Screening for Lynch Syndrome in the Saudi population

Masood Alqahtani
Hi diploma, BS, MSc

This thesis is presented for the degree of Doctor of Philosophy at the University of Western Australia

School of Biomedical Science

August, 2018
I, Masood Alqahtani, certify that:

This thesis has been substantially accomplished during enrolment in the degree.

This thesis does not contain material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution.

No part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of The University of Western Australia and where applicable, any partner institution responsible for the joint-award of this degree.

This thesis does not contain any material previously published or written by another person, except where due reference has been made in the text.

The work(s) are not in any way a violation or infringement of any copyright, trademark, patent, or other rights whatsoever of any person.

The research involving human data reported in this thesis was assessed and approved by The University of Western Australia Human Research Ethics Committee. Approval #: RA/4/1/7322

Written patient consent has been received and archived for the research involving patient data reported in this thesis.

The following approvals were obtained from three tertiary health institutes prior to commencing the relevant work described in this thesis:

A: Initial Screening (Chapter 3)

- King Fahad Specialist Hospital-Dammam, IRB LAB 055 and IRB Lab055
- King Khaled University Hospital, 15/0148/IRB
- King Fahad University Hospital, IRB-2014-01-297
B: Ethics approval for patient follow up and germline testing of MSI cases (Chapter 4) was obtained from three referral hospitals

- Fahad Specialist Hospital-Dammam, IRB no. LAB0305
- King Khaled University Hospital, IRB no. E-15-1468
- King Fahad University Hospital, IRB no. 2014-01-297

C: Ethics approval for the *In-Silico* project (Chapter 5) was obtained from the Saudi Health Council, Saudi Cancer Registry scientific committee (May 01, 2016)

Saudi Health Council supported the research with 5 letters directed to the 5 referral hospitals and granted from the following centres:

- King Saudi University, Institutional review board ref: 17/0427/IRB
- King Fahad Medical City, Institutional review board ref: 17-412E

The work described in this thesis was funded by the Saudi Arabia Cultural Mission, Canberra, Australia.

Technical assistance was kindly provided by:

- PathWest laboratory at Sir Charles Gairdner Hospital for allowing full access and use of the molecular pathology laboratory in Anatomical Pathology
- Histocompatibility and Immunogenetics laboratory at King Fahad Specialist Hospital-Dammam, Saudi Arabia for full access and use of the molecular sciences equipment
- Department of Pathology and Laboratory medicine, King Saud University for access to the histopathology lab
- Department Of Pathology and Laboratory Medicine, Dammam University for access to the laboratory

Signature:

Date: Aug 12,
1. This thesis does not contain work that I have published, nor work under review for publication.

2. This thesis contains only sole-authored work, some of which has been published and/or prepared for publication under sole authorship. The bibliographical details of the work and where it appears in the thesis are outlined below.

3. This thesis contains published work and/or work prepared for publication, some of which has been co-authored. The bibliographical details of the work and where it appears in the thesis are outlined below.

The student must attach to this declaration a statement for each publication that clarifies the contribution of the student to the work. This may be in the form of a description of the precise contributions of the student to the published work and/or a statement of percent contribution by the student. This statement must be signed by all authors. If signatures from all the authors cannot be obtained, the statement detailing the student’s contribution to the published work must be signed by the coordinating supervisor.

Manuscripts published:


The work presented in this manuscript is included in chapters 2, 3, and 4.

Student’s contribution:

- Initiated and completed all ethics application forms
• Generated patient information sheet, patient informed consent and laboratory agreement form
• Conducted all experiments in the manuscript with assistance from each hospital lab supervisor and supervision from the remaining authors. I also wrote the draft manuscript that was subsequently edited by other authors.

**Student Contribution 80%**


The work presented in this manuscript is included in chapters 2 and 4 of the thesis.

**Student’s contribution**

• Initiated and completed all ethics application forms.
• Created patient information sheet, patient informed consent and Laboratory agreement form
• Created genetic pedigree form
• I conducted all experiments in this manuscript with assistance from each hospital lab supervisor and supervision from the remaining authors. I also wrote the draft manuscript that was subsequently edited by other authors.

**Student Contribution 75%**

**Presentations**

Dedication

I would like to dedicate this thesis to my mother: Amrah, and my Father: Turki and my wife: Felwah, my children, my brothers and sisters, and PhD supervisor Professor Barry Iacopetta
ABSTRACT

Hereditary non-polyposis colorectal cancer, also known as Lynch syndrome (LS), is an autosomal dominant genetic condition associated with a lifetime risk of approximately 80% for the development of colorectal cancer (CRC). Other extra-colonic cancers within the Lynch spectrum of cancers include endometrial, gastric, ovarian and renal tumors. Approximately 1-2% of CRC cases are thought to be due to LS, which is characterised clinically by an early age of onset (average of 45 years) and the occurrence of multiple related cancers within first- and second-degree family members. LS arises because of pathogenic germline variants in mismatch repair (MMR) genes, most commonly \( MLH1 \), \( MSH2 \), \( MSH6 \) and \( PMS2 \). Almost all tumors associated with LS are characterised by ubiquitous changes in the length of DNA microsatellite repeats (microsatellite instability, or MSI) and loss of expression of MMR genes as detected by immunohistochemistry (IHC). Regular screening and early identification of pathogenic variants carriers has been demonstrated to improve survival. Also, the implementation of proper genetic counselling for mutation carriers and regular screening by colonoscopy has been shown to reduce mortality from CRC.

The proportion of LS cases amongst all CRC patients has been reported as up to 5%, however this percentage is likely to vary from one population to another. Until recently, CRC patients who were at high risk of having LS were identified through the use of clinical criteria such as the Amsterdam and Bethesda guidelines, which rely on obtaining a detailed family history of cancer. However, use of these guidelines has been associated with low sensitivity for the detection of LS and their implementation in routine clinical practice has been poor. The Saudi Arabian population is young compared to
Western countries, and over half of the CRC cases in this country are aged less than 60 years at diagnosis. This is alarming and, together with high levels of consanguinity, may indicate the presence of familial involvement and in particular LS.

The overall aim of this thesis was therefore to determine the prevalence of LS in the Saudi national population of CRC patients using MSI screening as the initial test. A secondary aim was to determine whether routine screening for MSI in CRC from young patients leads to the identification of previously unrecognised cases of LS in this population. In the study described in Chapter 3, CRC from a retrospective cohort of 284 young Saudi patients (<60 years at diagnosis) were tested for MSI in the absence of knowledge of family history of cancer. The frequency of MSI was 11.6% (33/284), which is similar to previous studies of large cohorts of CRC patients. In Chapter 4, 13 of the 33 MSI patients were successfully recruited for germline mutation testing and information on family history of cancer. The number of germline mutation carriers identified was 9. Of these, 8 were confirmed as pathogenic variants according to the American College of Medical Genetics and Genomics (ACMG) and one was classified as a variant of unknown pathogenicity. These results suggest that about 7% (11.6% x 8/13) of young CRC patients from Saudi Arabia have LS.

Chapter 5 describes a novel strategy to identify family clusters among young patients reported to the Saudi Arabia Cancer Registry (SCR) and diagnosed with one or more of the LS spectrum of cancers, i.e. CRC, endometrial, ovarian, gastric or kidney. Demographic details for 20,081 young cancer patients including full name and address, date and place of birth, together with their tumor type based on clinical and/or histopathological diagnosis, were extracted from SCR database. This information was
exported to a specially designed Microsoft Access database that allowed datamining using several queries aimed at identifying family cancer clusters that could indicate the presence of LS. The algorithm used to mine this database proved to be reasonably accurate, since 13 of the 17 (76%) potential family cancer clusters followed up to date were confirmed as being first degree relatives. Of these, tumors from 8 family clusters were tested for MSI and two were found to be positive, with these being very strong candidates for LS.

**Conclusions:** The work presented in this thesis suggests that population screening using MSI testing on young (<60 years) CRC patients is a feasible, sensitive, cost-effective and clinically acceptable strategy to identify LS, either retrospectively or prospectively. Also, further refinement of the *In-Silico* approach may also be useful for the identification of family cancer clusters with LS. Together, these approaches to identifying LS in the Saudi population could help to significantly reduce the economic burden by allowing early detection and treatment in mutation carriers. Other major findings of this work were:

- The prevalence of LS amongst young Saudi Arabian CRC patients is significantly higher than in equivalent West Australian patients
- The frequency and length of microsatellite deletions are different between Saudi and West Australian CRC patients with MSI. To our knowledge, such geographic differences in MSI tumors have not been previously reported
- Current limitations to the identification of LS amongst Saudi CRC patients include incomplete information on family history of cancer in medical records, the lack of
routine screening for MSI or IHC in CRC, and the frequent migration of patients from city to city which poses a major challenge for record keeping and follow up.
ACKNOWLEDGEMENTS

In the name of Allah, the Most Gracious and the most Merciful this thesis has been possible.

Words of wisdom: He who does not thank the people is not thankful to the Lord, Prophet Mohammed peace be upon him

I would like to thank all collaborators: Dr Rabab Alarras, Mrs Jan Kazan, Dr. Miral Mashhoor, Saber Alzahrani, Ahmed Alotaibi, Mansor Almohi, Dr Khaled Alsaleh, Dr Abdulmalik Alsheikh, and Dr. Dalal Altamimi for their support and facilities at their institute throughout this PhD project. I also would like to thank his Excellency the General Secretary of the Saudi Health Council Dr Yagob Al-Mazrou, his deputy Dr Abdulllah Alhariri, Dr Ali Al-Zahrani, Dr Hind Almutlaq and Mr Abdulrahman Albaibi for their support within the Saudi health council and cross the concerns Oncology centre. Your belief in the significance and importance of familial cancer screening and prevention has helped in the accomplishment of this project.

I would like to acknowledge my sincere appreciation to the Academic Affairs, the financial body, King Fahad Specialist hospital-Dammam and Ministry Of Health in Saudi Arabia for funding this PhD. In addition, my sincere appreciation goes to all the staff in the Saudi Cultural Bureau for their support.

I would like to thank Dr Benhur Amanuel, Amerigo Carrello and Fabienne Grieu for providing me with full access to their molecular pathology laboratory and for their feedback.

To the only one, Professor Barry Iacopetta. My supervisor, there are no words can describe my sincere gratitude to you for being available to me every day and in every step of the way throughout my PhD project. Your way of communication, thinking and feedback was not only helpful but it encouraged my research skills. Moreover, I became
more involved in other areas of research within the field of cancer prevention and screening allowing me to have more relationship with Saudi Cancer Registry research team and Saudi Health council. Thank you for your critical analysis, quick feedback and recommendations when reviewing this thesis. Your very caring personality, understanding and emotional support had really helped me and my family during my hard time. I will not forget the time you spent reviewing and monitoring my project progress. And I will not forget your support when my uncle passed away back in 2015.

To my brother Abdulaziz who raised me and who chose to drop from the college to take care of our family after my dad passed away, I will never forget your stand as a father, brother and mentor. You’re full of kindness and joy which makes our life easy and hopeful. You certified your futures for ours to flourish. I wrote these lines with tears as it indeed came from heart. Thank you for everything. And I ask all mighty Allah that I can payback what you have done for me.

To my beloved sisters and brothers, thank you so much for being supportive and encouraging which boosted my energy and helped me to work hard. The gathering you organised during my visit to Saudi Arabia brought a joy and happiness in a time where I was in despite need.

To my soulmate and lovely wife, Felwah, Thank you for always helping me think clearly, for tolerating our kid’s bouncing and playing to provide me with a quiet time, for believing in me and my competence. Thank you for making my ambition your own. This work is yours and for you, and I am yours. I would like to thank my kids, Abduaziz, Omar, Ibrahim and Abdulrahman for being helpful and supportive, Abduaziz for being my right hand, Omar for being the clown of the house, Ibrahim for being emotionally so close to your parents and Abdulrahman for being always ready when asked.

I am speechless as there are no words can will justify your Thank you so much no one will.

To my paradise, my mother Amrah and father Turki, I called you paradise because that
how I felt when I see you. No one believed that I will survive when I received the polio vaccine in 1974 except you. You slept at the hospital park during wintertime to be close to me, you prayed day and night for my survival. Thank you so much for never giving up-on our family or on me. You are exceptional parents who pay a lot of time teaching me how to be independent, polite and respectful. You lived poor to provide a clean home and food for me and my siblings, you watch us day and night to make sure that we have a peaceful sleep. You have helped us financially, emotionally and encouraged us to do the best and to be proud of who we are. From wooden house to mud house and to a new concrete villa, your words of wisdom “Oh son, sharpen your skills with good education, will result in a better future for you, your family and your country”. You have set an example I wish to match someday as a father.

“My Lord, have mercy upon them as they brought me up [when I was] young”

Finally, to the one, and the only one professor Barry Iacopetta
List of Figures

Figure 1.1  The 5 most common cancer types in the Saudi Arabian population (2013) compared to the worldwide incidence ........................................2

Figure 1.2  Crude incidence rate of cancer in the Saudi Arabian population by sex over the last 15 year .................................................................4

Figure 1.3  Overall age-standardized rates (ASR) for CRC in Western Australian (WA) and in Saudi Arabia (SA) in 2013 ..............................................6

Figure 1.4  Age distribution of CRC patients diagnosed in the Western Australian and Saudi Arabian populations in 2013 ..............................................6

Figure 1.5.  Age-standardized rates (ASR, per 100,000) for endometrial cancer in Saudi Arabia, 1997-201 .................................................................14

Figure 3.1  F-SSCP gel electrophoresis showing deletions in the mononucleotide repeat BAT-26 .................................................................35

Figure 3.2  Capillary electrophoresis of pentaplex microsatellite panel ............37

Figure 3.3  Average length and frequency of deletion in 5 mononucleotide microsatellite markers .................................................................38

Figure 3.4  BRAF V600E oncogene mutation screening using pyrosequencing and Sanger sequencing .................................................................39

Figure 3.5  KRAS mutations in codons 12 and 13 revealed by F-SSCP and by pyrosequencing .................................................................40

Figure 3.6  Immunohistochemical staining for mismatch repair proteins ............41
| Figure 4.1 | Flowchart outlining the strategy for identifying patients with Lynch syndrome | 57 |
| Figure 4.2 | Change in the number of cases (A) and incidence (B) of CRC in young (<60 years) Saudi patients over the past two decades | 58 |
| Figure 4.3 | Age distribution of CRC patients from Saudi Arabia (A) and from Western Australia (B) | 59 |
| Figure 5.1 | Flowchart outlining the workflow used to identify potential family clusters of LS-related cancers in the Saudi population | 69 |
| Figure 5.2 | Predicted family cluster sizes with LS-spectrum of cancers amongst Saudi nationals reported to the SCR (2003-2014) | 69 |
| Figure 5.3 | Age distribution of cases identified as putative familial clusters with LS-related cancers reported to the SCR (2003-2014) | 70 |
| Figure 5.4 | Distribution of cancer types amongst young patients belonging to putative family clusters with LS-related cancers | 70 |
| Figure 5.5 | Capillary electrophoresis analysis for MSI using the pentaplex panel of 5 mononucleotide markers | 71 |
| Figure 5.6 | Changes in the incidence of CRC in Saudi Arabia over the past two decades | 72 |
List of Tables

Table 1.1  Amsterdam Clinical Criteria for Lynch Syndrome…………………………15

Table 1.2  Bethesda Guidelines for testing CRC for MSI…………………………17

Table 3.1  Clinicopathological features of MSI tumors in young
            Saudi CRC patients………………………………………………………………36

Table 4.1  Clinical characteristics of 13 young Saudi CRC patients with
            MSI tumors tested for germline variants in MMR genes…………………..53

Table 4.2  Summary of MMR gene variants and their reporting in
            InSiGHT and ClinVar databases ……………………………………………54

Table 4.3  In-Silico functional prediction analysis of variants ……………………..55

Table 5.1  Overall cohort of cancers from the SCR used to identify
            family clusters………………………………………………………………66

Table 5.2  Software prediction variables used to crossmatch related
            patients and to group them into clusters ……………………………….67

Table 5.2  LS-related cancers amongst young patients in the SCR database……68
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACI &amp; II</td>
<td>Amsterdam criteria I and II</td>
</tr>
<tr>
<td>ACMG</td>
<td>American College of Medical Genetics and Genomics</td>
</tr>
<tr>
<td>BRAF</td>
<td>B-type Raf kinase</td>
</tr>
<tr>
<td>CIMP</td>
<td>CpG island methylation</td>
</tr>
<tr>
<td>CRC</td>
<td>Colorectal cancer</td>
</tr>
<tr>
<td>CIN</td>
<td>Chromosomal instability</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>EC</td>
<td>Endometrial cancer</td>
</tr>
<tr>
<td>FAP</td>
<td>Familial adenomatous polyposis</td>
</tr>
<tr>
<td>FCP</td>
<td>Familial cancer program</td>
</tr>
<tr>
<td>GP</td>
<td>General practitioner</td>
</tr>
<tr>
<td>GSWA</td>
<td>Genetic Services of Western Australia</td>
</tr>
<tr>
<td>H&amp;E</td>
<td>Hematoxylin and eosin stain</td>
</tr>
<tr>
<td>HNPCC</td>
<td>Hereditary non-polyposis colorectal cancer</td>
</tr>
<tr>
<td>ICD-O</td>
<td>International Classification of Diseases for Oncology</td>
</tr>
<tr>
<td>IHC</td>
<td>Immunohistochemistry</td>
</tr>
<tr>
<td>LS</td>
<td>Lynch syndrome</td>
</tr>
<tr>
<td>MLH1</td>
<td>MutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)</td>
</tr>
<tr>
<td>MLPA</td>
<td>Multiplex Ligation-dependent Probe Amplification</td>
</tr>
<tr>
<td>MMR</td>
<td>Mismatch repair</td>
</tr>
<tr>
<td>MSH2</td>
<td>MutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli)</td>
</tr>
<tr>
<td>MSH6</td>
<td>MutS homolog 6 (E. coli)</td>
</tr>
<tr>
<td>MSI</td>
<td>Microsatellite instability</td>
</tr>
<tr>
<td>NATA</td>
<td>National Association of Testing Authorities (Australia)</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PMS2</td>
<td>Postmeiotic segregation increased 2 (S. cerevisiae)</td>
</tr>
<tr>
<td>SA</td>
<td>Saudi Arabia</td>
</tr>
<tr>
<td>SCR</td>
<td>Saudi Cancer Registry</td>
</tr>
<tr>
<td>SSCP</td>
<td>Fluorescent-single strand conformation polymorphism</td>
</tr>
</tbody>
</table>
List Of Appendices


Appendix :  Patient information sheet for genetic research-Part I

Appendix :  Patient Informed Consent-Part II

Appendix :  Family history and pedigree MQ

Supplementary Figure 1: Aged distribution of Saudi Arabia population, 2016
# TABLE OF CONTENTS

Declaration........................................................................................................i

Abstract...........................................................................................................vi

Acknowledgements............................................................................................x

List of Figures..................................................................................................xiii

List of Tables...................................................................................................xv

Abbreviations ..................................................................................................xvi

Appendices ......................................................................................................xvii

## 1. Introduction and Literature Review..........................................................1

1.1. Background to the Saudi Arabian population...........................................1

1.2. Cancer incidence in Saudi Arabia............................................................1

1.3. Global Incidence and Mortality from Colorectal Cancer.........................4

1.4. Colorectal Cancer in Saudi Arabia ..........................................................5

1.5. Clinical and Pathological Features of Colorectal Cancer .......................7

1.6. Molecular Pathways and Pathogenesis of Colorectal Cancer ..................7

1.7. Chromosomal Instability and Methylator phenotypes.............................8

1.8. Microsatellite Instability and Defective DNA Mismatch Repair ...............9

1.9. BRAF oncogene mutation in Colorectal Cancer.......................................10

1.10. Familial Colorectal Cancer Syndromes ..................................................11

1.11. Lynch Syndrome and the Risk of Colorectal Cancer............................11

1.12. Lynch Syndrome and the Risk of Endometrial and Ovarian Cancer .......13

1.13. Screening and Detection of Lynch Syndrome: Amsterdam criteria..........14
1.14 Screening and Detection of Lynch Syndrome: Bethesda criteria ..........16
1.15 Cancer Registries for the Identification of Hereditary Cancer Syndrome ........18
1.16 The Saudi Cancer Registry for the Detection of LS ..................................19
1.17 Aims of the Thesis ..................................................................................20

2. Methods ........................................................................................................23
2.1 Patient cohorts ..........................................................................................23
2.2 Ethical use of human tissues for MSI and germline DNA testing ............23
2.3 DNA extraction from formalin-fixed, paraffin-embedded tissues ..........24
2.4 Microsatellite Instability analysis .............................................................25
2.5 \textit{BRAF} and \textit{KRAS} mutation analysis ..............................................25
2.6 Immunohistochemistry ...........................................................................26
2.7 Statistical analysis ....................................................................................26
2.8 Next Generation Sequencing to identify germline mutations ................27
2.9 \textit{In-Silico} method for identification of family clusters with possible LS ......29

3.0 Screening for Lynch Syndrome in Young Colorectal Cancer

Patients from Saudi Arabia Using MSI as the Initial Test .........................30
3.1 Abstract ....................................................................................................30
3.2 Introduction ..............................................................................................31
3.3 Methods ....................................................................................................34
3.4 Results ......................................................................................................34
3.5 Discussion .................................................................................................42
6.0 General Discussion ........................................................................................................75
6.1 Overview..................................................................................................................................75
6.2 Significance of the study ............................................................................................................75
6.3 Lynch Syndrome amongst young (<60 years) Saudi CRC patients.............................76
6.4 Frequency of MSI amongst young Saudi CRC patients .........................................................77
6.5 Amsterdam criteria and revised Bethesda guidelines to identify LS.............................79
6.6 Retrospective MSI screening of young CRC patients to detect LS..............................80
6.7 In-Silico approach to identify LS families using cancer registry data.............................81
6.8 Limitations of this research ....................................................................................................83
6.9 Recommendations for future work .......................................................................................84
6.10 Conclusions ..........................................................................................................................85

7.1 Bibliography ...........................................................................................................................86
1.0 Introduction and Literature Review

1.1. Background to the Saudi Arabian population

Saudi Arabia the largest country in the Middle East with an area of 2.15 million square kilometres. In 2016, The Saudi authority of statistics revealed the total population was about 31.7 million (M), with Saudi nationals accounting for 20M (61%) people distributed in 13 administrative regions. Over 60% of the Saudi national population is located in Riyadh (4.6M), Makkah (4.4M) and the Eastern Region (3.1M). Approximately one quarter of the total population is less than 15 years of age, while 72% are between 15-64 years of age, with only 3.2% older than 65 years (Supplementary Figure 1). It is projected the Saudi population will reach 40.4M in 2050 due to the high birth rate (23.7/1,000). The health care system in Saudi Arabia has shown rapid improvement in the last 10 years, resulting in increased life expectancy to 76.4 years. However, there have been steady increases in some diseases such as diabetes, hypertension, obesity and cancer. This can be attributed to the lack of development and implementation of effective prevention programs for these diseases (Almalki, 2009; MOH, Statistical year book 2016).

1.2. Cancer incidence in Saudi Arabia

Cancer is a major health problem worldwide. In 2012, GLOBOCAN data projected over 14 M newly diagnosed cancer cases worldwide, excluding non-melanoma skin cancer (Torre et al, 2015). The most common cancer types in males are lung (16.7% of total), prostate (15.0%), colorectal (CRC, 10%), stomach (8.5%) and liver (7.5%)
In females, breast cancer predominates (25.2%) followed by colorectal (9.2%), lung (8.7%), cervix (7.9%) and stomach (4.8%). In contrast, the most common cancer types in Saudi males are CRC (13.9%), non-Hodgkin’s lymphoma (NHL, 8.2%), leukemia (7.6%), liver (6.4%) and lung (6.2%). For Saudi females, the most common cancer types are breast (29.1%) followed by thyroid (11%), CRC (10.2%), NHL (4.9%) and corpus uteri (4.5%).

**Fig 1.1** The 5 most common cancer types in the Saudi Arabian population (2013) compared to the worldwide incidence (GLOBOCAN 2012).

The Saudi cancer registry (SCR) reported 15,653 new cancer cases in 2013 amongst all Saudi nationals and non-nationals. Amongst these, 11,645 (74.4%) arose in Saudis, of which 5,281 (45.3%) were in males and 6,364 (54.7%) in females. The crude incidence rate (CIR) of all cancers in the Saudi population was 57.5/100,000, while the overall age-standardized incidence rate (ASR) was 70.4 in
males and 81.4 in females. The highest ASR reported was in the Najran region (145.5/100,000 in females, 135.7 in males) followed by the Eastern region (118.9 in females, 113.7 in males), the Riyadh region (114.4 in females, 106.3 in males) and Tabuk (92.6 in females, 96.7 in males) (Al-Madouj et al, 2013).

The Saudi Arabian population cancer registry shows a steady increase in the Cancer Incident Rate (CIR) and Age-Standardized Rate (ASR) of almost all cancers in both males and females during the past 15 years. The overall CIR and ASR incidence rates amongst Saudis were 37.2 and 62.6 respectively in 2000, and steadily increased to reach 57.1 and 81.3 in 2013. Although the Saudi population increased 19% from 2000 to 2013 (16.9M to 20.1M), the number of cancer cases increased by 54% in the same period. The Saudi Arabian Ministry of Health uses several indicators of health promotion to implement strategic planning for the following years (Statistical year book 2016). Figure 1.2 shows the steady increase for ASR amongst the male and female Saudi population over the last 15 years. This increase is quite alarming considering the population growth rate is about 2.5% annually (Al-Madouj et al, 2013; General authority of statistic, 2016). The median age at cancer diagnosis was 58 years for men and 51 years for females.

