestinal stromal tumour                                                      |
| GnomAD      | Genome Aggregation Database                                                          |
| HGVS        | Human Genome Variation Society                                                       |
| <i>HRAS</i> | <i>Harvey Rat Sarcoma Virus</i>                                                      |
| <i>IGF2</i> | <i>Insulin-like Growth Factor 2</i>                                                  |
| IGV         | Integrative Genomics Viewer                                                          |
| MAF         | Minor allele frequency                                                               |
| <i>MDM2</i> | <i>Murine Double Minute 2</i>                                                        |
| <i>MDM4</i> | <i>Murine Double Minute 4</i>                                                        |
| <i>MED4</i> | <i>Mediator complex subunit 4</i>                                                    |
| NF1         | Neurofibromatosis Type 1                                                             |
| <i>NF1</i>  | <i>Neurofibromin1</i>                                                                |
| NLS         | Nuclear Localisation Signal                                                          |
| PeCan       | Paediatric Cancer Project                                                            |
| PolyPhen-2  | Polymorphism Phenotyping                                                             |
| <i>RB1</i>  | <i>Retinoblastoma1</i>                                                               |
| SDH         | Succinate dehydrogenase                                                              |
| SDHAF1      | Succinate dehydrogenase assembly factor 1                                            |
| <i>SDHB</i> | <i>Succinate dehydrogenase complex iron sulphur subunit B</i>                        |
| SIFT        | Sorting Intolerant from Tolerant                                                     |
| SNV         | Single nucleotide variant                                                            |
| <i>TP53</i> | <i>Tumour Protein p53</i>                                                            |

|            |                                   |
|------------|-----------------------------------|
| TSG        | Tumour-suppressor genes           |
| UniProt    | Universal Protein Resource        |
| VCF        | Variant Call Format               |
| VEP        | Variant Effect Predictor          |
| VUS        | Variant of Uncertain Significance |
| WES        | Whole Exome Sequencing            |
| WGAR       | Wilms tumour-aniridia syndrome    |
| WGS        | Whole Genome Sequencing           |
| <i>WT1</i> | <i>Wilms' tumour 1</i>            |

Research report in the format of a “submissible” paper

To be submitted to the Frontiers journal.

# Identification of germline variants in a cohort of South African paediatric cancer patients

Lana Naudé<sup>1</sup>, Thandiswa Ngcungcu<sup>1</sup>, Lindie Lamola<sup>1</sup>

<sup>1</sup> Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

Corresponding author: Lana Naudé

Email address: lananude95@gmail.com/ 2740639@students.wits.ac.za

Keywords: Paediatric cancer, germline variants, Whole Exome Sequencing, Virtual Gene Panel, retinoblastoma, rhabdomyosarcoma, South Africa

## Abstract for Journal Submission

**Introduction:** Paediatric cancers are characterised by genetic alterations that disrupt normal cellular growth. In children and adolescents, inherited germline variants thought to play a key role in tumour development. Germline variants in cancer predisposition genes- such as tumour suppressor genes (TSGs), proto-oncogenes and DNA repair genes- are associated with an increased cancer risk. This study aimed to identify germline single nucleotide variants (SNVs) in paediatric cancer patients, utilising Whole Exome Sequencing (WES) data. The focus of this study was on four common childhood solid cancers, namely: nephroblastoma, retinoblastoma, rhabdomyosarcoma and osteosarcoma.

**Methods:** WES was performed on DNA extracted from blood samples of 14 patients diagnosed with one of the four paediatric cancers. A custom virtual gene panel of 64 cancer predisposition genes was created. Subsequently, variants were prioritised based on factors such as minor allele frequencies, variant types, variant locations, and in silico prediction scores. Identified variants were further assessed using public databases, including ClinVar and AMDB, and classified according to ACMG/ AMP guidelines.

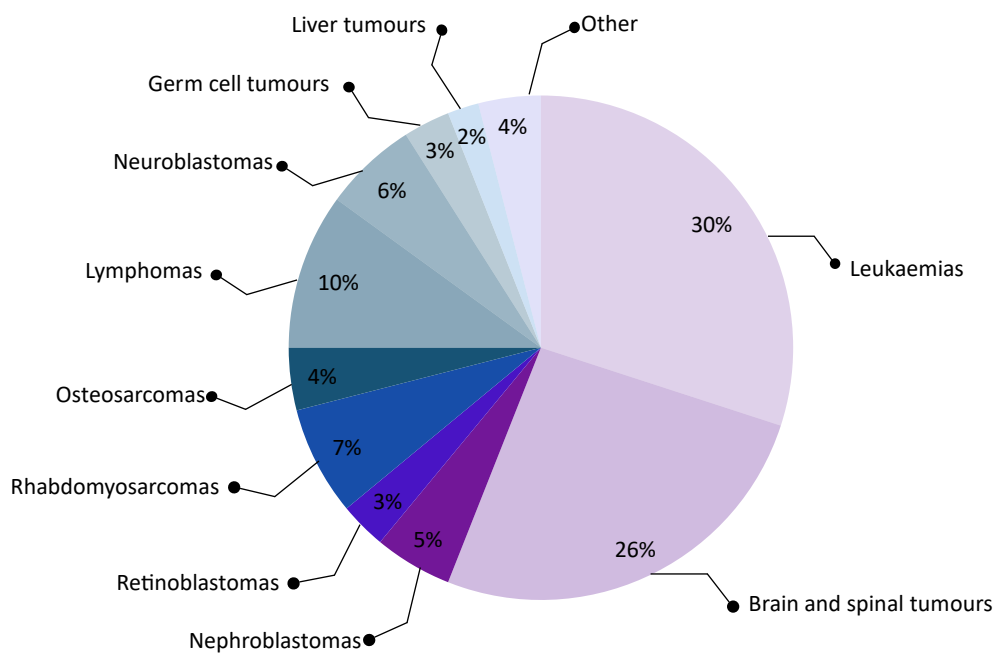
**Results:** A total of 212,825 variants were processed across the 14 patients, with the majority being intronic, followed by synonymous, missense, splice-site and indel variants. Two missense variants were prioritised: one in the *TP53* gene (NM\_000546.6:c.923T>C) in a retinoblastoma patient (PC79), and another in the *SDHB* gene (NM\_003000.3:c.455C>T) in a rhabdomyosarcoma patient (PC2), both of African ancestry. These two variants were absent on the GnomAD database and confirmed as true variants through visualisation in the Integrative Genomics Viewer (IGV), where heterozygous allele ratios were observed. Based on the available evidence, both variants were classified as variants of uncertain significance (VUSs).

**Conclusion:** The identification of two VUSs provide insights into the possible germline genetic causes of these paediatric cancers, however the lack of definitive pathogenic variants underscores the need for further investigations. The study also emphasises the challenges of interpreting genetic variants in underrepresented populations, such as those of African descent, where VUSs are more prevalent. Despite these challenges, WES proved to be valuable in uncovering germline variants and advancing our understanding of paediatric cancer genetics in South Africa.

## 1. Introduction:

Paediatric cancers are defined by genetic alterations that disrupt normal cellular growth in children and adolescents under the age of 18 years (Ren et al., 2022). As individuals age, the likelihood of accumulating DNA alterations increases (Kuhlen et al., 2019). Therefore, the etiological factors contributing to paediatric cancers may differ from those of adult-onset cancers (Seo et al., 2024). Instead, inherited germline genetic variants are thought to play a more significant role in development of paediatric tumours (Uchimura et al., 2015). The global prevalence of paediatric cancers is estimated to be approximately 175,000 new cases annually (Ren et al., 2022).

The most common paediatric cancers include leukaemia's, brain and spinal tumours, nephroblastomas, retinoblastomas, rhabdomyosarcomas, osteosarcomas, lymphomas, neuroblastomas, germ cell tumours and liver tumours (Figure 1) (Wu et al., 2022). Paediatric cancers are categorised into solid and non-solid tumours (Huang et al., 2018). Leukaemia, a non-solid tumour, is particularly common due to the rapid division of blood cells, which increases the likelihood of mutations (Huang et al., 2018). For this sub-study, the solid tumours nephroblastoma, retinoblastoma, rhabdomyosarcoma and osteosarcoma were selected due to their high incidence, collectively accounting for approximately 19% of all paediatric cancer cases (Figure 1) (Wu et al., 2022).



**Figure 1:** Proportional distribution of paediatric cancer types worldwide. Nephroblastoma, retinoblastoma, rhabdomyosarcoma and osteosarcoma collectively represent approximately 19% of all paediatric cancer cases globally. Figure adapted from Wu et al., 2022.

While most cancers arise sporadically, due to acquired mutations, individuals carrying germline variants in cancer predisposition genes have an increased risk of developing cancer (Kuhlen et al., 2019). Cancer predisposition syndromes (CPSs) refer to inherited conditions caused by variants in cancer predisposition genes (Steliarova-Foucher et al., 2017). These genes include tumour-suppressor genes (TSGs), proto-oncogenes and DNA repair genes (Davidoff, 2010). TSGs regulate cell growth and prevent uncontrolled proliferation, variants in these genes can lead to tumour formation (Davidoff, 2010). The two-hit hypothesis explains that cancers due to germline variants in TSGs typically require two mutations; one inherited and one acquired (Knudson, 1971). Proto-oncogenes normally regulate cell division, but variants can turn them into oncogenes, driving cancer development by stimulating cell growth and survival (Davidoff, 2010). DNA repair genes, responsible for correcting DNA damage, play a role in preventing the accumulation of variants, thus reducing the risk of tumour formation (Davidoff, 2010). Variants in the genes investigated in this study fell within these categories of cancer predisposition genes.

Nephroblastoma generally originates from abnormal kidney development, with a median age at diagnosis between two and three years (Steliarova-Foucher et al., 2017). Germline variants in oncogenes, such as *Wilms' tumour 1 (WT1)*, are typically implicated in bilateral nephroblastoma (Parsons et al., 2016). These in *WT1* variants also increase the risk of several CPSs, including Wilms tumour-aniridia syndrome (WGAR) and Denys-Drash syndrome (DDS) (Table 1) (Ruteshouser et al., 2008). Additionally, the *Insulin-like Growth Factor 2 (IGF2)* oncogene have been implicated in the aetiology of nephroblastoma (Murphy et al., 2023). Beckwith-Wiedemann syndrome (BWS), which involves multiple germline variants in the *IGF2* gene, is a CPS associated with an increased risk of nephroblastoma (Table 1) (Ruteshouser et al., 2008).

Retinoblastoma is an intraocular paediatric cancer, primarily affecting the neural retina (Pandey, 2014). The median diagnostic age is between one and two years, and its development is strongly associated with germline variants in the *Retinoblastoma1 (RB1)* TSG (Steliarova-Foucher et al., 2017). Additionally, a CPS impacting the onset of retinoblastoma has been observed in 13q deletion syndrome, characterised by a deletion including the *RB1* gene (Table 1) (Parsons et al., 2016). New studies have indicated that the development of retinoblastoma may be influenced by variants in

genetic modifier genes such as *Murine Double Minute 2 (MDM2)*, *Murine Double Minute 4 (MDM4)* or *Mediator complex subunit 4 (MED4)* that alters the expression of the *RB1* gene (Steliarova-Foucher et al., 2017).

Rhabdomyosarcoma is a type of soft tissue sarcoma that primarily affects skeletal muscle, often resulting from the inability of skeletal myoblasts to mature (Kim et al., 2021). It typically affects adolescents, with a median age of diagnosis between 16 and 17 years (Liu et al., 2020). Germline variants in oncogenes such as *Harvey Rat Sarcoma Virus (HRAS)*, are implicated in the development of rhabdomyosarcoma (Kim et al., 2021). Additionally, Castello syndrome, which is characterised by variants in *HRAS*, is a recognized CPS that increases the risk of rhabdomyosarcoma (Table 1) (Pal et al., 2019). Several TSGs have also been associated with rhabdomyosarcoma, including *Neurofibromin1 (NF1)*, which is linked to Neurofibromatosis type 1 (NF1), and further elevates the risk of rhabdomyosarcoma development (Table 1) (Kim et al., 2021).

Osteosarcoma is a bone cancer that affects osteoblast cells (Liu et al., 2020). It most commonly presents in adolescents, with a median diagnostic age of 15 years (Liu et al., 2020). Like retinoblastoma, a study on various sarcomas identified germline variants in the *RB1* TSG to be linked to osteosarcoma (Pal et al., 2019). As a result, individuals with congenital retinoblastoma, caused by *RB1* variants, are at an increased risk of developing osteosarcoma (Table 1) (Pal et al., 2019). In addition, germline variants in TSGs, such as *TP53* and *NF1*, that are associated with rhabdomyosarcoma are also linked to osteosarcoma (Alba-Pavón et al., 2023). Consequently, NF1 is a CPS that is associated with an increased risk of osteosarcoma (Table 1) (Pal et al., 2019).

**Table 1:** Genes most commonly associated with four cancer types and their related CPSs

| Cancer Type      | Gene        | Gene Type | Associated CPS        |
|------------------|-------------|-----------|-----------------------|
| Nephroblastoma   | <i>WT1</i>  | Oncogene  | WGAR                  |
|                  |             |           | DDS                   |
|                  | <i>IGF2</i> |           | BWS                   |
| Retinoblastoma   | <i>RB1</i>  | TSG       | 13q deletion syndrome |
| Rhabdomyosarcoma | <i>HRAS</i> | Oncogene  | Castello syndrome     |
|                  | <i>NF1</i>  | TSG       | NF1                   |
| Osteosarcoma     | <i>RB1</i>  | TSG       | Retinoblastoma        |
|                  | <i>NF1</i>  |           | NF1                   |

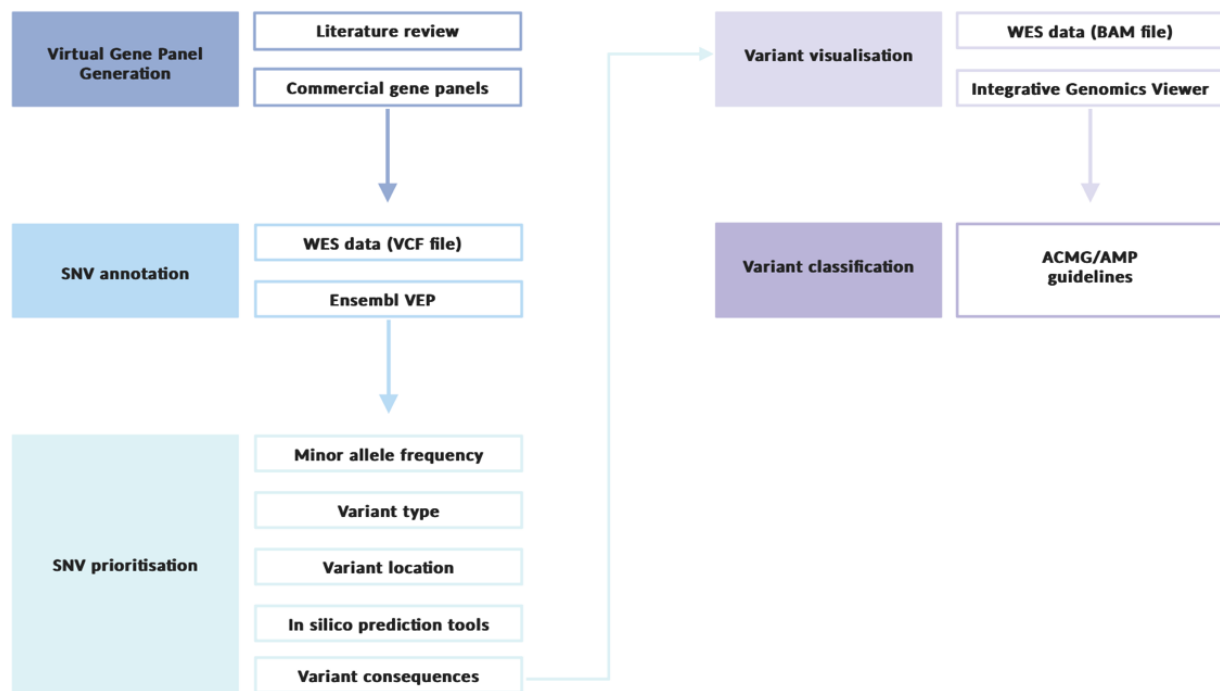
The primary variant types found in cancer predisposition genes are single nucleotide variants (SNVs) and copy number variants (CNVs) (Gröbner et al., 2018). SNVs involve changes to a single nucleotide within the DNA sequence and have been implicated in altering protein structure or function, potentially driving cancer development (Gröbner et al., 2018). SNVs are associated with 93% of nephroblastoma, 58% of retinoblastoma, 68% of rhabdomyosarcoma and 74% of osteosarcoma cases (Singh et al., 2016; Xu et al., 2018; Gröbner et al., 2018; Czarnecka et al., 2020). CNVs, which include deletions or duplications of DNA segments, can affect gene dosage, disrupting normal cellular processes and facilitating tumour formation (Gröbner et al., 2018). CNVs account for approximately 7% of nephroblastoma, 42% of retinoblastoma, 32% of rhabdomyosarcoma and 26% of osteosarcoma cases (Singh et al., 2016; Xu et al., 2018; Gröbner et al., 2018; Czarnecka et al., 2020).

These variant types are key to understanding the genetic landscape of paediatric cancers, but the full range of genes that predispose children to cancer is only partially understood (Huang et al., 2018). The precise proportion of paediatric malignancies attributable to underlying CPSs remains unclear (Huang et al., 2018). Whole Exome Sequencing (WES) has proven to be effective in identifying germline variants, including SNVs and CNVs, in paediatric cancers (Zhang et al., 2015). Understanding the role of germline variants in paediatric cancer development can help implement familial genetic counselling, to offer screening and management options for individuals with an increased risk (Villani et al., 2023).

Cancer is fundamentally defined by genetic alterations, thus scientific studies investigating its causes, diagnosis and treatment should prioritise genetics and genomics research (Rotimi et al., 2021). Africa's population history reveals that individuals of African ancestry are the most genetically diverse group in the world; however, they remain significantly underrepresented in cancer research, accounting for just 0.016% of global cancer publications (Rotimi et al., 2021). As a result, Variants of Uncertain Significance (VUSs) are more prevalent in this population because of the limited genetic data available, leading to a lack of characterisation for many variants (McDonald and Ricks-Santi, 2022). Therefore, the aim of this study was to analyse WES data and identify germline SNVs in South African paediatric cancer patients with four of the most common types of childhood solid cancers, namely; nephroblastoma, retinoblastoma, rhabdomyosarcoma and osteosarcoma.

## 2. Materials and Methods

This study formed part of a larger study, namely the PeCan project, which is investigating the genetics of paediatric cancers in South Africa. The workflow for this study involved the creation of a virtual gene panel, followed by variant annotation, prioritisation, visualisation and ultimately, variant classification (Figure 2).



**Figure 2:** Schematic overview of the methodology employed in this study. The figure outlines the key steps, including Virtual Gene Panel Generation, SNV annotation, SNV prioritisation, Variant visualisation and Variant classification, with sub-steps that were followed in the research.

### 2.1 Study cohort

A total of 14 patients were included in this study, all whom were recruited through the PeCan project. These patients were mainly recruited from Charlotte Maxeke academic hospital and Chris Hani Baragwanath hospital between 2019 and 2023. The inclusion criteria were children diagnosed with solid tumours that had not previously undergone genetic screening. Four patients were diagnosed with neuroblastoma, retinoblastoma or rhabdomyosarcoma, while two were diagnosed with osteosarcoma (Table 2). With the exception of two patients, all participants are of African ancestry. One patient with rhabdomyosarcoma has an unknown ancestry, and one osteosarcoma patient is of European ancestry (Table 2). The age at diagnosis ranged from three months to 11 years.

