# Genomics of Lynch Syndrome and Constitutional Mismatch Repair Deficiency Syndrome

**Lindiwe Lamola M.Sc. (Med)**



Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

Division of Human Genetics, Department of Pathology

UNIVERSITY OF CAPE TOWN

July 2017

Supervisor: Prof. Raj Ramesar

# Abstract

**Introduction:** The mismatch repair system plays an important role in maintaining the genome integrity as it functions to correct mismatches during DNA replication. Heterozygous mutations in one of the mismatch repair (MMR) genes e.g. *MLH1, MSH2, MSH6* and *PMS2* cause the dominant adult cancer syndrome termed Lynch syndrome (or hereditary non-polyposis colorectal cancer). In our South African cohort, the *MLH1* exon 13 c.1528C>T mutation is the most common Lynch syndrome-causing variant in the Mixed Ancestry population. Recently, a patient homozygous for this mutation, diagnosed with Constitutional mismatch repair deficiency (CMMR-D) syndrome was described within this extended cohort. CMMRD syndrome results in an increased predisposition to a range of cancers, most commonly brain and hematological tumours in early childhood. The aims of this thesis were: (i) to determine the rate of extra-colonic cancers in the cohort of Lynch syndrome families in our colorectal cancer registry, (ii) to determine if *MLH1* c.1528C>T is a founder mutation, and (iii) to focus on the CMMR-D syndrome as a branch of Lynch syndrome and to potentially use the hypermutability-status in CMMR-D to understand the diverse carcinogenesis in Lynch syndrome.

**Methods:** The registry consisting of Lynch syndrome families was interrogated and analysed to address the aim (i). Haplotype analysis was performed using microsatellite markers around the *MHL1* c.1528C>T mutation to determine founder effect for aim (ii). For aim (iii) whole exome sequencing was also performed in a Lynch/ CMMR-D syndrome family in order to investigate the extent of hypermutability in CMMR-D syndrome, and to develop a working hypothesis for carcinogenesis in CMMR- D and Lynch syndromes.

**Results**: From the analysis of the registry it was noted that 396 individuals carried a disease-causing mutation in either *MLH1* or *MSH2*; females have a relatively later age of onset (for cancer) than males and *MLH1* mutation carriers develop cancers relatively earlier in life than in individuals with *MSH2* mutations. The most common extra-colonic cancers were endometrial and breast in females; in males small bowel cancer was most common, after CRC. The cohort study revealed a large founder effect with the *MLH1* c.1528C>T mutation, with the most common inferred (disease-associated) haplotype found in 25 of the 30 subjects tested; the disease-associated haplotype was not present in controls. The mutation aging analysis traced the mutation to be ~225 years old. The WES investigation of the nuclear family within which the CMMR-D patient, including acquired and germline mutations in tissues from the child with CMMR-D, revealed a range of pathways including the extracellular matrix, WNT signaling, TGFβ and p53 as acquiring significant numbers of variants as a result of the MMR deficiency.

**Discussion and Conclusion:** The results which are indicative of the need to improve the Lynch syndrome mutation testing and management for all patients, also suggests the need to develop surveillance programs for extra-colonic cancers, which will improve compliance and disease-free survival. WES investigation of the nuclear family containing a child with CMMR-D point to the potential involvement of a range of pathways associated with cancer development which may be indirectly invoked in the process of tumorigenesis by the wide range of variants acquired as a result of mismatch repair deficiency. It is likely that some of these processes are also involved in the emergence of extracolonic cancers in individuals affected with Lynch syndrome (i.e. heterozygous for mutations in MMR genes).

# Plagiarism Declaration

"This thesis/dissertation has been submitted to the Turnitin module (or equivalent similarity and originality checking software) and I confirm that my supervisor has seen my report and any concerns revealed by such have been resolved with my supervisor."

Name of student:

**Student number: LMLLIN001**

Signed by candidate

**Signature:** Signature Removed

**Date: 18 July 2017**

# Dedication

"Ke habile lehodimong
Tseleng ha kena bodutu
Ha ke tsamaye ke le mong
Ke na le Morena Jesu
Lefelleng moo ke tsamayang
Ho lahleha ba bangata
Empa nna ya ntsamaisang
Ke Jeso, ke monga tsela
**Ntate ke mang ya kang ka wena**
O tshwanatswe ke thoriso
O morena wa marena
O kgosi ya dikgosi
Molefatsheng le Mahodimong
Hao ya tshwanang le wena
Hao ya tshwanang le wena"

**This thesis is dedicated to:**

Elizabeth Mamankoane Lamola
Teboho Phillip Lamola
Marriam Monyane
Nonhlanhla Neano Lamola

## Acknowledgments

To my supervisor, Prof. R. Ramesar, thank you for allowing me to be part of this project, thank you for all the support and confidence you had in me to take on this project.
To Alvera Anna Vorster, thank you for the smoke breaks and coffee moments that have, to date, shaped and defined projects I've worked on. God knows some of the most brilliant ideas and results were discussed during those moments.
To Emilson Chimusa, thank you for all the academic support and knowledge you have provided during this journey.
To HumGen, thank you so much for the support, the answers, the questions, the laughter and food. Every single one of you made so much difference during my PhD journey.
To Ursula Algar, thank you very much for all the work you put in for this project.
To the patients, nurses, doctors and many more individuals who made this project what it is, I thank you.
To Violet Lamola and Gilbert Matanhire (Mantwi le Guy) my number one fans and supporters, the pillars of strength, my reason to remain sane and the sidekick to the "starring" in this journey- Thank you.
To Vangile Lamola (Mami), there are many versions, which this story could have turned out. I am glad o kgethile nna.
To my best friend, the voice of reason and the person who put more fuel to my love of science, my inspiration, Dr. Shareefa Dalvie, thank you so much. The words will never be enough to express how thankful I am for everything you have done for me.
To my best friend Macy Kotola, thank you for your continued support through my journey.
To my journey mate, Horacia Naidoo, we made it!
To Lemogang Seale, this journey would have been twice as hard if I didn't have you by my side, thank you for always putting a smile on my face.
To Thulani Lamola and Nonhlanhla Lamola, you brought laugher and for giving me reasons to become a better person and added to my reasons to get up every morning, thank you.
Thank you to the National Research Foundation (NRF) for the funding.

# Table of Contents

## List of Figures:

## List of Tables:

# Abbreviations

| | |
|---|---|
| A | Adenine |
| ALL | Acute Lymphoblastic Leukemia |
| AML | Acute Myeloid Leukemia |
| APC | Adenomatous Polyposis Coli |
| ARID1A | AT-rich interaction domain 1a |
| bMMRD | Biallelic mismatch repair deficiency |
| BRCA1 | Breast cancer 1 |
| BRAF | B-raf proto-oncogene |
| C | Cytosine |
| CAL | Café au lait |
| CHEK2 | Checkpoint kinase 2 |
| cM | Centimorgan |
| CML | Chronic Myeloid Leukemia |
| CMMRD | Constitutional Mismatch Repair Deficiency |
| COSMIC | Catalogue of Somatic Mutations In Cancer |
| COX | Cyclooxygenase |
| CpG | CpG island is a short stretch of DNA in which the frequency of the C-G sequence |
| CRC | Colorectal cancer |
| dbSNP | Database of Single nucleotide polymorphisms |
| DNA | Deoxyribonucleic acid |
| ECM | Extracellular matrix |
| EDTA | Ethylenediaminetetraacetic acid |
| emPCR | Emulsion-based clonal amplification |
| EPCAM | Epithelial cell adhesion molecule |
| EXO1 | Exonuclease 1 |
| FANC | Fanconi Anemia Complementation Group |
| G | Guanine |
| GBM | Glioblastoma |
| gDNA | Genomic DNA |
| GO | Gene ontology |
| GWAS | Genome-wide association study |
| H&E | Hematoxylin and Eosin stain |
| HREC | Human research ethics committee |
| IHC | Immunohistochemistry |
| ILS | Internal lane standard |
| Indel | Insertion-deletion |
| Kb | Kilo-bases |
| KO | Knock out |
| KRAS | K-Ras proto-oncogene |
| LOH | Loss of heterozygosity |
| LS | Lynch syndrome |
| MAF | Minor allele frequency |
| MLH1 | Mutl homolog 1 |
| MLH3 | Mutl homolog 3 |
| MMR | Mismatch repair |
| MSH2 | Muts alpha 2 |
| MSH3 | Muts alpha 3 |
| MSH6 | Muts alpha 6 |
| µg | Microgram |
| µL | Microlitre |
| MSI | Microsatellite instability |
| MSI-H | Microsatellite instability high |

| | |
|---|---|
| MSI-L | Microsatellite instability low |
| MSS | Microsatellite stable |
| MWM | Molecular weight marker |
| NAT10 | N-acetyltransferase 10 |
| NF1 | Neurofibromatosis type i |
| ng | Nano gram |
| NGS | Next generation sequencing |
| NHL | Non-hodgkin's lymphoma |
| nM | Nano mole |
| NPC | Sample code =nonpolyposis cancer |
| OMIM | Online mendelian inheritance in man |
| P53 | Tp53 or tumor protein |
| PCR | Polymerase chain reaction |
| PIK3CA | Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha |
| PKA | Protein kinase a |
| PKC | Protein kinase c |
| PLEC | Plectin |
| PPI | Protein -protein interactions |
| PTEN | Phosphatase and tension homolog |
| QC | Quality check |
| RBC | Red blood cell |
| RNA | Ribonucleic acid |
| S.A. | South Africa |
| SHH | The hedgehog signaling pathway |
| SIRT1 | Sirtuin 1 |
| SNP | Single nucleotide polymorphisms |
| SPNET | Supratentorial primitive neuroectodermal tumor |
| T | Thymine |
| $T_a$ | Annealing temperature |
| TBE | Tris/borate/edta buffer |
| TE | Tris-edta |
| $T_m$ | Melting temperature |
| UBC | Ubiquitin c |
| U.S.A. | United States of America |
| UTR | Untranslated region |
| VCF | Variant call format |
| VEGF | Vascular endothelial growth factor |
| WES | Whole exome sequencing |
| WGS | Whole genome sequencing |
| WHO | World health organisation |
| WT | Wild type |

# 1. Chapter 1: Introduction to Cancer

## 1.1 Cancer in general

Cancer is the most common death-causing disease in the world, with more people dying of cancer than Acquired Immune Deficiency Syndrome (AIDS), tuberculosis and malaria combined. In 2012, 14.1 million new cases were diagnosed and an estimated 8.2 million people died of cancer (http://canceratlas.cancer.org/ the-burden/). Cancer occurs as a result of the accumulation of multiple mutations that promote clonal selection of cells with increased aggressive growth (Fearon 1997). However, over the years it has been established that both biology and the environment play an important role in the process of carcinogenesis (Fearon 1997). Biologically, the development of cancer has been described as a series of steps, known as hallmarks of cancer (discussed in the section below), which involve genetic alterations that transform a cell from normal to cancerous. This process has made an enormous contribution to the understanding of cancer. Additionally, technological developments in molecular genetics have not only allowed for the identification of the genetic factors associated with disease initiation, but have also led to the discovery of markers of disease progression and prognosis.

## 1.2 Hallmarks of cancer

Over a decade ago, Hanahan and Weinberg (2001) first described the process which became known as the 'hallmarks of cancer' (**Figure 1.1**). The process starts with the acquisition of an initiating mutation (either germ-line or somatic) and the subsequent accumulation of further mutations. These mutations lead to changes in gene activity which in turn, provide the growing population of cells the ability to: disregard the normal control of proliferation, evade growth suppressors, activate invasion/metastasis, enable replicative immortality, induce angiogenesis and resist cell death (Hanahan and Weinberg, 2011). As cancer progresses, the surrounding microenvironment also changes allowing the cells to proliferate (Pietras and Ostman, 2010), evading the typical control of territory and proliferation (Ponder, 2001). In 2011, Hanahan and Weinberg added 'emerging hallmarks', such as deregulating cellular energetics and avoiding immune destruction, as well as 'enabling characteristics'; which included tumour-promoting inflammation and genome instability. Of interest is the role

of the enabling characteristics in tumorigenesis, particularly genome instability. Genomic instability has been described previously as being able to both initiate the development and facilitate the progression of tumours (Negrini et al., 2010). This process leads to the accumulation of mutations, which often compromise the important repair systems. These mutations often occur in the caretaker genes that function as: 1) damage/alteration detectors, 2) employers of mechanisms of correction, and 3) inactivating mutagenic molecules before the damage occurs. This characteristic is important in inherited cancers and will be discussed further in the next section.

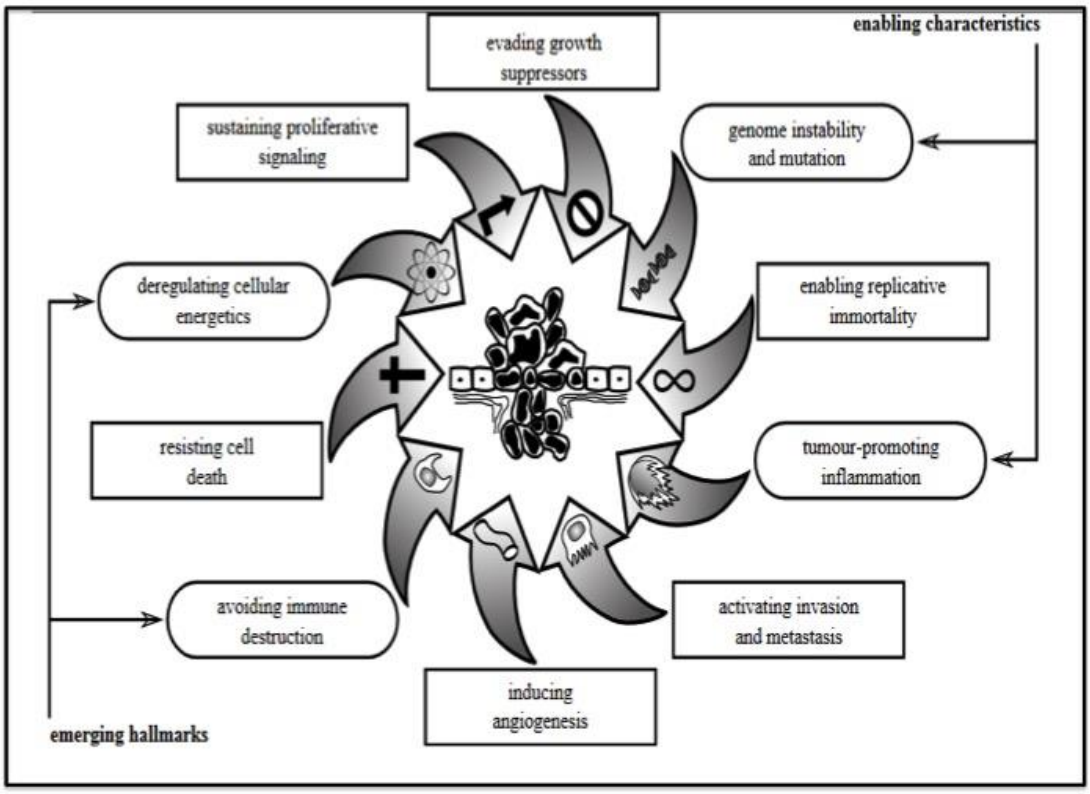



**Figure 1.1: Hallmarks of Cancer** [described by Hanahan and Weinberg (2011)]. The figure depicts the six original hallmarks of cancers as well as the four, later described, enabling characteristics and emerging hallmarks. **https://commons.wikimedia.org/wiki/File:Hallmarks_of_cancer.svg**

## 1.3 Genomic instability and cancer development

Chromosomal instability (CIN) (which is a component of genomic instability) is described as the high rate of change in chromosomal structure over time in neoplastic cells, compared to normal cells (Arends, 2013). The other form of genomic instability is microsatellite instability (MSI/ MIN) characterised by the changes in the number of nucleotide repeats per microsatellite sequence

(Arends, 2013). In inherited cancers, genomic instability has been associated with the accumulation of germline mutations in the DNA repair genes (Negrini et al., 2010). DNA repair genes encode for proteins that form part of the mismatch repair system, whose function is to correct mistakes or alterations, which occur during the DNA replication process (Corcos, 2012; Martín-López and Fishel, 2013).

Negrini et al., (2010) suggested that in inherited cancers, genomic instability is probably the initiating event, which then facilitates the establishment of all of the other hallmarks. In accordance with the mutator phenotype, the presence of genomic instability precedes acquisition of mutations in oncogenes and tumour suppressor genes. For instance, the deactivation of the mismatch repair (MMR) system can lead to the genome-wide accumulation of DNA replication errors generally, or at specific runs of nucleotide sequences resulting in microsatellite instability (MSI) (Corcos, 2012; Martín-López and Fishel, 2013). One of the most commonly inherited cancer syndromes associated with this phenomenon is Lynch syndrome.

### 1.4 MMR deficiency – Introduction to Lynch Syndrome

Lynch syndrome (OMIM: 120435) is the most common form of inherited cancer (Loeb, 2001). Warthin first described this syndrome about a century ago, after observing a large family with several cases of colorectal cancers (CRC) which manifest without the premonitory polyps characteristic of familial polyposis (Warthin, 1925; Lynch et al., 2009; Boland and Lynch, 2013). This syndrome is characterized by an increased risk of cancer, most commonly CRC. While individuals diagnosed with Lynch syndrome have about 70% - 85% lifetime risk of developing CRC, they also have an increased risk of developing other cancers (Lynch et al., 2003; Barrow et al., 2013; Lynch et al., 2015). Since CRC is the most common type of cancer in Lynch syndrome, much research has been conducted in understanding the factors including pathways, involved in disease development. A subset of these pathways will be discussed further in the next section.

### 1.4.1 Pathways involved in the development of CRC

Several pathways involved in colorectal carcinogenesis have been identified over the years (Arends, 2013). Three main contributors to the disease progression are: 1) chromosomal instability (CIN), 2) microsatellite instability (MSI), and 3) CpG island methylation-mutator phenotype (CIMP) (resulting in gene silencing). The chromosomal instability or CIN, often associated with mutations in the *Adenomatous Polyposis Coli* (*APC)* gene, is the major pathway. The gene product, APC, is an authentic gatekeeper of cell proliferation in the colonic epithelium. Microsatellite instability or MSI is often associated with a defective or altered DNA repair pathway. The CIMP, which involves the hyper-methylation of the promoters of genes leads to the transcriptional silencing of potentially a number of genes, representing the key epigenetic mechanism of inactivation of tumour suppressor genes.

#### *1.4.1.1 Chromosomal instability (CIN)*

CIN is often associated with mutation of the *APC* gene. The most commonly identified mutations lead to a truncated inactive APC protein. APC inactivation represents a major pathway of adenoma formation in CRC. The APC protein regulates WNT signalling by interacting with β-Catenin. As a result of the defective APC protein, β-Catenin is not degraded, resulting in the accumulation of this latter molecule. The accumulation of β-Catenin leads to activation of the WNT/APC/β-Catenin signalling pathway. The WNT signalling pathway has many targets - it is involved in many processes (including cell proliferation), which are affected by these changes, and which may result in cancer development (Fearon, 1997; Chung, 2010; Arends, 2013).

#### *1.4.1.2 Microsatellite instability (MSI)*

As mentioned previously, the presence of MSI is a result of mutations in any one of the DNA MMR genes. Mutations in any of the DNA MMR genes (*DNA MMR pathway*), result in the mutator phenotype, which renders the genome 'error prone'. Repetitive regions, such as microsatellite sequences, have been shown to preferentially accumulate these alterations, resulting in expansions or contractions of the sequences, hence the appellation 'microsatellite instability'

(Dietmaier et al., 1997; Imai and Yamamoto, 2008; Jiricny and Pen, 2015). This phenomenon is often observed in inherited CRC or MMR-deficient cancers.

#### *1.4.1.3 The CpG island methylator phenotype*

In sporadic CRC, MSI is often a result of the hyper-methylation of the *MLH1* promoter, which affects gene expression, and which is an example of the CIMP (Hitchins et al., 2005; Hitchins, 2013; Lynch et al., 2015). However, methylation of other promoters have been identified for genes such as *MGMT, PTEN*, *and DNMT3B* involved in the WNT/APC/β-Catenin signalling pathway – and known to be involved in colorectal carcinogenesis (Fearon, 1997; Felsberg et al., 2011; Arends, 2013).

#### *1.4.1.4 Other CRC-associated pathways*

Other associated pathways include the *sessile neoplasia pathway*, which involves the transition of hyperplastic polyps to serrated adenomas then to adenocarcinomas, with most lesions appearing in the right or proximal bowel (often also associated with Lynch syndrome) due to mutations or promoter hyper-methylation (Bansidhar and Silinsky, 2012). The *serrated neoplasia pathway* referred to as the 'traditional serrated neoplasia pathway' is associated with the *CIN/KRAS* mutations (Arends, 2013) (**Figure 1.2).**

KRAS-activation mutations occur early on in the adenoma progression. These mutations affect the enzymatic function of this protein, reducing the cleavage of termination of the phosphate group. Failure to dephosphorylate this compound leads to locking of the KRAS protein into the active state, mediating the excessive signalling through the RAS/RAF/MEK/ERK signalling pathway, which then drives cell proliferation (Arends, 2013). However, KRAS also feeds into the PI3K pathway with activation of serine/threonine kinase (*AKT)* signalling which supresses apoptosis. This pathway can also be activated by *PIK3CA*-activating mutations or it may be inactivated by deletions, mutations or promoter methylation of *PTEN* (Arends, 2013).

Another important role player in colorectal carcinogenesis is the p53 and TGF-β signalling pathway. Both p53 and TGF-β are key tumour suppressors, which

regulate a number of cellular responses (Pardali and Moustakas, 2007). TGF-β signals via the SMAD signal-transduction pathway, while p53 and SMAD physically interact and work together to induce transcription of a number of key tumour suppressor genes. Ultimately, mutant p53 generally subverts tumour suppressive TGF-β responses, diminishing transcriptional activation of key TGF-β target genes. Mutant p53 can also interact with SMAD and this enables complex formation and blocks p53-mediated activation of metastasis-suppressing genes to promote tumour progression (Baker et al., 2006; Arends, 2013). All of these pathways play important roles in colorectal carcinogenesis; however, they are also involved in the development of cancers in other organs and tissues, some of which form part of the Lynch syndrome spectrum (of cancers).

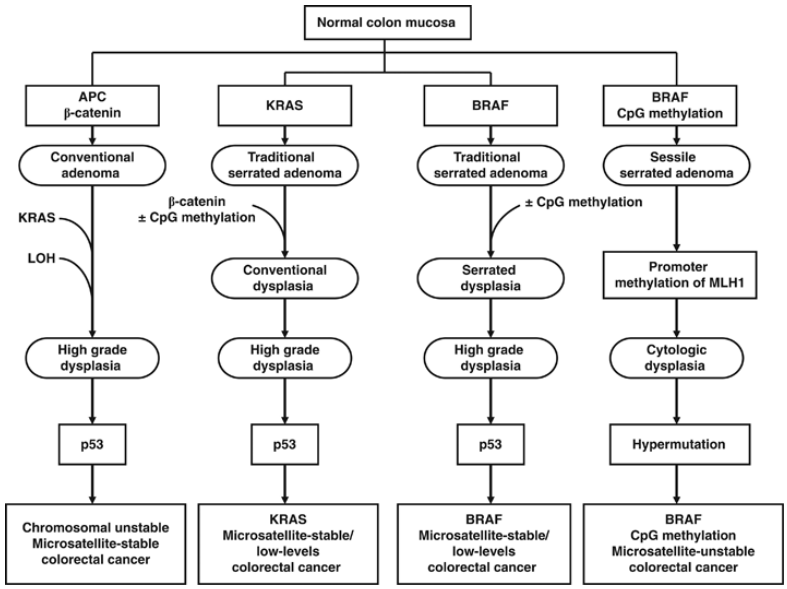



**Figure 1.2 Molecular pathways to CRC:** The figure depicts the four most common pathways associated with the development of CRC. Firstly, the APC pathway, which involves KRAS (loss of heterozygosity) and, lastly, p53 resulting in chromosomal instability but microsatellite stable CRC. The KRAS and BRAF pathways, which both involve promoter methylation and p53 inactivation, result in conventional and serrated dysplasia pathways, respectively. The final pathway involves BRAF CpG methylation, which also involves the *MLH1* promoter methylation causing a Lynch syndrome-like phenotype which is often microsatellite (high) unstable (MSI-H) CRC.

### 1.4.2 Molecular basis of Lynch syndrome

Lynch syndrome occurs as a result of a deficient MMR system. The MMR system has proof reading functionality, recognizes and repairs errors which occur during DNA replication. Mutations of the genes coding for components of the MMR system can lead to suboptimal post-replication error detection and repair

(Kosinski et al., 2010), resulting in the embedding of a higher than normal rate of errors into the genome (Fearon, 2011).

Many types of mutations in MMR genes have been identified and implicated in predisposition to Lynch syndrome (Tang et al., 2009; Borràs et al., 2010; Barrow et al., 2013; Lynch et al., 2015). These mutations include the frame-shift, nonsense and splicing mutations, which result in truncated proteins. The International Society of Gastrointestinal and Hereditary Tumours (InSIGHT) have created a database which catalogues most reported mutations (Peltomaki 2003; http://insight-group.org ). Alterations in five genes within the MMR system have been associated with Lynch syndrome, including the *MutS homolog2 (MSH2), MutS homolog6 (MSH6)*, *MutL homolog1 (MLH1), MutL homolog3 (MLH3)* and *Post-meiotic Segregation increased2 (PMS2)* (Cheah 2009; Peltomäki 2001; Peltomaki 2003). However, mutations in just *MLH1* and *MSH2* account for over 90% of disease-causing mutations in Lynch syndrome (Peltomaki 2003; Clarens, 2004).

In the MMR system, to initiate the repair process, the MutS proteins, namely MSH2 and MSH6, or MSH2 and MSH3, form a heterodimer (MutS-α), which recognizes (DNA) single base-pair mismatches (insertions and deletions) (**Figure 1.3**). Two to four base pair mismatches (insertions and deletions) are recognized by a heterodimer of MSH2 and MSH3 (Rustgi, 2007). Following the recognition of the mismatch, the MSH2/MSH6 complex binds to the DNA base-pair mismatch, the heterodimer recruits other complexes, such as the MLH1 and PMS2, which in turn trigger a series of events which lead to the excision of the mutated strand (Rustgi, 2007). Exonuclease 1 (EXO1) is involved in binding the heterodimers together during this process (Rustgi, 2007). A defective MMR system would result in the accumulation of somatic mutations in the rest of the genome during replication; such a mutation within either a tumour suppressor or oncogene leads to the subsequent development of Lynch syndrome-associated cancers (Calvert and Frucht, 2002; Martín-López et al., 2012).

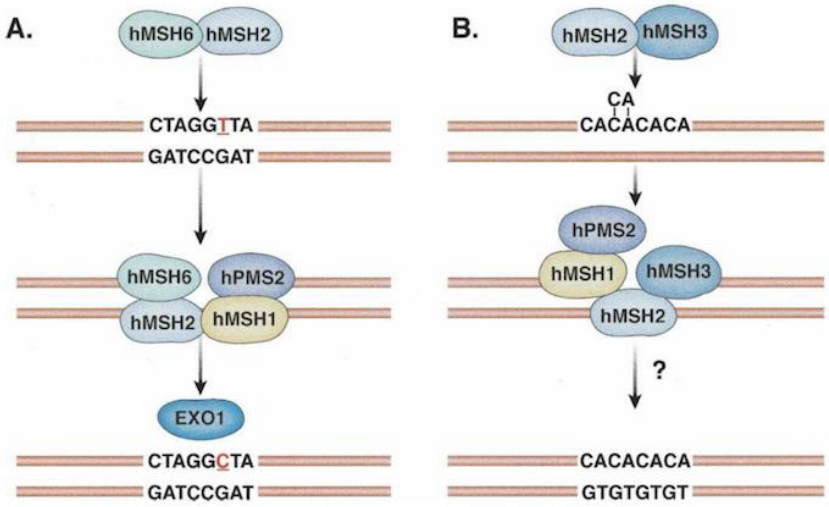



**Figure 1.3 Schematic representation of the MMR model in repairing DNA damage.** A) Depicts the key MMR proteins involved in correcting single base-pair errors; while B) Depicts those involved in the correction of 2-4 bp insertions and deletions (together with single base-pair mismatches).

### 1.4.3 MMR genes

The DNA MMR system is conserved in bacteria and humans, indicative of its evolutionary importance in the functioning of most life forms. There are a number of eukaryotic homologs of bacterial MutS and MutL. In prokaryotes MutS-homologs include; MSH1, 2, 3, 4, 5 and 6. The MutL-homologs include: MLH1, PMS1, MLH2 and MLH3. The human homologs for these proteins include MSH2, MSH6, MSH3, MLH1 and PMS2 for MutS and MutL, respectively (Iyer et al., 2006). The difference between the prokaryotes and eukaryotes is in how these proteins perform their functions. In prokaryotes, the MutL and MutS systems seem to function as homo-dimers, whereas in eukaryotes they function as heterodimers. For instance, MutS α consists of the heterodimer formed between MSH2 and MSH6. The MutS β is made up of MSH2 and MSH3. A similar functional need has been reported for the MutL α, where a heterodimer between MLH1 and PMS2 is formed in order for the protein to perform its function. The MutS α, binds to the DNA mismatches and short insertion/deletion loops (IDLs), and MutS β binds to the larger IDLs. The MutL α is a mismatch-specific endonuclease, and is the intermediary for activation of the down-stream mismatch gap repair (Iyer et al., 2006).

#### *1.4.3.1 MutL α (PMS2 and MLH1)*

Post-meiotic segregation increased-2 (PMS2) is a post replicative DNA mismatch-repair protein. The primary function of PMS2 involves its heterodimer formation (MutL α) with the MLH1 protein, and this complex then interacts with other complexes bound to the mismatched bases. The process of repair for these proteins is initiated by the MutS α or MutS β, binding to the double-stranded DNA. MutL α is recruited to the hetero-duplex. The assembly of the MutL-MutS hetero-duplex ternary complex in the presence of Replication Factor C Subunit 1 *(*RFC1), a five subunit DNA polymerase accessory protein, and Proliferating Cell Nuclear Antigen (PCNA), a cofactor of DNA polymerase delta, is sufficient to activate the exonuclease activity of PMS2, which introduces single-strand breaks near the mismatch, thereby facilitating new entry points for EXO1 for the strand containing the mismatch. PMS2 and MLH1 interact physically with other subunits of DNA polymerase III suggesting that they may play a role recruiting the polymerase to the site of the mismatch (http://www.genecards.org/).

#### *1.4.3.2 MutS α (MSH6 and MSH2)*

Both MSH6 and MSH2 proteins help with the recognition of mismatched nucleotides prior to their repair. These proteins contain a conserved region of about 150 amino acids called a "Walker-A-adenine" nucleotide-binding motif. The encoded proteins hetero-dimerize with each other to form a MMR-recognition complex (MutS α). This heterodimer functions as a bidirectional molecular switch, exchanging the ADP and ATPase; resulting in a conformational transition that converts MutS α into a sliding clamp capable of hydrolysis-independent diffusion. This transition is crucial for MMR. MutS α may also play a role in DNA homologous-recombination repair (http://www.genecards.org/ ).

### 1.5 Cancer risk in Lynch syndrome

Lynch syndrome-related CRC accounts for about 1-5% of all CRC, worldwide (Barrow et al., 2013). Characteristically, these cancers develop on the ascending or right side of the colon. In addition, Lynch syndrome-related CRC may be

synchronous, meaning two or more histologically distinct primary cancers developing at the same time, or metachronous, i.e. separate multiple primary cancers developing over time. Histologically, these tumours are mucinous with signet ring cells; they have poorly differentiated cell types and have infiltrating lymphocytes; these characteristics possibly influence the survival advantage of MMR-related CRC patients compared to those that are sporadic, or non MMR-related (Novelli, 2014; Lynch et al., 2015; Walsh, 2015). Although cancers of the colorectum and endometrium are the most common in Lynch syndrome, the risk of developing other (extra-colonic) cancers is significantly higher than in the background population (Win et al., 2012).

### 1.5.1 Lynch syndrome spectrum of cancers

Individuals diagnosed with Lynch syndrome have about a 70 to 85% chance of developing CRC throughout their lives. Endometrial cancers, which are the most common extra-colonic cancers in the Lynch syndrome spectrum (females have a 30-40% lifetime risk) occur at a higher rate than in the background population (Kamat, 2013). Other cancers such as small bowel cancer are considered rare (1-4% lifetime risk) in Lynch syndrome, but are significant when compared to an almost 0.01% lifetime risk in the background population (Win et al., 2012; Win et al., 2013). Other cancers suggested to be part of the Lynch syndrome spectrum include pancreatic, breast, prostate and rare adrenocortical tumours, as these have been reported to occur at a higher frequency in MMR-mutation positive individuals compared to the background population (Win et al., 2012; Win et al., 2013).

## 1.6 Diagnosis of Lynch syndrome

A number of guidelines have evolved to ensure optimal diagnosis and management of Lynch syndrome, mostly towards early detection of neoplasms in those at highest risk. Initially, the Amsterdam criteria (**Table 1.1**) did not include presence of extra-colonic cancers and MSI. Following the identification of the molecular basis of Lynch syndrome in 1994, the criteria were reviewed and revised (Lynch et al., 2009). In 1996, the Bethesda criteria was introduced and included all clinical characteristics of Lynch syndrome; i.e. early age of onset,

MSI, extra-colonic tumours, trend of first- and/or second-degree relatives diagnosed with Lynch syndrome-related cancer at an early age. **Table 1.1** shows the list of characteristics used to select patients for mutation analysis (Lynch et al., 2009). Currently, only when a mutation in one of the MMR genes is identified, will the individual be diagnosed with Lynch syndrome, and blood relatives considered to be at risk.

**Table 1.1 Amsterdam I, Amsterdam II, and Bethesda Guidelines for diagnosis of Lynch syndrome** (Vasen et al., 1999; Rodrrguez-Bigas et al., 1997; Umar et al., 2004)

| |
|---|
| **Amsterdam I Criteria** |
| 1. An affected individuals should be a first-degree relative of other two other affected relatives<br>2. At least two successive generations are affected<br>3. At least one of the relatives with CRC is diagnosed at <50 years of age<br>4. FAP has been excluded |
| **Amsterdam II Criteria** |
| 1. An affected individual should be a first-degree relative of two other affected relatives<br>2. At least two successive generations are affected<br>3. At least one of the Lynch syndrome-associated cancers should be diagnosed at <50 years of age<br>4. FAP should be excluded in any CRC cases<br>5. Tumours should be verified by pathology whenever possible |
| **Bethesda Guidelines for testing of colorectal cancers** |
| 1. CRC diagnosed in a patient who is <50 years of age;<br>2. Presence of synchronous or metachronous colorectal or other LS-associated tumours, regardless of age;<br>3. CRC with MSI-high (MSI-H) histology diagnosed in a patient who is <60 years of age;<br>4. CRC or Lynch syndrome-associated tumour diagnosed <50 years of age in at least one first-degree relative |

The description of the molecular basis of Lynch syndrome has revolutionized the understanding and diagnosis of this disorder. For instance, the role of characteristics such as MSI associated with tumours and the use of immunohistochemistry (IHC) to detect the (lack of) expression of MMR proteins in tumour tissue have made the diagnosis more accurate and specific (Boland and Lynch, 2013; Zhang and Li, 2013). Together, these molecular tests (i.e. protein expression and MSI testing) provide an indication of the gene most likely to harbor a pathogenic mutation. For IHC, the loss of expression of one of the proteins implies the presence of a pathogenic mutation in that particular gene (Stone, 2001). For instance, loss of expression of *PMS2* and *MSH6* is indicative of

the possible pathogenic mutation in either gene, respectively. However, a dual loss of expression of *MLH1* and *PMS2* would suggest a mutation in *MLH1* as PMS2 is destabilized by the absence of the MLH1 protein (Mohd et al. 2006; Lynch et al. 2015). A similar principle applies to the dual loss of MSH2 and MSH6, where the corresponding mutation will most likely be in *MSH2*, because the absence of the MSH2 protein destabilizes MSH6 (Lynch et al., 2015).

Although MSI is generally reflective of mutations in an MMR gene, a proportion of sporadic cancers may also manifest this feature. These sporadic tumours are sometimes the result of somatically acquired hyper-methylation of the promoters of either *MLH1* or *MSH2* (Hitchins et al., 2005; Hitchins et al., 2011; Hitchins, 2013); other sporadic cancers may also be associated with the presence of somatically-acquired mutations within the *p*roto-oncogene, *Serine-threonine kinase (BRAF)* and *Epithelial Adhesion Molecule (EPCAM)* genes (Rajagopalan et al., 2002; Hitchins et al., 2005; Hitchins, 2013). However, the diagnosis of Lynch syndrome is normally confirmed by the presence of germline mutations in the MMR genes and presence of significant familial clustering (Boland and Lynch, 2013); in this instance, sporadically-acquired mutations resulting in MSI tumours are excluded.

### 1.7 Treatment of cancers associated with Lynch syndrome

Patients with Lynch syndrome may be treated with radiation and/or chemotherapy as an adjuvant to reduce the tumour mass before the surgery. Opinions about the usefulness of *5 Flurouracil* (5FU) as an adjuvant chemotherapy have been contradictory (Zhang et al., 2015), with most reports showing that microsatellite stable tumours respond better to the administration of 5FU, compared to MSI tumours (Ribic et al., 2003; Sargent et al., 2010; Sinicrope et al., 2011). A recent report suggested that the combination of 5FU with one of the platinum-based drugs (e.g. Oxaliplatin) as an adjuvant, increased cancer-free survival period in patients with MSI tumours (Lynch et al., 2015). However, more studies need to be done to assess the effectiveness of the various drugs between the hereditary MMR-related and sporadic cancers.

From a preventative perspective, aspirin has been shown to significantly delay the onset of neoplasms in patients with Lynch syndrome (Macrae et al. 2008; Burn et al., 2010; Barrow et al., 2013; Boland, 2013). Regular use of aspirin in combination was reported to reduce the risk of developing CRC and adenomas (Barrow et al., 2013). Furthermore, aspirin has shown promising results with regards to reducing development of metastatic CRC in individuals at risk of developing CRC (Rothwell et al., 2010; 2011). Current trials are aimed at determining the minimal and optimal doses particularly because of the major side effect of high doses of aspirin resulting in gastrointestinal bleeding (Lynch et al., 2014). Thus, for the moment, the recommendation of aspirin should be taken with caution.

### 1.7.1 Surveillance and management of Lynch syndrome

A few measures are recommended for families with Lynch syndrome with the aim of preventing or reducing tumour development. These measures have been divided into longitudinal surveillance programs and surgical prophylaxis (**Table 1.2**). For monitoring purposes, a colonoscopic surveillance program for those testing mutation positive has been the most successful, as reflected by a number of studies that have shown early detection and an extended cancer-free survival period (Stupart et al., 2009; Barrow et al., 2013; Zhang et al., 2015; Leenen et al., 2016). For extra-colonic cancers, a range of surveillance programs have been proposed (Lindor et al., 2006). For instance, for gynaecological tumours, which are the most common in females, recommended examinations include trans-vaginal ultrasound, hysteroscopy and endometrial biopsy, at two yearly intervals (Dreyer et al., 2012). However, there is not enough evidence to support the benefits of performing any of the latter procedures (Lindor et al., 2006).

**Table 1.2 Lynch syndrome cancer surveillance recommendations** (Adapted from a systematic review by Zhang et al., 2015 & Lynch and de la Chapelle 2003)

| Cancer | Recommendations | From Age |
|---|---|---|
| **Colon cancer** | Colonoscopy every 1–2 years | 20–25 years |
| | Prophylactic colectomy generally not recommended | |
| **Endometrium/ovary** | Consider prophylactic TAH/BSO | After childbearing is complete |
| **Upper urinary tract** | Annual urinalysis | 30–35 years |
| **Upper GI tract Other** | Consider EGD every1–2 years | 30–35 years |
| **Other** | Annual physical exam | 21 years |
| | ROS for related cancers | |
| | Skin exam | |

TAB= Total abdominal hysterectomy; BSO= Bilateral salpingo oophorectomy;
EGD= Esophogogastroduodenoscopy; ROS= Reactive oxygen species

For gastric cancers, gastroscopy is recommended once every two years, from the age of 30-35 years (Zhang et al., 2015). For surgical prophylaxis, subtotal colectomy or segmental resection is advised for individuals diagnosed with Lynch syndrome-associated CRC (Stupart et al., 2011; Vasen et al., 2014). Total colectomy has been recommended as an option for individuals who find it hard to bear the pain and discomfort of the colonoscopy. Total colectomy has been advised for young individuals (less than 45 years) who have developed primary CRC, to reduce the development of a secondary CRC (Barrow et al., 2013). For other cancers, prophylactic hysterectomy is advised for individuals who have passed childbearing age (Lindor et al., 2006). This is to reduce the risk of developing both ovarian and endometrial tumours.

### 1.8 Lynch syndrome in South Africa (S.A.)

In S.A., Lynch syndrome as a disease is not common. However, cancers that are related to Lynch syndrome are common. For instance, CRC is among the top six cancers leading to deaths in this country. The incidence of CRC for males, according to the 2010 (S.A.) National Cancer Registry report, ranges from 8 to more than 50 per 100 000, within the age range of 45-75 years, and from 4 to 25 per 100 000 for females in the same age range (http://www.nioh.ac.za/assets/files/NCR Final 2010 tables(1).pdf; National Cancer registry 2010). Although most of the reported cases of CRC may be of the sporadic (not inherited) form, a recent study of MMR deficiency-associated CRC in the Northern Cape Province of S.A. indicated that there was a significantly high proportion (21.8%) of tumours with MMR deficiency (Vergouwe et al., 2013). This report is indicative of a possible increased rate of inherited/ MMR-deficient CRC within the low incidence disease areas.

Other Lynch syndrome-related cancers, such as endometrial cancer, are yet to be more fully investigated in S.A. Although, in general, the incidence of endometrial cancer was estimated to be about 4/100 000, the cases diagnosed seem to increase with age, from 10/ 100 000 at the age of 50 years to approximately 32/ 100 000 by the age of 74 years. In Lynch syndrome, the cancers generally develop at a much younger age than their sporadic counterparts. Thus,

understanding what types of cancer and how they develop within Lynch syndrome will not only benefit those who are at risk of inherited cancers, but may also have an implication for sporadic forms of cancers within the Lynch syndrome spectrum. It is thought that the research in this area is likely to provide knowledge towards strategizing how resource-limited countries, such as S.A., may reduce morbidity and mortality associated with cancers.