Globally, overall cancer related deaths in 2012 exceeded 8 million (GLOBOCAN, 2012). The cancer mortality rate in Western societies is relatively low compared to their high ASR. Jemal et al (2011) reported the ASR in developed countries was almost twice that of developing countries, however the mortality rate for all cancers was only 20% higher in males and 2% higher in females.
Fig 1.2 Crude incidence rate of cancer in the Saudi Arabian population by sex over the last 15 years. A steady increase can be seen in the crude incidence rate for all cancers among Saudi nationals. (From SCR reports 2000-2014)

1.3. Global Incidence and Mortality from Colorectal Cancer

CRC is a major global health problem and is ranked the third most common cancer type in men and the second in women. GLOBOCAN estimated 1.36M new cases of CRC in 2012 and 670,000 deaths worldwide from this disease (Ferlay, 2012). Australia, New Zealand, Europe and North America reported the highest CRC incidence rates (35.3-45.7 per 100,000), whereas the lowest incidence rates are reported in Africa and South Central Asia, with ASR’s of between 4.3-7.0 per 100,000 (Jemal, 2011). The mortality rate from CRC varies considerably across the world, but is significantly lower in North America, Australia and Europe. This is thought to reflect a combination of factors, including colorectal screening leading to earlier diagnosis and better treatments (Center, 2009; Siegel, 2011).
1.4. **Colorectal Cancer in Saudi Arabia**

CRC is a major health problem in Saudi Arabia and ranks as the most common cancer type in males (736 new cases and 14.5% of all cancers in 2013) and the second-most common in females (651 new cases, 10.2% of all cancers). The overall ASR was 11.7 for males and 10.1 for females (Al-Madouj, 2013). The Eastern region reported the highest ASR of 17.6 followed by Riyadh (17.1), Asir (11.4), Tabuk (10.5) and Madinah (10.3) (SCR 2013). Improved screening, surveillance and treatment have decreased the mortality rate from CRC, however the prevalence of CRC has nevertheless increased over the past 50 years (Saeed *et al.*, 2013). Epidemiological studies of CRC have mostly been conducted in developed nations and provide information about the incidence, mortality, pathology, diagnosis and management of the disease in these countries. In contrast, relatively few epidemiological studies have so far been conducted in the developing world (Al-Jaberi *et al.*, 1996). The overall incidence of CRC in Saudi Arabia is about 4-fold lower compared to Western Australia (Figure 1.3). However, a striking difference is the earlier onset of this disease in the Arab world. For example, in Saudi Arabia 26% of CRC patients are aged <50 years at diagnosis (Fig 1.4) compared to only 13% in Western Australia (Threlfall and Thompson, 2009)
Fig 1.3 Overall age-standardized rates (ASR) for CRC in Western Australian (WA) and in Saudi Arabia (SA) in 2013 for males and females (Threlfall and Thompson, 2013; Al-Madouj, 2013)

Fig 1.4 Age distribution of CRC patients diagnosed in the West Australian (WA) and Saudi Arabian (SA) populations in 2013 (Threlfall and Thompson, 2013; Al-Madouj, 2013)
1.5. Clinical and Pathological Features of Colorectal Cancer

Most CRC cases are asymptomatic and can evolve for years before developing a malignant phenotype. The most common symptoms include rectal bleeding together with weight loss, changes in bowel habit, abdominal pain and iron deficiency anemia. These manifestations may vary according to the age at onset (Del Giudice, 2014). Most CRC are diagnosed with tissue biopsy, while the clinical staging is determined from the surgically resected specimen. Most colorectal tumor types (>90%) are adenocarcinoma, which are subsequently classified as mucinous carcinomas if more than 50% of the tumor consists histologically of mucin. Adenocarcinomas are characterised by glands that differ in size and shape (Numata et al, 2012; Secco, 1994).

1.6. Molecular Pathways and Pathogenesis of Colorectal Cancer

CRC carcinogenesis is considered to be one of the most understood pathways in terms of cancer genetics. CRC evolves through an accumulation of genetic and epigenetic alterations that subsequently lead to the transformation of normal colonic epithelial cells to colon adenocarcinoma. In 1990, Fearon and Vogelstein proposed a model of the molecular pathway for CRC carcinogenesis whereby the tumor develops following activation of oncogenes coupled with inactivation of tumor suppressor genes. In their normal state, tumor suppressor genes encode proteins that negatively regulate cell growth and proliferation. Therefore, inactivating mutations to this class of gene can trigger carcinogenesis (Bardhan & Liu, 2013). Development and progression of malignancies require pathogenic mutations in 4 to
5 critical tumor suppressor genes and oncogenes, with the overall accumulation of these genetic changes responsible for the tumor phenotype properties. Our understanding of the molecular mechanisms involved in CRC has improved tremendously over the past two decades. The three main pathways for CRC development and progression are now recognized as chromosomal instability (CIN), CpG island methylation (CIMP) and microsatellite instability (MSI) (Fearon, 1990).

1.7 Chromosomal Instability and Methylator phenotypes

CIN tumors account for approximately 70% of all CRC and may be caused by incomplete chromosome segregation due to a mutation at the spindle assembly checkpoint, abnormality of centrosome number and function, or improper function of the telomerase and/or DNA damage response (Pino, 2010; Ku et al, 2012). The characteristic feature of CIN tumors is loss of heterozygosity of parental alleles, indicating a loss of chromosomal genetic material including tumor suppressor genes. CIMP tumors account for approximately 20% of all CRC and are characterized by frequent hypermethylation of gene promoter regions. These tumors are believed to originate from hyperplastic polyps and show frequent mutation of the BRAF oncogene. Ahuja et al (1997) reported that MSI CRC are often linked to CIMP due to hypermethylation of the MutL homolog 1 (hMLH1) gene. It has been suggested there are five molecular subtypes, each with unique clinicopathological and molecular characteristics (Jass, 2007):

1. CIMP-high, hMLH1 methylated, v-Raf murine sarcoma viral oncogene homolog B1 (BRAF) mutation, MSI, chromosomally stable, origin in serrated polyps
(approximately 12% of all CRC)

2. CIMP-high, incomplete methylation of $hMLH1$, microsatellite stable (MSS), $BRAF$ mutation, chromosomally stable, origin in serrated polyps (8%)

3. CIMP-low, MSS, $KRAS$ mutation, chromosomal instability, origin in either adenomas or serrated polyps (20%)

4. CIMP-negative, MSS, chromosomal instability, origin in adenomas, sporadic or associated with a polyposis syndrome (FAP or MUTYH) (57%)

5. Lynch syndrome, CIMP-negative, $BRAF$ wildtype, chromosomally stable, MSI positive, origin in adenomas (3%)

1.8 Microsatellite Instability and Defective DNA Mismatch Repair

MSI was first reported in CRC by Ionov et al in 1993. These groups reported the presence of thousands of somatic genetic alterations in the length of DNA microsatellite repeats in about 15% of sporadic CRC and in almost all tumors from hereditary non-polyposis colorectal cancer patients, now referred to as Lynch syndrome (LS). Many studies that used dinucleotide microsatellite repeat markers to evaluate MSI showed inconsistent findings, which led to confusion regarding the clinico-pathological features of MSI CRC. On the other hand, the use of mononucleotide microsatellite markers to evaluate MSI CRC revealed consistent clinico-pathological findings. Sporadic MSI CRC arise almost exclusively in the proximal colon and are more frequent in elderly women. MSI is due to a defective DNA mismatch repair (MMR) system following inactivation of one of the 4 major DNA MMR genes ($MLH1$, $MSH2$, $MSH6$ or $PMS2$). Inactivation of DNA MMR
leading to MSI can arise through methylation-induced silencing of MLH1 expression, somatic mutation of one of the MMR genes, or more rarely through germline variants in one of these genes in the case of LS. In recent studies, it was revealed that mutation in the *EPCAM* gene lead to epigenetic silencing of *MSH2* which subsequently lead to LS. Thus identifying carrier with germline mutation in *EPCAM* gene can be included in the LS screening and prevention program (Kempers *et al.*, 2011). MSI screening and identifying of EPCAM mutation carriers can thus be used as a very sensitive marker to screen CRC patients for the presence of LS, although it is not a specific marker because MSI can also occur in sporadic cases following *MLH1* methylation or somatic mutations. (Murphy *et al.*, 2006; Suraweera *et al.*, 2002)

### 1.9 *BRAF* oncogene mutation in CRC

Cellular homeostasis in CRC is maintained through a complex network of genes and signalling pathways. Over the years, significant advances in molecular genetic testing have allowed researchers to study phenotypes such as CIN, CIMP, MSI and others as predictors of tumour progression and treatment efficacy (Fearon, 1990; Jass, 2000). One such candidate predictive marker is the *BRAF* oncogene that belongs to the Ras-MAPK pathway. The *BRAF* gene codes for a serine/threonine kinase (Yoon, 2006) and is mutated in approximately 7% of human tumors, usually through point mutation leading to a single amino acid substitution. Over 90% of *BRAF* mutations are V600E, where glutamic acid is substituted for valine at position 600 (Wajapeyee, 2008). In Lynch syndrome, the presence of a *BRAF* mutation can be used to exclude sporadic tumors from further follow up as possible familial cases. The frequency of *BRAF* mutation reported in a consecutive cohort of 757 CRC
patients from Saudi Arabia was estimated at 2.5% (Siraj et al, 2014), which is lower than the 5-10% frequency reported in Western countries (Samowitz et al, 2005; Tejpar et al, 2010). Since the patient cohorts in these studies were not selected for age, this variable cannot account for the different BRAF mutation frequencies in CRC between Saudi Arabia and western countries.

1.10 Familial Colorectal Cancer Syndromes

Between 1895 and 1925, Dr Warthin in the US investigated many families with cancer and concluded there may be a familial predisposition to this disease. He found the susceptibility was particularly strong for tumors from the gastrointestinal tract and uterus, while also noting an early age of onset of cancer in these families (Boland et al, 2013). He proposed two types of hereditary CRC that could be classified according to their location in the large bowel. Later work revealed that familial CRC in the distal colon arises due to germline mutations in the adenomatous polyposis coli (APC) gene, with this condition referred to as familial adenomatous polyposis (FAP). FAP is easily recognizable by the presence of hundreds of small polyps lining the distal bowel wall. The second type of hereditary CRC is LS, previously referred to as HNPCC. In this condition, CRC arise mostly in the proximal colon and are due to germline mutations in one of the DNA MMR genes described above. CRC from LS patients are histologically indistinguishable from sporadic CRC, but are characterized by always having MSI. Approximately 15% of all CRC demonstrate some familial clustering, while 2-7% show clinical characteristics that are suggestive of LS (Lynch et al, 2004; Halvarsson et al, 2004). However, the
presence of LS in a patient with an MSI CRC can only be confirmed once a pathogenic germline variants is identified in one of the DNA MMR genes.

1.11  Lynch Syndrome and the Risk of Colorectal Cancer

LS is a dominantly inherited cancer syndrome characterized by a greatly increased lifetime risk of CRC and several other cancer types including endometrial (EC), gastric, ovarian and kidney (Lynch et al, 1997). LS can be recognized clinically by an early age of onset of CRC and the occurrence of multiple related cancers within families. The genes that are mutated in LS patients encode DNA MMR proteins, with the majority of mutations (>90%) occurring in the MLH1 and MSH2 genes (Peltomaki, 2003). The absence of readily defined morphological features in CRC from LS patients prevents the use of histopathological criteria to help diagnose this syndrome. However, all CRC from LS individuals are characterized by having MSI in their tumor DNA. This is almost always accompanied by the loss of expression of one or more of the DNA MMR proteins, usually a combination of MLH1 and PMS2, or of MSH2 and MSH6. Loss of expression of MMR proteins can be readily detected on clinical tumor specimens using standard immunohistochemical (IHC) technique.

LS is associated with a lifetime risk of approximately 80% for the development of CRC, with an average age of disease onset of 43 years for CRC. The lifetime risk of EC for a mutation carrier is 25%-60%, with the age of onset usually 48-62 years. The lifetime risk for gastric cancer is 6%-12% with average age at onset of 56 years, while for ovarian cancer the lifetime risk is 4%-12% with average age at onset of 43 years (Aarnio et al, 1999; Obermair et al, 2010). There
is strong evidence that surveillance and management of individuals with germline mutations in DNA MMR genes (i.e. LS) is cost-effective (Lindgren et al, 2002). Colonoscopies performed at 3-year intervals reduces the mortality from CRC by as much as 70% due to the early detection and removal of premalignant polyps (Jarvinen et al, 2000). Current best practice recommendations for the screening of “at risk” individuals in LS families (i.e. mutation carriers) include full colonoscopy every 1-2 years and beginning at the age of 25 yrs. More extensive surgery is also recommended for mutation carriers affected by CRC because of the increased risk of developing metachronous tumors. (Karimi et al, 2018)

1.12 Lynch Syndrome and the Risk of Endometrial and Ovarian Cancers

EC and ovarian cancer are the most common gynecological cancers in the world with over 319,000 cases reported in 2012 (Ferlay, 2015). In Saudi Arabia, EC was the fourth most common cancer type in women in 2014, accounting for around 5.8% of all cancer cases. Approximately 90% of EC are thought to be sporadic in origin, with the remainder having a possible hereditary cause (Al-Madouj et al, 2014). In women with LS, the risk of developing EC may in some cases be even higher than the risk of developing CRC (Sorosky et al 2012; Lu et al, 2005). Soliman et al (2005) reported that approximately 7% of patients with EC or ovarian cancer had molecular or clinical features that were suggestive of LS. These workers concluded the incidence of LS was low unless there was a strong family history of LS-associated tumors. Several studies have suggested that women with LS are at high risk of developing gynecologic cancers at an early age (Daniels et al, 2013;
Burke, 2014)

The American Cancer Society recommends that women with EC who are at risk of having LS should undergo genetic counselling and their tumors should be tested for the presence of MSI (American cancer society, 2015; Smith et al, 2009).

As with CRC, the incidence of EC has increased at an alarming rate in Saudi Arabia (Figure 1.5), demonstrating the urgency to identify familial cases so that closer surveillance can be offered. Women with LS have a high lifetime risk for EC (20–60%) and for ovarian cancer (4-12%) and are encouraged to have annual endometrial screening from the age of 30 to 35 years (Kohlmann and Adam, 2004).

**Figure 1.5** Age-standardized rates (ASR, per 100,000) for endometrial cancer in Saudi Arabia, 1997-2013. From Al-Eid et al, 2014.

1.13 Screening and Detection of Lynch Syndrome: Amsterdam criteria

The proportion of LS cases amongst all CRC patients may vary from country to country, but has been estimated at 1-5% (Aarnio et al, 1999). This apparent variation could be linked to limited access to germline mutation testing due to the high cost
and technical challenges, and also to differences in the incidence of founder mutations amongst various populations. During the 1990s, The International Collaborative Group on HNPCC proposed clinical criteria to classify and identify patients with LS. The so-called Amsterdam clinical criteria (Table 1.1) were proposed as a guide to help detect LS (Soliman, 1998). Families must have experienced at least three cases of CRC over at least two generations, with at least one case diagnosed before the age of 50 years (3:2:1 criteria). However, the difficulty in obtaining a full family history of cancer and the lack of accessible genetic counselling and testing has meant that many LS families remain undetected in the population, even in developed countries. Subsequently, the Amsterdam-II criteria (AC-II; Table 1.1) were developed which incorporated a spectrum of extra-colonic, LS-related tumors (Vasen et al, 1999). These were subsequently modified again to take into consideration the small size of families who failed to fulfil AC-II. Vasen et al reported the modified AC-II increased the specificity for the detection of LS, however the sensitivity was low.

Table 1.1 Amsterdam Clinical Criteria for Lynch Syndrome

**Amsterdam I**

- At least three cases of CRC in the same family
- One case a first-degree relative to the other two
- At least two successive generations affected
- At least one case diagnosed before the age of 50 years
- Exclusion of FAP

**Amsterdam II**

- There should be at least three relatives with an HNPCC-associated cancer
(CRC, EC, or cancer of the small bowel, ureter or renal pelvis)
• One case of first-degree relative to the other two
• At least two successive generations affected
• At least one case diagnosed before the age of 50 years
• FAP should be excluded in the CRC cases
• Tumors should be verified by pathological examination

1.14 Screening and Detection of Lynch Syndrome: Bethesda criteria

The National Cancer Institute organised an international conference in 1996 to elucidate the role of MSI testing to detect LS in patients who failed to meet Amsterdam criteria. The so-called Bethesda guidelines (Table 1.2) were proposed in order to increase the sensitivity for detection of families with LS (Boland et al., 1998; Umar et al., 2004). These included testing tumor DNA for MSI in suspected cases of LS using a standard panel of microsatellite markers (Iacopetta et al., 2010). The Bethesda guidelines recommended a standard reference panel for the evaluation of MSI that contained five microsatellite markers comprised of two mononucleotide repeats (BAT25 and BAT26) and three dinucleotide repeats (D2S123, D17S250 and D5S346) (Rodriguez, 1997; Boland, 1998). At the same congress, the Bethesda guidelines defined tumors as positive for MSI if these showed instability in two or more of the 5 panel markers (Boland, 1998). Patients with MSI tumors were recommended to undergo germline mutation testing for LS. These criteria were subsequently revised and improved in order to enhance the sensitivity for detection of families with LS (Table 1.2) (Umar et al., 2004).

The sensitivity of the revised Bethesda guidelines was found to be higher than the Amsterdam criteria for the identification of mutation carriers, albeit with lower specificity. About 20% of CRC cases meet the revised Bethesda criteria, in
which both MSI screening and loss of MMR protein expression by immunohistochemistry (IHC) is recommended (Pinol et al, 2005). Unfortunately, the proportion of cases with MSI or loss of MMR protein expression that subsequently attend familial cancer clinics is very low in the absence of good follow up (Wong et al, 2008). There are many challenges involved in implementing the Amsterdam and Bethesda guidelines for the detection of LS, resulting in failure to identify many pathogenic germline mutation carriers amongst the affected population (Lynch et al, 2004; Terdiman, 2005). Some of these challenges include the low number of patients that attend familial cancer clinics due to the lack of referral processes in remote hospitals, the lack of updated patient demographic data and the lack of familial clinics in some hospitals. Additionally, some patients have since died and others refuse to attend the genetic counselling clinic.

**Table 1.2 Bethesda Guidelines for testing CRC for microsatellite instability (MSI)**

**Original Guidelines**

- Patients with cancer in families that meet Amsterdam criteria
- Patients with two Lynch Syndrome related cancers *, including synchronous and metachronous CRCs or associated extra-colonic cancers
- Patients with CRC and a first-degree relative with CRC and/or other Lynch Syndrome spectrum of cancer and/or a colorectal adenoma, with one of the cancers diagnosed before 45 years of age, and the adenoma diagnosed before 40 years of age
- Patients with CRC or EC diagnosed before 45 years of age
- Patients with right-sided CRC with an undifferentiated pattern (solid/cribriform) on histopathology diagnosed before the age of 45
- Patients with signet-ring type CRC diagnosed before 45 years of age
- Patients with adenomas diagnosed before the 40 years of age

**Revised Bethesda criteria**
• Patients who were diagnosed with CRC below 50 years of age
• Presence of synchronous, metachronous colorectal, or other Lynch Syndrome related cancers, regardless of age
• CRC with the MSI-H histology** diagnosed in a patient who are less than 60 years
• CRC diagnosed with one or more first-degree relatives with an Lynch Syndrome related cancers, with one of the cancers being diagnosed before 50 years of age
• CRC diagnosed in two or more first or second degree relatives with HNPCC related tumor, regardless of age

* Lynch Syndrome spectrum of cancer include: colorectal, endometrial, stomach, ovarian, pancreas, ureter and renal pelvis, biliary tract, and brain

** MSI-H refers to deletion in two or more of the five markers approved by the NCI

1.15 Cancer Registries for the Identification of Hereditary Cancer Syndromes

Identification of hereditary cancer syndromes has been the subject of extensive research since the discovery of tumor susceptibility genes (Lichtenstein et al, 2000). Large population studies rely on published guidelines such as the Amsterdam and Bethesda guidelines to identify LS. However, many LS cases are likely to go undetected due to the stringency of those guidelines. Widespread screening for LS based on reflex MSI testing in young CRC patients has so far been implemented in only a few geographic regions, including the state of Western Australia (Schofield, 2009), although results from several hospital-based screening programs have also been reported (Samowitz et al, 2001; Piñol et al, 2005; Ward et al, 2013).

Rapid advances in information technology have helped researchers to perform data gathering, analysis and reporting. The establishment of nationwide,
electronic cancer registries has for example assisted government health authorities to plan and establish preventive measures aimed at reducing disease incidence. Many countries have now established a flow of continuing, authentic and robust clinical data for cancer incidence and mortality in order to monitor the impact of this disease on their population. The oldest model of the modern cancer registry was started in Germany in the 1920s and gathered not only clinical data of patients with cancer, but was also used to estimate the economic and public health impacts (Wagner, 1991). In the 1940’s the states of New York and Connecticut as well as Denmark, Canada and Britain established population-based cancer registries. In 1950 the World Health Organization founded the International Association of Cancer Registries (IACR) which is responsible for standardizing the cancer registry information and ensuring the quality of clinical and histopathological data. The International Classification of Diseases for Oncology (ICD-O) subsequently published the first report on cancer classification in 1976 and generated a topographical code that describes the tumor site and a morphological code that describes the histology of the tumor (ICD-O online, 2013).

1.16 The Saudi Cancer Registry for the Detection of LS

The Saudi Cancer Registry (SCR) is a population-based registry administered under the jurisdiction of the Ministry of Health and which began mandatory reporting of cancer cases in January 1994. The stored data includes the patient’s personal and demographic information, as well as the tumor details including primary site and histology coded according to the ICD-O. The patient’s name is recorded in the SCR
using the patient’s first, second, third, fourth and sometimes the fifth name (family name). In addition, the referring hospital data for each patient is reported as raw data, but is not published in the SCR reports for patient privacy reasons. Routine MSI screening of CRC for the detection of LS is not carried out by any hospital in Saudi Arabia or the Middle East and hence detection of LS is still reliant on clinical criteria such as a strong family history of cancer.

The SCR reported over 1,300 CRC cases in 2014, of which 52% were diagnosed in patients aged <60 years. This compares to less than 25% of CRC cases diagnosed in patients aged <60 years in Western Australia (Threlfall, 2014). It is not known whether the younger age of onset of CRC in the Saudi population is due to a higher incidence of LS or other familial cancer syndromes relative to Western populations, possibly also linked to the high incidence of consanguineous marriage. Approximately 51% of marriages in the Riyadh region were reported to be consanguineous i.e. first cousin marriages (al Husain M, 1997). To our knowledge, data contained within the SCR, or any other national cancer registry, has yet to be used as a starting point to identify families with hereditary cancer syndromes such as LS.

Although the incidence of cancer in Saudi Arabia is still relatively low compared to Western countries, it has risen significantly over the last two decades. The implementation of screening tests could significantly reduce the morbidity and mortality associated with hereditary cancers such as LS, through increased surveillance of affected family members and prophylactic surgery (Schofield et al, 2014).
1.17 Aims of this Thesis

Saudi Arabia comprises 13 geographic regions covering 2 million square kilometres. The recent demographic survey in 2016 estimated the population of Saudi nationals had reached 20 million, of which 96.8% were aged <64 years. The relatively early age of onset of CRC and the high incidence of consanguineous marriage indicate that hereditary cancer syndromes such as LS may be important.

There is currently a lack of studies on LS amongst CRC patients in Saudi Arabia, thus creating a challenge to the medical profession in terms of establishing preventive measures for these families in the Saudi population. The overarching goal of this thesis was to identify strategies aimed at identifying individuals and families with LS in the Saudi population. The thesis is divided into three aims, with each being addressed in a corresponding Results chapter.

Aim I

To identify potential LS cases (“red flag” cases) amongst young Saudi CRC patients by using MSI as the initial screening test.

For this study, tumors from 284 young (<60 years) CRC patients from different referral hospitals in the Eastern region were screened retrospectively for MSI (Chapter 3).

Aim II

To follow-up the “red flag” cases identified previously for genetic counselling and for
germline mutation testing in order to confirm the presence of LS.

A total of 33 red flag CRC cases (MSI, <60 years) identified in Chapter 3 were followed up, and germline testing for MMR gene variants was carried out in 13. Of these, 8 (62%) were found to carry a pathogenic germline variants (Chapter 4).

Aim III
To trial a novel strategy for the identification of LS families that uses the Saudi Cancer Registry to identify high-risk candidates amongst cancer patients diagnosed with CRC or an LS-related cancer type at an early age.

The SCR was used to identify 140 potential family “clusters” for LS containing two of more first-degree family members diagnosed with CRC or another LS-related cancer aged <55 years. In a pilot study, tumors from 8 of these family clusters were tested for MSI and 2 were found to be positive, thus strongly indicating the presence of LS (Chapter 5).
2.0 Methods

2.1 Patient cohorts

All patients with primary CRC diagnosed at three hospitals in Dammam (n=191; King Fahad Specialist Hospital, King Fahad University Hospital, Dammam Regional Laboratory) and the King Khaled University Hospital in Riyadh (n=93) between 2006 and 2015 were eligible. Because the large majority of LS patients are diagnosed with CRC at a young age (Lynch et al, 2003), only patients aged <60 years were included. No information was available regarding the family history of cancer. Clinicopathological information including gender, age, stage, tumor site and histological grade were obtained from pathology records.