**Table 2:** Patient characteristics

| Patient ID | Diagnosed cancer type | Patient sex | Patient ancestry | Age at diagnosis |
|------------|-----------------------|-------------|------------------|------------------|
| PC12       | Nephroblastoma        | Female      | African          | 3 years          |
| PC14       | Nephroblastoma        | Female      | African          | 3 years          |
| PC18       | Nephroblastoma        | Male        | African          | 6 months         |
| PC63       | Nephroblastoma        | Male        | African          | 3 years          |
| PC79       | Retinoblastoma        | Male        | African          | 4 months         |
| PC104      | Retinoblastoma        | Female      | African          | 8 months         |
| PC120      | Retinoblastoma        | Male        | African          | 3 months         |
| PC123      | Retinoblastoma        | Male        | African          | 9 months         |
| PC2        | Rhabdomyosarcoma      | Female      | African          | 8 years          |
| PC11       | Rhabdomyosarcoma      | Female      | African          | 3 years          |
| PC74       | Rhabdomyosarcoma      | Male        | Unknown          | 3 years          |
| PC121      | Rhabdomyosarcoma      | Female      | African          | 3 years          |
| PC58       | Osteosarcoma          | Female      | African          | 6 years          |
| PC108      | Osteosarcoma          | Male        | European         | 11 years         |

## 2.2 Virtual gene panel generation

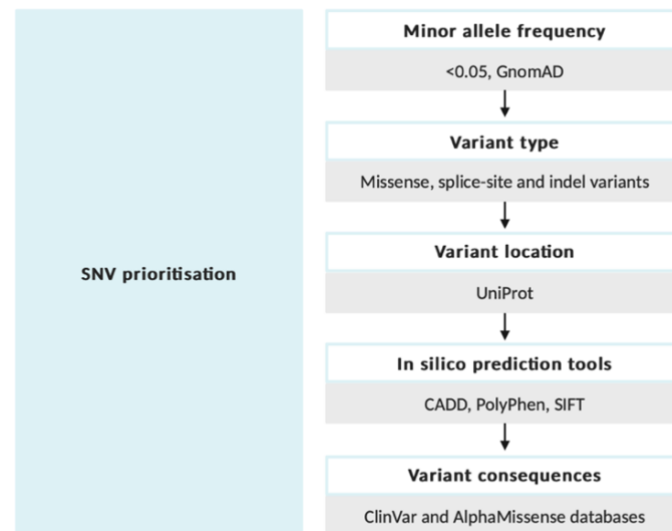
This study involved the creation of a virtual gene panel by identifying genes associated with each of the paediatric cancers through a comprehensive review of relevant paediatric oncology literature. The goal was to identify the 10% most common genes associated with each of the four cancers. Additionally, genes included in established commercial panels, such as the Invitae multi-cancer panel (<https://www.invitae.com>) and Prevention Genetics cancer panels (<https://www.preventiongenetics.com>), were also incorporated. Germline variants in a total of 64 candidate genes were investigated, with the full list of genes provided in Supplementary Table 1 (Appendix A).

## 2.3 Variant annotation and prioritisation

Blood samples were collected from each patient, from which DNA was extracted. WES was then performed using the Ion GeneStudio S5™ Series Sequencer to generate data for variant analysis. Subsequently, each patient's Variant Call Format (VCF) file, that stores gene sequence variations, was

annotated using Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016). Annotation involves assigning functional information to each variant, including its functional consequences, chromosomal locations, and minor allele frequencies (MAFs) (Bartha and Györfy, 2019).

Variants with a global MAF of less than 5% were prioritised based on data from the 1000 Genomes Project. When available, their frequencies were further validated using the Genome Aggregation database (GnomAD v2.1.1) (<https://gnomad.broadinstitute.org/>) (Figure 3). Missense, splice-site and indel variants were further filtered in, since they may impact protein function (Figure 3). The chromosomal location of the variants was validated using UniProt database, to determine whether the variants are located within mutation hot spot regions (<https://www.uniprot.org>) (Figure 3). *In silico* tools, including Combined Annotation Depletion (CADD) (<https://cadd.gs.washington.edu>), Polymorphism Phenotyping (PolyPhen-2) (<https://genetics.bwh.harvard.edu/pph2/>) and Sorting Intolerant from Tolerant (SIFT) (<https://bio.tools/sift>) were used to predict the effect of amino acid substitutions on protein structure and function (Figure 3). The pathogenicity of the variants was assessed using ClinVar and AlphaMissense (AMDB) databases (Figure 3).



**Figure 3:** SNV Prioritisation Workflow. SNVs were filtered and prioritised based on factors including minor allele frequency, variant type, variant location, in silico prediction tools and variant consequences.

#### 2.4 Variant visualisation

The Ion Torrent instrument software, Torrent Suite™ (Thermo Fisher Scientific, USA), was used to retrieve a Binary Alignment Map (BAM) file for each of the 14 patient samples. These BAM files represent DNA sequences that are aligned against the GRCh37 reference sequence and were

visualised using Integrative Genomics Viewer (IGV) (<https://software.broadinstitute.org/software/igv/>). The IGV tool was used to assess variant zygosity, read depth and coverage. Heterozygous variants were identified by an allele ratio of approximately 1:1 between the reference and alternate alleles. A variant was considered true if it had an allele depth exceeding 30x and was absent in the control file.

## 2.5 Variant classification

Variants were classified according to the guidelines set by the American College of Medical Genetics and Genomics and Association of Medical Pathology (ACMG/AMP) (Richards et al., 2015). Using a combination of these ACMG/AMP criteria, variants were assigned to one of five categories: (1) pathogenic, (2) likely pathogenic, (3) uncertain significance, (4) likely benign or (5) benign (Richards et al., 2015).

## 3. Results

The demographic characteristics of the 14 patients in this study are summarised in Table 2. The cohort included four patients diagnosed with nephroblastoma, four with retinoblastoma, four with rhabdomyosarcoma and two with osteosarcoma. The group encompassed seven male and seven female patients. The majority of patients are of African ancestry, and most patients were diagnosed between the ages of one and six years. Detailed patient characteristics are provided in Supplementary Table 2 (Appendix A).

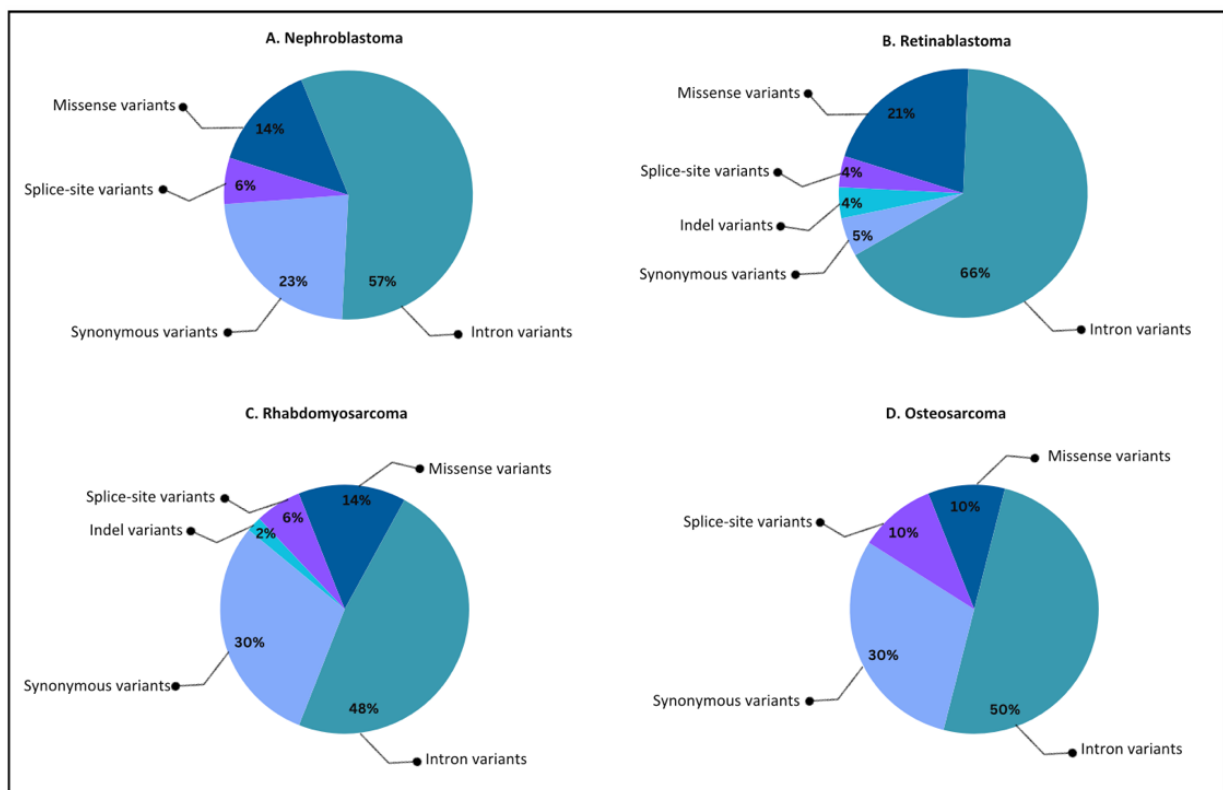
**Table 3:** Demographic characteristics of the study cohort (n=14).

|                  | Nephroblastoma<br>n=4 | Retinoblastoma<br>n=4 | Rhabdomyosarcoma<br>n=4 | Osteosarcoma<br>n=2 | Total<br>n=14 |
|------------------|-----------------------|-----------------------|-------------------------|---------------------|---------------|
| <b>Sex</b>       |                       |                       |                         |                     |               |
| Male             | 2                     | 3                     | 1                       | 1                   | 7             |
| Female           | 2                     | 1                     | 3                       | 1                   | 7             |
| <b>Ethnicity</b> |                       |                       |                         |                     |               |
| African          | 4                     | 4                     | 3                       | 1                   | 12            |
| European         | 0                     | 0                     | 0                       | 1                   | 1             |
| Other            | 0                     | 0                     | 1                       | 0                   | 1             |
| <b>Age</b>       |                       |                       |                         |                     |               |

|      |   |   |   |   |   |
|------|---|---|---|---|---|
| <1   | 1 | 4 | 0 | 0 | 5 |
| 1-6  | 3 | 0 | 3 | 1 | 7 |
| 7-12 | 0 | 0 | 1 | 1 | 2 |

### 3.1 Frequency and types of variants across different paediatric cancers

A total of 212,825 variants within the candidate genes were processed among the 14 patients. The majority of variants were intronic (55%), followed by synonymous (22%), missense (15%), splice-site (6.5%) and indel (1.5%) variants (Figure 4).



**Figure 4:** Distribution of variant types prior to variant prioritisation. **(A)** Nephroblastoma, **(B)** Retinoblastoma, **(C)** Rhabdomyosarcoma and **(D)** Osteosarcoma patients. Variant types include intron variants, synonymous variants, missense variants, splice-site variants and indel variants.

### 3.2 Variants of uncertain significance identified in retinoblastoma and rhabdomyosarcoma patients

No variants were found in the nephroblastoma and osteosarcoma patients based on the filtering criteria. However, two missense variants were identified: one in a retinoblastoma patient (PC79) and another in a rhabdomyosarcoma patient (PC2). In the retinoblastoma patient, the *TP53*

NM\_000546.6:c.923T>C (p.Leu308Pro) variant was identified, while in the rhabdomyosarcoma patient, the *SDHB* NM\_003000.3:c.455C>T (p.Ser152Phe) variant was found (Table 3). Both variants were classified as VUSs according to ACMG/AMP guidelines (Table 3). Detailed information on these two shortlisted variants is provided in Supplementary Table 3 (Appendix A).

**Table 4:** Details of the variants identified in retinoblastoma and rhabdomyosarcoma patients

| Patient ID | Gene        | Refseq      | Coding mutation | Protein change | AF    | Zygosity | IGV coverage | ACMG/AMP codes | ACMG/AMP classification |
|------------|-------------|-------------|-----------------|----------------|-------|----------|--------------|----------------|-------------------------|
| PC79       | <i>TP53</i> | NM_000546.6 | c.923T>C        | p.Leu308Pro    | <0.01 | Het      | 203x         | PM2, PM1, BP4  | VUS                     |
| PC2        | <i>SDHB</i> | NM_003000.3 | c.455C>T        | p.Ser152Phe    | <0.01 | Het      | 118x         | PM2, PM1, BP6  | VUS                     |

*TP53*, Tumour Protein P53; *SDHB*, Succinate Dehydrogenase Subunit B; *RefSeq*, reference sequence; *AF*, allele frequency; *Het*, heterozygous; *ACMG/AMP*, American College of Medical Genetics and Genomics and the Association of Molecular Pathology; *VUS*, Variant of Uncertain Significance.

### 3.2.1 The NM\_000546.6:c.923T>C variant in the *TP53* gene of a retinoblastoma patient

The NM\_000546.6:c.923T>C (p.Leu308Pro) missense variant in the *TP53* gene was identified in a male retinoblastoma patient of African ancestry (PC79), who was diagnosed at four months of age. This variant is absent from the GnomAD database (17-7673605-T-C), indicating a low allele frequency in the general population and suggesting a potential pathogenic role (Table 3). As a result, the ACMG/AMP moderate pathogenic code PM2 was applied. Additionally, this variant is found within a critical and well-established functional domain of the P53\_HUMAN protein (UniProt: P04637), further supporting its potential relevance, leading to the application of the ACMG/AMP moderate pathogenic PM1 code. A deleterious SIFT score of 0 also implies a likely disruptive effect on protein structure and function. While this variant has not been reported on ClinVar, it is classified as benign on AMDB (hg19: 17-7673605-A-G), prompting the application of the ACMG/AMP supporting benign code BP4.

To gain additional insight into this variant, the A:G allele ratio on IGV was determined as 48:52, consistent with a heterozygous genotype (Table 3). Visualisation of this variant with IGV showed an allele depth of 203x (Table 3). Comparison with a control BAM file confirmed the absence of this variant in the control sample, validating that it is a true variant. These IGV results are illustrated in

Supplementary Figure 1 (A) (Appendix A). Based on observations, according to ACMG/AMP guidelines, the missense variant NM\_000546.6:c.923T>C (p.Leu308Pro) in the *TP53* gene is classified as a VUS using the codes PM2, PM1 and BP4 (Table 3).

### 3.2.2 The NM\_003000.3:c.455C>T variant in the *SDHB* gene of a rhabdomyosarcoma patient

The NM\_003000.3:c.455C>T (p.Ser152Phe) missense variant in *SDHB* was identified in a female rhabdomyosarcoma patient of African ancestry (PC2), who was diagnosed at eight years of age. This variant is absent from the GnomAD database (1-17354329-C-T) and exhibits a low allele frequency in the general population, suggesting a potential pathogenic role (Table 3). Consequently, the ACMG/AMP moderate pathogenic code PM2 was applied. This variant lies within a critical and well-established functional domain of the SDHB\_HUMAN protein (UniProt: P21912), further supporting its potential relevance, which led to the application of the ACMG/AMP moderate pathogenic code PM1. While its overall classification remains uncertain significance on ClinVar, a high confidence submitter has classified this variant as likely benign (Variation number: 412492). Additionally, it is listed as ambiguous on AMDB (hg19: 1-17354329-G-A), prompting the application of the ACMG/AMP supporting benign code BP6.

To assess this variant in more detail, an A:G allele ratio of 61:39 was observed on IGV, which deviates slightly from the expected 50:50 ratio typically seen for a heterozygous genotype (Table 3). This deviation could be due to factors such as sequencing depth, allele-specific amplification, or an imbalance in allele representation during sequencing (Zhang et al., 2015). Nonetheless, the ratio is still consistent with a heterozygous genotype (Table 3). Visualisation of this variant on IGV revealed an allele depth of 118x (Table 3). Comparison with a control BAM file confirmed that the variant is absent in the control sample, validating it as a true variant. These IGV findings are presented in Supplementary Figure 1 (B) (Appendix A). Based on results, following ACMG/AMP guidelines, the missense variant NM\_003000.3:c.455C>T (p.Ser152Phe) in the *SDHB* gene is classified as a VUS using codes PM2, PM1, and BP6 (Table 3).

## 4. Discussion

This study employed a virtual gene panel-based approach to identify germline variants in a cohort of 14 paediatric cancer patients diagnosed with neuroblastoma, retinoblastoma, rhabdomyosarcoma

or osteosarcoma. Two germline VUSs were identified, one in the PC79 retinoblastoma patient and another in the PC2 rhabdomyosarcoma patient. A VUS is defined as a genetic variation for which the association of disease risk is unclear (McDonald and Ricks-Santi, 2022). The remaining variants were classified as benign or likely benign.

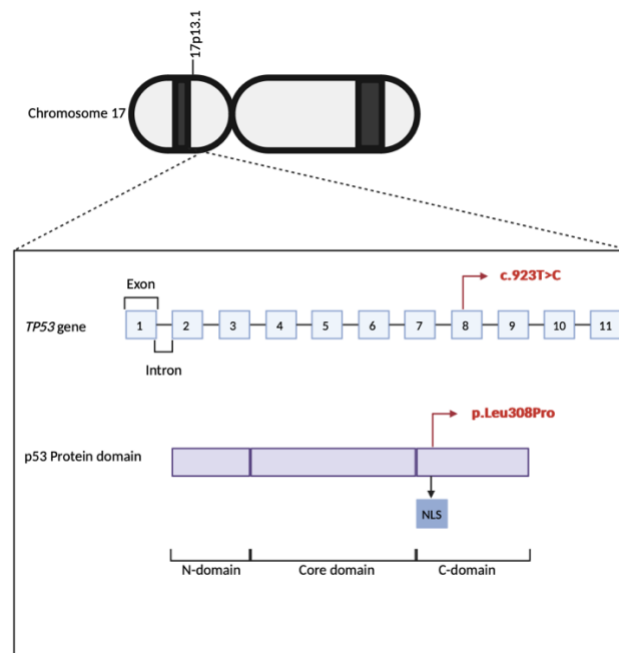
The early ages of diagnosis in this study cohort, ranging from three months to 11 years, suggests limited time for the accumulation of acquired DNA alterations. This implies that inherited germline variants may play a more significant role in the development of these paediatric cancers (Uchimura et al., 2015). This was particularly evident in the cases of the PC79 retinoblastoma and PC2 rhabdomyosarcoma patients. Retinoblastoma typically has a median diagnostic age between one and two years, while rhabdomyosarcoma usually presents in adolescents, with a median age of diagnosis between 16 and 17 years (Capasso et al., 2020; Liu et al., 2020; Parsons et al., 2016). In contrast, the PC79 retinoblastoma patient was diagnosed at four months of age, and the PC2 rhabdomyosarcoma patient at eight years of age, both considerably younger than the expected median age for these cancers.

#### *4.1 The TP53 NM\_000546.6:c.923T>C variant in retinoblastoma patient PC79*

The NM\_000546.6:c.923T>C (p.Leu308Pro) missense VUS identified in the PC79 male retinoblastoma patient, is located within the *TP53* gene on the short arm of chromosome 17 (17p13.1) (Olivier et al., 2010). While this specific variant has not been reported in ClinVar, *TP53* is a well-established TSG that encodes the p53 protein, which is commonly associated with various cancers (Olivier et al., 2010). Given the role of TSGs in regulating cell growth and preventing uncontrolled proliferation, this variant may be linked to an increased risk of retinoblastoma (Davidoff, 2010). According to the two-hit hypothesis, cancers caused by germline variants in TSGs typically require two mutations; one inherited and one acquired (Murphy et al., 2023). In this case, patient PC79 is heterozygous for the NM\_000546.6:c.923T>C germline variant, that predisposed them to develop cancer at an earlier age by inactivating one allele, leaving only one functional copy of *TP53*. A subsequent somatic mutation in the second allele, such as in retina cells, could lead to complete loss of p53 function, potentially contributing to the development of retinoblastoma (Olivier et al., 2010).