## 1.9 Motivation for study

The Division of Human Genetics at the University of Cape Town (UCT) in S.A. together with Groote Schuur Hospital (GSH) have developed a registry of patients with early-onset CRC (i.e. diagnosed at 50 years of age, or younger). This project began in the 1980s with the aim of identifying the genetic factors associated with early onset inherited cancer in the Northern and Western Cape regions of S.A. The registry and its associated bio-repository capture clinical information on these subjects relating to their medical history, population demographics, pedigree (or family history), disease-causing or predisposing mutation, colonoscopic surveillance, history of neoplasms and related pathology, radiation or chemotherapeutic treatment and general follow-up history, together with biological material (ideally venous blood and tumour tissue). To date (January 2017), the registry has captured details and biological material from more than 700 families who have been recruited. A number of disease-causing mutations have been identified in some of these individuals, including a S.A.-specific mutation (c.1528C>T) in exon 13 of the *MLH1* gene. In addition, this is the mutation which has been identified in the homozygous state (c.1528C>T: c.1528C>T) in a child, who was later diagnosed with Constitutional Mismatch Repair Deficiency (CMMR-D) syndrome, and - described later in Chapter 3 of this dissertation. The number of individuals who had developed the various kinds of cancers, due primarily to the c.1528C>T mutation in *MLH1*, provided the opportunity to interrogate potential pathways involved in disease and rate at which mutations accumulate in the presence of this specific mutation.

The aim of this research was to provide a better understanding of the genetic factors associated with both Lynch and CMMR-D syndromes.

The objectives were:

1. To determine the rate of extra-colonic cancers in a cohort of Lynch syndrome families
2. To provide evidence for the founder effect of the *MLH1* c.1528C>T mutation
3. To focus on the CMMR-D syndrome as a branch of Lynch syndrome and determine the genomics surrounding this phenomenon

The objectives were carried out to address the aim using the workflow in **Figure 1.4.**

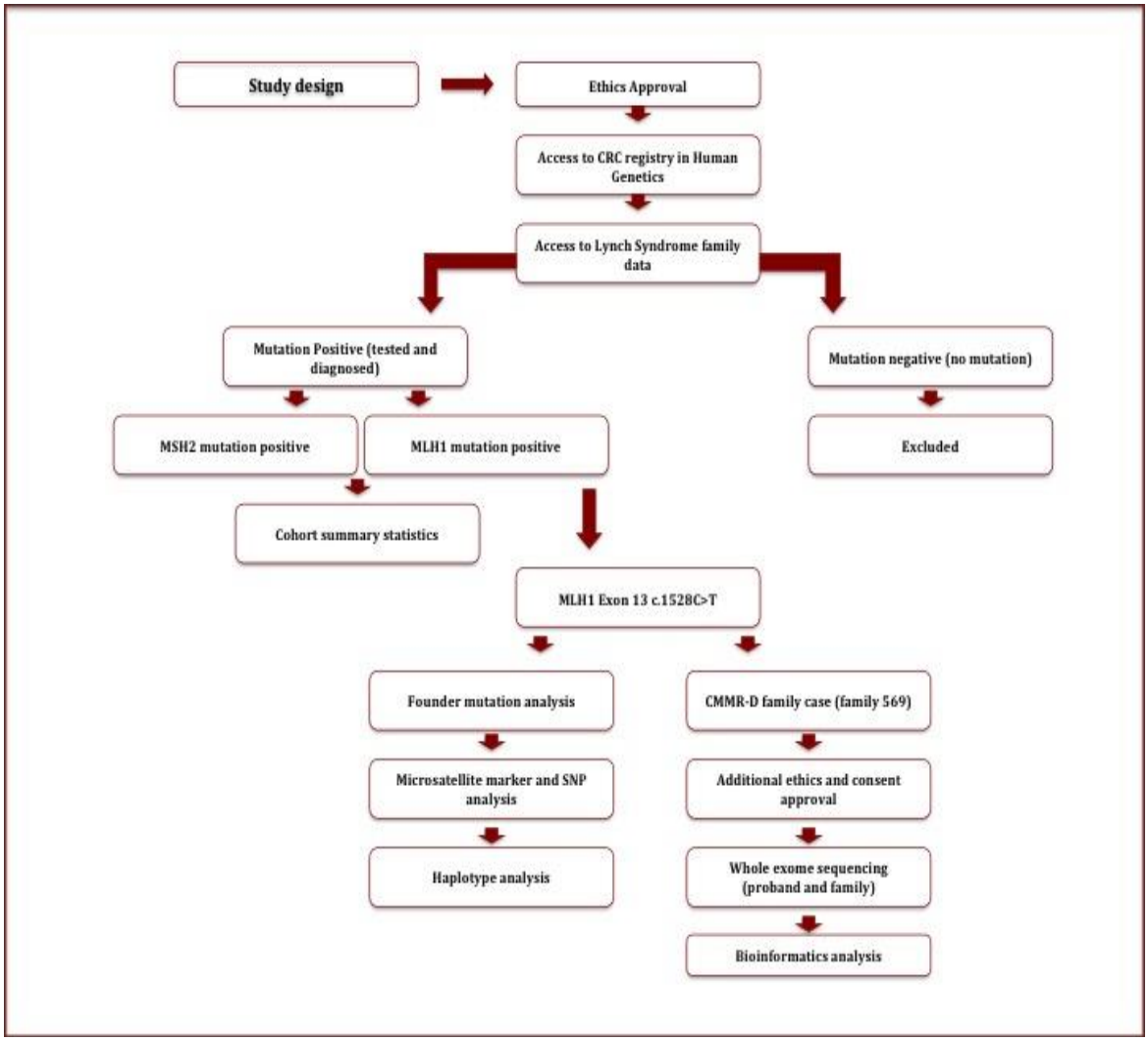



**Figure 1.4** Illustration of the strategy and workflow of the research described in this dissertation.

# Chapter 2: Lynch Syndrome - The Cohort Study

## 2.1 Brief background

Previously, Lynch syndrome was thought to manifest as two clinical subtypes; Lynch syndrome type I and type II (https://www.omim.org/entry/120435). Lynch syndrome type I was characterised by autosomal dominantly inherited nonpolypotic CRC with an early age of onset. The CRC was generally site specific (mostly in the ascending/ or right-hand side of the colon) and was observed in multiple generations. Lynch syndrome type II was characterized by families in which other types of malignancy (i.e. extracolonic), such as endometrial and urinary tract cancers occurred, in addition to CRC. The most common types of cancer following CRC in type II Lynch syndrome were those of the endometrium, stomach, ovaries, small intestine, pancreas, urinary tract (i.e. kidney, ureter or bladder), bile duct, skin (usually sebaceous adenomas), and brain. More recently, however, and as data has accumulated longitudinally – it has been recognised that extracolonic cancers are a general feature of Lynch syndrome (Vasen et al., 1999), and that the type I classifier may have had to do with the mode of ascertainment i.e. through a gastroenterologist or related specialist – since CRCs are most common in Lynch syndrome, and due to limited pedigree sizes.

### 2.1.1 Cancer risk

As previously described, mutations in the MMR genes increase the risk of developing a range of cancers in carriers. Over the years there have been a number of studies on genotype-phenotype correlations, where some mutations have been shown to confer a much higher risk of specific cancers within the Lynch syndrome spectrum. However, apart from a high risk for CRC, especially in *MLH1* and *MSH2* mutation-carriers, there is no clear correlation between the MMR gene, the nature/domain affected by the mutation and the range of cancers –especially when one looks at the larger pedigrees or patient collections, internatonally.

In one such study (involving patients from Denmark, Holland, Finland and U.S.A.) Watson et al., (2008) investigated the incidence of cancers in 6041

individuals with *MLH1* and *MSH2* mutations. The most common type of malignancy, after CRC and endometrial cancer, were urologic cancers with an overall lifetime risk of 8.4% in males; and these were more likely to occur in *MLH1* mutation carriers (Watson et al., 2008).

In a separate study of 147 families (55 with mutations in *MLH1*, 81 with mutations in *MSH2* and 11 with mutations in *MSH6*), the cumulative risk for cancer development was 66% in males and 42% in females (Stoffel et al., 2009). CRC occurred more often in males (68.3%) than in females (56.6%) with *MLH1* mutations, and 5.1% of the females had endometrial cancer (at a median age of 47.5 years). Lifetime-risk for CRC and endometrial cancer in MMR-mutation carriers was high even after adjusting for an ascertainment bias (Stoffel et al., 2009). In a more recent study of 2118 cases of MMR-mutation carriers, the seven most common cancers observed in this cohort were those of the stomach, small bowel, urinary tract, breast, ovary and prostate (Engel et al., 2012).

Risk of primary cancers in Lynch syndrome has been the focus of research for many years. However, considering that individuals with this disorder are at risk of developing multiple cancers throughout their lives, it is of value to determine the risk of developing extracolonic cancers. As already mentioned, there is some contention that the range of cancers may be dependent on the actual gene or specific mutation, or a range of modifier factors (both environmental and genetic) (Bansidhar and Silinsky, 2012).

In a cohort of 46 mutation-positive individuals from 22 different families with Lynch syndrome the manifestation of the range of cancers (whether site specific or affecting a wide range of organs) seemed to depend on the actual gene that was mutated, rather than the type of mutation in the gene (Pérez-Cabornero et al., 2013). Similarly, Geocke et al., (2006) observed that *MSH2* mutations were associated with multiple extra-colonic cancers whereas *MLH1*-mutation carriers, and specifically males, had a relatively younger age at diagnosis, (Goecke et al., 2006). In addition, and relatively speaking, there is an increased risk of CRC in *MLH1* and *MSH2* mutation carriers when compared to the other MMR genes, for

example *MSH6* (Pérez-Cabornero et al., 2013). Lastly, individuals with mutations in *MSH6* generally have a later age of onset/diagnosis - almost 10 years later than both *MLH1* and *MSH2* mutation carriers (Pérez-Cabornero et al., 2013) .

In Africa, Lynch syndrome is one of the most understudied disorders, there is not much literature regarding the spectrum of extra-colonic cancers in African populations, which is why the selected background literature mostly focuses on the reported North American and European cases. Work published on a Congolese cohort of 89 individuals with early onset (median age of 35 years) CRC found that 38.2% of the cases had abnormal immunohistochemistry staining for both MLH1 and MSH2 (Poaty et al., 2017). No specific mutations were reported in this study and most of the tested cases were that of CRC and no extra-colonic cancers. In a Tunisian study conducted in 2014, it was reported that of all the suspected Lynch syndrome cases, only 30% of the tested samples were MSI while the 70% were MSS. The loss of expression in these cases was seen in all the common MMR genes (MLH1, MSH2, MSH6 and PMS2) (Amira et al., 2014). In another study from Tunisia, Sana et al. (2014), identified 10 germline mutations in CRC cases in 11 of the 31 HNPCC cases. Most of these mutations were in *MLH1* and *MSH2*, similar to what we have seen here in South Africa and what has been reported internationally (Sana, et al., 2014).

Another recent study by Ziada-Bouchaar (2017), identified 3 novel variants in the MLH1 gene, 4 variants MSH2, a large exon1 deletion in MSH6; none of the identified mutations have been seen in our population. Of interest in this cohort was the fact that there was similar age of onset regardless of the gene that was mutated. This is slightly different from what has been reported elsewhere, where MLH1 and MSH2 have an earlier age of onset compared to other MMR genes, such as MSH6 (Pérez-Cabornero et al., 2013). Often the HNPCC/Lynch Syndrome studies in African cohorts have CRC as the most common cancer, however cases of endometrial cancer have been reported as well, but little is known about other Lynch syndrome-related extra-colonic cancers in African cohorts.

Although there are no absolute phenotypes for Lynch syndrome, family history plays an important role in identifying cases for this disorder. It is true that the set of Bethesda criteria **(Chapter 1; Table 1.1)** is an important tool for

identifying these families; however, some of the extra-colonic cancers have only relatively recently been seen in the very large cohorts, and had not previously been included as part of the Lynch syndrome spectrum of cancers. Such cancers include adrenocortical adenocarcinoma, thyroid carcinoma, peritoneal mesothelioma, malignant fibrohistiocytoma, rhabdomyo-sarcoma, dermatofibrosarcoma, leiomyosarcoma, liposarcoma, carcinoid tumour, non-Hodgkin lymphoma, malignant melanoma, and those of the pancreas, prostate and breast (Lynch, 2003; Lynch et al., 2009). The list that has been included in the Lynch syndrome spectrum of tumours are often those which conceivably have a significantly increased risk as a result of MMR deficiency.

### 2.1.2 The *MLH1* c.1528C>T mutation

A founder mutation or variant is a genetic change that is observed at a higher frequency in geographically or culturally isolated populations. The frequency of Lynch syndrome in the Cape Mixed Ancestry population, specifically in the Western and Northern Cape Provinces of S.A. is high compared to other parts of the country or in other populations. In addition, the observation of a relatively high frequency of the *MLH1* c.1528C>T mutation among the Cape Mixed Ancestry population raises the issue of whether this particular site of the genome is prone to mutagenesis or whether it may be the result of a founder effect, which leads to questions about its origin and distribution in the southern African population.

Hundreds of *MLH1* mutations have been identified in patients with Lynch syndrome in populations around the world, mostly, in Europe (Moisio et al., 1996; Borràs et al., 2010; Pinheiro et al., 2011; Tomsic et al., 2012; Borelli et al. 2014). Many of these mutations have suggested a founder effect due to their frequencies in specific populations. Borreli et al., (2014) investigated 11 unrelated Lynch syndrome families to evaluate the clinical consequence of the *MLH1* exon19 c.2252_2253delAA terminal mutation and to determine if this was a possible founder. The index and a further approximately 300 cases, identified over a period of 12 years, were investigated with 7 SNP and 10 microsatellite markers to determine whether they shared a common origin (or a founder

mutation). The same mutation was previously observed to share founder status in disparate geographic locations including Australia, Germany, the United Kingdom and Denmark (Borreli et al., 2014). The analysis of all accessible individuals with this particular mutation showed a shared haplotype of over 1.7Mb in at least eight of the 11 families and was deemed to be a founder.

Several other founder mutations in *MLH1* have been identified in a number of global populations which explains the widespread geographic occurrence of Lynch syndrome. However, these geographically widespread founder mutations have been described as 'heterogeneous' in function and frequency. Most of the mutations result in a truncated protein, which are almost always confirmed to be pathogenic. However, with the introduction of the next generation sequencing platforms, many more potential mutations have been identified with unknown pathogenicity (i.e. variants of unknown significance), and will prove interesting in understanding the extent of founder mutations and associated phenotypes in future.

A separate study characterised two relatively frequent *MLH1* mutations (c.306+5G>A and c.1865T>A) among Spanish families with Lynch syndrome (Borras et al., 2010). In some Lynch syndrome families around Europe, one of the two mutations was reported to be responsible for 28% of all *MLH1* mutations and represented 17% of all families with mutations in this gene in the Ebro Basin region (which is geographically isolated). Borras et al., (2010), showed that families with specific mutations clustered in specific regions because of geographic isolation. One mutation, *MLH1* c.306+5G>A, was estimated to be about 2000 years old while the other, *MLH1* c.1865T>A, was found to have arisen relatively recently (~384 years) (Borras et al., 2010).

In 2009, Tang et al., investigated 93 Taiwanese Lynch syndrome families. Among the identified mutations was the *MLH1* c.793C>T mutation, in 13 of the families. Genetic analysis showed that the mutation occurred on two separate extended haplotypes, with a shared region of about only 30kb, perhaps suggestive of a distant common origin. The ancestry for this particular mutation was traced

back to mainland China, and was further confirmed by the fact that this particular mutation had been previously reported in mainland Chinese Lynch syndrome patients (Tang et al., 2009).

Other MMR gene mutations, have been observed at relatively high frequencies in specific populations such as Finland (Nyström-Lahti et al., 1994), China (Chan et al., 2004), U.S.A. (Wagner et al., 2003), Canada (Froggatt et al., 1999), Sweden (Cederquist et al., 2005), Korea (Shin et al., 2004) and Italy (Caluseriu et al., 2004). Interestingly, even if the mutations may be shared in these geographically and ethnically diverse populations, the profile of Lynch syndrome-related cancers in these patients may differ slightly. For instance, gastric cancer was the most common extra-colonic cancer in the 93 Taiwanese families, consistent with the reports from other Asian countries (such as Japan, Korea and China), while in Europe the most common extra-colonic cancer is endometrial cancer (Tang et al., 2009).

#### *2.1.2.1 MutL-Homolog1 (MLH1)*

The *MLH1* gene was mapped to chromosome 3p21.3 in 1994 (Bronner et al., 1994) followed by a description of the structure of the gene (Han et al., 1995). The gene consists of 19 coding exons, spanning about 100kb. This gene codes for a protein, which forms a heterodimer with PMS2 (**Figure 1.3, Chapter 1**) to form a Mut-L α complex, a component of the post-replicative DNA MMR-system. The *MLH1* gene was identified as frequently mutated in HNPCC consistent with the alterations detected in microsatellite sequences. Common conditions associated with this gene also include Muir-Torre syndrome (OMIM: 158320), Turcot syndrome (OMIM: 276300) and the more recently described, CMMR-D syndrome (OMIM: 276300). MLH1 is involved in many biological pathways including, the cell cycle, DNA repair, mismatch repair, meiosis and mitosis (Ellison et al., 2004; Cohen et al., 2006). One of the listed functions of this protein includes, ATP-, ATPase- and protein- binding; it also contributes to MutS α complex binding (Plotz et al., 2002, 2006). In addition to the functions already described, this protein is also involved with a number of processes such as double stranded-break repair via non-homologous end joining and in the

intrinsic apoptotic-signalling pathway in response to DNA damage (Schofield and Hsieh, 2003).

Of particular interest is the exon 13 codon 510 mutation, also referred to as *rs63749923* (NM_0002493: c.1528C>T nonsense). The change in base (C>T) results in a premature stop codon, leading to a truncated transcript. This mutation is one of more than 1000 ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/?term=mlh1%5Bgene%5D) *MLH1* mutations, also recognized by InSIGHT (https://www.insight-group.org/variants/databases/), as a Lynch syndrome predisposing mutation (Vasen, 2004).

#### *2.1.2.2 MLH1 c.1528C>T Mutation*

The history of the ascertainment of families with the *MLH1* exon 13 c.1528 C>T mutation began in 1987, when a 30-year old male with a family history of CRC was identified. The CRC seemed to be autosomal dominantly inherited but affecting only males (Goldberg et al., 1998). However, this was largely due to an ascertainment error, since the medical records ascertained from the hospital in the diamond-mining village of Kleinzee in the Northern Cape Province, treated only the workers on the mine – who were generally only men who were bussed in to the mine from a village some 50 kilometres away. In due course with tracking family lineages through to the place of residence of the miners, both males and females were found to have manifest CRC in previous generations. The family fitted the Amsterdam criteria for HNPCC. Molecular genetic studies led to the identification of the pathogenic mutation in exon 13 of *MLH1* in 1994 (subsequently described by Goldberg et al., 1998). The discovery of this mutation has led to the introduction of a programme of genetic counselling and genetic testing for individuals closely related to those previously affected by disease. This programme is aimed at early detection of neoplastic lesions through colonoscopic surveillance of those carrying the disease-causing mutation, and the exclusion from surveillance of those with no risk (mutation negative) for this familial condition. For mutation carriers, colonoscopic

surveillance has been offered with the objective of increasing the cancer-free survival period.

### 2.1.3 Objectives of this section:

1. To analyse the registry with regards to the risk of MMR-mutation carriers developing a range of Lynch syndrome associated cancers
2. To perform survival analysis on the data
3. To investigate whether the *MLH1* exon 13 c.1528C>T mutation is due to a founder mutation.

## 2.2 Method

### 2.2.1 Study cohort

Although the main research focus was to be on the *MLH1* exon13 c.1528C>T mutation positive individuals/families, in the first instance, data on the whole cohort of MMR-mutation positive individuals/families was accessed and analysed. As with all other Lynch syndrome cases, the same mutation may have a variable penetrance in individuals between and within families. However, this c.1528C>T subset of the cohort was thought to be relatively homogenous, which theoretically ought to put them at similar risk of developing disease.

For this section of the study, research ethics approval was previously obtained for the access to patient demographic data and clinical information, which included, disease information, age at onset, and the type of cancer, as well as the pathology reports (pertaining to diagnosed cancers or biopsies of neoplastic tissue).

The third section of this study, investigated the possibility of the c.1528C>T mutation as a founder in 30 probands, each of whom was from a distinct 'family' or 'lineage' with no obvious connection to any of the other pedigrees/families, from the Northern and Western Cape provinces of S.A. segregating the c.1528C>T mutation. Ethical clearance was obtained to genotype probands with this mutation for five previously selected microsatellite markers flanking the mutation in order to determine the haplotypes of these cases. In addition, 98 healthy controls were also genotyped for all the markers to determine their haplotypes. The controls consisted of "unaffected" individuals of Mixed Ancestry from the Western Cape, unrelated to the Lynch syndrome families.

### 2.2.2 Statistical analysis (R)

R version 3.1.3 is an open source statistical and graphic computing software environment. This programme is available with a number of codes and scripts written and ready to be used; these are stored in a number of packages on CRAN (https://cran.r-project.org/web/packages/), which is easily accessible and can be installed and used freely. Currently, the CRAN package repository consists of more than 9000 packages available for use. The package "survival" was

downloaded and installed for use for this part of the study. Survival analysis is a non-parametric method used to estimate the survival function from data using several analysis features where the outcome variable is the time until the occurrence of the event of interest, which can, for example, be disease or death, amongst other factors. The Cox proportional hazard model analysis allows for the estimation of risk of an event (cancer development) for individuals given the prognostic variables (https://cran.r-project.org/web/packages/survival/survival.pdf). For survival analysis, status was set as the "development of cancer" and the gender as well as type of cancer were analysed as independent variables.

### 2.2.3 Microsatellite genotyping and haplotype testing

Microsatellite genotyping was performed to measure the potentially shared region flanking the *MLH1* c.1528C>T mutation. Five markers flanking the c.1528C>T mutation (**Table 2.1**) were identified and selected for use based on their distribution either 5' or 3' to the disease-causing mutation. Following marker identification, the relevant marker-related primers were designed to amplify the fragments of interest using standard PCR, and genotyped. For subsequent resolution on a ABI3130xl Genetic Analyzer (Applied Biosystems; Foster City, CA), one of the primers (per pair) was designed to incorporate a fluorescent label

**Table 2.1 Microsatellite Markers with marker positions relative to the mutation**

| NC_000003.12 GRCh38 accessed 2014-07-18 | |
|---|---|
| **Marker name** | **Position with respect to mutation** |
| **D3S3512** | 34594179 |
| **D3S1561** | 36484119 |
| **c.1528C>T*** | **37028881** |
| **D3S1611** | 37068501 |
| **D3S3623** | 37443531 |
| **D3S3527** | 39345373 |

*Mutation (*MHL1* c.1528C>T)

#### *2.2.3.1 Polymerase Chain Reaction (PCR)*

The PCR is a method used to amplify regions of interest in the genome. The locus-specific primers were used to initialize the synthesis of DNA using the thermo-stable DNA (*Taq)* polymerase, derived from the organism, *Thermus aquaticus*. The reagents required to perform PCR included: template DNA at a concentration of 100ng/µL, 1x GoTaq buffer (Promega, Madison, WI, U.S.A.) at a

pH of 8.5 and $MgCl_2$ (concentration of 1.5mM), 5Units/µL GoTaq DNA polymerase (Promega®, Madison, WI, U.S.A.), 200µM of each deoxynucleoside triphosphate (dNTPs) (BIOLINE, London, UK), a final concentration of 0.4µM primer per reaction and distilled water (Sabax, Adocock Ingram, JHB, S.A.), in a total reaction volume of 25µL in a 0.2mL microfuge tube. The reaction was conducted in a thermal-cycler (BioRad, Hercules, CA, U.S.A.), comprising of cycles of denaturation, primer annealing and extension of the product under conditions set out in **Table 2.2** (Schochetman et al., 1988). All reactions were conducted in multiplex format.

**Table 2.2 Thermo-cycler conditions for standard PCR**

| Step | Temp (°C) | Duration |
|---|---|---|
| **Initial step** | | |
| *Denaturation* | 95 | 5 min |
| **25-30 cycles** | | |
| *Denaturation* | 94 | 30 sec |
| *Primer annealing* | 50-60 | 30 sec |
| *Extension* | 72 | 40 sec |
| **Final Step** | | |
| *Final extension* | 72 | 7 min |

### *2.2.3.2 Automated capillary electrophoresis*

Subsequent to PCR, genotyping was performed on the 3130xl Genetic Analyzer (Applied Biosystems; Foster City, CA). In a standard capillary electrophoresis reaction, a 1 in 10 dilution of the PCR product was made for all amplicons, 1 µL of this was added to 8µL Hi-Di™ formamide (HiDi) (Applied Biosystems, Foster City, CA), the formamide was used to keep the DNA as a single strand (Sambrook and Russell, 2001a). Results were analysed using the GeneMapper® software (Applied Biosystems, Foster City, CA).

### *2.2.3.3 Statistical analysis PHASE II*

PHASE is freely available statistical software that implements a Bayesian statistical method for reconstruction of haplotypes using genotype data (Stephen and Donelly, 2003). This software allows the user to input data generated from SNPs, microsatellites and other multi-allelic loci. For the analysis, data is generated and input files prepared; the analysis can be performed on case only

or case-control data. This regression model could be estimated through Markov Chain Monte Carlo (MCMC) methods. To implement an MCMC simulation, the process starts at some value for the parameter of interest, which can be defined by the user. In order to determine this starting point, a number of 'burn-in' iterations was specified, which is a grace period during which the Markov Chain wanders to the most probable region of the sample space. The samples obtained during the burn-in are discarded. After the burn-in, the simulation will continue to generate samples from the distributions of the parameters of interest for the specified number of iteration times. For example, an MCMC model was run with 300 iterations and a burn-in of 30, indicating 330 iterations; but the first 30 were discarded. The challenge of using MCMC is specifying an appropriate number of iterations and appropriate burn-in. For the analysis, the run was conducted for five microsatellite markers and results were analysed for both cases (30 probands) and controls (98 controls). The run was set to have 1000 burn-in iterations, followed by 1000 iterations and recorded after one hundred iterations.

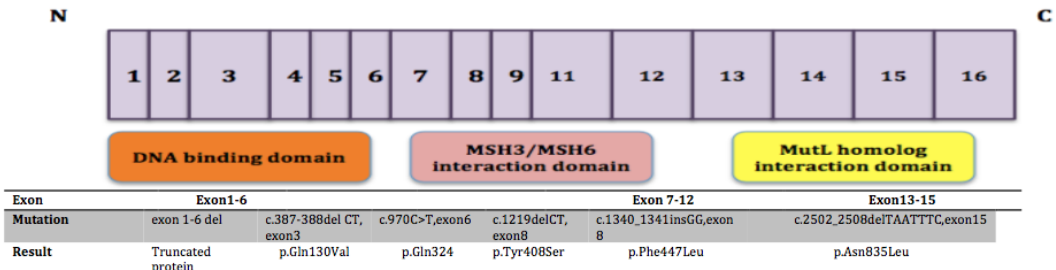
### 2.2.4 Estimating the time of origin or age of c.1528C>T mutation

To estimate the age of the South African (Cape Mixed ancestry) *MLH1* c.1528C>T mutation the DMLE+2.3 software program (www.dmle.org) was used. This programme was designed for high resolution mapping of disease mutation and estimation of its age. The method is based on the observed linkage disequilibrium between the five markers (relative to the mutation) in both cases and controls. The software uses the MCMC algorithm (Hamra et al., 2013) to allow Bayesian estimation of the mutation age, based on parameters including; the observed haplotypes, in samples of unrelated cases or controls, distances between genotyped markers and position of the mutation, the estimated population growth rate and an estimation of the proportion of disease-causing mutation. Following the haplotypic phasing of all 30 probands using Phase version 2.1, the chromosome map distances were obtained from the NCBI (https://www.ncbi.nlm.nih.gov). The population growth rate was obtained from the Statistics S.A. report (https://www.statssa.gov.za/publications/P0302/P03022014.pdf). A 25-year intergenerational interval was used to calculate number of years per generation.

## 2.3 Results

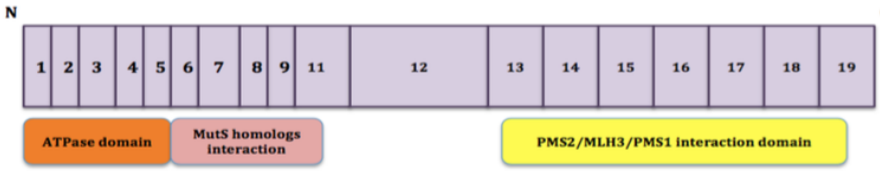
### 2.3.1 Cohort demographics

The Division of Human Genetics at the UCT, with the Division of Gastroenterology at GSH have maintained a registry, which consists of probands and families with early onset (50 years of age or younger) CRC. These families, predominantly from the Northern and Western Cape provinces of S.A., have been enrolled in this research study. The on-going search of the individuals at risk in the Northern and Western Cape provinces, has led to the identification of almost 400 individuals with more than 20 different disease-causing mutations in just two of the MMR genes, namely *MLH1* and *MSH2*. This cohort is sub-divided into those carrying mutations in either *MSH2* (**Figure 2.1)** or *MLH1* **(Figure 2.2).** The *MLH1*/*MSH2* schematic representations show the list of mutations, the domains of the protein affected by the mutation, and types of cancers observed in the cohort. As is evident from **Figure 2.2**, most individuals in this cohort have a mutation in exon13 of the *MHL1* gene.




| Exon | Exon1-6 | | | | Exon 7-12 | Exon13-15 |
|---|---|---|---|---|---|---|
| Mutation | exon 1-6 del | c.387-388del CT, exon3 | c.970C>T,exon6 | c.1219delCT, exon8 | c.1340_1341insGG,exon 8 | c.2502_2508delTAATTTC,exon15 |
| Result | Truncated protein | p.Gln130Val | p.Gln324 | p.Tyr408Ser | p.Phe447Leu | p.Asn835Leu |
| No. With mutation | 4 | 5 | 1 | 15 | 27 | 5 |
| CRC | | Y | Y | Y | Y | Y |
| Endometrium | | Y | | | Y | Y |
| Breast | | Y | | | Y | |
| Ovary | | | | | | Y |
| Small bowel | | Y | | | Y | Y |
| Liver | | | | | | |
| Cervix | | | | | | |
| Prostate | | | | | | |
| Stomach | | Y | | | | |
| Pancreas | | | | | | |
| Testicles | | | | | | |
| Skin | | | | | | Y |
| Urethra | | | | | | |
| Bladder | | | | | | Y |
| Neuroendocrine | | | | | | |
| kidney | | | | | | |
| Bone | | | | | | |
| Thyroid | | | | | | |

**Figure 2.1 Site of cancers associated with the mutations in the *MSH2* gene in the cohort 1**. Indicated are exons (in mauve), domains (orange, pink and yellow), specific genotype of mutations and the type of cancers observed in this cohort. C=coding region, p. =amino acid, N= the N-terminal, C=the N-terminal, Y= yes.




| Exon | exon 1-5 | exon 6-11 | exon 13-19 | | | |
|---|---|---|---|---|---|---|
| Mutation | c.350C>T, exon4 | c.790+1G>C, exon9 | c.1409G>A, intron12 | c.1459C>T,exon13 | c.1521insT, exon13 | c.1528C>T,exon13 |
| Result | p.Thr117Met | p.Glu227_Serdel | aberrant splicing | p.Arg487* | p.Leu507Phe | p.Gln510* |
| No. with mutation. | 6 | 2 | 3 | 1 | 1 | 309 |
| CRC | Y | Y | Y | | Y | Y |
| Endometrium | | | | | | Y |
| Breast | | | | | | Y |
| Ovary | | | | | | Y |
| Small bowel | | | | | | Y |
| Liver | | | | | | Y |
| Cervix | | | | | | Y |
| Prostate | | | | | | Y |
| Stomach | | | | | Y | Y |
| Pancreas | | | | | | Y |
| Testicle | | | | | | Y |
| Skin | | | | | | Y |
| Urethra | | | | | | Y |
| Bladder | | | | | | Y |
| Neuroendocrine | | | | | | Y |
| Kidney | | | | | | Y |
| Bone | | | | | | Y |
| Thyroid | | | | | | Y |
| Lipoma | | | | | | Y |

**Figure 2.2 Site of cancers associated with the mutations in the *MLH1* gene in the cohort 1.** Indicated are exons (in mauve), domains (orange, pink and yellow), specific mutations and the types of cancers observed in this cohort. C=coding region, p. =amino acid, N= the N-terminal, C=the N-terminal, Y= yes.

In total, there were 396 mutation positive individuals identified in the UCT/GSH cohort (cohort 1). Of those, 309 individuals were confirmed to have the *MLH1* exon 13 (c.1528C>T) mutation (from here on referred to as cohort 2). To date, this mutation has only been reported in the Mixed Ancestry population of S.A., specifically located in the Northern and Western Cape Provinces of S.A. The mutation-carrying individuals form part of 30 different families. In the registry, the *MLH1* c.1528C>T sub-cohort (cohort 2) is comprised of 53% females and 47% males. Lynch syndrome has been described to cause early onset disease (cancer), usually in individuals younger than 50 years of age. In cohort 2, the majority of individuals (62%) are under the age of 50 years

Cohort 3 (carrying mutations other than the *MLH1* exon 13 mutation) accounted for 86 cases in 23 families diagnosed with Lynch syndrome (from hereon referred to as cohort 3). The most common mutation after the *MLH1* c. 1528C>T is the *MSH2*- exon8 *c.1340InsGG* mutation, in two families with 27 mutation-positive individuals. This is followed by the *MSH2 c.*1219delCT mutation (in exon7), which has been identified in 15 individuals from five families (**Table**

**2.3**). Cohort 3 was made up of 61 females and 25 males, of whom 37 (43%) have developed cancers; these are made up of more than half of the mutation-carrying males and 40% of the mutation-carrying females. As expected, the most common malignancy is CRC accounting for about 60% of the individuals who have been diagnosed with cancer. However, there are a significant number of cases with multiple cancers, with extra-colonic cancers making up 24% of cancers diagnosed in cohort 3. The most common of the extra-colonic cancers observed in cohort 3 is that of the endometrium and breast. Other cancers observed, included cancers of the small bowel, and the stomach, amongst others.

**Table 2.3 Cohort demographics for Lynch syndrome registry**

| **Cohort demographics** | | |
|---|---|---|
| | ****Cohort 3** | ***Cohort 2** |
| **Gender** | | |
| ***Males*** | 25 | 147 |
| ***Females*** | 61 | 162 |
| | | |
| **MMR gene mutation** | | |
| ***MLH1*** | 14 | 309 |
| ***MSH2*** | 73 | 0 |
| | | |
| **Age** | | |
| **Age range** | 17-88 years | 4-96 years |
| | | |
| **Cancer** | 37 | 111 |
| ***CRC/ CRC+other*** | 28 | 95 |
| ***CRC only*** | 23 | 73 |
| ***Extra-colonic cancers (only)*** | 9 | 16 |

*cohort 2: individuals carrying the *MLH1* c. 1528C>T;
** cohort 3: individuals carrying any of the other mutations in *MLH1* and *MSH2*

In cohort 2 consisting of 309 individuals, 104 have been diagnosed with cancer(s). In most instances, probands had been treated, usually for CRC, before the disease pre-disposing mutation was identified in their particular family. Again, as expected, the most common cancer observed in cohort 2 was CRC.

In addition, about 30% of the individuals with CRC have had secondary extra-colonic cancers, including cancers of the liver, small bowel, endometrium and

breast (**Figure 2.3 and 2.4**). Overall, 43% of the individuals who are affected in cohort 2 have developed extra-colonic cancers. Perhaps understandably, most extra-colonic cancers in this cohort have been reported in females (62%). Analysis of the cohort data shows more cases of small bowel cancer in males (35% of the reported extra-colonics); the most common type of extra-colonic cancer in females is that of the endometrium observed in 29% of all extra-colonic cancers (12/41), with an age range of 50 to 64 years. This was followed by breast cancer which was observed in 24% (10/41) of the affected individuals with an age range of 35 to 72 years. These latter two types of cancers together accounted for over 50% of the reported extra-colonic malignancies in females.

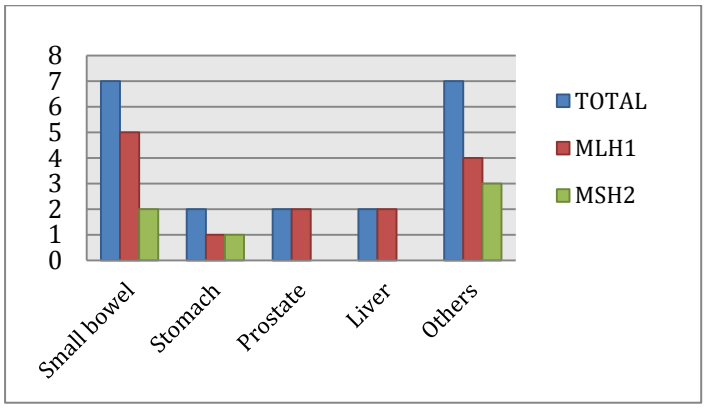



**Figure 2.2 Bar graph showing the range of extra-colonic cancers in males with mutations in *MLH1* (in red) and *MSH2* (in green).** The most common cancers observed were in small bowel for both *MLH1* and *MSH2* mutation carriers. No cases of prostate or liver cancer were observed in individuals with a mutation in the *MSH2* gene.

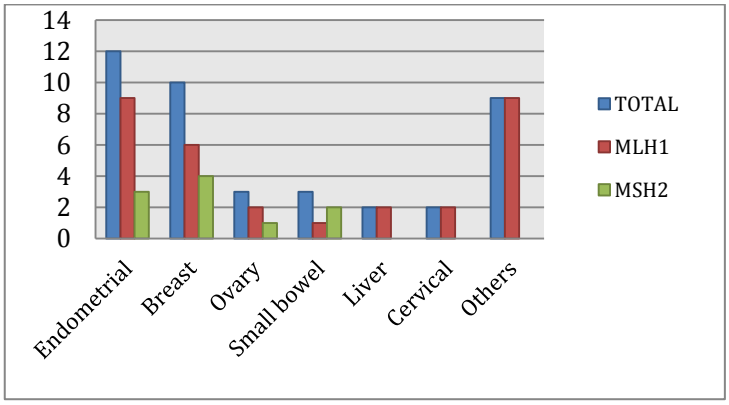



**Figure 2.3 Bar graph showing the range of extra-colonic cancers in females with mutations in *MLH1* (red) and *MSH2* (green).** The most common cancers observed were those of the endometrium and breast. Most of the other cancers were only seen in *MLH1* mutation carriers.

The average age at diagnosis for cohort 1 was approximately 48 years for males and 49 years for females (all cancers) (**Table 2.4**). However, *MLH1* c.1528C>T mutation carriers, had a younger age at onset for CRC (**Figure 2.5**) compared to the background population (66 years) as reported in the National Cancer registry 2010 (http://www.nioh.ac.za/assets/files/NCR_Final_2010_tables(1).pdf;).

**Table 2.4 Average ages at onset for cancer in mutation positive cases**

| | Average age of onset (in years) | |
|---|---|---|
| ***Type of cancers*** | **Males** | **Females** |
| **CRC** | 44.19 | 46.5 |
| **Other cancers*** | 52.8 | 53.54 |
| ***MLH1*-CRC** | 43.13 | 45.2 |
| ***MLH1*-other cancers** | 53.15 | 45.6 |
| ***MSH2*-CRC** | 45.2 | 49.1 |
| ***MSH2*-other cancers** | 52.5 | 56.42 |

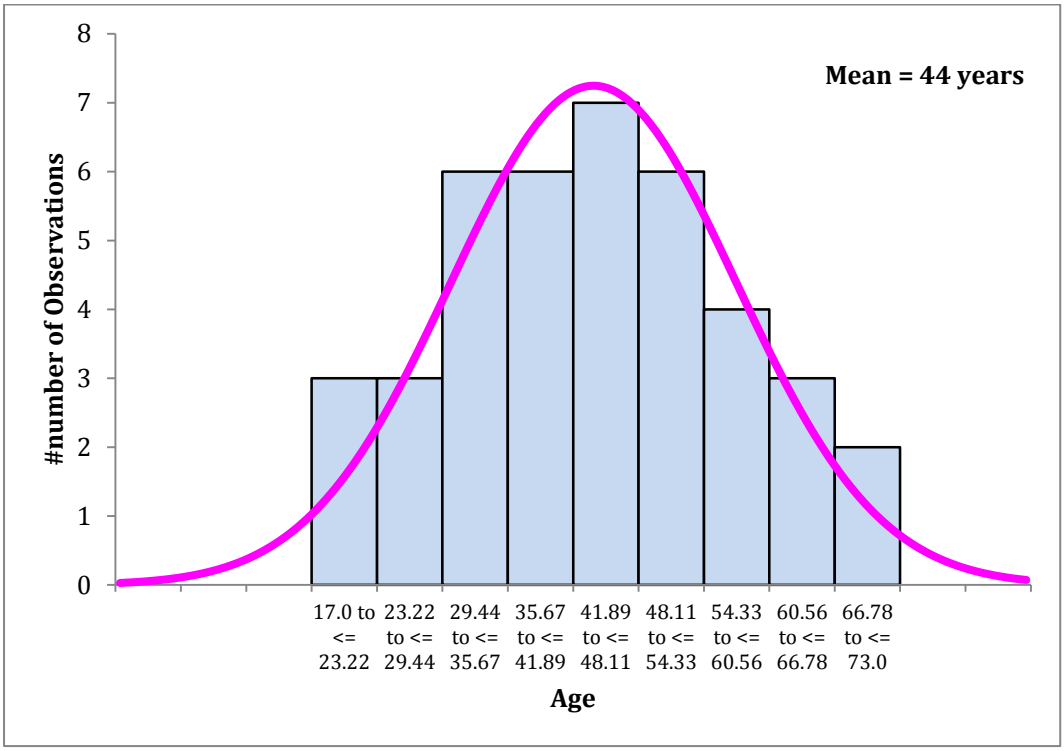



**Figure 2.5 Distribution of age at onset/diagnosis for CRC in Lynch syndrome patients in cohort 2.** The mean age is 44 years. Indicated is the number of observations per age grouping. The age range was from 17 to 69 years for (CRC cases only)

### 2.3.2 Survival Analysis

Survival analysis was performed for cohort 2, consisting only of individuals with the *MLH1 c.1528C>T* mutation. The cohort was also stratified by gender and type of cancer to determine if there was a significant difference in the survival period amongst those who had developed cancer and those who had not developed cancer.