2.2 Ethical use of human tissues for MSI screening and germline DNA testing

The MSI screening study (Chapter 3) was approved by the human research ethics committees of four tertiary health institutes: King Fahad Specialist Hospital-Dammam (IRB LAB 055), King Khaled University Hospital (15/0148/IRB), King Fahad University Hospital (IRB-2014-01-297) and the Dammam Regional Laboratory (approval date 18/08/2014). Ethics approval for patient follow up and germline testing of MSI cases (Chapter 4) was obtained from three referral hospitals: King Khaled University Hospital (IRB no. E-15-1468), King Fahad Specialist Hospital-Dammam (IRB no. LAB0305) and King Fahad University Hospital (IRB no. 2014-01-297). Ethics approval was also received from the University of Western Australia Human Research Ethics Committee (RA/4/1/7322). Patients who provided
consent gave 10 ml of peripheral blood for germline DNA analysis, as well as information on family history of cancer to their treating physician.

Patient privacy was maintained throughout this project by assigning each subject with an anonymous code. All tissue samples used in this study were given a unique code that could be identified only by the principle investigator and by the delegated genetic counsellor or treating oncologist. Project data was logged into password-protected, secure Excel and SSPS data sheets. Patient samples were anonymously coded and all testing and analysis was performed blinded to the patient identification. Results were sent to the genetic counsellor using the same anonymous code. Ethics issues regarding patient contact were discussed in an earlier publication (Zeps et al, 2007).

2.3 DNA extraction from formalin-fixed and paraffin-embedded archival tissues

For each case, sections of 10μm thickness (for DNA extraction) or 4μm thickness (for IHC) were cut from archival, paraffin-embedded tumour and normal tissue blocks obtained from the surgically resected specimen. Each block was verified by a pathologist for maximal tumor cell content (>50%) by examination of H&E stained slides. DNA extraction from paraffin-embedded tissue sections was performed using a kit and automated DNA extractor as described by the manufacturer (MagNA Pure Compact, Roche, USA). DNA purity and concentration were evaluated by NanoDrop spectrophotometer and adjusted to 50 ng/μl for each sample as stock concentration.

2.4 Microsatellite Instability analysis

Initial MSI screening was performed using PCR and fluorescent-single strand
conformation polymorphism (F-SSCP) analysis to detect deletions in the BAT-26 mononucleotide repeat (Iacopetta et al, 2000). All positive cases were confirmed using the commercially available pentaplex MSI analysis system as described by the manufacturer (Promega, Australia). This assay contains five mononucleotide repeat markers (BAT-25, BAT-26, NR-21, NR-24 and MONO-27) that are highly susceptible to somatic deletions in tumors with defective MMR (Suraweera et al, 2002), as well as two pentanucleotide markers used to ensure correct sample matching. PCR products were run on an ABI 3500 genetic analyzer and the length of allelic deletions was quantified using Genemapper software (ABI, California, USA). Deletion lengths and frequencies were compared to those of MSI cases detected in young CRC patients in WA using the pentaplex system during routine evaluation.

2.5 **BRAF and KRAS mutation analysis**

Samples were screened for mutations in codon 600 of the *BRAF* oncogene using competitive allele-specific Taqman (CAST)-PCR (Life Technologies, USA) as described previously (Richter et al, 2013). Data was collected during 40 cycles of amplification and analysed using the Mutation DetectorTM software v.2.0 (Life Technologies, USA). Samples with a Δ(d)Ct of less than 9.96 were considered positive for mutation, where ΔCt= Ct mut – Ct ref.

Mutations in codon 600 of *BRAF* were then confirmed using the Therascreen® *BRAF* Pyrosequencing Kit (Qiagen, Australia) according to the manufacturer’s instructions. The *KRAS* Pyrosequencing kit (Qiagen, Australia) was also used to
detect mutations in codons 12, 13 and 61 of the \textit{KRAS} gene. Pyrosequencing was performed on the PyroMark Q24 platform (Qiagen) using Therascreen buffers and reagents (v1). Readouts were generated with the PyroMark Q24 software (v. 2.0.6.) and data was analyzed manually or with a plug-in tool provided by Qiagen.

2.6 \textbf{Immunohistochemistry}

Tumor samples that were found to be positive for MSI and wildtype for \textit{BRAF} (thus candidates for LS) were investigated for loss of expression of the MLH1, PMS2, MSH2 and MSH6 proteins using IHC as described previously (Schofield \textit{et al}, 2009). Briefly, 4\textmu m tissue sections were cut serially from the same tumor blocks used to provide tissue samples for DNA extraction. IHC for MLH1, PMS2, MSH2 and MSH6 expression was performed using commercially available antibodies (clones M1, EPR3947, G219-1129 and 44, respectively) at the recommended dilutions (Ventana, Australia). Normal colonic epithelium located adjacent to tumor cells served as the internal positive control for MMR protein expression. Tumors were scored as showing normal expression, partial loss of expression or complete loss of expression.

2.7 \textbf{Statistical analysis}

Results for the length of deletion of mononucleotide repeats were analyzed by SPSS. Comparison of deletions between MSI cases from SA and WA were performed using the Student's t-test. Associations between the presence of MSI and clinicopathological features were evaluated using Fisher's exact test. Significance
was assumed at $P<0.05$.

### 2.8 Next Generation Sequencing to Identify Germline variants

DNA was extracted from peripheral blood using an automated DNA extractor and kit as per the manufacturer's instructions (Cat# 03730972001, MagNA Pure Compact, Roche, USA). Approximately 50 ng of DNA was processed using the Illumina Trusight Cancer Panel (Illumina Inc. San Diego, CA, 92122, USA) and sequenced on the MiSeq system (Illumina) according to the manufacturer's instructions. This protocol provides comprehensive sequence coverage of 94 cancer-related genes, including the DNA mismatch repair genes MLH1, MSH2, MSH6 and PMS2. DNA was fragmented and tagged, followed by the application of sequencing adaptors and indices by PCR. Sample libraries were denatured with subsequent hybridization to labeled probes specific to the targeted region. Several rounds of hybridization and enrichment of these DNA fragments were performed prior to PCR amplification. The products were then pooled and loaded onto the MiSeq (Illumina) for cluster generation and 2 x 150-bp paired-end sequencing.

Data obtained from the MiSeq was analysed via BaseSpace (https://basespace.illumina.com) to generate BAM and VCF files, allowing alignment against the Burrows-Wheeler Aligner 1 (BWA) and variant calling by GATK. The depth of coverage was assessed by the coverageBed tool (v2.14.2). Regions with a depth of <20 reads were detected by the Cancer_coverage_v2.py script and noted for gap filling by Sanger sequencing. VCF files were imported into Cartegenia (http://www.agilent.com) for variant annotation and cross-matching of
variants against the dbSNP, Human Gene Mutation Database (HGMD 2014.2), COSMIC (v71) and NCB1 ClinVar databases, in addition to the population databases ESP6500, 1000 Genomes and ExAC. Further interrogation, including BAM file assessment, was performed using Alamut Visual (v2.7.2), which also provided functional prediction analysis by Align GVGD, SIFT, MutationTaster and PolyPhen2. In addition, variants were cross-referenced to the International Society for Gastrointestinal Hereditary Tumors Incorporated (InSiGHT) database. All reported variants were confirmed by Sanger sequencing and classified according to the current ACMG guidelines.

Multiplex Ligation-dependent Probe Amplification (MLPA) of MLH1, MSH2, MSH6 and PMS2 was performed according to the manufacturer’s instructions (MRC-Holland, Amsterdam). MLPA was used due to its ability to analyze up to 50 DNA sequences in a single reaction and to detect copy number variations in specific genes such as MMR, including small intragenic rearrangements. This was targeted to genes that demonstrated loss of MMR gene expression, as detected by IHC staining of tumor tissues carried out previously. DNA was denatured and subsequently hybridized to the MLPA probes followed by a ligation reaction. PCR was performed and the amplified products were separated by capillary electrophoresis. Data was analysed using the Coffalyser.Net software (v.140721.1958). All reported variants were confirmed by repeat MLPA using a confirmation probe-mix and classified according to the current ACMG guidelines (Richards et al, 2015).

Sanger sequencing using a long-range PCR method was also performed for
patients showing loss of IHC staining for PMS2 in their tumor tissue. This method avoids amplification of the pseudogene (Vaughn et al, 2010). Briefly, PMS2 was amplified in 3 segments by long range PCR using the TaKaRa enzyme, interspaced by exons 6 and 10, which were then amplified by usual PCR methods. These segment products were subsequently used for nested PCR and sequencing of each individual exon (van der Klift et al, 2016)

2.9  *In-Silico* method for Identification of Family Clusters with Possible LS

*In-Silico* is a prediction strategy that uses software to identify familial cancer clusters that may be due to LS (Chapter 5). It uses detailed data contained within the Saudi Cancer Registry (SCR) to identify families at-risk of LS from amongst colorectal, endometrial, small intestine and renal cancer patients. Essentially, candidates for LS are identified using the SCR, archival tumor tissue from these candidates is tested for MSI, and positive cases are then followed up for counselling and germline variants analysis as described above in order to confirm the presence of LS. Ethics approval to access raw data in the SCR was obtained from the SCR Scientific Committee followed by letters of support from the five largest referral hospitals in Riyadh and the Eastern Province of SA (King Fahad Medical City, King Faisal Specialist Hospital and research Centre, King Saud Medical city, Prince Sultan Military Medical City, King Fahad Specialist Hospital). A full description of the *in-silico* approach used to identify candidate family clusters with cancers suggestive of LS is given in Chapter 5.
3.0 Screening for LS in Young Colorectal Cancer Patients from Saudi Arabia Using Microsatellite Instability as the Initial Test

3.1 Abstract

**Background:** Lynch Syndrome (LS) is a familial cancer condition caused by germline mutations in DNA mismatch repair genes. Individuals with LS have a greatly increased risk of developing colorectal cancer (CRC) and it is therefore important to identify mutation carriers so they can undergo regular surveillance. Tumor DNA from LS patients characteristically shows microsatellite instability (MSI). The aim of work described in this chapter was to screen young CRC patients for MSI as a first step in the identification of unrecognized cases of LS in the Saudi population. **Materials and Methods:** Archival tumor tissue was obtained from 284 CRC patients treated at 4 institutes in Dammam and Riyadh between 2006 and 2015 and aged less than 60 years at diagnosis. MSI screening was performed using the BAT-26 microsatellite marker and positive cases confirmed using the pentaplex MSI analysis system. Positive cases were screened for *BRAF* oncogene mutations to exclude sporadic CRC and were also evaluated for loss of expression of four DNA mismatch repair proteins using immunohistochemistry. **Results:** MSI was found in 33/284 (11.6%) cases, of which only one showed *BRAF* mutation. Saudi MSI cases showed similar instability in the BAT-26 and BAT-25 markers to Australian MSI cases, but significantly lower frequencies of instability in 3 other microsatellite markers. **Conclusions:** MSI screening of young Saudi CRC patients reveals that
approximately 1 in 9 are candidates for LS. Patients with MSI are strongly recommended to undergo genetic counselling and germline mutation testing for LS. Other affected family members can then be identified and offered regular surveillance for early detection of LS-associated cancers.

3.2 Introduction

Hereditary nonpolyposis colorectal cancer (HNPCC), more commonly known as Lynch syndrome (LS), is an autosomal dominant genetic condition associated with a high risk of colorectal cancer (CRC) (Lynch et al., 2003). Approximately 1-2% of all CRC are thought to be due to LS. In addition to CRC, LS is also associated with increased risk for endometrial, gastric, ovarian and other rarer cancer types. LS is caused by germline mutations in DNA mismatch repair (MMR) genes, with the most frequent being MLH1, MSH2, MSH6, and PMS2. The incidence of MMR gene mutation carriers is about 1 in 500 in Western populations. Identification of these individuals is crucial because it allows them to undergo early and regular surveillance for cancer. It also allows their extended family to be screened for additional mutation carriers. Regular screening by colonoscopy has been shown to reduce mortality from CRC in mutation carriers (Jarvinen et al., 2000).

Until recently, CRC patients who were suspected of being at risk for LS were identified through the use of clinical criteria which rely on obtaining a detailed family history of cancer, as outlined in the Amsterdam and Bethesda guidelines (Umar et al., 2004). However, these guidelines have been associated with low sensitivity for the detection of LS and their implementation in routine clinical practice has been
poor (Lynch et al., 2004). Consequently, there have been calls to introduce laboratory-based screening tests for LS that do not require the clinician to obtain a detailed family history of cancer (Terdiman, 2005).

Tumors that arise in patients with LS have a defective DNA MMR system, resulting in ubiquitous small deletions in DNA repeat regions that are referred to as microsatellite instability (MSI) (Iacopetta et al., 2010). In addition to MSI, these tumors almost always show loss of expression of one or more MMR proteins as seen with immunohistochemistry (IHC) (Marcus et al., 1999). MSI and IHC tests are therefore essential for the selection of CRC patients to undergo further germline mutation testing for LS. However, a positive MSI test and loss of MMR expression by IHC are also observed in approximately 10% of sporadic CRC, meaning they are not specific markers for the presence of LS. Fortuitously, sporadic MSI CRC cases often display a hotspot mutation (V600E) in the \textit{BRAF} oncogene, whereas MSI CRC from LS patients never show mutations in this oncogene. The presence of \textit{BRAF} mutation can therefore be used to identify MSI CRC cases that are sporadic in origin and can thus be excluded from further testing for possible germline mutations (Domingo et al., 2004).

The results of MSI testing in young CRC patients as a first screen to detect LS cases in the state of Western Australia (WA) were previously reported (Schofield et al., 2009). This laboratory-based screening program was conducted in the absence of any information on the family history of cancer. It was established that initial screening for MSI, followed by testing for \textit{BRAF} mutation in positive cases, was an effective strategy to identify LS mutation carriers in the WA population.
Routine MSI and IHC testing was subsequently implemented for all CRC patients aged <60 years in WA starting in 2008. The rate of diagnosed LS cases in this state (population 2.5 million) has since increased from 2-3 per year prior to routine MSI screening to an average of 8 cases per year over the past 7 years (Schofield et al., 2014; Threlfall, 2014).

In Saudi Arabia (SA), CRC is the most frequent cancer type in males (13% of all cancer cases) and the second most common cancer type in females (9%), with a total of almost 1,200 cases reported in 2011 (Al-Madouj et al., 2011). Interestingly, the mean age at diagnosis (55-58 years) is approximately 12-15 years younger compared to Western populations (Aljebreen, 2007; Mosli et al., 2012; Amin et al., 2012). Compared to WA, the age-standardized rate for CRC is about 3-4-fold lower in SA, although the incidence appears to be increasing quite rapidly (Ibrahim et al., 2008), probably due to the adoption of a more Western diet. Little is known however about the incidence of LS in the Saudi population and there are currently no reflex MSI- or IHC-based screening programs to help detect LS in the routine clinical setting. A recent publication involving 807 CRC cases from Riyadh reported an MSI frequency of 11.3% and LS frequency of 0.9% (Siraj et al., 2015). The LS frequency is similar to previous results from large Australian (Schofield et al., 2009; Ward et al., 2013), Spanish (Pinol et al., 2005) and American (Samowitz et al., 2001) studies, but lower than estimates of 5.1% (Zeinalian et al., 2015) and 2% (Nemati et al., 2011) from Iranian studies. It should be noted however that the latter studies were based on clinical findings (Amsterdam criteria) rather than genetic findings.

In the work described in this chapter, Saudi Arabian CRC cases from the
Eastern Province and from Riyadh were retrospectively screened for MSI and \textit{BRAF} mutation as a first step in the identification of LS cases. The results were compared with those of a similar study conducted in the WA population (Schofield \textit{et al}, 2009), which has a 3-fold higher incidence of CRC and an older mean patient age compared to the Saudi population.

3.3 Methods

A total of 284 tumors from young (<60 yrs) CRC patients from Saudi Arabia were screened for MSI and \textit{BRAF} mutation as described in Chapter 2 (Methods).

3.4 Results

All 284 cases were successfully tested for MSI using PCR and F-SSCP analysis for deletions in the BAT-26 marker. Representative results obtained with this screening technique are shown in Figure 3.1 below. Thirty three cases (11.6\%) showed deletions in BAT-26. These were all confirmed as having MSI by using the commercial and standardized pentaplex system (Promega, Australia) that analyzes 5 mononucleotide repeats (Figure 3.2).

Associations between the presence of MSI and clinicopathological features are shown in Table 3.1. Younger and male patients showed a trend for higher MSI frequency. Right-sided tumors also showed a higher frequency of MSI compared to left-sided tumors, however this did not reach statistical significance ($P=0.13$).
Figure 3.1 F-SSCP gel electrophoresis showing deletions in the mononucleotide repeat BAT-26 (arrows) in two tumor samples from young Saudi CRC patients (cases 3 and 6), indicating the presence of MSI.

The length and frequency of allelic deletions was quantified for each of the 5 mononucleotide repeats evaluated in the pentaplex assay. This analysis was performed in 33 MSI cases from SA and in 56 age-matched MSI cases from WA (Figure 3.3). BAT-26 was deleted in all cases in both cohorts, with a similar average length of deletion (9.1bp in WA and 9.7bp in SA). Interestingly, the average length of deletion for the NR-21, NR-24 and MONO-27 markers was significantly shorter in MSI tumors from SA (Figure 3.3 A). This was due to the lower frequency of deletion of these 3 markers in MSI tumors from SA (Figure 3.3 B). Nine of the 33 MSI tumors from SA (27.3%) showed deletions only in the BAT-26 and BAT-25 markers. As shown in Fig. 3.3A, the size of deletions were largest for BAT-26, even in the Saudi CRC cases, making it highly likely this was the most sensitive marker.
### Table 3.1 Clinicopathological features of MSI tumors in young Saudi CRC patients

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>%</th>
<th>MSI</th>
<th>%</th>
<th>MSS</th>
<th>%</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total no. of cases</td>
<td>284</td>
<td></td>
<td>33</td>
<td>11.6</td>
<td>251</td>
<td>88.4</td>
<td></td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>under 40</td>
<td>43</td>
<td>15</td>
<td>6</td>
<td>14</td>
<td>37</td>
<td>86.1</td>
<td>0.37</td>
</tr>
<tr>
<td>41-50</td>
<td>88</td>
<td>31</td>
<td>13</td>
<td>14.8</td>
<td>75</td>
<td>85.2</td>
<td></td>
</tr>
<tr>
<td>51-60</td>
<td>153</td>
<td>54</td>
<td>14</td>
<td>9.2</td>
<td>139</td>
<td>90.9</td>
<td></td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>146</td>
<td>51</td>
<td>20</td>
<td>13.7</td>
<td>126</td>
<td>86.3</td>
<td>0.26</td>
</tr>
<tr>
<td>Female</td>
<td>138</td>
<td>49</td>
<td>13</td>
<td>9.4</td>
<td>125</td>
<td>90.6</td>
<td></td>
</tr>
<tr>
<td><strong>Tumor Site</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left Colon</td>
<td>201</td>
<td>71</td>
<td>20</td>
<td>9.9</td>
<td>181</td>
<td>90.1</td>
<td>0.13</td>
</tr>
<tr>
<td>Right Colon</td>
<td>81</td>
<td>29</td>
<td>13</td>
<td>16</td>
<td>66</td>
<td>81.5</td>
<td></td>
</tr>
<tr>
<td><strong>Grade</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorly differentiated</td>
<td>18</td>
<td>6</td>
<td>2</td>
<td>11.1</td>
<td>16</td>
<td>88.9</td>
<td>0.93</td>
</tr>
<tr>
<td>Mod. differentiated</td>
<td>238</td>
<td>84</td>
<td>27</td>
<td>11.3</td>
<td>211</td>
<td>88.7</td>
<td></td>
</tr>
<tr>
<td>Well differentiated</td>
<td>28</td>
<td>10</td>
<td>4</td>
<td>14.3</td>
<td>24</td>
<td>88.7</td>
<td></td>
</tr>
<tr>
<td>**Stage **</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>23</td>
<td>8.7</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>100</td>
<td>0.30</td>
</tr>
<tr>
<td>II</td>
<td>145</td>
<td>54.5</td>
<td>15</td>
<td>10.3</td>
<td>130</td>
<td>89.7</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>61</td>
<td>22.9</td>
<td>6</td>
<td>9.8</td>
<td>55</td>
<td>90.2</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>37</td>
<td>13.9</td>
<td>7</td>
<td>18.9</td>
<td>30</td>
<td>81.1</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>18</td>
<td>11</td>
<td>5</td>
<td>27.9</td>
<td>13</td>
<td>72.2</td>
<td></td>
</tr>
</tbody>
</table>

* Tumor site was not reported for 2 cases

** Tumor stage was not reported for 18 cases
Figure 3.2 Capillary electrophoresis (ABI 3500) showing results obtained with the pentaplex panel of 5 mononucleotide repeat markers (NR-24, BAT-26, BAT-25, NR-24, MONO-27) used to confirm the presence of MSI. The x-axis is the PCR product size (bp) while the y-axis is the fluorescence intensity. Green, blue, and black peaks are amplification products from different microsatellite loci. Note the additional, shortened alleles (arrows) present in the tumor sample (B) compared to the normal mucosa sample (A), indicating MSI.
Figure 3.3  The average length of deletion in 5 mononucleotide repeat markers (NR-21, BAT-26, BAT-25, NR-24, MONO-27) was compared between MSI tumors from the WA and Saudi populations (A). The frequency of deletions in these markers was also compared (B). Statistically significant differences between the two populations are indicated by an asterix ($P<0.05$).
The 33 MSI cases were next screened for \textit{BRAF} mutations using CAST real time PCR. Only one tumor was found to have a \textit{BRAF} mutation and this was confirmed using pyrosequencing and Sanger sequencing (Figure 3.4A). The mutant allele frequency was estimated to be just 5-10\% and this was independently confirmed using Sanger DNA sequencing (Figure 3.4B).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{BRAF-mutation.png}
\caption{BRAF mutation (arrow) with low allele frequency is shown using pyrosequencing (A) and confirmed by Sanger sequencing (B). The mutation is C.1799T>A, p.Val600Glu in exon 15.}
\end{figure}
The incidence of \textit{KRAS} mutation amongst the MSI cases was determined by initial screening with F-SSCP followed by pyrosequencing to confirm and identify mutations (Figure 3.5). The observed frequency of \textit{KRAS} mutation was 27.3\% (9/33).

\textbf{A. \textit{KRAS} mutation detection by SSCP}

![SSCP Image]

\textbf{A. \textit{KRAS} mutation detection by Pyrosequencing}

![Pyrosequencing Image]

\textbf{Figure 3.5} \textit{KRAS} mutations in codons 12 and 13 revealed by F-SSCP (A) and confirmed by pyrosequencing (B). The mutation shown by pyrosequencing is c.35G>A, p.Gly12Asp.
Loss of expression of the four major MMR proteins (MLH1, PMS2, MSH2, MSH6) was examined by IHC in the MSI cases, except for three cases not stained for MLH1 due to lack of available tissue. Representative results for IHC are shown in Figure 3.6. Seven cases showed loss of both MSH2 and MSH6, 9 showed loss of both MLH1 and PMS2, 11 showed loss of PMS2 only, while 3 showed apparently normal staining with no loss of expression for any of the MMR proteins. Tissue was not available for MLH1 analysis in 3 of the 11 tumors that showed loss of PMS2 only.

Figure 3.6 Immunohistochemical staining for the mismatch repair proteins MSH2 (A and
B) and MLH1 (C and D). A and C show areas of positive staining in adjacent normal mucosa, while B and D show complete loss of staining of the respective MMR protein in the corresponding tumors. Magnification: 200x and 500x (inset).

3.5 Discussion

The incidence of CRC in the Saudi population is increasing rapidly and is currently the most frequent cancer type in males (Ibrahim et al, 2008; Al-Madouj et al, 2011). The average age of Saudi CRC patients at diagnosis is approximately 15 years younger than in Western CRC patients (55 vs 70 years, respectively) (Aljebreen, 2007; Mosli et al, 2012; Amin et al, 2012). Although the incidence of consanguinity in SA is over 50% (el-Hazmi et al, 1995), it is unclear whether this contributes to a higher incidence of familial cancer syndromes. The aim of this work was therefore to establish whether MSI screening of young CRC patients (<60 years) could help to identify unrecognized cases of LS in SA. Screening for MSI was performed in the absence of information on the family history of cancer in these patients. A similar approach used in a Western population was shown to increase the detection rate of LS in young CRC patients approximately 3-fold and this strategy has since been adopted as routine clinical practice in WA (Schofield et al, 2009; Schofield et al, 2014).