The NM\_000546.6:c.923T>C variant is located within exon 8 of *TP53*, specifically within the Nuclear Localisation Signal (NLS) domain of p53 (Figure 5) (Wang et al., 2023). The NLS domain is essential for the nuclear import of the p53 protein, allowing it to regulate the expression of various target genes

involved in controlling the cell cycle, apoptosis and maintaining genomic stability (Wang et al., 2023). Variants in this domain disrupt the nuclear import of p53, which has been reported to cause genomic instability and contribute to cancer progression (Wang et al., 2023).

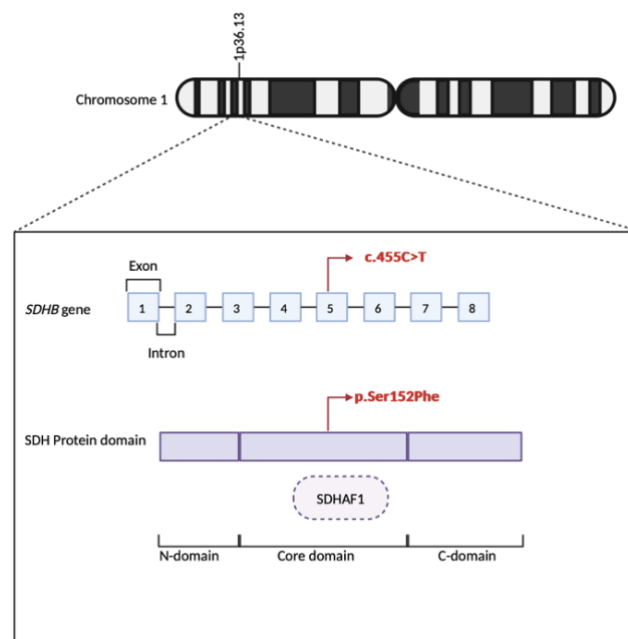


**Figure 5:** Genomic location of the *TP53* NM\_000546.6:c.923T>C (p.Leu308Pro) variant. This variant is located in exon 8 of 11 in the *TP53* gene. It is present within the NLS binding site in the C-domain of the p53 protein.

#### 4.2 The *SDHB* NM\_003000.3:c.455C>T variant in rhabdomyosarcoma patient PC2

The NM\_003000.3:c.455C>T (p.Ser152Phe) missense VUS identified in the PC2 female rhabdomyosarcoma patient is located within the *SDHB* gene on the short arm of chromosome 1 (1p36.13) (Liu et al., 2023). This specific variant has been reported in seven submissions on ClinVar. While one high confidence submitter classified it as likely benign, the overall classification on ClinVar remains as uncertain significance. Notably, none of the submissions have associated this variant with rhabdomyosarcoma. The most recent submitter (Accession number: SCV004202975.2) classified it as uncertain significance regarding its association with gastrointestinal stromal tumours (GISTs), which are tumours that develop in the digestive tract (Jiang et al., 2015). A study by Jiang et al described the possibility of rhabdomyosarcoma differentiation within GISTs, where differentiation refers to the process by which a tumour may adopt characteristics of a different tissue type, such as skeletal muscle tissue (Jiang et al., 2015).

Similarly to *TP53*, the *SDHB* gene is a TSG that encodes the succinate dehydrogenase (SDH) enzyme complex. Patient PC2 is heterozygous for the NM\_003000.3:c.455C>T germline variant, therefore a subsequent somatic mutation in the second allele, such as in skeletal muscle cells, could lead to complete loss of SDH function, potentially contributing to the progression of rhabdomyosarcoma (Liu et al., 2023). This variant is located within the succinate dehydrogenase assembly factor 1 (SDHAF1) region in exon 5 of *SDHB* (Figure 6) (Liu et al., 2023). The SDHAF1 region is crucial for stabilising and assembling the SDH complex by interacting with the core domain of the SDH protein (Liu et al., 2023). This interaction is essential for regulating cellular metabolism and preventing the accumulation of metabolic intermediates like succinate (Liu et al., 2023). Variants in the SDHAF1 region disrupts SDH protein function, resulting in metabolic dysfunction (Liu et al., 2023). These variants have been associated with succinate buildup and the activation of oncogenic pathways that contribute to cancer progression (Liu et al., 2023).



**Figure 6:** Genomic location of the *SDHB* NM\_003000.3:c.455C>T (p.Ser152Phe) variant. This variant is located in exon 5 of 8 in the *SDHB* gene. It is present within the SDHAF1 region, which primarily affects the core domain of the SDH protein.

#### 4.3 Variants of Uncertain Significance

The classification of two VUSs, in the PC79 retinoblastoma and PC2 rhabdomyosarcoma patients of African ancestry, was primarily due to conflicting criteria or insufficient evidence being available. Notably, as the majority of this cohort (12/ 14 patients) is of African ancestry, it is important to

recognise that individuals of African descent are more likely to receive inconclusive genetic test results (McDonald and Ricks-Santi, 2022). This is due to a higher prevalence of VUSs in this population, which is attributed to the greater genetic diversity in this group (McDonald and Ricks-Santi, 2022).

The absence of clinically significant variants, classified as pathogenic or likely pathogenic, within cancer predisposition genes commonly associated with these four paediatric cancers in European populations is particularly striking. For instance, variants in *WT1* (nephroblastoma), *RB1* (retinoblastoma), *HRAS* (rhabdomyosarcoma) and *NF1* (osteosarcoma), were absent from this study cohort (Capasso et al., 2020). Subsequently, CPSs such as WGAR, DDS, BWS, 13q deletion syndrome, Castello syndrome, and NF1 were also not identified (Steliarova-Foucher et al., 2017). While this could be attributed to the small cohort size, it may also reflect the underrepresentation of African genetic data in cancer research (McDonald and Ricks-Santi, 2022).

These findings underscore the need to expand variant profile data for common paediatric cancers within the diverse African context. WES offers a valuable approach for this, as it enables the discovery of mutations in genes not previously associated with paediatric cancers (Zhang et al., 2015). In resource constrained African settings, WES is more cost-effective than Whole Genome Sequencing (WGS) and provides broader gene coverage than targeted panels (Zhang et al., 2015).

#### *4.4 Limitations*

The low representation of African data in public genetic databases may have contributed to the identification of VUSs and impeded the identification of potentially relevant variants (McDonald and Ricks-Santi, 2022). Additionally, technical difficulties prevented the use of Torrent Suite™ Software for CNV identification. This limited our ability to identify potential CNVs that may be present in this study cohort and contributed to paediatric cancer development.

#### *4.5 Future work*

None of the VUSs identified in this study were determined to be ‘hot’, which are those where sufficient evidence could potentially reclassify the variant as pathogenic. Given that nephroblastoma, retinoblastoma, rhabdomyosarcoma and osteosarcoma collectively account for approximately 19% of all paediatric cancers globally, ascertaining a genetic cause for these common paediatric cancers in

this South African study cohort is crucial for characterising their genetic profile (Wu et al., 2022). Future analysis of each patient's BAM file that has been generated using Torrent Suite software could detect clinically significant CNVs, particularly in cancers like retinoblastoma, rhabdomyosarcoma and osteosarcoma, where CNVs contribute up to 42%, 32% and 26% of cases respectively (Gröbner et al., 2018). This approach could also aid in identifying the proportion of these paediatric malignancies possibly linked to underlying CPSs, such as 13q deletion syndrome, which increases the risk of retinoblastoma due to a *RB1* gene deletion (Capasso et al., 2020). Additionally, it would be valuable to investigate genes outside of the current virtual gene panel. Identifying clinically significant variants in this study cohort would facilitate genetic counselling and provide screening and management options for at-risk family members, ultimately improving patient outcomes (Villani et al., 2023).

## 5. Conclusion

This study highlights the potential role of inherited germline variants in cancer predisposition genes, particularly TSGs, in the development of early onset of paediatric cancers. While the identification of two VUSs in key TSGs, *TP53* and *SDHB*, provides insight into the possible germline genetic causes of these cancers, the lack of definitive pathogenic variants underscores the need for further research. The findings also point to the challenges of interpreting genetic data in populations with limited representation in global databases, such as those of diverse African descent, where VUSs are common. Moreover, the absence of variants in genes typically associated with these cancers in European populations, such as *WT1*, *RB1*, *HRAS* and *NF1*, suggests that genetic predispositions in this South African study cohort may differ. Despite technical limitations, the use of WES proved valuable in uncovering potential genetic contributors to these paediatric cancers, by generating a large amount of variants for analyses. Future studies should focus on integrating CNVs and expanding the candidate gene list to refine our understanding of the genetic basis of these cancers. This approach will also aid in improving risk assessments and facilitate genetic counselling for the affected families of this study cohort.

## **Acknowledgement**

Dr Phelelani T. Mpangase assisted Ms L Naudé with creating a Virtual Gene Panel with bioinformatics. Prof. A. Krause, the head of division of Human Genetics. Dr N. Mabaso, a medical geneticist at the National Health Laboratory Services, Johannesburg. Prof. J. Geel and Prof. G. Naidu for the recruitment of patients.

## **Statement of Ethics**

This sub-study protocol was reviewed and approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand (approval number M240759), under the parent study (approval number M240262). The ethical clearance certificate is provided in Appendix B.

## **Conflict of Interest**

The authors have no conflicts of interest to declare.

## **Funding Sources**

Funding for the pre-generated WES data for this project has been made available by the; PeCan Project, National Research Foundation, South African Medical Research Council, Female Academic Leaders Fellowship, National Health Laboratory Service, and the University of the Witwatersrand Faculty of Health Science.

## **Author Contributions**

Dr L Lamola designed the study. Ms L Naudé, Dr L Lamola and Dr T Ngcungcu contributed to acquisition and interpretation of the data. Ms L Naudé analysed the data. Ms L Naudé drafted the manuscript. Ms L Naudé, Dr L Lamola and Dr T Ngcungcu revised the manuscript for important intellectual content. Dr L Lamola and Dr T Ngcungcu supervised the research.

## **Data Availability Statement**

Data available upon request.

## References

- Alba-Pavón, P., Alaña, L., Gutierrez-Jimeno, M., García-Obregón, S., Imízcoz, T., Panizo, E., et al. (2023). Identification of germline cancer predisposition variants in pediatric sarcoma patients from somatic tumor testing. *Sci Rep* 13. <https://doi.org/10.1038/s41598-023-29982-2>
- Baldovini, C., Rossi, G., Ciarrocchi, A. (2019). Approaches to tumor classification in pulmonary sarcomatoid carcinoma. *Lung Cancer: Targets and Therapy*. <https://doi.org/10.2147/LCTT.S186779>
- Bartha, Á., Gyórfy, B. (2019). Comprehensive outline of whole exome sequencing data analysis tools available in clinical oncology. *Cancers (Basel)*. <https://doi.org/10.3390/cancers11111725>
- Burdick, K.J., Cogan, J.D., Rives, L.C., Robertson, A.K., Koziura, M.E., Brokamp, E., et al. (2020). Limitations of exome sequencing in detecting rare and undiagnosed diseases. *Am J Med Genet A* 182, 1400–1406. <https://doi.org/10.1002/ajmg.a.61558>
- Capasso, M., Montella, A., Tirelli, M., Maiorino, T., Cantalupo, S., Iolascon, A. (2020). Genetic Predisposition to Solid Pediatric Cancers. *Front Oncol*. <https://doi.org/10.3389/fonc.2020.590033>
- Czarnecka, A.M., Synoradzki, K., Firlej, W., Bartnik, E., Sobczuk, P., Fiedorowicz, M., et al. (2020). Molecular biology of osteosarcoma. *Cancers (Basel)*. <https://doi.org/10.3390/cancers12082130>
- Davidoff, A.M. (2010). Pediatric oncology. *Semin Pediatr Surg* 19, 225–233. <https://doi.org/10.1053/j.sempedsurg.2010.03.007>
- Foong, J., Girdea, M., Stavropoulos, J., Brudno, M. (2015). Prioritizing clinically relevant copy number variation from genetic interactions and gene function data. *PLoS One* 10. <https://doi.org/10.1371/journal.pone.0139656>
- Gargallo, P., Oltra, S., Yáñez, Y., Juan-Ribelles, A., Calabria, I., Segura, V., et al. (2021). Germline predisposition to pediatric cancer, from next generation sequencing to medical care. *Cancers (Basel)* 13. <https://doi.org/10.3390/cancers13215339>
- Gröbner, S.N., Worst, B.C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva, V.A., et al. (2018). The landscape of genomic alterations across childhood cancers. *Nature* 555, 321–327. <https://doi.org/10.1038/nature25480>

- Guernier-Cambert, V., Chamings, A., Collier, F., Alexandersen, S. (2021). Diverse Bacterial Resistance Genes Detected in Fecal Samples From Clinically Healthy Women and Infants in Australia—A Descriptive Pilot Study. *Front Microbiol* 12. <https://doi.org/10.3389/fmicb.2021.596984>
- Huang, K., Mashl, R.J., Wu, Y., Ritter, D.I., Wang, J., Oh, C., et al. (2018). Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell* 173, 355-370.e14. <https://doi.org/10.1016/j.cell.2018.03.039>
- Jiang, X., Anderson, H.B., Guy, C.D., Mosca, P.J., Riedel, R.F., Cardona, D.M. (2015). Rhabdomyosarcomatous Transformation of a Gastrointestinal Stromal Tumor following Treatment with Imatinib. *Case Rep Oncol Med* 2015, 1–5. <https://doi.org/10.1155/2015/317493>
- Kim, J., Light, N., Subasri, V., Young, E.L., Wegman-Ostrosky, T., Barkauskas, D.A., et al. (2021). Pathogenic Germline Variants in Cancer Susceptibility Genes in Children and Young Adults With Rhabdomyosarcoma. <https://doi.org/10.1200/PO.20>
- Knudson, A.G. (1971). Mutation and Cancer: Statistical Study of Retinoblastoma.
- Kuhlen, M., Taeubner, J., Brozou, T., Wieczorek, D., Siebert, R., Borkhardt, A. (2019). Family-based germline sequencing in children with cancer. *Oncogene*. <https://doi.org/10.1038/s41388-018-0520-9>
- Liu, C., Zhou, D., Yang, K., Xu, N., Peng, J., Zhu, Z. (2023). Research progress on the pathogenesis of the SDHB mutation and related diseases. *Biomedicine and Pharmacotherapy*. <https://doi.org/10.1016/j.biopha.2023.115500>
- Liu, J., Wu, S., Xie, X., Wang, Z., Lei, Q. (2020). Identification of potential crucial genes and key pathways in osteosarcoma. *Hereditas* 157. <https://doi.org/10.1186/s41065-020-00142-0>
- McDonald, J.T., Ricks-Santi, L.J. (2022). Hereditary variants of unknown significance in African American women with breast cancer. *PLoS One* 17. <https://doi.org/10.1371/journal.pone.0273835>
- Murphy, A.J., Cheng, C., Williams, J., Shaw, T.I., Pinto, E.M., Dieseldorff-Jones, K., et al. (2023). Genetic and epigenetic features of bilateral Wilms tumor predisposition in patients from the Children's Oncology Group AREN18B5-Q. *Nat Commun* 14. <https://doi.org/10.1038/s41467-023-43730-0>
- Nicora, G., Zucca, S., Limongelli, I., Bellazzi, R., Magni, P. (2022). A machine learning approach based on ACMG/AMP guidelines for genomic variant classification and prioritization. *Sci Rep* 12. <https://doi.org/10.1038/s41598-022-06547-3>
- Olivier, M., Hollstein, M., Hainaut, P. (2010). TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb Perspect Biol*. <https://doi.org/10.1101/cshperspect.a001008>

- Pal, A., Chiu, H.Y., Taneja, R. (2019). Genetics, epigenetics and redox homeostasis in rhabdomyosarcoma: Emerging targets and therapeutics. *Redox Biol.* <https://doi.org/10.1016/j.redox.2019.101124>
- Pandey, A.N. (2014). Retinoblastoma: An overview. *Saudi Journal of Ophthalmology.* <https://doi.org/10.1016/j.sjopt.2013.11.001>
- Parsons, D.W., Roy, A., Yang, Y., Wang, T., Scollon, S., Bergstrom, K., et al. (2016). Diagnostic yield of clinical tumor and germline whole-exome sequencing for children with solid tumors. *JAMA Oncol* 2, 616–624. <https://doi.org/10.1001/jamaoncol.2015.5699>
- Ren, H.M., Liao, M.Q., Tan, S.X., Cheng, C., Zhu, S., Zheng, L., et al. (2019). Analysis of the Global Burden of Disease Study 2019. *Front Public Health* 10. <https://doi.org/10.3389/fpubh.2022.910641>
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., 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, 405–424. <https://doi.org/10.1038/gim.2015.30>
- Rotimi, S.O., Rotimi, O.A., Salhia, B. (2021). A Review of Cancer Genetics and Genomics Studies in Africa. *Front Oncol.* <https://doi.org/10.3389/fonc.2020.606400>
- Ruteshouser, E.C., Robinson, S.M., Huff, V. (2008). Wilms tumor genetics: Mutations in WT1, WTX, and CTNNB1 account for only about one-third of tumors. *Genes Chromosomes Cancer* 47, 461–470. <https://doi.org/10.1002/gcc.20553>
- Seo, E.S., Lee, J.W., Lim, J., Shin, S., Cho, H.W., Ju, H.Y., et al. (2024). Germline functional variants contribute to somatic mutation and outcomes in neuroblastoma. *Nat Commun* 15, 8360. <https://doi.org/10.1038/s41467-024-52128-5>
- Singh, J., Mishra, A., Jayamuruga Pandian, A., Mallipatna, A.C., Khetan, V., Sripriya, S., et al. (2016). Next-generation sequencing-based method shows increased mutation detection sensitivity in an Indian retinoblastoma cohort.
- Steliarova-Foucher, E., Colombet, M., Ries, L., Moreno, F., Dolya, A., Bray, F., et al. (2017). International incidence of childhood cancer, 2001–10: a population-based registry study. *Lancet Oncol* 18, 719–731. [https://doi.org/10.1016/S1470-2045\(17\)30186-9](https://doi.org/10.1016/S1470-2045(17)30186-9)
- Uchimura, A., Higuchi, M., Minakuchi, Y., Ohno, M., Toyoda, A., Fujiyama, A., et al. (2015). Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type

laboratory mice and mutator mice. *Genome Res* 25, 1125–1134.  
<https://doi.org/10.1101/gr.186148.114>

Villani, A., Davidson, S., Kanwar, N., Lo, W.W., Li, Y., Cohen-Gogo, S., et al. (2023). The clinical utility of integrative genomics in childhood cancer extends beyond targetable mutations. *Nat Cancer* 4, 203–221. <https://doi.org/10.1038/s43018-022-00474-y>

Wang, H., Guo, M., Wei, H., Chen, Y. (2023). Targeting p53 pathways: mechanisms, structures, and advances in therapy. *Signal Transduct Target Ther*. <https://doi.org/10.1038/s41392-023-01347-1>

Wu, Y., Deng, Y., Wei, B., Xiang, D., Hu, J., Zhao, P., et al. (2022). Global, regional, and national childhood cancer burden, 1990–2019: An analysis based on the Global Burden of Disease Study 2019. *J Adv Res* 40, 233–247. <https://doi.org/10.1016/j.jare.2022.06.001>

Xu, L., Zheng, Y., Liu, J., Rakheja, D., Singleterry, S., Laetsch, T.W., et al. (2018). Integrative Bayesian Analysis Identifies Rhabdomyosarcoma Disease Genes. *Cell Rep* 24, 238–251. <https://doi.org/10.1016/j.celrep.2018.06.006>

Zhang, J., Walsh, M.F., Wu, G., Edmonson, M.N., Gruber, T.A., Easton, J., et al. (2015). Germline Mutations in Predisposition Genes in Pediatric Cancer. *New England Journal of Medicine* 373, 2336–2346. <https://doi.org/10.1056/nejmoa1508054>