There was no significant difference in cumulative survival for the individuals who have developed cancer versus those who had not (data not shown). When the cohort was stratified by gender, there was no significant difference in overall survival (p=0.15); neither was there a significant difference (p=0.209) in cancer-free survival period between males and females (**Figure 2.6**). However, when individuals were divided by cancer type (**Figure 2.7**), there was a trend towards significance (p=0.0782), with individuals who had developed CRC having a better survival compared to those with extra-colonic cancers.

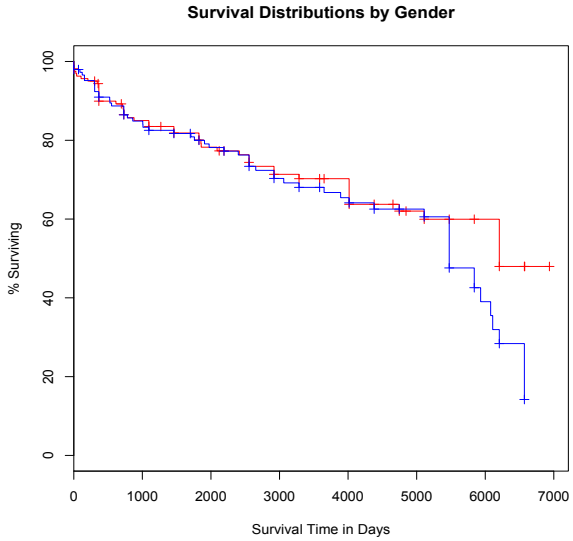



**Figure 2.6 Kaplan Meier Survival analysis for cohort 2 showing males vs. females (Red:** males, **Blue**: Females)

X axis: Survival time in days, Y axis = cumulative survival

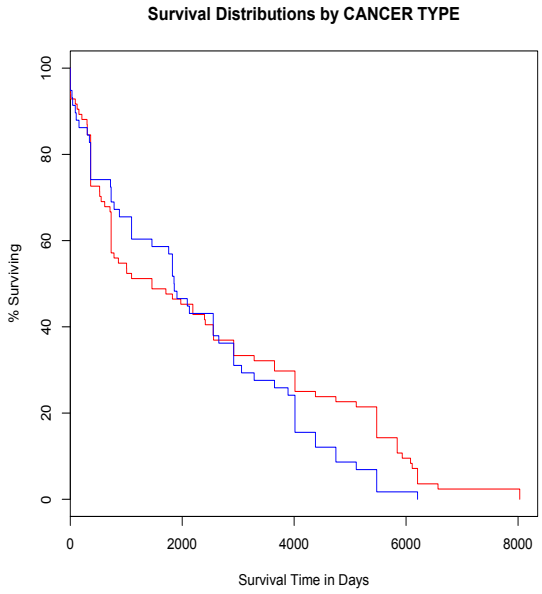



**Figure 2.7 Kaplan Meier Survival analysis of cohort 2 by cancer type** (**Red**: CRC, **Blue:** Extra-colonic)

X axis: Survival time in days, Y axis = cumulative survival

Even though not significantly different, there is a seemingly better survival period recorded for males when compared to females (**Figure 2.6**). However it is noteworthy that there are more cases of CRC in males (than in females), and more extra-colonic cancers in females (than in males). That said, it is likely that, historically, extra colonic cancers were not part of the routine surveillance – and they may be diagnosed at a relatively advanced stage – with a drop in their cancer-free survival period (**Figure 2.7**). This becomes clearer specifically when observing survival comparing CRC and extra-colonic cancers; a better survival margin was noted for individuals with CRC compared to those with extra-colonic cancers.

### 2.3.3 Founder mutation phase analysis

A founder effect is expected to result in shared polymorphisms in *cis* closest to the mutation of interest. To test whether the c.1528C>T mutation represented a founder effect, five-microsatellite markers (**listed in Table 2.5**) around the *MLH1* gene were genotyped and analysed. Phase II (http://stephenslab.uchicago.edu/phase/download.html ) was used to construct the haplotypes. The haplotype analysis was performed in 30 probands and 98 controls. The most common inferred haplotype (**1-6-c.1528C>T-8-1-6**) (for

markers: D3S3512, D3S1561, **c.1528C>T**, D3S1611, D3S3623, and D3S3527) was observed in 25 of the 30 probands. The other five individuals shared the alleles closest to the mutation (**Table 2.5**) but had a different extended haplotype. To determine the prevalence of the haplotype (**1-6-c.1528C>T-8-1-6**), 98 Cape Mixed Ancestry, unrelated controls were genotyped for the same markers. The observed haplotype was not observed in the control population.

**Table 2.5: Microsatellite marker results relative to the c.1528C>T mutation for the 30 probands**. 25 of the 30 probands had the same haplotype, whilst the five remaining cases (indicated in red) share only parts of the haplotype on either side of the mutation.

| PROBAND | D3S3512 | D3S1561 | c.1528C>T | D3S1611 | D3S3623 | D3S3527 |
|---|---|---|---|---|---|---|
| NC_000003.12 | 34594179 | 36484119 | 37028881 | 37068501 | 37443531 | 39345373 |
| n1 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n10 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n100 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n101 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n102 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n103 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n104 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n105 | 1 | 6 | c.[1528C>T] | 8 | 1 | **11** |
| n106 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n106 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n107 | 1 | 6 | c.[1528C>T] | 8 | **3** | **3** |
| n108 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n109 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n11 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n111 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n44 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n45 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n62 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n8 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n87 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n9 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n90 | **7** | **5** | c.[1528C>T] | 8 | 1 | 6 |
| n91 | 1 | 6 | c.[1528C>T] | 8 | **3** | **3** |
| n92 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n93 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n94 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n95 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n96 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n97 | **5** | **8** | c.[1528C>T] | 8 | 1 | 6 |
| n98 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |
| n99 | 1 | 6 | c.[1528C>T] | 8 | 1 | 6 |

In the present study the probable carrier frequency was calculated for the *MLH1 c.1528C>T* mutation in the selected geographical areas in the Northern and Western Cape provinces of S.A. To estimate the prevalence of mutation carriers in a random sample, the carrier frequency of heterozygotes was estimated (**Table 2.6**). The theoretical probability of a homozygous mutation carrier is even lower. However, considering how small and how remote these communities are, the probability of a homozygous case with CMMR-D-syndrome may be less than 1 in 18000 of mutation positive in some of these communities (**Table 2.6**).

**Table 2.6: Carrier frequency estimations for c.1528C>T in the Northern Cape populations**

| Location | Population | Number of Carriers | Frequency (het) | Frequency (Hom) | 1 in: |
|---|---|---|---|---|---|
| Northern Cape | 1166700 | 137/309 | 0.000011 | 0.000000000121 | 826 446 280 |
| Hondeklipbaai | 543 | 4 | 0.0073 | 0.000053 | **18867** |
| Komaggas | 3116 | 23 | 0.0073 | 0.000053 | **18867** |
| Kleinzee | 728 | 5 | 0.0068 | 0.000046 | **21739** |
| Buffelsrivier | 1123 | 7 | 0.0062 | 0.000038 | **26315** |
| Steinkopf | 7842 | 9 | 0.0011 | 0.0000012 | 83333 |
| Garies | 2105 | 6 | 0.0028 | 0.0000078 | 128205 |
| Aggeneys | 2262 | 6 | 0.0026 | 0.0000067 | 149253 |
| Okiep | 6304 | 11 | 0.0017 | 0.0000029 | 344827 |
| Port Nolloth | 6092 | 9 | 0.00016 | 0.000000025 | 4000000 |

### 2.3.4 Estimating the time of origin or age of c.1528C>T mutation

The founder mutation (c.1528C>T) age was estimated using the DMLE+2.3 program (Reeve and Rannala, 2002). A total of 10 000 000 real iterations were performed with 1 000 000 burn-in iterations. The population growth was set at 1.1% in accordance with the current statistics in the Northern Cape province of S.A. The age analysis of the founder mutation showed a range from minimum to nine generations (95% credible set 8.14-9.5) when calculated using the haplotype frequencies (**Figure 2.8**). The results were obtained by algorithmic metric mean of two simultaneous runs shown in **Figure 2.8**. Therefore, if the generation span is considered to be 25 years (Fenner, 2005) and that 2014 is the year study/sampling, the mutation dates back to between 1789 and 1814.

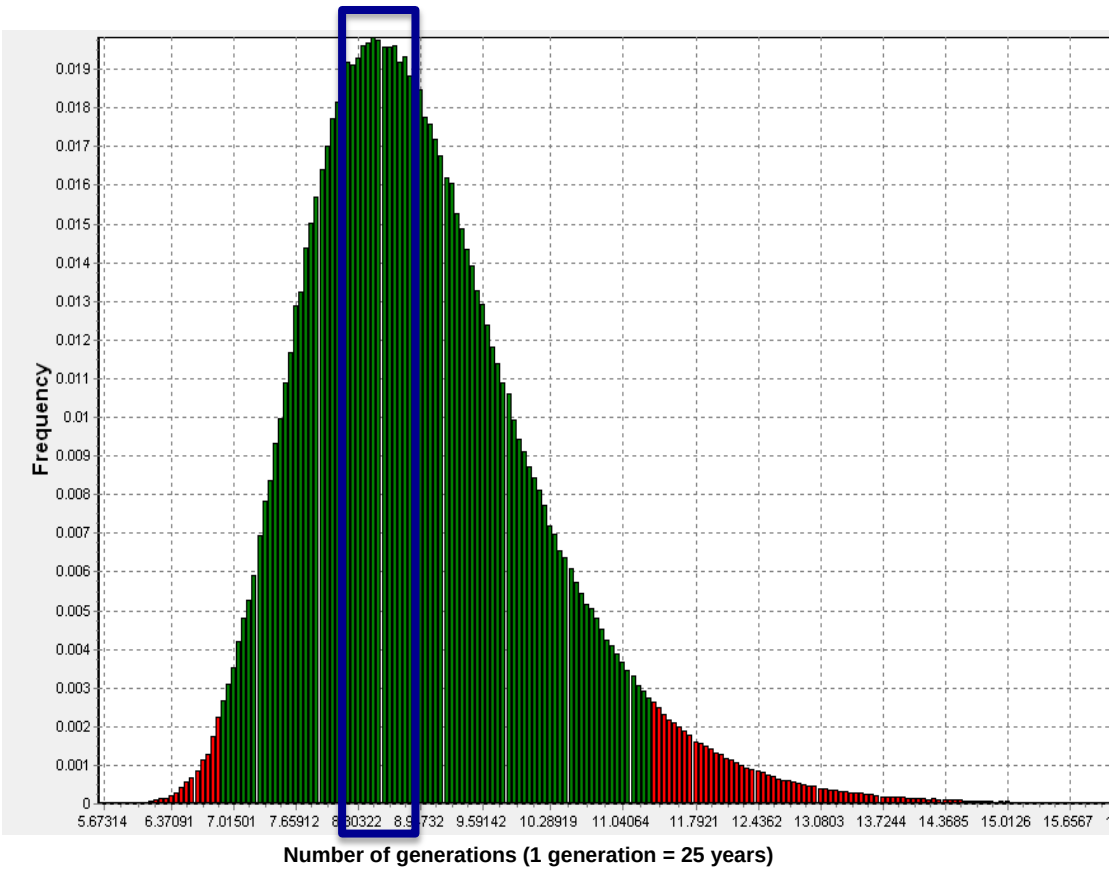



**Figure 2.8: Age estimation of the *MLH1* c.1528C>T mutation.** The probability distribution plots (obtained by two simultaneous runs) of the mutation (in generations), as estimated by the software DMLE+2.3. The analyses were conducted with haplotype frequencies. The blue rectangle shows the age values with the highest probability.

## 2.4 Discussion

The objectives of this chapter involved a detailed analysis of all cases with Lynch syndrome in the UCT Human Genetics registry, and to investigate whether the *MLH1* exon 13 *c.1528C>T* mutation is due to a founder effect. Mutations in both *MLH1* and *MSH2* contribute to the entire cohort of Lynch syndrome cases (396) in the registry. These families have been followed up for over three decades and records of the oldest family in the registry spans several generations. As mentioned previously, the most commonly occurring malignancies in Lynch syndrome are CRC and endometrial cancers. Other malignancies which occur at a higher frequency within Lynch syndrome compared to the general population; include small bowel, gastric, urothelial/renal, brain, biliary, pancreatic and skin cancers. Overall, of the total of 111 cancers in cohort 2, 95 (or 85%) were CRC (range of age at disgnosis: 17-69 years). In cohort 3, CRC was observed in 28 of 37 (76%) individuals (range of age at diagnosis: 25-59 years). This observation of an overwhelming number of CRCs is a feature of most larger collections of Lynch syndrome (Lynch et al., 2009; Pérez-Cabornero et al., 2013; Boland and Lynch, 2013). For males; extra-colonic sites for cancers included the small bowel, stomach and prostate; while for females cancers of the endometrium, ovary, small bowel, liver and cervix were observed, although in smaller numbers.

The most common extra-colonic cancer observed in males was that of the small bowel. Cases of liver cancer were also observed, however reports on the incidence of liver cancer in Lynch syndrome are inconsistent (Levine and Allen, 2014), with further clarification required with regards to whether liver cancer was a primary or metastatic neoplasm. Small bowel cancers have an estimated lifetime risk of 0.01% in the general population (Aarnio et al., 1995; Bansidhar, 2012) while individuals with Lynch syndrome have a lifetime risk of 1 to 4% ( Kate et al., 2007; Bansidhar, 2012). In cohort 1, small bowel cancer was diagnosed in 7 of 21 (33%) (age range: 33- 69 years) cases of mutation positive extra-colonic cancers in males.

For females, the most commonly observed extracolonic cancers in the entire study cohort were those of the endometrium (29%) and breast (24%). A lifetime

risk of endometrial cancer in women with *MLH1* or *MSH2* mutations has been reported to be approximately 30~40 %, with a median age at diagnosis of 49 years (Barrow et al., 2013). In a study of women with Lynch syndrome who developed both a colon and a gynaecologic cancer, Lu and Daniels (2013) found that 50 % of cases first presented with a gynaecologic cancer. Bonaı et al. (2013) investigated 537 French families with Lynch syndrome or MMR gene mutations and reported a lifetime risk of 35% for endometrial cancer with the highest risk predicted for women with *MLH1* mutations. In the present study (cohort 2), 12 of 41 extra-colonic cancer cases in females were that of the endometrium, which is similar to that reported in the literature.

Breast cancer was the second most common extracolonic cancer in females (24% of the cases) in the overall study cohort (cohort 1). The incidence is relatively high considering that this type of cancer had not previously been considered part of the Lynch syndrome spectrum of cancers. The breast cancers in the present cohort have been immunohistologically shown to have the loss of staining of the relevant MMR gene product. There has been considerable debate as to whether breast cancer should be added to the list of the Lynch syndrome spectrum. There have also been several studies documenting MSI in breast tumors, as well as a loss of MMR-gene product staining, histologically, thereby confirming their status as an important feature of Lynch syndrome. Because of the evidence of breast cancer being a feature in our Lynch syndrome cohort, over the past four years, the non-profit organization, Pink Drive (www.pinkdrive.co.za), which advocates screening for breast cancers, has been part of the annual clinical surveillance trip to remote regions of the Western and Northern Cape provinces. This clinical surveillance trip was previously wholly focused on colonoscopic examination of MMR mutation carriers.

A number of studies have shown conflicting results regarding the relative risk of breast cancer among individuals with MMR deficiency (Walsh, 2010; Win et al., 2012; Win et al., 2013). Most of the previously conducted studies were retrospective, however, recently a prospective study (Win et al., 2012) showed that women with MMR deficiency had a four fold increase (4-fold) in the risk of

developing breast cancer compared to the background population. Since the presence of an MMR gene mutation leads to deficiency in the functioning of the protein in correcting all MMR-related errors, it is possible that the individuals with MMR deficiency may have an accumulation of mutations in genes involved in breast cancer (such *as BRCA1, BRCA2, CHEK2 or ATM,* amongst others). The accumulation of these mutations may well accelerate tumour development. As with the range of Lynch syndrome cancers, the development of the disease may not necessarily be the primary result of MMR deficiency, but rather as a secondary effect of the increase in the number of genome-wide mutations which are not efficiently repaired. The suggested modest increase of breast cancers in Lynch syndrome is important to confirm, given that breast cancer is a common disease. In South Africa the age specific incidence of breast cancer in 2012 was about 25/ 100 000 individuals between the ages of 35 and 40 years, (http://www.nioh.ac.za/assets/files/NCR%202012%20results.pdf). This is relatively high compared to other cancers. A lack of prospective studies, and differences in population ethnicity, and a range of environmental and other risk factors makes it difficult to estimate the role of genetic factors contributing towards the disease development.

Win et al. (2012) calculated that as much as 50% of breast cancers in families with MMR gene mutations showed evidence of MMR-deficiency. Some histopathological features such as poor tumour differentiation and presence of tumour-infiltrating lymphocytes, specifically in CRC and endometrial cancers have been reported in the breast cancers in mutation-positive individuals in families with Lynch syndrome (Win et al., 2012; 2013). In addition, some of the breast cancers with MMR deficiency have an increased incidence of mucinous differentiation, another characteristic of cancers with MMR-deficiency. Interestingly, a separate study by Walsh et al. (2013),which investigated the MSI status of MMR-deficient breast cancers, reported these cancers as predominantly of higher nuclear grade, *in situ*, supporting the aberrations of MMR functioning and that they may lead to a more aggressive type of breast cancer. The indication that 50% of the cases of breast cancer in Lynch syndrome families did

not have classic MMR deficiency was no surprise since they may be part of the population burden of a reasonably common cancer (Win et al., 2012).

A study which investigated 90 families with occurrence of both CRC and breast cancer cases, found that about 60% of the families had a mutation in an MMR gene (Walsh et al., 2010). In the observed breast cancer cases, 18/107 cases had abnormal IHC results for MMR gene products. Usually for Lynch syndrome testing, the negative IHC result is indicative of the presence of a mutation in the gene of interest. In this particular study, there was an over representation of the *MLH1* mutation positive cases who had developed breast cancer (Walsh et al., 2010). This has been observed in other cohorts, where individuals with *MLH1* mutations have an increased risk of developing breast cancer when compared to individuals with mutations in other MMR genes. The Brazilian study found breast cancer was one of the most common extra-colonic cancers, even exceeding endometrial cancers in their cohort (Oliveira-Ferreira et al., 2004).

A higher lifetime risk of CRC for males was observed in the UCT/GSH cohort (cohort 1)– this is similar to that reported in other MMR cohorts (Lynch et al., 2009). However, what was evident in the UCT/GSH cohort was that males had a relatively better survival compared to females; this is indicative of the need to develop and employ improved methods of surveillance and /or management of the extracolonic cancers, especially those affecting females. The results showed that individuals with an *MLH1* mutation have a younger age of onset/diagnosis for cancer (**Table 2.6**). This is in accordance with the literature – but may well reflect that this cohort is under better surveillance, and the neoplastic lesions are possibly being detected and treated earlier. Also noted is the age and gender difference in the UCT/GSH cohort, where female *MSH2* mutation carriers have a later age at onset/diagnosis especially for extra-colonic cancers. Again, this may be explained by a lack of adequate surveillance for the extracolonic cancers in this cohort. Even though this is the case, there was no significant difference in survival for *MHL1* or *MSH2* mutation carriers.

A number of factors determine the frequency of Lynch syndrome related cancers; these include: geography, ethnicity, gender, the underlying range of somatic mutations and the specific germ-line mutation (Bansidhar and Silinsky, 2012). Many studies have been conducted with the aim of identifying some of the modifier effects which may impact on Lynch syndrome-related cancers. Talseth-Palmer et al. (2013a) identified candidate modifiers as follows: 1) xenobiotic clearance and micronutrient metabolism, 2) cell cycle control, 3) DNA repair, 4) immunological function, 5) growth factors, 6) other modifiers which included the role of the *MTHFR* gene and the DNA (cytosine-5-)-methyltransferase-3-beta (*DNMT3B*) gene.

The search for modifier genes that influence disease expression in Lynch syndrome has revealed a number of potential candidates. By identifying and including modifier genes/loci in risk algorithms it would be possible to tailor individualized surveillance options for patients, allowing for better outcomes and possibly reducing morbidity and mortality (Felix et al., 2006; Scott et al., 2012; Bellido et al., 2013; Talseth-Palmer et al., 2013a).

The region around the c.1528C>T variant, spanning 4.8 Mb, was characterised by genotyping the microsatellite markers; *D3S3512, D3S1561, D3S1611, D3S3623 and D3S3527*. Based on the allele frequencies for the five markers, one hundred and eighty seven different haplotypes were inferred for 98 (apparently healthy) 'control' individuals from the Cape Mixed Ancestry population and 30 probands harboring the c.1528C>T DNA variant. The most common inferred extended haplotype was observed in 83% of the probands. This disease-associated haplotype was not inferred for any of the persons tested from the background Cape Mixed Ancestry population. Thus, it can be concluded that the disease-associated haplotype is not on a common haplotype present in the relevant population. The mutation age was estimated to be between 8 and 9 generations, which is approximately between 200 and 225 years old (1789-1814). The South African Mixed Ancestry population makes up about 9% of the entire South African population. This population group has its origins in the late 1700s, with the colonization of the Western Cape (De Wit et al., 2010). The mutation has not

been identified anywhere else in the world, suggesting that it originated here in S.A. perhaps more recently than the arrival of European immigrants.

The carrier frequency of this mutation (c.1528C>T) is not high in the Cape Mixed Ancestry population, although it is very likely that the numbers are underestimated. It is to be noted that only 98 persons (background control) were genotyped to determine the carrier frequency of this mutation. In South Africa, mixed ancestry individuals account for just over 8% of the population (~ 5 million), thus to accurately determine the carrier frequency for this mutation, more individuals will need to be genotyped. Future studies should address this limitation. The c.1528C>T mutation has only ever been observed in the South African Cape mixed Ancestry population. This is one of the most admixed populations in South Africa. The population has genetic contributions from indigenous Africans such as Khoesan (or Khoisan) and Xhosa (Bantu speaking populations), Europeans (German, Dutch, French and British) and Asians (De Wit et al., 2010). Thus determining the origin of this mutation will require the haplotype analysis of the genetic markers flanking the c.1528C>T mutation in other populations, perhaps indigenous communities in S.A. and abroad. Nonetheless, since the mutation has not been reported in international disease-specific databases, it is reasonable to presume a local African origin.

Lastly, admixture mapping is a valuable method that can be used to help study the relationship of the admixture of disease risk and to control for admixture as a confounder. McKeingue et al. (2000) described the method to analyse data from an admixed population that allows the effects of linkage and population structure to be distinguished. However, this analysis could not be performed in the current study because of the limited number of individuals genotyped, in addition, only one population was included in the current study. Thus, future work should determine the genotypes for the selected markers (in other populations that may have contributed to this admixed population) in order to estimate the level of admixture in this population, but also to accurately determine the markers associated with disease.

# Chapter 3: Constitutional Mismatch Repair- Deficiency (CMMR-D) Syndrome

## 3.1 Introduction to CMMR-D syndrome

Heterozygous mutations within the DNA mismatch repair (MMR) genes cause Lynch syndrome. In 1999, homozygous mutations in the DNA mismatch repair genes were described in another cancer syndrome: "CMMR-D syndrome" (OMIM: 276300) (Wimmer and Etzler, 2008; Wang et al., 1999; Ricciardone and Tayfun, 1999). This disorder manifests as a result of constitutional biallelic/homozygous or compound heterozygous mutations in the genes encoding MMR proteins (Leung et al., 1998; Menko et al., 2004; Poley et al., 2007; Amayiri et al., 2015; Amayiri et al., 2015; Elhasid et al., 2015; Lavoine et al., 2015; Li et al., 2015). Unlike Lynch syndrome, this disorder presents predominantly in paediatric patients, most of whom do not survive the disease.

CMMR-D syndrome was only recognised relatively recently as a 'component' of Lynch syndrome – it is likely that earlier cases were classified as another disease, since CMMR-D patients have similar characteristics to other paediatric cancer syndromes. The original report by Ricciardone and Tayfun (1999) described three affected children of consanguineous Turkish parents who, although unaffected, both carried the same cancer-predisposing mutation in one of their *MMR* genes, *MLH1*. The pedigree information revealed that both parents had a family history of *hereditary non-polyposis colorectal cancer* (HNPCC -OMIM: 120435). Each of the three offspring had developed haematological cancers by the age of three years and had characteristics of neurofibromatosis type 1 (NF1:OMIM; 162200) e.g. *café-au-lait* (CAL) spots. Results of mutation testing in each of the children revealed an inherited homozygous mutation in *MLH1*. Mutation testing to understand the presence of NF1 symptoms revealed somatic and not germline mutations in the *NF1* or *Neurofibromin* gene on chromosome 17. The presence of the haematological malignancies also suggested that these cases were not traditional NF1 patients. It was speculated that homozygous mutations in *MLH1* may have created a mutator phenotype, which resulted in somatic mutations in the relatively large *NF1* gene on chromosome 17 (374,244

base pairs in length) (Genecards: http://www.genecards.org/cgi-bin/carddisp.pl?gene=NF1).

Soon after this report, Wang et al. (1999) also described paediatric patients with similar phenotypes, where the parents were first cousins. The children had constitutional homozygous mutations in the *MLH1* gene and the extended family was from North Africa, with a confirmed history of cancer. The family characteristics were consistent with the Amsterdam criteria of HNPCC, wherein eight family members developed cancer before the age of 50 years. In the family, there were cases of both neoplastic (indicative of cancer growth in a tissue) and non-neoplastic diseases. The first case of CMMR-D in this family was that of a child, who developed malignant non-Hodgkin's lymphoma (OMIM: 605027) at the age of two years. Subsequently, her sister was diagnosed with acute myeloid leukaemia (OMIM: 61626) at the age of six years, followed by medulloblastoma (OMIM: 155255) at the age of seven years. Both sisters had developed CAL dermatologic macules with no previous family history of NF1. Parents had heterozygous mutations in *MLH1* previously described as pathogenic and disease causing. Genetic testing in one of the children confirmed the presence of the homozygous mutation in *MLH1* (Wang et al., 1999).

One of the main characteristics of MMR deficiency is the presence of MSI in the repeat (microsatellite) sequences of the genome. MSI testing of DNA from buccal mucosal cells of one of the affected sisters showed evidence of MSI (Wang et al., 1999). Although no DNA was available for testing for the other sibling – the observation that the children had similar characteristics and disease presentations led the authors to conclude that both children had inherited the *MLH1* mutation homozygously from their parents. Based on these cases, the authors suggested that penetrance for CMMR-D was 100% by the age of five years. It was speculated that signs of NF1 could be due to the accumulation of *de novo* mutations in the *neurofibromin* gene during the cell/genome replication process due to the mutator phenotype, and that this was simply a function of the relatively large size of the *NF1* gene. It was plausible that the absence of the MLH1-associated post-replication DNA repair system resulted in the embedding

of these non-repaired mutations in the *NF1* gene, resulting in the 'acquired' NF1 features in these subjects.

The first two reported cases of CMMR-D had homozygous *MLH1* mutations, and their parents were related. In 2002, the first case of a homozygous mutation in another MMR gene, *MSH2*, was reported (Whiteside et al., 2002). The mutation was novel and resulted in exon skipping. The patient presented with acute lymphoblastic leukaemia (OMIM: 613065) and signs of NF1 (OMIM: 162200). Phenotypically, the subject described in this report did not meet the more comprehensive criteria for an NF1 diagnosis. The parents were not related and, interestingly, there was no family history of cancer, which added to the complexity of understanding the aetiology of the disease. To date, this syndrome has shown a variable phenotype and it is possible that some features and phenotypes are yet to be described.

### 3.1.1 Epidemiology of CMMR-D syndrome

In 2008, a report of 78 cases of CMMR-D syndrome from 46 different families was published (Wimmer and Etzler, 2008). Subsequently, a comprehensive report by Wimmer et al. (2014), reviewed 146 cases from 91 different families. Most of the cases reported to date are from North America and Europe, although more recent reports from countries in the Middle East and Africa have emerged (Leung et al., 1998; Kratz et al., 2009; Bruwer et al., 2013; Bakry et al., 2014; Antelo et al., 2015; Shlien et al., 2015).

In Saudi Arabia, the numbers of CMMR-D cases were higher than expected, possibly due to cultural practices of consanguinity in that region (Uddin et al., 2013). A retrospective study covering a ten-year period identified 42 cases that met the CMMR-D syndrome criteria (Amayiri et al., 2015). Of those, 39% were diagnosed with high-grade gliomas, approximately 80% of which had negative IHC staining, indicating the loss of expression of more than one MMR protein. From this data, the authors concluded that CMMR-D occurs in approximately 51% of childhood gliomas (Amayiri et al., 2015).

### 3.1.2 Clinical features associated with CMMR-D

The neoplastic phenotypes associated with CMMR-D consist of four categories; (i) haematological malignancies, (ii) brain tumours, (iii) gastric cancers (including the Lynch syndrome spectrum of cancers), and (iv) others (Baas et al., 2013). As with most cancer syndromes, these tumours manifest early in life. The age range for brain tumours in CMMR-D was estimated at five and a half to eight years, while the age of onset for the Lynch syndrome spectrum of cancers in CMMR-D syndrome, was estimated at about 15 -30 years (Jasperson et al., 2011).

As mentioned previously, individuals diagnosed with CMMR-D also present with signs of NF1, such as CAL macules on the skin. NF1 is an autosomal dominant disorder, which results from mutations in the *NF1* gene. This large tumour-suppressor gene, comprising of 60 exons has a high mutation rate (Campos et al., 2013). Although NF1 is usually an inherited disease, for individuals with CMMR-D syndrome, the signs of NF1 are quite distinct, comprising largely of e.g. CAL macules on the skin. In CMMR-D there is hyperpigmentation of the CAL macules with spots of hypo-pigmented borders; the skin lesions are more diffuse and irregular when compared to the classic CALs in inherited NF1. The number of skin lesions in CMMR-D varies, ranging from mostly one to two focal areas, and sometimes more areas of skin pigmentation (Whiteside et al., 2002). From the cases that have been published, there is no evidence of systemic disease as one would see in NF1 which is the result of inherited mutations in the *NF1* gene (Wang et al., 2003).

Since CMMR-D syndrome is relatively rare, one of the ways in which researchers have begun to investigate it, has been through research consortia. A consortium is an association of several entities, which in this case includes a multidisciplinary team including paediatric and adult oncologists from different countries, working together with other clinicians and scientists to identify, evaluate and investigate the causes of the various phenotypes within the syndrome, to develop relevant monitoring and surveillance modalities, and to develop new treatments for children and young adults with CMMR-D. To date, two international consortia have been established, namely, the Biallelic

Mismatch Repair Deficiency (BMMRD) syndrome consortium (www.sickkids.ca/MMRD) and the Care for Constitutional Mismatch Repair Deficiency (C4CMMRD) consortium (Wimmer et al., 2014), both of which have similar goals. In a recent report by Bakry et al. (2014), 97% of CMMR-D patients who were followed up as part of the *BMMRD* consortium, had developed CAL macules (Bakry et al., 2014).

However, it should be noted that because of the shared characteristics with many other inherited syndromes, in some instances, individuals with CMMR-D syndrome are misdiagnosed or misclassified. As noted above, the specific features of the CAL macules differ from the classic NF1 – however this may only be evident when highly experienced clinicians at tertiary and quaternary health-care facilities see patients. The CMMR-D-related NF1 skin phenotype consists of distinct irregular hyper- and hypo-pigmented lesions (Rauen et al., 2014).

The non-recognition or misdiagnosis of CMMR-D also means that appropriate surveillance measures are not engaged with timeously and individuals end up dying of potentially preventable malignancies. Another challenge with CMMR-D is that individuals who survive the first malignancy, usually develop other tumours soon afterwards. For these individuals, the progression from adenoma to adeno-carcinoma is rapid with most individuals likely going unnoticed until the cancer has progressed to an advanced stage. Turcot syndrome is an additional rare childhood MMR-deficiency syndrome, which further complicates CMMR-D diagnosis. The disorder is characterised by multiple polyps and adenomas of the colon, and an increased risk of both colon and brain cancers, specifically medulloblastoma in children. In OMIM (http://www.omim.org) Turcot syndrome is classified under the broad category of all MMR cancer syndromes (OMIM: 276300), together with CMMR-D. When it was initially described in 1959, Turcot syndrome was divided into dominantly and recessively inherited types (Hamilton and Liu, 1995). The dominantly inherited type is marked by the presence of heterozygous mutations in the *APC* gene (OMIM: 611731), which is primarily associated with FAP (OMIM: 175100), another colon cancer-predisposing disease. A number of distinct features have

been identified between these two syndromes, including: type of tumour developed, presence of NF1-like characteristics and haematological cancers for CMMR-D (Järvinen et al., 2000). Of note is that cases of Turcot syndrome, due to *APC* gene mutations, often develop medulloblastomas compared to those that have MMR deficiency, most of whom develop glioblastomas (Mori et al., 1994; Hamilton and Liu, 1995; Huang et al., 2000). Following the description of CMMR-D syndrome, some of the previously described cases of Turcot syndrome have been re-classified as CMMR-D.

### 3.1.3 Spectrum of tumours in CMMR-D

#### *3.1.3.1 Haematological malignancies*

In general, haematological malignancies are a group of heterogeneous conditions, which originate from bone marrow and plasma cells (Rodriguez-Abreu et al., 2007; Hall et al., 2013) and are divided into leukaemias, lymphomas and plasma cell neoplasms (Hall et al., 2013). According to Wimmer & Etzler (2008) , the most prevalent of haematological cancers in CMMR-D are non-Hodgkin's lymphoma and acute lymphoblastoid leukaemia. Non-Hodgkin's lymphoma is a lymph-proliferative disease with distinct clinical and histological characteristics (Evans and Hancock, 2003). This type of cancer is heterogeneous and 85-90% of the cases arise from B-lymphocytes, which can occur in lymph nodes and white blood cells (Evans & Hancock, 2003; Shankland et al., 2012). Acute lymphoblastoid leukaemia is also a heterogeneous cancer, derived from both B and T lymphoid progenitors. Sporadic cases of this leukaemia have been observed in both children and adults ( Onciu, 2009; Inaba et al., 2013). It has been proposed that development of other types of haematological malignancies, such as acute myeloid leukaemia might be secondary, due to the administration of chemotherapy and/or the development of NF1 (of which acute myeloid leukaemia is a feature) in the CMMR-D cases (Wimmer and Etzler, 2008).

#### *3.1.3.2 Brain tumours*

Brain tumours are the most common malignancy in CMMR-D cases. These include high-grade gliomas (HGG), primitive neuroectodermal tumours and medulloblastomas. In a recent study of 18 individuals with CMMR-D, Barky et al. (2014) reported the presence of brain tumours in 74% of cases. The most

prevalent type of brain tumours are the glioblastomas (Wimmer and Etzler, 2008). Glioblastomas are one of the most common brain cancers in young children, and these may develop from lower grade astrocytic tumours (mostly, the secondary type), while the primary glioblastomas most commonly occur *de novo* (Lee et al., 2008; Huse and Holland, 2010; Epple et al., 2012; Archer et al., 2013). This type of brain tumour displays a highly heterogeneous cellular composition, which in turn increases the invasiveness of the tumour (Epple et al., 2012; Archer et al., 2013).

One type of brain cancer reported as part of the CMMR-D spectrum is medulloblastoma, also a relatively common brain malignancy in children (Huse and Holland, 2010). Medulloblastomas are believed to arise in the cerebellum because of an aberrant activation of the Sonic hedgehog (SHh) pathway. The Hedgehog pathway is a signalling system, which regulates a number of important processes such as cell differentiation and cell proliferation, amongst others. An inappropriate activation of this pathway has been implicated in a number of cancers including those of the lung, breast and brain (Kool et al., 2008; Gupta et al., 2010). Other types of brain tumours, which have been reported as part of the syndrome, are listed in **Table 3.1** (Wimmer et al., 2008).

#### *3.1.3.3 Lynch syndrome-related tumours*

Not many instances of the Lynch syndrome spectrum of cancers have been reported in individuals with CMMR-D although, when present, these cancers develop at an earlier age e.g. in childhood or adolescence (mean age: 17 years) (Wimmer and Etzler, 2008). Individuals with gastrointestinal tumours often present with polyps. CRC is also seen in these cases, with a mean age of about 16.4 years at diagnosis (range from 8 to 35 years), with most cases being paediatric. Like Lynch syndrome, individuals with CMMR-D often have CRC diagnosed in the ascending colon (Levi et al., 2015). Compared to other CMMR-D cancers, the CRC cases often occur in relatively older individuals with an age range of 23-35 years (Wimmer and Etzler, 2008; Wimmer et al., 2014). A low occurrence of renal and bladder cancers have also been reported in CMMR-D, in

individuals ranging from 19 to 21 years of age (Wimmer et al., 2014). Other cancers within this syndrome are also classified under 'emerging phenotypes', including hepatic carcinomas, neuroblastomas, Wilms tumor and rhabdomyosarcoma (Wimmer et al., 2014).

**Table 3.1 Spectrum of cancers reported in CMMR-D syndrome**

Table adapted from (Wimmer and Etzler, 2008)

| Type of cancer | Median age at onset in years (age range where available) |
|---|---|
| **Haematological Malignancies** | |
| **Acute lymphoblastoid leukaemia** | 4 |
| **Acute myeloid leukaemia** | 9 |
| **Lymphomas** | 5 |
| **Atypical chronic myeloid leukemia** | 1 |
| Overall average age | **5.5** |
| **Brain Tumours** | |
| **Glioblastoma and other astrocytic tumours** | 8 |
| **Supratentorial primitive neuroectodermal tumour** | 8 |
| **Medulloblastoma** | 7 |
| **Unspecified** | 23-24 |
| Overall average age | **8** |
| **Lynch Syndrome-associated Tumours** | |
| **CRC** | 16(8-35) |
| **Endometrial carcinoma** | 24(23-35) |
| **Duodenum/jejunum carcinoma** | 16(11-41) |
| **Ureter/renal pelvis carcinoma** | 15 |
| Overall average age | **16** |
| **Other Cancers** | |
| **Neuroblastoma** | 13 |
| **Wilms tumour** | 4 |
| **Ovarian neuroectodermal tumour** | 21 |
| **Infantile myofibromatosis** | 1 |
| **Rhabdomyosarcoma** | 4 |
| **Mamma carcinoma** | 35 |
| **Sarcoma** | 65 |

### 3.1.4 Family history and CMMR-D

Consanguineous parents are a relatively common feature of CMMR-D cases. In some instances, CMMR-D families may already have a history of the Lynch syndrome spectrum of cancers, or at least meet the Amsterdam criteria. However, from a review of the literature thus far, parents of offspring with CMMR-D are often unaffected. This observation may be due to CMMR-D-related

cancer diagnosis in very young children (e.g. <5 years of age) where parents may well be in their third (20s) or fourth (30s) decade of life, prior to any signs or clinical diagnosis of Lynch syndrome. It is also possible that there is a decreased penetrance associated with certain genes and mutations (Drost et al., 2013). Only a few studies have been performed to assess the penetrance of disease in individuals with *PMS2*, which only account for about 10% of all cases of Lynch syndrome, compared to other MMR genes, such as *MLH1* and *MSH2*. In one such study, of 377 *PMS2*-mutation-positive cases from more than 2000 European families: the lifetime risk for disease development was estimated to be about 19% and 11% for CRC in males and females, respectively, and about a 12% risk for endometrial cancer in females. The estimated risk of a malignancy with *PMS2* mutations was still higher compared to the background population, but was relatively lower when compared to *MLH1*- and *MSH2*-mutation carriers. This could, at least to some extent, explain the lack of a clear dominant family history among individuals who harbour biallelic *PMS2* mutations and who manifest with CMMR-D.

### 3.1.5 Genetics of CMMR-D syndrome

#### *3.1.5.1 Mutations associated with CMMR-D syndrome*

Biallelic mutations in the genes coding for the four main MMR proteins, namely MLH1, PMS2, MSH6 and MSH2 have been associated with CMMR-D syndrome. All four of these genes play an important role in MMR activity (**Chapter 1, section 1.4.2**). Mutations (including monoalleic) in these genes result in the suboptimal functioning of the DNA post replication-repair mechanism and repair efficiency. This biological compromise has been associated with many cancers and cancer syndromes as previously described. The "InSight" variant website (http://insight-group.org) maintains a database of all mutations in these MMR genes that have previously been reported to be associated with cancers. As would be expected, not all of the thousands of mutations reported in the MMR genes have been associated with CMMR-D. The mutations reported in association with CMMR-D are listed in **Table 3.2.** Most of these are compound heterozygous mutations located in the *PMS2* gene. Mutations reported in the *MLH1* and *MSH2* genes are mostly homozygous and indels, respectively.