Somatic deletion in the BAT-26 mononucleotide repeat is a sensitive marker for MSI in CRC (Iacopetta et al, 2010). Using a rapid F-SSCP technique (Figure 3.1), MSI was found in 11.6% (33/284) of young CRC patients from the Eastern province of Dammam and from Riyadh. This frequency is almost identical to that reported by Siraj et al. (2015) in an unselected cohort of 807 CRC patients from
Riyadh (11.3%). In comparison, the MSI frequency found in young CRC patients from WA was significantly lower (7.8%, 105/1,344; \(P=0.045\)) (Schofield et al, 2009). The higher frequency of MSI observed in young CRC patients from SA may reflect a higher prevalence of LS compared to Western populations, however this awaits germline testing of the 32 MSI/BRAF wildtype cases identified here (described further in Chapter 4). The anatomical distribution of MSI cases was also different between the SA and WA populations. In WA, the majority of MSI tumors from young patients arise in the proximal colon (90/118, 76%) (Schofield et al, 2009). In contrast, less than half of the MSI tumors were right-sided in the current study (13/33, 39%; \(P<0.001\)) and in another study of unselected Saudi CRC patients (38/90, 42%) (Siraj et al, 2015). These results indicate there are likely to be differences in the etiology and pathogenesis of MSI tumors between Saudi and Western populations.

The pentaplex system for evaluating MSI status (Suraweera et al, 2002) is considered superior to the Bethesda panel that includes both dinucleotide and mononucleotide repeat markers (Xicola et al, 2007). This system was used here to confirm the MSI status of all 33 cases identified by F-SSCP screening with BAT-26 (Figure 3.2). For each of the 5 mononucleotide repeats, the average length of allelic deletion and the frequency of allelic deletion between SA and WA patients was also compared. While BAT-25 and BAT-26 showed a similar pattern of somatic deletion between the two cohorts, the NR-21, NR-24 and MONO-27 markers were deleted less frequently in MSI tumors from SA patients (Figure 3.3B), thus accounting for the shorter average length of deletions (Figure 3.3A). A sizeable minority of the Saudi MSI cases (9/33, 27%) showed no deletions in any of the NR-21, NR-24 and
MONO-27 markers, whereas these were each deleted in at least 95% of MSI cases from WA. Such geographic differences in the frequency of allelic deletion have not been previously reported for MSI tumors. The present findings confirm that BAT-26 is an excellent marker for MSI screening because the large size of deletions (average of 9-10 bp in WA and SA CRC) facilitates the detection of MSI.

The presence of \textit{BRAF} mutation is used to distinguish MSI tumors that are sporadic in origin from those that may arise due to LS (Domingo \textit{et al}, 2004). Of the 33 MSI cases identified in this cohort of young Saudi CRC patients, only 1 (3\%) contained a \textit{BRAF} mutation, therefore excluding it as a potential LS case. A very low frequency of \textit{BRAF} mutation (2.5\%, 19/757) was also reported by Siraj \textit{et al} (2014) in an unselected cohort of Saudi CRC patients, whereas a higher \textit{BRAF} mutation frequency (7/105, 6.7\%) was reported for MSI cases from young Australian CRC patients (Schofield \textit{et al}, 2009). The single \textit{BRAF} mutation detected here showed a very low allelic frequency (Figure 3.4), suggesting the presence of tumor heterogeneity. The remaining 32 MSI cases with wildtype \textit{BRAF} are candidates for germline mutation in MMR genes as the underlying cause of their MSI phenotype.

The presence of \textit{BRAF} mutation in CRC is mutually exclusive with that of \textit{KRAS} mutation (Li \textit{et al}, 2006). Because of the low \textit{BRAF} mutation frequency observed here and in a previous study of CRC in the Saudi population (Siraj \textit{et al}, 2015), we investigated whether this was compensated by a higher frequency of \textit{KRAS} mutation. The \textit{KRAS} mutation frequency observed here for MSI CRC (27\%, 9/33) was similar to that reported in a previous study of unselected Saudi CRC patients (30\%) (Beg \textit{et al}, 2015), but less than reported in a study of young Saudi
CRC patients (40%) (Elsamany et al, 2014). We are not aware of any other reports that have investigated the frequency of KRAS mutation specifically in MSI CRC.

MSI is almost always accompanied by loss of expression of one or more MMR proteins, usually as combinations of MLH1/PMS2 or MSH2/MSH6 loss. This information is important for informing germline mutation analysis to the appropriate gene(s). In the present study, all but 3 of the 30 (10%) MSI cases examined by IHC showed loss of expression of MMR proteins. In an earlier study on CRC from the WA population, 2 of 97 (2.1%) MSI cases showed no apparent loss of expression (Schofield et al, 2009), while a Saudi study also found discordant IHC results in 2 of 73 (2.7%) MSI tumors (Siraj et al, 2015). Such rare discordant cases may be explained by failure of the MMR gene mutation to alter protein antigenicity, or by mutation of a different MMR gene family member as the cause of MSI. In the present study, a similar proportion of MLH1/PMS2 and MSH2/MSH6 double loss cases (9 and 7, respectively) was observed in SA MSI cases compared to Australian MSI cases (42 and 30, respectively) (Schofield et al, 2009). However, a higher percentage of cases showing only the loss of PMS2 was found in SA cases (37%) compared to Australian MSI cases (7%). Tissue was not available in 3 of the 11 cases with PMS2 loss to test for concomitant loss of MLH1. Nevertheless, this result suggests that PMS2 mutations may be more frequent in the Saudi population.

3.6 Conclusions

In conclusion, the present work has shown that screening with the BAT-26 mononucleotide microsatellite marker is an efficient way to detect MSI in archival
tumor samples from Saudi CRC patients. The MSI frequency was significantly higher compared to a similar age cohort of Australian CRC patients. Analysis of results obtained with 5 mononucleotide repeat markers revealed novel differences in the frequency and length of allelic deletions between MSI tumors from Saudi and Australian patients. The anatomic distribution of MSI cases was also markedly different, with a more even distribution of MSI cases between the left and right colon in Saudi patients. The present results confirm the low frequency of \textit{BRAF} mutations in MSI tumors reported earlier in unselected Saudi CRC patients. This marker allows any rare MSI/\textit{BRAF} mutant patients to be excluded from further follow up as possible LS cases. The next step in this work is to follow up the surviving MSI patients identified here and to perform germline mutation analysis of MMR genes in individuals who give consent for genetic testing and following appropriate pre-test counselling (Chapter 4).
Screening for Lynch Syndrome in Young Saudi Colorectal Cancer patients using Microsatellite Instability Testing and Next Generation Sequencing

4.1 Abstract

Background: Individuals with Lynch Syndrome (LS) have germline variants in DNA mismatch repair (MMR) genes that confer a greatly increased risk of colorectal cancer (CRC), often at a young age. Identification of these individuals has been shown to increase their survival through improved surveillance. In the previous chapter, 33 high-risk cases for LS were identified amongst 284 young Saudi CRC patients by screening for microsatellite instability (MSI) in their tumor DNA. The aim of work presented in this chapter was to identify MMR gene variants within this cohort of MSI patients.

Methods: Peripheral blood DNA was obtained from 13 individuals who were at high risk of LS due to positive MSI status and young age (<60 years at diagnosis). Next generation sequencing, Sanger sequencing and Multiplex Ligation-dependent Probe Amplification were used to screen for germline variants in the MLH1, MSH2, MSH6 and PMS2 MMR genes. These were cross-referenced against several variant databases, including the International Society for Gastrointestinal Hereditary Tumors Incorporated database.

Results: Variants with pathogenic or likely pathogenic significance were identified in 8 of 13 (62%) high-risk cases, comprising 4 in MLH1 and 4 in MSH2. All carriers had a positive family history for CRC or endometrial cancer.
Conclusions: Next generation sequencing is an effective strategy for identifying young CRC patients who are at high risk of LS because of positive MSI status. Based on this work, it is estimated that 7% of CRC’s that arise in patients aged <60 years in Saudi Arabia are due to LS, accounting for around 50 new cases per year.

4.2 Introduction

LS, formerly known as hereditary non-polyposis colorectal cancer, is the most common inherited form of CRC. This condition is associated with a lifetime risk of approximately 80% for the development of CRC, as well as an increased risk for several other cancer types including endometrial, gastric, ovarian and kidney tumors (Lynch et al, 2009). Clinical features of LS include an early age of onset of CRC (average of 45 years) and the occurrence of multiple, related cancers within families (Vasen et al, 1999). The genes responsible for LS encode DNA mismatch repair (MMR) proteins and the majority of pathogenic variants occur in MLH1 and MSH2 (Lynch et al, 2015). There are no readily defined morphological features that distinguish CRCs arising in LS patients from non-familial cases. However, the large majority of tumors from LS patients are characterized by ubiquitous changes in the length of DNA microsatellite repeats due to a defective DNA MMR system. This is referred to as MSI and can be used to help identify CRC patients with LS (Baudhuin et al, 2005). The loss of expression of one or more MMR proteins in tumor cells, as detected by IHC, often also indicates the presence of an underlying pathogenic germline variant in MMR genes. Epigenetic changes to MMR genes may also account for some families with no apparent pathogenic germline variants but which
nevertheless show loss of tumor expression of MMR proteins and MSI (Lynch et al, 2015).

The proportion of LS cases amongst all CRC patients has been estimated to be up to 5%, however this incidence is likely to vary from country to country (de la Chapelle, 2005). The so-called Amsterdam clinical criteria were proposed as a guide to help detect LS (Vasen et al, 1991). Families must have experienced at least 3 cases of CRC over at least two generations, with at least one case diagnosed before the age of 50 years. However, the difficulty in obtaining a full family history of cancer and the lack of accessible genetic counselling and testing has meant that many LS families remain undetected, even in countries with advanced health care systems (Terdiman, 2005). The Bethesda criteria were subsequently proposed to increase the sensitivity of detection by incorporating the molecular screening of tumor DNA for MSI in suspected cases of LS (Umar et al, 2004; Matloff et al, 2013). Close surveillance and management of individuals with pathogenic germline variants in DNA MMR genes is cost-effective (Jarvinen et al, 2000) and colonoscopies performed at 3-year intervals reduce the mortality from CRC by as much as 70% (Dinh et al, 2011; Singh et al, 2010).

The results of a state-wide screening program to detect LS in the West Australian (WA) population has been previously reported (Schofield et al, 2009; Schofield et al, 2014). This was based on routine testing for MSI in CRC from all young patients (<60 yrs at diagnosis), irrespective of knowledge or information on family cancer history. Using this approach, the rate of detection of LS families in the WA population was increased by approximately three-fold (Schofield et al, 2014).
LS was estimated to account for slightly less than 1% of all CRC cases in the WA population. This concurs with the results reported by several other large, hospital-based studies in Western countries that also employed MSI testing of tumor DNA as the initial screen for the detection of LS (Samowitz et al, 2001; Piñol et al, 2005; Ward et al, 2013).

The average age of CRC patients in Saudi Arabia at the time of diagnosis is approximately 15 years younger than in Western populations (Aljebreen et al, 2007). For example, approximately one quarter of CRC patients in WA are aged <60 years at diagnosis (Threlfall, 2014) whereas slightly more than half of all CRC patients in SA are aged <60 years at diagnosis (SACR reports, 1994-2013). In the preceding chapter, almost 12% of CRC diagnosed in young (<60 years) Saudi patients were found to show a positive tumor MSI status (Chapter 3). Because this was approximately 50% more than observed in Australian CRC patients from the equivalent age cohort (Schofield et al, 2009), it was hypothesized the incidence of LS may also be higher in Saudi CRC patients. In this Chapter, next generation (massively parallel) sequencing was used to screen for germline DNA variants in Saudi CRC patients identified earlier as being at high risk of LS due to their young age and positive MSI status (Alqahtani et al, 2016).

4.3 Methods

Peripheral blood DNA was obtained from 13 consenting individuals who were at high risk of LS due to positive MSI status and young age (<60 years at diagnosis). The methodology for next generation sequencing and variant analysis of MMR
genes (MLH1, MSH2, MSH6 & PMS2) in this germline DNA was described in detail in Chapter 2.

4.4 Results and Discussion

Of the 13 red flag cases tested for germline variants in MMR genes using next generation sequencing, 8 (62%) were found to harbor pathogenic or likely pathogenic variants according to the American College of Medical Genetics and Genomics (ACMG) guidelines. Four of these were present in MSH2 and 4 in MLH1 (Tables 4.1 and 4.2). The average age of patients with these variants was 44.9 years at diagnosis and the gender distribution was even (4 males and 4 females). All 4 cases with MSH2 variants showed loss of expression for both MSH2 and MSH6 in tumor cells (Table 4.1). Three cases with MLH1 variants showed loss of both MLH1 and PMS2 expression, while the fourth showed a partial loss of PMS2 only (A3). Amsterdam II criteria were met in 4 cases (A3, A4, A5, A8), while a fifth case (A2) was borderline with two family members being diagnosed with CRC at age 50 yrs. All 13 cases evaluated here met the Bethesda criteria by virtue of being aged <60 yrs at diagnosis and having tumors that showed both MSI and loss of MMR gene expression.

Four of the 13 red flag cases revealed no significant MMR gene variants (A7, A9, A10, A11), with none of these meeting Amsterdam II criteria (Tables 4.1 and 4.2). Interestingly, one showed loss of PMS2 only (A7), another showed only partial loss of PMS2 (A9), while another showed focal loss of MLH1 and PMS2 (A11). Additionally, only one of these 4 cases included another family member who had
been diagnosed with CRC or endometrial cancer (A11).

The final case, A1, showed an MSH2 variant of unknown pathogenic significance (class 3), no family history of cancer, but loss of expression of MLH1 and PMS2 rather than MSH2 (Tables 4.1 and 4.2). However, in silico analysis provided evidence for a deleterious functional effect associated with this missense variant (Table 4.3). The five cases without apparent pathogenic germline variant are likely to represent sporadic CRC with MSI found in young patients and also showing loss of expression of MMR protein in IHC.

Seven of the 9 germline variants found here in young Saudi CRC patients were previously reported in the InSiGHT database as pathogenic variants, with 6 of these also reported in the ClinVar database (Table 4.2). Although not yet recorded in either of these databases, the variant found in case A2 was also recently reported in a study of Saudi CRC patients by Siraj et al (2015). In silico functional prediction analyses confirmed the deleterious and disease-causing effects of the four missense variants (A3, A4, A6, A8) previously reported in the InSiGHT and ClinVar databases (Table 4.3).

The final case, A1, showed an MSH2 variant of unknown pathogenic significance (class 3), no evidence of any family history of cancer, and tumor loss of MLH1 and PMS2 rather than MSH2 (Tables 4.1 and 4.2). However, in silico analysis suggested there was a deleterious functional effect associated with this missense variant (Table 4.3). Interestingly, this variant of unknown functional significance was previously reported in a study of CRC patients from the same population (Siraj et al, 2015).
Table 4.1  Clinical characteristics of 13 young Saudi CRC patients with MSI tumors tested for germline variants in MMR genes.

<table>
<thead>
<tr>
<th>Case</th>
<th>Age at Dx</th>
<th>Sex</th>
<th>IHC result</th>
<th>MMR gene variant</th>
<th>Family history of cancer</th>
<th>Amsterdam II</th>
<th>Bethesda</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>40</td>
<td>M</td>
<td>loss of MLH1</td>
<td>MSH2&lt;sup&gt;1&lt;/sup&gt;</td>
<td>none</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>A2</td>
<td>50</td>
<td>F</td>
<td>loss of MLH1</td>
<td>MSH2</td>
<td>mother died of CRC at 64 yrs</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>A3</td>
<td>48</td>
<td>M</td>
<td>partial loss of PMS2</td>
<td>MSH2</td>
<td>brother dx with CRC at 51 yrs</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>A4</td>
<td>47</td>
<td>M</td>
<td>loss of MLH1</td>
<td>MLH1</td>
<td>two sisters dx with CRC one died before age of 50 yrs</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>A5</td>
<td>34</td>
<td>F</td>
<td>loss of MLH1</td>
<td>MSH2</td>
<td>patient dx with breast cancer at 42 yrs</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>A6</td>
<td>55</td>
<td>F</td>
<td>loss of MLH1</td>
<td>MLH1</td>
<td>niece died of endometrial cancer at 42 yrs</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>A7</td>
<td>55</td>
<td>F</td>
<td>loss of PMS2</td>
<td>none found</td>
<td>father died of brain cancer at unknown age</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>A8</td>
<td>53</td>
<td>M</td>
<td>loss of MLH1</td>
<td>MSH2</td>
<td>mother died of endometrial cancer at 74 yrs</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>A9</td>
<td>41</td>
<td>M</td>
<td>partial loss of PMS2</td>
<td>none found</td>
<td>uncle died of unknown cancer at 60 yrs</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>A10</td>
<td>30</td>
<td>F</td>
<td>loss of MLH1</td>
<td>none found</td>
<td>none</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>A11</td>
<td>45</td>
<td>M</td>
<td>focal loss of MLH1</td>
<td>none found</td>
<td>sister diagnosed with CRC at 43 years</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>A12</td>
<td>29</td>
<td>M</td>
<td>loss of MLH1</td>
<td>MLH1</td>
<td>mother dx with polyps at &lt;60 yrs</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>A13</td>
<td>43</td>
<td>F</td>
<td>loss of MLH1</td>
<td>MSH2</td>
<td>sister dx with CRC at 40 yrs</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>

<sup>1</sup> Variant of unknown pathogenicity
Table 4.2  Summary of MMR gene variants and their reporting in InSiGHT and ClinVar databases.

<table>
<thead>
<tr>
<th>Case</th>
<th>Variant</th>
<th>Variant type</th>
<th>Variant class</th>
<th>InSiGHT</th>
<th>ClinVar</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>NM_000251.2(MSH2): c.[737A&gt;T][=] p.[(Lys246Ile)][(=)]</td>
<td>missense variant</td>
<td>3</td>
<td>NR ²</td>
<td>NR</td>
</tr>
<tr>
<td>A2 ³</td>
<td>NM_000251.2(MSH2): c.[2262_2267del][=] p.[(Ser755_Thr756del)][(=)]</td>
<td>in-frame deletion</td>
<td>4</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>A3</td>
<td>NM_000249.3(MLH1): c.[62C&gt;T][=] p.[(Ala21Val)][(=)]</td>
<td>missense variant</td>
<td>5</td>
<td>P ⁴</td>
<td>P</td>
</tr>
<tr>
<td>A4</td>
<td>NM_000249.3(MLH1): c.[1961C&gt;T][=] p.[(Pro654Leu)][(=)]</td>
<td>missense variant</td>
<td>5</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>A5</td>
<td>NM_000251.2(MSH2): c.[367-?_1276+?del][=] p.[(?)][(=)]</td>
<td>large deletion</td>
<td>5</td>
<td>P</td>
<td>NR</td>
</tr>
<tr>
<td>A6</td>
<td>NM_000249.3(MLH1): c.[677G&gt;A][=] p.[(Arg226Gln)][(=)]</td>
<td>missense variant ⁵</td>
<td>5</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>A8</td>
<td>NM_000251.2(MSH2): c.[2089T&gt;C][=] p.[(Cys697Arg)][(=)]</td>
<td>missense variant</td>
<td>5</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>A12</td>
<td>NM_000249.3(MLH1): c.[454-?_545+?del][=] p.[(?)][(=)]</td>
<td>large deletion</td>
<td>5</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>A13</td>
<td>NM_000251.2(MSH2): c.[2038C&gt;T][=] p.[(Arg680*)][(=)]</td>
<td>point/nonsense</td>
<td>5</td>
<td>P</td>
<td>P</td>
</tr>
</tbody>
</table>

1 According to ACMG guidelines
2 NR, not reported
3 Also reported by Siraj et al
4 P, pathogenic
5 This variant has been shown to cause aberrant splicing rather than a missense change in the protein
Table 4.3 *In silico* functional prediction analysis of variants (bioinformatic methods are described in Chapter 2.9).

<table>
<thead>
<tr>
<th>Case</th>
<th>Variant</th>
<th>Align GVGD</th>
<th>SIFT</th>
<th>MutTaster</th>
<th>PolyPhen2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>NM_000251.2(MSH2): c.[737A&gt;T];[=] p.[(Lys246Ile)];[=]</td>
<td>C35</td>
<td>deleterious</td>
<td>disease causing</td>
<td>possibly damaging (0.924)</td>
</tr>
<tr>
<td>A2</td>
<td>NM_000251.2(MSH2): c.[2262_2267del];[=] p.[(Ser755_Thr756del)];[=]</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>A3</td>
<td>NM_000249.3(MLH1): c.[62C&gt;T];[=] p.[(Ala21Val)];[=]</td>
<td>C65</td>
<td>deleterious</td>
<td>disease causing</td>
<td>probably damaging (0.998)</td>
</tr>
<tr>
<td>A4</td>
<td>NM_000249.3(MLH1): c.[1961C&gt;T];[=] p.[(Pro654Leu)];[=]</td>
<td>C65</td>
<td>deleterious</td>
<td>disease causing</td>
<td>probably damaging (1.0)</td>
</tr>
<tr>
<td>A5</td>
<td>NM_000251.2(MSH2): c.[367-?_1276+?del];[=] p.[(?)];[=]</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>A6</td>
<td>NM_000249.3(MLH1): c.[677G&gt;A];[=] p.[(Arg226Gln)];[=]</td>
<td>C0</td>
<td>deleterious</td>
<td>disease causing</td>
<td>probably damaging (0.991)</td>
</tr>
<tr>
<td>A8</td>
<td>NM_000251.2(MSH2): c.[2089T&gt;C];[=] p.[(Cys697Arg)];[=]</td>
<td>C65</td>
<td>deleterious</td>
<td>disease causing</td>
<td>probably damaging (1.0)</td>
</tr>
<tr>
<td>A12</td>
<td>NM_000249.3(MLH1): c.[454-_545+?del];[=] p.[(?)];[=]</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>A13</td>
<td>NM_000251.2(MSH2): c.[2038C&gt;T];[=] p.[(Arg680*)];[=]</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

1 Classification ranges from Class 0 (C0) to Class 65 (C65), with C0 less likely to interfere with function and C65 most likely to interfere with function.
Based on the earlier observation that 11.6% of 284 Saudi CRC patients aged <60 years were MSI positive (Chapter 3) and that 62% (8/13) of these were now subsequently found to harbor a pathogenic germline variant (see Figure 4.1 for flowchart), we estimate that approximately 7% of CRC cases arising in young Saudi patients are due to LS (i.e. 62% x 11.6%). The overall number (Figure 4.2A) and incidence (Figure 4.2B) of CRC in this population has increased approximately four-fold and two-fold, respectively, over the past 20 years (SCR, 1994-2013). The most recent data from the Saudi National Cancer Registry shows that almost 720 CRC cases were diagnosed nationwide in the young (<60 yrs) Saudi population in 2013 (SCR, 1994-2013). Therefore, the present findings suggest that about 50 new CRC cases per year in Saudi Arabia can be attributed to LS (7% of 720), with this number likely to keep increasing. Moreover, this estimate should be considered as a minimum since some LS-associated cases almost certainly arise in patients aged >60 years at diagnosis.

The two-fold increase in CRC incidence evident in young Saudis over the past 20 years (Figure 4.2B) may in part be due to improved diagnostic strategies and patient ascertainment. However, more likely causes are changes in dietary and lifestyle practices to resemble those found in Western countries with a high CRC incidence, including increased consumption of meat and rising levels of obesity. The four-fold increase in the overall number of CRC cases arising in young individuals over this period (Figure 4.2A) also reflects the rapidly increasing population of Saudi Arabia. Of note is the almost identical incidence of CRC between young (<60 yrs) male and female Saudis (Figure 4.2B). This contrasts with the approximately 60:40 gender distribution observed in Australian CRC patients from the same age group, respectively (Schofield et al, 2009).
Figure 4.1  Flowchart outlining the strategy for identifying patients with Lynch syndrome from within a large cohort of young Saudi CRC patients. The initial MSI screening, IHC and \textit{BRAF} mutation testing was described in Chapter 3.
Figure 4.2  Changes in the total number of cases (A) and incidence (B) of CRC in young (<60 years) patients in Saudi Arabia over the past two decades. The results shown are derived from the Saudi national cancer registry (SCR, 1994-2013).

The age of CRC patients in SA (Figure 4.3A) is considerably younger than that of CRC patients from WA (Figure 4.3B), with a mean age at diagnosis of 55 years compared to 70 years, respectively (Threlfall et al, 2007; SCR, 1994-2013).

Using the same strategy of MSI and BRAF screening of a very large (n=1,344) and unselected cohort of primary CRC, it was previously estimated that 3.6% of all CRC arising in young (<60 yrs) CRC patients in WA were attributable
to LS (Schofield et al, 2009). We therefore conclude that the contribution of LS to the overall burden of CRC is approximately two-fold higher in young patients from Saudi Arabia compared to a Western population (7% vs 3.6%).

Figure 4.3 Age distribution of CRC patients from Saudi Arabia (A) and Western Australia (B). The results shown are derived from the Saudi national (Al-Madouj et al, 1994-2013) and Western Australian state (Western Australian Cancer Registry, 2007-2013) cancer registries. Just over half of Saudi CRC patients are aged <60 years at diagnosis, compared to only one quarter of West Australian patients.

Siraj et al recently investigated a cohort of 807 unselected CRC patients from Riyadh, SA, for the prevalence of LS. These workers observed an almost identical incidence of MSI to the results described in Chapter 3 (11.3% vs 11.6%,
respectively). However, they reported a lower prevalence of germline variants in their overall CRC cohort (0.99%, 8/807) compared with the present estimate of 7% in a young cohort of Saudi CRC patients. This discrepancy could in part be due to the different ages of the study cohorts, as well as to differences in the methodology used to screen for genetic variants. The bioinformatics and variant classification strategy applied in the current study are conservative and therefore unlikely to explain the higher incidence of pathogenic variants observed in our cohort. The relatively low prevalence of LS amongst Saudi CRC patients reported by Siraj et al is in fact similar to results from a number of previous studies in Western countries, including from WA (Schofield et al, 2009; Samowitz et al, 2001; Piñol et al, 2005; Ward et al, 2013). These studies were also based on the initial screening of tumors for MSI and they each found that 0.8-1% of all CRC cases were due to LS.