Zhao, L., Liu, H., Yuan, X., Gao, K., Duan, J. (2020). Comparative study of whole exome sequencing-based copy number variation detection tools. *BMC Bioinformatics* 21. <https://doi.org/10.1186/s12859-020-3421-1>

## Appendices

### Appendix A: Supplementary Tables and Figures

|                                                                         |    |
|-------------------------------------------------------------------------|----|
| Supplementary Table 1: Virtual Gene Panel (n= 64).....                  | 39 |
| Supplementary Table 2: Details of the shortlisted variants .....        | 39 |
| Supplementary Figure 1:. Variant visualisation using the IGV tool ..... | 40 |

### Appendix B: Ethics Clearance Certificate

### Appendix C: Approved Research Protocol

### Appendix D: Turnitin Report and Plagiarism Declaration

### Appendix E: Author Guidelines for *Frontiers*

## Appendix A: Supplementary Tables and Figures

**Supplementary Table 1:** Virtual Gene Panel (n= 64)

*AMER1, APC, ASXL1, BLM, BRCA2, BUB1B, CBL, CDC73, CDKN1A, CDKN1C, CDKN2A, CHEK2, CTR9, DICER1, DGCR8, DIS3L2, DROSHA, EPCAM, EXT1, EXT2, FH, FBXW7, FGFR4, GPC3, HACE1, HRAS, IGF2, KDM3B, KIT, MLH1, MDM2, MDM4, MED4, MSH2, MSH6, NBN, NF1, NF2, NYNRIN, PALB2, PDGFRA, PIK3CA, PMS2, POT1, PRKAR1A, PTCH1, POU6F2, RECQL4, REST, RB1, SDHA, **SDHB**, SDHC, SDHD, SMARCA4, SMARCB1, SUFU, **TP53**, TRIM28, TRIM37, TRIP13, WRN, XPO5, WT1*

**Supplementary Table 2:** Details of the shortlisted variants

| Patient ID           | PC79                                                                                                                                                                                    | PC2                                                                                                                                                                    |
|----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Cancer type          | Retinoblastoma                                                                                                                                                                          | Rhabdomyosarcoma                                                                                                                                                       |
| Gene                 | <i>TP53</i>                                                                                                                                                                             | <i>SDHB</i>                                                                                                                                                            |
| RefSeq               | NM_000546.6                                                                                                                                                                             | NM_003000.3                                                                                                                                                            |
| Coding mutation      | c.923T>C                                                                                                                                                                                | c.455C>T                                                                                                                                                               |
| Protein change       | p.Leu308Pro                                                                                                                                                                             | p.Ser152Phe                                                                                                                                                            |
| GnomAD AF            | - (<0.01)                                                                                                                                                                               | - (<0.01)                                                                                                                                                              |
| UniProt location     | P53_HUMAN                                                                                                                                                                               | SDHB_HUMAN                                                                                                                                                             |
| CADD/ PolyPhen/ SIFT | -/ -/ deleterious (0)                                                                                                                                                                   | -/ benign (0.048)/<br>deleterious (0.01)                                                                                                                               |
| ClinVar/ AMDB        | -/ benign                                                                                                                                                                               | Uncertain significance/<br>Ambiguous                                                                                                                                   |
| Location             | 17:7673605                                                                                                                                                                              | 1:17354329                                                                                                                                                             |
| Zygosity             | Heterozygous                                                                                                                                                                            | Heterozygous                                                                                                                                                           |
| IGV coverage         | 203x                                                                                                                                                                                    | 118x                                                                                                                                                                   |
| ACMG/AMP criteria    | PM2_Moderate:<br>Absent from controls.<br>PM1_Moderate:<br>Located in a critical and<br>well-established<br>functional domain.<br>BP4_Supporting:<br>Multiple lines of<br>computational | PM2_Moderate:<br>Absent from controls.<br>PM1_Moderate: Located<br>in a critical and well-<br>established functional<br>domain.<br>BP6_Supporting:<br>Reputable source |

|                          |                             |                                     |
|--------------------------|-----------------------------|-------------------------------------|
|                          | evidence suggest no impact. | recently reports variant as benign. |
| ACMG/ AMP classification | VUS                         | VUS                                 |

RefSeq, reference sequence; ACMG/ AMP, American College of Medical Genetics and Genomics and the Association of Molecular Pathology; *TP53*, *Tumour Protein P53*; *SDHB*, *Succinate Dehydrogenase Subunit B*; VUS, Variant of Uncertain Significance.

### Supplementary Figure 1: Variant visualisation using the IGV tool



**(A)** IGV results for chromosomal coordinate chr 17:7673605. The first panel shows the control BAM file where this variant is absent. The second panel indicates the missense variant NM\_000546.6:c.923T>C (p.Leu308Pro) identified in retinoblastoma patient PC79. **(B)** IGV results for chromosomal coordinate chr1:17354329. The first panel indicates the missense variant NM\_003000.3:c.455C>T (p.Ser152Phe) identified in rhabdomyosarcoma patient PC2. The second panel shows the control BAM file where this variant is absent.

## Appendix B: Ethics Clearance Certificate

|                                                                                                                                          |                                              |
|------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------|
| <br>UNIVERSITY OF THE<br>WITWATERSRAND,<br>JOHANNESBURG | HUMAN RESEARCH ETHICS<br>COMMITTEE (MEDICAL) |
|------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------|

Office of the Deputy Vice-Chancellor (Research and Innovation)

**TO:** Ms L Naude  
School of Pathology  
Division of Human Genetics  
Faculty of Health Sciences

E-mail: [2740639@students.wits.ac.za](mailto:2740639@students.wits.ac.za)

**CC:** Supervisor: Drs L Lamola and T Ngcungcu  
[Lindie.Lamola@wits.ac.za](mailto:Lindie.Lamola@wits.ac.za)  
and <HREC-Medical Research Office@wits.ac.za>

**FROM:** Mr Iain Burns  
Human Research Ethics Committee (Medical)  
Tel: 011 717 1252

E-mail: [Iain.Burns@wits.ac.za](mailto:Iain.Burns@wits.ac.za)

**DATE:** 20 August 2024

**REF:** R14/49

**PROTOCOL NO:** **M240759** (This is your ethics application reference number. Please quote it in all enquiries, oral or written, relating to this study.)

**PROJECT TITLE:** *Identification of germline variants in a cohort of South African paediatric cancer patients*

Please find attached the Clearance Certificate for the above project. I hope it goes well and that an article in a recognized publication comes out of it. This will reflect well on your professional standing and contribute to Government funding of the University.



MSWorks2000/Iain0007/Clearscan.wps

## Appendix C: Approved Research Protocol



Private Bag 3 Wits, 2050  
Fax: 027117172119  
Tel: 02711 7172076

Reference: Mrs Sandra Benn  
E-mail: [sandra.benn@wits.ac.za](mailto:sandra.benn@wits.ac.za)

10 July 2024  
Person No: 2740639  
PAG

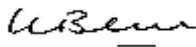
Miss L Naude  
6 Keyes Avenue  
Rosebank  
2196  
South Africa

Dear Miss Lana Naude

### **Master of Science in Medicine: Approval of Title**

We have pleasure in advising that your proposal entitled *Identification of germline variants in a cohort of South African paediatric cancer patients* has been approved. Please note that any amendments to this title have to be endorsed by the Faculty's higher degrees committee and formally approved.

Yours sincerely

A handwritten signature in black ink, appearing to read 'Sandra Benn'.

Mrs Sandra Benn  
Faculty Registrar  
Faculty of Health Sciences





**University of the Witwatersrand Student Ethics Declaration Form**

**(To be completed during the protocol assessor meeting)**

**Background**

All Research conducted by a University of the Witwatersrand student, with human subjects or animals, requires approval by the Wits Human Research Ethics Committee or Animal Research Ethics Committee, respectively.

If research has been undertaken without the necessary ethics approvals, this is considered an ethics violation. This will be reported to the relevant structures, the data will have to be discarded, and in the case of students, they cannot use the data towards their degree.

To prevent any ethics violations, the ethics requirements for the proposed project will be discussed with you at the protocol assessment.

**Declaration**

Based on the current protocol assessment (and any proposed changes suggested by the assessor committee), we, the undersigned, understand that the proposed research requires:

1. Human Research Ethics clearance certificate

a. Covered under existing supervisor ethics

b. Requires a new HREC application

2. Animal Research Ethics clearance certificate

3. No Human or Animal Ethics Clearance

4. Unclear, will seek appropriate guidance from the HREC/AREC committees (whichever relevant)

|                                      |                                     |
|--------------------------------------|-------------------------------------|
| <input checked="" type="radio"/> Yes | <input type="radio"/> No            |
| <input type="radio"/> Yes            | <input checked="" type="radio"/> No |
| <input checked="" type="radio"/> Yes | <input type="radio"/> No            |
| <input type="radio"/> Yes            | <input type="radio"/> No            |
| <input type="radio"/> Yes            | <input type="radio"/> No            |
| <input type="radio"/> Yes            | <input type="radio"/> No            |

**Signatures**

Supervisor/s: T. Ngweni

Student: [Signature]

[Signature]

Date: 22/04/2024

11 March 2019/MP

|                                                                                                                                |               |                                        |            |                    |         |
|--------------------------------------------------------------------------------------------------------------------------------|---------------|----------------------------------------|------------|--------------------|---------|
| CANDIDATE'S SURNAME:<br>[Please print]                                                                                         | Naudé         | FIRST NAME/S:                          | Lana       | STUDENT NUMBER:    | 2740639 |
| CURRENT QUALIFICATIONS: BSc(UNIVERSITY OF PRETORIA), BHSc (Hons) (UNIVERSITY OF THE WITWATERSRAND)                             |               |                                        |            |                    |         |
| TEL: 0768108488                                                                                                                | CELL:         | E-MAIL:<br>2740639@students.wits.ac.za | FAX:       |                    |         |
| DEGREE FOR WHICH PROTOCOL IS BEING SUBMITTED: MSc (Med) Genomic Medicine                                                       |               |                                        |            |                    |         |
| PART-TIME OR FULL-TIME: Full-time                                                                                              |               |                                        |            |                    |         |
| FIRST REGISTERED FOR THIS DEGREE:                                                                                              | TERM : 1 Year |                                        | YEAR: 2024 |                    |         |
| DEPARTMENT: Human Genetics                                                                                                     |               |                                        |            |                    |         |
| TITLE OF PROPOSED RESEARCH: Germline SNV and CNV detection through WES data to analyse four paediatric cancers in South Africa |               |                                        |            |                    |         |
| CANDIDATE'S SIGNATURE:                       |               |                                        |            | DATE: 11/ 04/ 2024 |         |
| SUPERVISOR 1 (NAME & SURNAME): Lindie Lamola                                                                                   |               |                                        |            | % Supervision 50%  |         |
| SUPERVISOR'S QUALIFICATIONS BSc, BSc(Med)Hons, MSc, Phd(CAPE TOWN)                                                             |               |                                        |            |                    |         |
| SUPERVISOR'S DEPARTMENT Human Genetics                                                                                         |               |                                        |            |                    |         |
| SUPERVISOR'S ADDRESS / TEL / E-MAIL: 0114899222/ lindie.lamola@wits.ac.za                                                      |               |                                        |            |                    |         |
| SUPERVISOR 2 (NAME & SURNAME): Thandiswa Ngcungcu                                                                              |               |                                        |            | % Supervision 50%  |         |
| SUPERVISOR'S QUALIFICATIONS BSc, BScHons, MSc(Med), PhD(UNIVERSITY OF THE WITWATERSRAND)                                       |               |                                        |            |                    |         |
| SUPERVISOR'S ADDRESS / TEL / E-MAIL: 0114899293/ thandiswa.ngcungcu@wits.ac.za                                                 |               |                                        |            |                    |         |
| SUPERVISOR 3 (NAME & SURNAME):                                                                                                 |               |                                        |            | % Supervision      |         |
| SUPERVISOR'S QUALIFICATIONS                                                                                                    |               |                                        |            |                    |         |
| SUPERVISOR'S ADDRESS / TEL / E-MAIL:                                                                                           |               |                                        |            |                    |         |

**SYNOPSIS OF RESEARCH:** (Brief summary of proposed research project; between 200-300 words only; with sub-headings: an introduction and justification for study, aim/s, proposed methodology and expected outcome/s)  
[Use reverse side of this page if more space is required]

**Introduction:**

Cancer is caused by genetic alterations changing normal cellular growth and functioning due to somatic or germline variants. While somatic variants are linked to sporadic cancers and may be influenced by environmental factors, germline variants are present in reproductive cells that can be inherited. Pediatric cancers are less influenced by environmental factors, suggesting an underestimated proportion may be due to germline variants and cancer predisposition syndromes.

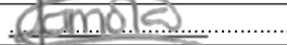
WITS ETHICS NOT REQUIRED:  
WITS ETHICS PENDING:  
WITS ETHICS APPROVED:  
(circle appropriate symbol)\*

|            |    |
|------------|----|
| Yes        | No |
| <u>Yes</u> | No |
| Yes        | No |

IF Y SUPPLY ETHICS  
CLEARANCE CERTIFICATE AS  
ATTACHMENT AND INCLUDE  
ETHICS NUMBER HERE:

**\*Please note the final human ethics clearance certificate or animal ethics certificate must be available prior to starting research**

**As supervisor/s, I/we confirm that I have read the protocol which has been submitted for assessment.**

|                            |                                                                                     |                                                                                       |
|----------------------------|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| SIGNATURE OF SUPERVISOR/S: |  |  |
| SIGNATURE PG OFFICE STAFF  | REGISTERED<br>YES..... NO.....                                                      | STAMP                                                                                 |

**SYNOPSIS OF RESEARCH CONTINUED**

**Study justification:**

Over 80% of the 429 000 pediatric cancers that are annually diagnosed occur in low- and middle-income countries (LMICs). Additionally, the lack of accurate population-based data in LMICs hinders efforts to develop better therapies and improve survival rates. Whole Exome Sequencing (WES) enables the detection of both single nucleotide variants (SNVs) and copy number variants (CNVs).

**Aim:**

The aim of this study is to analyze WES data and identify germline SNVs and CNVs in pediatric cancer patients with four of the most common types of childhood solid cancers, namely: nephroblastoma, retinoblastoma, rhabdomyosarcoma and osteosarcoma.

**Proposed methodology:**

Sixteen patients were selected for this study. Whole Exome Sequencing was performed from blood samples using the Ion GeneStudio S5™ Series Sequencer. Candidate genes associated with these cancers will be identified through scientific literature, whereafter germline SNVs and CNVs will be annotated and prioritized with the use of various bioinformatic tools. These variants will be classified according to American College of Medical Genetics and Genomics and Association of Medical Pathology codes. Binary Alignment Map files will be visualized to assess the variant zygosity and coverage.

**Expected outcomes:**

The advantages of investigating the role of germline variants in childhood cancers includes the prospect of personalized medical treatments, better donor selections for transplants and genetic counselling for families. The knowledge gained in this study will advance our understanding of the germline genetic changes that underlie childhood cancers.

11 March 2019/MP

## **Table of Contents:**

|                                                                     |           |
|---------------------------------------------------------------------|-----------|
| <b>1. Introduction.....</b>                                         | <b>50</b> |
| <b>1.1 Common paediatric malignancies .....</b>                     | <b>50</b> |
| 1.1.1 Genetic characteristics of nephroblastoma .....               | 52        |
| 1.1.2 Genetic characteristics of retinoblastoma .....               | 52        |
| 1.1.3 Genetic characteristics of rhabdomyosarcoma .....             | 52        |
| 1.1.4 genetic characteristics of osteosarcoma .....                 | 53        |
| <b>1.2 Study rationale.....</b>                                     | <b>53</b> |
| <b>1.3 Aim .....</b>                                                | <b>53</b> |
| <b>2. Objectives .....</b>                                          | <b>53</b> |
| <b>3. Methods .....</b>                                             | <b>54</b> |
| <b>3.1 Study cohort.....</b>                                        | <b>54</b> |
| <b>3.2 Identifying candidate genes .....</b>                        | <b>54</b> |
| <b>3.3 Bioinformatic analysis.....</b>                              | <b>54</b> |
| 3.3.1 Single nucleotide variant annotation and prioritisation ..... | 54        |
| 3.3.2 Copy number variant annotation and prioritisation .....       | 54        |
| 3.3.3 Variant visualisation .....                                   | 55        |
| 3.3.4 Theoretical variant classification.....                       | 55        |
| <b>4. Ethics .....</b>                                              | <b>55</b> |
| <b>5. Timing.....</b>                                               | <b>55</b> |
| <b>6. Funding.....</b>                                              | <b>56</b> |
| <b>7. Study challenges .....</b>                                    | <b>56</b> |
| <b>8. Bibliography.....</b>                                         | <b>57</b> |

## **List of Figures:**

|                                                                     |           |
|---------------------------------------------------------------------|-----------|
| <b>Figure 1: The most common paediatric cancers worldwide .....</b> | <b>51</b> |
|---------------------------------------------------------------------|-----------|

### List of Abbreviations:

| Abbreviation | Definition                                                                         |
|--------------|------------------------------------------------------------------------------------|
| ACMG-AMP     | American College of Medical Genetics and Genomics Association of Medical Pathology |
| BAM          | Binary Alignment Map                                                               |
| CADD         | Combined Annotation Depletion                                                      |
| CNV          | Copy number variant                                                                |
| CPS          | Cancer predisposition syndrome                                                     |
| DDS          | Denys-Drash syndrome                                                               |
| GnomAD       | Genome Aggregation Database                                                        |
| HGVS         | Human Genome Variation Society                                                     |
| HMM          | Hidden Markov Model                                                                |
| <i>HRAS</i>  | <i>Harvey Rat Sarcoma Virus</i>                                                    |
| HREC         | University of the Witwatersrand Human Research Ethics Committee                    |
| <i>IGF2</i>  | <i>Insulin-like Growth Factor 2</i>                                                |
| IGV          | Integrative Genomics Viewer                                                        |
| LMICs        | Low- and middle-income countries                                                   |
| LOH          | Loss of heterozygosity                                                             |
| <i>MDM2</i>  | <i>Murine Double Minute 2</i>                                                      |
| <i>MDM4</i>  | <i>Murine Double Minute 4</i>                                                      |
| <i>MED4</i>  | <i>Mediator complex subunit 4</i>                                                  |
| NF1          | Neurofibromatosis Type 1                                                           |
| <i>NF1</i>   | <i>Neurofibromin1</i>                                                              |
| PeCan        | Paediatric Cancer Project                                                          |
| PolyPhen-2   | Polymorphism Phenotyping                                                           |
| <i>RB1</i>   | <i>Retinoblastoma1</i>                                                             |
| SIFT         | Sorting Intolerant from Tolerant                                                   |
| SNV          | Single nucleotide variant                                                          |
| <i>TP53</i>  | <i>Tumour Protein p53</i>                                                          |
| TSG          | Tumour-suppressor genes                                                            |
| UniProt      | Universal Protein Resource                                                         |
| UV           | Ultraviolet                                                                        |
| VCF          | Variant Call Format                                                                |
| VEP          | Variant Effect Predictor                                                           |
| VUS          | Variant of Uncertain Significance                                                  |
| WES          | Whole Exome Sequencing                                                             |
| WGAR         | Wilms tumour-aniridia syndrome                                                     |
| <i>WT1</i>   | <i>Wilms' tumour 1</i>                                                             |

## **1. Introduction:**

Cancer is caused by genetic alterations changing normal cellular growth and functioning due to somatic or germline variants (Kuhlen et al., 2019). Sporadic cancers are not passed on from parent to child and variants arise largely due to environmental factors (Kuhlen et al., 2019). Germline variants are present in reproductive cells, therefore if they are inherited, they will be present in every cell that contains the offspring's DNA. (Uchimura et al., 2015).