**Table 3.2 List of mutations associated with CMMR-D syndrome (CompHet = compound heterozygote)**

| No. | Gene | Zygosity | Sequence change | Protein change |
|---|---|---|---|---|
| **1** | *MLH1* | Homozygous | c.676C>T | ARG2265STP |
| **2** | *MLH1* | Homozygous | c.199G>T | GLY67TRP |
| **3** | *PMS2* | CompHet | c.2113G>A | GLU705LYS |
| | *PMS2* | CompHet | C del-upstream codon 1 | p.? |
| **4** | *PMS2* | CompHet | c.400C>T | AGR134STP |
| | *PMS2* | CompHet | c.delAAG | LYS-DEL618 |
| **5** | *PMS2* | CompHet | c.1221delG | THR408LEU |
| | *PMS2* | CompHet | c.2361delCTTT | p.? |
| **6** | *PMS2* | Homozygous | c.1169ins20 | p.? |
| **8** | *MSH2* | Homozygous | c.1662-1G>A (exon skipping) | p.? |
| **9** | *MSH2* | CompHet | c.exon3 deletion | p.? |
| | *MSH2* | CompHet | c.153-1del1bp (exon3) | p.? |
| **10** | *MLH1* | Homozygous | c.1943C>T | P648S |
| **11** | *MLH1* | Homozygous | c.2059C>T | ARG687TRP |
| **12** | *MLH1* | Homozygous | nonsense mutation | GLU268GLY |
| **13** | *PMS2* | Homozygous | | R802STP |
| **14** | *MSH6* | Homozygous | c. 3386-3388delGTG | p.? |
| **15** | *MSH6* | CompHet | c.3609-3612del | HIS1203GLN |
| | *MSH6* | CompHet | c.3073G>A | p.? |
| **16** | *PMS2* | CompHet | c.1951 C>T | Q643STP |
| | *PMS2* | CompHet | c.161C>T | S46I |
| **17** | *MSH2* | Homozygous | c.2006-5T>A | GLY669ASP |
| **18** | *MLH1* | CompHet | c.677G>A | R226STP |
| | *MLH1* | CompHet | c.2146G>A | V716M |
| **19** | *PMS2* | Homozygous | | R802STP |
| **20** | *MSH6* | CompHet | c.3226C>T | CYT1076ARG |
| | *MSH6* | CompHet | c.3991C>T | ARG1331STP |
| **21** | *MSH2* | CompHet | c.229G>A | C765W |
| | *MSH2* | CompHet | | V878A |
| **22** | *MSH6* | CompHet | c.642C>G | Y214STP |
| | *MSH6* | CompHet | c.458-1 G>A(splice site) | p.? |
| **23** | *PMS2* | Homozygous | c.400kb del. (exon9-15) | p.? |
| **24** | *MSH6* | CompHet | c.1596-1597T | GLU533fs |
| | *MSH6* | CompHet | c.3261delC | PRO1087fs |
| **25** | *MLH1* | CompHet | c.595delAG(frameshift) | p.? |
| | *MLH1* | CompHet | c.104T>G | MET35ASN |
| **26** | *MSH2* | Homozygous | c.226C>T | Q76X |
| **27** | *MSH2* | Homozygous | c.1906G>C | ALA63PRO |
| **28** | *PMS2* | Homozygous | c.1306dupA | SER436LYSfs |
| **29** | *MSH6* | Homozygous | c.4002-3_4002_8 indel | p.? |
| **30** | *PMS2* | Homozygous | c.182delA | TRY61LEU |
| | *PMS2* | Homozygous | c.234A>G(exon4) | p.? |
| | *PMS2* | Homozygous | c.2340C>T(unclassified) | (p.?) r.2340c>u |
| **31** | *PMS2* | Homozygous | c.182delA | TRY61LEU |
| **32** | *MSH6* | CompHet | c.1806-1809delAAG | GLU604LEU |
| | *MSH6* | CompHet | c.3226C>T | ARG1076CYS |
| **33** | *PMS2* | Homozygous | | P1590STP |
| **34** | *PMS2* | Homozygous | | G271V |
| **35** | *PMS2* | CompHet | c.1408C>T | LYS614STP |
| | *PMS2* | CompHet | c.1454C>A | THR485LYS |
| **36** | *PMS2* | Homozygous | del Exon7 | p.? |
| **37** | *PMS2* | Homozygous | c.137G>A | S46N |
| **38** | *PMS2* | Homozygous | c.989-1G>T | p.? |
| **39** | *PMS2* | CompHet | del Exon 1-6 | p.? |
| | *PMS2* | CompHet | c.1A>G Met1? | Met1? |
| **40** | *PMS2* | Homozygous | c.219A>T | CYS73STP |
| **41** | *PMS2* | CompHet | c.137G>T | SER46ILE |
| | *PMS2* | CompHet | c.804-2A>G(splice site) | p.? |
| **42** | *MSH6* | CompHet | c.1634-1635AAdel | LYS545ARG |
| | *MSH6* | CompHet | c.357-3958insTCAAA… | p.? |
| **43** | *MSH6* | Homozygous | c.691delG | VAL231STP |
| **44** | *MSH6* | CompHet | c.3226C>T | R1076C |
| | *MSH6* | CompHet | c.1422insTG | GLN475CYS |
| **45** | *PMS2* | Homozygous | c.989-1G>T | p.? |

| No | Gene | Zygosity | Sequence change | Protein change |
|---|---|---|---|---|
| **46** | *MSH6* | Homozygous | c.326insC | F1088L |
| **47** | *PMS2* | CompHet | c.1687C>T | R563STP |
| | *PMS2* | CompHet | Maternal allele dropout | p.? |
| **48** | *PMS2* | CompHet | c.137G>T | SER46ILE |
| | *PMS2* | CompHet | c.736_741del6insTGTG… | PRO246fs |
| **49** | *PMS2* | CompHet | c.1239dupA | LYS413fs |
| | *PMS2* | CompHet | c.1927C>T | GLN643STP |
| **50** | *PMS2* | Homozygous | c.2174+1G>A(intron12) | p.? |
| **51** | *PMS2* | CompHet | c.137G>T | SER46ILE |
| | *PMS2* | CompHet | c.2174+1G>A(intron12) | p.? |
| **52** | *PMS2* | CompHet | del PMS2 | p.? |
| | *PMS2* | CompHet | c.736-741insGTGTG… | PRO246CYSfs |
| **53** | *PMS2* | CompHet | c.137G>A | SER46ILE |
| | *PMS2* | CompHet | c.2174+1G>A(intron12)(aberrant splicing) | p.? |
| **54** | *PMS2* | CompHet | del Exon1-11 | p.? |
| | *PMS2* | CompHet | del Exon3-7 | p.? |
| **55** | *PMS2* | Homozygous | c.2174+1G>A(intron12) | p.? |
| **56** | *PMS2* | CompHet | c.delExon7-14 | p.? |
| | *PMS2* | CompHet | c.delExon7-11 | p.? |
| **57** | *PMS2* | CompHet | c.1687C>T | ARG563STP |
| | *PMS2* | CompHet | r.2007_2445del439(delExons12-14) | p.? |
| **58** | *PMS2* | CompHet | c.325dupG | p.? |
| | *PMS2* | CompHet | r.804_825del(silent premature stop-RNA) | ILE269ALA(STP) |
| **59** | *PMS2* | CompHet | c.1221delG | THR408LUE(STP) |
| | *PMS2* | CompHet | c.2361_2364delCTTC | PHE788CYS |
| **60** | *PMS2* | Homozygous | c.1169_1170ins20(frame shift) | p.? |
| **61** | *PMS2* | CompHet | c.137G>T | SER46ILE |
| | *PMS2* | CompHet | c.1927C>T | GLN643STP |
| **62** | *PMS2* | CompHet | c.137G>T | SER46ILE |
| | *PMS2* | CompHet | c.1730dupA; 1732C>T | ARG578VAL |
| **63** | *PMS2* | CompHet | c.400G>T | ARG134STP |
| | *PMS2* | CompHet | c.2184-2185del | LEU729GLN |
| **64** | *PMS2* | CompHet | c.2361-2364del | PHE788CYS |
| | *PMS2* | CompHet | c.1221delG | THR408LUE(STP) |
| | *PMS2* | Homozygous | c.1169_1170ins20(frameshift) | p.? |
| **65** | *PMS2* | Homozygous | c.2407C>T | ARG802STP |
| | *PMS2* | Homozygous | c.1768del | ILE590PHE |
| **66** | *PMS2* | Homozygous | c.812G>T | GLY271VAL |
| **67** | *PMS2* | Homozygous | c.1840A>T | LYS614STP |
| **68** | *PMS2* | Homozygous | del Exon7 | p.? |
| **69** | *PMS2* | CompHet | c.2249G>A | GLY750ASP |
| | *PMS2* | CompHet | del PMS2 | p.? |
| **70** | *PMS2* | CompHet | c.1A>G Met1?(5'trancation) | Met1? |
| | *PMS2* | CompHet | del Exon9-10 | p.? |
| **71** | *PMS2* | CompHet | c.1A>G Met1?(5'trancation) | Met1? |
| | *PMS2* | CompHet | c.614A>C | GLN205PRO |
| **72** | *PMS2* | CompHet | c.1A>G Met1?(5'trancation) | Met1? |
| | *PMS2* | CompHet | c.251A>G(aberrant splicing) | p.? |
| **73** | *PMS2* | Homozygous | c.949G>T | GLN317STP |
| **74** | *MLH1* | CompHet | c.1852-1853CC>GG | LYS618ALA |
| | *MLH1* | CompHet | c.546A>G(aberrant splicing) | ARG182SER |
| **75** | *MLH1* | Homozygous | c.2059C>T | ARG687TRP |
| **76** | *MLH1* | Homozygous | c.806C>G | SER269STP |
| **77** | *MLH1* | CompHet | c.2146G>A | VAL716MET |
| | *MLH1* | CompHet | c.676C>T | ARG226STP |
| **78** | *MSH2* | Homozygous | c.2006-5T>A(aberrant splicing) | GLY669ASP |
| **79** | *MSH2* | Homozygous | c.1906G>C | ALA636PRO |
| **80** | *MSH2* | CompHet | del Exons1-6 | p.? |
| | *MSH2* | CompHet | c.1A>G (5'trancation) | Met1? |
| **81** | *MSH6* | Homozygous | c.3386-3388del | CYS1129_VAL1130indelLEU |
| **82** | *MSH6* | Homozygous | c.3635dupT | ASP1213GLY |
| **83** | *MSH6* | CompHet | c.2633T>C | VAL787ALA |
| | *MSH6* | CompHet | c.2295C>G | CYS765TRP |
| **84** | *MSH6* | CompHet | c.3226C>T | ARG1076CYS |
| | *MSH6* | CompHet | c.3991C>T | ARG1331STP |
| **85** | *MLH1* | Homozygous | c.1528C>T | GLN510STP |

### 3.1.6 Genotype/Phenotype correlation

Wimmer et al. (2014) recently reviewed 146 reported cases of CMMR-D from 91 different families, globally. These subjects ranged in age from about six months to 39 years. A total of 145 of the 146 cases already had malignancies or neoplasias following 'mutational' diagnosis. *PMS2* and *MSH6* were by far the most commonly mutated genes reported in patients diagnosed with CMMR-D. As mentioned earlier, CMMR-D patients with *MSH6* and *PMS2* mutations often have no family history of cancer. According to Wimmer et al. (2014), individuals with either *MLH1* or *MSH2* mutations were more likely to develop haematological primary malignancies, and at a much earlier age than those with mutations in either *MSH6* or *PMS2*. About 60% of the *PMS2* and *MSH6* mutation carriers, compared to only about 30% of the *MLH1 or MSH2* mutation carriers, developed brain tumours as their primary malignancy.

Although patients with either *MSH6* or *PMS2* mutations were likely to survive their initially diagnosed malignancy, they tend to be diagnosed with secondary tumours relatively shortly thereafter (Wimmer et al., 2014). There is an emerging literature on CMMR-D syndrome – however, it is clear that more research is needed to establish clear genotype and phenotype correlations for this devastating group of diseases.

### 3.1.7 Diagnosis of CMMR-D

Unlike the molecular diagnosis of Lynch syndrome, which requires identification of a mutation in one of the alleles in one of the MMR genes, there are challenges associated with diagnosis of CMMR-D (Leenen et al., 2011). Traditionally for Lynch syndrome, testing is performed if the individual meets the Bethesda criteria as previously described in Chapter 1 (literature review). For the individuals with CMMR-D, a family history (as defined within the Bethesda criteria) is not very helpful; and patients may well be identified 'accidentally', and where there is a specialist interest in e.g. Lynch syndrome. Most studies – when already suspecting CMMR-D, perform three main tests; 1) IHC, 2) MSI testing, and 3) molecular genetic testing (Wimmer and Kratz, 2010).

### *3.1.7.1 Immunohistochemistry (IHC)*

IHC tests are performed with antibodies which check for the integrity of the common protein products of the MMR genes associated with the disease i.e. *MLH1*, *MSH2*, *MSH6* and *PMS2*. The tumour analysis for IHC in Lynch syndrome has been shown to be specific and sensitive in pointing to the respective 'mutated' gene (Peltomäki and Gylling, 2011). For IHC, the loss of expression in one of the genes implies the presence of a pathogenic germline mutation in that particular gene (with a somatic second hit occurring – and perhaps being the precipitating event) leading to tumorigenesis. Following IHC, sequencing of the indicated gene is performed to confirm the mutation.

### *3.1.7.2 Microsatellite instability (MSI)*

Tumours with MMR deficiency are often characterised by the presence of MSI. However, with CMMR-D, the role or presence of MSI is controversial. A review of 31 brain cancers (diagnosed with CMMR-D) by the C4CMMRD consortium in Europe, reported that most cases had microsatellite stable tumours. In a study by Giunti et al., (2009), the MSI mononucleotide marker panel was tested in gliomas in order to gauge whether patterns of MSI correlated with the family history of cancer. It was reported that the degree of instability is lower in MMR-related gliomas versus colon cancer cells, perhaps because of the small allelic shift in the CNS (where there are fewer actively proliferating cells) compared to the colon/CRC (Giunti et al., 2009). Thus, depending on which tissue is being tested, the MSI status could be uninformative and this calls for improvement in methods used to diagnose CMMR-D syndrome.

### *3.1.7.3 Molecular genetic testing*

In Lynch syndrome, the MMR proteins shown to be compromised during IHC generally point to the gene which may be mutated; this gene is then screened through DNA sequencing to identify the causative mutation. Usually the parents of the affected individual will also be tested to confirm the origin of the mutation manifesting in the child. The identified mutations range from single base pair changes to insertions and deletions of larger sections of the gene. As previously mentioned, to date mutations in the *PMS2* gene have been identified to underlie

cases of CMMR-D syndrome more often than in other MMR genes. Other tests performed are those for the mutations in the *NF1* gene. This has been done mostly to confirm that the *NF1* mutations are not germline, but somatic, reflecting that the NF1 features are a secondary result of the mutator phenotype, and a consequence of MMR mutations.

### 3.1.8 Recommended surveillance and treatment

For individuals with CMMR-D syndrome, very little is known about the efficiency of chemotherapy for the treatment of childhood haematological malignancies (Ripperger et al., 2010). Regular follow-up by an oncologist (Leenen et al., 2011), assessment of immunoglobulin levels, colonoscopy (Ilencikova et al., 2011; Herkert et al., 2011), video capsule small bowel endoscopy from the age of eight years (Herkert et al., 2011) and magnetic resonance imaging (MRI) (Ilencikova et al., 2011; Leenen et al., 2011; Durno et al., 2012 ) may be considered options for surveillance. There are, as yet, no established management surveillance programs or guidelines to monitor these patients to ensure optimal prognosis and survival.

### 3.1.9 Chapter synopsis

As reported earlier, researchers in the Division of Human Genetics at the UCT have found the primary predisposing gene defect in the *MLH1 (7127)* gene, which is a component of the post-replication MMR system for Lynch syndrome in the Mixed Ancestry population of the Northern and Western Cape Provinces of S.A. Currently, several communities are being managed (pre-symptomatically) with the use of genetic testing, genetic counselling and surveillance colonoscopies of MMR-gene mutation carriers. A study by Stupart et al., (2009) indicated the effectiveness of this programme through improved mortality and morbidity. However, more recently, a four year old child in this S.A. cohort demised due to a grade IV astrocytoma in the brainstem, despite early pre-symptomatic molecular diagnosis (Bruwer et al., 2013). This subject was the child of non-consanguineous parents, who were both heterozygous for the *MLH1* c.1528C>T mutation, with a history of Lynch syndrome in their respective

families. As indicated in the pedigree in **Figure 3.1** the proband (III-1) was homozygous for the c.1528C>T; c.1528C>T mutation.

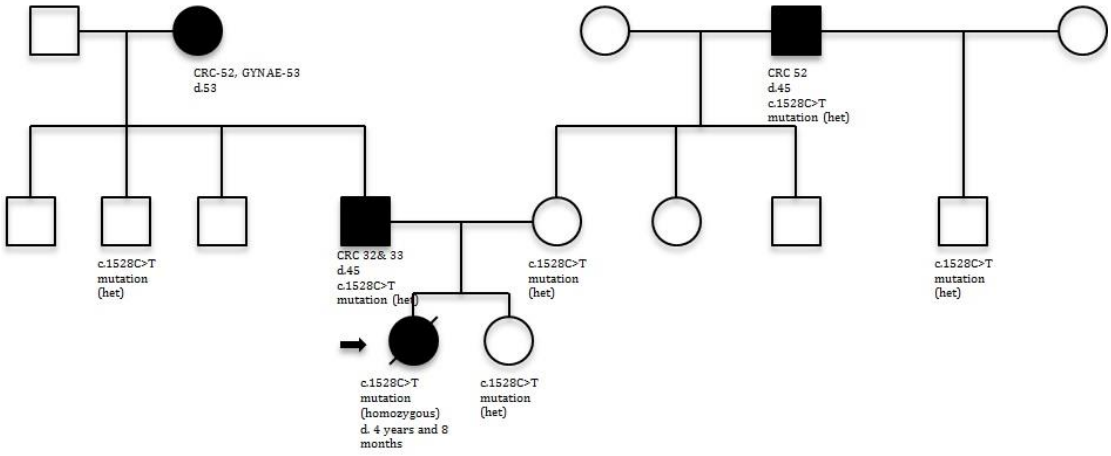



**Figure 3.1: Pedigree indicating clinical features in a family with Lynch syndrome and CMMR-D**. Also indicated are the types of cancers, age at diagnosis and mutation status for family members. In the figure, squares represent males, and circles represent females, deceased individuals are indicated by a diagonal line through the square/circle. Black/shaded symbols represent affected individuals with the type of cancer indicated below the symbol e.g. CRC or gynae for a gynaecological cancer. The specific *MLH1* c.1528C>T (het= heterozygous) mutation has only been observed in families of Mixed Ancestry in S.A. and has not been reported in any other population (Ramesar et al., 2000). (**Figure adapted from Bruwer et al., 2013**)

As discussed by Bruwer et al. (2013), the management as well as the surveillance applied to this patient was unsuccessful, highlighting the need to identify new ways to manage CMMR-D cases. The questions worthy of investigation include: what are the biological consequences of having a homozygous mutation in a MMR gene, and whether one might be able to predict which cancers to optimally screen for in these patients. The clinical proximal research will be supported by asking the following questions:

(i) what are the range of genes that are most susceptible to acquiring somatic mutations as a result of both a heterozygous and/or homozygous mutation in the *MLH1* gene,

(ii) which parts of these genes are most likely to be mutated i.e. consensus nucleotide sequence or specific domain-coding regions of these genes, and

(iii) what are the range of secondary mutations which may lead to the development of the different types of cancers

This chapter will also provide a means of identifying the possible rate of mutagenesis in the mutation carrier. This chapter will also provide new channels to assist in understanding the pathways involved in MMR-defective tumorigenesis.

It is known that the transition from normal to a cancerous state is a result of accumulation of mutations (Frank, 2003; Frank and Nowak, 2004) and that the mutation rate is a key biological feature of somatic cells that determines risk for malignant transformation (Araten et al., 2005). An average mutation rate in a somatic cell in humans is about $1.02 \times 10^{-9}$, which is 17 times higher than the germline rate and 3.3 times higher than the average for yeast and *E. coli* (Lynch, 2010). Under normal circumstances, the repair systems are able to cope with such rates. However, the loss of MMR leads to a mutator phenotype for all cells and this loss of repair functioning does not only affect the frequency at which mutations occur, it also changes the pattern of mutations (Klasen et al., 2005). Lang et al., (2013) used yeast to generate the genome view of the rate, spectrum and distribution of mutations in the absence of MMR function. Of the 19 strains used, they found insertions and deletions at homopolymeric sites to be more common (87%) and predicted that these homopolymer sites, with proximal repeats, host potential drivers of tumorigenesis in MMR-defective cells. Identification of these drivers in human cells could improve the understanding of the mutagenic processes which are likely to occur in the absence of MMR functioning, and also assist in identifying target loci that contribute to carcinogenesis.

#### *3.1.9.1 Research Aims*

The main aim of this chapter of the study was to identify the underlying genetic features involved in initiation and progression of neoplasia in CMMR-D syndrome. The hypothesis is that there are regions in the human genome that

are more susceptible to accumulation of somatic mutations in the absence of MMR function which, as a result, enhance carcinogenesis.

#### *3.1.9.2 Research objectives:*

1) To determine whether CMMR-D individuals exhibit microsatellite instability
2) To determine the germline mutation rate by using whole exome sequencing.
3) To employ bioinformatics tools for pathway-based analysis in order to identify the gene clusters in existing pathways which may be involved in disease initiation and progression

## 3.2 Methods

### 3.2.1 Ethics and consent

All subjects investigated in this study previously provided informed consent as part of the main study (HREC REF 225/2010) and in the case of minors, parents consented on behalf of the children (**Appendix A**). The purpose of the additional consent was to provide permission to use post-mortem material of the demised proband for research purposes. This research was approved by the Human Research Ethics Committee, at UCT (**HREC REF333/2014**) and biological material was collected according to the guidelines of the Declaration of Helsinki (World Medical Association, 2013).

### 3.2.2 Biological material

The biological materials for all subjects in the current study were previously collected with informed consent. For this investigation, eligible subjects included individuals who met the following criteria: a confirmed disease-causing mutation in one of the MMR genes and, where applicable, knowledge of their mutation status following a genetic counseling session. Several different biological material types were used for various aspects of this study. For the

germline genetic testing, DNA was isolated from both saliva and whole blood. For the tissue extractions to identify somatic mutations, DNA was isolated from formalin-fixed paraffin-embedded (FFPE) and freshly-frozen tissue. Described below are the methods used to isolate DNA from various biological specimens.

#### *3.2.2.1 DNA Isolation*

##### *3.2.2.1.1 Whole-blood*

An adjusted version of the Helms' salting-out DNA extraction method (Miller et al., 1988) (http://hdklab.wustl.edu/lab_manual/dna/dna2.html) was used to extract DNA from whole blood. Briefly, the process involved a series of salt-saturated buffers to separate the DNA from the rest of the blood, followed by protein precipitation/removal, and the precipitation of DNA. The DNA was hydrated in TE buffer. Subsequently, samples were stored at either -20°C (short-term storage, or working stock) or -80°C (long-term storage).

##### *3.2.2.1.2 Saliva*

The Oragene® manual purification of DNA (prep IT.L2P) was used to isolate genomic DNA from 0.5mL saliva as per the manufacturer's protocol (http://www.dnagenotek.com/US/pdf/PD-PR-006.pdf). Briefly, the process involved heating samples at 50°C to ensure adequate extraction and nuclease inactivation. This was followed by ethanol precipitation and removal of the co-contaminants such as phenol and proteins. The samples were rehydrated using the DNA storage buffer or 1X TE buffer, then stored at either -20°C (short-term storage) or -80°C (long-term storage).

##### *3.2.2.1.3 Tissue*

DNA isolation from FFPE tissue was carried out using three methods, the ROCHE® DNA FFPE Tissue isolation method, Ion Ampliseq Direct FFPE DNA kit™ and the frozen tissue was isolated using a derivative of salting out method, each of which is explained briefly, below. RAN was isolated using the Quick-RNA mini-prep Zymo kit method.

#### *3.2.2.1.3.1 ROCHE ® DNA FFPE Tissue isolation method*

The isolation was carried out as per manufacturer's protocol and recommendation. The principle of isolation involved removing formalin using xylene, and proteinase K to lyse the cells. The samples were heated to remove the remaining formalin; this was followed by a series of washing steps, with washing buffer to remove the co-contaminants. Finally, DNA was eluted into a tube using an elution buffer, after which it was stored at either -20°C (short-term storage) or -80°C (long-term storage).

#### *3.2.2.1.3.2 Ion Ampliseq Direct FFPE DNA kit™*

The method was performed as per the manufacturer's protocol (**Appendix B**). This method did not make use of deparaffinising with xylol or xylene. The principle of isolation involved the use of two solutions; 'transfer solution' and 'direct reagent' (whose ingredients are proprietory). The tissue was removed from the microscope slide using transfer solution, in to a microfuge tube containing direct reagent. The tube was then incubated in a thermocycler at 65°C for 15 minutes and then at 20°C for 30 minutes, for the purpose of digesting the cell walls/membranes to release DNA. Subsequently, the samples were stored at -20°C until required.

#### *3.2.2.1.3.3 Frozen tissue extraction (DNA)*

This method was adapted from the Helms' salting-out DNA extraction method (Miller et al., 1988). The tissue was dissected into small fragments with a sterile scalpel and transferred into a sterile microfuge tube, followed by the addition of a buffer containing proteinase K and incubation at 37°C to allow for tissue lysis. Subsequently, the samples were subjected to a series of steps of precipitation with ethanol and elution using 1X TE buffer. The DNA samples were stored at -20°C until required.

### *3.2.2.2 DNA Quantification and Quality Control*

To determine the quality and quantity of the DNA samples, spectrophotometry and/or Qubit® fluorometry and gel electrophoresis were performed. Spectrophotometry was performed using the NanoDrop® ND-1000 (Thermo Scientific, Wilmington, U.S.A.). In addition to obtaining the concentrations of DNA in the stock solution, this instrument measures the purity of the solution.

The intensity for DNA was measured at A260, whilst the wavelength of maximum absorbance for protein was A280, and other solvent(s) at A230. The ratio of A260/A280 measures protein levels and a ratio of less than 1.8 indicated protein and phenol contaminants. An A260/A230 ratio below the range of 1.6-2.0 may indicate the presence of co-purified products such as inorganic solvents (http://www.nanodrop.com/Library/nd-1000-v3.7-users-manual-8.5x11.pdf).

Qubit® 3.0 Fluorometry (Thermo Fisher Scientific, MA, U.S.A) is a method used to quantify the nucleic acid and protein concentrations in a given sample. It uses fluorescent dyes, each of which is specific to the type of the molecule (DNA, RNA or protein). The 'free' dyes have low fluorescence but once bound to the target, their fluorescence increases. At a specific level of fluorescence, signals from the dye and DNA mixture are directly proportional to the concentration of the DNA in the solution. The Qubit®3.0 Fluorometer performs the calculations automatically and then produces a reading in ng/uL. The Qubit® dsDNA HS (high sensitivity) Assay kit (Thermo Fisher Scientific, MA, U.S.A.) was used. The assay was carried out as described by the manufacturer (https://tools.thermofisher.com/content/sfs/manuals/Qubit_dsDNA_HS_Assay_UG.pdf).

Subsequent to spectrophotometry and/ or Qubit® readings, the DNA was diluted to a working concentration of 100ng/µL using sterile, distilled water, (Adcock Ingram, Johannesburg, South Africa). Working solutions were stored at –20°C for an integrity test using gel-electrophoresis, as outlined in the following paragraphs.

#### *3.2.2.3 Integrity check: Gel electrophoresis*

Gel-electrophoresis is a method used to separate and visualize nucleic acids according to size. A sample of DNA (usually 100ng/µL in distilled water) was loaded onto an agarose gel submerged in TBE buffer. The gel was then subjected to an electric field, which draws the negatively charged DNA across it. The molecules travel at different speeds depending on their net charge and size (larger fragments travel slower while smaller fragments travel faster and

further) and therefore end up at different positions in the gel, as a function of their size, after a given time.

To prepare the gel, a 1% or 2% [weight per volume agarose, Seakem® LE Agarose Lonza (Rocklands, ME, U.S.A.)], was mixed with TBE buffer (**Appendix C**). In order to enable visualization of the DNA fragments in the gel, a nucleic acid stain, SYBR® Safe stain (1mg/ml) (Sigma, MO, U.S.A.), was added directly into the agarose gel. The DNA stain binds to DNA by intercalating between base pairs, to facilitate visibility of the sample in the gel. Five microliters (5 µL) of GeneRulerTM 100bp DNA Ladder Plus molecular weight marker (0.05ug/µL) (Fermentas®, Vilnius, Lithuania) was loaded into one of the wells on the gel in order to gauge the approximate size of the DNA molecules (i.e. in this case to check if DNA was relatively intact or degraded) (**Appendix C**). Gel electrophoresis was conducted between 110-160 volts (V) for 30 to 40 minutes. The resulting gel was visualized using the Ultra-Violet trans-illuminator (UVI-Tech, Cambridge, UK), and the image processing was performed using the UVI-Pro software (version 12.3).

### 3.2.3 Microsatellite genotyping

#### *3.2.3.1 Marker amplification*

To determine the microsatellite status (i.e. extent of microsatellite instability) in the germline and tumour material, the Promega (Madison, WI, U.S.A.) MSI analysis system, version 1.2, was used. This assay includes labelled primers for co-amplification of five mononucleotide repeat markers (BAT25, BAT26, NR21, NR24 and MONO27) and two penta-nucleotide markers (penta C and penta D) (**Table 3.3**)

The mononucleotide markers were used for MSI testing while the penta-nucleotides were used to ensure accurate amplification and sizing and as a control in case of sample mix-ups and contamination.

**Table 3.3 List of microsatellite Markers for MSI testing (Promega)**

| Marker Name | GenBank® Number | Size Range (bp) |
|---|---|---|
| **NR-21** | XM_033393 | 94–101 |
| **BAT26** | U41210 | 103-115 |
| **BAT-25** | L04143 | 114–124 |
| **NR24** | X60152 | 130-133 |
| **MONO-27** | AC007684 | 142–154 |
| **Penta C** | ALI38752 | 143-194 |
| **Penta D** | AC000014 | 135–201 |

For microsatellite genotyping, the PCR involved the use of nuclease free water, Gold Star 10X buffer (Promega Madison, WI, U.S.A.), MSI 10X primer pair and AmpliTaq Gold DNA polymerase (Promega Madison, WI, U.S.A.), which is a specialized polymerase that performs well in multiplex PCR conditions. The PCR was conducted under conditions described in **Appendix D.** Once the amplification was completed, capillary electrophoresis was carried out.

### *3.2.3.2 Automated capillary electrophoresis*

To prepare samples for capillary electrophoresis, a mixture of an internal lane standard (ILS600) and Hi-Di™ formamide (HiDi), (Applied Biosystems, Foster City, CA) was made to maintain the DNA as a single strand (Sambrook and Russell, 2001a). An internal lane standard 600 (ILS600) serves as a molecular weight marker during this process. The ILS600 specifically, contains 22 DNA fragments ranging from 60 to 600 bases in length. Each fragment is labelled and detected separately in the presence of the MSI during electrophoresis. For this particular experiment, 1uL of ILS600 was added to 9uL of HiDi formamide, to make up the cocktail, and 1uL of the PCR product (of each sample) was then added.

The combination of the sample and cocktail was denatured at $95^0$C for five minutes and immediately snap frozen in ice for three minutes (this is to allow the DNA fragments to remain single stranded in preparation for the automated electrophoresis). Results were analysed using the GeneMapper® software version 4 (Applied Biosystems, Foster City, CA).

### 3.2.4 Next Generation sequencing

Next-generation sequencing (NGS) has emerged as a powerful tool to determine the wide range of DNA variations, from single nucleotide changes to insertions and deletions and other structural changes (Majewski et al., 2011).

The workflow is divided into three major steps (**Figure 3.2**); sample preparation and sequencing, primary data analysis and secondary data processing.

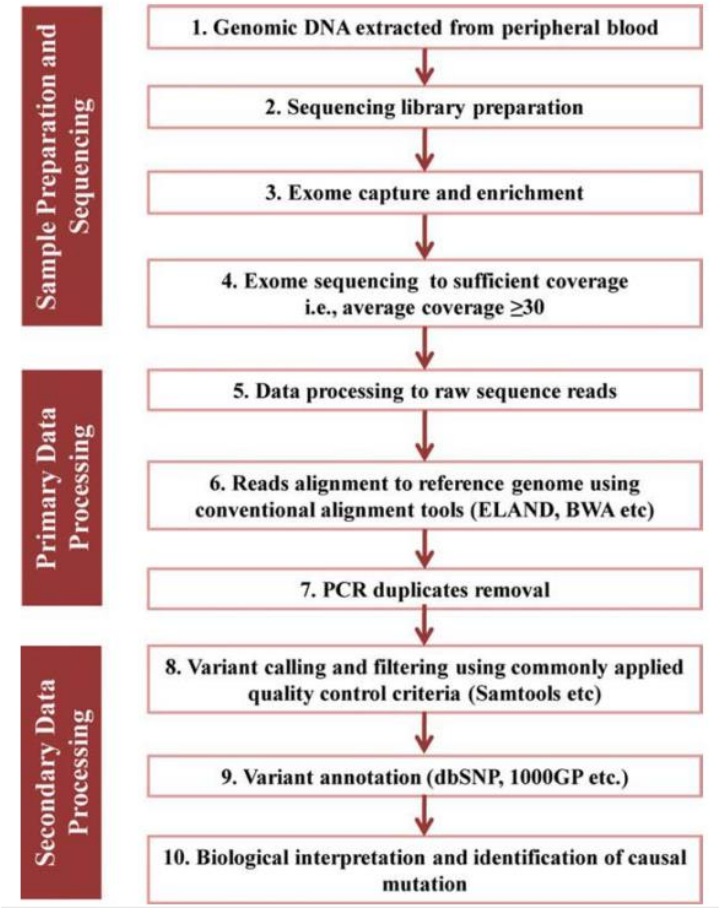



**Figure 3.2: NGS workflow**. Indicated above are the three important stages of NGS, sample preparation and sequencing, primary data processing and secondary data analysis [Image taken from Ku et al. 2012)]

For the purpose of this study, the Ion Torrent, Ion Express™ sequencing platform was utilized to perform WES. The workflow for this platform is explained below.

The samples used in the sequencing included NPC569.1 (father), NPC1.77 (mother). NPC569.3 (proband) and NPC569.4 (sibling). For the parents (NPC569.1 and NPC1.77) DNA was extracted from whole blood, and for the sibling DNA was extracted from saliva. For the proband, DNA was extracted from saliva, and for the subsequent experiments, DNA was extracted from

several tissues (adrenal, bowel, cerebellum, cerebrum, kidney, liver and spleen) which were obtained post mortem.

### *3.2.4.1 Library Construction*

In preparation of the library construction, the quality checks were performed to ensure that the input DNA quality was good enough for library construction. For Ion Torrent WES, the TaqMan RNaseP detection Reagent Kit (Thermo Fisher Scientific, MA, U.S.A.) is recommended to quantify the amplifiable DNA. For each exome library it was recommended that ~50-100ng of genomic DNA (gDNA) be used as an input.

Library construction involved fragmenting the DNA into uniform sizes of about 200 to 400 base pairs, followed by the addition of sequencing adapters (**Figure 3.3**). Sequencing adapters are short pieces of DNA added to the ends of the fragmented DNA. During the sequencing step, these adapters were used to prime DNA replication. The Ion AmpliSeq™ Exome kit was used to construct the library.

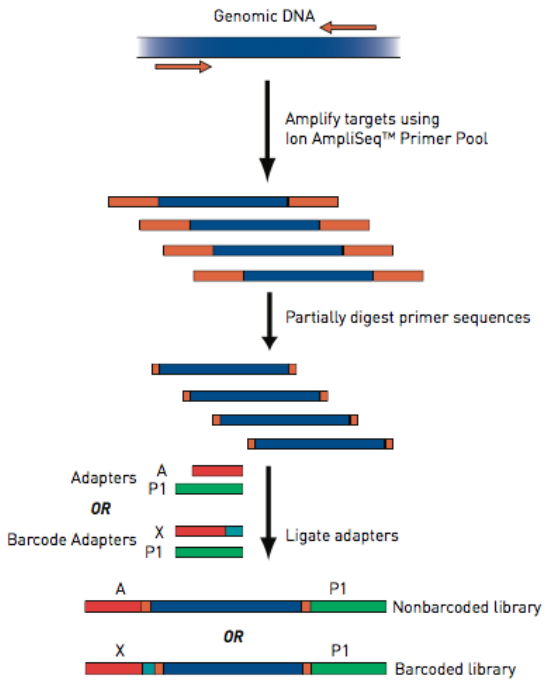



**Figure 3.3: Library construction Ion Torrent sequencing platform**. A step-by-step process from DNA sample to generation of libraries. The generated library for the study was bar-coded, and two samples were run at the same time (https://tools.thermofisher.com/content/sfs/manuals/MAN0010084_AmpliSeq_ExomeRDY_LibraryPrep_UG.pdf)

#### *3.2.4.2 Template Preparation/amplification*

Following the library construction, templates were prepared for amplification. This process starts with attachment of beads to the fragments generated in the library preparation, described above and subsequent amplification of the fragments using emulsion PCR (emPCR). The beads, coated with complementary primers, were mixed with a dilute aqueous solution containing the fragments to be sequenced along with the necessary PCR reagents. This solution was then mixed with oil to form an emulsion of micro-droplets. The clonal amplification of each fragment was then performed within the micro-droplets. The sample emPCR, emulsion breaking and enrichment were performed using the Ion X-press template kit according to the manufacturer's instructions (Thermo Fisher Scientific, MA, U.S.A.). Following amplification, the emulsion was broken by organic extraction and centrifugation, and the amplified beads were enriched in a glycerol gradient.

#### *3.2.4.3 Sequencing*

The Ion Torrent sequencing technology is based on the standard pyro-sequencing chemistry. The nucleotide bases are introduced one at a time and incorporated by the DNA polymerase. The machine then measures the direct release of protons ($H^+$) from the reaction of base introduction. Nucleotide incorporation into the growing complimentary DNA strand causes a release of a Hydrogen ion, which is sensed by a hypersensitive sensor. The technology is faster to run because it does not use optics and processes about 200 reads in two hours. Sequencing was subsequently performed using the sequencing kit v2.0 (Thermo Fisher Scientific, MA, U.S.A.), on the Ion Torrent PGM (Thermo Fisher Scientific, MA, U.S.A.) for 65 cycles with barcoded samples, with two DNA samples/libraries per chip.

### 3.2.5 Data analysis

Data from the PGM runs were processed using the Ion Torrent platform specific pipeline, Torrent Suite software™ (version 4.2.1) (Thermo Fisher Scientific, MA, U.S.A.), to generate sequencing reads, trim the adapter sequences, filter and remove poor signal profile reads. The Torrent Suite software™ (version 4.2.1) is a web based platform which included the analysis pipeline optimized for Torrent

raw data, alignments and variant calling purposes and for this process it uses a plug-in function Torrent variant base caller version 4.2-18/6b3fd1b, configuration: Generic Proton- Germline – high stringency. The analysis was performed using the factory default settings. The BAM files and variant call files (VCF) were generated using the Torrent Suit Software with plug-in variant caller. The filter was set at a depth average of >100, each variant calling of >20, a variant frequency of each sample >5% and p value of <0.01.

### *3.2.5.1 WES Analysis workflow (family analysis)*

The bioinformatics analysis for the germline DNA WES of family NPC569 (**Figure 3.4**) was completed on the Torrent variant caller software version 4.2-18/6b3fd1b using the default settings. The BAM files and VCFs were exported for further analysis. VCF is a text file format (most likely stored in a compressed manner) consisting of a list of all variants identified from a specific genome.

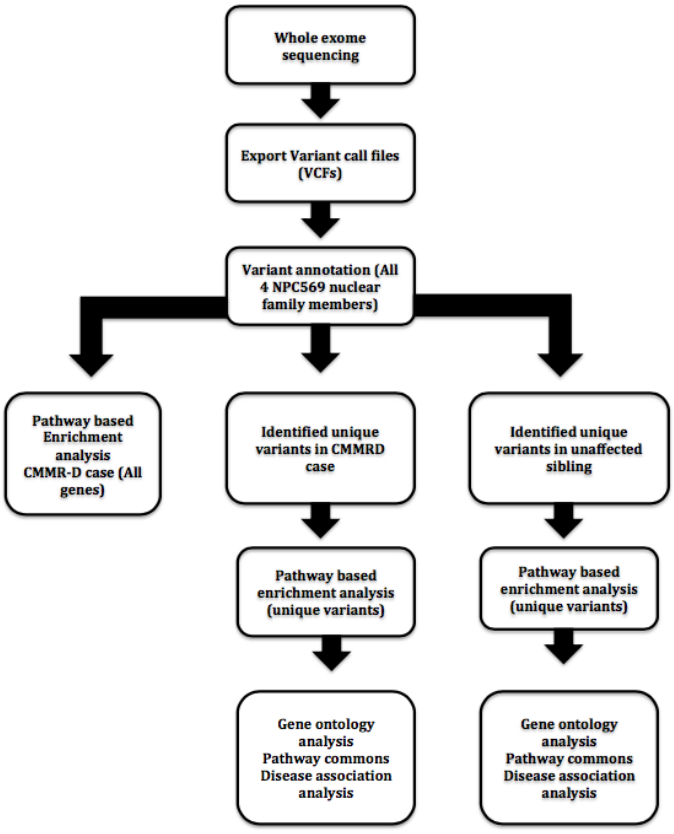



**Figure 3.4: Analysis workflow for the NGS analysis** part of the study with particular interest in identifying novel variants in the NPC569.3 and NPC569.4. These variants were identified by excluding all variants observed in both parents, then put through a pathway based analysis to identify the most enriched pathways.

### *3.2.5.2 Variant annotation and filtering*

Subsequent to variant calling, variant annotation was performed on each sample VCF file and the merged family VCF files. Variant annotation is a process whereby the raw recognized and data was combined to determine certain information such as definition of the variant and genotype call to the variants identified. ANNOVAR was used to independently perform gene-based annotation in each sample and the merged family VCF files were analysed in order to catalogue whether SNPs cause protein-coding changes and whether the amino acids are affected http://annovar.openbioinformatics.org/en/latest/. ANNOVAR also predicts whether the identified variants result in any protein coding changes and if the affected codons may result in a change of amino acids.

This program also performs filter-based annotation, whereby frequency of variants which have been previously reported in other databases, for instance, dbSNP, and 1000 Genomes Project are reported. Annovar is also able to calculate the pathogenicity score for each of the variants. The following ANNOVAR settings were used to annotate variants:

I. Population frequency information for each variant were obtained from 1000 Genomes exome (http://www.internationalgenome.org) and targeted exon datasets and COSMIC (http://cancer.sanger.ac.uk/cosmic),
II. Gene function was obtained from RefGene (http://refgene.com),
III. Functional predictions were obtained from SIFT (http://sift.bii.a-star.edu.sg), PolyPhen 2 (http://genetics.bwh.harvard.edu/pph2/), MutationTaster (http://www.mutationtaster.org), MutationAssessor (http://www.ngrl.org.uk/Manchester/page/mutation-assessor ), LRT (http://www.genetics.wustl.edu/jflab/lrt_query.html), FATHMM (http://fathmm.biocompute.org.uk/fathmmMKL.htm), MetaSVM (http://bioinform.github.io/metasv/), GERP++ (https://omictools.com/genomic-evolutionary-rate-profiling-tool), MetaLR (http://annovar.openbioinformatics.org/en/latest/), PhyloP (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/phyloP46way/) and SiPhy (http://portals.broadinstitute.org/genome_bio/siphy/),

IV. Conserved and segmental duplication sites, dbSNP code and clinical relevance reported in dbSNP138 were also ascertained (https://www.ncbi.nlm.nih.gov/projects/SNP/ )

In order to identify novel variants in the proband (NPC569.3), following the annotation of variants, all variants observed in both parents were excluded. The identified variants were then put through a pathway-based analysis to identify the most enriched pathways. The purpose of the pathway-based analysis was to identify key regulators of disease related gene networks. In addition, this type of analysis provides a bird's eye view of affected biological systems.