A national MSI screening program for young CRC patients in SA, together with IHC and BRAF mutation analysis for the resulting MSI positive cases, should enable the large majority of newly diagnosed LS-associated cases to be identified. This would allow a Saudi national LS registry to be established, similar to other national (Vasen et al, 2016) or state (Siraj et al, 2015) programs. The benefits of such registries include the identification of LS families and the encouragement of long-term participation in surveillance programs for high risk individuals. Registration and regular colonoscopic surveillance has been demonstrated to substantially reduce the mortality from CRC in LS-affected families (Vasen et al, 2016).

4.5 Conclusions
In summary, MSI testing of primary tumors from young Saudi CRC patients was used as the initial screen for the detection of LS (Chapter 3). Following IHC and \textit{BRAF} mutation testing of MSI positive cases, next generation sequencing and MLPA were performed in order to identify patients with germline variants in MMR genes. All 8 patients with pathogenic variants identified in this study showed a positive family history for LS-related cancers, with 4 meeting the Amsterdam II criteria. The prevalence of LS in young CRC patients (<60 yrs) is estimated to be two-fold higher in SA compared to WA (7% vs 3.6%, respectively). If this incidence is confirmed by further studies in large patient cohorts, the present results suggest that at least 50 new CRC cases arise each year in SA because of LS. Identification of these individuals and of other affected family members, in conjunction with close clinical surveillance, should reduce the mortality from CRC. In conclusion, MSI testing of young CRC patients followed by next generation sequencing and MLPA is a feasible and cost-effective strategy to identify Saudi families with LS.
5.0 Identification of Lynch Syndrome in the Saudi Population
Using a Novel Strategy Involving the Saudi Cancer Registry

5.1 Background
The familial cancer condition known as LS is difficult to detect because clinicians often fail to record an accurate family history of cancer. Furthermore, there are no distinguishing clinical or pathological features between familial and sporadic cases of CRC. However, once mutation carriers for LS are identified, other family members can be screened for the mutation and protocols for regular surveillance (eg. colonoscopy) can be implemented with the aim of detecting any developing cancers at an early and therefore potentially curable stage (Jarvinen et al, 2000; Vasen et al, 1998; Umar et al, 2004). Unfortunately, it has been reported that many LS cases are missed despite the implementation of AC or Bethesda guidelines. Hampel et al (2008) revealed in a large cohort (n=500) that the sensitivity of AM and Bethesda criteria for identifying LS were 39% and 72%, respectively, suggesting a need to search for LS using other approaches.

In Chapters 3 and 4, a strategy was described for the detection of LS that was based upon retrospective screening for MSI in consecutive cases of CRC diagnosed in younger (<60 years) Saudi patients. This approach was not dependent on any prior information regarding family history of cancer. In the current chapter, an alternative strategy for the identification of LS families in the Saudi population will be investigated. This proposal will use the Saudi Cancer Registry to identify first-degree relatives who were diagnosed at a young age with CRC and/or another cancer type from the LS spectrum of tumors, namely endometrial, gastric, renal and ovarian. Such early-onset, familial clustering of
CRC and other LS-related tumors is strongly suggestive of LS. Family cancer clusters will be determined using an algorithm that identifies shared names, addresses and treating hospitals. Confirmation of LS amongst these family clusters will then involve the screening of archival tumor tissue from a cancer-affected family member for MSI, followed by genetic testing for MMR gene pathogenic variants in surviving individuals whose tumor tissue has tested positive for MSI. To our knowledge, this is the first attempt to use a population-based Cancer Registry to identify families with LS.

5.2 Aim

The aim of this work is to trial a novel approach for the identification of previously unknown families with LS in the Saudi population. This is based on using the Saudi national cancer registry to identify first-degree relatives amongst young patients diagnosed with CRC or another LS-related cancer type.

5.3 Potential Significance

Large, population-based studies suggest that only about 1% of CRC are due to LS. However, the large majority of these cases arise in young individuals (median age 45 years) in the prime of their life and hence the cost to society is very significant. By identifying LS families in the Saudi population, this research should reduce the mortality from CRC and other LS-related cancer types in relatively young individuals through the subsequent implementation of comprehensive surveillance programs.

Although this project should be considered a pilot study for the detection of LS, it was nevertheless anticipated that several families with this condition would be identified. Each of these families is likely to have several members that
are mutation carriers and have either already been diagnosed with cancer in the past, or are at a very high risk of developing this disease in future.

5.4 Methods

5.4.1 The Saudi Cancer Registry as an instrument for the detection of LS

The national Saudi cancer registry (SCR) was initiated in 1992 and includes 10 regional offices throughout Saudi Arabia. It registers separately all diagnosed cancer cases in the country for both nationals and foreigners. The first report from the SCR was published in 1996 and combines data for the three consecutive years 1994-1996. This was followed by reports for 1997-1998, 1999-2000 and subsequently on a yearly basis. The stored data includes the patient’s personal identification details, demographic information and tumor morphology details coded according to the International Classification of Diseases for Oncology 3rd and 10th Editions (ICD-O-3 and ICD-10). The SCR was used in the current study to identify individuals who met the following case definition:

*Person diagnosed at 55 years or younger with one or more cancers (including multiple of the same organ) belonging to the LS spectrum of tumors, i.e. colorectal carcinoma (ICD-10 codes C18-C20, C21), endometrial carcinoma (C54.1), gastric carcinoma (C16), ovarian carcinoma (C56), small bowel carcinoma (C17) and renal carcinoma (C64).*

Data extracted from the SCR included the patient’s full name and address, date of birth, and tumor type based on clinical and/or histopathological diagnosis. Each patient had a minimum of 4 names, comprising their first name, father’s name, grandfather’s name and family name. Many Saudis also carry the name of their great-grandfather, which further enhances the prediction for a family cluster. Compared to other Arab populations, Saudi females keep their paternal family
names even after they get married, thus providing additional information in the
determination of family clusters. Other data included in the national identification
record is date and place of birth, as well as the issuing authority for ID. These
comprise 108 offices in 13 administrative regions around the country.

5.4.2 **Strategy for In-Silico detection of LS using the Saudi Cancer Registry**

**Selection criteria**

A total of 20,081 cases of LS-related cancers were reported to the SCR between
2004 and 2014 and were included in this study. The data were received in
separate files for the different cancer types (CRC, endometrial, ovarian, gastric,
renal) and then merged into a single datasheet. The total number of cases for
each cancer type, together with their ICD-10 code, is shown in Table 5.1. As LS
syndrome is characterized by the early onset of disease, a filter was included
such that only patients aged 55 or younger at diagnosis were included. The 55-
year age limit was selected as an arbitrary compromise between sensitivity and
specificity for the detection of LS. Increasing the age limit would increase the
number of LS cases detected, but at the expense of increasing the size of the
study population and hence the workload involved in confirmation of family
history, tumor block retrieval and MSI analysis. Reducing the age limit decrease
the study population and subsequent workload, but also reduces the sensitivity
for LS detection.
Table 5.1  Overall cohort of individual cancer cases from the SCR used for the identification of family clusters

<table>
<thead>
<tr>
<th>Cancer Type</th>
<th>ICD-10 code</th>
<th>Number of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorectal cancer</td>
<td>C18-C20 &amp; C21</td>
<td>9,764</td>
</tr>
<tr>
<td>Gastric cancer</td>
<td>C16 &amp; C17</td>
<td>4,115</td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>C56</td>
<td>1,686</td>
</tr>
<tr>
<td>Endometrial cancer</td>
<td>C54</td>
<td>1,981</td>
</tr>
<tr>
<td>Renal cancer</td>
<td>C64</td>
<td>2,553</td>
</tr>
</tbody>
</table>

Software and data mining

A Microsoft access database was generated (In-Silico LS) to analyze the data using several queries aimed at identifying potential family clusters of cancers. Table 5.2 shows the identifiable data used during data mining (described in Chapter 2.9). These data included the full name, demographic data, national ID number, date of birth and Medical Record Number (MRN) of the primary, secondary and referral hospitals. The software also included data from the histology report and the stage of each case. A total of 22 variables were used to generate queries for the prediction of family clusters.

5.4.3  Confirmation of family clusters, genetic counselling and laboratory testing

The candidate family clusters identified as described above were verified through their MRN or through telephone interview by a medical oncologist at each hospital. Confirmed family clusters were considered to be “red flags” for LS and hence an archival tumor paraffin block was sought for DNA extraction and MSI testing from one or more family members. Patients whose tumors are found to show MSI will be recruited in order to obtain a full family history of cancer, undergo genetic counselling and obtain consent for germline mutation testing. The genetic counselling will follow routine practice at each hospital and a unified Form 2.3
Family Cluster Pedigree and History will be used for all cases. MSI and IHC testing were carried out on archival tumor tissue from at least one member of the family cluster as described in Chapter 2.

Table 5.2 Identifiable data used for software prediction of family clusters of cancer

<table>
<thead>
<tr>
<th>Variable</th>
<th>Family cluster case I</th>
</tr>
</thead>
<tbody>
<tr>
<td>First name</td>
<td>AB</td>
</tr>
<tr>
<td>First middle name</td>
<td>EF</td>
</tr>
<tr>
<td>Second middle name</td>
<td>GH</td>
</tr>
<tr>
<td>Third middle name</td>
<td>IJ</td>
</tr>
<tr>
<td>Family name</td>
<td>KMN</td>
</tr>
<tr>
<td>First name</td>
<td>AB</td>
</tr>
<tr>
<td>First middle name</td>
<td>EF</td>
</tr>
<tr>
<td>Second middle name</td>
<td>GH</td>
</tr>
<tr>
<td>Third middle name</td>
<td>IJ</td>
</tr>
<tr>
<td>Family name</td>
<td>KMN</td>
</tr>
<tr>
<td>Sex</td>
<td>M</td>
</tr>
<tr>
<td>DOB</td>
<td>1/01/1960</td>
</tr>
<tr>
<td>Place of birth</td>
<td>Dammam</td>
</tr>
<tr>
<td>Age difference</td>
<td>28 years 10 months 20 days</td>
</tr>
<tr>
<td>Address</td>
<td>A st street Khobar34217, Eastern province, Saudi Arabia</td>
</tr>
<tr>
<td>Region</td>
<td>Eastern province</td>
</tr>
<tr>
<td>General hospital</td>
<td>KFUH</td>
</tr>
<tr>
<td>MRN</td>
<td>30xxxxx</td>
</tr>
<tr>
<td>Ref hospital</td>
<td>KFMC</td>
</tr>
<tr>
<td>MRN</td>
<td>56xxxxx</td>
</tr>
<tr>
<td>Histology sample ID</td>
<td>HI-xxx</td>
</tr>
<tr>
<td>Tumor</td>
<td>CRC</td>
</tr>
<tr>
<td>Score of Similarity</td>
<td>80%</td>
</tr>
<tr>
<td>Predicted relationship</td>
<td>Son</td>
</tr>
<tr>
<td>% of relationship</td>
<td>90%</td>
</tr>
</tbody>
</table>

The similarities identified by the software are shown. For example, the first, second and third names of the subject in the left column match the second, third and fourth names of the subject in the right column, indicating a father and son relationship. Additionally, the address, state and referral hospitals were matched, indicating a high probability of positive family
relationship.

5.5 Results

A total of 20,081 cases of potentially LS-related cancers were reported in the SCR between 2004 and 2014 and included in this study. The data was received in 4 separate files and merged into a single datasheet, followed by export to the In-Silico LS database for further analysis. The distribution of cancers amongst the five LS-related cancer types investigated here is shown in Table 5.1. Since the large majority of elderly cases were presumably sporadic and not familial in origin, an age limit of 55 years or less at cancer diagnosis was used to enhance the prediction for true familial cancer clusters. This resulted in a total of 8,703 young patients being selected, together with 22 corresponding variables, for inclusion in the In-Silico LS database for further data-mining to identify putative family cancer clusters (Table 5.3).

Table 5.3 LS-related cancers amongst young patients from the SCR database

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>ICD-10 code</th>
<th>Number of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorectal</td>
<td>C18-C20 &amp; C21</td>
<td>4,333</td>
</tr>
<tr>
<td>Gastric</td>
<td>C16 &amp; C17</td>
<td>1,362</td>
</tr>
<tr>
<td>Ovarian</td>
<td>C56</td>
<td>931</td>
</tr>
<tr>
<td>Endometrial</td>
<td>C54</td>
<td>682</td>
</tr>
<tr>
<td>Renal</td>
<td>C64</td>
<td>1,395</td>
</tr>
</tbody>
</table>

The in-silico analysis identified a total of 330 patients who were potentially related. These comprised a total of 139 putative family clusters, as shown in Figure 5.1. The clusters varied from 2 to 6 individuals in size (Figure 5.2). Although the large majority comprised siblings, 25 clusters of putative first cousins were also
identified.

**Figure 5.1** Flowchart outlining the workflow used to identify potential family clusters of LS-related cancers in the Saudi population using information from the Saudi Cancer registry database for the period 2003-2014.

**Figure 5.2** Distribution of predicted family clusters with LS-spectrum of cancers amongst Saudi nationals reported in the SCR from 2003-2014.

* 25 groups of first cousins were identified from the same 139 putative family clusters
The age distribution for the 335 cases belonging to putative LS-related familial cancer clusters is shown in Figure 5.3. Over half the cases (181/335, 58%) were aged 45 years or less at cancer diagnosis.

**Figure 5.3** Age distribution for cases identified as belonging to putative familial clusters with LS-related cancer types in young Saudi cancer patients reported to the SCR from 2003-2014.
The *in-silico* prediction analysis revealed that CRC accounted for over half the cases (180/335, 53%) in the putative family cancer clusters (Figure 5.4). To date, an attempt has been made to confirm a familial relationship in 17 of the 139 putative family clusters. Of these, 13 (76%) were confirmed as first-degree relatives through either medical records or telephone interview by clinical staff. Tumor tissue for MSI testing was available for 8 of these family clusters, with two (25%) found to be positive for MSI (Figure 5.5). Both these families are highly likely to be LS.

**Figure 5.4** Distribution of cancer types amongst 335 young patients belonging to putative family clusters with LS-related cancers.

**Figure 5.5** Capillary electrophoresis (ABI 3500) showing results obtained with the pentaplex panel of 5 mononucleotide repeat markers (NR-24, BAT-26, BAT-25, NR-24, MONO-27) used to confirm the presence of MSI in a tumor block from a member of a putative family cancer cluster. The additional, shortened alleles (arrows) present in the tumor sample (B) compared to the normal mucosa sample (A) indicates the presence of MSI.
5.6 Discussion

The work described in Chapters 3 and 4 demonstrated that retrospective screening for MSI in consecutive cases of younger Saudi CRC patients can be used to identify previously unrecognized cases of LS. This result confirms similar findings reported in the West Australian population (Schofield et al, 2009). The aim of work presented in this chapter was to investigate whether an alternative approach that involves data mining of a national cancer registry can be used to identify unknown LS families. To our knowledge, this is the first study to use data from any national or state cancer registry to identify putative familial cancer clusters suggestive of LS. Since the establishment of the SCR in 1992 there has been a 3- and 4-fold increase in the annual number of reported CRC (Figure 5.6) and endometrial cancer cases, respectively. Many of these cases arise in younger patients, hence the urgent need to identify all families with LS in order to implement regular surveillance programs.

![Incidence of CRC in Saudi Arabia, 1994-2013](image)

**Fig 5.6** Incidence of CRC in Saudi Arabia over the past two decades. The results
Although this work should be considered preliminary, some interesting results were nevertheless obtained. It appears the algorithm used to identify family clusters of cancer is reasonably accurate, since 13 of the 17 (76%) clusters followed up to date were confirmed as being first-degree relatives. Of the 8 family clusters tested so far in the laboratory, two were found to be positive for MSI and are therefore highly likely to be LS. The first of these two familial clusters with MSI tumors comprised a female diagnosed with endometrial cancer at 44 yrs and her brother diagnosed with CRC at 37 yrs. Interestingly, amongst the 8 pathogenic germline variants carriers identified in Chapter 4 through retrospective MSI screening of CRC, two (25%) had first degree relatives previously diagnosed with endometrial cancer. The second familial cluster with MSI tumors comprised two sisters diagnosed with CRC at 26 yrs and 23 yrs of age respectively. These two families are currently being contacted by clinicians at their treating hospital in order to undergo further genetic counselling and to be offered genetic testing.

The yield of MSI positive cases from the family clusters tested to date was surprisingly low (2/8, 25%), although the sample size was quite small. Further work is required to confirm the family relationships in the remaining putative family cancer clusters, followed by MSI testing of tumors from confirmed families. From the preliminary findings to date, about 104 family clusters will be confirmed (76% of 139) as first-degree relatives. Of these, about 26 families may be positive for MSI (25% of 104) and therefore highly likely to have LS. Considering the large size of most Saudi families, this would imply the existence of many more mutation carriers who are at high risk of developing CRC or another LS-related cancer type.
One of the limitations of the *in-silico* approach to detect LS families is that the time span for data collection was only 11 years (2004 to 2014 inclusive). Extension of this period would allow parent-child clusters to be detected in addition to sibling or cousin clusters. Another limitation was the inconsistent spelling in English and Arabic names between different reporting hospitals. Standardization of the English spelling of Arabic names is likely to increase the accuracy of the prediction algorithm for family clusters. Use of the family name only instead of the Arabic triple name might make the prediction more accurate. Frequent population migration can also limit the accuracy of the algorithm due to treatment of family members at different hospitals and in different regions. Active collaboration of the Department of Civil Affair will expedite and ensure the accuracy of predicted family clusters. Consolidation of the ethics approvals into one full board application for use across all government hospitals will also greatly reduce the administrative burden. Identification of unrecognised LS families in the young Saudi population (approximately 80% of inhabitants are < 55 years of age) should significantly impact the health and economic cost of this familial cancer condition. This research will hopefully encourage the Ministry of Health to support researchers and clinicians in generating a national screening program for LS in Saudi Arabia.
6.0 **General Discussion**

6.1 **Overview**

In the Saudi population, more than half of all CRC cases occur in patients aged <60 years at diagnosis (Fig 1.4), suggesting a possible hereditary influence for this disease and in particular LS. The primary aim of this study was therefore to estimate the prevalence of LS amongst young (<60 years) CRC patients in the Saudi population. Research toward this aim allowed identification of the frequency of MSI amongst young CRC patients in Saudi Arabia, comparison of MSI with CRC patients from Western Australia, and finally the identification of pathogenic variants in MMR genes in this population. This research also trialled a novel approach to identifying LS in the population by using data from the national Saudi cancer registry (SCR) to identify family clusters affected by the LS spectrum of cancers.

6.2 **Significance of the study**

At the start of this work there were no published studies reporting the frequency of LS amongst young CRC patients of Arab ethnicity. Although only 1-2% of CRC are thought to be due to LS, these cases often arise in young individuals (median age 45 years) who are in the prime of their life. The rapid increase in the Saudi population and the steady increase in cancer incidence in the younger population (<60 years) over the past 15 years have given rise to an alarming situation. Health researchers and government officials are required to take urgent steps to prevent a very significant disease and economic burden to the Saudi Arabian society. These steps include identifying unrecognised LS families in the population, often without the assistance of detailed information on family history of cancer. As
suggested by this research, routine screening for MSI, together with IHC, will assist in determining which CRC patients should be targeted for germline mutation testing to identify LS. Identifying these families in Saudi Arabia should help to reduce the mortality from CRC and other LS-related cancer types by allowing regular endoscopic surveillance in mutation carriers.

6.3 LS amongst young (<60 years) CRC patients in the Saudi population

At the beginning of this study, a comprehensive survey of the literature revealed no published studies on screening for LS in the Saudi population. According to the SCR, the incidence of CRC in this population has increased over the last 25 years, especially in younger patients. Compared to CRC patients from Western countries, the average age of onset of CRC in Saudi Arabia is about 15 years younger (70 vs 55 years, respectively) (Ibrahim et al., 2008; Al-Madouj et al., 2011). Interestingly, more than half of all Saudi CRC patients are aged <60 years at diagnosis compared to only 23% in Western Australia (Fig 4.3). This observation suggests there may be a hereditary involvement (Vasen et al., 1999). In the current study, the MSI frequency amongst young Saudi CRC patients aged <60 years) was 11.6% (33/284), which is very similar to a previous finding (11.3%) in this population by Siraj et al. (2015). Of the 33 MSI cases identified here, 13 patients were successfully recruited and consented for germline mutation testing of MMR genes using next generation sequencing. Eight of these 13 patients were found to harbour pathogenic variants according to the ACMG reporting guidelines: 4 in MLH1 and 4 in MSH2 (Table 4.1). The average age of CRC onset for these patients was 44.9 years and there was an equal gender distribution (4 male and 4 female). All 13 recruited MSI cases met Bethesda criteria but only 4 met the Amsterdam-II
criteria, with an additional case considered as borderline due to two family members being diagnosed with CRC at age 50 years.

The number and incidence of CRC in the Saudi population has increased rapidly over the last 2 decades. Based on the findings from this work, it is estimated that approximately 7% of CRC cases diagnosed in young Saudi CRC patients are due to LS (i.e. 8/13 × 11.6%). Therefore, this result predicts that approximately 50 CRC cases per year in the younger Saudi population (<60 years) are due to LS (7% of 714 CRC cases) (SCR 2013-2014). Since some LS-associated CRC also occur in patients aged >60 years, this number is likely to be an underestimate. Thus, routine MSI testing for CRC patients aged <60 years followed by IHC and subsequent germline mutation testing using NGS for positive (red flag) cases may be a feasible and cost-effective strategy to identify unrecognised LS families. This strategy can be used equally for both retrospective and prospective cases, but clearly the patient follow up is much easier for newly diagnosed cases.

6.4 Frequency of MSI amongst young Saudi CRC patients

Despite the different microsatellite markers used to identify MSI in CRC patients from Western populations, the consensus is that about 7-20% of CRC tumors show this phenotype (Aaltonen et al, 1998; Salovaara et al, 2000; Cunningham et al, 2000; Samowitz et al, 2001; Hampel et al, 2005; Pinol et al, 2005; Schofield et al, 2009). Here, the MSI frequency amongst young CRC patients from Saudi Arabia was found to be 11.6%, which is almost identical to that reported for an unselected cohort in this population (11.3%) (Siraj et al, 2015). In comparison, the MSI frequency in young CRC patients (<60 years) from the Western Australian population was significantly lower (P=0.036) at 7.8% (Schofield et a, 2009), with
Pinol et al (2005) reporting 7%. The clinicopathological features of MSI CRC in the West Australian study revealed that only 23% (28/122) were located in the proximal colon (Schofield et al, 2009). This compares to 40% (13/33) in the current study and 42% (38/90) in another study of unselected Saudi CRC (Siraj et al, 2015). This suggests there may be etiological differences in the pathogenesis of MSI tumors between Arab and Caucasian populations.

The different clinicopathological features of MSI tumors from Saudi CRC patients were also accompanied by differences in the frequency and length of deletions in several microsatellite markers. To evaluate MSI status, a panel of 5 quasi-monomorphic mononucleotide markers (BAT-25, BAT-26, NR-21, NR-24 and MONO-27) was used in a single multiplex PCR reaction. In tumors with defective MMR (i.e. MSI), these markers are highly susceptible to somatic deletions (Surweera et al, 2002). For each of the 5 mononucleotide repeats, the average length of allelic deletion and the frequency of allelic deletion was compared between Saudi and West Australian CRC patients with MSI. The two cohorts showed similar patterns of somatic deletions for BAT-25 and BAT-26, however the NR-21, NR-24 and MONO-27 markers were deleted less frequently (Fig 3.3 B) and had shorter average lengths of deletion (Fig 3.3A). A significant proportion (9/33, 27%) of Saudi MSI patients showed no deletions in any of the NR-21, NR-24 and MONO-27 markers, whereas less than 5% of MSI cases from Western Australia showed no somatic deletions in these markers (Fig 3.3B). To our knowledge, such geographic differences in the frequency of allelic deletion have not been previously reported for MSI tumors.

BAT-26 was found to be an excellent marker for MSI screening because the large size of deletions (average of 9-10 bp in Saudi and West Australian CRC)
facilitates detection. Xicola et al (2007) concluded that the pentaplex of mononucleotide repeats was more accurate for the detection of MSI than the Bethesda panel of markers. For routine screening, the use of BAT-25 and BAT-26 markers alone are sufficient to establish MSI-H status in the absence of DNA from normal tissue because of the large somatic deletions observed in the tumor DNA and because of their quasi-monomorphic size in almost all populations (Suraweera et al, 2002).

6.5 Amsterdam criteria and revised Bethesda guidelines to identify LS

Although MSI testing facilitates the screening for potential LS cases, recruitment of individuals for germline mutation testing remains a major challenge. An international collaborative group first established the so-called Amsterdam criteria to assist with the diagnosis of LS. However, the selection criteria required for CRC patients to undergo germline mutation testing was quite stringent, resulting in low sensitivity due to many individuals not meeting the criteria. Subsequently, amendments were made to include extracolonic LS-associated cancers such as endometrial and renal tumors.