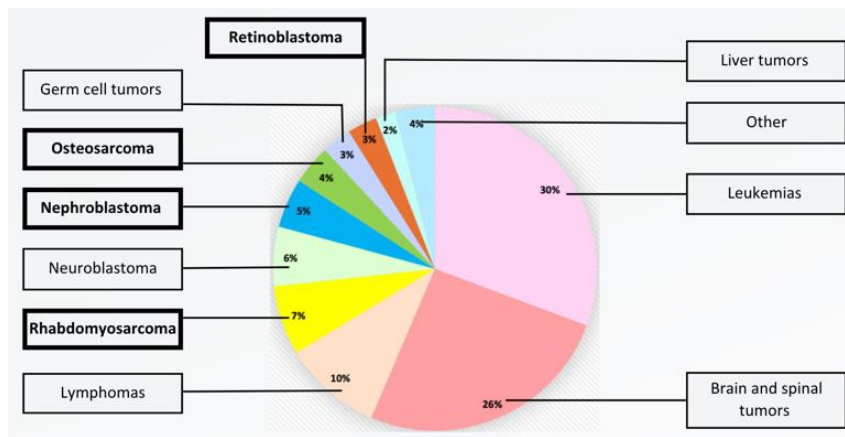
In 1991, Narod *et al.*, reported that only 10% of paediatric cancer cases may be attributed to a germline genetic predisposition (as reviewed in Gargallo et al., 2021). Unlike in adults, environmental factors such as Ultraviolet (UV) exposure, smoking and alcohol consumption are less relevant in children, suggesting an underestimated proportion of paediatric cancers may result from germline variants in cancer predisposition genes (Kuhlen et al., 2019).

Recent studies have additionally highlighted that 8.5% of paediatric cancers may be associated with cancer predisposition syndromes (CPSs), where individuals who carry a mutant variant within a group of genes, called cancer predisposition genes, have an increased susceptibility to develop cancer (Kuhlen et al., 2019). Paediatric cancers often result from variants in tumour-suppressor genes (TSGs) and proto-oncogenes (reviewed in Davidoff, 2010). The normal function of TSGs is to regulate cell growth, and variants in these genes can lead to uncontrolled cellular proliferation and tumour formation (reviewed in Davidoff, 2010). Proto-oncogenes, which regulate normal cell division, can mutate into oncogenes, promoting cancer development by stimulating cell growth and inhibiting cell death (reviewed in Davidoff, 2010).

The full range of variants that predispose children to cancer remains partially understood, and the precise proportion of paediatric malignancies attributable to underlying CPS remains unclear (Gargallo et al., 2021).

### **1.1 Common paediatric malignancies:**

The most common paediatric cancers include leukaemia's, brain and spinal tumours, lymphomas, rhabdomyosarcoma, neuroblastoma, nephroblastoma, osteosarcoma, germ cell tumours and retinoblastoma (Figure 1) (Wu et al., 2022). The focus of the Paediatric Cancer (PeCan) parent study is on solid paediatric tumours that excludes leukaemia's. For this sub-study nephroblastoma, retinoblastoma, rhabdomyosarcoma and osteosarcoma were selected due to them being amongst the most common paediatric cancers observed, accounting for approximately 19% of all cancers seen in children (Figure 1) (Wu et al., 2022).



**Figure 1:** The most common paediatric cancers worldwide. Adapted from Wu et al., 2022

Nephroblastoma is typically characterised by its origin from abnormal kidney development (reviewed in Capasso et al., 2020). Retinoblastoma is an intraocular paediatric cancer, primarily affecting the neural retina (reviewed in Pandey, 2014).

The two-hit hypothesis by Knudson and Strong applies to all cancers that develop due to germline variants in TSGs (Murphy et al., 2023). Heritable cases of nephroblastoma and retinoblastoma involves a pre-zygotic germline variant as a first hit, followed by a second variant post-zygotically, while somatic cases result when both variants occur post-zygotically in a somatic cell (Murphy et al., 2023).

Rhabdomyosarcoma affects skeletal muscle and often results from the inability of skeletal myoblasts to mature (Kim et al., 2021). Osteosarcoma is a bone cancer affecting osteoblast cells (Liu et al., 2020). Since these paediatric cancers exhibit distinct genetic differences, they present unique challenges in diagnosis and treatment options (reviewed in Baldovini et al., 2019).

Single nucleotide variants (SNVs), characterised by alterations in a single nucleotide within the DNA sequence, have been implicated in causing changes in protein structure or function, potentially driving cancer development (Gröbner et al., 2018). Previous studies indicate that SNVs contribute to approximately 93%, 58%, 68% and 74% of Nephroblastoma, Retinoblastoma, Rhabdomyosarcoma and Osteosarcoma cases, respectively (Czarnecka et al., 2020; Gröbner et al., 2018; Singh et al., 2016; Xu et al., 2018).

Conversely copy number variants (CNVs), including deletions or duplications of DNA segments, can modulate changes in gene dosage, disrupting normal cellular processes and facilitating tumour formation (Gröbner et al., 2018). Previous studies suggest that CNVs account for approximately 7%, 42%, 32% and 26% of Nephroblastoma, Retinoblastoma, Rhabdomyosarcoma and Osteosarcoma cases, respectively (Czarnecka et al., 2020; Gröbner et al., 2018; Singh et al., 2016; Xu et al., 2018).

### 1.1.1 Genetic characteristics of nephroblastoma

Pre-zygotic germline variants in bilateral nephroblastoma often involve variants in oncogenes such as *Wilms' tumour 1 (WT1)*. Several CPSs that increases the risk of nephroblastoma development are caused by variants in *WT1*, including Wilms tumour-aniridia syndrome (WGAR) and Denys-Drash syndrome (DDS) (Ruteshouser et al., 2008).

In addition to variants that effect the DNA, epigenetic modifications have also been reported in the aetiology of nephroblastoma including hypermethylation at the 11p15.5 *H19/ICR1* locus or loss of heterozygosity (LOH) at the 11p15.5 locus, which contains the *Insulin-like Growth Factor 2 (IGF2)* oncogene (Murphy et al., 2023). Beckwith-Wiedemann syndrome, characterised by multiple germline variants at the 11p15 locus, is also associated with an increased risk of nephroblastoma (Ruteshouser et al., 2008).

### 1.1.2 Genetic characteristics of retinoblastoma

Germline variants in retinoblastoma are notably associated with germline variants in the *Retinoblastoma1 (RB1)* TSG (reviewed in Capasso et al., 2020). In addition, a CPS impacting the onset of retinoblastoma has been observed in 13q deletion syndrome, characterised by a deletion including the *RB1* gene (reviewed in Capasso et al., 2020).

New studies have indicated that the development of retinoblastoma may be influenced by variants in genetic modifier genes such as *Murine Double Minute 2 (MDM2)*, *Murine Double Minute 4 (MDM4)* or *Mediator complex subunit 4 (MED4)* that alters the expression of the *RB1* gene (reviewed in Capasso et al., 2020).

### 1.1.3 Genetic characteristics of rhabdomyosarcoma

Germline variants in oncogenes such as *Harvey Rat Sarcoma Virus (HRAS)* have been identified in the development of rhabdomyosarcoma (Kim et al., 2021). Additionally, Castello syndrome, characterised by a variant in *HRAS*, is a recognized CPS associated with an elevated risk of rhabdomyosarcoma development (reviewed in Pal et al., 2019). Germline variants in several TSGs associated with rhabdomyosarcoma have also been identified, including *Tumour Protein p53 (TP53)* and *Neurofibromin1 (NF1)* (Kim et al., 2021). Neurofibromatosis type 1 (NF1), linked to deletions in *NF1*, is further associated with an increased risk of rhabdomyosarcoma (Kim et al., 2021).

#### 1.1.4 Genetic characteristics of osteosarcoma

Similarly to retinoblastoma, a study on various sarcomas identified germline variants in the *RB1* TSG to be linked to osteosarcoma (reviewed in Pal et al., 2019). Therefore, patients with congenital retinoblastoma have an increased risk of developing osteosarcoma due to *RB1* variants (reviewed in Pal et al., 2019). Additionally, germline variants in TSGs that are associated with rhabdomyosarcoma including *TP53* and *NF1* are also linked to osteosarcoma (Alba-Pavón et al., 2023). Therefore, *NF1* is a CPSs that is also associated with an increased risk of osteosarcoma (reviewed in Pal et al., 2019).

#### 1.2 Study rationale

The advantages of elucidating the role of germline variants in the development of paediatric cancers includes the implementation of familial genetic counselling, to offer screening and management options for individuals with an increased risk (Villani et al., 2023). Whole Exome Sequencing (WES) has proven to be a valuable tool in identifying germline variants in paediatric cancers, since it enables the detection of both SNVs and CNVs (Zhang et al., 2015).

Since cancer is a genetic disease, scientific studies investigating its causes, diagnosis and treatment should focus more on genetics and genomics to include diverse African populations (reviewed in Rotimi et al., 2021). African individuals are represented in 0.016% of cancer publications globally (reviewed in Rotimi et al., 2021). The goal of this study is to partake in increasing the variant profile data for common paediatric cancers in an African context.

#### 1.3 Aim

The aim of this study is to analyse WES data and identify germline SNVs and CNVs in paediatric cancer patients with four of the most common types of childhood solid cancers, namely; nephroblastoma, retinoblastoma, rhabdomyosarcoma and osteosarcoma.

### 2. Objectives

1. To identify candidate genes associated with the respective cancers
2. To annotate and prioritise SNVs
3. To annotate and prioritise CNVs
4. To visualise variants using Integrative Genomics Viewer (IGV)
5. To classify the variants according to American College of Medical Genetics and Genomics (ACMG) and Association of Medical Pathology (AMP) guidelines

### **3. Methods**

#### **3.1 Study cohort**

This study forms part of a larger study, namely the PeCan project, which is investigating the genetics of paediatric cancers in South Africa. Sixteen patients were selected for this study, that were recruited for the PeCan project, and have not previously undergone genetic screening. Four patients have been diagnosed with each of the four cancers, either nephroblastoma, retinoblastoma, rhabdomyosarcoma or osteosarcoma. All patients, except two, are of African ancestry. One Osteosarcoma patient is of European ancestry, and one Rhabdomyosarcoma patient's ancestry is unknown.

#### **3.2 Identifying candidate genes**

Research on paediatric oncology will be performed with the use of reliable journal articles to write a literature review. The purpose of the review is to identify the 10% most common genes associated with each of the four cancers. These genes will be further be compared and prioritised with genes in commercial cancer gene panels, such as the Invitae multi-cancer panel (<https://www.invitae.com/us/providers/test-catalog/test-01101>). Germline SNVs and CNVs in an approximate 100 candidate genes will be investigated.

#### **3.3 Bioinformatic analysis**

##### **3.3.1 Single nucleotide variant annotation and prioritisation**

After sequencing, SNVs will be annotated using Ensembl Variant Effect Predictor (VEP). Annotation is the process of assigning functional information including the location, functional consequences and minor allele frequencies for each variant (reviewed in Bartha and Györfy, 2019). The Ensembl VEP input data is each patient's Variant Call Format (VCF) file that stores gene sequence variations (reviewed by Bartha and Györfy, 2019). The effect of SNVs could be further predicted using various *in silico* tools such as Combined Annotation Depletion (CADD), Polymorphism Phenotyping (PolyPhen-2) and Sorting Intolerant from Tolerant (SIFT) to prioritise SNVs with a damaging effect on protein function.

##### **3.3.2 Copy number variant annotation and prioritisation**

Primary data analysis for CNV detection and annotation will be performed using Torrent Suite™ Software. Raw sequencing data will be imported and aligned against a reference genome with Torrent Suite™, whereafter this software will be integrated with Ion Reporter™ Software 5.20 to further detect CNVs with the use of a Hidden Markov Model (HMM). A HMM classifies CNVs as either a deletion or duplication based on the sequencing depth of exons (Guernier-Cambert et al., 2021). In addition to

CNV type, Torrent Suite™ Software provides further annotation information such as the size, location and functional impact of CNVs. The effect of CNVs on protein function could be further predicted using *in silico* tools such as CNVkit, cm.MOPS and DECoN (Foong et al., 2015).

### 3.3.3 Variant visualisation

A Binary Alignment Map (BAM) file will be retrieved for each of the 16 study samples. These BAM files will represent DNA sequences that are aligned against the reference sequence and will be visualised using Integrative Genomics Viewer (IGV). The IGV tool will be used to assess the zygosity of the variants and the coverage of the target sequence, to assess whether the variants are true positives.

### 3.3.4 Theoretical variant classification

American College of Medical Genetics and Genomics (ACMG) and Association of Medical Pathology (AMP) codes will be assigned to all the shortlisted variants (Nicora et al., 2022). A combination of these ACMG-AMP codes will be used to classify the pathogenicity of the shortlisted variants (Nicora et al., 2022). The classification of these variants will lastly be evaluated with the use of scientific literature.

## 4 Ethics

The ethics certificate numbers for the PeCan project are M180855 and M240262. An ethics application has been submitted to the University of the Witwatersrand Human Research Ethics Committee (HREC) on the 15<sup>th</sup> of March 2024, to be assessed on the 19<sup>th</sup> of April 2024. This study will commence as soon as ethics approval is obtained.

## 5. Timing

|                       | March | April | May | June | July | August | September | October | November |
|-----------------------|-------|-------|-----|------|------|--------|-----------|---------|----------|
| Literature Review     |       |       |     |      |      |        |           |         |          |
| Ethics application    |       |       |     |      |      |        |           |         |          |
| Protocol preparation  |       |       |     |      |      |        |           |         |          |
| Protocol presentation |       |       |     |      |      |        |           |         |          |
| Protocol assessment   |       |       |     |      |      |        |           |         |          |
| Ethics clearance      |       |       |     |      |      |        |           |         |          |
| Data analysis         |       |       |     |      |      |        |           |         |          |
| Write-up              |       |       |     |      |      |        |           |         |          |

|                       |  |  |  |  |  |  |  |  |  |
|-----------------------|--|--|--|--|--|--|--|--|--|
| Final presentation    |  |  |  |  |  |  |  |  |  |
| Final project hand-in |  |  |  |  |  |  |  |  |  |

## **6. Funding**

Funding for the pre-generated WES data for this project has been made available by the; PeCan Project, National Research Foundation, South African Medical Research Council, Female Academic Leaders Fellowship, National Health Laboratory Service, University of the Witwatersrand Faculty of Health Science. No additional funding is needed for this project.

## **7. Study challenges**

Whole Exome Sequencing enables the sequencing of all protein-coding regions in the human genome, therefore variants that are found in non-coding regions will not be detected with this study design (Burdick et al., 2020). CNV sizes between approximately 10 kb and 100 kb can be detected with WES (Zhao et al., 2020). Another limitation of WES is therefore to detect large CNVs that span across coding and non-coding regions (Burdick et al., 2020). A variant of uncertain significance is when the analysis of a patient's genome identifies a variant with insufficient or conflicting evidence to classify the variant as pathogenic, likely pathogenic, likely benign or benign (Nicora et al., 2022). Due to the quantity and complexity of the genetic information generated by WES, the number of VUS makes it challenging to interpret the results (Nicora et al., 2022).

## **8. Bibliography**

- Alba-Pavón, P., Alaña, L., Gutierrez-Jimeno, M., García-Obregón, S., Imízcoz, T., Panizo, E., González-Urdiales, P., Echebarria-Barona, A., Lopez Almaraz, R., Zaldumbide, L., Astigarraga, I., Patiño-García, A., Villate, O., 2023. Identification of germline cancer predisposition variants in pediatric sarcoma patients from somatic tumor testing. *Sci Rep* 13. <https://doi.org/10.1038/s41598-023-29982-2>
- Baldovini, C., Rossi, G., Ciarrocchi, A., 2019. Approaches to tumor classification in pulmonary sarcomatoid carcinoma. *Lung Cancer: Targets and Therapy*. <https://doi.org/10.2147/LCTT.S186779>
- Bartha, Á., Gyórfy, B., 2019. Comprehensive outline of whole exome sequencing data analysis tools available in clinical oncology. *Cancers (Basel)*. <https://doi.org/10.3390/cancers11111725>
- Burdick, K.J., Cogan, J.D., Rives, L.C., Robertson, A.K., Koziura, M.E., Brokamp, E., Duncan, L., Hannig, V., Pfothenhauer, J., Vanzo, R., Paul, M.S., Bican, A., Morgan, T., Duis, J., Newman, J.H., Hamid, R., Phillips, J.A., 2020. Limitations of exome sequencing in detecting rare and undiagnosed diseases. *Am J Med Genet A* 182, 1400–1406. <https://doi.org/10.1002/ajmg.a.61558>
- Capasso, M., Montella, A., Tirelli, M., Maiorino, T., Cantalupo, S., Iolascon, A., 2020. Genetic Predisposition to Solid Pediatric Cancers. *Front Oncol*. <https://doi.org/10.3389/fonc.2020.590033>
- Czarnecka, A.M., Synoradzki, K., Firlej, W., Bartnik, E., Sobczuk, P., Fiedorowicz, M., Grieb, P., Rutkowski, P., 2020. Molecular biology of osteosarcoma. *Cancers (Basel)*. <https://doi.org/10.3390/cancers12082130>
- Davidoff, A.M., 2010. Pediatric oncology. *Semin Pediatr Surg* 19, 225–233. <https://doi.org/10.1053/j.sempedsurg.2010.03.007>
- Foong, J., Girdea, M., Stavropoulos, J., Brudno, M., 2015. Prioritizing clinically relevant copy number variation from genetic interactions and gene function data. *PLoS One* 10. <https://doi.org/10.1371/journal.pone.0139656>
- Gargallo, P., Oltra, S., Yáñez, Y., Juan-Ribelles, A., Calabria, I., Segura, V., Lázaro, M., Balaguer, J., Tormo, T., Dolz, S., Fernández, J.M., Fuentes, C., Torres, B., Andrés, M., Tasso, M., Castel, V., de Mora, J.F., Cañete, A., 2021. Germline predisposition to pediatric cancer, from next generation sequencing to medical care. *Cancers (Basel)* 13. <https://doi.org/10.3390/cancers13215339>

Gröbner, S.N., Worst, B.C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva, V.A., Johann, P.D., Balasubramanian, G.P., Segura-Wang, M., Brabetz, S., Bender, S., Hutter, B., Sturm, D., Pfaff, E., Hübschmann, D., Zipprich, G., Heinold, M., Eils, J., Lawerenz, C., Erkek, S., Lambo, S., Waszak, S., Blattmann, C., Borkhardt, A., Kuhlen, M., Eggert, A., Fulda, S., Gessler, M., Wegert, J., Kappler, R., Baumhoer, D., Burdach, S., Kirschner-Schwabe, R., Kontny, U., Kulozik, A.E., Lohmann, D., Hettmer, S., Eckert, C., Bielack, S., Nathrath, M., Niemeyer, C., Richter, G.H., Schulte, J., Siebert, R., Westermann, F., Molenaar, J.J., Vassal, G., Witt, H., Zapatka, M., Burkhardt, B., Kratz, C.P., Witt, O., Tilburg, C.M.V., Kramm, C.M., Fleischhack, G., Dirksen, U., Rutkowski, S., Frühwald, M., Hoff, K. Von, Wolf, S., Klingebiel, T., Koscielniak, E., Landgraf, P., Koster, J., Resnick, A.C., Zhang, J., Liu, Y., Zhou, X., Waanders, A.J., Zwijnenburg, D.A., Raman, P., Brors, B., Weber, U.D., Northcott, P.A., Pajtler, K.W., Kool, M., Piro, R.M., Korbel, J.O., Schlesner, M., Eils, R., Jones, D.T.W., Lichter, P., Chavez, L., Pfister, S.M., 2018a. The landscape of genomic alterations across childhood cancers. *Nature* 555, 321–327. <https://doi.org/10.1038/nature25480>