The pathway-based analysis was performed using two web-based programmes, WebGestalt (http://bioinfo.vanderbilt.edu/webgestalt/) and Consensus PathDB (http://consensuspathdb.org). Using both programs provided a different but useful approach to interpreting the data. The WebGestalt program, short for *Web-based Gene Set Analysis Toolkit*, is a combination of a functional genome, proteomic and large-scale genetic studies from which a list of large number of genes was generated. WebGestalt incorporates information from a number of different sources and provides an efficient way to make sense of a seemingly disparate list of genes.

The list of identified genes was uploaded to WebGestalt. For the enrichment analysis, *Homo sapiens* was selected as the organism of choice, the gene list was uploaded and then for the analysis, options of the type of enrichment analysis was selected, as were the selection ranges from GO analysis to phenotype-analysis. First, GO analysis was performed; the statistical method used was hyper-geometric, which was the default method for the program. This was followed by the selection of the multiple testing statistical tests (and corrections/adjustments) to be performed, by default the program was set on BH, but it ranges from none to the Bonferroni adjustment for multiple testing. The significance level can also be selected; for the purpose of this study, it was left at 'top ten', as in the top ten results, which the program will produce

regardless of their p-value. This is good for eliminating the unnecessary output. The minimum number of genes per category was set at two by default (and left as that), but can range from two to 10. For this analysis, there were no results with less than 10 genes as an ouput. As mentioned earlier, the program is web-based; as soon as the results are ready, they can be accessed on the web. The analysis for this study was performed using the whole list of genes identified to have variants, and separately, also for a list of genes that were found to have variants only in the proband.

The second pathway-based analysis was performed using the Consensus Path DB (http://consensuspathdb.org) program; also a web-based program, which integrates networks such as binary and complex protein–protein genetic metabolic signaling, gene metabolic signaling, gene regulatory, and drug target as well as biochemical pathways. The data is sourced from 32 different publicly-available databases and the interactions are curated from the literature. Overrepresentation analysis, enrichment and induced network modules analysis were performed. Although the WebGestalt is not the most up to date and does each analysis separately, it allows for additional analyses, whereby one can determine which disease or phenotype has been previously associated with any of the genes on the list.

#### *3.2.5.3 Sub-network analysis*

The aim of this particular part of the analysis was to identify sub-networks of interacting genes, generated from a list of candidate variants (workflow in **Figure 3.5**). From the VCF files of the proband (NPC569.3), sibling (NPC 569.4), father (NPC569.1) and mother (NPC1.77), candidate variants were identified. Declared candidate variants, were those that were ranked through at least six of the eight algorithms (i.e. SIFT, PolyPhen 2, MutationTaster, Mutation Assessor, LRT, FATHMM, MetaSVM and MetaLR) predicting significant impact on functional significance i.e. "deleterious" (D), "probably damaging" (D), "disease causing-automatic" (A) or "disease-causing"(D). The identified variants and

genes were investigated further to identify the phenotypes associated with these sub-networks.

The protein–protein interaction (PPI) was performed using the list of candidate variants identified. To determine this, the protein interaction network analysis platform (http://cbg.garvan.unsw.edu.au /pina/) was used to generate the potential sub-networks (derived from the identified candidate genes) in the four individual samples, as well as in the merged data. The sub-networks identified were compared and subjected to an enrichment analysis using Enrichr (http://amp.pharm.mssm.edu/Enrichr/) (Chen et al., 2013; Kuleshov et al., 2016), to identify the pathways associated with the set of genes, biological processes and associated human phenotypes. Enrichr is a freely available comprehensive resource for curated gene sets and a search engine that accumulates biological knowledge for further biological discoveries. The dataset contains over 180 000 annotated gene sets from over 100 gene set libraries (Kuleshov et al., 2016; Chen et al., 2013).

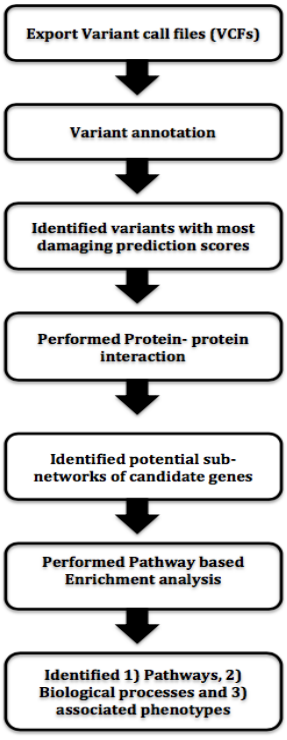



**Figure 3.5: Analysis workflow for sub-network analysis.** The variants with most damaging prediction scores were used to map protein-protein interactions. Followed by the mapping pathways they are involved in, and the biological processes as well as phenotypes they are associated with.

### 3.2.6 WES Analysis workflow: Mutations in different tissues

Exome sequencing of the samples extracted from the five different tissues (adrenal gland, bowel, cerebrum, kidney, and saliva) from the proband was conducted as previously described using the Ion Torrent sequencing technology (Section 3.2.4). Data from the PGM runs were processed using the Ion Torrent platform specific pipeline, Torrent Suite software™ (version 4.2.1), to generate sequencing reads, trim the adapter sequences, filter and remove poor signal profile reads. ANNOVAR (http://annovar.openbioinformatics.org /en/latest/) was used to analyse the generated VCFs. Further analysis was conducted to identify variants only observed in individual tissues, and shared variants amongst all the tissue types. To identify the most biologically enriched pathways, pathway-based analysis was carried out using EnrichR as described in Section 3.2.5.3.

### 3.2.7 NGS panel for Lynch syndrome susceptibility genes

In this NGS design, 13 Lynch syndrome/ cancer susceptibility genes were sequenced. Most of the amplicons had a mean depth of 900 per sample. This part of the sequencing was performed to validate some of the sequencing results obtained in the WES previously. The 13 genes which probes were designed for included: *APC, BMPPR1, EPCAM, MLH1, MSH2, MSH6, MUTYH, PMS2, POLD1, POLE, PTEN, STK11, TP53*. To analyze the sequences; the Torrent Suite software™ (version 4.2.1) was used to generate sequencing reads, trim the adapter sequences, filter and remove poor signal profile reads. This was followed by the annotation of variants using ANNOVAR http://annovar.openbioinformatics.org /en/latest/). The identified variants were compared to those obtained in the WES analysis, for validation purposes.

## 3.3 Results

### 3.3.1 DNA quantification and quality control

Several methods were used to isolate DNA from the different types of samples. These methods were generally successful in yielding good quality DNA i.e. DNA concentrations of >100ng/µL (**Table 3.4)** and good DNA integrity as indicated by gel electrophoresis (**Figure 3.6)**.

**Table 3.4 DNA Sample quality analysis output**

| Sample ID | Sample origin | Sample type | Isolation Method | Nano-drop concentration | Qubit results * | Gel Electrophoresis |
|---|---|---|---|---|---|---|
| **NPC569.1** | Blood | DNA | Salting out | 369ng/µL | - | ✓ |
| **NPC1.77** | Blood | DNA | Salting out | 229ng/µL | - | ✓ |
| **NPC569.4** | Saliva | DNA | Oragene | 238ng/µL | - | ✓ |
| **NPC569.3** | Saliva | DNA | Oragene | 219ng/µL | - | ✓ |
| **NPC569.3** | Tumour | DNA | RocheDNA FFPE Tissue Isolation method | 55ng/µL | - | ✓ |
| **NPC569.3** | Adrenal gland | DNA | Salting out | 618ng/µL | 55ng/µL | ✓ |
| **NPC569.3** | Bowel | DNA | Salting out | 390ng/µL | 54ng/µL | ✓ |
| **NPC569.3** | Cerebellum | DNA | Zymo-Spin column DNA isolation Kit | 272ng/µL | 49.3ng/µL | ✓ |
| **NPC569.3** | Cerebrum | DNA | Salting out | 152ng/µL | 49.6ng/µL | ✓ |
| **NPC569.3** | Kidney | DNA | Salting out | 465ng/µL | 55ng/µL | ✓ |
| **NPC569.3** | Liver | DNA | Zymo-Spin column DNA isolation Kit | 404ng/µL | 40.5ng/µL | ✓ |
| **NPC569.3** | Spleen | DNA | Zymo-Spin column DNA isolation Kit | 352ng/µL | 75ng/µL | ✓ |

*Qubit ® 3.0 Fluorometer DNA concentrations were only determined for samples that were subsequently used for sequencing.

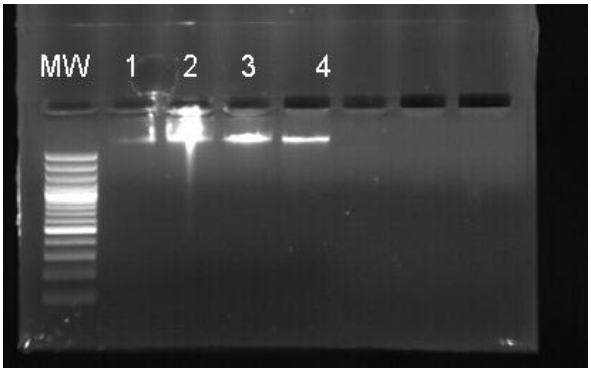



**Figure 3.6: Agarose gel electrophoresis showing integrity of DNA from members of Family NPC569.** MW= molecular weight marker, lane 1: NPC569.1, lane 2: NPC1.77, lane 3: NPC569.3 and lane 4: NPC569.4.

Both the Roche® and the Ion Ampliseq® Direct FFPE DNA kits failed to produce sufficient DNA from the FFPE sample for WES. However, the quality of the

isolated DNA using the Roche® FFPE kit was satisfactory for microsatellite status testing.

The standard salting out method for DNA isolation was performed on all seven samples of freshly frozen tissue (obtained post mortem) from the deceased subject. However, integrity checks using gel electrophoresis (**Figure 3.7**) showed that three of the seven samples were degraded to some extent and appeared to have protein contaminants. These contaminants would have the potential to inhibit some of the downstream processing such as PCR. For this reason, new samples were sectioned and the Zymo-Spin® column DNA isolation kit (Irvine, CA, U.S.A.) was used to extract DNA for which the yield was satisfactory.

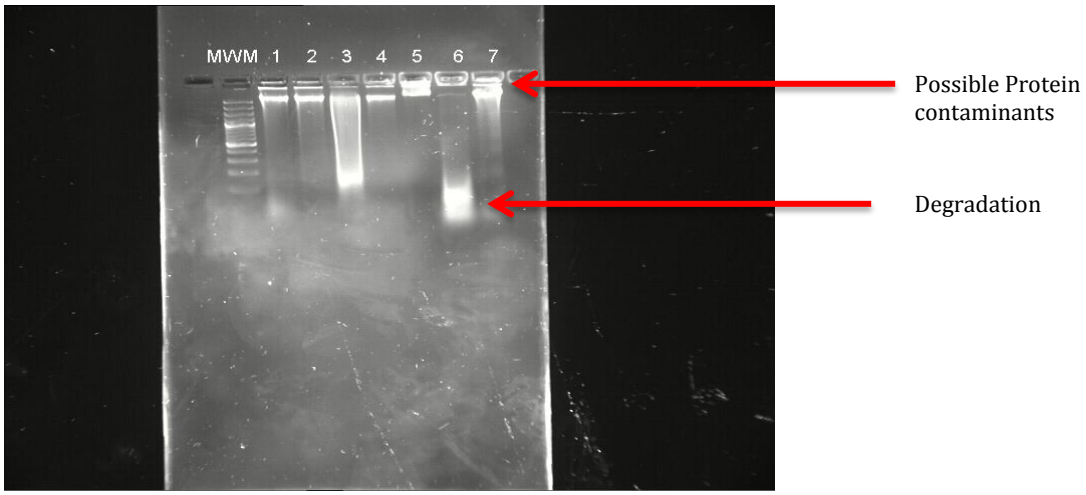



**Figure 3.7: Integrity gel of DNA isolated from frozen tissue from subject NPC569.3** MWM=Molecular weight marker, Lane 1: Adrenal gland, Lane 2: Colon, Lane 3: Cerebrum, Lane 4: Cerebellum, Lane 5: Kidney, Lane 6: Liver, and Lane 7: Spleen.

DNA concentrations measured using a Nano-drop spectrophotometer indicated that the samples were all satisfactory for further analysis. Subsequently, Qubit® 3.0 Fluorometry was performed only on samples used for WES **(Table 3.4**). Since the Nano-drop and Qubit® 3.0 Fluorometer quantification methods only determine the concentration of the DNA, gel-electrophoresis was used to check the integrity of the samples. This process determines the level of degradation and/or contaminants, as seen in **Figure 3.6** and **Figure 3.7.**

The working concentration for all reactions was 100ng/µL. However, for WES, DNA concentrations were required to be at least 750ng/µL. For this process i.e. WES, additional checks were carried out to ensure quality of samples for amplification. These checks included: (i) the TaqMan® RNAse P detection assay which determined the quality of the sample prior to library preparation and, (ii) qPCR assay which determined the concentration of the unamplified libraries before sequencing.

### 3.3.2 MSI genotype testing

All markers amplified successfully for all samples. There was no significant difference observed in the microsatellite status between germline DNA (saliva) and somatic DNA (tumour) of the CMMR-D patient (NPC569.3) (**Figure 3.8)**, i.e. all markers appeared stable for both the germline and somatic samples. In addition, when comparing germline marker status between NPC569.3, her sibling (NPC569.4) and both parents, (NPC569.1 and NPC1.77), (all three of the latter who are heterozygous for the c.1528C>T mutation), no significant difference in microsatellite status was observed (data not shown).

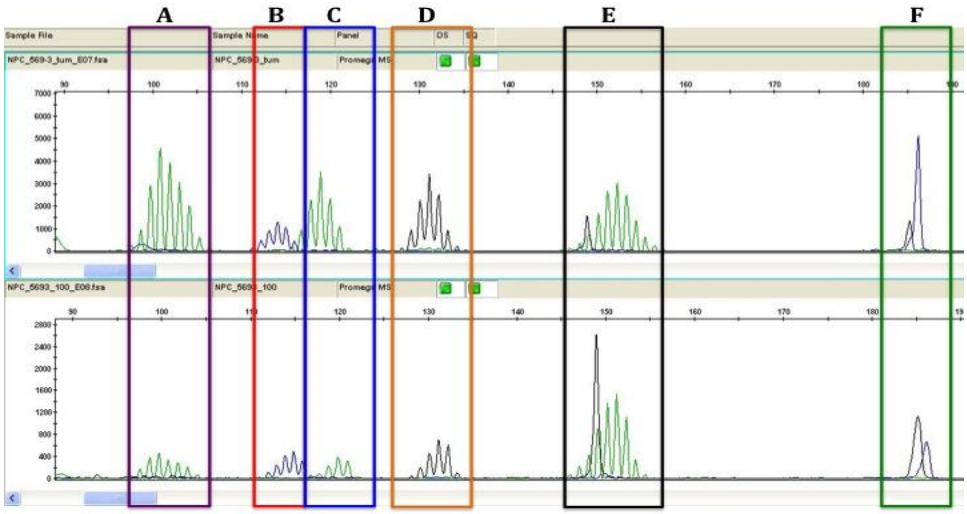



**Figure 3.8: Microsatellite instability analysis results. Tumour (somatic) DNA in the top panel versus normal (germline – as seen in DNA from saliva), in the bottom panel, of NPC 569.3.** Seven markers were genotyped, *A: NR21, B: BAT-26, C: BAT-25, D: NR24, E: Mono-27 (green) & Penta-C (black), F: Penta-D (blue) & Penta C (black).*

### 3.3.3 WES results of the family study

Following WES and variant annotation, more than 25 000 variants were identified in each individual (**Table 3.5**). Notably, the proband (NPC569.3) and

her mother (NPC.177) had more variants (26 523 and 26 948, respectively) compared to her father (NPC 569.1) who had 24 828 and her sibling (NPC 569.4) who had 25 157 variants.

**Table 3.5 WES Variant call summary (CMMR-D proband highlighted in red)**

| Sample no.<br>Type of Variant | NPC569.1<br>(father) | NPC569.3<br>(proband) | NPC569.4<br>(sibling) | NPC1.77<br>(mother) |
|---|---|---|---|---|
| Total variants | 24 828 | 26 523 | 25 157 | 26 948 |
| Intronic | 1 694 | 1 834 | 1 666 | 1 792 |
| Exonic | **22 452** | **23 874** | **22 789** | **24 334** |
| UTR | 479 | 594 | 492 | 580 |
| Splicing | 70 | 64 | 64 | 80 |
| *Others | 133 | 167 | 131 | 162 |
| Non-synonymous | 10 166 | 10 769 | 10 214 | 10 871 |
| Synonymous | 11 481 | 12 288 | 11 816 | 12 637 |
| Frameshift | 118 | 131 | 114 | 113 |
| Non-frameshift | 8 | 9 | 7 | 12 |
| Stop gain/loss | 90 | 88 | 85 | 94 |
| Unknown | 353 | 371 | 343 | 378 |
| N/A | 2 612 | 2 837 | 2 578 | 2 843 |

*Others= intergenic, upstream, downstream and ncRNA, N/A=”.” never been reported mapped variants

In the first instance, WES was performed to determine if there were any obvious genetic differences observed between the CMMR-D proband (NPC569.3) and related individuals (NPC569.4, NPC569.1 and NPC1.77). The WES was performed on DNA isolated from both saliva and blood, extracting DNA from FFPE samples (tumour DNA) for the purpose of exome sequencing was unsuccessful; all the samples isolated did not yield desirable amounts or quality of DNA to conduct high throughput sequencing.

Having identified 26 523 variants in NPC569.3 (proband), who was homozygous for the *MLH1* disease-causing mutation, the next step was to determine if there were regions of the genome that might have accumulated more variants compared to the variant distribution in the rest of her family members – each of whom was heterozygous for the disease-causing mutation. The gross distribution of variants across the genome indicated no obvious preferential region of variant accumulation (**Figure 3.9**). Chromosome 1 and 19 had the most number of variants amongst all of the chromosomes, with an average of

more than 2000 variants per chromosome (chromosome 1 and 19) in each of the four individuals (**Figure 3.9**). There was no significant difference in the number of variants per chromosome between the proband and her sibling.

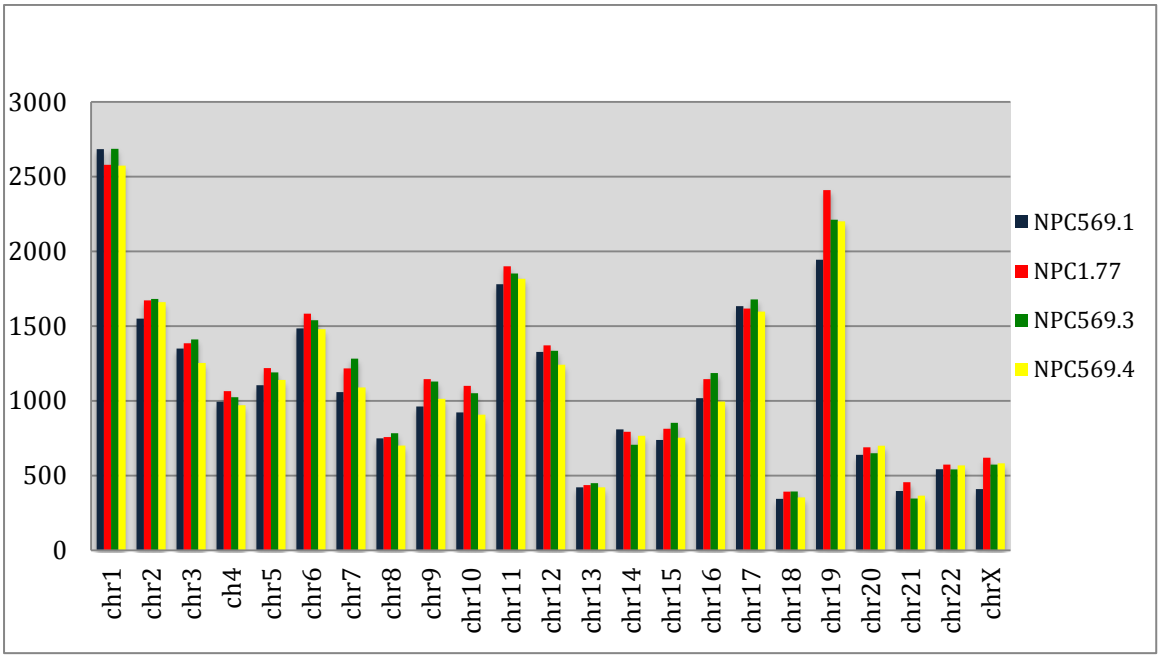



**Figure 3.9 Distribution of variants per chromosome.** Variants observed in all four members of the family NPC569 are listed by chromosome, each bar / colour represents an individual. NPC569.1 (father), NPC1.77 (mother), NPC569.3 (proband) and NPC569.4 (sibling).

### *3.3.3.1 Number of new "de novo" variants*

One of the main consequences of having a deficient mismatch repair system is the accumulation of mutations with every cycle of DNA replication. It has been hypothesised that individuals with CMMR-D accumulate an even higher number of mutations/variants over time. This section of the study investigated whether the number of *de novo* variants identified in the proband, NPC569.3, is significantly higher than the sibling (NPC569.4) who has one defective copy (heterozygous form) of the MMR gene. To do this, all variants that the proband inherited from her parents (i.e. all variants shared between the proband and her parents) were excluded for this analysis. In total, 964 of the total of 26 523 variants (3.63%) were unique to the proband. Similar analysis for the sibling revealed 755 of a total of 25 157 (3%) variants that were unique to the sibling. The sibling had 17.3% less variants than the proband. Some of the *de novo* variants observed (n=185) were common to both the proband, NPC569.3, and the sibling, NPC569.4 (but were not in their parents). The distribution of new

variants indicated that NPC569.3 (in blue) had more variants compared to NPC569.4 (red) (**Figure 3.10**), across most but not all of the chromosomes (**Figure 3.10**).

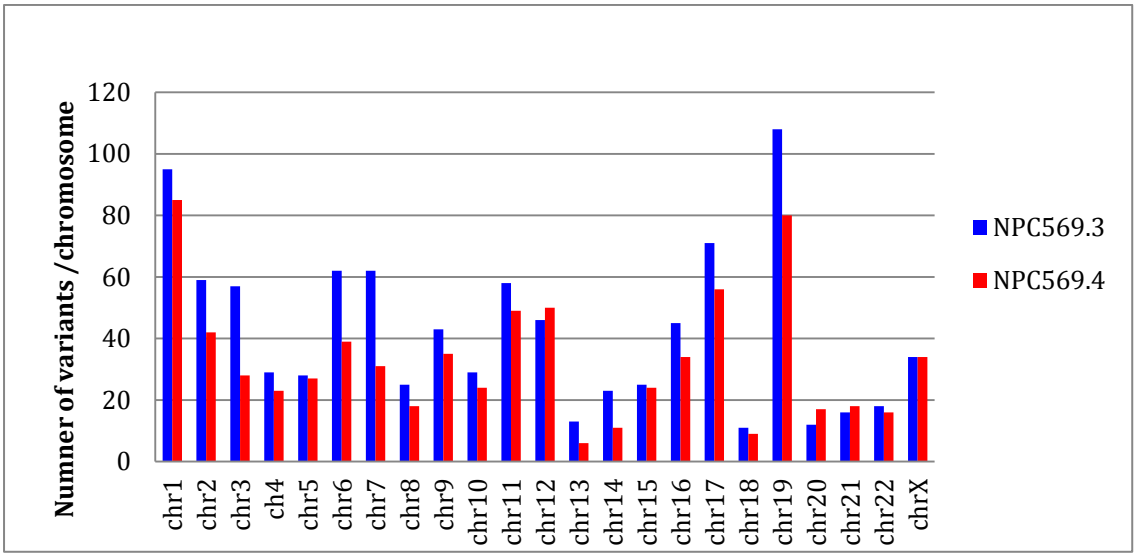



**Figure 3.10 Distribution of new variants between NPC569.3 and NPC569.4.** Blue = proband (NPC569.3) and Red = sibling (NPC569.4). The figure depicts the distribution of all non-shared variants between the siblings across all the chromosomes**.**

When comparing NPC569.3 and NPC569.4, there was no significant difference in the type of variants observed. Even though the number of variants differed, for both samples, 60% of the variants were made up of non-synonymous variants and about 2% of all the new variants had been previously reported (**Figure 3.11**).

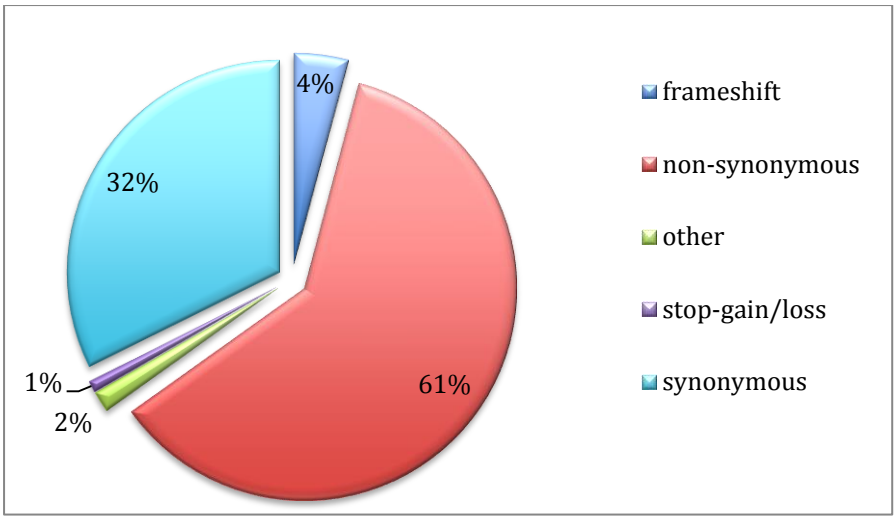



**Figure 3.11 Types of new variants in NPC 569.3** – in (germline DNA from the) saliva sample. This figure indicates the types of variants observed in the proband. Most were non-synonymous variants accounting for 61% of all variants.

For NPC569.3, most (569/964) of the *de novo* variants had not been reported; whereas 369 had been reported in dbSNP previously. Similarly, for NPC569.4, of all the variants, 479 did not have a dbSNP identifier. Of those that had been reported in both individuals, only 14 variants (**Table 3.6**) had been reported in the catalogue of somatic mutations in cancer (COSMIC) (http://cancer.sanger.ac.uk/cosmic).

**Table 3.6 List of shared COSMIC variants in NPC569.3 and NPC569.4**

| Chr | REF | ALT | *GENE* | rs ID | COSMIC ID (Type of cancer associated with variant) |
|---|---|---|---|---|---|
| **chr1** | C | T | *NOTCH2NL* | rs140871032 | ID=COSM166701,COSM166700;OCCURENCE=1 (haematopoietic_and_lymphoid_tissue) |
| **chr1** | T | C | *F5* | rs6028 | ID=COSM146676;OCCURENCE=1(stomach) |
| **chr5** | C | T | *MYO10* | rs16868979 | ID=COSM449271;OCCURENCE=1(breast) |
| **chr5** | C | T | *SPINK5* | rs33920397 | ID=COSM149975;OCCURENCE=1(stomach) |
| **chr6** | T | C | *DSP* | rs36087964 | ID=COSM451787;OCCURENCE=2(breast) |
| **chr6** | C | A | *HLA-B* | rs1131215 | ID=COSM1443267;OCCURENCE=1(colon) |
| **chr7** | C | A | *LAMB4* | rs9690688 | ID=COSM452142;OCCURENCE=1(breast) |
| **chr7** | G | A | *ZNF775* | rs7780011 | ID=COSM452592;OCCURENCE=1(breast) |
| **chr8** | G | A | *SBSPON* | rs150036727 | ID=COSM1101550;OCCURENCE=1(endometrium) |
| **chr9** | T | C | *ROR2* | rs10820900 | ID=COSM1569878;OCCURENCE=1(colon) |
| **chr10** | C | T | *ADARB2* | rs2271275 | ID=COSM1560925;OCCURENCE=1(colon) |
| **chr11** | C | T | *MAML2* | rs113349418 | ID=COSM1561901;OCCURENCE=1(colon) |
| **chr17** | C | A | *EPN3* | rs4794159 | ID=COSM436852;OCCURENCE=1(breast) |
| **chr19** | A | C | *SRRM5* | rs3815422 | ID=COSM148658;OCCURENCE=1(stomach) |

**Chr= chromosome, REF= reference allele, ATL= Alternate allele, rs ID= SNP identity number, COSMIC ID= COSMIC identification number**

"Mutational signature" is a concept which has been described recently, the aim of which is to distinguish the type of cancers by the type of mutations they accumulate. There have been more than 30 signatures described (and proven) to date (Nik-Zainal et al., 2012a, 2012b; Alexandrov et al., 2013; Alexandrov and Stratton, 2014). Amongst these is a 'mismatch repair deficiency'- signature, which is linked to the presence of the type of changes commonly seen in MMR-deficient tumours. Signatures 6 and 16, associated with MMR deficiency, were investigated amongst the *de novo* variants, in germline DNA. These signatures are commonly seen with the predominance of the **C>T** and **T>C** changes. This was observed in our results (**Figure 3.12**).

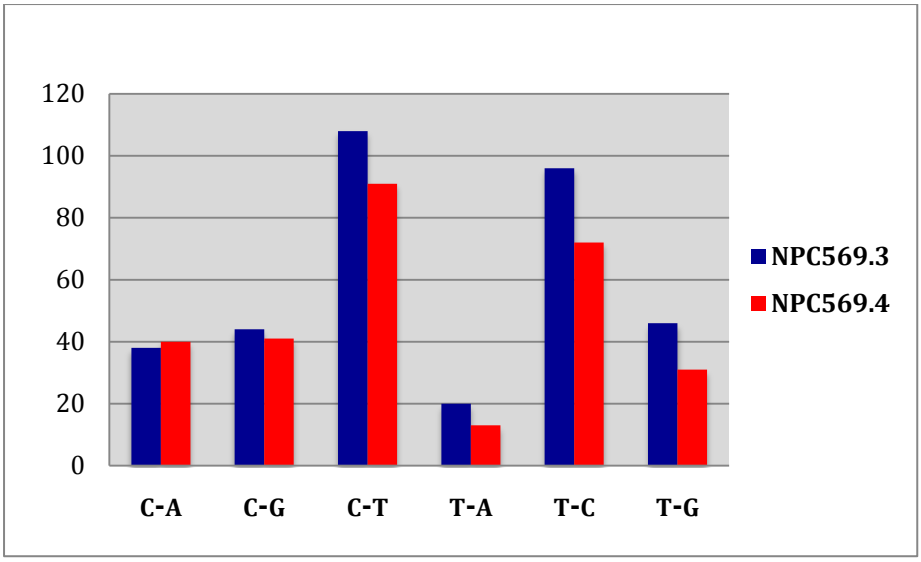



**Figure 3.12 Illustration of the mutational signatures.** Relative occurrence of specific base-pair changes amongst the *de novo* variants in the DNA of the proband (NPC569.3) and sibling (NPC 569.4)

### *3.3.3.2 Pathway based analysis*

In an effort to identify whether specific genes or regions of the genome may be more susceptible to mutations, the purpose of this section was to analyse the most 'variant-enriched' genes and pathways. Two pathway-based analysis programs were used: ConsensusPathDB and WebGestalt. The list of genes with unique variants was investigated using ConsensusPathDB. In total, 964 variants in 776 different genes were uploaded. For the enriched pathway based set, 57 genes were identified and the top ten enriched pathways included collagen biosynthesis, collagen formation and the extracellular matrix (**Table 3.7**).

**Table 3.7 Output of the pathway-based analysis for unique variants in subject NPC569.3**

| Pathway | p-value | Source |
|---|---|---|
| **Collagen biosynthesis and modifying enzymes** | 4.5319822194e-08 | Reactome |
| **Termination of O-glycan biosynthesis** | 2.32210853602e-06 | Reactome |
| **Collagen formation** | 2.40720716677e-06 | Reactome |
| **Protein digestion and absorption - *Homo sapiens* (human)** | 3.64596367622e-06 | KEGG |
| **O-linked glycosylation** | 4.89505355362e-06 | Reactome |
| **Signalling by NOTCH** | 2.52534301889e-05 | Reactome |
| **Integrin** | 3.76067907245e-05 | INOH |
| **Extracellular matrix organization** | 0.000207526993005 | Reactome |
| **NOTCH1 Intracellular Domain Regulated Transcription** | 0.000240906460416 | Reactome |
| **Extracellular matrix-receptor interaction – *Homo sapiens* (human)** | 0.000274061819978 | KEGG |

For the Gene Ontology based set (**Table 3.8**), the extracellular matrix, complex collagen trimmers and proteinaceous extracellular matrix were amongst the most enriched Gene Ontology sets. Lastly, for the enriched protein complex-based sets, the extracellular matrix, integrin-signalling and ALL-1, (a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation) were among the top enriched pathways.

**Table 3.8 Gene-Ontology Enriched based set for NPC 569.3 unique variants**

| Term name | p-value | GO-id |
|---|---|---|
| **Extracellular matrix structural constituent** | 1.93652371537e-08 | GO:0005201 |
| **Complex of collagen trimers** | 1.08250495618e-07 | GO:0098644 |
| **Proteinaceous extracellular matrix** | 1.50391421233e-06 | GO:0005578 |
| **Structural molecule activity conferring elasticity** | 2.75039949804e-05 | GO:0097493 |
| **Fibrillar collagen trimer** | 0.00010947904853 | GO:0005583 |

The second pathway analysis program used was WebGestalt. For this, all variants identified during the WES as well as the list of genes which were identified during the unique variants analysis were analysed.

For the Gene Ontology analysis, the results were divided into three sections: 1) biological process, 2) molecular function, and 3) cellular component. Under the rubric of 'biological process', genes involved in biological adhesion, detection of chemical stimuli, cellular adhesion, and extracellular structure and organisation emerged. Under 'molecular function', molecular adhesion, receptor activity, transmembrane signalling, ion binding and calcium binding, emerged. For 'cellular component', the extracellular membrane, intrinsic to membrane, cell membrane and proteinaceous extracellular membrane, were identified. The extracellular membrane was the most enriched pathway in the unique variant list as well, and this is a trend seen in the independent 'molecular function' and 'cellular component' analyses. Worthy and of note, is that the top five diseases associated with the list of genes are diseases of the brain/ central nervous system.

The same analysis was performed using all of the genes that have been found to have variants in them, and the results showed disorders pertinent to collagen

and the brain (**Table 3.9**) for the proband (NPC569.3), while the analysis for the sibling (NPC569.4) revealed disorders/processes pertinent to collagen, adhesion and myocardial infarction, amongst others (**Table 3.10**).

**Table 3.9 Disease-association analysis for variants in subject NPC569.3**

| Disease | Number of gene | Statistics |
|---|---|---|
| **Collagen Diseases** | 14 | C=96;O=14;E=1.61;R=8.69;rawP=8.40e-10;adjP=8.37e-07 |
| **Brain Diseases** | 25 | C=411;O=25;E=6.90;R=3.62;rawP=4.42e-08;adjP=4.41e-05 |
| **Nervous System Diseases** | 33 | C=694;O=33;E=11.65;R=2.83;rawP=1.22e-07;adjP=0.0001 |
| **Central Nervous System Diseases** | 24 | C=438;O=24;E=7.35;R=3.26;rawP=5.48e-07;adjP=0.0005 |
| **Mitral Valve Prolapse** | 6 | C=21;O=6;E=0.35;R=17.02;rawP=9.60e-07;adjP=0.0010 |
| **Tremor** | 8 | C=54;O=8;E=0.91;R=8.82;rawP=3.20e-06;adjP=0.0032 |
| **Genetic Predisposition to Disease** | 33 | C=808;O=33;E=13.56;R=2.43;rawP=3.45e-06;adjP=0.0034 |
| **Myasthenia Gravis** | 7 | C=39;O=7;E=0.65;R=10.69;rawP=3.52e-06;adjP=0.0035 |
| **Schizophrenia** | 20 | C=360;O=20;E=6.04;R=3.31;rawP=3.86e-06;adjP=0.0038 |
| **Mental Disorders** | 26 | C=564;O=26;E=9.47;R=2.75;rawP=4.59e- |

**Table 3.10 Disease association analysis of variants in the sibling, NPC569.4**

| Disease | Number of genes | Statistics |
|---|---|---|
| **Collagen Diseases** | 12 | C=96;O=12;E=1.33;R=9.00;rawP=1.00e-08; adjP=5.77e-06 |
| **Adhesion** | 30 | C=647;O=30;E=8.99;R=3.34;rawP=1.31e-08; adjP=5.77e-06 |
| **Myocardial Infarction** | 15 | C=242;O=15;E=3.36;R=4.46;rawP=1.81e-06; adjP=0.0004 |
| **Bronchial Diseases** | 17 | C=307;O=17;E=4.26;R=3.99;rawP=1.78e-06; adjP=0.0004 |
| **Vascular Diseases** | 18 | C=357;O=18;E=4.96;R=3.63;rawP=3.35e-06; adjP=0.0006 |
| **Hypertension** | 14 | C=227;O=14;E=3.15;R=4.44;rawP=4.21e-06; adjP=0.0006 |
| **Infarction** | 14 | C=236;O=14;E=3.28;R=4.27;rawP=6.58e-06; adjP=0.0008 |
| **Curvature of spine NOS** | 7 | C=53;O=7;E=0.74;R=9.51;rawP=8.53e-06; adjP=0.0009 |
| **Cardiovascular Diseases** | 19 | C=425;O=19;E=5.90;R=3.22;rawP=1.01e-05; adjP=0.0010 |
| **Lupus Erythematosus, Systemic** | 13 | C=218;O=13;E=3.03;R=4.29;rawP=1.32e-05; adjP=0.0012 |

#### *3.3.3.3 Sub-network analysis for predicted damaging variants*

In order to identify the sub-network of interactive genes and the pathways which they are involved in, a list of genes was compiled from the variants with the most damaging prediction scores (**Table 3.11);** also included in the list was the disease causing variant, *MLH1* c.1528C>T (**Table 3.11**).

This list was compiled using data from all four individuals. Some of the variants identified as most damaging, were shared amongst all four individuals (highlighted in red in **Table 3.11**). There were 11 variants identified in the proband (NPC569.3), 13 from the sibling (NPC569.4), 9 and 12 from each of the parents, NPC569.1 and NPC1.77, respectively. This list of variants was utilised to determine the networks and interactions (**Figures 3.13 and 3.14**).