It has been well documented that mutations in MMR genes almost always lead to loss of protein expression for one or usually two of these genes (MLH1/PMS2, or MSH2/MSH6). Consequently, most clinicians include IHC testing as part of the work up for suspected LS cancers. An ongoing debate is whether MSI or IHC testing is preferable for routine screening to detect LS. Both techniques have advantages and disadvantages, with MSI in particular being highly specific and sensitive and therefore suited for large cohorts. On the other hand, IHC that includes antibodies against all 4 MMR genes is considered to be as sensitive as
MSI, but has several limitations such as being labour intensive and sometimes also resulting in difficult to interpret staining patterns without standardised guidelines (Muller et al, 2004; Arends et al, 2008; Barzi et al, 2015; Markow et al, 2017; Svrcek et al, 2019).

Although the revised Amsterdam-II guidelines for LS were widely accepted, their sensitivity remains low compared to other criteria that involve MSI screening, such as the Bethesda criteria or routine MSI screening of young (<50 years) CRC patients. In response to the criticism of Amsterdam-II criteria being too stringent, the Bethesda guidelines and subsequent revised Bethesda guidelines were established and incorporate 5 microsatellite markers (two mononucleotide repeats and three dinucleotide repeats). It was later recognised that mononucleotide markers (BAT-25 and BAT-26) had more sensitivity and specificity than the dinucleotide markers in the Bethesda panel. Suraweera et al (2002) identified three additional mononucleotides repeat markers (NR-21, NR-24, MONO-27) that were as sensitive as BAT-25 and BAT-26 for the detection of MSI in tumors. Application of the Amsterdam clinical criteria in large patient cohorts is time consuming and has relatively low sensitivity for the detection of LS. On the other hand, MSI screening of younger CRC patients using mononucleotide repeats in the pentaplex assay has greater sensitivity and specificity and a low cost.

6.6 Retrospective MSI screening of young CRC patients for LS detection

MSI has prognostic and predictive significance for CRC other than its use in helping to identify families with LS. Several large retrospective studies have followed either the Amsterdam-II or Bethesda guidelines, or used MSI screening with the Bethesda microsatellite panel in order to determine the eligibility of CRC
patients for further genetic testing. However some LS cases can be missed due to three main reasons: (1) the high stringency of Amsterdam I & II criteria, (2) the challenges involved in obtaining family history of cancer in a large, retrospective cohort, and (3) misclassification of MSI due to the low specificity of the dinucleotide repeats in the Bethesda panel. Although there is ongoing debate about the superiority of routine MSI and IHC screening for the detection of LS, these methods show similar sensitivity and specificity when used by competent scientists and experienced laboratories (Halvarsson et al, 2004).

6.7 In-Silico approach to identifying LS using cancer registries

To our knowledge, an In-Silico approach using cancer registry data to identify family clusters of LS-related cancers has not previously been used to help identify LS. The granting of government (SCR) and health provider approval for this approach was challenging, as was the obtaining of human research ethics approval. However, there was also strong awareness and concern by government officials regarding the burden of familial cancer syndromes such as LS for the Saudi health system. The In-silico approach employed here successfully imported comprehensive data on 20,081 Saudi nationals diagnosed with one or more tumors from the LS-spectrum of cancers and reported between 2004 and 2014 (Chapter 5). For practical purposes, the family cluster analysis used a patient age threshold of ≤55 years at diagnosis in order to increase the specificity for LS. This resulted in the identification of 139 distinct, putative family clusters, each containing 2-6 relatives.

The major variable used for the cluster analysis was the full Arabic name as a unique identifier. Moreover, the issuing authority for the civil ID pass can identify
the geographical residence of the patient as well as their place of birth. The In-
Silico strategy used here was somewhat limited because the relatively short period
for cancer diagnosis (2004-2014) meant it was unlikely that parent-child
relationships were detected. Indeed, the large majority of family clusters involved
siblings. Grouping of the family clusters by referral hospital expedited their follow-
up for confirmation of the family relationship through patient medical records. This
allowed confirmation of a familial relationship in 13 of the 17 clusters that were
followed up. Of these, tumors from 8 clusters were tested for MSI and two were
found to show MSI. Cluster A: a female diagnosed with endometrial cancer at age
44, and her brother diagnosed with CRC at age 37. Cluster B: two sisters, both
diagnosed with CRC at ages of 26 and 23 years respectively. Both families are
very likely to have LS, although germline testing has yet to be carried out to confirm
this.

The accuracy of the In-Silico approach for identifying true family clusters
varied between hospital centres from 20% (1/5) to 50% (6/12). This may be
explained by the quality of record keeping and information systems at individual
centres. Standardised patient health care ID (Medicare ID) should increase the
accuracy of predicting familial relationships, particularly in remote hospitals with
poor IT systems. Moreover, expansion of the In-Silico analysis to a period of 20
years or more (eg. 1995-2014) will likely result in many additional family clusters
due to capture of parent-child and first cousin relationships. The In-Silico strategy
revealed a relatively low incidence of positive MSI cases amongst confirmed family
clusters (2/8), however this should be considered as a pilot study only. Covering a
longer period in the SCR (>20 years) and incorporating additional features to
identify first-degree relatives should enhance the prediction for true family cancer
clusters and therefore LS detection.

6.8 Limitations of the study

This research has established that identification of unrecognised LS amongst Saudi CRC patients using MSI testing as the initial screen is feasible. However, there are several limitations to the strategies employed in this work:

- Little or no information on family history of cancer was available in patient medical records

- The human research ethics committees in certain institutes were reluctant to approve retrospective studies for MSI screening, thus jeopardising efforts to identify pathogenic variants carriers for LS

- Misspelling of patient name and incorrect national ID number by the hospital information system data entry sometimes made it difficult or impossible to follow up patients

- Incomplete patient demographic data and contact details

- The absence of a familial cancer clinic at many hospitals made patient follow up more difficult and meant that specialized genetic counselling services were not available and had to be obtained from elsewhere

- Exporting of tissue or DNA sample to an overseas laboratory in Australia required additional approval from a government organisation (King Abdulaziz City for Science and Technology)
- Lack of molecular pathology laboratory setup for MSI and \textit{BRAF} mutation in many regional Saudi hospitals meant that samples had to be tested elsewhere.

6.9 Recommendations for future work

From the results of this study, it is recommended that identification of LS in Saudi Arabia could be vastly improved through retrospective MSI screening, prospective routine MSI screening, and \textit{In-Silico} database prediction methodology. Recommendations for improvement include:

- Establishment of centralised, state-wide genetic services that provide a free clinical service following referral from a GP or hospital. This service should include genetic counselling, germline mutation testing and advice on surveillance programs.

- Enhanced collaboration between health centres, clinicians, pathologists and medical laboratory scientists to ensure routine MSI screening, IHC and follow up of red flag cases.

- Standardization of the English spelling of Arabic names and use of the family name to increase the accuracy of the \textit{In-Silico} prediction algorithm for family clusters.

- Establish a unique personal health ID and centralise the patient data either in a secure cloud or government server.

6.10 Conclusions

The work described in this thesis has established the frequency of LS amongst
young Saudi CRC patients and found a higher incidence compared to Western populations. This work supports the routine use of MSI screening to identify LS, regardless of the availability of information about family history of cancer. The major findings can be summarized as:

- Saudi Arabia has a high proportion of young population with elevated risk of developing LS

- Routine screening of young CRC patients with MSI is not currently available in large referral hospitals but was shown here to be an effective strategy for identifying LS cases

- MSI screening of retrospective CRC patient cohorts for the identification of LS is feasible, regardless of information on the family history of cancer

- Use of an In-Silico approach involving national or state cancer registries can identify previously unknown LS families. Refinement of the prediction tools for family cancer clusters may provide a cost-effective option for identifying LS that is based on initial analysis of cancer registry data. However, direct MSI screening of young CRC patients is likely to be the most cost effective way to identify LS families in the Saudi population.
7. Bibliography


Terdiman J (2005). It is time to get serious about diagnosing Lynch syndrome (hereditary nonpolyposis colorectal cancer with defective DNA mismatch repair) in the general population. *Gastroenterology* 129:741-744


RESEARCH ARTICLE

Screening for Lynch Syndrome in Young Colorectal Cancer Patients from Saudi Arabia Using Microsatellite Instability as the Initial Test

Masood Alqahtani\textsuperscript{1,2}, Fabienne Grieu\textsuperscript{3}, Amerigo Carrello\textsuperscript{3}, Benhur Amanuel\textsuperscript{3}, Miral Mashour\textsuperscript{2}, Rabab Alattas\textsuperscript{2}, Khalid Al-Saleh\textsuperscript{4}, Abdulmalik Alsheikh\textsuperscript{4}, Sarah Alqahtani\textsuperscript{5}, Barry Iacopetta\textsuperscript{1*}

Abstract

Background: Lynch Syndrome (LS) is a familial cancer condition caused by germline mutations in DNA mismatch repair genes. Individuals with LS have a greatly increased risk of developing colorectal cancer (CRC) and it is therefore important to identify mutation carriers so they can undergo regular surveillance. Tumor DNA from LS patients characteristically shows microsatellite instability (MSI). Our aim here was to screen young CRC patients for MSI as a first step in the identification of unrecognized cases of LS in the Saudi population.

Materials and Methods: Archival tumor tissue was obtained from 284 CRC patients treated at 4 institutes in Dammam and Riyadh between 2006 and 2015 and aged less than 60 years at diagnosis. MSI screening was performed using the BAT-26 microsatellite marker and positive cases confirmed using the pentaplex MSI analysis system. Positive cases were screened for BRAF mutations to exclude sporadic CRC and were evaluated for loss of expression of 4 DNA mismatch repair proteins using immunohistochemistry.

Results: MSI was found in 33/284 (11.6%) cases, of which only one showed a BRAF mutation. Saudi MSI cases showed similar instability in the BAT-26 and BAT-25 markers to Australian MSI cases, but significantly lower frequencies of instability in 3 other microsatellite markers.

Conclusions: MSI screening of young Saudi CRC patients reveals that approximately 1 in 9 are candidates for LS. Patients with MSI are strongly recommended to undergo genetic counselling and germline mutation testing for LS. Other affected family members can then be identified and offered regular surveillance for early detection of LS-associated cancers.

Keywords: Colorectal cancer - microsatellite instability - Lynch syndrome - screening - Saudi Arabia

Asian Pac J Cancer Prev, 17 (4), 1917-1923

Introduction

Hereditary nonpolyposis colorectal cancer (HNPCC), more commonly known as Lynch syndrome (LS), is an autosomal dominant genetic condition associated with a high risk of colorectal cancer (CRC) (Lynch et al., 2003). Approximately 1-2% of all CRC are thought to be due to LS. In addition to CRC, LS is also associated with increased risk for endometrial, gastric, ovarian, small bowel, and other rarer cancer types. LS is caused by germline mutations in DNA mismatch repair (MMR) genes, with the most frequent being MLH1, MSH2, MSH6, and PMS2. The incidence of MMR gene mutation carriers is about 1 in 500 in Western populations. Identification of these individuals is crucial because it allows them to undergo early and regular surveillance for cancer. It also allows their extended family to be screened for additional mutation carriers. Regular screening by colonoscopy has been shown to reduce mortality from CRC in mutation carriers (Jarvinen et al., 2000). Until recently, CRC patients who were suspected of being at risk for LS were identified through the use of clinical criteria which rely on obtaining a detailed family history of cancer, as outlined in the Amsterdam and Bethesda guidelines (Umar et al., 2004). However, these guidelines have been associated with low sensitivity for the detection of LS and their implementation in routine clinical practice has been poor (Lynch et al., 2004). Consequently, there have been calls to introduce laboratory-based screening tests for LS that do not require the clinician to obtain a detailed family history of cancer (Terdiman, 2005).

Tumors that arise in patients with LS have a defective
DNA MMR system, resulting in ubiquitous small deletions in DNA repeat regions that are referred to as microsatellite instability (MSI) (Iacopetta et al., 2010). In addition to MSI, these tumors almost always show loss of expression of one or more MMR proteins as seen with immunohistochemistry (IHC) (Marcus et al., 2009). MSI and IHC tests are therefore essential for the selection of CRC patients to undergo further germline mutation testing for LS. However, a positive MSI test and loss of MMR expression by IHC are also observed in approximately 10% of sporadic CRC, meaning they are not specific markers for the presence of LS. Fortunately, sporadic MSI CRC cases often display a hotspot mutation in the BRAF oncogene, whereas MSI CRC from LS patients never show mutations in this oncogene. The presence of BRAF mutation can therefore be used to identify MSI CRC cases that are sporadic in origin and can thus be excluded from further testing for possible germline mutations (Domingo et al., 2004).

We previously reported the results of MSI testing in young CRC patients as a first screen to detect LS in the state of Western Australia (WA) (Schofield et al., 2009). This laboratory-based screening program was conducted in the absence of any information on the family history of cancer. It was established that initial screening for MSI, followed by testing for BRAF mutation in positive cases, was an effective strategy to identify LS mutation carriers in the WA population. Routine MSI and IHC testing was subsequently implemented for all CRC patients aged <60 years in WA starting in 2008. The rate of diagnosed LS cases in this state (population 2.5 million) has since increased from 2-3 per year prior to routine MSI screening to an average of 8 cases per year over the past 7 years (Schofield et al., 2014).

In Saudi Arabia (SA), CRC is the most frequent cancer type in males (13% of all cancer cases) and the second most common cancer in females (9%), with a total of almost 1,200 cases reported in 2011 (Al-Madouj et al., 2011). Interestingly, the mean age at diagnosis (55-58 years) is approximately 12-15 years younger compared to Western populations (Alijbreen, 2007; Mosli et al., 2012; Amin et al., 2012). Compared to WA, the age standardized rate for CRC is about 3-4-fold lower in SA, although the incidence appears to be increasing quite rapidly (Ibrahim et al., 2008), probably due to the adoption of a more Western diet. Little is known however about the incidence of LS in the Saudi population and to our knowledge there are no reflex MSI- or IHC-based screening programs to help detect LS in the routine clinical setting. A recent publication involving 807 CRC cases from Riyadh reported an MSI frequency of 11.3% and LS frequency of 0.9% (Siraj et al., 2015). The LS frequency is similar to previous results from large Australian (Schofeld et al., 2009; Ward et al., 2013), Spanish (Pinol et al., 2005) and American (Samowitz et al., 2001) studies, but lower than estimates of 5.1% (Zeinalian et al., 2015) and 2% (Nemati et al., 2011) from Iranian studies. It should be noted however that the latter studies were based on clinical findings (Amsterdam criteria) rather than genetic findings.

In the present study we have retrospectively screened Saudi CRC cases from the Eastern Province and from Riyadh for MSI and BRAF mutation as a first step in the identification of LS cases. The results are compared with those of a similar study conducted in the WA population (Schofield et al., 2009), which has a 3-fold higher incidence of CRC and a much older mean patient age compared to the Saudi population.

Materials and Methods

Patient cohorts

All patients with primary CRC diagnosed at three hospitals in Dammam (n=191; King Fahad Specialist Hospital, King Fahad University Hospital, Dammam Regional Laboratory) and the King Khaled University Hospital in Riyadh (n=93) between 2006 and 2015 were eligible. Because the large majority of LS patients are diagnosed with CRC at a young age (Lynch et al., 2003), only patients aged <60 years were included. No information was available regarding the family history of cancer. Clinicopathological information including gender, age, stage, tumor site and histological grade were obtained from pathology records. For each case, sections of 10μm thickness (for DNA extraction) or 4μm thickness (for IHC) were cut from archival, paraffin-embedded tumor and normal tissue blocks were obtained from surgically resected specimens. Each block was verified for maximal tumor cell content (>50%) by examination of H&E stained slides by a pathologist. The study was approved by the human research ethics committees of the King Fahad Specialist Hospital-Dammam (IRB LAB 055), King Khaled University Hospital (15/0148/IRB), King Fahad University Hospital (IRB-2014-01-297) and Dammam Regional Laboratory (approval date 18/08/2014).

DNA extraction

DNA extraction from paraffin-embedded tissue sections was performed using a kit and automated DNA extractor as described by the manufacturer (MagNA Pure Compact, Roche, USA). DNA purity and concentration were evaluated by NanoDrop spectrophotometer.

MSI analysis

Initial MSI screening was performed using PCR and fluorescent-single strand conformation polymorphism (F-SSCP) analysis to detect deletions in the BAT-26 mononucleotide repeat (Iacopetta et al., 2000). All positive cases were confirmed using the commercially available pentaplex MSI analysis system as described by the manufacturer (Promega, Australia). This contains five mononucleotide repeat markers (BAT-25, BAT-26, NR-21, NR-24 and MONO-27) that are highly susceptible to somatic deletions in tumors with defective MMR (Suraweera et al., 2002), as well as two pentanucleotide markers used to ensure correct sample matching. PCR products were run on an ABI 3500 genetic analyzer and the length of allelic deletions was quantified using Genemapper software (ABI, California, USA). Deletion lengths and frequencies were compared to those of MSI cases detected in young CRC patients in WA using the pentaplex system during routine evaluation.
BRAF and KRAS mutation analysis

Samples were screened for mutations in codon 600 of BRAF using competitive allele-specific Taqman (CAST)-PCR (Life Technologies, USA) as described previously by our group (Richter et al., 2013). Data was collected during 40 cycles of amplification and analysed using the Mutation DetectorTM software v.2.0 (Life Technologies, USA). Samples with a Δ(d)Ct of less than 9.96 were considered positive for mutation, where ΔCt= Ct mut – Ct ref.

Mutations in codon 600 of BRAF were confirmed using the Therascreen® BRAF Pyrosequencing Kit (Qiagen, Australia) according to manufacturer’s instructions. The KRAS Pyrosequencing kit (Qiagen, Australia) was also used to detect mutations in codons 12, 13 and 61 of the KRAS gene. Pyrosequencing was performed on the PyroMark Q24 platform (Qiagen) using Therascreen buffers and reagents (v1). Readouts were generated with the PyroMark Q24 software (v. 2.0.6.) and data was analyzed manually or with a plug-in tool provided by Qiagen.

Immunohistochemistry

Tumor samples that were found to be positive for MSI and wildtype for BRAF were investigated for loss of expression of the MLH1, PMS2, MSH2 and MSH6 proteins using IHC as described previously (Schofield et al., 2009). Briefly, 4μm tissue sections were cut serially from the same tumor blocks used to provide tissue samples for DNA extraction. IHC for MLH1, PMS2, MSH2 and MSH6 expression was performed using commercially available antibodies (clones M1, EPR3947, G219-1129 and 44, respectively) at the recommended dilutions (Ventana, Australia). Normal colonic epithelium located adjacent to tumor cells served as the internal positive control for MMR protein expression. Tumors were scored as showing normal expression, partial loss of expression or complete loss of expression.

Statistical analysis

Comparison between SA and WA MSI cases for the length of allelic deletions was performed using the Student’s t-test. Associations between the presence of MSI and clinicopathological features was evaluated using Fisher’s exact test. Significance was assumed at P<0.05.

Results

A total of 284 cases were successfully tested for MSI using PCR and F-SSCP analysis for deletions in the BAT-26 marker. Representative results obtained with this screening technique are shown in Figure 1. Thirty three cases (11.6%) showed deletions in BAT-26 and all were confirmed as having MSI using the pentaplex system that analyzes 5 mononucleotide repeats (Figure 2). Associations between MSI and clinicopathological features are shown in Table 1. Younger and male patients showed a trend for higher MSI frequency. Right-sided tumors also showed a higher frequency of MSI compared

Table 1. Clinicopathological Features of MSI Tumors in Young Saudi Arabian Colorectal Cancer Patients

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>%</th>
<th>MSI</th>
<th>%</th>
<th>MSS</th>
<th>%</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total no. of cases</td>
<td>284</td>
<td></td>
<td>33</td>
<td>11.6</td>
<td>251</td>
<td>88.4</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>under 40</td>
<td>43</td>
<td>15</td>
<td>6</td>
<td>14</td>
<td>37</td>
<td>86.1</td>
<td>0.37</td>
</tr>
<tr>
<td>41-50</td>
<td>88</td>
<td>31</td>
<td>13</td>
<td>14.8</td>
<td>75</td>
<td>85.2</td>
<td></td>
</tr>
<tr>
<td>51-60</td>
<td>153</td>
<td>54</td>
<td>14</td>
<td>9.2</td>
<td>139</td>
<td>90.9</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>146</td>
<td>51</td>
<td>20</td>
<td>13.7</td>
<td>126</td>
<td>86.3</td>
<td>0.26</td>
</tr>
<tr>
<td>Female</td>
<td>138</td>
<td>49</td>
<td>13</td>
<td>9.4</td>
<td>125</td>
<td>90.6</td>
<td></td>
</tr>
<tr>
<td>Tumor Site*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left Colon</td>
<td>201</td>
<td>71</td>
<td>20</td>
<td>9.9</td>
<td>181</td>
<td>90.1</td>
<td>0.13</td>
</tr>
<tr>
<td>Right Colon</td>
<td>81</td>
<td>29</td>
<td>13</td>
<td>16</td>
<td>66</td>
<td>81.5</td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorly differentiated</td>
<td>18</td>
<td>6</td>
<td>2</td>
<td>11.1</td>
<td>16</td>
<td>88.9</td>
<td>0.93</td>
</tr>
<tr>
<td>Mod. differentiated</td>
<td>238</td>
<td>84</td>
<td>27</td>
<td>11.3</td>
<td>211</td>
<td>88.7</td>
<td></td>
</tr>
<tr>
<td>Well differentiated</td>
<td>28</td>
<td>10</td>
<td>4</td>
<td>14.3</td>
<td>24</td>
<td>88.7</td>
<td></td>
</tr>
<tr>
<td>Stage **</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>9</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0.30</td>
</tr>
<tr>
<td>II</td>
<td>122</td>
<td>65</td>
<td>12</td>
<td>9.8</td>
<td>110</td>
<td>90.2</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>25</td>
<td>13</td>
<td>4</td>
<td>16</td>
<td>21</td>
<td>84</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>32</td>
<td>17</td>
<td>6</td>
<td>18.8</td>
<td>26</td>
<td>81.3</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>96</td>
<td>11</td>
<td>11</td>
<td>6.6</td>
<td>85</td>
<td>88.5</td>
<td></td>
</tr>
</tbody>
</table>

*Tumor site was not reported for 2 cases; ** Tumor site was not reported for 96 samples.
The length and frequency of allelic deletions was quantified in each of the 5 mononucleotide repeats of the pentaplex assay. This analysis was performed in 33 MSI cases from SA and in 56 age-matched MSI cases from WA (Figure 3). BAT-26 was deleted in all cases in both cohorts, with a similar average length of deletion (9.1bp in WA and 9.7bp in SA). Interestingly, the average length of deletion for the NR-21, NR-24 and MONO-27 markers was significantly shorter in MSI tumors from SA (Figure 3A). This was due to the lower frequency of deletion of these 3 markers in MSI tumors from SA (Figure 3B). Nine of the 33 MSI tumors from SA (27.3%) showed deletions only in the BAT-26 and BAT-25 markers.

We next screened the 33 MSI cases from SA for
BRAF mutations using CAST real time PCR. Only one tumor was found to have a BRAF mutation and this was confirmed using pyrosequencing (Figure 4A). The mutant allele frequency was estimated at just 5-10% and this was independently confirmed using Sanger DNA sequencing (Figure 4B). Therefore, all MSI cases in this series except one were possible candidates for LS.

The incidence of KRAS mutation amongst the MSI cases was determined by initial screening with F-SSCP followed by pyrosequencing to confirm and identify mutations (Figure 5). The observed KRAS mutation frequency was 27.3% (9/33).

Loss of expression of the four major MMR proteins (MLH1, PMS2, MSH2, MSH6) was examined by IHC in 30 MSI cases for which tissue was available. Representative results for IHC are shown in Figure 6. Seven cases showed loss of both MSH2 and MSH6, 9 showed loss of both MLH1 and PMS2, 11 showed loss of PMS2 only, while 3 showed apparently normal staining with no loss of expression for any of the MMR proteins. Tissue was not available for MLH1 analysis in 3 of the 11 tumors that showed loss of PMS2 only.

Discussion

The incidence of CRC in the Saudi population is increasing rapidly and is currently the most frequent cancer type in males (Ibrahim et al., 2008; Al-Madouj et al., 2011). The average age of Saudi CRC patients at diagnosis is approximately 15 years younger than Western CRC patients (55 vs 70 years, respectively) (Aljebreen, 2007; Mosli et al., 2012; Amin et al., 2012). Although the incidence of consanguinity in SA is over 50% (El-Hazmi et al., 2007; Mosli et al., 2012; Amin et al., 2012), it is unclear whether this contributes to a higher incidence of familial cancer syndromes. The aim of this work was therefore to establish whether MSI screening of young CRC patients (<60 years) could help to identify unrecognized cases of LS in SA. Screening for MSI was performed in the absence of information on the family history of cancer in these patients. A similar approach used in a Western population was shown to increase the detection rate of LS in young CRC patients 3-fold and this strategy has now been adopted as routine clinical practice in WA (Schofield et al., 2009; Schofield et al., 2014).

Somatic deletion in the BAT-26 mononucleotide repeat is a sensitive marker for MSI in CRC (Iacopetta et al., 2010). Using a rapid F-SSCP technique (Figure 1), MSI was found in 11.6% (33/284) of young CRC patients from the Eastern province of Dammam and from Riyadh. This frequency is almost identical to that reported (11.3%) in an unselected cohort of 807 CRC patients from Riyadh (Siraj et al., 2015). In comparison, the MSI frequency found in young CRC patients from WA was significantly lower (7.8%, 105/1,344, P=0.045) (Schofield et al., 2009). The higher frequency of MSI observed in young patients from SA may reflect a higher prevalence of LS compared to Western populations, however this awaits germline testing of the 32 MSI/BRAF wildtype cases found here. The anatomical distribution of MSI cases was also different. In WA, the majority of MSI tumors from young patients arise in the proximal colon (90/118, 76%) (Schofield et al., 2009). In contrast, less than half of the MSI tumors were right-sided in the current study (13/33, 39%; P<0.001) and in another study of unselected Saudi CRC patients (38/90, 42%) (Siraj et al., 2015). These results suggest there may be differences in the etiology and pathogenesis of MSI tumors between Saudi and Western populations.