Gröbner, S.N., Worst, B.C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva, V.A., Johann, P.D., Balasubramanian, G.P., Segura-Wang, M., Brabetz, S., Bender, S., Hutter, B., Sturm, D., Pfaff, E., Hübschmann, D., Zipprich, G., Heinold, M., Eils, J., Lawerenz, C., Erkek, S., Lambo, S., Waszak, S., Blattmann, C., Borkhardt, A., Kuhlen, M., Eggert, A., Fulda, S., Gessler, M., Wegert, J., Kappler, R., Baumhoer, D., Burdach, S., Kirschner-Schwabe, R., Kontny, U., Kulozik, A.E., Lohmann, D., Hettmer, S., Eckert, C., Bielack, S., Nathrath, M., Niemeyer, C., Richter, G.H., Schulte, J., Siebert, R., Westermann, F., Molenaar, J.J., Vassal, G., Witt, H., Zapatka, M., Burkhardt, B., Kratz, C.P., Witt, O., Tilburg, C.M.V., Kramm, C.M., Fleischhack, G., Dirksen, U., Rutkowski, S., Frühwald, M., Hoff, K. Von, Wolf, S., Klingebiel, T., Koscielniak, E., Landgraf, P., Koster, J., Resnick, A.C., Zhang, J., Liu, Y., Zhou, X., Waanders, A.J., Zwijnenburg, D.A., Raman, P., Brors, B., Weber, U.D., Northcott, P.A., Pajtler, K.W., Kool, M., Piro, R.M., Korbel, J.O., Schlesner, M., Eils, R., Jones, D.T.W., Lichter, P., Chavez, L., Pfister, S.M., 2018b. The landscape of genomic alterations across childhood cancers. *Nature* 555, 321–327. <https://doi.org/10.1038/nature25480>

Guernier-Cambert, V., Chamings, A., Collier, F., Alexandersen, S., 2021. Diverse Bacterial Resistance Genes Detected in Fecal Samples From Clinically Healthy Women and Infants in Australia—A Descriptive Pilot Study. *Front Microbiol* 12. <https://doi.org/10.3389/fmicb.2021.596984>

Huang, K. lin, Mashl, R.J., Wu, Yige, Ritter, D.I., Wang, Jiayin, Oh, C., Paczkowska, M., Reynolds, S., Wyczalkowski, M.A., Oak, N., Scott, A.D., Krassowski, M., Cherniack, A.D., Houlahan, K.E., Jayasinghe, R., Wang, L.B., Zhou, D.C., Liu, D., Cao, S., Kim, Y.W., Koire, A., McMichael, J.F., Huchtagowder, V., Kim, T.B., Hahn, A., Wang, C., McLellan, M.D., Al-Mulla, F., Johnson, K.J.,

Caesar-Johnson, S.J., Demchok, J.A., Felau, I., Kasapi, M., Ferguson, M.L., Hutter, C.M., Sofia, H.J., Tarnuzzer, R., Wang, Z., Yang, L., Zenklusen, J.C., Zhang, J. (Julia), Chudamani, S., Liu, J., Lolla, L., Naresh, R., Pihl, T., Sun, Q., Wan, Y., Wu, Ye, Cho, J., DeFreitas, T., Frazer, S., Gehlenborg, N., Getz, G., Heiman, D.I., Kim, J., Lawrence, M.S., Lin, P., Meier, S., Noble, M.S., Saksena, G., Voet, D., Zhang, Hailei, Bernard, B., Chambwe, N., Dhankani, V., Knijnenburg, T., Kramer, R., Leinonen, K., Liu, Y., Miller, M., Reynolds, S., Shmulevich, I., Thorsson, V., Zhang, W., Akbani, R., Broom, B.M., Hegde, A.M., Ju, Z., Kanchi, R.S., Korkut, A., Li, J., Liang, H., Ling, S., Liu, W., Lu, Y., Mills, G.B., Ng, K.S., Rao, A., Ryan, M., Wang, Jing, Weinstein, J.N., Zhang, J., Abeshouse, A., Armenia, J., Chakravarty, D., Chatila, W.K., de Bruijn, I., Gao, J., Gross, B.E., Heins, Z.J., Kundra, R., La, K., Ladanyi, M., Luna, A., Nissan, M.G., Ochoa, A., Phillips, S.M., Reznik, E., Sanchez-Vega, F., Sander, C., Schultz, N., Sheridan, R., Sumer, S.O., Sun, Y., Taylor, B.S., Wang, Jioajiao, Zhang, Hongxin, Anur, P., Peto, M., Spellman, P., Benz, C., Stuart, J.M., Wong, C.K., Yau, C., Hayes, D.N., Parker, J.S., Wilkerson, M.D., Ally, A., Balasundaram, M., Bowlby, R., Brooks, D., Carlsen, R., Chuah, E., Dhalla, N., Holt, R., Jones, S.J.M., Kasaian, K., Lee, D., Ma, Y., Marra, M.A., Mayo, M., Moore, R.A., Mungall, A.J., Mungall, K., Robertson, A.G., Sadeghi, S., Schein, J.E., Sipahimalani, P., Tam, A., Thiessen, N., Tse, K., Wong, T., Berger, A.C., Beroukhir, R., Cibulskis, C., Gabriel, S.B., Gao, G.F., Ha, G., Meyerson, M., Schumacher, S.E., Shih, J., Kucherlapati, M.H., Kucherlapati, R.S., Baylin, S., Cope, L., Danilova, L., Bootwalla, M.S., Lai, P.H., Maglinte, D.T., Van Den Berg, D.J., Weisenberger, D.J., Auman, J.T., Balu, S., Bodenheimer, T., Fan, C., Hoadley, K.A., Hoyle, A.P., Jefferys, S.R., Jones, C.D., Meng, S., Mieczkowski, P.A., Mose, L.E., Perou, A.H., Perou, C.M., Roach, J., Shi, Y., Simons, J. V., Skelly, T., Soloway, M.G., Tan, D., Veluvolu, U., Fan, H., Hinoue, T., Laird, P.W., Shen, H., Zhou, W., Bellair, M., Chang, K., Covington, K., Creighton, C.J., Dinh, H., Doddapaneni, H.V., Donehower, L.A., Drummond, J., Gibbs, R.A., Glenn, R., Hale, W., Han, Y., Hu, J., Korchina, V., Lee, S., Lewis, L., Li, W., Liu, X., Morgan, M., Morton, D., Muzny, D., Santibanez, J., Sheth, M., Shinbrot, E., Wang, L., Wang, M., Wheeler, D.A., Xi, L., Zhao, F., Hess, J., Appelbaum, E.L., Bailey, M., Cordes, M.G., Ding, L., Fronick, C.C., Fulton, L.A., Fulton, R.S., Kandoth, C., Mardis, E.R., Miller, C.A., Schmidt, H.K., Wilson, R.K., Crain, D., Curley, E., Gardner, J., Lau, K., Mallery, D., Morris, S., Paulauskis, J., Penny, R., Shelton, C., Shelton, T., Sherman, M., Thompson, E., Yena, P., Bowen, J., Gastier-Foster, J.M., Gerken, M., Leraas, K.M., Lichtenberg, T.M., Ramirez, N.C., Wise, L., Zmuda, E., Corcoran, N., Costello, T., Hovens, C., Carvalho, A.L., de Carvalho, A.C., Fregnani, J.H., Longatto-Filho, A., Reis, R.M., Scapulatempo-Neto, C., Silveira, H.C.S., Vidal, D.O., Burnette, A., Eschbacher, J., Hermes, B., Noss, A., Singh, R., Anderson, M.L., Castro, P.D., Ittmann, M., Huntsman, D., Kohl, B., Le, X., Thorp, R., Andry, C., Duffy, E.R., Lyadov, V., Paklina, O., Setdikova, G., Shabunin, A., Tavobilov, M., McPherson, C., Warnick, R., Berkowitz, R., Cramer, D., Feltmate, C., Horowitz, N., Kibel, A., Muto,

M., Raut, C.P., Malykh, A., Barnholtz-Sloan, J.S., Barrett, W., Devine, K., Fulop, J., Ostrom, Q.T., Shimmel, K., Wolinsky, Y., Sloan, A.E., De Rose, A., Giuliente, F., Goodman, M., Karlan, B.Y., Hagedorn, C.H., Eckman, J., Harr, J., Myers, J., Tucker, K., Zach, L.A., Deyarmin, B., Hu, H., Kvecher, L., Larson, C., Mural, R.J., Somiari, S., Vicha, A., Zelinka, T., Bennett, J., Iacocca, M., Rabeno, B., Swanson, P., Latour, M., Lacombe, L., Têtu, B., Bergeron, A., McGraw, M., Staugaitis, S.M., Chabot, J., Hibshoosh, H., Sepulveda, A., Su, T., Wang, T., Potapova, O., Voronina, O., Desjardins, L., Mariani, O., Roman-Roman, S., Sastre, X., Stern, M.H., Cheng, F., Signoretti, S., Berchuck, A., Bigner, D., Lipp, E., Marks, J., McCall, S., McLendon, R., Secord, A., Sharp, A., Behera, M., Brat, D.J., Chen, A., Delman, K., Force, S., Khuri, F., Magliocca, K., Maithel, S., Olson, J.J., Owonikoko, T., Pickens, A., Ramalingam, S., Shin, D.M., Sica, G., Van Meir, E.G., Zhang, Hongzheng, Eijckenboom, W., Gillis, A., Korpershoek, E., Looijenga, L., Oosterhuis, W., Stoop, H., van Kessel, K.E., Zwarthoff, E.C., Calatozzolo, C., Cuppini, L., Cuzzubbo, S., DiMeco, F., Finocchiaro, G., Mattei, L., Perin, A., Pollo, B., Chen, C., Houck, J., Lohavanichbutr, P., Hartmann, A., Stoehr, C., Stoehr, R., Taubert, H., Wach, S., Wullich, B., Kyler, W., Murawa, D., Wiznerowicz, M., Chung, K., Edenfield, W.J., Martin, J., Baudin, E., Bubley, G., Bueno, R., De Rienzo, A., Richards, W.G., Kalkanis, S., Mikkelsen, T., Noushmehr, H., Scarpace, L., Girard, N., Aymerich, M., Campo, E., Giné, E., Guillermo, A.L., Van Bang, N., Hanh, P.T., Phu, B.D., Tang, Y., Colman, H., Evason, K., Dottino, P.R., Martignetti, J.A., Gabra, H., Juhl, H., Akeredolu, T., Stepa, S., Hoon, D., Ahn, K., Kang, K.J., Beuschlein, F., Breggia, A., Birrer, M., Bell, D., Borad, M., Bryce, A.H., Castle, E., Chandan, V., Cheville, J., Copland, J.A., Farnell, M., Flotte, T., Giama, N., Ho, T., Kendrick, M., Kocher, J.P., Kopp, K., Moser, C., Nagorney, D., O'Brien, D., O'Neill, B.P., Patel, T., Petersen, G., Que, F., Rivera, M., Roberts, L., Smallridge, R., Smyrk, T., Stanton, M., Thompson, R.H., Torbenson, M., Yang, J.D., Zhang, L., Brimo, F., Ajani, J.A., Gonzalez, A.M.A., Behrens, C., Bondaruk, J., Broaddus, R., Czerniak, B., Esmaeli, B., Fujimoto, J., Gershenwald, J., Guo, C., Lazar, A.J., Logothetis, C., Meric-Bernstam, F., Moran, C., Ramondetta, L., Rice, D., Sood, A., Tamboli, P., Thompson, T., Troncoso, P., Tsao, A., Wistuba, I., Carter, C., Haydu, L., Hersey, P., Jakrot, V., Kakavand, H., Kefford, R., Lee, K., Long, G., Mann, G., Quinn, M., Saw, R., Scolyer, R., Shannon, K., Spillane, A., Stretch, J., Synott, M., Thompson, J., Wilmott, J., Al-Ahmadie, H., Chan, T.A., Ghossein, R., Gopalan, A., Levine, D.A., Reuter, V., Singer, S., Singh, B., Tien, N.V., Broudy, T., Mirsaidi, C., Nair, P., Drwiega, P., Miller, J., Smith, J., Zaren, H., Park, J.W., Hung, N.P., Kebebew, E., Linehan, W.M., Metwalli, A.R., Pacak, K., Pinto, P.A., Schiffman, M., Schmidt, L.S., Vocke, C.D., Wentzensen, N., Worrell, R., Yang, H., Moncrieff, M., Goparaju, C., Melamed, J., Pass, H., Botnariuc, N., Caraman, I., Cernat, M., Chemencedji, I., Clipca, A., Doruc, S., Gorincioi, G., Mura, S., Pirtac, M., Stancul, I., Tcaciuc, D., Albert, M., Alexopoulou, I., Arnaout, A., Bartlett, J., Engel, J., Gilbert, S., Parfitt, J., Sekhon, H., Thomas, G., Rassl, D.M., Rintoul, R.C., Bifulco, C.,

Tamakawa, R., Urba, W., Hayward, N., Timmers, H., Antenucci, A., Facciolo, F., Grazi, G., Marino, M., Merola, R., de Krijger, R., Gimenez-Roqueplo, A.P., Piché, A., Chevalier, S., McKercher, G., Birsoy, K., Barnett, G., Brewer, C., Farver, C., Naska, T., Pennell, N.A., Raymond, D., Schilero, C., Smolenski, K., Williams, F., Morrison, C., Borgia, J.A., Liptay, M.J., Pool, M., Seder, C.W., Junker, K., Omberg, L., Dinkin, M., Manikhas, G., Alvaro, D., Bragazzi, M.C., Cardinale, V., Carpino, G., Gaudio, E., Chesla, D., Cottingham, S., Dubina, M., Moiseenko, F., Dhanasekaran, R., Becker, K.F., Janssen, K.P., Slotta-Huspenina, J., Abdel-Rahman, M.H., Aziz, D., Bell, S., Cebulla, C.M., Davis, A., Duell, R., Elder, J.B., Hilty, J., Kumar, B., Lang, J., Lehman, N.L., Mandt, R., Nguyen, P., Pilarski, R., Rai, K., Schoenfeld, L., Senecal, K., Wakely, P., Hansen, P., Lechan, R., Powers, J., Tischler, A., Grizzle, W.E., Sexton, K.C., Kastl, A., Henderson, J., Porten, S., Waldmann, J., Fassnacht, M., Asa, S.L., Schadendorf, D., Couce, M., Graefen, M., Huland, H., Sauter, G., Schlomm, T., Simon, R., Tennstedt, P., Olabode, O., Nelson, M., Bathe, O., Carroll, P.R., Chan, J.M., Disaia, P., Glenn, P., Kelley, R.K., Landen, C.N., Phillips, J., Prados, M., Simko, J., Smith-McCune, K., VandenBerg, S., Roggin, K., Fehrenbach, A., Kendler, A., Sifri, S., Steele, R., Jimeno, A., Carey, F., Forgie, I., Mannelli, M., Carney, M., Hernandez, B., Campos, B., Herold-Mende, C., Jungk, C., Unterberg, A., von Deimling, A., Bossler, A., Galbraith, J., Jacobus, L., Knudson, M., Knutson, T., Ma, D., Milhem, M., Sigmund, R., Godwin, A.K., Madan, R., Rosenthal, H.G., Adebamowo, C., Adebamowo, S.N., Boussioutas, A., Beer, D., Giordano, T., Mes-Masson, A.M., Saad, F., Bocklage, T., Landrum, L., Mannel, R., Moore, K., Moxley, K., Postier, R., Walker, J., Zuna, R., Feldman, M., Valdivieso, F., Dhir, R., Luketich, J., Pinero, E.M.M., Quintero-Aguilo, M., Carlotti, C.G., Dos Santos, J.S., Kemp, R., Sankarankuty, A., Tirapelli, D., Catto, J., Agnew, K., Swisher, E., Creaney, J., Robinson, B., Shelley, C.S., Godwin, E.M., Kendall, S., Shipman, C., Bradford, C., Carey, T., Haddad, A., Moyer, J., Peterson, L., Prince, M., Rozek, L., Wolf, G., Bowman, R., Fong, K.M., Yang, I., Korst, R., Rathmell, W.K., Fantacone-Campbell, J.L., Hooke, J.A., Kovatich, A.J., Shriver, C.D., DiPersio, J., Drake, B., Govindan, R., Heath, S., Ley, T., Van Tine, B., Westervelt, P., Rubin, M.A., Lee, J. II, Aredes, N.D., Mariamidze, A., Lichtarge, O., Boutros, P.C., Raphael, B., Zhang, W., Wendl, M.C., Govindan, R., Jain, S., Wheeler, D., Kulkarni, S., Dipersio, J.F., Reimand, J., Meric-Bernstam, F., Chen, K., Shmulevich, I., Plon, S.E., Chen, F., Ding, L., 2018. Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell* 173, 355-370.e14. <https://doi.org/10.1016/j.cell.2018.03.039>

Jiang, X., Anderson, H.B., Guy, C.D., Mosca, P.J., Riedel, R.F., Cardona, D.M., 2015. Rhabdomyosarcomatous Transformation of a Gastrointestinal Stromal Tumor following Treatment with Imatinib. *Case Rep Oncol Med* 2015, 1–5. <https://doi.org/10.1155/2015/317493>

Kim, J., Light, N., Subasri, V., Young, E.L., Wegman-Ostrosky, T., Barkauskas, D.A., Hall, D., Philip, J., Lupo, J., Rajesh Patidar, J., Luke, J., Maese, D., Jones, K., Wang, J., Mingyi, Cancer, J., Laboratory, G.R., Sean,

- ;; Tavtigian, V., Wu, D., Shlien, ; Adam, Telfer, ; Frank, Goldenberg, A., Skapek, S.X., Jun, ;; Wei, S., Wen, ; Xinyu, Daniel Catchpoole, ; Douglas, ; Hawkins, S., Schiffman, J.D., Javed Khan, ; Malkin, D., Stewart, D.R., 2021. Pathogenic Germline Variants in Cancer Susceptibility Genes in Children and Young Adults With Rhabdomyosarcoma. <https://doi.org/10.1200/PO.20>
- Knudson, A.G., 1971. Mutation and Cancer: Statistical Study of Retinoblastoma.
- Kuhlen, M., Taeubner, J., Brozou, T., Wieczorek, D., Siebert, R., Borkhardt, A., 2019. Family-based germline sequencing in children with cancer. *Oncogene*. <https://doi.org/10.1038/s41388-018-0520-9>
- Liu, C., Zhou, D., Yang, K., Xu, N., Peng, J., Zhu, Z., 2023. Research progress on the pathogenesis of the SDHB mutation and related diseases. *Biomedicine and Pharmacotherapy*. <https://doi.org/10.1016/j.biopha.2023.115500>
- Liu, J., Wu, S., Xie, X., Wang, Z., Lei, Q., 2020. Identification of potential crucial genes and key pathways in osteosarcoma. *Hereditas* 157. <https://doi.org/10.1186/s41065-020-00142-0>
- McDonald, J.T., Ricks-Santi, L.J., 2022. Hereditary variants of unknown significance in African American women with breast cancer. *PLoS One* 17. <https://doi.org/10.1371/journal.pone.0273835>
- Murphy, A.J., Cheng, C., Williams, J., Shaw, T.I., Pinto, E.M., Dieseldorff-Jones, K., Brzezinski, J., Renfro, L.A., Tornwall, B., Huff, V., Hong, A.L., Mullen, E.A., Crompton, B., Dome, J.S., Fernandez, C. V., Geller, J.I., Ehrlich, P.F., Mulder, H., Oak, N., Maciejsek, J., Jablonowski, C.M., Fleming, A.M., Pichavaram, P., Morton, C.L., Easton, J., Nichols, K.E., Clay, M.R., Santiago, T., Zhang, J., Yang, J., Zambetti, G.P., Wang, Z., Davidoff, A.M., Chen, X., 2023. Genetic and epigenetic features of bilateral Wilms tumor predisposition in patients from the Children's Oncology Group AREN18B5-Q. *Nat Commun* 14. <https://doi.org/10.1038/s41467-023-43730-0>
- Nicora, G., Zucca, S., Limongelli, I., Bellazzi, R., Magni, P., 2022. A machine learning approach based on ACMG/AMP guidelines for genomic variant classification and prioritization. *Sci Rep* 12. <https://doi.org/10.1038/s41598-022-06547-3>
- Olivier, M., Hollstein, M., Hainaut, P., 2010. TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb Perspect Biol*. <https://doi.org/10.1101/cshperspect.a001008>
- Pal, A., Chiu, H.Y., Taneja, R., 2019. Genetics, epigenetics and redox homeostasis in rhabdomyosarcoma: Emerging targets and therapeutics. *Redox Biol*. <https://doi.org/10.1016/j.redox.2019.101124>