**Table 3.11 List of filtered variants with most damaging prediction scores**

| Sample | dbSNP ID | #CHROM | *refGene* | FuncGene | SIFT | HDIV | HVAR | LRT | MT | MA | FA | RI | LR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **NPC569.3** | | | | | | | | | | | | | |
| | rs113718290 | chr1 | *ARID1A* | exonic | T | P | B | D | D | L | T | T | T |
| | rs113576347 | chr2 | *PLEKHH2* | exonic | T | B | B | D | D | L | T | T | T |
| | rs2306369 | chr4 | *KIAA1109* | exonic | T | B | B | D | D | N | T | T | T |
| | rs33977775 | chr8 | *NPBWR1* | exonic | D | D | D | D | P | M | D | T | T |
| | rs17417407 | chr10 | *PLCE1* | exonic | T | D | D | N | P | L | T | T | T |
| | rs36006049 | chr11 | *NAT10* | exonic | T | B | B | D | D | M | T | T | T |
| | rs11539445 | chr12 | *YARS2* | exonic | D | P | B | D | P | N | T | T | T |
| | rs4778078 | chr15 | *RGMA* | exonic | D | B | B | U | P | L | D | T | T |
| | rs112038587 | chr17 | *FAAP100* | exonic | D | D | P | N | N | N | T | T | T |
| | rs13047478 | chr21 | *C21orf58* | exonic | D | D | P | N | P | L | T | T | T |
| | rs63749923 | chr3 | *MLH1* | exonic | . | . | . | . | . | . | . | . | . |
| **NPC569.4** | | | | | | | | | | | | | |
| | rs373793845 | chr1 | *PLEKHG5* | exonic | T | D | D | D | D | L | T | T | T |
| | rs344141 | chr4 | *SHROOM3* | exonic | T | D | D | N | P | M | T | T | T |
| | rs2306369 | chr4 | *KIAA1109* | exonic | T | B | B | D | D | N | T | T | T |
| | rs17850652 | chr6 | *RARS2* | exonic | D | B | B | D | P | N | T | T | T |
| | rs17417407 | chr10 | *PLCE1* | exonic | T | D | D | N | P | L | T | T | T |
| | rs7081076 | chr10 | *SORBS1* | exonic | D | D | D | N | D | N | T | T | T |
| | rs10895991 | chr11 | *CAPN1* | exonic | T | D | D | D | P | M | D | T | T |
| | rs34982553 | chr17 | *RILP* | exonic | T | D | D | N | P | L | T | T | T |
| | rs144623089 | chr18 | *DCC* | exonic | D | B | B | D | D | N | T | T | T |
| | rs13047478 | chr21 | *C21orf58* | exonic | D | D | P | N | P | L | T | T | T |
| | rs6151415 | chr22 | *ARSA* | exonic | D | D | P | N | D | N | D | T | T |
| | rs375687463 | chrX | *PHKA2* | exonic | D | D | D | D | D | M | D | D | D |
| | rs63749923 | chr3 | *MLH1* | exonic | . | . | . | . | . | . | . | . | . |
| **NPC569.1** | | | | | | | | | | | | | |
| | rs2306369 | chr4 | *KIAA1109* | exonic | T | B | B | D | D | N | T | T | T |
| | rs76067797 | chr7 | *CDHR3* | exonic | D | P | P | N | N | M | D | T | T |
| | rs17417407 | chr10 | *PLCE1* | exonic | T | D | D | N | P | L | T | T | T |
| | rs2277339 | chr12 | *PRIM1* | exonic | D | D | P | D | P | L | T | T | T |
| | rs3743398 | chr15 | *ACAN* | exonic | T | D | D | N | P | M | T | T | T |
| | rs34745339 | chr16 | *GALNS* | exonic | D | P | B | D | D | M | D | D | D |
| | rs144623089 | chr18 | *DCC* | exonic | D | B | B | D | D | N | T | T | T |
| | rs10425488 | chr19 | *OCEL1* | exonic | T | D | D | N | P | N | T | T | T |
| | rs63749923 | chr3 | *MLH1* | exonic | . | . | . | . | . | . | . | . | . |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **NPC1.77** | | | | | | | | | | | | | |
| | rs113718290 | chr1 | *ARID1A* | exonic | T | P | B | D | D | L | T | T | T |
| | rs344141 | chr4 | *SHROOM3* | exonic | T | D | D | N | P | M | T | T | T |
| | rs33977775 | chr8 | *NPBWR1* | exonic | D | D | D | D | P | M | D | T | T |
| | rs7081076 | chr10 | *SORBS1* | exonic | D | D | D | N | D | N | T | T | T |
| | rs36006049 | chr11 | *NAT10* | exonic | T | B | B | D | D | M | T | T | T |
| | rs192902098 | chr13 | *PDX1* | exonic | D | D | D | D | D | M | D | D | D |
| | rs4778078 | chr15 | *RGMA* | exonic | D | B | B | U | P | L | D | T | T |
| | rs8074498 | chr17 | *ASPSCR1* | exonic | D | D | D | D | P | M | T | T | T |
| | rs12978266 | chr19 | *DOCK6* | exonic | D | D | P | D | P | M | T | T | T |
| | rs13047478 | chr21 | *C21orf58* | exonic | D | D | P | N | P | L | T | T | T |
| | rs375687463 | chrX | *PHKA2* | exonic | D | D | D | D | D | M | D | D | D |
| | rs63749923 | chr3 | *MLH1* | exonic | . | . | . | . | . | . | . | . | . |

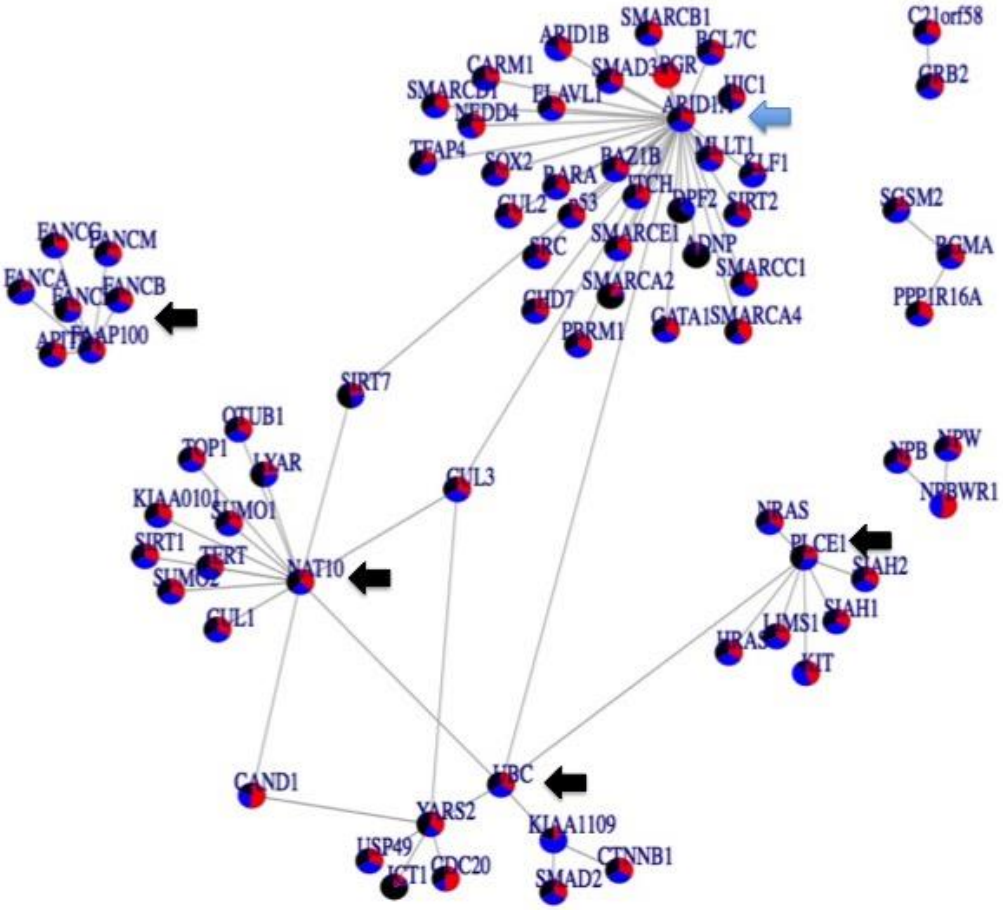



**Figure 3.13 Extracted sub-network of interactions for NPC 569.3**
The extracted sub-network of interactions for the list of variants with the most damaging predicted scores. The identified hubs for the proband (NPC 569.3) include *FANC, NAT10, UBC, PLEC1 and ARID1A*, indicated by arrows. *Figure key: each pie chart is colour co-ordinated according to the frequency of the each allele represented. The three different populations represented herein include, African (black), East Asian (blue) and European (red). The Mixed Ancestry population of a Western Cape are an admixture of these populations.

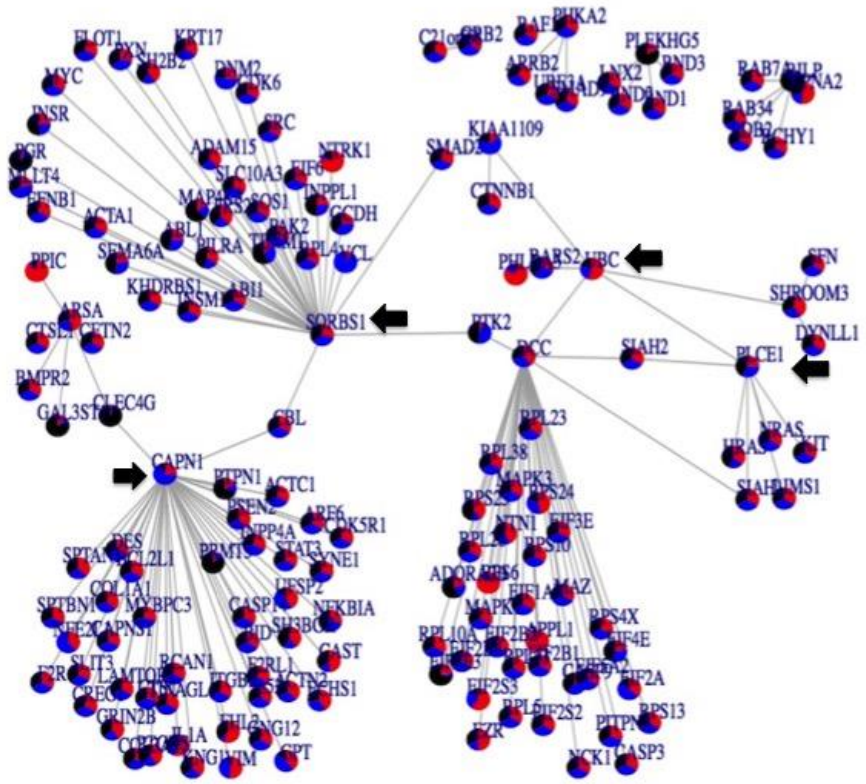



**Figure 3.14 Extracted sub-network of interactions for NPC 569.4.** The extracted sub-network of interactions for the list of variants with the most damaging predicted scores. The identified hubs in the sibling NPC 569.4 include *UBC, DCC, SORBS, CAPIN1, and PLCE1*, indicated by arrows. *Figure key: each pie chart is colour co-ordinated according to the frequency of the each allele represented. The three different populations represented herein include, African (black), East Asian (blue) and European (red). The Mixed Ancestry population of a Western Cape are an admixture of these populations.

Several hubs were identified for both NPC569.3 (**Figure 3.13**) and NPC569.4 (**Figure 3.14**), some of which were shared, for instance UBC (Ubiquitin C), and NAT10 (N-Acetyltranferase 10). However, there were also hubs, which were not connected to the main hubs, for example the FANC hub (top left in **Figure 3.13)** and PNA2, (top right in **Figure 3.14)**. FANC (Fanconi Anemia Complementation group) is a cluster of Fanconi anemia-related proteins which play an important role in chromosomal instability and breakage, defective DNA repair mechanisms, and have been shown to cause a number of recessive disorders, including Fanconi anemia (OMIM: 227650).

In the next step the sub-networks were characterized according to relevant biological pathways (**Table 3.12**) and biological processes (**Table 3.13**). The

biological pathways associated with the top sub-network in NPC569.3, such as the Wnt-signalling pathway, p53 feed-back loops and TGF-β signalling pathway are enriched with relevant molecular functions, implicated in human cancers.

**Table3.12 Results of EnrichR Analysis for Biological Pathways in NPC569.3**

| Term | Overlap | p-value | Adjusted P-value | Z-score | Combined Score | Genes |
|---|---|---|---|---|---|---|
| **Wnt signaling pathway_** | 10/278 | 0.000 | 0.008 | -1.77 | 8.39 | *SMARCE1;SMARCD1;SMARCC1;SMARCB1;SMARCC2;SIAH1;CTNNB1;ARID1A;SMARCA2;SMARCA4* |
| **p53 pathway feedback loops** | 4/45 | 0.001 | 0.020 | -1.53 | 5.98 | *NRAS;SIAH1;CTNNB1;HRAS* |
| **TGF-beta signalling pathway** | 4/88 | 0.014 | 0.131 | -1.28 | 2.60 | *SMAD2;NRAS;SMAD3;HRAS* |
| **p53 pathway** | 3/71 | 0.041 | 0.282 | -1.22 | 1.55 | *SUMO1;SIAH1;SIRT1* |

**Table3.13 Results of EnrichR Analysis for Biological Processes in proband, NPC 569.3**

| Term | Overlap | p-value | Adjusted p-Value | z-score | Combined Score | Genes |
|---|---|---|---|---|---|---|
| **Chromatin remodelling (GO:0006338)** | 14/118 | 0.00 | 0.00 | -2.18 | 58.32 | *SMARCE1;SMARCD1;SMARCC1;PBRM1;SMARCC2;SMARCB1;BAZ1B;ARID1A;ARID1B;SIRT1;SMARCA2;SMARCA4;KLF1;TOP1* |
| **Protein-DNA complex disassembly (GO:0032986)** | 7/17 | 0.00 | 0.00 | -2.96 | 56.12 | *SMARCE1;SMARCD1;SMARCC1;SMARCC2;SMARCB1;ARID1A;SMARCA4* |
| **Nucleosome disassembly (GO:0006337)** | 7/17 | 0.00 | 0.00 | -2.95 | 55.97 | *SMARCE1;SMARCD1;SMARCC1;SMARCC2;SMARCB1;ARID1A;SMARCA4* |
| **Chromatin modification (GO:0016568)** | 19/475 | 0.00 | 0.00 | -2.40 | 50.02 | *SMARCE1;SMARCD1;SMARCC1;PBRM1;SMARCC2;USP49;SMARCB1;CHD7;SIRT7;BAZ1B;ARID1A;ARID1B;SMARCA2;SIRT1;SIRT2;KLF1;SMARCA4;CARM1;TOP1* |
| **DNA repair (GO:0006281)** | 14/403 | 0.0 | 0.00 | -2.45 | 31.19 | *SMARCB1;FANCM;FANCL;FANCA;FANCB;BAZ1B;SIRT1;FANCG;APITD1;SUMO1;KIAA0101;UBC;OTUB1;FAAP100* |
| **Ubiquitin-dependent protein catabolic process (GO:0006511)** | 13/355 | 0.00 | 0.00 | -2.44 | 29.83 | *USP49;CUL3;SIAH2;CUL2;SIAH1;CUL1;SIRT1;SIRT2;CDC20;SOX2;ITCH;NEDD4;UBC* |
| **Modification-dependent macromolecule catabolic process (GO:0043632)** | 13/368 | 0.00 | 0.00 | -2.44 | 29.58 | *USP49;CUL3;SIAH2;CUL2;SIAH1;CUL1;SIRT1;SIRT2;CDC20;SOX2;ITCH;NEDD4;UBC* |

The biological pathways associated with the sub-network for the proband's sibling, i.e. NPC569.4, in **Table 3.14** were pertinent to "angiogenesis", VEGF-signalling and integrin-signalling. Angiogenesis plays an important role in cellular growth, neoplastic progression, invasion and metastasis, all important hallmarks of cancers. The VEGF- (vascular endothelial growth factor) signalling pathway is a regulator of tumour angiogenesis. It stimulates endothelial growth, survival and proliferation; it also regulates the growth of specific vascular tissue in blood and lymph vessels. The integrin-signalling pathway functions as a cell surface receptor that interacts with the extracellular matrix; this pathway also

mediates intercellular signals in response to the extracellular matrix including cellular shape and mobility. **Table 3.15** indicates some of the biological processes identified during the analysis. These included the processes of translational initiation and neuron projection guidance, amongst others.

**Table 3.14 Results of EnrichR Analysis for Biological Pathways in sibling, NPC 569.4**

| Term | Overlap | p-value | Adjusted P-value | Z-score | Combined Score | Genes |
|---|---|---|---|---|---|---|
| Angiogenesis | 17/142 | 0.00 | 0.00 | -1.66 | 14.99 | *SRC;F2R;PXN;STAT3;FRS2;PTK2;CASP9;EFNB1;NRAS;CTNNB1;MAPK1;RAF1;SOS1;PAK2;HRAS;MAPK3;NCK1* |
| Integrin signalling pathway | 17/156 | 0.00 | 0.00 | -1.68 | 14.366 | *ACTN2;SRC;PXN;RND2;RND3;PTK2;RND1;COL1A1;NRAS;COL3A1;RAF1;SOS1;HRAS;VCL;LIMS1;ARF6;MAPK3* |
| **CCKR signalling map ST_*Homo sapiens*_P06959** | 16/165 | 0.00 | 0.001 | -1.61 | 11.14 | *SRC;RPS6;PXN;STAT3;ARRB2;PTK2;NFKBIA;MYC;CASP3;CTNNB1;MAPK1;RAF1;SOS1;EIF4E;BCL2L1;MAPK3* |
| VEGF signaling pathway_ | 8/54 | 0.00 | 0.005 | -1.45 | 7.47 | *CASP9;NRAS;PXN;MAPK1;RAF1;HRAS;PTK2;MAPK3* |
| **T cell activation_ *Homo sapiens*_P00053** | 9/73 | 0.000 | 0.006 | -1.20 | 5.98 | *NFKBIA;NRAS;MAPK1;SOS1;RAF1;PAK2;HRAS;MAPK3;NCK1* |
| **Ras Pathway_ *Homo sapiens*_P04393** | 8/69 | 0.001 | 0.016 | -1.15 | 4.72 | *NRAS;STAT3;MAPK1;SOS1;RAF1;PAK2;HRAS;MAPK3* |
| **B cell activation_ *Homo sapiens*_P00010** | 7/57 | 0.002 | 0.021 | -1.03 | 4.01 | *NFKBIA;NRAS;MAPK1;SOS1;RAF1;HRAS;MAPK3* |
| **EGF receptor signaling pathway_ *Homo sapiens*_P00018** | 9/109 | 0.008 | 0.03 | -0.78 | 2.53 | *NRAS;STAT3;MAPK1;SFN;RAF1;SOS1;CBL;HRAS;MAPK3* |
| **Interleukin signaling pathway_ *Homo sapiens*_P00036** | 8/86 | 0.006 | 0.035 | -0.74 | 2.47 | *IL1A;NRAS;MYC;STAT3;MAPK1;SOS1;RAF1;MAPK3* |
| PDGF signalling pathway | 9/112 | 0.009 | 0.042 | -0.59 | 1.88 | *NRAS;MYC;STAT3;MAPK1;RAF1;SOS1;HRAS;MAPK3;NCK1* |

**Table 3.15 Results of EnrichR Analysis Biological Processes NPC569.4**

| Term | Overlap | p-value | Adjusted P-value | Z-score | Combined Score | Genes |
|---|---|---|---|---|---|---|
| **Translational initiation (GO:0006413)** | 24/139 | 0.00 | 0.00 | -2.18 | 85.78 | *RPL4;RPL5;EIF2B4;EIF2B2;EIF4E3;EIF1AX;RPL23;RPS6;RPL10A;EIF2S2;RPS4X;EIF2S3;EIF6;RPL13;EIF3E;RPL38;EIF2B1;EIF4E;RPS10;RPL28;RPS24;RPS13;EIF2A;RPS23* |
| **Neuron projection guidance (GO:0097485)** | 28/367 | 0.00 | 0.00 | -2.41 | 68.31 | *SRC;ITGB3;PITPNA;NTN1;RND1;EFNB1;NRAS;ABL1;MAPK1;SLIT3;HRAS;PAK2;SPTAN1;SPTBN1;NCK1;MAPK3;NTRK1;SEMA6A;DCC;SIAH2;SIAH1;PTK2;COL3A1;GRB2;EZR;RAF1;SOS1;CDK5R1* |
| **Axon guidance (GO:0007411)** | 28/367 | 0.00 | 0.000 | -2.40 | 68.244 | *SRC;ITGB3;PITPNA;NTN1;RND1;EFNB1;NRAS;ABL1;MAPK1;SLIT3;HRAS;PAK2;SPTAN1;SPTBN1;NCK1;MAPK3;NTRK1;SEMA6A;DCC;SIAH2;SIAH1;PTK2;COL3A1;GRB2;RAF1;SOS1;EZR;CDK5R1* |
| **Cellular response to organo-nitrogen compound (GO:0071417)** | 26/411 | 0.00 | 0.000 | -2.55 | 56.81 | *PXN;NRAS;ABL1;MAPK1;EIF4E;HRAS;SH2B2;APPL1;MAPK3;NTRK1;PTPN1;INSR;RPS6;STAT3;FRS2;SORBS1;GNG12;PTK2;COL1A1;COL3A1;CTNNB1;GRB2;LAMTOR1;RAF1;SOS1;BCL2L1* |
| **Cellular component disassembly (GO:0022411)** | 25/350 | 0.00 | 0.00 | -2.34 | 55.13 | *RPL4;RPL5;RPL10A;RPS4X;CAPNS1;CASP3;FLOT1;RPL13;RPL38;CAPN1;SPTAN1;PAK2;RPS10;RPS13;RPL23;RPS6;PTK2;COL1A1;COL3A1;ADAM15;CTNNB1;VIM;RPL28;RPS24;RPS23* |
| **Cellular response to nitrogen compound (GO:1901699)** | 26/438 | 0.00 | 0.00 | -2.53 | 54.43 | *PXN;NRAS;ABL1;MAPK1;HRAS;EIF4E;SH2B2;APPL1;MAPK3;NTRK1;PTPN1;INSR;RPS6;STAT3;FRS2;SORBS1;GNG12;PTK2;COL1A1;COL3A1;CTNNB1;GRB2;LAMTOR1;RAF1;SOS1;BCL2L1* |
| **Response to peptide (GO:1901652)** | 25/384 | 0.00 | 0.00 | -2.43 | 53.4 | *PXN;INPPL1;NRAS;MAPK1;EIF2B1;HRAS;EIF4E;SH2B2;APPL1;MAPK3;PTPN1;EIF2B4;EIF2B2;INSR;RPS6;STAT3;FRS2;SORBS1;GNG12;PTK2;COL1A1;NFKBIA;GRB2;RAF1;SOS1* |

To determine whether different tissues could be acquiring and embedding mutations at different rates, WES of DNA from adrenal gland, colon, cerebrum, kidney and saliva of the deceased proband NPC569.3 was performed. The aim was to determine whether there were differences in the number and type of variants in the different tissues of the proband, and specifically whether brain

tissue may be more prone to certain types of mutation not seen in other tissues – taking into account that the proband demised as a result of a glioblastoma/astrocytoma. As noted in **Figure 3.15** there were no regions within the genome, which harboured more variants than the other, with the exception of both chromosome 1 and 19, which had a similar pattern in the family data.

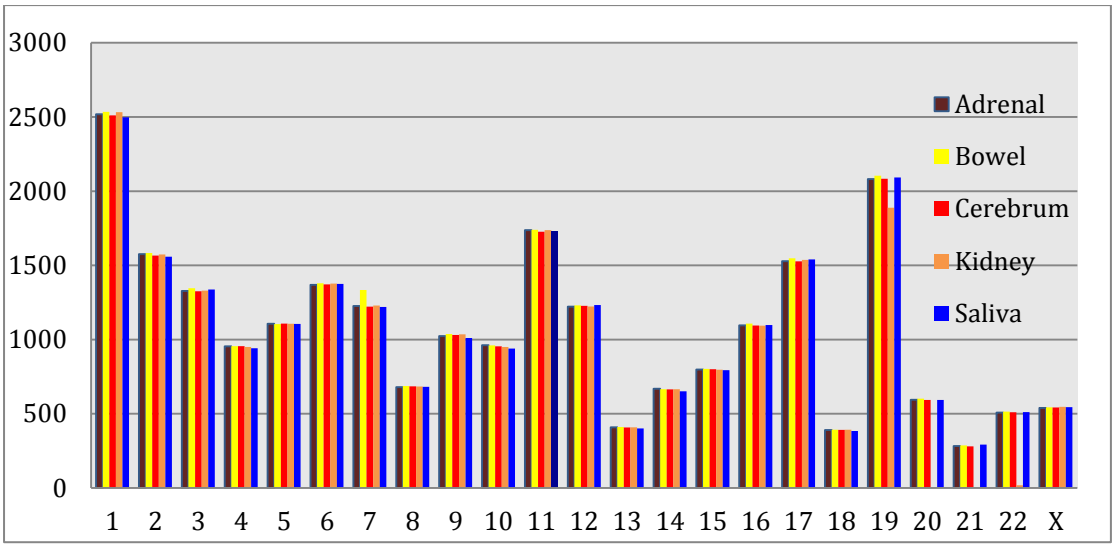



**Figure 3.15 Number of variants observed per chromosome inWES analysis of five tissues for NPC569.3.** This variant distribution was generated from the WES data using DNA extracted from different tissues of the proband. Only variants in the exons and splicing sites were included in this figure.

The analysis also involved investigating the number of de-novo variants in each tissue. About 500 variants were unique to each tissue, although most of the observed variants were missense or thought to be benign. A pathway-based analysis for each tissue was then performed from this list of variants. The results indicated that the most enriched pathways in the cerebrum, which was the closest tissue to the tumour, which was analysed in this section, included the extracellular matrix among others, which was a similar finding to the previous analysis of variants in germline DNA.

To validate some of these findings of the study we conducted a targeted gene panel sequencing of 13 genes that had previously been described to be instrumental in cancer development (**Figure 3.16**). The number of variants observed in the panel coincided with the number of variants seen in these genes

on WES, with one exception. *POLD1* had an average of 13 variants across the tissues yet, none were seen in the VCF output of the WES data. Upon checking for 'pathogenicity', none of the 13 variants were predicted to have an impact on the translated protein structure.

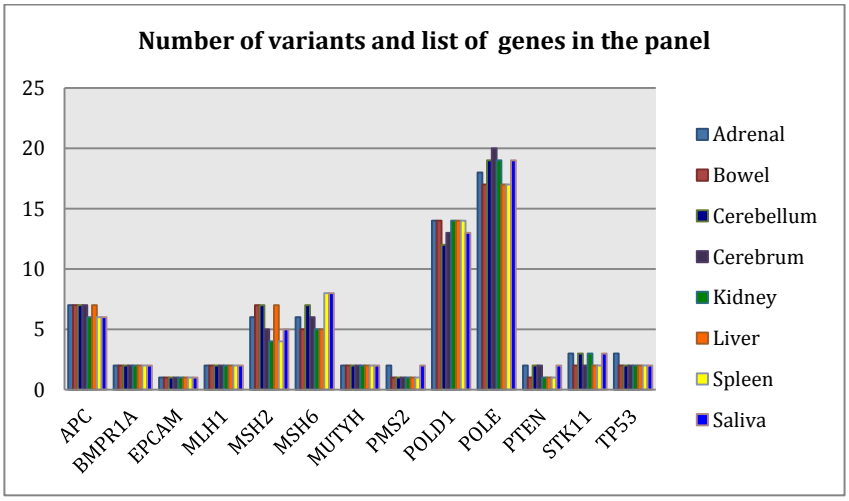



**Figure 3.16** The number of mutations identified in the 'cancer panel of genes' across eight tissues of NPC596.3. Indicated in the figure are the list of genes (X-axis) in the panel (sequenced), the number of variants observed for each sample (Y-axis) and the list of tissues (from which DNA was sequenced), which are colour coded on the right of the bar graph.

## 3.4 Discussion

The main aim of this section of the study was to identify the state of mutagenicity associated with mismatch repair deficiency. It was thought that at least some genes and pathways involved in cancers would be exposed more easily in an individual who was constitutionally mismatch repair deficient i.e. homozygous for a deleterious *MLH1* mutation. In this preliminary analysis - there were seemingly no specific regions within the genome which were identified to be more prone to accumulating mutations. However, the results presented here suggest that, even though most of the individual variants seem benign and not disease causing, some of these genes are involved in the most important known cancer pathways.

In the first instance for CMMR-D, the intention was to determine whether microsatellite instability could be a feature that can be used to diagnose this syndrome. There was no observed difference in the microsatellite status in

germline (saliva) or tumour DNA of the CMMR-D subject of the study (NPC569.3). This was unsurprising as other investigations have previously also reported microsatellite stable results for CMMR-D cases (Leenen et al. 2011; Baris et al. 2016). Nonetheless, improved methodology for detection of MSI has been suggested (Ingham et al., 2013; Wimmer et al., 2014). This method involves using a program Peak Heights (http://dna.leeds.ac.uk/peakheights/guide/) to determine peak ratios of germline MSI (gMSI) from three microsatellite markers (D2S123, D17S250 and D17S791). By determining the peak ratio of the markers, authors found that there was a significant difference in the gMSI of CMMR-D patients compared to the controls (Ingham et al., 2013).

For the mutational signature analysis, signatures 6 and 16, associated with MMR deficiency, were observed in our results. These signatures are commonly seen with the predominance of the C>T and T>C changes (Nik-Zainal et al., 2012a, 2012b; Alexandrov et al., 2013; Alexandrov and Stratton, 2014). Of interest is that the C>T change abundance is also present in signature 2. Signature 2 has been attributed to the activity of the activation-induced Cytidine Deaminase (AID)/ APOBEC family of cytidine deaminase. Cytidine deaminase functions by deaminating the cytidine coupled with base excision repair and mismatch repair machinery as a natural mutator, and play an important role in adaptive immunity (Kumar et al., 2014). The AID expression and activity is tightly regulated, and its deregulation is associated with diseases including cancers (Kumar et al. 2014; Rogozin et al., 2016).

Gene annotation and pathway analysis for the family study identified the potential dysregulation of several biologically important genes and pathways. Several cancer-related pathways including those pertinent to collagen biosynthesis, cellular adhesion and the extracellular matrix (organization and structural constituent) were over-represented in the samples. Pathways involving cell adhesion and the extracellular matrix were implicated repeatedly. Similar results were obtained in the analysis of the exome sequencing of the DNA derived from the cerebrum - the most proximal tissue to where the tumour was

located. Both, the extracellular matrix pathway and cell substrate adhesion were implicated as the most enriched pathways. Together, these results could suggest that cell adhesion and extracellular matrix interaction pathways are inadequately modulated or may be dysregulated early during tumorigenesis of MMR-deficiency-related cancers. Similar findings were reported by Emery et al. (2009), who investigated the expression profiles of breast cancers, and reported that for Atypical Ductal Hyperplasia (ADH), the genetic changes related to cellular adhesion and extracellular matrix pathways suggest that these regions are poorly modulated before cancer invasion and, in malignancy.

The extracellular matrix is important in modulating cellular functions as it directly interacts with the cellular-signalling receptors, and functions as a provider of growth factors and signalling molecules (Venning et al., 2015). Extracellular matrix signalling is important for cellular migration, invasion and angiogenesis; these are important functions, which makes this process (i.e. extracellular matrix signalling) instrumental in the metastasis/spreading of cancer. Since the extracellular matrix is so important in tumorigenesis, it has been suggested as a promising target in the development of adjuvant therapeutics (Emery et al., 2009; Lu et al., 2012; Venning et al., 2015; Xiong and Xu, 2016).

A list of hubs was identified from the sub-network analysis (from the protein-protein interactions): for example, *UBC* or ubiquitin C is part of the cell cycle checkpoint pathway and has been implicated in diseases such as apocrine adenoma and congenital granular cell tumour (Hsu et al., 2003; Zerener et al., 2013). Another identified hub, was *small ubiquitin-like modifier* (SUMO1), which is involved in translational modification and cellular processes such as nuclear transport, transcriptional regulation, apoptosis and protein stability. Both UBC and SUMO are important in maintaining genome stability (Jackson and Durocher, 2013).

N-Acetyltranferase 10 (NAT10), also identified as a hub, is a regulator of p53 activation, previously associated with CRC. After DNA damage, NAT10

translocates into nucleoplasm and activates p53-mediated cell cycle control and apoptosis. This protein inhibits cell proliferation through p53 activation by counteracting MDM2 action, providing an alternative pathway, which in turn activates the p53 cellular stress (Zhang et al., 2014; Liu et al., 2016). The AT-rich interactive domain 1A (ARID1A) gene encodes a member of the switch/sucrose non-fermentable (SWI-SNF) chromatin-remodelling complex, and is considered to work as a tumour suppressor in concert with p53. Loss of *ARID1A* is more common in advanced gastric cancers, sporadic CRC with microsatellite instability and MMR deficiency of both colon cancers and endometrial cancers (Chou et al., 2014). Variants in the gene encoding phospholipase C epsilon 1 or PLEC1 protein have been associated with CRC susceptibility and oesophageal cancer. In addition, specific mutations have been implicated in upper gastric cancers (Malik et al., 2014; Xue et al., 2015).

Some pathways of known cancer relevance were not prominent among the first gene sets which emerged in the analysis (e.g. pathways involved in maintenance of genomic stability, apoptosis, and damage repair), of unique variants in subjects NPC569.3 (proband) and NPC569.4 (sibling). However, during the sub-network analysis, when extracting the list of variants with the most damaging protein prediction scores, a number of pathways associated with cancers were identified. These included the Wnt-signalling pathway (Ponder, 2001; Zhan et al., 2016; Yoda et al., 2015), p53-pathway (Markowitz and Robertst, 1996; Vogelstein et al. 2000; ; Wakefield and Roberts, 2002; Harris and Levine 2005; Pardali and Moustakas, 2007; Lampropoulos et al., 2012), and the TGF-β pathway, suggesting that most of these pathways may be collectively involved, perhaps through incremental individual effects.

Even though NPC569.4 was not affected by any MMR-deficiency disorder, in the presence of the heterozygous *MLH1 c.1528C>T* variant, the germ-line variant analysis of her DNA showed cancer-related pathways to be most enriched. These included, the Integrin-signalling pathway (Guo and Giancotti, 2004), VEGF-signalling pathway (McMahon, 2000), and the PDGF-signalling pathway (Yu et al., 2003; Liu et al., 2011; Heldin et al., 2013). For instance, the angiogenesis

pathway (which pertains to one of the main hallmarks of cancers) is very important in growth, progression, invasion and metastasis of cancers). Also, VEGF stimulates endothelial growth, survival and proliferation. In addition, the PDGF pathway, which has been described as a model system for growth factors, also regulates the biological processes such as the activity of receptor kinases and pathways that drive cellular responses. Of note, also, are previously mentioned pathways such as the N-RAS and CTNNB1, which have previously been associated with CRC (Caluseriu et al., 2004; Haigis et al., 2008).

Following the exome sequencing of the five tissues of the CMMR-D proband, there were on average of 60,000 variants observed for each tissue. Most of the variants (99%) were shared (identical for each tissue) which is expected since it derived from the same individual, and only 1% (i.e. about 500) were unique to each tissue sample. In addition, when characterising these variants most were benign with no obvious impact on translated protein. Again, there were no particular regions of the genome which showed bias towards accumulation of these variants.

For validation purposes, the panel sequencing of the 13 genes, that been described to be instrumental in cancer development, was performed. *POLD1* had an average of 13 variants across the tissues yet, none were observed in the WES data. Upon analysis of the raw data, it was concluded that this was a result of the coverage for this particular region of the genome, the variants were excluded because the generated data did not pass the quality control process during filtering steps of WES analysis. *POLD1* is a catalytic and proof reading subunit of DNA polymerase delta (Weedon et al., 2013), which is responsible for the synthesis of the lagging strand during DNA replication (Palles et al., 2013;Weedon et al., 2013). Mutations in this gene have been implicated in a number of cancers including CRC, endometrial cancer (Palles et al., 2013) as well as brain tumours in CMMR-D patients (Shlien et al., 2015).

The appearance of novel changes in the *POLD1 and POLE* genes is notable since variants in these genes have previously been associated with cancers of the CNS.

The fact that none of the variants identified here and in the various tissues is pathogenic ought not to detract from the observation of increased mutagenicity especially within this gene i.e. *POLD1*. As mentioned previously, the mutations in this gene have been implicated in a number of cancers, and in the report by Shlien et al., (2015), two mutations (*POLD1 L606M* and *C319Y*) both of which reduced the replication fidelity of this enzyme, were identified as drivers of tumours in individuals with CMMRD-syndrome. It is possible that the glioblastoma was initiated by a pathogenic mutation, which would have arisen 'de novo' as with the other *POLD1* variants described in the various tissues analysed in this study. Unfortunately, at this stage of the work up, there was no tumour-material or DNA from tumour remaining from the proband for further investigation. Nonetheless, what the preliminary findings and their interpretation represent is the possible accumulation of a range of mutations in a range of genes/pathways which may be either directly pathogenic, or whose expression may be deleteriously affected – and which either individually or collectively expose CMMR-D children to early childhood cancers. This work will be ongoing in the Division of Human Genetics with the Departments of Paediatric Oncology and Anatomical Pathology at Red Cross Children's Hospital in order to gather additional cases/familes with Lynch syndrome and CMMR-D.

Furthermore, the above approach and findings provide a strong motivation to establish a tumour versus normal tissue bank for all cancers from patients admitted to the academic hospitals platform here in Cape Town. Of particular relevance to the present project would be the sequencing (WES, and then whole genome sequencing of tumour versus normal tissue) for the range of cancers that form the Lynch syndrome phenotype. The regular surveillance of communities carrying the *MLH1* c.1528C>T mutation, and their diagnoses with cancers of different organs/tissues makes the prospect of identifying intersecting pathways and processes for cancers of these different tissues - perhaps shedding further light on the process of carcinogenesis for Lynch syndrome generaly, but for the bigger burden of e.g. non Lynch endometrial, breast and other tissues.

# Chapter 4. General Discussion and Conclusions

## 4.1 Summary of main findings

The main aim of this research was to provide a clearer understanding of the genetic factors associated with Lynch syndrome on a macro scale, and CMMR-D syndrome, on a micro scale. To reach this aim, our first objective was to investigate the major clinical features associated with Lynch syndrome within our CRC registry in the Division of Human Genetics at UCT, comparing and contrasting a relatively large cohort of individuals with a common founder *MLH1* c.1528C>T mutation with the rest of the cohort with a variety of mutations in both *MLH1* and *MSH2*. The second objective was to investigate whether the *MLH1* c.1528C>T mutation was due to a founder effect – and if it was, to identify its possible 'age' or date of origin. The third and final objective was to perform an intensive investigation of the status of genomic changes in CMMR-D syndrome in a nuclear Lynch syndrome family where both parents were heterozygous for the *MLH1* c.1528C>T mutation, and who produced a child who had inherited the mutation from both her parents towards understanding the scale of increased mutagenicity on a genomic scale in individuals who are heterozygous and homozygous for the *MLH1* c.1528C>T – and how this, in turn, might predispose to increased tumorigenicity.

In summary, the Lynch syndrome cohort consisted of 396 mutation positive cases (both *MLH1* and *MSH2* mutation carriers). The most common mutation, *MLH1* c.1528C>T, was confirmed in 309 individuals, from 30 families. CRC was the most common cancer observed in both *MLH1* and *MSH2* carriers. In accordance with what has been previously reported, the age at diagnosis (of cancers) is later for *MSH2* mutation carriers compared to those with mutations in *MLH1*. Overall, the most common extra-colonic cancer among males was small bowel cancer, while endometrial and breast cancers were the most common in females. No significant difference in the survival period was evident when comparing males and females with Lynch syndrome. Although not significant, there was an observed trend towards improved survival for individuals with CRC versus extra-colonic cancers. This may be understandable since most of the

individuals who are tested for this study have been part of a long standing preventative programme involving genetic counselling, genetic testing and specifically colonoscopic surveillance of those testing positive for mutations in the MMR genes. The regular colonoscopic surveillance has been shown to identify premalignant lesions and their successful removal – thereby preventing progression and the associated morbidity and mortality (Stupart et al., 2009), which might accompany gynaecological cancers which had not historically been actively screened for.

For the investigation of the founder effect of the *MLH1 c.1528C>T* mutation in a group of Lynch syndrome probands, evidence was found suggesting that this mutation was indeed due to a founder effect. Five microsatellite markers surrounding the mutations were tested in 30 probands and 98 healthy controls. The most common extended haplotype covering all five markers tracked with the disease-causing mutation in 25 of the 30 probands and was not observed in any of healthy controls. Following haplotype analysis, the age estimation predicted the mutation to have arisen approximately nine generations ago; if each generation is considered 25 years, the mutation dates back to between 1789 and 1814 AD. The time line coincides with the origins and diversification of the "Coloured" or Mixed Ancestry population in South Africa in the 1700s. The origin of the mutation is still to be determined, this would require testing of the haplotype in different populations that have contributed to this admixed population such as the indigenous Khoisan, Bantu and immigrant European, and Asian populations. It is reasonably clear however, that because this mutation has not been reported in the countries of origin of the original European and Asian immigrants (to the knowledge of the investigator)– that the origin is likely to be here in Africa. There is much fieldwork still to be done – tracking families beyond the northern borders of South Africa- and perhaps into Namibia, especially amongst the native Nama and other indigenous communities.

The work on CMMR-D syndrome focused mainly in identifying genetic factors associated with this syndrome. The aim of this section was to investigate the level of dynamic change in the genomes of individuals carrying either a single

functional copy of the *MLH1* gene or no functional copy of the gene. The goal was to identify the underlying genetic features involved in the initiation and progression of neoplasia in a hypermutating environment. It had been hypothesised that specific regions within the genome may be more susceptible to mutagenesis in the absence of mismatch repair function. We investigated the alternative pathways involved in disease manifestation and to identify genetic changes/mutations which may drive the premature manifestation of disease in CMMR-D, and perhaps also hint at the mechanisms leading to tumorigenesis in Lynch syndrome (and perhaps even sporadic cancers).

In the first instance, a range of microsatellite markers were tested on various tissues of a proband with CMMR-D and established that MSI testing could not be used to diagnose CMMR-D. The purpose of the second part of the CMMR-D study was to identify genes which may be prone to mutations, and to link these genes to pathways associated with carcinogenesis. For this section, DNA from the proband (NPC569.3), the sibling (NPC569.4) and the parents (NPC569.1 and NPC1.77) were subjected to WES. The pathways within genes which had embedded the most mutations which were biologically significant because of MMR deficiency, included the 'collagen-related' and the 'extracellular matrix organisation' pathways. The extracellular matrix is one of the most important components in the tumour microenvironment and has previously been implicated in tumorigenesis. In the process of mapping protein-protein interactions a number of pathways such as, 'Wnt signalling', 'p53 pathway feed-back' and 'TGF-β' were observed to be the most enriched pathways in the CMMR-D proband (NPC569.3), while the genome of the sibling, NPC569.4 (heterozygous for the MLH1 mutation), reflected variants pertinent to pathways such as 'angiogenesis', 'integrin-' and 'VEGF- signalling'; all of which have also been previously associated with cancer development and progression. Therefore, even though there were no obvious genes and/or identified regions susceptible to accumulation of mutations, the data suggests that the variants, although seemingly random, occur in or influence specific pathways which are involved in the development and spread of cancers. The development of cancers has been described as a series of steps starting with the initiating mutation, which often

results in a cascade of events that leads to a cell's ability to disregard the normal control of proliferation, evasion of growth suppressors and resisting cell death. The data produced in this study shows that the presence of the identified variants result in an indirect influence on tumurigenesis thereby contributing, indirectly, to the series of events in disease development. The unfortunate lack of adequate amount of tumour tissue, for WES was a major drawback in this study.

WES was also performed on DNA samples from diverse tissues from the deceased proband (NPC569.3). These tissues included adrenal gland, bowel, cerebrum, kidney and saliva. The objective of this experiment was to further assess the state of mutagenicity in an individual with CMMR-D, by determining if there were differences in number of genetic variants which might be acquired between different tissues. There were a number of unique variants present in each of the tissues that were subjected to sequencing. Again, the pathway-based analysis for the variants indicated that the extracellular matrix was the most enriched pathway, in agreement with the previous analysis. For further validation, targeted sequencing of a panel of 13 most common cancer- related genes was done. The results were consistent with those observed in the WES, except for one gene, *POLD1*, which showed an average of 13 variants across the series of eight tissue samples (n=8). Interestingly, however, no variants were detected in this gene for WES. This was thought to be due to the 'on board' filtering which resulted in variants being excluded or discarded during quality check processes of the on-board computational sequencing software. Equally, it could be that certain critical regions of *POLD1* are not captured during the 'amplification/library construction phase' for WES. Nonetheless, *POLD1* is one of the genes that has previously been labelled as a driver in cases of glioblastoma in CMMR-D syndrome (Shlien et al., 2015) – and is very likely to contain the disease-causing event in the tumor which resulted in the demise of the subject of this study, i.e. Proband NPC569.3.