The pentaplex system for evaluating MSI status (Suraweera et al., 2002) is considered superior to the Bethesda panel that includes both dinucleotide and mononucleotide repeat markers (Xicola et al., 2007). We used this system to confirm the MSI status of all 33 cases identified by F-SSCP screening with BAT-26 (Figure 2). For each of the 5 mononucleotide repeats we also compared the average length of allelic deletion and the frequency of allelic deletion between SA and WA patients. While BAT-25 and BAT-26 showed a similar pattern of somatic deletion between the two cohorts, the NR-21, NR-24 and MONO-27 markers were deleted less frequently (Figure 3B), thus accounting for the shorter average length of deletions (Figure 3A). A sizeable minority of the Saudi MSI cases (9/33, 27%) showed no deletions in any of the NR-21, NR-24 and MONO-27 markers, whereas these were each deleted in at least 95% of MSI cases from WA. To our knowledge, such geographic differences in the frequency of allelic deletion have not been previously reported for MSI tumors. Our findings confirm that BAT-26 is an excellent marker for MSI screening because the large size of deletions (average of 9-10 bp in WA and SA CRC) facilitate detection.

The presence of BRAF mutation is used to distinguish MSI tumors that are sporadic in origin from those that may be LS (Domingo et al., 2004). Of the 33 MSI cases identified in this cohort of young Saudi CRC patients, only 1 (3%) contained a BRAF mutation, therefore excluding it as a potential LS case. A very low frequency of BRAF mutation (2.5%, 19/757) was also reported by Siraj et al (2014) in an unselected cohort of Saudi CRC patients.

DOI:http://dx.doi.org/10.7314/APJCP.2016.17.4.1917

Lynch Syndrome Screening in Young Colorectal Cancer Patients from Saudi Arabia Using MSI as the Initial Test

Figure 6. Immunohistochemical Staining for the Mismatch Repair Proteins MSH2 (A and B) and MLH1 (C and D). A and C show areas of positive staining in adjacent normal mucosa, while B and D show complete lack of staining of the respective MMR protein. Magnification: 200x and 500x (inset)
patients, whereas a higher BRAF mutation frequency (7/105, 6.7%) was reported for MSI cases from young Australian CRC patients (Schofield et al., 2009). The single BRAF mutation detected here showed a very low allelic frequency (Figure 4), suggesting the presence of tumor heterogeneity. The remaining 32 MSI cases with wildtype BRAF are candidates for germline mutation in MMR genes as a cause of their MSI phenotype.

The presence of BRAF mutation in CRC is mutually exclusive with that of KRAS mutation (Li et al., 2006). Because of the low BRAF mutation frequency observed here and in a previous study of Saudi CRC (Siraj et al., 2015), we investigated whether this was compensated by a higher frequency of KRAS mutation. The KRAS mutation frequency observed here (27%, 9/33) was similar to that reported in a study of unselected Saudi CRC patients (30%) (Beg et al., 2015), but less than reported in a study of young Saudi CRC patients (40%) (Elsamany et al., 2014). We are not aware of any reports that have investigated the frequency of KRAS mutation specifically in MSI CRC tumors.

MSI is almost always accompanied by loss of expression of one or more MMR proteins, usually as MLH1/PMS2 or MSH2/MSH6 combinations. This information is important for directing germline mutation analysis to the appropriate gene(s). In the present study, all but 3 of the 30 MSI cases examined by IHC showed loss of expression of MMR proteins. In our earlier study, 2 of 97 MSI cases showed no apparent loss of expression (Schofield et al., 2009), while a Saudi study also found discordant IHC results in 2 of 73 MSI tumors (Siraj et al., 2015). Such rare discordant cases may be explained by failure of the MMR gene mutation to alter protein antigenicity, or by mutation of a different MMR gene family member as the cause of MSI. In the present study we found a similar proportion of MLH1/PMS2 and MSH2/MSH6 double loss cases compared to Australian MSI cases (9 and 7 vs 42 and 30, respectively) (Schofield et al., 2009), but a higher percentage with loss of PMS2 only (37% vs 7%, respectively). Tissue was not available in 3 of the 11 cases with PMS2 loss to test for concomitant loss of MLH1. Nevertheless, this result suggests that PMS2 mutations may be more frequent in the Saudi population.

In conclusion, we have shown that screening with the BAT-26 mononucleotide repeat marker is an efficient way to detect MSI in archival tumor samples from Saudi CRC patients. The MSI frequency was significantly higher compared to a similar age cohort of Australian CRC patients. Analysis of results obtained with 5 mononucleotide repeat markers revealed novel differences in allelic deletions between MSI tumors from Saudi and Australian patients. The anatomic distribution of MSI cases was also markedly different, with a more even distribution of MSI cases between the left and right colon in Saudi patients. Our results confirm the low frequency of BRAF mutations in MSI tumors reported earlier in unselected Saudi CRC patients. This marker allows exclusion of rare MSI/BRAF mutant patients from further follow up as possible LS cases.

The next step in this work is to follow up surviving MSI patients and to perform germline mutation analysis of MMR genes in individuals who give consent for genetic testing and following appropriate pre-test counselling. The IHC results should help to inform this analysis. The incidence of LS amongst Saudi CRC patients was recently estimated to be approximately 1% (8/807) (Siraj et al., 2015). This is similar to the incidence of 0.8-1% reported in CRC cohorts from Western populations (Samowitz et al., 2001; Pinol et al., 2005; Schofield et al., 2009; Ward et al., 2013). The present study may help to justify the introduction of routine MSI screening of young CRC patients as a cost-effective way (Snowsill et al., 2015) to identify LS in the Saudi population.

Acknowledgements

The authors would like to thank Dr Dalal Altamimi, Dr Mohamed El-Shawarby, Dr Ramez Atassi, Ahmed Alotaibi, Saber Alzahrani, Clara Kuti, Edison Tan, Abdulrahman Hazzazzi, Hala Almadani and Shaker Ahmed for their technical support and provision of clinical data. We are also grateful to the Anatomical Pathology Department at King Fahad University Hospital, Dammam University, the Histopathology Department at Dammam Regional Laboratory & Blood Bank, Ministry of Health, and the Department of Pathology and Laboratory Medicine, King Fahad Specialist Hospital-Dammam.

References


Lynch Syndrome Screening in Young Colorectal Cancer Patients from Saudi Arabia Using MSI as the Initial Test


Screening for Lynch syndrome in young Saudi colorectal cancer patients using microsatellite instability testing and next generation sequencing

Masood Alqahtani1,2 · Caitlin Edwards3 · Natasha Buzzacott3 · Karen Carpenter3 · Khalid Alsaleh4 · Abdulmalik Alsheikh4 · Waleed Abozeed4,5 · Miral Mashhour2 · Afnan Almousa2 · Yousef Housawi2 · Shareefa Al Hawwaj2 · Barry Iacopetta1

Abstract  Individuals with Lynch syndrome (LS) have germline variants in DNA mismatch repair (MMR) genes that confer a greatly increased risk of colorectal cancer (CRC), often at a young age. Identification of these individuals has been shown to increase their survival through improved surveillance. We previously identified 33 high risk cases for LS in the Saudi population by screening for microsatellite instability (MSI) in the tumor DNA of 284 young CRC patients. The aim of the present study was to identify MMR gene variants in this cohort of patients. Peripheral blood DNA was obtained from 13 individuals who were at high risk of LS due to positive MSI status and young age (<60 years at diagnosis). Next generation sequencing, Sanger sequencing and Multiplex Ligation-dependent Probe Amplification were used to screen for germline variants in the MLH1, MSH2, MSH6 and PMS2 MMR genes. These were cross-referenced against several variant databases, including the International Society for Gastrointestinal Hereditary Tumors Incorporated database. Variants with pathogenic or likely pathogenic significance were identified in 8 of the 13 high risk cases (62%), comprising 4 in MLH1 and 4 in MSH2. All carriers had a positive family history for CRC or endometrial cancer. Next generation sequencing is an effective strategy for identifying young CRC patients who are at high risk of LS because of positive MSI status. We estimate that 7% of CRC patients aged <60 years in Saudi Arabia are due to LS, potentially involving around 50 new cases per year.

Keywords  Lynch syndrome · Colorectal cancer · Microsatellite instability · Screening · Saudi

Introduction

Lynch syndrome (LS), formerly known as hereditary non-polyposis colorectal cancer, is the most common inherited form of colorectal cancer (CRC). LS is associated with a lifetime risk of approximately 80% for the development of CRC, as well as an increased risk for several other cancer types including endometrial, gastric, ovarian and kidney tumors [1]. Clinical features of LS include an early age of onset of CRC (average of 45 years) and the occurrence of multiple related cancers within families [2]. The genes responsible for LS encode DNA mismatch repair (MMR) proteins and the majority of pathogenic variants occur in MLH1 and MSH2 [3]. There are no readily defined morphological features that distinguish CRCs arising in LS patients from non-familial cases. However, the large majority of tumors from LS patients are characterized by ubiquitous changes in the length of DNA microsatellite repeats due to a defective DNA MMR system. This is referred to as microsatellite instability (MSI) and can be used to help identify CRC patients with LS [4]. The loss of expression of one
or more MMR proteins in tumor cells, as detected by immunohistochemistry (IHC), often also indicates the presence of an underlying pathogenic germline variant in MMR genes. Epigenetic changes to MMR genes may also account for some families with no apparent pathogenic germline variants but which nevertheless show loss of tumor expression of MMR proteins and MSI [3].

The proportion of LS cases amongst all CRC patients has been estimated to be up to 5%, however this incidence is likely to vary from country to country [5]. The so-called Amsterdam clinical criteria were proposed as a guide to help detect LS [6]. Families must have experienced at least three cases of CRC over at least two generations, with at least one case diagnosed before the age of 50 years. However, the difficulty in obtaining a full family history of cancer and the lack of accessible genetic counselling and testing has meant that many LS families remain undetected, even in countries with advanced health care systems [7]. The Bethesda criteria were subsequently proposed to increase the sensitivity of detection by incorporating the molecular screening of tumor DNA for MSI in suspected cases of LS [8, 9]. Close surveillance and management of individuals with pathogenic germline variants in DNA MMR genes is cost-effective [10] and colonoscopies performed at 3-year intervals reduce the mortality from CRC by as much as 70% [11, 12].

We previously reported the results of a state-wide screening program to detect LS in the West Australian population [13–15]. This was based on routine testing for MSI in CRC from all young CRC patients (<60 years at diagnosis), irrespective of information on family cancer history. Using our approach, the rate of detection of LS families in this population increased by approximately three-fold [15]. We estimated that LS accounted for slightly less than 1% of all CRC in the West Australian population. This concurs with the incidence reported by several other large, hospital-based studies in Western countries that also employed MSI testing of tumor DNA as the initial screen [16–18].

The average age of CRC patients in Saudi Arabia at the time of diagnosis is approximately 15 years younger than in Western populations [19]. Approximately one quarter of CRC patients in Western Australia are <60 years at diagnosis [20], whereas slightly more than half of all CRC patients in Saudi Arabia are aged <60 years at diagnosis [21]. We recently reported that almost 12% of CRC diagnosed in young (<60 years) Saudi patients showed a positive tumor MSI status [22]. Because this was approximately 50% more than observed in Australian CRC patients from the same age cohort [13], we hypothesized the incidence of LS may also be higher in Saudi CRC patients. In the present study we therefore performed next generation (massively parallel) sequencing of germline DNA from Saudi CRC patients identified in our earlier study [22] as being at high risk of LS due to their young age and positive MSI status.

**Methods**

Tumor tissue from 284 Saudi national patients aged <60 years at diagnosis of CRC were previously screened for MSI as described earlier [22]. These comprised all patients who met the age criteria and who were diagnosed with primary CRC between 2006 and 2015 at three hospitals in Dammam (n = 191; King Fahad Specialist Hospital, King Fahad University Hospital) and Riyadh (n = 93; King Khaled University Hospital). In all, 33 cases showed MSI (11.6%) and were also wildtype for *BRAF*, thus defining them as “red flag” and therefore strong candidates for LS. Ethics approval for patient follow up was obtained from two referral hospitals (King Khaled University Hospital, IRB no: E-15-1468; King Fahad Specialist Hospital-Dammam, IRB no: LAB0305). Of the 33 red flag cases, 9 had since died, 8 could not be contacted and 3 did not give consent for genetic testing. The remaining 13 patients gave consent and provided 10 ml of peripheral blood for germline DNA analysis. Information on family history of cancer was also obtained for the red flag cases that could be contacted. The workflow used in this and our previous study [22] to identify LS cases based on initial screening of young CRC patients for MSI is shown in Fig. 1.

DNA was extracted from peripheral blood using an automated DNA extractor and kit as per the manufacturer's instructions (Cat# 03730972001, MagNA Pure Compact, Roche, USA). Approximately 50 ng of DNA was processed

![Fig. 1 Flowchart outlining the strategy for identifying patients with Lynch syndrome from within a large cohort of young Saudi CRC patients. The initial MSI screening, IHC and *BRAF* mutation testing was described in an earlier report [22]]
using the Illumina TruSight Cancer Panel (Illumina Inc. San Diego, CA, 92122, USA) and sequenced on the MiSeq system (Illumina) according to the manufacturer’s instructions. This protocol provides comprehensive sequence coverage of 94 cancer-related genes, including the DNA mismatch repair genes MLH1, MSH2, MSH6 and PMS2. DNA was fragmented and tagged, followed by the application of sequencing adaptors and indices by PCR. Sample libraries were denatured with subsequent hybridization to labeled probes specific to the targeted region. Several rounds of hybridization and enrichment of these DNA fragments were performed prior to PCR amplification. The products were then pooled and loaded onto the MiSeq (Illumina) for cluster generation and 2 × 150-bp paired-end sequencing.

Data obtained from the MiSeq was analysed via BaseSpace (https://basespace.illumina.com) to generate BAM and VCF files, allowing alignment against the Burrows-Wheeler Aligner 1 (BWA) and variant calling by GATK. The depth of coverage was assessed by the coverageBed tool (v2.14.2). Regions with a depth of <20 reads were detected by the Cancer_coverage_v2.py script and noted for gap filling by Sanger sequencing. VCF files were imported into Cartegenia (http://www.agilent.com) for variant annotation and cross-matching of variants against the dbSNP, Human Gene Mutation Database (HGD) 2014.2, COSMIC (v71) and NCB1 ClinVar databases, in addition to the population databases ESP6500, 1000 Genomes and ExAC. Further interrogation, including BAM file assessment, was performed using Alamut Visual (v2.7.2), which also provided functional prediction analysis by Align GVGD, SIFT, MutationTaster and PolyPhen2. In addition, variants were cross-referenced to the International Society for Gastrointestinal Hereditary Tumors Incorporated (InSiGHT) database. All reported variants were confirmed by Sanger sequencing and classified according to the current ACMG guidelines [23].

Multiplex Ligation-dependent Probe Amplification (MLPA) of MLH1, MSH2, MSH6 and PMS2 was performed according to the manufacturer’s instructions (MRC-Holland, Amsterdam). This was targeted to genes that demonstrated loss of MMR gene expression, as detected by IHC staining of tumor tissues carried out previously [22]. DNA was denatured and subsequently hybridized to the MLPA probes followed by a ligation reaction. PCR was performed and the amplified products were separated by capillary electrophoresis. Data was analysed using the Coffalyser.Net software (v.140721.1958). All reported variants were confirmed by repeat MLPA using a confirmation probe-mix and classified according to the current ACMG guidelines [23].

Sanger sequencing using a long-range PCR method was also performed for patients showing loss of IHC staining for PMS2 in their tumor tissue. This method avoids amplification of the pseudogene [24]. Briefly, PMS2 was amplified in 3 segments by long range PCR using the TaKaRa enzyme, interspersed by exons 6 and 10, which were then amplified by usual PCR methods. These segment products were subsequently used for nested PCR and sequencing of each individual exon.

Results and discussion

Eight of the 13 red flag cases (62%) tested for germline variants in MMR genes using next generation sequencing were found to harbor pathogenic or likely pathogenic variants according to ACMG guidelines: 4 in MSH2 and 4 in MLH1 (Tables 1, 2). The average age of patients with these variants was 44.9 years at diagnosis and there was an even gender distribution (4 males and 4 females). All 4 cases with MSH2 variants showed loss of tumor expression for both MSH2 and MSH6, while 3 cases with MLH1 variants showed loss of both MLH1 and PMS2 expression and the other showed a partial loss of PMS2 only (A3). Amsterdam II criteria were met in 4 cases (A3, A4, A5, A8), while a fifth case (A2) was borderline with two family members being diagnosed with CRC at age 50 years. All 13 cases evaluated here met the Bethesda criteria by virtue of being aged <60 years at diagnosis and having tumors that showed MSI and loss of MMR gene expression.

Four of the 13 red flag cases revealed no significant MMR gene variants (A7, A9, A10, A11) and none of these met Amsterdam II criteria (Tables 1, 2). Interestingly, one showed loss of PMS2 only (A7), another showed only partial loss of PMS2 (A9), while another showed focal loss of MLH1 and PMS2 (A11). Only one of these four cases included another family member who had been diagnosed with CRC or endometrial cancer (A11).

The final case, A1, showed an MSH2 variant of unknown pathogenic significance (class 3), no family history of cancer, and tumor loss of MLH1 and PMS2 rather than MSH2 (Tables 1, 2). However, in silico analysis provided evidence for a deleterious functional effect associated with this missense variant (Table 3).

Seven of the 9 germline variants found here in young Saudi CRC patients were previously reported in the InSiGHT database as pathogenic variants, while 6 were reported in the ClinVar database (Table 2). Although not yet recorded in either of these databases, the variant found in case A2 was recently reported in a study of Saudi CRC patients [25]. In silico functional prediction analyses confirmed the deleterious and disease-causing effects of the 4 missense variants (A3, A4, A6, A8) previously reported in the InSiGHT and ClinVar databases (Table 3).

Based on our earlier observation that 11.6% of 284 Saudi CRC patients aged <60 years were MSI positive
and that 62% (8/13) of these were now subsequently found to harbor a pathogenic germline variant (Fig. 1), we estimate that approximately 7% of CRC cases arising in young Saudi patients are due to LS (i.e. 62 × 11.6%). The number (Fig. 2a) and incidence (Fig. 2b) of CRC in this population have increased approximately four-fold and two-fold, respectively, over the past 20 years [21]. The most recent data from the Saudi National Cancer Registry shows that almost 720 CRC cases were diagnosed nationwide in the young (<60 years) Saudi population in 2013 [21].

Table 1 Clinical characteristics of 13 young Saudi CRC patients with MSI positive tumors tested for germline variants in MMR genes

<table>
<thead>
<tr>
<th>Case</th>
<th>Age at Dx</th>
<th>Sex</th>
<th>IHC result</th>
<th>MMR gene variant</th>
<th>Family history of cancer</th>
<th>Amsterdam II</th>
<th>Bethesda</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>40</td>
<td>M</td>
<td>Loss of MLH1</td>
<td>MSH2</td>
<td>None</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>A2</td>
<td>50</td>
<td>F</td>
<td>Loss of MSH2</td>
<td>MSH2</td>
<td>Mother died of CRC at 64 years</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>A3</td>
<td>48</td>
<td>M</td>
<td>Partial loss of PMS2</td>
<td>MLH1</td>
<td>Brother dx with CRC at 51 years</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>A4</td>
<td>47</td>
<td>M</td>
<td>Loss of MLH1</td>
<td>MLH1</td>
<td>Two sisters dx with CRC</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>A5</td>
<td>34</td>
<td>F</td>
<td>Loss of MSH2</td>
<td>MSH2</td>
<td>Patient dx with breast cancer at 42 years</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>A6</td>
<td>55</td>
<td>F</td>
<td>Loss of MLH1</td>
<td>MLH1</td>
<td>Niece died of endometrial cancer at 42 years</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>A7</td>
<td>55</td>
<td>F</td>
<td>Loss of PMS2</td>
<td>None found</td>
<td>Father died of brain cancer at unknown age</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>A8</td>
<td>53</td>
<td>M</td>
<td>Loss of MSH2</td>
<td>MSH2</td>
<td>Mother died of endometrial cancer at 74 years</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>A9</td>
<td>41</td>
<td>M</td>
<td>Partial loss of PMS2</td>
<td>None found</td>
<td>Uncle died of unknown cancer at 60 years</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>A10</td>
<td>30</td>
<td>F</td>
<td>Loss of MSH2</td>
<td>None found</td>
<td>None</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>A11</td>
<td>45</td>
<td>M</td>
<td>Focal loss of MLH1</td>
<td>None found</td>
<td>Sister diagnosed with CRC at 43 years</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>A12</td>
<td>29</td>
<td>M</td>
<td>Loss of MLH1</td>
<td>MLH1</td>
<td>Mother dx with polyps at &lt;60 years</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>A13</td>
<td>43</td>
<td>F</td>
<td>Loss of MSH2</td>
<td>MSH2</td>
<td>Sister dx with CRC at 40 years</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*Variant of unknown pathogenicity

Table 2 Summary of MMR gene variants and their reporting in InSiGHT and ClinVar databases

<table>
<thead>
<tr>
<th>Case</th>
<th>Variant</th>
<th>Variant type</th>
<th>Variant classa</th>
<th>InSiGHT</th>
<th>ClinVar</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>NM_000251.2(MSH2): c.[2262_2267del];[=] p.[(Ser755_Thr756del)];[=]</td>
<td>In-frame deletion</td>
<td>4</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>A5</td>
<td>NM_000251.2(MSH2): c.[367-?_1276+?del];[=] p.[(?)];[=]</td>
<td>Large deletion</td>
<td>5</td>
<td>P</td>
<td>NR</td>
</tr>
<tr>
<td>A12</td>
<td>NM_000249.3(MLH1): c.[454-?_545+?del];[=] p.[(??)];[=]</td>
<td>Large deletion</td>
<td>5</td>
<td>P</td>
<td>P</td>
</tr>
</tbody>
</table>

NR not reported, P pathogenic

aAccording to ACMG guidelines [23]
bAlso reported by Siraj et al. [25]
cThis variant has been shown to cause aberrant splicing rather than a missense change in the protein

[22] and that 62% (8/13) of these were now subsequently found to harbor a pathogenic germline variant (Fig. 1), we estimate that approximately 7% of CRC cases arising in young Saudi patients are due to LS (i.e. 62 × 11.6%). The number (Fig. 2a) and incidence (Fig. 2b) of CRC in this population have increased approximately four-fold and two-fold, respectively, over the past 20 years [21]. The most recent data from the Saudi National Cancer Registry shows that almost 720 CRC cases were diagnosed nationwide in the young (<60 years) Saudi population in 2013 [21].

© Springer
Therefore, our results suggest that about 50 CRC cases per year in Saudi Arabia can be attributed to LS (7% of 720), with this number likely to keep increasing. Moreover, this estimate should be considered as a minimum since some cases almost certainly occur in patients aged >60 years at diagnosis.

The two-fold increase in CRC incidence evident in young Saudis over the past 20 years (Fig. 2b) may in part be due to improved diagnostic strategies and patient ascertainment. However, the more likely causes are changes in dietary and lifestyle practices to resemble those found in Western countries with a high CRC incidence, including increased consumption of meat and rising levels of obesity. The four-fold increase in the overall number of CRC cases arising in young individuals over this period (Fig. 2a) also reflects the rapidly increasing population of Saudi Arabia. Of note is the almost identical incidence of CRC between young (<60 years) male and female Saudis (Fig. 2b). This contrasts with the approximately 60:40 gender distribution observed in Australian CRC patients from the same age group, respectively [13].

The age of CRC patients in Saudi Arabia (Fig. 3a) is considerably younger than that of CRC patients from Western Australia (Fig. 3b), with a mean age at diagnosis of 55 years compared to 70 years, respectively [20, 21]. Using the same strategy of MSI and BRAF screening of a very large (n = 1344) and unselected cohort of primary CRC, we previously estimated that 3.6% of all CRC arising in young (<60 years) CRC patients in Western Australia was attributable to LS [13]. We therefore conclude the contribution of LS to CRC is approximately two-fold higher in young patients from Saudi Arabia compared to a Western population (7 vs. 3.6%).