- Pandey, A.N., 2014. Retinoblastoma: An overview. *Saudi Journal of Ophthalmology*. <https://doi.org/10.1016/j.sjopt.2013.11.001>
- Parsons, D.W., Roy, A., Yang, Y., Wang, T., Scollon, S., Bergstrom, K., Kerstein, R.A., Gutierrez, S., Petersen, A.K., Bavle, A., Lin, F.Y., López-Terrada, D.H., Monzon, F.A., Hicks, M.J., Eldin, K.W., Quintanilla, N.M., Adesina, A.M., Mohila, C.A., Whitehead, W., Jea, A., Vasudevan, S.A., Nuchtern, J.G., Ramamurthy, U., McGuire, A.L., Hilsenbeck, S.G., Reid, J.G., Muzny, D.M., Wheeler, D.A., Berg, S.L., Chintagumpala, M.M., Eng, C.M., Gibbs, R.A., Plon, S.E., 2016. Diagnostic yield of clinical tumor and germline whole-exome sequencing for children with solid tumors. *JAMA Oncol* 2, 616–624. <https://doi.org/10.1001/jamaoncol.2015.5699>
- Ren, H.M., Liao, M.Q., Tan, S.X., Cheng, C., Zhu, S., Zheng, L., Ma, J.R., Mu, Y.J., Li, W.L., Zhang, S.W., OuYang, R.Q., Li, S.N., Cui, Y.F., Ke, X.Y., Luo, Z.Y., Xiong, P., Liu, J., Li, L.P., Liang, X.F., Zeng, F.F., Su, X.F., Han, L.Y., 2022. Global, Regional, and National Burden of Cancer in Children Younger Than 5 Years, 1990–2019: Analysis of the Global Burden of Disease Study 2019. *Front Public Health* 10. <https://doi.org/10.3389/fpubh.2022.910641>
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H.L., 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, 405–424. <https://doi.org/10.1038/gim.2015.30>
- Rotimi, S.O., Rotimi, O.A., Salhia, B., 2021. A Review of Cancer Genetics and Genomics Studies in Africa. *Front Oncol*. <https://doi.org/10.3389/fonc.2020.606400>
- Ruteshouser, E.C., Robinson, S.M., Huff, V., 2008. Wilms tumor genetics: Mutations in WT1, WTX, and CTNNB1 account for only about one-third of tumors. *Genes Chromosomes Cancer* 47, 461–470. <https://doi.org/10.1002/gcc.20553>
- Seo, E.S., Lee, J.W., Lim, J., Shin, S., Cho, H.W., Ju, H.Y., Yoo, K.H., Sung, K.W., Park, W.Y., 2024. Germline functional variants contribute to somatic mutation and outcomes in neuroblastoma. *Nat Commun* 15, 8360. <https://doi.org/10.1038/s41467-024-52128-5>
- Singh, J., Mishra, A., Jayamuruga Pandian, A., Mallipatna, A.C., Khetan, V., Sripriya, S., Kapoor, S., Agarwal, S., Sankaran, S., Katragadda, S., Veeramachaneni, V., Hariharan, R., Subramanian, K., Mannan, A.U., 2016. Next-generation sequencing-based method shows increased mutation detection sensitivity in an Indian retinoblastoma cohort.

Steliarova-Foucher, Eva, Colombet, Murielle, Ries, L.A.G., Moreno, Florencia, Dolya, Anastasia, Bray, F., Hesselning, P., Shin, H.Y., Stiller, C.A., Bouzbid, S., Hamdi-Cherif, M., Hablas, A., Chirpaz, E., Buziba, N., Chesumbai, G., Manraj, S., Reynders, D., Wabinga, H., Chokunonga, E., Moreno, F, Lima, C., Asturian Laporte, C., de Oliveira, J., de Aquino, J.P., Gallagher, S.V., Uribe, C., Bravo, L., Yopez Chamorro, M., Torres Alvarado, G., Galán Alvarez, Y., Martinez Reyes, F., Castillo Calvas, J., Mendoza Alava, M., Cueva Ayala, P., Hanchard, B., Fajardo-Gutiérrez, A., Zavala Zegarra, D., Barrios, E., Nikiforuk, C., Woods, R., Turner, D., MacIntyre, M., Corriveau, A., Navaneelan, T., Bertrand, C., Stuart-Panko, H., Wilson, R., Kosary, C., Shen, X., Brockhouse, J., Yee, G., Mitchell, T., Snipes, K., West, D., Rao, C., Bolick, S., Rycroft, R., Mueller, L., Zheng, Y., Dosch, K., Brown, H., Vargas, A., Levin, G., Bayakly, R., Johnson, C., Shen, T., Ruppert, L., Lynch, C., Lai, S., Tucker, T., Wu, X., Schwenn, M., Stern, K., Gershman, S., Copeland, G., Bushhouse, S., Rogers, D., Jackson Thompson, J., Lemons, D., Frederick, S., Harris, J., Riddle, B., Stroup, A., Wiggins, C., Schymura, M., Giljahn, L., Sheikh, A., Schubert, S., Aldinger, W., Fulton, J., Whiteside, M., Nogueira, L., Sweeney, C., Johnson, A., Martin, J., Farley, S., Harrelson, D., Malicki, R., Espinoza, J., Hernandez, B., Abulfateh, N., Wang, N., Ngan, R., Lingegowda, K., Swaminathan, R., Koyande, S., Silverman, B., Ozasa, K., Kanemura, S., Soda, M., Miyashiro, I., Shibata, A., Nimri, O., Won, Y., Kim, C., Hong, N., Nam, H., Kweon, S., Kim, W., Huh, J., Jung, K., Yoo, C., Elbasmy, A., Laudico, A., Lumague, M., AlMutlag, H., Buasom, R., Srisukho, S., Tanabodee, J., Wiangnon, S., Pongnikorn, D., Sriplung, H., Dirican, O., Eser, S., Le Hoang, M., Hackl, M., Zborovskaya, A., Dimitrova, N., Valerianova, Z., Sekerija, M., Pavlou, P., Dušek, M., Mägi, M., Clavel, J., Lacour, B., Guizard, A., Bouvier, V., Troussard, X., Woronoff, A., Tretarre, B., Colonna, M., Molinié, F., Bara, S., Velten, M., Marrer, E., Ganry, O., Grosclaude, P., Kaatsch, P., Zeissig, S., Holleczeck, B., Katalinic, A., Jakab, Z., Birgisson, H., Walsh, P., Mangone, L., Merletti, F., Magoni, M., Mangone, L., Ferretti, S., Serraino, D., Spagnoli, G., Fusco, M., Michiara, M., Tumino, R., Falcini, F., Sensi, F., Tisano, F., Piffer, S., Stracci, F., Tagliabue, G., Smailyte, G., Agius, D., Visser, O., Ursin, G., Didkowska, J., Trojanowski, M., Wojciechowska, U., Forjaz de Lacerda, G., Silva, M., Laranja Pontes, J., da Costa Miranda, A., Kaiserova, E., Primic Žakelj, M., Peris-Bonet, R., Vicente Raneda, M., Almar Marqués, E., Quirós García, J., Ramos Monserrat, M., Errezola Saizar, M., Alemán Herrera, A., Díaz García, J., Marcos-Gragera, R., Sanchez-Perez, M., Ardanaz Aicua, E., Galceran, J., Klint, A., Kuehni, C., Bouchardy, C., Levi, F., Bordoni, A., Konzelmann, I., Rohrmann, S., Stiller, C., Gavin, A., Brewster, D., Phung, H., Rushton, S., Guthridge, S., Aitken, J., D’Onise, K., Venn, A., Farrugian, H., Threlfall, T., Laumond, S., Yen Kai Sun, L., Hendrix, J., Ballantine, K., Colombet, M., Dolya, A., Masuyer, E., Steliarova-Foucher, E, 2017. International incidence of childhood cancer, 2001–10: a population-based registry study. *Lancet Oncol* 18, 719–731. [https://doi.org/10.1016/S1470-2045\(17\)30186-9](https://doi.org/10.1016/S1470-2045(17)30186-9)

- Uchimura, A., Higuchi, M., Minakuchi, Y., Ohno, M., Toyoda, A., Fujiyama, A., Miura, I., Wakana, S., Nishino, J., Yagi, T., 2015. Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. *Genome Res* 25, 1125–1134. <https://doi.org/10.1101/gr.186148.114>
- Villani, A., Davidson, S., Kanwar, N., Lo, W.W., Li, Y., Cohen-Gogo, S., Fuligni, F., Edward, L.M., Light, N., Layeghifard, M., Harripaul, R., Waldman, L., Gallinger, B., Comitani, F., Brunga, L., Hayes, R., Anderson, N.D., Ramani, A.K., Yuki, K.E., Blay, S., Johnstone, B., Inglese, C., Hammad, R., Goudie, C., Shuen, A., Wasserman, J.D., Venier, R.E., Eliou, M., Lorenti, M., Ryan, C.A., Braga, M., Gloven-Brown, M., Han, J., Montero, M., Spatare, F., Whitlock, J.A., Scherer, S.W., Chun, K., Somerville, M.J., Hawkins, C., Abdelhaleem, M., Ramaswamy, V., Somers, G.R., Kyriakopoulou, L., Hitzler, J., Shago, M., Morgenstern, D.A., Tabori, U., Meyn, S., Irwin, M.S., Malkin, D., Shlien, A., 2023. The clinical utility of integrative genomics in childhood cancer extends beyond targetable mutations. *Nat Cancer* 4, 203–221. <https://doi.org/10.1038/s43018-022-00474-y>
- Wang, H., Guo, M., Wei, H., Chen, Y., 2023. Targeting p53 pathways: mechanisms, structures, and advances in therapy. *Signal Transduct Target Ther*. <https://doi.org/10.1038/s41392-023-01347-1>
- Wu, Y., Deng, Y., Wei, B., Xiang, D., Hu, J., Zhao, P., Lin, S., Zheng, Y., Yao, J., Zhai, Z., Wang, S., Lou, W., Yang, S., Zhang, D., Lyu, J., Dai, Z., 2022. Global, regional, and national childhood cancer burden, 1990–2019: An analysis based on the Global Burden of Disease Study 2019. *J Adv Res* 40, 233–247. <https://doi.org/10.1016/j.jare.2022.06.001>
- Xu, L., Zheng, Y., Liu, J., Rakheja, D., Singleterry, S., Laetsch, T.W., Shern, J.F., Khan, J., Triche, T.J., Hawkins, D.S., Amatruda, J.F., Skapek, S.X., 2018. Integrative Bayesian Analysis Identifies Rhabdomyosarcoma Disease Genes. *Cell Rep* 24, 238–251. <https://doi.org/10.1016/j.celrep.2018.06.006>
- Zhang, J., Walsh, M.F., Wu, G., Edmonson, M.N., Gruber, T.A., Easton, J., Hedges, D., Ma, X., Zhou, X., Yergeau, D.A., Wilkinson, M.R., Vadodaria, B., Chen, X., McGee, R.B., Hines-Dowell, S., Nuccio, R., Quinn, E., Shurtleff, S.A., Rusch, M., Patel, A., Becksfort, J.B., Wang, S., Weaver, M.S., Ding, L., Mardis, E.R., Wilson, R.K., Gajjar, A., Ellison, D.W., Pappo, A.S., Pui, C.-H., Nichols, K.E., Downing, J.R., 2015. Germline Mutations in Predisposition Genes in Pediatric Cancer. *New England Journal of Medicine* 373, 2336–2346. <https://doi.org/10.1056/nejmoa1508054>
- Zhao, L., Liu, H., Yuan, X., Gao, K., Duan, J., 2020. Comparative study of whole exome sequencing-based copy number variation detection tools. *BMC Bioinformatics* 21. <https://doi.org/10.1186/s12859-020-3421-1>

## Appendix D: Turnitin Report and Plagiarism Declaration



### PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I Lana Naudé (Student number: 2740639) am a student registered for the degree of MSc (Med) Genomic Medicine in the academic year 2024.

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: *Lana Naudé*

Date: 29/ 11/ 2024

## Lana\_Final\_Research\_Report.pdf

### ORIGINALITY REPORT

8%

SIMILARITY INDEX

4%

INTERNET SOURCES

4%

PUBLICATIONS

5%

STUDENT PAPERS

### PRIMARY SOURCES

1

Submitted to University of Witwatersrand

Student Paper

3%

2

[www.frontiersin.org](http://www.frontiersin.org)

Internet Source

1%

3

"33rd European Congress of Pathology – Abstracts", Virchows Archiv, 2021

Publication

1%

4

J. K. Cowell. "Molecular Genetics of Cancer", CRC Press, 2003

Publication

<1%

5

[www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)

Internet Source

<1%

6

Submitted to University of Cambridge

Student Paper

<1%

7

Catherine Collin, Masaru Minami, Hasan Parvez, Hideya Saito, Simone Parvez, Qureshi, Claude Reiss. "Advances in Neuroregulation and Neuroprotection", CRC Press, 2019

Publication

<1%

[www.researchgate.net](http://www.researchgate.net)

|    |                                                                                                                                                                                                                             |      |
|----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 8  | Internet Source                                                                                                                                                                                                             | <1 % |
| 9  | Submitted to University of Limerick<br>Student Paper                                                                                                                                                                        | <1 % |
| 10 | <a href="http://www.aaabg.org">www.aaabg.org</a><br>Internet Source                                                                                                                                                         | <1 % |
| 11 | <a href="http://academic.oup.com">academic.oup.com</a><br>Internet Source                                                                                                                                                   | <1 % |
| 12 | Patience Obasaju, Shubin Shahab, Emily Dunn, Daniel S. Rhee et al. "BRAF V600E-mutated metastatic pediatric Wilms tumor with complete response to targeted RAF/MEK inhibition", Molecular Case Studies, 2020<br>Publication | <1 % |

Exclude quotes Off  
Exclude bibliography Off

Exclude matches Off

## **Appendix E: Author Guidelines for *Frontiers***

### **Original Research**

Original Research articles report on primary and unpublished studies. Original Research may also encompass confirming studies and disconfirming results which allow hypothesis elimination, reformulation and/or report on the non-reproducibility of previously published results. Original Research articles are peer-reviewed, have a maximum word count of 12,000 and may contain no more than 15 Figures/Tables. Authors are required to pay a fee (A-type article) to publish an Original Research article. Original Research articles should have the following format: 1) Abstract, 2) Introduction, 3) Materials and Methods, 4) Results, 5) Discussion.

### **Author guidelines**

### **General standards**

### **Article type**

Frontiers requires authors to select the appropriate article type for their manuscript and to comply with the article type descriptions defined in the journal's 'Article types' page, which can be seen from the 'For authors' menu on every Frontiers journal page. Please pay close attention to the word count limits.

### **Templates**

If working with Word please use our [Word templates](#). If you wish to submit your article as LaTeX, we recommend our [LaTeX templates](#). For LaTeX files, please ensure all relevant manuscript files are uploaded: .tex file, PDF, and .bib file (if the bibliography is not already included in the .tex file).

During the [interactive review](#), authors are encouraged to upload versions using track changes. Editors and reviewers can only download the PDF file of the submitted manuscript.

### **Manuscript length**

Frontiers encourages the authors to closely follow the article word count lengths given in the 'Article types' page of the journals. The manuscript length includes only the main body of the text, footnotes,

and all citations within it, and excludes the abstract, section titles, figure and table captions, funding statement, acknowledgments, and references in the bibliography. Please indicate the number of words and the number of figures and tables included in your manuscript on the first page.

### Language editing

Frontiers requires manuscripts submitted to meet international English language standards to be considered for publication.

For authors who would like their manuscript to receive language editing or proofreading to improve the clarity of the manuscript and help highlight their research, Frontiers recommends the language-editing services provided by the following external partners.

Note that sending your manuscript for language editing does not imply or guarantee that it will be accepted for publication by a Frontiers journal. Editorial decisions on the scientific content of a manuscript are independent of whether it has received language editing or proofreading by these partner services or other services.

**Editage** Frontiers is pleased to recommend the language-editing service provided by our external partner Editage to authors who believe their manuscripts would benefit from professional editing. These services may be particularly useful for researchers for whom English is not the primary language. They can help to improve the grammar, syntax, and flow of your manuscript prior to submission. Frontiers authors will receive a 10% discount by visiting the following link: [editage.com/frontiers](https://editage.com/frontiers).

**The Charlesworth Group** Frontiers recommends the Charlesworth Group's author services, who has a long-standing track record in language editing and proofreading. This is a third-party service for which Frontiers authors will receive a 10% discount by visiting the following link: [www.cwauthors.com/frontiers](https://www.cwauthors.com/frontiers).

Frontiers推荐您使用在英语语言编辑和校对领域具有悠久历史和良好口碑的查尔斯沃思作者服务。此项服务由第三方为您提供，Frontiers中国作者通过此链接提交稿件时可获得10%的特别优惠: [www.cwauthors.com.cn/frontiers](https://www.cwauthors.com.cn/frontiers)

### Language style

The default language style at Frontiers is American English. If you prefer your article to be formatted in British English, please specify this on the first page of your manuscript. For any questions regarding style, Frontiers recommends authors to consult the [Chicago Manual of Style](#).

## Search engine optimization (SEO)

There are a few simple ways to maximize your article's discoverability. Follow the steps below to improve the search results of your article:

- include a few of your article's keywords in the title of the article
- do not use long article titles
- pick 5-8 keywords using a mix of generic and more specific terms on the article subject(s)
- use the maximum amount of keywords in the first two sentences of the abstract
- use some of the keywords in level 1 headings.

## CrossMark policy

[CrossMark](#) is a multi-publisher initiative to provide a standard way for readers to locate the current version of a piece of content. By applying the CrossMark logo Frontiers is committed to maintaining the content it publishes and to alerting readers to changes if and when they occur. Clicking on the CrossMark logo will tell you the current status of a document and may also give you additional publication record information about the document.

## Title

The title should be concise, omitting terms that are implicit and, where possible, be a statement of the main result or conclusion presented in the manuscript. Abbreviations should be avoided within the title.

Witty or creative titles are welcome, but only if relevant and within measure. Consider if a title meant to be thought-provoking might be misinterpreted as offensive or alarming. In extreme cases, the editorial office may veto a title and propose an alternative. Authors should avoid:

- titles that are a mere question without giving the answer
- unambitious titles, for example starting with 'Towards,' 'A description of,' 'A characterization of' or 'Preliminary study on'
- vague titles, for example starting with 'Role of', 'Link between', or 'Effect of' that do not specify the role, link, or effect
- including terms that are out of place, for example the taxonomic affiliation apart from species name.

For Corrigenda, General Commentaries, and Editorials, the title of your manuscript should have the following format:

- 'Corrigendum: Title of Original Article'
- General Commentaries: 'Commentary: Title of Original Article' 'Response: Commentary: Title of Original Article'
- 'Editorial: Title of Research Topic'

The running title should be a maximum of five words in length.

### **Authors and affiliations**

All names are listed together and separated by commas. Provide exact and correct author names as these will be indexed in official archives. Affiliations should be keyed to the author's name with superscript numbers and be listed as follows:

- Laboratory, Institute, Department, Organization, City, State abbreviation (only for United States, Canada, and Australia), and Country (without detailed address information such as city zip codes or street names).