Overall, this study of the Lynch syndrome cohort (comprising of individuals with mutations in *MLH1* and *MSH2*) has provided an up to date status report of cancers in our cohort, and provides data for further analysis of the effectiveness

of the surveillance programme. The range of cancers now recognised to be part of Lynch syndrome benefits from data emerging from large cohorts of indviduals carrying the same pathogenic MMR gene mutation. The access to an individual with CMMR-D in amongst the large cohort with the *MLH1* c.1528C>T requires closer investigation of variants in tumors (versus normal tissue) in the range of cancers seen in this cohort, and their overlap with varinats seen in the subject with CMMR-D. A considerable amount of effort is being invested in collaboraring with the disciplines of Paediatric Oncology and Anatomical Pathology at Red Cross Children's Hospital to molecularly investigate a range of cancers seen in CMMR-D patients, starting with e.g. clinical awareness of e.g. signs of café au lait macules, a family history of cancers, and then progressing to immunohistochemistry of MMR proteins and then WES (and when affordable WGS).

### 4.2 Limitations of the study

Although the *MLH1* c.1528C>T cohort is significantly large to investigate disease phenomenology- the entire cohort contained relatively small numbers of disparate mutations in *MLH1* and *MSH2* to get an idea of any correlation. However, ongoing work – is very likely to yield extensive lineages of founder effects amongst CRC patients in S.A. to be able to do this kind of work in future. This study is valuable as it analysed the c.1528C>T cohort, and reflects the spread of Lynch-spectrum cancers, age at onset/diagnosis, and the success or otherwise of Lynch syndrome surveillance in an African setting.

The section on CMMR-D syndrome only focused on a single family, with one proband. This study would be most powerful if there were more families and affected individuals included. However, it should be noted that CMMR-D syndrome is one of the rare cancer syndromes with less than 200 cases reported worldwide, therefore recruitment of families is challenging. There is little doubt, however, that many more paediatric cancers may be identified as CMMR-D in the future.

For the intended list of tissues which were sequenced, unfortunately, no fresh tumour tissue was sampled – because of the relatively small size of the tumour. Therefore, only preserved material (FFPE) with all of its attendant shortcomings (for large scale DNA analysis) was available. Following a series of protocols used to extract DNA from the sample, none yielded enough DNA to allow for WES or even sequencing of the panel of genes described in Chapter 3. As a last resort, whole genome amplification of tumour DNA was attempted, in vain.

Pathway-based analysis was meant to provide a bird's eye view of the systems most impacted by the genes/ variants identified in the proband and sibling, however, this information depends greatly on what data is available at the time of analysis. Thus, even though the pathway analysis system uses the most updated information on specific diseases, it still may change as and when more information has been discovered on the genes their interactions and/or improved understanding of biological mechanisms. One of the main limitations of the pathway based analysis databases is understandable incomplete and inaccurate annotations (Khatri et al., 2014)

### 4.3 Future directions

The on-going interest in hereditary cancers and notably Lynch syndrome in the Division of Human Genetics and the Division of Surgical Gastroenterology at UCT and its Affiliated Hospitals – bodes well for identifying larger lineages of families and communities affected with Lynch syndrome. This will add to numbers through which one may get an improved idea of the emerging varieties of cancer which constitute Lynch syndrome in the African setting. With the current effort – it is likely that individuals with other MMR gene mutations will also be identified and add to our knowledge of disease phenomenology – towards improved management. There is currently a major drive to recruit individuals of indigenous African origin – especially where there is anecdotal information on earlier onset cancers, which are more aggressive. A visiting co-worker has been successful in identifying characteristics of Lynch syndrome in at least 20% of cancers in individuals diagnosed with cancer, under the age of 50 years in Zimbabwe, which is a neighbouring southern African country (Katsidzira et al.,

2017a; b). Of particular interest will be an understanding of whether the genetics or environment may contribute to differential manifestation in ethnically diverse populations (taking into account the issues of penetrance and selection which influence and are influenced by human migration and their environmental exposures). South Africa is an excellent international migratory cul-de-sac to study this phenomenon.

There are further aspects of the project which should be considered for the future, in the first instance every effort should be made to collect and store fresh as well as preserved tumour tissue; WES should be applied to tumours and related normal tissue from individuals under 50 years of age in order to identify the driver mutations, as well as to compare the acquired mutations; this would be equally effective for any child suspected of or identified with CMMR-D syndrome. This process requires very careful attention to detail in terms of workflows pertinent to patient care, as well as their recruitment for research. Although most of this is in place – there needs to be an institutional commitment to ethically recruiting all patients for research and ensuring that the resources are in place to get this done. For this to materialise, other FFPE DNA extraction methods should be tried in order to yield good quality DNA from samples, methods such as the QIAamp FFPE DNA extraction method has been shown to yield better quality DNA for downstream exome analysis (Hedengaard et al., 2014).

Identification of other CMMR-D cases in order to add to the findings of the study would be ideal. This would also allow for identification of genetic factors that may later be developed as potential biomarkers in persons with CMMR-D syndrome. Relationships have been developed with the Paediatric Oncologists, Neuorsurgeons, and Haematologists to consider CMMR-D as a differential diagnosis in their patients.

For the current study we were unable to provide evidence for the use of MSI testing as a diagnostic tool for CMMR-D. Thus future research should consider more sensitive methods, which can be used to diagnose CMMR-D syndrome.

The data generated in this study should be compared with other data, such as the Cancer Genome Atlas (TCGA) (https://tcga-data.nci.nih.gov/docs/publications/tcga/) data on childhood brain cancers and other Lynch syndrome related cancers, to identify shared variants/genes that may be involved in disease manifestation or progression.

## References:


Aarnio, M, Mecklin, J-P, Aaltonen, LA, Nyström-Lahti, M, Järvinen, HJ. 1995. Life-time risk of different cancers in hereditary non-polyposis colorectal cancer (hnpcc) syndrome. Int. J. Cancer 64: 430–433.

Alexandrov, LB, Nik-Zainal, S, Wedge, DC, Aparicio, S a JR, Behjati, S, Biankin, A V, Bignell, GR, Bolli, N, Borg, A, Børresen-Dale, A-L, Boyault, S, Burkhardt, B, Butler, AP, Caldas, C, Davies, HR, Desmedt, C, Eils, R, Eyfjörd, JE, Foekens, J a, Greaves, M, Hosoda, F, Hutter, B, Ilicic, T, Imbeaud, S, Imielinsk, M, Jäger, N, Jones, DTW, Jones, D, Knappskog, S, Kool, M, Lakhani, SR, López-Otín, C, Martin, S, Munshi, NC, Nakamura, H, Northcott, P a, Pajic, M, Papaemmanuil, E, Paradiso, A, Pearson, J V, Puente, XS, Raine, K, Ramakrishna, M, Richardson, AL, Richter, J, Rosenstiel, P, Schlesner, M, Schumacher, TN, Span, PN, Teague, JW, Totoki, Y, Tutt, ANJ, Valdés-Mas, R, van Buuren, MM, van 't Veer, L, Vincent-Salomon, A, Waddell, N, Yates, LR, Zucman-Rossi, J, Futreal, PA, McDermott, U, Lichter, P, Meyerson, M, Grimmond, SM, Siebert, R, Campo, E, Shibata, T, Pfister, SM, Campbell, PJ, Stratton, MR. 2013. Signatures of mutational processes in human cancer. Nature 500: 415–21.

Alexandrov, LB, Stratton, MR. 2014. Mutational signatures: The patterns of somatic mutations hidden in cancer genomes. Curr. Opin. Genet. Dev. 24: 52–60.

Amira, A.T., Mouna, T., Ahlem, B., Raoudha, A., Majid, B.H., Amel, H., Rachida, Z., Nadia, K., 2014. Immunohistochemical expression pattern of MMR protein can specifically identify patients with colorectal cancer microsatellite instability. Tumor Biol. 35, 6283–6291.

Amayiri, N, Al-Hussaini, M, Swaidan, M, Jaradat, I, Qandeel, M, Tabori, U, Hawkins, C, Musharbash, A, Alsaad, K, Bouffet, E. 2015. Synchronous glioblastoma and medulloblastoma in a child with mismatch repair mutation. Child's Nerv. Syst. 32: 553–557.

Amayiri, N, Tabori, U, Campbell, B, Bakry, D, Aronson, M, Durno, C, Rakopoulos, P, Malkin, D, Qaddoumi, I, Musharbash, A, Swaidan, M, Bouffet, E, Hawkins, C, Al-Hussaini, M. 2016. High frequency of mismatch repair deficiency among pediatric high grade gliomas in Jordan. Int. J. Cancer 138: 380–385.

Antelo, M, Milito, D, Rhees, J, Roca, E, Barugel, M, Cuatrecasas, M, Moreira, L, Liz,

M. 2015. Pitfalls in the diagnosis of biallelic PMS2 mutations.

Araten, DJ, Golde, DW, Zhang, RH, Thaler, HT, Gargiulo, L, Notaro, R, Luzzatto, L. 2005. A quantitative measurement of the human somatic mutation rate. Cancer Res. 65: 8111–7.

Archer, TC, Weeraratne, SD, Pomeroy, SL, Boston, H. 2013. Hedgehog-GLI Pathway in Medulloblastoma. 30: 2154–2156.

Arends, MJ. 2013. Pathways of colorectal carcinogenesis. Appl. Immunohistochem. Mol. Morphol. 21: 97–102.

Baas, AF, Gabbett, M, Rimac, M, Kansikas, M, Raphael, M, Nievelstein, RA, Nicholls, W, Offerhaus, J, Bodmer, D, Wernstedt, A, Krabichler, B, Strasser, U, Nyström, M, Zschocke, J, Robertson, SP, van Haelst, MM, Wimmer, K. 2013. Agenesis of the corpus callosum and gray matter heterotopia in three patients with constitutional mismatch repair deficiency syndrome. Eur. J. Hum. Genet. 21: 55–61.

Baker, K, Chong, G, Foulkes, WD, Jass, JR. 2006. Transforming growth factor-beta pathway disruption and infiltration of colorectal cancers by intraepithelial lymphocytes. Histopathology 49: 371–380.

Bakry, D, Aronson, M, Durno, C, Rimawi, H, Farah, R, Alharbi, QK, Alharbi, M, Shamvil, A, Ben-Shachar, S, Mistry, M, Constantini, S, Dvir, R, Qaddoumi, I, Gallinger, S, Lerner-Ellis, J, Pollett, A, Stephens, D, Kelies, S, Chao, E, Malkin, D, Bouffet, E, Hawkins, C, Tabori, U. 2014. Genetic and clinical determinants of constitutional mismatch repair deficiency syndrome: Report from the constitutional mismatch repair deficiency consortium. Eur. J. Cancer: 1–10.

Bansidhar, BJ, Silinsky, J. 2012. History and pathogenesis of lynch syndrome. Clin. Colon Rectal Surg. 25: 63–66.

Bansidhar, BJ. 2012. Extracolonic manifestations of lynch syndrome. Clin. Colon Rectal Surg. 25: 103–110.

Baris, HN, Barnes-Kedar, I, Toledano, H, Halpern, M, Hershkovitz, D, Lossos, A, Lerer, I, Peretz, T, Kariv, R, Cohen, S, Half, EE, Magal, N, Drasinover, V, Wimmer, K, Goldberg, Y, Bercovich, D, Levi, Z. 2016. Constitutional Mismatch Repair Deficiency in Israel: High Proportion of Founder Mutations in MMR Genes and Consanguinity. Pediatr. Blood Cancer 63: 418–427.

Barrow, E, Hill, J, Evans, DG. 2013. Cancer risk in Lynch Syndrome. Fam. Cancer.

Bellido, F, Guinó, E, Jagmohan-Changur, S, Seguí, N, Pineda, M, Navarro, M, Lázaro, C, Blanco, I, Vasen, HF a, Moreno, V, Capellá, G, Wijnen, JT, Valle, L. 2013. Genetic variant in the telomerase gene modifies cancer risk in Lynch syndrome. Eur. J. Hum. Genet. 21: 511–6.

Boland, CR. 2013. The Mystery of Mismatch Repair Deficiency: Lynch or Lynch-like? Gastroenterology 144: 868–70.

Boland, CR, Lynch, HT. 2013. The History of Lynch Syndrome. Fam. Cancer.

Bonaı, B, Grandjouan, S, Huiart, L, Caron, O, Colas, C, Bonaı, C. 2013. and MSH6 Genes in Lynch Syndrome. JAMA 305: 2304–10.

Borelli, I, Casalis Cavalchini, GC, Del Peschio, S, Micheletti, M, Venesio, T, Sarotto, I, Allavena, A, Delsedime, L, Barberis, MA, Mandrile, G, Berchialla, P, Ogliara, P, Bracco, C, Pasini, B. 2014. A founder MLH1 mutation in Lynch syndrome families from Piedmont, Italy, is associated with an increased risk of pancreatic tumours and diverse immunohistochemical patterns. Fam. Cancer 13: 401–413.

Borràs, E, Pineda, M, Blanco, I, Jewett, EM, Wang, F, Teulé, À, Caldés, T, Urioste, M, Martínez-Bouzas, C, Brunet, J, Balmaña, J, Torres, A, Ramón Y Cajal, T, Sanz, J, Pérez-Cabornero, L, Castellví-Bel, S, Alonso, Á, Lanas, Á, González, S, Moreno, V, Gruber, SB, Rosenberg, NA, Mukherjee, B, Lázaro, C, Capellá, G. 2010. MLH1 founder mutations with moderate penetrance in Spanish Lynch syndrome families. Cancer Res. 70: 7379–7391.

Bronner, CE, Baker, SM, Morrison, PT, Warren, G, Smith, LG, Lescoe, MK, Kane, M, Earabino, C, Lipford, J, Lindblom, A, Tannergard, P, Bollag, RJ, Godwin, AR, Ward, DC, Nordenskj[osol]ld, M, Fishel, R, Kolodner, R, Liskay, RM. 1994. Mutation in the DNA mismatch repair gene homologue hMLH 1 is associated with hereditary non-polyposis colon cancer. Nature 368: 258–261.

Bruwer, Z, Algar, U, Vorster, A, Fieggen, K, Davidson, A, Goldberg, P, Wainwright, H, Ramesar, R. 2013. Predictive Genetic Testing in Children: Constitutional Mismatch Repair Deficiency Cancer Predisposing Syndrome. J. Genet. Couns.

Burn, J, Bishop, DT, Mecklin, J, Macrae, F, Möslein, G, Olschwang, S, Bisgaard, M, Ramesar, R, Eccles, D, Maher, ER, Bertario, L, Jarvinen, HJ, Lindblom, A, Evans, DG, Lubinski, J, Morrison, PJ, Ho, JWC, Vasen, HFA, Side, L, Thomas, HJW. 2008. Effect of aspirin or resistant starch on colorectal neoplasia in the

Lynch syndrome. N. Engl. J. Med. 359: 2567–2578.

Burn, J, Gerdes, A, Macrae, F, Mecklin, J, Moeslein, G, Olschwang, S, Eccles, D, Evans, DG, Morrison, PJ, Murday, V, Ramesar, R, Side, L, Scott, RJ, Thomas, HJW, Vasen, HF, Barker, G, Crawford, G. 2010. Long-term eff ect of aspirin on cancer risk in carriers of hereditary colorectal cancer : an analysis from the CAPP2 randomised controlled trial. Lancet 378: 2081–2087.

Caluseriu, O, Di Gregorio, C, Lucci-Cordisco, E, Santarosa, M, Trojan, J, Brieger, A, Benatti, P, Pedroni, M, Colibazzi, T, Bellacosa, A, Neri, G, Ponz de Leon, M, Viel, A, Genuardi, M. 2004. A founder <em>MLH1</em> mutation in families from the districts of Modena and Reggio-Emilia in northern Italy with hereditary non-polyposis colorectal cancer associated with protein elongation and instability. J. Med. Genet. 41: e34 LP-e34.

Calvert, PM, Frucht, H. 2002. The Genetics of Colorectal Cancer. Ann. Intern. Med. 137: 603–612.

Campos, B, Balmaña, J, Gardenyes, J, Valenzuela, I, Abad, O, Fàbregas, P, Volpini, V, Díez, O. 2013. Germline mutations in NF1 and BRCA1 in a family with neurofibromatosis type 1 and early-onset breast cancer. Breast Cancer Res. Treat. 139: 597–602.

Cederquist, K, Emanuelsson, M, Wiklund, F, Golovleva, I, Palmqvist, R, Grönberg, H. 2005. Two Swedish founder MSH6 mutations, one nonsense and one missense, conferring high cumulative risk of Lynch syndrome. Clin. Genet. 68: 533–541.

Chan, TL, Wai Chan, Y, Ho, JWC, Chan, C, Chan, ASY, Chan, E, Lam, PWY, Wah Tse, C, Cheong Lee, K, Wai Lau, C, Gwi, E, Yi Leung, S, Yuen, ST. 2004. MSH2 c.1452–1455delAATG Is a Founder Mutation and an Important Cause of Hereditary Nonpolyposis Colorectal Cancer in the Southern Chinese Population. Am. J. Hum. Genet. 74: 1035–1042.

Cheah, PY. 2009. Recent advances in colorectal cancer genetics and diagnostics. Crit. Rev. Oncol. Hematol. 69: 45–55.

Chen, EY, Tan, CM, Kou, Y, Duan, Q, Wang, Z, Meirelles, GV, Clark, NR, Ma'ayan, A. 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14: 128.

Chou, A, Toon, CW, Clarkson, A, Sioson, L, Houang, M, Watson, N, Desilva, K, Gill,

AJ. 2014. Loss of ARID1A expression in colorectal carcinoma is strongly associated with mismatch repair deficiency. Hum. Pathol. 45: 1697–1703.

Cohen, PE, Pollack, SE, Pollard, JW. 2006. Genetic Analysis of Chromosome Pairing, Recombination, and Cell Cycle Control during First Meiotic Prophase in Mammals. Endocr. Rev. 27: 398–426.

Corcos, D. 2012. Unbalanced replication as a major source of genetic instability in cancer cells. Am. J. Blood Res. 2: 160–9.

De Wit, E, Delport, W, Rugamika, CE, Meintjes, A, Möller, M, Van Helden, PD, Seoighe, C, Hoal, EG. 2010. Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. Hum. Genet. 128: 145–153.

Dietmaier, W, Wallinger, S, Bocker, T. 1997. Diagnostic Microsatellite Instability : Definition and Correlation with Mismatch Repair Protein Expression Diagnostic Microsateffite Instability : Definition and Correlation with Mismatch Repair Protein Expression ’. 4749–4756.

Dreyer, G, Pret, M, Pret, MOG, Sa, M. 2012. Best Practice & Research Clinical Obstetrics and Gynaecology Screening for gynaecologic cancers in genetically predisposed women. Best Pract. Res. Clin. Obstet. Gynaecol. 26: 267–282.

Elhasid, R, Dvir, R, Rosenfeld Keidar, H, Ben Shachar, S, Bitan, M, Solar, I, Durno, C, Aronson, M, Malkin, D, Hawkins, C, Bouffet, E, Tabori, U, BMMRD Consortium. 2015. Management of Acute Myeloblastic Leukemia in a Child With Biallelic Mismatch Repair Deficiency. J. Pediatr. Hematol. Oncol. 37: e490-3.

Ellison, AR, Lofing, J, Bitter, GA. 2004. Human Mutl homolog (MLH1) function in DNA mismatch repair: A prospective screen for missense mutations in the ATPase domain. Nucleic Acids Res. 32: 5321–5338.

Emery, L a, Tripathi, A, King, C, Kavanah, M, Mendez, J, Stone, MD, de las Morenas, A, Sebastiani, P, Rosenberg, CL. 2009. Early dysregulation of cell adhesion and extracellular matrix pathways in breast cancer progression. Am. J. Pathol. 175: 1292–302.

Engel, C, Loeffler, M, Steinke, V, Rahner, N, Holinski-Feder, E, Dietmaier, W, Schackert, HK, Goergens, H, Von Knebel Doeberitz, M, Goecke, TO,

Schmiegel, W, Buettner, R, Moeslein, G, Letteboer, TGW, Go???ez Garc??a, E, Hes, FJ, Hoogerbrugge, N, Menko, FH, Van Os, TAM, Sijmons, RH, Wagner, A, Kluijt, I, Propping, P, Vasen, HFA. 2012. Risks of less common cancers in proven mutation carriers with lynch syndrome. J. Clin. Oncol. 30: 4409–4415.

Epple, LM, Griffiths, SG, Dechkovskaia, AM, Dusto, NL, White, J, Ouellette, RJ, Anchordoquy, TJ, Bemis, LT, Graner, MW. 2012. Medulloblastoma exosome proteomics yield functional roles for extracellular vesicles. PLoS One 7: e42064.

Evans, LS, Hancock, BW. 2003. Non-Hodgkin lymphoma. Lancet 362: 139–46.

Fearon, ER. 1997. Human Cancer Syndromes: Clues to the Origin and Nature of Cancer. Science (80-. ). 278: 1043–1050.

Fearon, ER. 2011. Molecular genetics of colorectal cancer. Annu. Rev. Pathol. 6: 479–507.

Felix, R, Bodmer, W, Fearnhead, NS, van der Merwe, L, Goldberg, P, Ramesar, RS. 2006. GSTM1 and GSTT1 polymorphisms as modifiers of age at diagnosis of hereditary nonpolyposis colorectal cancer (HNPCC) in a homogeneous cohort of individuals carrying a single predisposing mutation. Mutat. Res. 602: 175–81.

Felsberg, J, Thon, N, Eigenbrod, S, Hentschel, B, Sabel, MC, Westphal, M, Schackert, G, Kreth, FW, Pietsch, T, Löffler, M, Weller, M, Reifenberger, G, Tonn, JC. 2011. Promoter methylation and expression of MGMT and the DNA mismatch repair genes MLH1, MSH2, MSH6 and PMS2 in paired primary and recurrent glioblastomas. Int. J. Cancer 129: 659–70.

Fenner, JN. 2005. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. Am. J. Phys. Anthropol. 128: 415–423.

Frank, S a. 2003. Somatic Mutation: Early Cancer Steps Depend on Tissue Architecture. Curr. Biol. 13: R261–R263.

Frank, S a, Nowak, M a. 2004. Problems of somatic mutation and cancer. Bioessays 26: 291–9.

Froggatt, N, Green, J, Brassett, C, Evans, D, Bishop, D, Kolodner, R, Maher, E. 1999. A common MSH2 mutation in English and North American HNPCC families:

origin, phenotypic expression, and sex specific differences in colorectal cancer. J. Med. Genet. 36: 97–102.

Giunti, L, Cetica, V, Ricci, U, Giglio, S, Sardi, I, Paglierani, M, Andreucci, E, Sanzo, M, Forni, M, Buccoliero, AM, Genitori, L, Genuardi, M. 2009. Type A microsatellite instability in pediatric gliomas as an indicator of Turcot syndrome. Eur. J. Hum. Genet. 17: 919–27.

Goecke, T, Schulmann, K, Engel, C, Holinski-Feder, E, Pagenstecher, C, Schackert, HK, Kloor, M, Kunstmann, E, Vogelsang, H, Keller, G, Dietmaier, W, Mangold, E, Friedrichs, N, Propping, P, Krüger, S, Gebert, J, Schmiegel, W, Rueschoff, J, Loeffler, M, Moeslein, G. 2006. Genotype-phenotype comparison of German MLH1 and MSH2 mutation carriers clinically affected with lynch syndrome: A report by the German HNPCC consortium. J. Clin. Oncol. 24: 4285–4292.

Goldberg, PA, Madden, M V., Harocopos, C, Felix, R, Westbrook, C, Ramesar, RS. 1998. In a resource-poor country, mutation identification has the potential to reduce the cost of family management for hereditary nonpolyposis colorectal cancer. Dis. Colon Rectum 41: 1250–1255.

Guo, W, Giancotti, FG. 2004. Integrin signalling during tumour progression. Nat Rev Mol Cell Biol 5: 816–826.

Haigis, KM, Kendall, KR, Wang, Y, Cheung, A, Haigis, MC, Glickman, JN, Niwa-Kawakita, M, Sweet-Cordero, A, Sebolt-Leopold, J, Shannon, KM, Settleman, J, Giovannini, M, Jacks, T. 2008. Differential effects of oncogenic K-Ras and N-Ras on proliferation, differentiation and tumor progression in the colon. Nat. Genet. 40: 600–608.

Hall, A, Lynagh, M, Bryant, J, Sanson-Fisher, R. 2013. Supportive care needs of hematological cancer survivors: A critical review of the literature. Crit. Rev. Oncol. Hematol.

Hamilton, SR, Liu, B. 1995. The molecular basis of turcot's syndrome s. N. Engl. J. Med. 332: 839–847.

Hamra, G, MacLehose, R, Richardson, D. 2013. Markov chain monte carlo: An introduction for epidemiologists. Int. J. Epidemiol. 42: 627–634.

Han, HJ, Maruyama, M, Baba, S, Park, JG, Nakamura, Y. 1995. Genomic structure of human mismatch repair gene, hMLH1, and its mutation analysis in patients with hereditary non-polyposis colorectal cancer (HNPCC). Hum. Mol. Genet.

4: 237–242.

Hanahan, D, Weinberg, R a. 2011. Hallmarks of cancer: the next generation. Cell 144: 646–74.

Harris, SL, Levine, AJ. 2005. The p53 pathway: positive and negative feedback loops. Oncogene 24: 2899–2908.

Heldin, C-H, Heldin, C-H, Westermark, B, Andrae, J, Gallini, R, Betsholtz, C, Heldin, C-H, Fredriksson, L, Li, H, Fieber, C, Li, X, Eriksson, U, Fredriksson, L, Li, H, Eriksson, U, Ustach, C, Kim, H, Ustach, C, Huang, W, Conley-LaComb, M, Lin, C, Che, M, Abrams, J, Kim, H, Omura, T, Heldin, C-H, Östman, A, Yang, Y, Yuzawa, S, Schlessinger, J, Baxter, R, Secrist, J, Vaillancourt, R, Kazlauskas, A, Irusta, P, Luo, Y, Bakht, O, Lai, C, Smith, S, DiMaio, D, Chiara, F, Bishayee, S, Heldin, C-H, Demoulin, J-B, Heldin, C-H, Östman, A, Rönnstrand, L, Bae, Y, Sung, J-Y, Kim, O-S, Kim, Y, Hur, K, Kazlauskas, A, Rhee, S, Sundaresan, M, Yu, Z, Ferrans, V, Irani, K, Finkel, T, Jurek, A, Amagasaki, K, Gembarska, A, Heldin, C-H, Lennartsson, J, Fantl, W, Escobedo, J, Martin, G, Turck, C, Rosario, M del, McCormick, F, Williams, L, Jurek, A, Heldin, C-H, Lennartsson, J, Haj, F, Markova, B, Klaman, L, Bohmer, F, Neel, B, Persson, C, Sävenhed, C, Bourdeau, A, Tremblay, M, Markova, B, Böhmer, F, Haj, F, Neel, B, Elson, A, Heldin, C-H, Kovalenko, M, Denner, K, Sandström, J, Persson, C, Groß, S, Jandt, E, Vilella, R, Böhmer, F, Östman, A, et al. 2013. Targeting the PDGF signaling pathway in tumor treatment. Cell Commun. Signal. 11: 1–18.

Herkert, JC, Niessen, RC, Olderode-Berends, MJW, Veenstra-Knol, HE, Vos, YJ, van der Klift, HM, Scheenstra, R, Tops, CMJ, Karrenbeld, A, Peters, FTM, Hofstra, RMW, Kleibeuker, JH, Sijmons, RH. 2011. Paediatric intestinal cancer and polyposis due to bi-allelic PMS2 mutations: case series, review and follow-up guidelines. Eur. J. Cancer 47: 965–82.

Hitchins, M, Owens, S, Kwok, CT, Godsmark, G, Algar, U, Ramesar, R. 2011. Identification of new cases of early-onset colorectal cancer with an MLH1 epimutation in an ethnically diverse South African cohort. Clin. Genet. 80: 428–434.

Hitchins, M, Williams, R, Cheong, K, Halani, N, Lin, VAP, Packham, D, Ku, S, Buckle, A, Hawkins, N, Burn, J, Gallinger, S, Goldblatt, J, Kirk, J, Tomlinson, I, Scott, R, Spigelman, A, Suter, C, Martin, D, Suthers, G, Ward, R. 2005. MLH1 germline

epimutations as a factor in hereditary nonpolyposis colorectal cancer. Gastroenterology 129: 1392–1399.

Hitchins, MP. 2013. The role of epigenetics in Lynch syndrome. Fam. Cancer.

Hsu, P-J, Liu, C-H, Huang, C-J. 2003. Mixed tubulopapillary hidradenoma and syringocystadenoma papilliferum occurring as a verrucous tumor. J. Cutan. Pathol. 30: 206–210.

Huang, H, Mahler-Araujo, BM, Sankila, A, Chimelli, L, Yonekawa, Y, Kleihues, P, Ohgaki, H. 2000. APC mutations in sporadic medulloblastomas. Am. J. Pathol. 156: 433–7.

Huse, JT, Holland, EC. 2010. Targeting brain cancer: advances in the molecular pathology of malignant glioma and medulloblastoma. Nat. Rev. Cancer 10: 319–31.

Ilencikova, D, Sejnova, D, Jindrova, J, Babal, P. 2011. High-Grade Brain Tumors in Siblings With Biallelic MSH6 Mutations. 1067–1070.

Imai, K, Yamamoto, H. 2008. Carcinogenesis and microsatellite instability: the interrelationship between genetics and epigenetics. Carcinogenesis 29: 673–80.

Inaba, H, Greaves, M, Mullighan, CG. 2013. Acute lymphoblastic leukaemia. Lancet 381: 1943–1955.

Ingham, D, Diggle, CP, Berry, I, Bristow, C a, Hayward, BE, Rahman, N, Markham, AF, Sheridan, EG, Bonthron, DT, Carr, IM. 2013. Simple detection of germline microsatellite instability for diagnosis of constitutional mismatch repair cancer syndrome. Hum. Mutat. 34: 847–52.

Iyer, RR, Pluciennik, A, Burdett, V, Modrich, PL. 2006. DNA mismatch repair: functions and mechanisms. Chem. Rev. 106: 302–23.

Jackson, SP, Durocher, D. 2013. Regulation of DNA Damage Responses by Ubiquitin and SUMO. Mol. Cell 49: 795–807.

Järvinen, HJ, Aarnio, M, Mustonen, H, Aktan-Collan, K, Aaltonen, L a, Peltomäki, P, De La Chapelle, A, Mecklin, JP. 2000. Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer. Gastroenterology 118: 829–834.

Jasperson, KW, Samowitz, WS, Burt, RW. 2011. Constitutional mismatch repair-deficiency syndrome presenting as colonic adenomatous polyposis: clues

from the skin. Clin. Genet. 80: 394–7.

Kamat, AM. 2013. Commentary on "Risks of primary extracolonic cancers following colorectal cancer in lynch syndrome. Urol. Oncol. Semin. Orig. Investig. 31: 716.

Kate, GL ten, Kleibeuker, JH, Nagengast, FM, Craanen, M, Cats, A, Menko, FH, Vasen, HFA. 2007. Is surveillance of the small bowel indicated for Lynch syndrome families? Gut 56: 1198–1201.

Katsidzira, L, Gangaidzo, I, Thomson, S, Rusakaniko, S, Matenga, J, Ramesar, R. 2017. The shifting epidemiology of colorectal cancer in sub-Saharan Africa. Lancet Gastroenterol. Hepatol. 2: 377–383.

Klasen, M, Spillmann, FJX, Marra, G, Cejka, P, Wabl, M. 2005. Somatic hypermutation and mismatch repair in non-B cells. Eur. J. Immunol. 35: 2222–9.

Kosinski, J, Hinrichsen, I, Bujnicki, JM, Friedhoff, P, Plotz, G. 2010. Identification of Lynch syndrome mutations in the MLH1-PMS2 interface that disturb dimerization and mismatch repair. Hum. Mutat. 31: 975–982.

Kratz, CP, Holter, S, Etzler, J, Lauten, M, Pollett, a, Niemeyer, CM, Gallinger, S, Wimmer, K. 2009. Rhabdomyosarcoma in patients with constitutional mismatch-repair-deficiency syndrome. J. Med. Genet. 46: 418–20.

Ku, C-S, Cooper, DN, Polychronakos, C, Naidoo, N, Wu, M, Soong, R. 2012. Exome sequencing: dual role as a discovery and diagnostic tool. Ann. Neurol. 71: 5–14.

Kuleshov, M V, Jones, MR, Rouillard, AD, Fernandez, NF, Duan, Q, Wang, Z, Koplev, S, Jenkins, SL, Jagodnik, KM, Lachmann, A, McDermott, MG, Monteiro, CD, Gundersen, GW, Ma'ayan, A. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 44: W90-7.

Kumar, R., J, D.L., Chaudhuri, J., Evans, T., 2014. Biological function of activation-induced cytidine deaminase {(AID).}. Biomed. J. 37, 269–283.

Kushner, BH, LaQuaglia, MP, Wollner, N, Meyers, PA, Lindsley, KL, Ghavimi, F, Merchant, TE, Boulad, F, Cheung, NK, Bonilla, MA, Crouch, G, Kelleher, JF, Steinherz, PG, Gerald, WL. 1996. Desmoplastic small round-cell tumor: Prolonged progression-free survival with aggressive multimodality therapy. J. Clin. Oncol. 14: 1526–1531.

Li, L, Hamel, N, Baker, K, Mcguf, MJ, Couillard, M, Gologan, A, Marcus, VA, Chodirker, B, Chudley, A, Stefanovici, C, Durandy, A, Hegele, RA, Feng, B,Goldgar, DE, Zhu, J, Rosa, M De, Gruber, SB, Wimmer, K, Young, B, Chong, G, Tischkowitz, MD, Foulkes, WD. 2015. A homozygous PMS2 founder mutation with an attenuated constitutional mismatch repair deficiency phenotype. J. Med. Genet. 52: 348–352.

Lampropoulos, P, Zizi-Sermpetzoglou, A, Rizos, S, Kostakis, A, Nikiteas, N, Papavassiliou, AG. 2012. TGF-beta signalling in colon carcinogenesis. Cancer Lett. 314: 1–7.

Lang, GI, Parsons, L, Gammie, AE. 2013. Mutation rates, spectra, and genome-wide distribution of spontaneous mutations in mismatch repair deficient yeast. Genes, Genomes Genetics (G3)(Bethesda). 3: 1453–65.

Lavoine, N, Colas, C, Muleris, M, Bodo, S, Duval, A, Coulet, F, Cabaret, O, Andreiuolo, F, Charpy, C, Sebille, G, Wang, Q, Lejeune, S, Guimbaud, R, Jedraszak, G, Bourdeaut, F, Grill, J, Caron, O, Tinat, J, Bougeard, G, Frébourg, T, Brugières, L. 2015. Constitutional mismatch repair de fi ciency syndrome : clinical description in a French cohort. 1–9.

Lee, Y, Scheck, AC, Cloughesy, TF, Lai, A, Dong, J, Farooqi, HK, Liau, LM, Horvath, S, Mischel, PS, Nelson, SF. 2008. Gene expression analysis of glioblastomas identifies the major molecular basis for the prognostic benefit of younger age. BMC Med. Genomics 1: 1–12.

Leenen, CHM, Geurts-Giele, WRR, Dubbink, HJ, Reddingius, R, van den Ouweland, a M, Tops, CMJ, van de Klift, HM, Kuipers, EJ, van Leerdam, ME, Dinjens, WNM, Wagner, a. 2011. Pitfalls in molecular analysis for mismatch repair deficiency in a family with biallelic pms2 germline mutations. Clin. Genet. 80: 558–65.

Leenen, CHM, Heijer, M Den, Meer, C Van Der. 2016. Genetic testing for Lynch syndrome : family communication and motivation. Fam. Cancer 15: 63–73.

Leung, SY, Chan, TL, Chung, LP, Chan, a S, Fan, YW, Hung, KN, Kwong, WK, Ho, JW, Yuen, ST. 1998. Microsatellite instability and mutation of DNA mismatch repair genes in gliomas. Am. J. Pathol. 153: 1181–8.

Levi, Z, Kariv, R, Barnes-Kedar, I, Goldberg, Y, Half, E, Morgentern, S, Eli, B, Baris, HN, Vilkin, A, Belfer, RG, Niv, Y, Elhasid, R, Dvir, R, Abu-Freha, N, Cohen, S.

2015. The gastrointestinal manifestation of constitutional mismatch repair deficiency syndrome: From a single adenoma to polyposis-like phenotype and early onset cancer. Clin. Genet. 88: 474–478.

Levine, EA, Allen, GO. 2014. Combined Genetic and Nutritional Risk Models of Triple Negative Breast Cancer Combined Genetic and Nutritional Risk Models of Triple. Nutr. Cancer 66: 955–963.

Lindor, NM, Petersen, GM, Hadley, DW, Kinney, AY, Miesfeldt, S, Lu, KH, Lynch, P, Burke, W, Press, N. 2006. Recommendations for the Care of Individuals With an Inherited Predisposition to Lynch Syndrome. A Systematic Review. Jama 296: 1507–1517.

Liu, K-W, Hu, B, Cheng, S-Y. 2011. Platelet-derived growth factor signaling in human malignancies. Chin. J. Cancer 30: 581–4.

Liu, X, Tan, Y, Zhang, C, Zhang, Y, Zhang, L, Ren, P. 2016. p 53 at K 120 and ubiquitinating Mdm 2. 17: 1–18.

Loeb, LA. 2001. Perspectives in Cancer Research A Mutator Phenotype in Cancer 1. 3230–3239.

Lu, KH, Daniels, M. 2013. Endometrial and ovarian cancer in women with Lynch syndrome: Update in screening and prevention. Fam. Cancer 12: 273–277.

Lu, P, Weaver, VM, Werb, Z. 2012. The extracellular matrix: A dynamic niche in cancer progression. J. Cell Biol. 196: 395–406.

Lynch, HT. 2003. 030306 Hereditary Colorectal Cancer. Color. Cancer 348: 919–932.

Lynch, HT, Cristofaro, G, Rozen, P, Vasen, H, Lynch, P, Mecklin, J, John, JS, Lynch, T, Lynch, J, Lynch, P, Cristofaro, G, John, JS. 2003. NonPolyposis Colorectal Cancer. 2: 3–5.

Lynch, HT, Drescher, K, Knezetic, J, Lanspa, S. 2014. Hereditary Cancer Syndrome Diagnosis , Heterogeneity and Treatment : A Review. 429–442.

Lynch, HT, Lynch, PM, Lanspa, SJ, Snyder, CL, Lynch, JF, Boland, CR. 2009. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. Clin. Genet. 76: 1–18.

Lynch, HT, Snyder, CL, Shaw, TG, Heinen, CD, Hitchins, MP. 2015. Milestones of Lynch syndrome: 1895–2015. Nat. Rev. Cancer 15: 181–194.

Lynch, M. 2010. Rate, molecular spectrum, and consequences of human

mutation. Proc. Natl. Acad. Sci. U. S. A. 107: 961–8.

Malik, M a, Srivastava, P, Zargar, S a, Mittal, B. 2014. Phospholipase C epsilon 1 (PLCE1) haplotypes are associated with increased risk of gastric cancer in Kashmir Valley. Saudi J. Gastroenterol. 20: 371–7.

Markowitz, SD, Roberts, AB. 1996. Tumor suppressor activity of the TGF-beta pathway in human cancers. Cytokine Growth Factor Rev. 7: 93–102.

Martín-López, J V, Barrios, Y, Medina-Arana, V, Andújar, M, Lee, S, Gu, L, Li, G-M, Rüschoff, J, Salido, E, Fishel, R. 2012. The hMSH2(M688R) Lynch syndrome mutation may function as a dominant negative. Carcinogenesis 33: 1647–54.

Martín-López, J V, Fishel, R. 2013. The mechanism of mismatch repair and the functional analysis of mismatch repair defects in Lynch syndrome. Fam. Cancer.

McKeigue, P.M., Carpenter, J.R., Parra, E.J., Shriver, M.D., 2000. Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. Ann. Hum. Genet. 64, 171–186. doi:DOI: undefined

McMahon, G. 2000. VEGF Receptor Signaling in Tumor Angiogenesis. 5: 3–10.

Menko, FH, Kaspers, GL, Meijer, G a, Claes, K, van Hagen, JM, Gille, JJP. 2004. A homozygous MSH6 mutation in a child with café-au-lait spots, oligodendroglioma and rectal cancer. Fam. Cancer 3: 123–7.

Miller, SA, Gallie, DR, Sleat, DE, Watts, JW, Turner, PC, Wilson, TM a. 1988. Volume 16 Number 3 1988 Nucleic Acids Research. Nucleic Acids Res. 16: 883–893.

Mohd, AB, Palama, B, Nelson, SE, Tomer, G, Nguyen, M, Huo, X, Buermeyer, AB. 2006. Truncation of the C-terminus of human MLH1 blocks intracellular stabilization of PMS2 and disrupts DNA mismatch repair. DNA Repair (Amst). 5: 347–361.

Moisio, AL, Sistonen, P, Weissenbach, J, de la Chapelle, A, Peltomäki, P. 1996. Age and origin of two common MLH1 mutations predisposing to hereditary colon cancer. Am. J. Hum. Genet. 59: 1243–51.

Mori, T, Nagase, H, Horii, A, Miyoshi, Y, Nakatsuru, S, Aoki, T, Arakawa, H, Nakamura, Y, Shimano, T, Yanagisawa, A, Ushio, Y, Takano, S, Ogawa, M, Nakamura, M, Shibuya, M, Nishikawa, R, Matsutani, M, Hayashi, Y, Takahashi,

H, Ikuta, F, Nishihira, T, Mori, S. 1994. Germ-line and somatic mutations of the APC gene in patients with turcot syndrome and analysis of APC mutations in brain tumors. Genes, Chromosom. Cancer 9: 168–172.

Negrini, S, Gorgoulis, VG, Halazonetis, TD. 2010. Genomic instability--an evolving hallmark of cancer. Nat. Rev. Mol. Cell Biol. 11: 220–8.

Nik-Zainal, S, Alexandrov, LB, Wedge, DC, Van Loo, P, Greenman, CD, Raine, K, Jones, D, Hinton, J, Marshall, J, Stebbings, L a, Menzies, A, Martin, S, Leung, K, Chen, L, Leroy, C, Ramakrishna, M, Rance, R, Lau, KW, Mudie, LJ, Varela, I, McBride, DJ, Bignell, GR, Cooke, SL, Shlien, A, Gamble, J, Whitmore, I, Maddison, M, Tarpey, PS, Davies, HR, Papaemmanuil, E, Stephens, PJ, McLaren, S, Butler, AP, Teague, JW, Jönsson, G, Garber, JE, Silver, D, Miron, P, Fatima, A, Boyault, S, Langerød, A, Tutt, A, Martens, JWM, Aparicio, S a JR, Borg, Å, Salomon, AV, Thomas, G, Børresen-Dale, A-L, Richardson, AL, Neuberger, MS, Futreal, PA, Campbell, PJ, Stratton, MR. 2012a. Mutational processes molding the genomes of 21 breast cancers. Cell 149: 979–93.