Siraj et al. [25] recently investigated a cohort of 807 unselected CRC patients from Riyadh, Saudi Arabia, for the prevalence of LS. These workers observed an almost identical incidence of MSI to our earlier study of young Saudi cases [22] (11.3 vs. 11.6%, respectively). However,

---

**Table 3** In silico functional prediction analysis of variants

<table>
<thead>
<tr>
<th>Case</th>
<th>Variant</th>
<th>Align GVGDb</th>
<th>SIFT</th>
<th>MutTaster</th>
<th>PolyPhen2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>NM_000251.2(MSH2): c.[737A &gt; T];[=] p.[(Lys246Ile)];[=(=)]</td>
<td>C35</td>
<td>Deleterious</td>
<td>Disease causing</td>
<td>Possibly damaging (0.924)</td>
</tr>
<tr>
<td>A2</td>
<td>NM_000251.2(MSH2): c.[2262_2267del];[=] p.[(Ser755_ Thr756del)];[=(=)]</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>A3</td>
<td>NM_000249.3(MLH1): c.[62C &gt; T];[=] p.[(Ala21Val)];[=(=)]</td>
<td>C65</td>
<td>Deleterious</td>
<td>Disease causing</td>
<td>Probably damaging (0.998)</td>
</tr>
<tr>
<td>A4</td>
<td>NM_000249.3(MLH1): c.[1961C &gt; T];[=] p.[(Pro654Leu)];[=(=)]</td>
<td>C65</td>
<td>Deleterious</td>
<td>Disease causing</td>
<td>Probably damaging (1.0)</td>
</tr>
<tr>
<td>A5</td>
<td>NM_000251.2(MSH2): c.[367-?_1276+?del];[=] p.[(?)];[=(=)]</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>A6</td>
<td>NM_000249.3(MLH1): c.[677G &gt; A];[=] p.[(Arg226Gln)];[=(=)]</td>
<td>C0</td>
<td>Deleterious</td>
<td>Disease causing</td>
<td>Probably damaging (0.991)</td>
</tr>
<tr>
<td>A8</td>
<td>NM_000251.2(MSH2): c.[2089T &gt; C];[=] p.[(Cys697Arg)];[=(=)]</td>
<td>C65</td>
<td>Deleterious</td>
<td>Disease causing</td>
<td>Probably damaging (1.0)</td>
</tr>
<tr>
<td>A12</td>
<td>NM_000249.3(MLH1): c.[454-?_545+?del];[=] p.[(?)];[=(=)]</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>A13</td>
<td>NM_000251.2(MSH2): c.[2038C &gt; T];[=] p.[(Arg680*)];[=(=)]</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

bClassification ranges from Class 0 (C0) to Class 65 (C65), with C0 less likely to interfere with function and C65 most likely to interfere with function

---

**Fig. 2** Changes in the total number of cases (a) and incidence (b) of CRC in young (<60 years) patients in Saudi Arabia over the past two decades. The results shown are derived from the Saudi national cancer registry [21].
they reported a lower prevalence of germline variants in their overall CRC cohort (0.99%, 8/807) compared to the present estimate of 7% in a young cohort of Saudi CRC patients. This discrepancy could in part be due to the different ages of the study cohorts, as well as to differences in the methodology used for the analysis of genetic variants. The bioinformatics and variant classification strategy that we applied are conservative and therefore unlikely to explain the higher variant incidence observed in the present cohort. The relatively low prevalence of LS amongst Saudi CRC patients reported by Siraj et al. is in fact similar to results from a number of previous studies in Western countries, including our own [13, 16–18]. These studies were also based on initial MSI screening of tumors and found that 0.8–1% of all CRC cases were due to LS.

A national MSI screening program for young CRC patients in Saudi Arabia, together with IHC and BRAF mutation analysis for the resulting MSI positive cases, should enable the large majority of newly diagnosed LS-associated cases to be identified. This would allow a Saudi national LS registry to be established, similar to other national [26] or state [15] programs. The benefits of such registries include the identification of LS families and the encouragement of long-term participation in surveillance programs for high risk individuals. Registration and regular colonoscopic surveillance has been demonstrated to substantially reduce the mortality from CRC in affected families [26].

In summary, we have used MSI testing of primary tumors from young Saudi CRC patients as an initial screen for the detection of LS. Following IHC and BRAF mutation testing of MSI positive cases, we performed next generation sequencing and MLPA in order to identify patients with germline variants in MMR genes. All eight patients with pathogenic variants identified in this study showed a positive family history for LS-related cancers, with four meeting Amsterdam II criteria. We estimate the prevalence of LS in young CRC patients (<60 years) is two-fold higher in Saudi Arabia compared to Western Australia (7 vs. 3.6%, respectively). If this incidence is confirmed by further studies in large patient cohorts, our results suggest that at least 50 CRC cases arise each year in Saudi Arabia because of LS. Identification of these individuals and of other affected family members, in conjunction with close surveillance, should reduce the mortality from CRC. We conclude that MSI testing of young CRC patients followed by next generation sequencing and MLPA is a feasible and cost-effective strategy to identify Saudi families with LS.

Acknowledgements The authors gratefully acknowledge the financial support provided by King Fahad Specialist Hospital-Dammam and the Saudi Arabia Cultural Mission of the Royal Embassy. We also extend our appreciation to Ahmed Alotaibi, Elhassan Khalafaalla, Elhassan Taha, Clara Kutty and the HIL laboratory for their technical support and use of laboratory facilities. The authors are grateful to Dr. Lyn Schofield for helpful discussions.

References

nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst 96:261–268

 Springer
Subject Informed Consent for genetic research

Title of the Study Proposal:
Screening for Lynch Syndrome in the Saudi population

A. Purpose of the Research:
You are invited to participate in a research project that involves genetic testing to possibly identify a gene mutation that is linked to Lynch Syndrome, such as colorectal cancer, endometrial, ovarian, gastric and renal. The research involves genetic testing on a sample of your tissue, DNA, or blood. The reason for this invitation is that you have been diagnosed with LS related cancer.

Genes are units of information inside the cells of our bodies, handed down from parent to offspring that determine the structure of our individual bodies for traits like hair, color, eye color, tendency for disease, and intelligence. Genes may also carry information about any medical conditions or diseases you may have or carry.

Researchers have identified certain genetic disorders that could predispose patients to cancers, like Lynch Syndrome. This syndrome happens if there is a mutation in one or all MMR gene. Lynch Syndrome increases the risks of certain cancers, mainly colorectal cancer.

Currently, diagnosis is made using standard clinical criteria (Bethesda, Amsterdam, or Duke), however these clinical criteria might not be sensitive in identifying familial cancer syndromes.

As result, we need to find a process that enables us to identify and diagnose those patient who potentially could have familial cancer syndromes, to intervene early and avoid complications of delayed therapy.

The most accurate way to diagnose familial cancer syndromes, like Lynch syndrome, might be via genetic testing. Since you have been diagnosed with colorectal cancer in the past, you might have this particular syndrome. Participating in this study might help finding out if you have it or not.

Version 3, 7/6/2016
B. Description of the Research:

The study involves genetic testing for Lynch syndrome through a 5 ml blood sample drawn at the clinic by a qualified phlebotomist. Your DNA will be extracted from the blood sample at the laboratory of KFSHD under supervision of the primary investigator, Mr. Masood Alqahtani. Part of this sample will be sent to the genetic lab in Australia. This sample contains no identifier (name, medical record no., Saudi ID). The result, whether positive or negative, will be discussed with you by your treating physician Dr. .................. (specialty of the physician) at hospital........... Based on that, you might be referred to the specialists on GI tract or oncology for proper management and screening, since you already have an existing medical record.

C. Potential Risks and Discomforts:

The risks and discomfort may be related to the blood drawing procedures. It may cause discomfort from the needle stick, bruising, or infection. Infection at the needle stick site is very rare. You may feel dizzy or faint, your doctor will provide all treatment and assistance for you during the sample withdrawing.

Anxiety is common while waiting for the result or after knowing the result. The research team will support you with what you need since you are an existing patient in the hospital.

There is a possibility of loss of privacy or breech of confidentiality. We will take measures to reduce this risk, such as assigning a study number to your data that is collected for the study.

If a potential employer learns that you have a high risk of developing colorectal cancer because you have Lynch Syndrome, you may lose your job or have difficulties getting a job. If you have a health insurance, you may lose it or have the premium raised.

D. Potential Benefits:

If you do have Lynch syndrome, you would benefit from frequent screening that takes place more than the regular screening as per guidelines. This screening would detect recurrence of the cancer early on (as Lynch syndrome significantly increases risk of recurrence). You will be tested for other diseases and tumors associated with this syndrome.
However, you may not get direct benefit but the information gained from this research may help scientists and doctors to learn more about this condition.

E. Alternative to Participation:

You are free to participate in this study. Alternative is not to participate and follow the routine screening process.

F. Termination of Participation:

If you prefer to stop your participation in this research, you may ask the doctor to destroy any record of your participation in this research and to destroy any sample with your name on it or that may be linked to you. Your identity will be removed from all data and research records. However, data obtained from your sample prior to your withdrawal may be used for publication and other research purposes after your personal information have been removed.

G. Compensation / Treatment:

In the event of any injury resulting from participation in the study, hospital......will offer you the needed care at no cost to you, since you have an existing medical record.

H. Voluntary Participation:

Participation in this study is voluntary. You will suffer neither penalty nor loss of any benefits to which you are otherwise entitled should you decide not to participate.

Withdrawal from this research study will not affect your ability to receive medical care available at hospital......

If you wish, significant new findings developed during the course of the research study that might be reasonably expected to affect your willingness to continue to participate in the research study will be provided to you.

I. Confidentiality:

The confidentiality of your information will be kept as per the policy of the hospital. A special number (code) will be assigned to your sample. Only the research team will have access to your information. There will be no identifying data.
We will not provide the results of your sample to any other party in such a way that you may be identified. If the results are published in a medical journal, or presented at a medical meeting, no identifying information will be included.

The results of this research will be kept separate from your general medical records.

Your identity and medical record will remain confidential with respect to any publications of the results of this study. Your medical record in connection with this study will be kept confidential to the extent permitted by the law. Furthermore, the IRB in accordance with applicable laws and regulations may review your medical record.

The results of this study will not be linked to you personally or to your family. The results (may/ may not) be linked to your ethnic or social background such as your tribe, nationality, or social status.

By signing this consent form you are authorizing such access to your medical records.

K. Contact Person(s):
You may call the hospital……..Health Research or Institutional Review Board (IRB) on tel +966 ............At ext ..... for general questions concerning research at the hospital or research subjects’ rights. For any specific questions with regard to this study, or in the event of a research-related injury, please contact Dr. ............ +966..................

لن يتمكشف نتائج الأبحاث المستخدمة فيها عيناتك لأي شخص بحيث يتم التعرف عليه، وإذا نشرت نتائج هذه الدراسة في مجلات علمية أو قدمت في مؤتمرات طبية، فإن تحتوي على معلومات شخصية تتعلق بك.

ستبقى دائماً نتائج هذا البحث محظوظة بطريقة منفصلة عن سجلك الطبي العام بالمستشفى.

ستظل عيناتك وسجلاتك الطبية، كمشارلك في هذه الدراسة سرية في أي منشورات أو تقارير ذات صلة بنتائج هذا البحث، كما سيبقى سجلك الخاص بهذه الدراسة سرياً في الحدود التي يسمح بها القانون. وكذلك يمكن الاسئلة على هذا السجل من قبل مجلس تقييم الأبحاث وذلك في حدود النظم والقوانين المطبقة بهذا الخصوص.

إن نتائج هذه الدراسة لن تكون فيها ما يعرف به شخصك أو أحد من أفراد أسرتك. من غير الممكن أن يُعرف بها أسمك العربي، أو خلفيتك الاجتماعية مثل القبيلة، أو الجنسية، أو حالتك الاجتماعية.

بتوفرك لهذه الموافقة المستنيرة فانك تسمح لفريق البحث الوصول لمواردك الطبية المتوفرة في سجلك الطبي.

م. الأشخاص الذين يمكن الاتصال بهم:

في حالة وجد أي مشكلة في استخدام الخدمة في مركز الأبحاث الصحية بمدينة الملك خالد الطبية أو مجلس مراجعة البحوث (IRB) على هاتف رقم +966............ + تحويلة رقم ...... أو توجه أي أسئلة عامة تصل بالبحث بالمستفيض أو تتعلق بحقوق المشارك.

في حالة وجود أسئلة محددة تتعلق بهذا البحث أو في حالة حدوث أي إصابات تصل بالدراسة، نرحب بالاتصال على الهاتف +966...........
I confirm that I have read, or had read to me, the foregoing authorization and that all blanks or statements requiring completion were properly completed before I signed.

Research Subject or Surrogate Signature:
Print Name: _____
Date: _____
Signature: ______________________________
Relationship: ____________________________

(If signed by person other than the research subject)

Investigator
I have fully explained to the above volunteer/relative/surrogate the nature and purpose of the above-mentioned research project, including the fact that the studies will not result in any direct therapeutic benefit and the extent to which the studies are experimental. I have also explained the possible complications which may arise from both known and unknown causes as a result and the consequences and risks, if the volunteer decides to discontinue participation. It is my belief that he/she understands the nature, purposes, and risks of these studies before he/she signs this informed consent. I have fully and completely answered all questions to the best of my ability. I have also offered to answer any questions relating to these studies that may arise in the future.

Signature: ______________________________
Print Name: ____________________________
Date: _________________________________
PART II: Authorization for Genetic Research

1. AUTHORIZATION OF VOLUNTARY PARTICIPANT WHO IS NOT EXPECTED TO OBTAIN ANY DIRECT BENEFIT

Name of Research Participant: ____________
MRN number or full address: ____________

I hereby volunteer to participate in a research program under the supervision of Dr. ____________ and his/her associates at Hospital ____________, that will involve:

Identifying whether I carried a genetic mutation that led to Lynch Syndrome which is a familial cancer or not. By this screening other family members can be screened as a preventive strategy to avoid late diagnosis.

2. INFORMATION RESULTING FROM THE RESEARCH PROJECT

Please tick (✓) and initial all of the following:

a. I would like for Hospital ____________ investigators to contact me and give me a summary of the results of this research project.
   ① Yes  ② No Initials: ______________

b. I would like for Hospital ____________ investigators to contact me and inform me of my own results (and/or family results) in this research project.
   ① Yes  ② No Initials: ______________

3. THE USE OF MY SAMPLE IN OTHER RESEARCH PROJECTS

Please tick (✓) and initial all of the following:

a. I would like for Hospital ____________ investigators to contact me to obtain my permission for any research project using my sample, if the Institutional Review Board approves such research.
   ① Yes  ② No Initials: ______________

b. I would like for my sample to be used for other research projects, without contacting me, if such projects are directly related to this research project, as judged by the Institutional Review Board.
   ① Yes  ② No Initials: ______________

c. I would like for my sample to be used for other research projects, without contacting me, even if such projects are related, but not directly so, to this research project, as judged by the Institutional Review Board.
   ① Yes  ② No Initials: ______________

الجزء الثاني: الموافقة

1 الموافقة على المشاركة في الدراسة التي لا تتوقع أن تكون لها فائدة مباشرة للمشارك

اسم المشارك في الدراسة: ____________
رقم السجل الطبي أو العنوان بالكامل: ____________

أتطوع للمشاركة في هذه الدراسة تحت كلا من الدكتور: ____________
والمشاركين معه في ____________, والتي تتضمن: تحديد ما إذا كان الورم المصاب/المصابي به هو أحد أنواع الأورام المصاحبة لمتلازمة لينش الوراثية والتي تثبت العلم أن الخلل الوراثي المصاحب لمتلازمة لينش قد ينتقل لأحد أفراد العائلة.

2 المعلومات الناتجة من الدراسة

الرجاء وضع علامة (✓) على أحد الخيارات التالية والتوقيع:

أرغب من الباحثين في ____________ التواصل معي وإعطائي ملخص عن نتائج الدراسة.

3 استخدام عيناتي في مشاريع بحثية أخرى

الرجاء وضع علامة (✓) على أحد الخيارات التالية والتوقيع:

أرغب من الباحثين في ____________ الحصول على موافقتى قبل استخدام عيناتي في أي دراسة أخرى وافتم عليها لجنة تقييم الأبحاث.

أرغب في استخدام عيناتي في دراسات أخرى دون الاتصال بي، وإن كانت هذه الدراسات لها علاقة مباشرة بهذه الدراسة وذلك إذا قصت لجنة تقييم الأبحاث بذلك.

أرغب في استخدام عيناتي في دراسات أخرى دون الاتصال بي، حتى إذا كانت هذه الدراسات لها علاقة بهذه الدراسة وإن كانت علاقة غير مباشرة، وإن كانت قصت لجنة تقييم الأبحاث بذلك.
d. I would like for my sample to be used for any other approved research projects without contacting me, as long as I cannot be linked, in any way, to the sample.

☐ Yes ☐ No Initials: ........................................

4. THE USE OF MY SAMPLE BY OTHER INVESTIGATORS FROM OTHER INSTITUTIONS:

Please tick (✓) and initial all of the following:

I agree that my sample can be used by investigators outside the Hospital……….. as long as such use is approved by the Institutional Review Board.

☐ Yes ☐ No Initials: ........................................

5. MONETARY/FINANCIAL GAIN FROM RESEARCH

Please initial: ....................................................

I understand that Hospital……….. may obtain monetary/financial Gain from the use of my sample in research and that such gain will not be shared with me.

Initials: .......................................................
vii I understand that, in order to provide the data by which to measure the effectiveness of these studies, Dr. and his/her associates may carry out certain routine preliminary diagnostic procedures that have been fully described and explained to me. Should these indicate any abnormality, my participation in the aforementioned studies will not be allowed. I am unaware of any preexisting medical or emotional problem that would make it unwise for me to participate in these studies.

ix I understand that I am free to withdraw this authorization and discontinue participation in these studies at any time. The consequences and risks of such withdrawal during the course of the studies have been explained to me. I understand that such withdrawal will not affect my ability to receive any medical care made necessary by the performance of such studies or to which I might be otherwise entitled.

xii grant this consent as a voluntary contribution in the interest of medical research.

xiii I confirm that I have read, or had read to me, the foregoing authorization and that all blanks or statements requiring completion were properly completed before I signed.

Research Subject or Surrogate Signature:
Print Name: _____
Date: _____
Signature: _______________________
Relationship: _____
(IF SIGNED BY PERSON OTHER THAN THE RESEARCH SUBJECT)

9. WITNESS/TRANSLATOR
I confirm that I have accurately translated and/or read the information to the subject:

Witness:
Print name: _____
KFSHD ID #: _____
Signature: _______________________
Date: _____

Whoever this document is written in English is to be translated and/or read to the subject, and the translation is to be signed by the translator/witness.

 Witness/Translator
Print name: _____
KFSHD ID #: _____
Signature: _______________________
Date: _____

Note: The witness/translator must be someone other than the research subject.

vii أفهم أنني أقدم معلومات يمكن من خلالها تقدير كفاءة وفعالية هذه الدراسة، فكل من الدكتور..... أو أحد المشاركين في الدراسات التخيفية التمهيدية بعد أن يتم توضيحها لي تفصيلًا، وأنه إذا كانت ملاحظات هذه الإجراءات غير طبيعية فإنه لن يسمح لي بالمشاركة في الدراسات المذكورة سابقاً.

ix أفهم بأنني مطلق الحرية لسحب هذا الموافقة الخطية وإنهاء مشاركتي في هذه الدراسات في أي وقت أشاء وقد تم شرح جميع العواقب والمخاطر المرتبطة على انسحاب من هذه الدراسة. كما أفهم أن هذا الانسحاب لن يؤثر على حقي في تلقي العناية الطبية اللازمة، والموافقة المقدمة في إنجاز هذه الدراسة التي أستحقها في الأحوال العادية.

xii أمنح هذه الموافقة من الدراسة متطوعًا في تطوعي في الإسهام في البحوث الطبية.

xiii أؤكد بأنني قد قرأت أو قرأت لي هذه الموافقة الخطية، وأن كل الفراغات والاقترارات قد تم تعبئتها قبل توقيعني عليها.

توقيع المريض أو ولي الأمر:
الاسم: _______________________
التاريخ: _______________________
التوقيع: _______________________
صلة القرابة: _______________________
( إذا كان الموقع غير المريض المشارك )

13. الشاهد / المترجم
أقر بأنه قد قرأت أو ترجمت للمشارك هذه المعلومات بشكل صحيح:
الاسم: _______________________
رقم البطاقة: _______________________
التاريخ: _______________________
التوقيع: _______________________

10. Investigator or Delegate
I have fully explained to the above volunteer/relative/surrogate the nature and purpose of the above-mentioned research project, including the fact that the studies will not result in any direct therapeutic benefit and the extent to which the studies are experimental. I have also explained the possible complications which may arise from both known and unknown causes as a result and the consequences and risks, if the volunteer decides to discontinue participation. It is my belief that he/she understands the nature, purposes, and risks of these studies before he/she signs this informed consent. I have fully and completely answered all questions to the best of my ability. I have also offered to answer any questions relating to these studies that may arise in the future.

Co-Investigators
Signature: __________________________
Print Name: Dr. ............
Employee I.D. Number: 
Date: 

Co-Investigators
Signature: __________________________
Print Name: 
Employee I.D. Number: 
Date: 

أقر بأنني قد شرحت للمتطوع/ل قريب/أو ولي أمره المذكور أعلاه بصورة كاملة طبيعة وأهداف الدراسة المذكورة. و أنه لا توجد فائدة مباشرة له أو إلى أي مدى هي دراسة تجريبية. وقد شرحت له أيضاً المضاعفات المحتملة حولها من جراء المشاركة في هذه الدراسة سواء كانت لأسباب معروفة أو غير معروفة، و العواقب و المخاطر المترتبة إذا ما قرر المتطوع إنهاء مشاركته بالدراسة. في اعتقادي أنه قد فهم/فهمت طبيعة الدراسة و الغرض منها و المخاطر الناتجة عنها و ذلك قبل توقيعه / توقيعها على الموافقة الخطية. و لقد قمت بتوضيح استعدادي للإجابة على أي أسئلة تتعلق بهذه الدراسة، و قمت فعلا بالإجراء الشفاف على جميع الأسئلة التي سئلت. وقد عرضت عليه استعدادي التام للإجابة على أي أسئلة تتعلق بهذه الدراسات في المستقبل.

ممثل الباحث
التوقيع: __________________________
الاسم: د ... 
رقم بطاقة المستشفى: 
التاريخ: 

ممثل الباحث
الباحث أو ممثله
التوقيع: __________________________
الاسم: 
رقم بطاقة المستشفى: 
التاريخ: 
Family history and pedigree

Cancer is a common illness in our community; however, only about 5% of cancers are due to an inherited predisposition. In order to provide you with the best genetic counselling and subsequently refer you unique specialist, it is very important to obtain as much information as possible about your family history of cancer. We guarantee that all released information by you will be treated as a confidential information and won’t be shared with anyone without your written approval. Also, the information will not be used to contact other family members.

Please feel free to contact Dr...... for any general questions concerning research at Hospital .......... or research subjects’ rights at +966 ............ Ext:....... For any specific questions with regard to this study, or in the event of a research-related issue, please contact Dr. ........... on telephone # +966..... Or +966 11 467......Ext......

Personal details

Did you hear about a research done on colorectal cancer?
What did you hope to gain by participating in this research?

Your personal history of cancer/ polyps/tumours

Have you had any polyps, cancers or tumours?

<table>
<thead>
<tr>
<th>Date of Diagnosis</th>
<th>Type of Cancer/Polyp/Tumour</th>
<th>Hospital Name and Address</th>
<th>Name of Doctor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please sign here if you give your consent for us to access medical records confirming your diagnosis.

(If you receive this form electronically, you may need to sign this form when you attend for an appointment.)

Signature: ________________________________ Date: ____________________
My Family Pedigree

Include other ancestors other than the one logged in the below family pedigree, please list the sex parental or maternal association and type of cancer.

*Also make sure to write the date of the onset for each affected family member*
If a relative has had polyps, cancer, or other tumours write the details in this table. Please include as much specific information as possible.
If the cancer started in one place and spread, please indicate where it started.
If there was more than one cancer, please list them all.
If the details are unknown please write? In the space provided.

<table>
<thead>
<tr>
<th>Name (first and last)</th>
<th>Date of birth /Age</th>
<th>Relationship to you</th>
<th>Maternal/ Paternal? M / P</th>
<th>Type of polyps/cancer/tumour breast left/right/both?</th>
<th>Age &amp; year of diagnosis</th>
<th>Alive? Y / N</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXAMPLE: Shatha</td>
<td>01-Jan-64</td>
<td>Aunt</td>
<td>M</td>
<td>Breast (left)</td>
<td>49 (2013)</td>
<td>Y</td>
</tr>
</tbody>
</table>
Your ancestry

Some conditions are more common in people with a shared background or ancestry, so it is important to be as specific as possible.

What is your family’s origin?
Your mother’s family: Your father’s family:

Is there any consanguinity between your parents

Do you have a non-Saudi ancestry?  Yes  No  Maternal / Paternal (tick one)

Is anyone in your family related by blood as well as by marriage?  Yes  No

Maybe  

Please give details:

Family genetic information

Members of your family may already have been assessed by a Genetic Service. Knowing about this may allow us to provide you with better consultation and your result might expedited.

Have any of your relatives ever attended genetic service clinic within or outside of Saudi Arabia

Yes  No

Please give details

<table>
<thead>
<tr>
<th>Name</th>
<th>Relation to you</th>
<th>Date of Birth</th>
<th>Genetic Clinic Location</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Contact details

Please feel free to contact us if you have any questions or need any assistance completing this form.

Dr. …….., FRCPC  
Masood Alqahtani  

Department of  
Principle investigator  

Institute name  
Translational Cancer research laboratory  

Address: ….Riyadh xxxx P.O.Box xxx  
University Of Western Australia (UWA)  

Telephone: (+966) xxxxxx  
Perth, Australia In collaboration with  

Ext. xxx  
Mobile:  +966 xxxxxxxxx  
Email:  

Supplementary Figure 1: Saudi population pyramid, Demographic survey, 2016 (https://www.stats.gov.sa/sites/default/files/en-demographic-research-2016_2.pdf)