Example: Max Maximus<sup>1</sup> <sup>1</sup> Department of Excellence, International University of Science, New York, NY, United States.

### **Correspondence**

The corresponding author(s) should be marked with an asterisk in the author list. Provide the exact contact email address of the corresponding author(s) in a separate section. Example: Max Maximus\* maximus@iuscience.edu If any authors wish to include a change of address, list the present address(es) below the correspondence details using a unique superscript symbol keyed to the author(s) in the author list.

### **Equal contributions**

The authors who have contributed equally should be marked with a symbol (†) in the author list of the doc/latex and pdf files of the manuscript uploaded at submission.

Please use the appropriate standard statement(s) to indicate equal contributions:

- **Equal contribution:** These authors contributed equally to this work
- **First authorship:** These authors share first authorship
- **Senior authorship:** These authors share senior authorship

- **Last authorship:** These authors share last authorship
- **Equal contribution and first authorship:** These authors contributed equally to this work and share first authorship
- **Equal contribution and senior authorship:** These authors contributed equally to this work and share senior authorship
- **Equal contribution and last authorship:** These authors contributed equally to this work and share last authorship

Example: Max Maximus 1<sup>†</sup>, John Smith<sup>2†</sup> and Barbara Smith<sup>1</sup> <sup>†</sup>These authors contributed equally to this work and share first authorship

### **Consortium/group and collaborative authors**

Consortium/group authorship should be listed in the manuscript with the other author(s).

In cases where authorship is retained by the consortium/group, the consortium/group should be listed as an author separated by a comma or 'and'. The consortium/group name will appear in the author list, in the citation, and in the copyright. If provided, the consortium/group members will be listed in a separate section at the end of the article.

For the collaborators of the consortium/group to be indexed in PubMed, they do not have to be inserted in the Frontiers submission system individually. However, in the manuscript itself, provide a section with the name of the consortium/group as the heading followed by the list of collaborators, so they can be tagged accordingly and indexed properly.

Example: John Smith, Barbara Smith and The Collaborative Working Group. In cases where work is presented by the author(s) on behalf of a consortium/group, it should be included in the author list separated with the wording 'for' or 'on behalf of.' The consortium/group will not retain authorship and will only appear in the author list.

Example: John Smith and Barbara Smith on behalf of The Collaborative Working Group.

## **Artificial intelligence**

These guidelines cover acceptable uses of generative AI technologies such as Large Language Models (ChatGPT, Jasper) and text-to-image generators (DALL-E 2, Midjourney, Stable Diffusion) in the writing or editing of manuscripts submitted to Frontiers.

## **AI use by authors**

Authors should not list a generative AI technology as a co-author or author of any submitted manuscript. Generative AI technologies cannot be held accountable for all aspects of a manuscript and consequently do not meet the criteria required for authorship.

If the author of a submitted manuscript has used written or visual content produced by or edited using a generative AI technology, this use must follow all Frontiers guidelines and policies. Specifically, the author is responsible for checking the factual accuracy of any content created by the generative AI technology. This includes, but is not limited to, any quotes, citations or references. Figures produced by or edited using a generative AI technology must be checked to ensure they accurately reflect the data presented in the manuscript. Authors must also check that any written or visual content produced by or edited using a generative AI technology is free from plagiarism.

If the author of a submitted manuscript has used written or visual content produced by or edited using a generative AI technology, such use must be acknowledged in the acknowledgements section of the manuscript and the methods section if applicable. This explanation must list the name, version, model, and source of the generative AI technology. We encourage authors to upload all input prompts provided to a generative AI technology and outputs received from a generative AI technology in the supplementary files for the manuscript.

## **Abstract**

As a primary goal, the abstract should make the general significance and conceptual advance of the work clearly accessible to a broad readership. The abstract should be no longer than a single paragraph and should be structured, for example, according to the IMRAD format. For the specific structure of the abstract, authors should follow the requirements of the article type or journal to which they're submitting. Minimize the use of abbreviations and do not cite references, figures or tables. For clinical trial articles, please include the unique identifier and the URL of the publicly-accessible website on which the trial is registered.

## **Keywords**

All article types require a minimum of five and a maximum of eight keywords.

## Text

The entire document should be single-spaced and must contain page and line numbers in order to facilitate the review process. The manuscript should be written using either Word or LaTeX. See above for templates.

## Nomenclature

The use of abbreviations should be kept to a minimum. Non-standard abbreviations should be avoided unless they appear at least four times, and must be defined upon first use in the main text. Consider also giving a list of non-standard abbreviations at the end, immediately before the acknowledgments.

Equations should be inserted in editable format from the equation editor.

Italicize gene symbols and use the approved gene nomenclature where it is available. For human genes, please refer to the HUGO Gene Nomenclature Committee ([HGNC](#)). New symbols for human genes should be submitted to the HGNC [here](#). Common alternative gene aliases may also be reported, but should not be used alone in place of the HGNC symbol. Nomenclature committees for other species are listed [here](#). Protein products are not italicized.

We encourage the use of Standard International Units in all manuscripts.

Chemical compounds and biomolecules should be referred to using systematic nomenclature, preferably using the recommendations by the International Union of Pure and Applied Chemistry (IUPAC).

Astronomical objects should be referred to using the nomenclature given by the International Astronomical Union (IAU) provided [here](#).

Life Science Identifiers (LSIDs) for ZOOBANK registered names or nomenclatural acts should be listed in the manuscript before the keywords. An LSID is represented as a uniform resource name (URN) with the following format: urn:lsid:<Authority>:<Namespace>:<ObjectID>[:<Version>]

For more information on LSIDs please see the 'Code' section of our [policies and publication ethics](#).

## Sections

The manuscript is organized by headings and subheadings. The section headings should be those appropriate for your field and the research itself. You may insert up to 5 heading levels into your manuscript (i.e.,: 3.2.2.1.2 Heading Title).

For Original Research articles, it is recommended to organize your manuscript in the following sections or their equivalents for your field.

**Introduction** Succinct, with no subheadings.

**Materials and methods** This section may be divided by subheadings and should contain sufficient detail so that when read in conjunction with cited references, all procedures can be repeated. For experiments reporting results on animal or human subject research, an ethics approval statement should be included in this section (for further information, see the 'Bioethics' section of our [policies and publication ethics](#).)

**Results** This section may be divided by subheadings. Footnotes should not be used and must be transferred to the main text.

**Discussion** This section may be divided by subheadings. Discussions should cover the key findings of the study: discuss any prior research related to the subject to place the novelty of the discovery in the appropriate context, discuss the potential shortcomings and limitations on their interpretations, discuss their integration into the current understanding of the problem and how this advances the current views, speculate on the future direction of the research, and freely postulate theories that could be tested in the future.

For further information, please check the descriptions defined in the journal's 'Article types' page, in the 'For authors' menu on every journal page.

## **Acknowledgments**

This is a short text to acknowledge the contributions of specific colleagues, institutions, or agencies that aided the efforts of the authors. Should the content of the manuscript have previously appeared online, such as in a thesis or preprint, this should be mentioned here, in addition to listing the source within the reference list.

## **Scope statement**

When you submit your manuscript, you will be required to summarize in 200 words your manuscript's scope and its relevance to the journal and/or specialty section you're submitting to. The aim is to

convey to editors and reviewers how the contents of your manuscript fit within the selected journal's scope. This statement will not be published with your article if it is accepted for publication. The information will be used during the initial validation and review processes to assess whether the manuscript is a suitable fit for the chosen journal and specialty. We encourage you to consider carefully where to submit your manuscript, as submissions to an unsuitable journal or specialty will result in delays and increase the likelihood of manuscript rejection. If you are submitting to a Research Topic, please also clarify how your submission is suited to the specific topic.

## **Figure and table guidelines**

### **CC-BY license**

All figures, tables, and images will be published under a Creative Commons [CC-BY license](#), and permission must be obtained for use of copyrighted material from other sources (including re-published/adapted/modified/partial figures and images from the internet). It is the responsibility of the authors to acquire the licenses, follow any citation instructions requested by third-party rights holders, and cover any supplementary charges.

For additional information, please see the 'Image manipulation' section of our [policies and publication ethics](#).

### **Figure requirements and style guidelines**

Frontiers requires figures to be submitted individually, in the same order as they are referred to in the manuscript; the figures will then be automatically embedded at the end of the submitted manuscript. Kindly ensure that each figure is mentioned in the text and in numerical order.

For figures with more than one panel, panels should be clearly indicated using labels (A), (B), (C), (D), etc. However, do not embed the part labels over any part of the image, these labels will be replaced during typesetting according to Frontiers' journal style. For graphs, there must be a self-explanatory label (including units) along each axis.

For LaTeX files, figures should be included in the provided PDF. In case of acceptance, our production office might require high-resolution files of the figures included in the manuscript in EPS, JPEG or TIF/TIFF format.

To upload more than one figure at a time, save the figures (labeled in order of appearance in the manuscript) in a zip file and upload them as 'Supplementary Material Presentation.'

Please note that figures not in accordance with the guidelines will cause substantial delay during the production process.

### ***Captions***

Captions should be preceded by the appropriate label, for example 'Figure 1.' Figure captions should be placed at the end of the manuscript. Figure panels are referred to by bold capital letters in brackets: (A), (B), (C), (D), etc.

### **Image size and resolution requirements**

Figures should be prepared with the PDF layout in mind. Individual figures should not be longer than one page and with a width that corresponds to 1 column (85 mm) or 2 columns (180 mm).

All images must have a resolution of 300 dpi at final size. Check the resolution of your figure by enlarging it to 150%. If the image appears blurry, jagged, or has a stair-stepped effect, the resolution is too low.

The text should be legible and of high quality. The smallest visible text should be no less than eight points in height when viewed at actual size.

Solid lines should not be broken up. Any lines in the graphic should be no smaller than two points wide.

Please note that saving a figure directly as an image file (JPEG, TIF) can greatly affect the resolution of your image. To avoid this, one option is to export the file as PDF, then convert into TIFF or EPS using a graphics software.

### **Format and color image mode**

The following formats are accepted: TIF/TIFF (.tif/.tiff), JPEG (.jpg), and EPS (.eps) (upon acceptance). Images must be submitted in the color mode RGB.

### **Chemical structures**

Chemical structures should be prepared using ChemDraw or a similar program. If working with ChemDraw please use our [ChemDraw template](#). If working with another program please follow the guidelines below.

- Drawing settings: chain angle, 120° bond spacing, 18% width; fixed length, 14.4 pt; bold width, 2.0 pt; line width, 0.6 pt; margin width, 1.6 pt; hash spacing, 2.5 pt. Scale 100% Atom Label settings: font, Arial; size, 8 pt
- Assign all chemical compounds a bold, Arabic numeral in the order in which the compounds are presented in the manuscript text.

### Table requirements and style guidelines

Tables should be inserted at the end of the manuscript in an editable format. If you use a word processor, build your table in Word. If you use a LaTeX processor, build your table in LaTeX. An empty line should be left before and after the table.

Table captions must be placed immediately before the table. Captions should be preceded by the appropriate label, for example 'Table 1.' Please use only a single paragraph for the caption.

Kindly ensure that each table is mentioned in the text and in numerical order.

Please note that large tables covering several pages cannot be included in the final PDF for formatting reasons. These tables will be published as supplementary material.

Tables which are not according to the above guidelines will cause substantial delay during the production process.

### Accessibility

Frontiers encourages authors to make the figures and visual elements of their articles accessible for the visually impaired. An effective use of color can help people with low visual acuity, or color blindness, understand all the content of an article.

These guidelines are easy to implement and are in accordance with the W3C Web Content Accessibility Guidelines ([WCAG 2.1](#)), the standard for web accessibility best practices.

**Ensure sufficient contrast between text and its background** People who have low visual acuity or color blindness could find it difficult to read text with low contrast background color. Try using colors that provide maximum contrast.

WC3 recommends the following contrast ratio levels:

- Level AA, contrast ratio of at least 4.5:1

- Level AAA, contrast ratio of at least 7:1

You can verify the contrast ratio of your palette with these online ratio checkers:

- [WebAIM](#)
- [Color Safe](#)

**Avoid using red or green indicators** More than 99% of color-blind people have a red-green color vision deficiency.

**Avoid using only color to communicate information** Elements with complex information like charts and graphs can be hard to read when only color is used to distinguish the data. Try to use other visual aspects to communicate information, such as shape, labels, and size. Incorporating patterns into the shape fills also make differences clearer; for an example please see below:

### Supplementary material

Data that are not of primary importance to the text, or which cannot be included in the article because they are too large or the current format does not permit it (such as videos, raw data traces, PowerPoint presentations, etc.), can be uploaded as supplementary material during the submission procedure and will be displayed along with the published article. All supplementary files are deposited to Figshare for permanent storage and receive a DOI.

Supplementary material is not typeset, so please ensure that all information is clearly presented without tracked changes/highlighted text/line numbers, and the appropriate caption is included in the file. To avoid discrepancies between the published article and the supplementary material, please do not add the title, author list, affiliations or correspondence in the supplementary files.

The supplementary material can be uploaded as:

- data sheet (Word, Excel, CSV, CDX, FASTA, PDF or Zip files)
- presentation (PowerPoint, PDF or Zip files)
- image (CDX, EPS, JPEG, PDF, PNG or TIF/TIFF),
- table (Word, Excel, CSV or PDF)
- audio (MP3, WAV or WMA)
- video (AVI, DIVX, FLV, MOV, MP4, MPEG, MPG or WMV).

Technical requirements for supplementary images:

- 300 DPIs
- RGB color mode.

For supplementary material templates (LaTeX and Word), see our [supplementary material templates](#).

## References

Frontiers' journals use one of two reference styles, either Harvard (author-date) or Vancouver (numbered). Please check our [help center](#) to find the correct style for the journal to which you are submitting.

- All citations in the text, figures or tables must be in the reference list and vice-versa
- The names of the first six authors followed by et al. and the DOI (when available) should be provided
- Given names of authors should be abbreviated to initials (e.g., Smith, J., Lewis, C.S., etc.)
- The reference list should only include articles that are published or accepted
- Unpublished data, submitted manuscripts, or personal communications should be cited within the text only, for article types that allow such inclusions
- For accepted but unpublished works use 'in press' instead of page numbers
- Data sets that have been deposited to an online repository should be included in the reference list. Include the version and unique identifier when available
- Personal communications should be documented by a letter of permission
- Website URLs should be included as footnotes
- Any inclusion of verbatim text must be contained in quotation marks and clearly reference the original source
- Preprints can be cited as long as a DOI or archive URL is available, and the citation clearly mentions that the contribution is a preprint. If a peer-reviewed journal publication for the same preprint exists, the official journal publication is the preferred source. See the preprints section for each reference style below for more information.

### Harvard reference style (author-date)

Many Frontiers journals use the Harvard referencing system; to find the correct reference style and resources for the journal you are submitting to, please visit our [help center](#). Reference examples are found below, for more examples of citing other documents and general questions regarding the Harvard reference style, please refer to the [Chicago Manual of Style](#).

#### *In-text citations*

- For works by a single author, include the surname, followed by the year
- For works by two authors, include both surnames, followed by the year

- For works by more than two authors, include only the surname of the first author followed by et al., followed by the year
- For humanities and social sciences articles, include the page numbers.

### ***Reference examples***

**Article in a print journal** Sondheimer, N., and Lindquist, S. (2000). Rnq1: an epigenetic modifier of protein function in yeast. *Mol. Cell.* 5, 163-172.

**Article in an online journal** Tahimic, C.G.T., Wang, Y., Bikle, D.D. (2013). Anabolic effects of IGF-1 signaling on the skeleton. *Front. Endocrinol.* 4:6. doi: 10.3389/fendo.2013.00006

**Article or chapter in a book** Sorenson, P. W., and Caprio, J. C. (1998). "Chemoreception," in *The Physiology of Fishes*, ed. D. H. Evans (Boca Raton, FL: CRC Press), 375-405.

**Book** Cowan, W. M., Jessell, T. M., and Zipursky, S. L. (1997). *Molecular and Cellular Approaches to Neural Development*. New York: Oxford University Press.

**Abstract** Hendricks, J., Applebaum, R., and Kunkel, S. (2010). A world apart? Bridging the gap between theory and applied social gerontology. *Gerontologist* 50, 284-293. Abstract retrieved from Abstracts in Social Gerontology database. (Accession No. 50360869)

**Website** World Health Organization. (2018). E. coli. <https://www.who.int/news-room/fact-sheets/detail/e-coli> [Accessed March 15, 2018].

**Patent** Marshall, S. P. (2000). Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity. U.S. Patent No 6,090,051. Washington, DC: U.S. Patent and Trademark Office.

**Data** Perdiguero P, Venturas M, Cervera MT, Gil L, Collada C. Data from: Massive sequencing of Ulms minor's transcriptome provides new molecular tools for a genus under the constant threat of Dutch elm disease. Dryad Digital Repository. (2015) <http://dx.doi.org/10.5061/dryad.ps837>

**Theses and dissertations** Smith, J. (2008) Post-structuralist discourse relative to phenomenological pursuits in the deconstructivist arena. [dissertation/master's thesis]. [Chicago (IL)]: University of Chicago

**Preprint** Smith, J. (2008). Title of the document. Preprint repository name [Preprint]. Available at: <https://persistent-url> (Accessed March 15, 2018).

### **Vancouver reference style (numbered)**

Many Frontiers journals use the numbered referencing system; to find the correct reference style and resources for the journal you are submitting to, please visit our [help center](#).

Reference examples are found below, for more examples of citing other documents and general questions regarding the Vancouver reference style, please refer to [Citing Medicine](#).

### ***In-text citations***

- Please apply the Vancouver system for in-text citations
- In-text citations should be numbered consecutively in order of appearance in the text – identified by Arabic numerals in the parenthesis (use square brackets for physics and mathematics articles).

### ***Reference examples***

**Article in a print journal** Sondheimer N, Lindquist S. Rnq1: an epigenetic modifier of protein function in yeast. *Mol Cell* (2000) 5:163-72.

**Article in an online journal** Tahimic CGT, Wang Y, Bikle DD. Anabolic effects of IGF-1 signaling on the skeleton. *Front Endocrinol* (2013) 4:6. doi: 10.3389/fendo.2013.00006

**Article or chapter in a book** Sorenson PW, Caprio JC. "Chemoreception". In: Evans DH, editor. *The Physiology of Fishes*. Boca Raton, FL: CRC Press (1998). p. 375-405.

**Book** Cowan WM, Jessell TM, Zipursky SL. *Molecular and Cellular Approaches to Neural Development*. New York: Oxford University Press (1997). 345 p.

**Abstract** Christensen S, Oppacher F. An analysis of Koza's computational effort statistic for genetic programming. In: Foster JA, editor. *Genetic Programming. EuroGP 2002: Proceedings of the 5th European Conference on Genetic Programming; 2002 Apr 3–5; Kinsdale, Ireland*. Berlin: Springer (2002). p. 182–91.

**Website** World Health Organization. *E. coli* (2018). <https://www.who.int/news-room/fact-sheets/detail/e-coli> [Accessed March 15, 2018].

**Patent** Pagedas AC, inventor; Ancel Surgical R&D Inc., assignee. Flexible Endoscopic Grasping and Cutting Device and Positioning Tool Assembly. United States patent US 20020103498 (2002).

**Data** Perdiguerio P, Venturas M, Cervera MT, Gil L, Collada C. Data from: Massive sequencing of Ulms minor's transcriptome provides new molecular tools for a genus under the constant threat of Dutch elm disease. Dryad Digital Repository. (2015) <http://dx.doi.org/10.5061/dryad.ps837>

### **Theses and dissertations**

Smith, J. (2008) Post-structuralist discourse relative to phenomological pursuits in the deconstructivist arena. [dissertation/master's thesis]. [Chicago (IL)]: University of Chicago

**Preprint** Smith, J. Title of the document. Preprint repository name [Preprint] (2008). Available at: <https://persistent-url> (Accessed March 15, 2018).