Nik-Zainal, S, Van Loo, P, Wedge, DC, Alexandrov, LB, Greenman, CD, Lau, KW, Raine, K, Jones, D, Marshall, J, Ramakrishna, M, Shlien, A, Cooke, SL, Hinton, J, Menzies, A, Stebbings, L a, Leroy, C, Jia, M, Rance, R, Mudie, LJ, Gamble, SJ, Stephens, PJ, McLaren, S, Tarpey, PS, Papaemmanuil, E, Davies, HR, Varela, I, McBride, DJ, Bignell, GR, Leung, K, Butler, AP, Teague, JW, Martin, S, Jönsson, G, Mariani, O, Boyault, S, Miron, P, Fatima, A, Langerød, A, Aparicio, S a JR, Tutt, A, Sieuwerts, AM, Borg, Å, Thomas, G, Salomon, AV, Richardson, AL, Børresen-Dale, A-L, Futreal, PA, Stratton, MR, Campbell, PJ. 2012b. The life history of 21 breast cancers. Cell 149: 994–1007.

Novelli, M. 2014. The pathology of hereditary polyposis syndromes. Histopathology.

Nyström-Lahti, M, Sistonen, P, Mecklin, JP, Pylkkänen, L, Aaltonen, LA, Järvinen, H, Weissenbach, J, de la Chapelle, A, Peltomäki, P. 1994. Close linkage to chromosome 3p and conservation of ancestral founding haplotype in hereditary nonpolyposis colorectal cancer families. Proc. Natl. Acad. Sci. 91: 6054–6058.

Oliveira-Ferreira, F, Napoli-Ferreira, CC, Rossi, BM, Nakagawa, WT, Aguiar, S, Monteiro-Santos, EM, Vieira-Costa, ML, Lopes, A. 2004. Frequency of extra-

colonic tumors in hereditary nonpolyposis colorectal cancer (HNPCC) and familial colorectal cancer (FCC) Brazilian families: An analysis by a Brazilian Hereditary Colorectal Cancer Institutional Registry. Fam. Cancer 3: 41–47.

Onciu, M. 2009. Acute Lymphoblastic Leukemia. Hematol. Oncol. Clin. North Am. 23: 655–674.

Päivi Peltomäki and Annette Gylling (2011). Brain Tumors and the Lynch Syndrome, In Management of CNS Tumors, Dr. Miklos Garami (Ed.), InTech, DOI: 10.5772/21293. p45-50 -Available from: https://www.intechopen.com/books/management-of-cns-tumors/brain-tumors-and-the-lynch-syndrome

Palles, C, Cazier, J-B, Howarth, KM, Domingo, E, Jones, AM, Broderick, P, Kemp, Z, Spain, SL, Guarino, E, Salguero, I, Sherborne, A, Chubb, D, Carvajal-Carmona, LG, Ma, Y, Kaur, K, Dobbins, S, Barclay, E, Gorman, M, Martin, L, Kovac, MB, Humphray, S, Lucassen, A, Holmes, CC, Bentley, D, Donnelly, P, Taylor, J, Petridis, C, Roylance, R, Sawyer, EJ, Kerr, DJ, Clark, S, Grimes, J, Kearsey, SE, Thomas, HJW, McVean, G, Houlston, RS, Tomlinson, I. 2013. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat Genet 45: 136–144.

Pardali, K, Moustakas, A. 2007. Actions of TGF- β as tumor suppressor and pro-metastatic factor in human cancer. Biochim. Biophys. Acta - Rev. Cancer. 1775: 21–62.

Peltomaki, P. 2001. Deficient DNA mismatch repair: a common etiologic factor for colon cancer. Hum. Mol. Genet. 10: 735–740.

Peltomaki, P. 2003. Role of DNA Mismatch Repair Defects in the Pathogenesis of Human Cancer. J. Clin. Oncol. 21: 1174–1179.

Peltomäki, P, Gylling, A. 2011. Brain Tumors and the Lynch Syndrome. In: Management of CNS Tumors., p 363–382.

Peña-Diaz, J, Jiricny, J. 2012. Mammalian mismatch repair: Error-free or error-prone? Trends Biochem. Sci. 37: 206–214.

Pérez-Cabornero, L, Infante, M, Velasco, E, Lastra, E, Miner, C, Durán, M. 2013. Genotype-phenotype correlation in MMR mutation-positive families with Lynch syndrome. Int. J. Colorectal Dis. 28: 1195–1201.

Pietras, K, Ostman, A. 2010. Hallmarks of cancer: interactions with the tumor stroma. Exp. Cell Res. 316: 1324–31.

Pinheiro, M, Pinto, C, Peixoto, A, Veiga, I, Mesquita, B, Henrique, R, Baptista, M, Fragoso, M, Sousa, O, Pereira, H, Marinho, C, Moreira Dias, L, Teixeira, MR. 2011. A novel exonic rearrangement affecting MLH1 and the contiguous LRRFIP2 is a founder mutation in Portuguese Lynch syndrome families. Genet. Med. 13: 895–902.

Pino, MS, Chung, DC. 2010. The Chromosomal Instability Pathway in Colon Cancer. Gastroenterology 138: 2059–2072.

Plotz, G, Piiper, A, Wormek, M, Zeuzem, S, Raedle, J. 2006. Analysis of the human MutL alpha MutS alpha complex. Biochem. Biophys. Res. Commun. 340: 852–859.

Plotz, G, Raedle, J, Brieger, A, Trojan, J, Zeuzem, S. 2002. hMutSalpha forms an ATP-dependent complex with hMutLalpha and hMutLbeta on DNA. Nucleic Acids Res. 30: 711–718.

Poley, J-W, Wagner, A, Hoogmans, MMCP, Menko, FH, Tops, C, Kros, JM, Reddingius, RE, Meijers-Heijboer, H, Kuipers, EJ, Dinjens, WNM. 2007. Biallelic germline mutations of mismatch-repair genes: a possible cause for multiple pediatric malignancies. Cancer 109: 2349–56.

Ponder, B a. 2001. Cancer genetics. Nature 411: 336–41.

Poaty, H., Aba Gandzion, C., Soubeyran, I., Gassaye, D., Peko, J.F., Nkoua Bon, J.B., Gombé Mbalawa, C., 2017. The identification of Lynch syndrome in Congolese colorectal cancer patients. Bull. Cancer 104, 831–839.

Rajagopalan, H, Bardelli, A, Lengauer, C, Kinzler, KW, Vogelstein, B, Velculescu, VE. 2002. Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. Nature 418: 934.

Ramesar, RS, Madden, M V, Felix, R, Harocopos, CJ, Westbrook, CA, Jones, G, Cruse, JP, Goldberg, PA. 2000. Molecular genetics improves the management of hereditary non-polyposis colorectal cancer. S. Afr. Med. J. 90: 709–714.

Rauen, K a, Huson, SM, Burkitt-Wright, E, Evans, DG, Farschtschi, S, Ferner, RE, Gutmann, DH, Hanemann, CO, Kerr, B, Legius, E, Parada, LF, Patton, M, Peltonen, J, Ratner, N, Riccardi, VM, van der Vaart, T, Vikkula, M, Viskochil, DH, Zenker, M, Upadhyaya, M. 2014. Recent developments in

neurofibromatoses and RASopathies: Management, diagnosis and current and future therapeutic avenues. Am. J. Med. Genet. A 167A: 1–10.

Reeve, JP, Rannala, B. 2002. DMLE+: Bayesian linkage disequilibrium gene mapping. Bioinforma. Appl. NOTE 18: 894–895.

Ribic, CM, Sargent, DJ, Moore, MJ, Thibodeau, SN, French, AJ, Goldberg, RM, Hamilton, SR, Laurent-Puig, P, Gryfe, R, Shepherd, LE, Tu, D, Redston, M, Gallinger, S. 2003. Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. N. Engl. J. Med. 349: 247–57.

Ricciardone, MD, Tayfun, O. 1999. Advances in Brief Human MLH1 Deficiency Predisposes to Hematological Malignancy and. Cancer Res. 59: 290–293.

Ripperger, T, Beger, C, Rahner, N, Sykora, KW, Bockmeyer, CL, Lehmann, U, Kreipe, HH, Schlegelberger, B. 2010. leukemia / lymphoma – report on a novel biallelic MSH6 mutation St ta Fe rra or ti Fo un da tio n Fo un tio n rra ta St or ti Fe. 841–844.

Rodney J. Scott, Stuart Reeves and Bente Talseth-Palmer (2012). The Role of Modifier Genes in Lynch Syndrome, In Colorectal Cancer Biology - From Genes to Tumor, Dr. Rajunor Ettarh (Ed.), InTech, DOI: 10.5772/28021. 36-58 -Available from: https://www.intechopen.com/books/colorectal-cancer-biology-from-genes-to-tumor/the-role-of-modifier-genes-in-lynch-syndrome.

Rodriguez-Abreu, D, Bordoni, A, Zucca, E. 2007. Epidemiology of hematological malignancies. Ann. Oncol. 18: i3–i8.

Rodrrguez-bigas, MA, Vasen, HFA, Ph, D. 1997. Rectal Cancer Risk in Hereditary Nonpolyposis Colorectal Cancer After Abdominal Colectomy. 225: 202–207.

Rogozin, I.B., Lada, A.G., Goncearenco, A., Green, M.R., De, S., Nudelman, G., Panchenko, A.R., Koonin, E. V, Pavlov, Y.I., 2016. Activation induced deaminase mutational signature overlaps with CpG methylation sites in follicular lymphoma and other cancers. Sci. Rep. 6, 38133.

Rothwell, PM, Fowkes, FGR, Belch, JF, Ogawa, H, Warlow, CP, Meade, TW. 2011. Effect of daily aspirin on long-term risk of death due to cancer: Analysis of individual patient data from randomised trials. Lancet 377: 31–41.

Rothwell, PM, Wilson, M, Elwin, CE, Norrving, B, Algra, A, Warlow, CP, Meade,

TW. 2010. Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials. Lancet 376: 1741–1750.

Rustgi, AK. 2007. The genetics of hereditary colon cancer. Genes Dev. 21: 2525–38.

Sana, A., Farid, Z., Amel, M., Nadia, K., Amel, M., Nabil, A., Taoufik, N., Farhat, B.A., Sarah, B.J., Mohamed, M., Marie-Pierre, B., 2014. [Molecular characterization of Lynch syndrome in Tunisia]. Tunis. Med. 92, 458–462.

Sargent, DJ, Marsoni, S, Monges, G, Thibodeau, SN, Labianca, R, Hamilton, SR, French, AJ, Kabat, B, Foster, NR, Torri, V, Ribic, C, Grothey, A, Moore, M, Zaniboni, A, Seitz, JF, Sinicrope, F, Gallinger, S. 2010. Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. J. Clin. Oncol. 28: 3219–3226.

Schofield, M,J, Hsieh,P. 2003. DNA mismatch repair: molecular mechanisms and biological function, Annu. Rev. Microbiol. 57: 579–608.

Scott, RJ, Reeves, S, Talseth-palmer, B. 2012. The Role of Modifier Genes in Lynch Syndrome. In: Colorectal Cancer Biology- from genes to tumour., p 36–58.

Shankland, KR, Armitage, JO, Hancock, BW. 2012. Non-Hodgkin lymphoma. Lancet 380: 848–57.

Shin, Y-K, Heo, S-C, Shin, J-H, Hong, S-H, Ku, J-L, Yoo, B-C, Kim, I-J, Park, J-G. 2004. Germline mutations in MLH1, MSH2 and MSH6 in Korean hereditary non-polyposis colorectal cancer families. Hum. Mutat. 24: 351.

Shlien, A, Campbell, BB, de Borja, R, Alexandrov, LB, Merico, D, Wedge, D, Van Loo, P, Tarpey, PS, Coupland, P, Behjati, S, Pollett, A, Lipman, T, Heidari, A, Deshmukh, S, Avitzur, N, Meier, B, Gerstung, M, Hong, Y, Merino, DM, Ramakrishna, M, Remke, M, Arnold, R, Panigrahi, GB, Thakkar, NP, Hodel, KP, Henninger, EE, Göksenin, a Y, Bakry, D, Charames, GS, Druker, H, Lerner-Ellis, J, Mistry, M, Dvir, R, Grant, R, Elhasid, R, Farah, R, Taylor, GP, Nathan, PC, Alexander, S, Ben-Shachar, S, Ling, SC, Gallinger, S, Constantini, S, Dirks, P, Huang, A, Scherer, SW, Grundy, RG, Durno, C, Aronson, M, Gartner, A, Meyn, MS, Taylor, MD, Pursell, ZF, Pearson, CE, Malkin, D, Futreal, PA, Stratton, MR, Bouffet, E, Hawkins, C, Campbell, PJ, Tabori, U. 2015. Combined hereditary and somatic mutations of replication error repair genes result in

rapid onset of ultra-hypermutated cancers. Nat. Genet. 47: 257–262.

Sinicrope, FA, Foster, NR, Thibodeau, SN, Marsoni, S, Monges, G, Labianca, R, Yothers, G, Allegra, C, Moore, MJ, Gallinger, S, Sargent, DJ. 2011. DNA mismatch repair status and colon cancer recurrence and survival in clinical trials of 5-fluorouracil-based adjuvant therapy. J. Natl. Cancer Inst. 103: 863–875.

Stone, JG. 2001. Immunohistochemistry for MSH2 and MHL1: a method for identifying mismatch repair deficient colorectal cancer. J. Clin. Pathol. 54: 484–487.

Stupart, D a, Goldberg, P a, Algar, U, Ramesar, R. 2009. Surveillance colonoscopy improves survival in a cohort of subjects with a single mismatch repair gene mutation. Colorectal Dis. 11: 126–30.

Stupart, DA, Goldberg, PA, Baigrie, RJ, Algar, U, Ramesar, R. 2011. Surgery for colonic cancer in HNPCC : total vs segmental colectomy. 1395–1399.

Talseth-Palmer, B a, Wijnen, JT, Grice, DM, Scott, RJ. 2013a. Genetic modifiers of cancer risk in Lynch syndrome: a review. Fam. Cancer.

Talseth-Palmer, B a, Wijnen, JT, Grice, DM, Scott, RJ. 2013b. Genetic modifiers of cancer risk in Lynch syndrome: a review. Fam. Cancer 12: 207–16.

Tang, R, Hsiung, C, Wang, JY, Lai, CH, Chien, HT, Chiu, LL, Liu, CT, Chen, HH, Wang, HM, Chen, SX, Hsieh, LL. 2009. Germ line MLH1 and MSH2 mutations in Taiwanese Lynch syndrome families: Characterization of a founder genomic mutation in the MLH1 gene. Clin. Genet. 75: 334–345.

Tomsic, J, Liyanarachchi, S, Hampel, H, Morak, M, Thomas, BC, Raymond, VM, Chittenden, A, Schackert, HK, Gruber, SB, Syngal, S, Viel, A, Holinski-Feder, E, Thibodeau, SN, De La Chapelle, A. 2012. An American founder mutation in MLH1. Int. J. Cancer 130: 2088–2095.

Uddin, S, Alkuraya, FS, Al-kuraya, KS. 2013. Constitutional Mismatch Repair-Deficiency Syndrome Is a Rare Cause of Cancer Even in a Highly Consanguineous. 2013: 996–1004.

Umar, a., Boland, CR, Terdiman, JP, Syngal, S, Chapelle, a. DL, Ruschoff, J, Fishel, R, Lindor, NM, Burgart, LJ, Hamelin, R, Hamilton, SR, Hiatt, R a., Jass, J, Lindblom, a., Lynch, HT, Peltomaki, P, Ramsey, SD, Rodriguez-Bigas, M a., Vasen, HF a., Hawk, ET, Barrett, JC, Freedman, a. N, Srivastava, S. 2004.

Revised Bethesda Guidelines for Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome) and Microsatellite Instability. JNCI J. Natl. Cancer Inst. 96: 261–268.

Umar, A, Boland, CR, Terdiman, JP, Syngal, S, De, A, Rüschoff, J, Fishel, R, Lindor, NM, Burgart, LJ, Hamilton, SR, Hiatt, RA, Jass, J, Lindblom, A, Henry, T, Peltomaki, P, Ramsey, SD, Rodriguez-bigas, MA, Vasen, HFA, Hawk, ET, Barrett, JC, Freedman, AN, Srivastava, S. 2010. NIH Public Access. 96: 261–268.

Vasen, H. 2004. Mutations associated with HNPCC predisposition – Update of ICG-HNPCC / INSiGHT mutation database. 20: 269–276.

Vasen, H, Watson, P, Mecklin, J, Lynch, H. 1999. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative Group on HNPCC☆. Gastroenterology 116: 1453–1456.

Vasen, HF a, Ghorbanoghli, Z, Bourdeaut, F, Cabaret, O, Caron, O, Duval, a, Entz-Werle, N, Goldberg, Y, Ilencikova, D, Kratz, CP, Lavoine, N, Loeffen, J, Menko, FH, Muleris, M, Sebille, G, Colas, C, Burkhardt, B, Brugieres, L, Wimmer, K. 2014. Guidelines for surveillance of individuals with constitutional mismatch repair-deficiency proposed by the European Consortium "Care for CMMR-D" (C4CMMR-D). J. Med. Genet. 51: 283–93.

Venning, FA, Wullkopf, L, Erler, JT. 2015. Targeting ECM Disrupts Cancer Progression. Front. Oncol. 5: 1–18.

Vergouwe, F, A Boutall, and D Stupart. 2013. Mismatch Repair Deficiency in Colorectal Cancer Patients in a Low-Incidence Area. S Afr J Surg. 51(1):16-21

Vogelstein, B, Lane, D, Levine, AJ. 2000. Surfing the p53 network. 408.

Wagner, A, Barrows, A, Wijnen, JT, van der Klift, H, Franken, PF, Verkuijlen, P, Nakagawa, H, Geugien, M, Jaghmohan-Changur, S, Breukel, C, Meijers-Heijboer, H, Morreau, H, van Puijenbroek, M, Burn, J, Coronel, S, Kinarski, Y, Okimoto, R, Watson, P, Lynch, JF, de la Chapelle, A, Lynch, HT, Fodde, R. 2003. Molecular Analysis of Hereditary Nonpolyposis Colorectal Cancer in the United States: High Mutation Detection Rate among Clinically Selected Families and Characterization of an American Founder Genomic Deletion of the MSH2 Gene. Am. J. Hum. Genet. 72: 1088–1100.

Wakefield, LM, Roberts, AB. 2002. TGF-beta signaling: positive and negative effects on tumorigenesis. Curr. Opin. Genet. Dev. 12: 22–29.

Walsh, MD, Buchanan, DD, Cummings, MC, Pearson, S, Arnold, ST, Clendenning, M, Walters, R, Mckeone, DM, Spurdle, AB, Hopper, JL, Jenkins, MA, Phillips, KD, Suthers, GK, George, J, Goldblatt, J, Muir, A. 2010. Lynch Syndrome – Associated Breast Cancers : Clinicopathologic Characteristics of a Case Series from the Colon Cancer Family Registry. 16: 2214–2225.

Walsh, MD, Cummings, MC, Pearson, S, Clendenning, M, Walters, RJ, Nagler, B, Hopper, JL, Jenkins, MA, Suthers, GK, Goldblatt, J, Tucker, K, Gattas, MR, Arnold, JL, Parry, S, Macrae, FA, McGuckin, M a, Young, JP, Buchanan, DD. 2013. Lynch syndrome-associated breast cancers do not overexpress chromosome 11-encoded mucins. Mod. Pathol. 26: 944–54.

Walsh, S. 2015. The pathology of Lynch Syndrome. Diagnostic Histopathol. 21: 161–164.

Wang, Q, Lasset, C, Desseigne, F, Frappaz, D, Bergeron, C, Navarro, C, Ruano, E, Puisieux, A. 1999. Neurofibromatosis and early onset of cancers in hMLH1-deficient children. Cancer Res. 59: 294–297.

Wang, Q, Montmain, G, Ruano, E, Upadhyaya, M, Dudley, S, Liskay, RM, Thibodeau, SN, Puisieux, A. 2003. Neurofibromatosis type 1 gene as a mutational target in a mismatch repair-deficient cell type. Hum. Genet. 112: 117–23.

Warthin, AS. 1925. The Further Study of a Cancer Family. Cancer Res.: 279–286.

Watson, P, Vasen, HFA, Mecklin, JP, Bernstein, I, Aarnio, M, Järvinen, HJ, Myrhoj, T, Sunde, L, Wijnen, JT, Lynch, HT. 2008. The risk of extra-colonic, extra-endometrial cancer in the Lynch syndrome. Int. J. Cancer 123: 444–449.

Weedon, MN, Ellard, S, Prindle, MJ, Caswell, R, Allen, HL, Oram, R, Godbole, K, Yajnik, CS, Sbraccia, P, Novelli, G, Turnpenny, P, McCann, E, Goh, KJ, Wang, Y, Fulford, J, McCulloch, LJ, Savage, DB, O'Rahilly, S, Kos, K, Loeb, LA, Semple, RK, Hattersley, AT. 2013. An in-frame deletion at the polymerase active site of POLD1 causes a multisystem disorder with lipodystrophy. Nat Genet 45: 947–950.

Whiteside, D, McLeod, R, Graham, G, Steckley, JL, Booth, K, Somerville, MJ, Andrew, SE. 2002. A Homozygous Germ-Line Mutation in the Human MSH2

Gene Predisposes to Hematological Malignancy and Multiple Café-au-Lait Spots Advances in Brief A Homozygous Germ-Line Mutation in the Human MSH2 Gene Predisposes to Hematological Malignancy and Multiple Ca. 359–362.

Wimmer, K, Etzler, J. 2008. Constitutional mismatch repair-deficiency syndrome: have we so far seen only the tip of an iceberg? Hum. Genet. 124: 105–22.

Wimmer, K, Kratz, CP. 2010. Constitutional mismatch repair-deficiency syndrome. Haematologica 95: 699–701.

Wimmer, K, Kratz, CP, Vasen, HF a, Caron, O, Colas, C, Entz-Werle, N, Gerdes, A-M, Goldberg, Y, Ilencikova, D, Muleris, M, Duval, A, Lavoine, N, Ruiz-Ponte, C, Slavc, I, Burkhardt, B, Brugieres, L. 2014. Diagnostic criteria for constitutional mismatch repair deficiency syndrome: suggestions of the European consortium “care for CMMRD” (C4CMMRD). J. Med. Genet. 51: 355–65.

Win, AK, Hopper, JL, Buchanan, DD, Young, JP, Tenesa, A, Dowty, JG, Giles, GG, Goldblatt, J, Winship, I, Boussioutas, A, Young, GP, Parry, S, Baron, J a, Duggan, D, Gallinger, S, Newcomb, P a, Haile, RW, Marchand, L Le, Lindor, NM, Jenkins, M a. 2013a. Are the common genetic variants associated with colorectal cancer risk for DNA mismatch repair gene mutation carriers? Eur. J. Cancer: 1–10.

Win, AK, Hopper, JL, Buchanan, DD, Young, JP, Tenesa, A, Dowty, JG, Giles, GG, Goldblatt, J, Winship, I, Boussioutas, A, Young, GP, Parry, S, Baron, JA, Duggan, D, Gallinger, S, Newcomb, PA, Haile, RW, Le Marchand, L, Lindor, NM, Jenkins, MA. 2013b. Are the common genetic variants associated with colorectal cancer risk for DNA mismatch repair gene mutation carriers? Eur. J. Cancer 49: 1578–87.

Win, AK, Lindor, NM, Jenkins, M a. 2013c. Risk of breast cancer in Lynch syndrome: a systematic review. Breast Cancer Res. 15: 1–9.

Win, AK, Young, JP, Lindor, NM, Tucker, KM, Ahnen, DJ, Young, GP, Buchanan, DD, Clendenning, M, Giles, GG, Winship, I, Macrae, FA, Goldblatt, J, Southey, MC, Arnold, J, Thibodeau, SN, Gunawardena, SR, Bapat, B, Baron, JA, Casey, G, Gallinger, S, Le Marchand, L, Newcomb, PA, Haile, RW, Hopper, JL, Jenkins, MA. 2012. Colorectal and other cancer risks for carriers and noncarriers

from families with a DNA mismatch repair gene mutation: A prospective cohort study. J. Clin. Oncol. 30: 958–964.

World Medical Association. 2013. World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. J. Am. Med. Assoc. 310: 2191–2194.

Xiong, G-F, Xu, R. 2016. Function of cancer cell-derived extracellular matrix in tumor progression. J. Cancer Metastasis Treat. 2: 357.

Xue, W, Zhu, M, Wang, Y, He, J, Zheng, L. 2015. Association between PLCE1 rs2274223 A > G polymorphism and cancer risk: proof from a meta-analysis. Sci. Rep. 5: 7986.

Yoda, Y, Takeshima, H, Niwa, T, Kim, JG, Ando, T, Kushima, R, Sugiyama, T, Katai, H, Noshiro, H, Ushijima, T. 2015. Integrated analysis of cancer-related pathways affected by genetic and epigenetic alterations in gastric cancer. Gastric Cancer 18: 65–76.

Yu, J, Ustach, C, Kim, H-RC. 2003. Platelet-derived growth factor signaling and human cancer. J. Biochem. Mol. Biol. 36: 49–59.

Zerener, T, Sencimen, M, Altun, C, Altug, HA. 2013. Congenital granular cell tumor in newborn. Eur. J. Dent. 7: 497–499.

Zhan, T, Rindtorff, N, Boutros, M. 2016. Wnt signaling in cancer. Oncogene 4(5). pii: a008052.

Zhang, H, Hou, W, Wang, HL, Liu, HJ, Jia, XY, Zheng, XZ, Zou, YX, Li, X, Hou, L, McNutt, MA, Zhang, B. 2014. GSK-3beta-regulated N-acetyltransferase 10 is involved in colorectal cancer invasion. Clin. Cancer Res. 20: 4717–4729.

Zhang, T, Boswell, EL, McCall, SJ, Hsu, DS. 2015. Mismatch repair gone awry: Management of Lynch syndrome. Crit. Rev. Oncol. Hematol. 93: 170–179.

Zhang, X, Li, J. 2013. Era of universal testing of microsatellite instability in colorectal cancer. World J. Gastrointest. Oncol. 5: 12–9.

Ziada-Bouchaar, H., Sifi, K., Filali, T., Hammada, T., Satta, D., Abadi, N., 2017. First description of mutational analysis of MLH1, MSH2 and MSH6 in Algerian families with suspected Lynch syndrome. Fam. Cancer 16, 57–66.

# Appendices

## REQUEST FOR MOLECULAR STUDIES (DNA)

**Molecular Laboratory**
**Division of Human Genetics**
**IIDMM, LEVEL 3**
**UCT Medical School, Observatory 7925**

Tel: (021) 406 6425 Fax: (021) 406 6826

Blood should be drawn in 2 plastic EDTA Tubes (Purple top) +/- 10ml each using a yellow barrel.
Each tube should be inverted to mix and should be clearly labelled with the patient's name and DOB
Keep blood in fridge at 4°C until able to send to laboratory

**Please DO NOT send specimens on ice or frozen.**

***Please fill in all the information requested:***

Surname:_____ First Name(s):_____

New Family: Yes ☐ No ☐ (If no, please fill in family name) Family name: _____

Medical Aid: _____ Medical Aid No: _____

Sex: M ☐ F ☐ Date of Birth: Year:_____ Month:_____ Day: _____

Number of children: _____

Ethnic Origin : ( please indicate ancestry of both your mother and father )_____

Contact Address: _____ Town: _____ Fax: Tel: _____

Referring Doctor/Sister: _____ Town: _____ Fax: Tel: _____

Hospital or Address: _____ Town: _____ Fax: Tel: _____

Reason for Referral (Clinical diagnosis):

Affected ☐ At Risk ☐ Carrier ☐ Spouse ☐ Query ☐ Unaffected ☐

| | | |
|---|---|---|
| Becker Muscular Dys. ☐ | Duchenne Muscular Dys ☐ | Colonic Carcinoma ☐ |
| Fragile-X Syndrome ☐ | Bipolar Disorder ☐ | Huntington Disease ☐ |
| Retinitis Pigmentosa ☐ | Spinocerebellar Ataxia ☐ | Waardenberg Syndrome ☐ |

Additional disorders (apparent or previously treated): _____

Additional family history _____

Clinical Details:

Physical disability ☐ Mental retardation ☐ Deafness ☐ Impaired vision ☐ Night blindness ☐

Other: _____

Have samples from this patient been sent to a DNA lab before? (DELETE WHERE NOT APPLICABLE) YES / NO / Don't Know

If Yes, where: _____

***For Laboratory use only:***

*DNA number:* _____ *Vol.Blood:* _____*(ml)* *Other:* _____

*Date Received: Year:* _____ *Month:* _____ *Day:* _____ *Computer Index No:*_____

## CONSENT FOR DNA ANALYSIS AND STORAGE

1. I, _____ , request that an attempt be made using genetic material to assess the probability that: I / my child / my unborn child (DELETE WHERE NOT APPLICABLE) might have inherited a disease-causing mutation in the gene for:_____

2. I understand that the genetic material for analysis is to be obtained from: blood cells/skin sample/other (specify) (DELETE WHERE NOT APPLICABLE) :

3. I request that **no** portion of the sample be stored for later use. ☐ (MARK IF APPLICABLE )
   Or
   I request that a portion of the sample be stored indefinitely for (DELETE WHERE NOT APPLICABLE):
   ( a ) possible re-analysis
   ( b ) analysis for the benefit of members of my immediate family
   ( c ) research purposes, subject to the approval of the University of Cape Town Research Ethics Committee, provided that any information from such research will remain confidential.

4. The results of the analysis carried out on this sample of stored biological material will be made known to me, via my doctor, in accordance with the relevant protocol, if and when available.
   In addition, I authorise that they may be made known to: (DELETE WHERE NOT APPLICABLE) :
   other doctors involved in my care
   the following family members: _____
   other: _____

5. I authorise / do not authorise my doctor(s) (DELETE WHERE NOT APPLICABLE) to provide relevant clinical details to the Division of Human Genetics, UCT.

6. I have been informed that:
   ( a ) there are risks and benefits associated with genetic analysis and storage of biological material and these have been explained to me.
   ( b ) the analysis procedure is specific to the genetic condition mentioned above and cannot determine the complete genetic makeup of an individual.
   ( c ) the genetics laboratory is under an obligation to respect medical confidentiality .
   ( d ) genetic analysis may not be informative for some families or family members.
   ( e ) even under the best conditions, current technology of this type is not perfect and could lead to incorrect results.
   ( f ) where biological material is used for research purposes, there may be no direct benefit to me.

7. I understand that I may withdraw my consent for any aspect of the above at any time without this affecting my future medical care.

8. **ALL OF THE ABOVE HAS BEEN EXPLAINED TO ME IN A LANGUAGE THAT I UNDERSTAND AND MY QUESTIONS ANSWERED BY:**

DATE:

**Patient signature** _____ **Witnessed consent** _____

**NOTE - PLEASE INSERT A FAMILY PEDIGREE DRAWING ON THE REVERSE OF THIS FORM**

## Appendix B: Nucleic acid extraction methods

**Preparation of samples for DNA extraction from freshly frozen tissue:**

Samples were kept frozen at all times
**Clear the area to ensure its is RNAse free
Label 1.5mL tubes and petri-dishes accordingly

**For DNA**
Add 600uL of cell lysis buffer into 1.5mL tube
Add 400uL of cell lysis buffer into the petri dish

Using surgical blades dissect enough tissue into the petri dish (small slices are preferred) a slice and a half should be sufficient, however for tissues full of lipids, such as the brain, you may need to get more slices.

Break the tissue into the smallest pieces as possible to ensure complete lysis
Transfer the tissue into the 1.5mL tube
Leave as much lysis buffer containing blood in the dish as possible

**For DNA isolation:**

1. Following the addition of tissues into the 1.5mL tube add 20uL of 10% SDS and 5uL of Proteinase-K , keep over night or over 2 days. The lysis period depends on the amount of tissue.
2. Add 200uL of the 6M NaCl and vortex thoroughly
3. Centrifuge at 3000rpm for 20minutes
4. Transfer the supernatant into the new 1.5mL tube
5. Add 1mL of absolute Ethanol to the supernatant and invert the tube 50times
6. Centrifuge for 10min at 5000rpm
7. Discard the supernatant
8. Ass 500uL ethanol and vortex
9. Centrifuge at 5000rpm for 10minutes
10. Decant the supernatant and air-dry the pellet for not more than 2hours

Add 50 to 100uL of 1XTE buffer to the pellet

**Ion Ampliseq FFPE DNA extraction protocol:**
https://tools.thermofisher.com/content/sfs/brochures/ion-ampliseq-direct-ffpe-dna-kit-flyer.pdf

**Prepare reagents**
• Equilibrate Transfer Solution to room temperature (15–30°C) before use.
• Keep Direct Reagent on ice prior to use.

**Prepare Direct FFPE DNA**

The recommended tissue area to be used for this protocol is **4–100 mm²** from a **5–10 µm** thick unstained section mounted on a slide.
If desired, scrape unwanted tissue from the slide before transfer.

## For each sample,

Pipette **30 µL** of **Transfer Solution** into a single well of a 96-well PCR plate/ PCR tube.
Note: Transfer Solution is viscous, pipet slowly.

**Note: Alternatively, label one nuclease-free 1.5-mL Eppendorf LoBindTM tube for each FFPE tissue sample and perform incubations in a heated block.**

Using a single 20-µL pipette tip for each sample:

a. Pipette **2–10 µL of the Transfer Solution** from the well onto the region of interest of the FFPE tissue section mounted on a slide.

b. Using the same 20-µL-pipette tip**, spread the Transfer Solution** to ensure complete coverage of the region of interest, then scrape and break up the tissue with the pipette tip.
The tissue should be a slurry of fine particles in the Transfer Solution.

c. **Pipette the slurry from the slide** back into the PCR tube containing **Transfer Solution.**

d. Pipette the slurry up and down at least **five times**, leaving as much tissue as possible in the PCR tube.

e. If needed, use the same tip to repeat steps b-d, transferring as much of the region of interest as possible into the PCR tube.

**Note: The total volume of Transfer Solution remaining in the 96-well plate may vary, but no volumetric adjustment is required.**

f. **Add 21 µL of Direct Reagent** to each well containing sample in the PCR tube.

g. Set a pipette to 30 µL, then **mix the Direct Reagent and slurry** by pipetting up and down **ten times**.

h. Seal the tube, then verify that the contents are at the bottom of the tube.

**Note: If necessary, gently tap the tube on a hard at surface to collect the contents at the bottom of the wells.**

i. Place a compression pad on the plate, load the plate into the thermal cycler, then run the following program:

**65** degrees Celsius for **15 minutes**
**20** degrees Celsius for **30 minutes**

## Appendix C: Gel-electrophoresis

### Reagents preparation

**10X TBE buffer**

Tris (MW = 121.134) 108g

Boric Acid 55g

EDTA (pH 8) 7.4gt

Made up to 1000ml with distilled water

**1X TE buffer**

10mM Tris (121.134) 12.11g

1mM Ethylene-diaminetetra-acetic acid (EDTA) (MW 372.24) 1.87g

Made up to 1000ml with distilled water

**Agarose loading dye**

Bromo phenol blue (0.25% w/v) 0.25g

Sucrose (40% w/v) 40g

EDTA, pH 8 4ml

Made up to 100ml distilled water

***Gel preparation***

**1% , 1.5% and 2% Agarose gel (100ml)**

1g/ 1.5g/ 2g Agarose

100ml 1X TBE

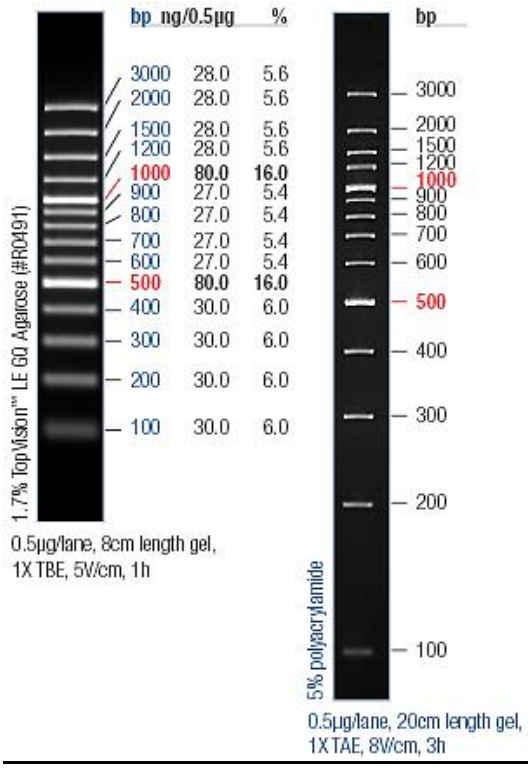
Heat until the agarose dissolved

Allow to cool and poor into the gel tray and allow setting

***Molecular weight Marker:***

***Gene-Ruler 100pb plus (Fermentas Inqaba)***

A 1/10 dilution of the marker solution was made. From the 1:10 solution 125µl of the marker was added to 125µl of loading dye.




**Range:**

**14 fragments (in bp): 3000, 2000, 1500, 1200, 1000, 900, 800, 700, 600, 500, 400, 300, 200, 100**

## Appendix D: Promega MSI analysis protocol

**DNA Amplification Using the MSI Analysis System**
https://www.promega.com/resources/protocols/technical-manuals/0/msi-analysis-system-version-12-protocol/

Materials

- thermal cycler [GeneAmp® System 9600 or 9700 (Applied Biosystems)]
- microcentrifuge
- 0.2ml (thin-walled) microcentrifuge tubes, MicroAmp® reaction tube strips or MicroAmp® optical 96-well reaction plates (Applied Biosystems)
- 1.5ml microcentrifuge tubes
- aerosol-resistant pipette tips
- AmpliTaq Gold® DNA polymerase (Life Technologies Cat.# N8080242)

The MSI Analysis System is optimized to amplify 1–2ng of genomic DNA in a 10µl reaction volume using the protocols detailed below. However, optimization of input DNA amounts should be performed to adjust for variations in DNA yield and quality due to differences in samples, DNA isolation methods or both. Using excessive amounts of DNA template may result in peak heights exceeding the linear detection range of the CE instruments. Use of insufficient DNA template can result in low PCR yields, and peak heights may fall below detection limits (50RFU). Accurate quantitation of template DNA is highly recommended. We recommend quantitating the FFPE template DNA using the Promega QuantiFluor® ONE dsDNA System (Cat.# E4870 or E4871) or QuantiFluor® dsDNA System (Cat.# E2670). Testing at Promega has shown that DNA quantitation using the QuantiFlour® ONE dsDNA System with the QuantusTM Fluorometer provided results more similar to those of qPCR from DNA samples extracted from three FFPE tissue types as compared to the NanoDrop® 2000 instrument (1).
The MSI Analysis System is optimized for use with the GeneAmp® PCR System 9600 and 9700 thermal cyclers.

4.A. Amplification Setup
Note: We strongly recommend using gloves and aerosol-resistant pipette tips to prevent cross-contamination. We recommend keeping all pre-amplification and post-amplification reagents in separate rooms. Prepare ampli cation reactions in a room dedicated for reaction setup. Use equipment and supplies dedicated for amplification setup.
Thaw the Gold STHR 10X Bu er, MSI 10X Primer Pair Mix and Nuclease-Free Water.

Mix these reagents by vortexing for 5–10 seconds before each use. A precipitate may form in the Gold STHR

10X Buffer. If this occurs, warm the buffer at 37°C, then vortex until the precipitate is in solution.

To prepare the Amplification Mix, determine the number of reactions to be set up. This should include positive and negative control reactions. Add 1 or 2 reactions to this number to compensate for losses during pipetting. This approach ensures that you will have enough PCR master mix for all samples. It also ensures that each reaction contains the same master mix.

Place one 0.2ml micro-centrifuge tube for each reaction into a rack, and label appropriately. Alternatively use MicroAmp® optical 96-well reaction plates.

Combine the volumes of Nuclease-Free Water, Gold STHR 10X Bu er, MSI 10X Primer Pair Mix and AmpliTaq Gold® DNA polymerase in a sterile, 1.5ml tube. Mix gently.

Transfer 8µl of Amplification Mix to the bottom of each reaction tube or well. Pipet 2µl of template DNA (1–2ng) for each sample into the bottom of the appropriate tube or well containing Amplification Mix. Mix by pipetting several times.
Be sure that the template DNA is mixed well before transferring it to the tube or well containing Amplification Mix.
Note: Store DNA templates in nuclease-free water or TE–4 buffer [10mM Tris HCl (pH 8.0), 0.1mM EDTA]. If the template DNA is stored in TE buffer that is not pH 8.0 or contains a higher EDTA concentration, the volume of DNA sample added should not exceed 20% of thermal reaction volume. PCR amplification efficiency and quality can be greatly altered by changes in pH (due to added Tris-HCl), available magnesium concentration (due to chelation by EDTA) or other PCR inhibitors, which may be present at low concentrations depending on the source of template DNA and the extraction procedure used.

For the positive amplification control, dilute the K562 High Molecular Weight DNA 1:10 to 1ng/µl in Nuclease- Free Water. Pipet 2ng of the diluted DNA into the bottom of the tube or well containing Amplification Mix. Mix by pipetting several times.
For the negative amplification control, pipet Nuclease-Free Water (instead of template DNA) into a microcentrifuge reaction tube or well containing Amplification Mix. Mix by pipetting several times.

Protocol for the GeneAmp® PCR System 9600 Thermal Cycler Cycling profile:
95°C for 11 minutes, then: 96°C for 1 minute, then:
94°C for 30 seconds ramp, 68 seconds to 58°C, hold for 30 seconds ramp 50 seconds to 70°C, hold for 1 minute for 10 cycles, then:
90°C for 30 seconds ramp, 60 seconds to 58°C, hold for 30 seconds ramp 50 seconds to 70°C, hold for 1 minute for 20 cycles, then:
60°C for 30 minutes and
4°C hold