Colorectal Cancer in those with a genetic susceptibility

A thesis submitted to the University of Manchester for the degree of Doctorate of Medicine in the Faculty of Medical and Human Sciences

Katherine Newton MB ChB (Manc), MRCS
2013

School of Medicine
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>10</td>
</tr>
<tr>
<td>Declaration</td>
<td>12</td>
</tr>
<tr>
<td>Copyright statement</td>
<td>12</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>13</td>
</tr>
<tr>
<td>Publications and Presentations relating to research</td>
<td>14</td>
</tr>
<tr>
<td>List of tables</td>
<td>17</td>
</tr>
<tr>
<td>List of figures</td>
<td>20</td>
</tr>
<tr>
<td>Preface</td>
<td>22</td>
</tr>
<tr>
<td>Hypothesis and aims</td>
<td>23</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>24</td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>24</td>
</tr>
<tr>
<td>1.2 Colorectal adenocarcinoma</td>
<td>25</td>
</tr>
<tr>
<td>1.2.1 Adenocarcinoma</td>
<td>25</td>
</tr>
<tr>
<td>1.2.2 Precursor lesions (polyps)</td>
<td>26</td>
</tr>
<tr>
<td>1.2.2.1 Adenomas</td>
<td>26</td>
</tr>
<tr>
<td>1.2.2.2 Serrated lesions</td>
<td>27</td>
</tr>
<tr>
<td>1.2.2.3 Hamartomas</td>
<td>27</td>
</tr>
<tr>
<td>1.2.2.4 Inflammatory polyps</td>
<td>28</td>
</tr>
<tr>
<td>1.3 Colorectal cancer molecular genetics</td>
<td>28</td>
</tr>
<tr>
<td>1.3.1 Fearon and Vogelstein’s genetic model for tumourogenesis</td>
<td>29</td>
</tr>
<tr>
<td>1.3.2 Jass’ molecular classification of colorectal cancer</td>
<td>29</td>
</tr>
<tr>
<td>1.3.2.1 Genetic Instability</td>
<td>30</td>
</tr>
<tr>
<td>1.3.2.2 DNA methylation</td>
<td>30</td>
</tr>
<tr>
<td>1.3.2.3 Jass’ classification</td>
<td>32</td>
</tr>
<tr>
<td>1.3.3 The chromosomal instability (CIN) pathway</td>
<td>35</td>
</tr>
<tr>
<td>1.3.4 The Lynch Syndrome pathway</td>
<td>37</td>
</tr>
<tr>
<td>1.3.5 The CpG Island Methylator Phenotype (CIMP) pathway</td>
<td>37</td>
</tr>
<tr>
<td>1.4 Biomarkers</td>
<td>38</td>
</tr>
<tr>
<td>1.4.1 Biomarker categories</td>
<td>39</td>
</tr>
<tr>
<td>1.4.2 Biomarker discovery and validation</td>
<td>40</td>
</tr>
<tr>
<td>1.4.2.1 Biomarker discovery</td>
<td>40</td>
</tr>
<tr>
<td>1.4.2.2 Biomarker validation</td>
<td>41</td>
</tr>
<tr>
<td>1.4.3 Biomarker discovery roadmaps</td>
<td>41</td>
</tr>
<tr>
<td>1.5 Biomarkers and colorectal cancer</td>
<td>42</td>
</tr>
<tr>
<td>1.5.1 Risk assessment (predisposition) biomarkers</td>
<td>42</td>
</tr>
<tr>
<td>1.5.1.1 Molecular biomarkers suggestive of Lynch Syndrome</td>
<td>44</td>
</tr>
<tr>
<td>1.5.1.2 Genome wide association studies (GWAS)</td>
<td>44</td>
</tr>
</tbody>
</table>
1.5.2 Screening biomarkers

1.5.2.1 Currently available screening modalities

1.5.3 Diagnostic biomarkers

1.5.4 Pharmacological biomarkers

1.5.5 Predictive biomarkers

1.5.5.1 Anti-EGFR therapy

1.5.5.1.1 KRAS

1.5.5.1.2 BRAF

1.5.5.2 Mismatch repair deficiency and 5-fluorouracil chemotherapy

1.5.6 Prognostic biomarkers

1.5.6.1 Mismatch repair deficiency

1.6 Familial colorectal cancer syndromes

1.6.1 Familial risk

1.6.2 Single gene disorders

1.6.3 Moderate familial risk

1.6.4 Low-penetrance, low risk susceptibility genes

1.7 Lynch Syndrome

1.7.1 Introduction and history

1.7.2 Molecular genetics

1.7.3 Diagnosis

1.7.3.1 Amsterdam and Bethesda

1.7.3.2 Other predictive models

1.7.4 Tumour biomarker analysis

1.7.4.1 Microsatellite instability

1.7.4.2 Mismatch repair immunohistochemistry

1.7.5 Screening

1.7.6 Treatment

1.7.6.1 Surgery

1.7.6.1 Adjuvant therapy

1.7.7 Lynch Syndrome registries

1.8 Familial adenomatous polyposis

1.8.1 Introduction and history

1.8.2 Diagnosis

1.8.3 Screening

1.8.4 Prophylactic surgery

1.8.5 Registries

1.9 St Mary’s Hospital department of medical genetics

1.10 Introduction to the three studies
Study One: MLH1 promoter region methylation and BRAF gene mutation testing - an alternative pre-screen strategy in Lynch Syndrome (HNPCC)

2.1 Background and Introduction

2.1.1 Introduction

2.1.2 Family history as a method as ascertainment

2.1.3 BRAF mutation and MLH1 promoter region methylation

2.1.4 The use of population based biomarkers

2.1.5 Applying pre-screening strategies to patients with different pre-test risk

2.1.6 Case and control selection

2.1.7 Critique of laboratory methods

2.1.7.1 Analysis of mismatch repair status

2.1.7.1.1 Immunohistochemistry

2.1.7.1.2 Microsatellite instability

2.1.7.2 BRAF and KRAS mutation testing

2.1.7.2.1 Pyrosequencing

2.1.7.3 MLH1 promoter region methylation testing

2.1.8 Aim of Study one

2.1.9 Hypothesis of Study one

2.2 Methods

2.2.1 Ethics approval

2.2.2 Participants

2.2.2.1 Cases

2.2.2.2 Controls

2.2.3 Tissue specimens

2.2.3.1 Tissue specimen preparation

2.2.3.1.1 Slide preparation (Manchester samples only)

2.2.3.1.2 Tissue preparation for DNA extraction

2.2.4 Histopathology

2.2.4.1 MMR protein IHC interpretation
2.2.5 Molecular genetics

2.2.5.1 Microsatellite instability

2.2.5.1.1 Interpretation of electropherograms produced by the ABI PRISM® 3100 genetic analyser

2.2.5.2 KRAS and BRAF gene mutation analysis

2.2.5.2.1 Interpreting pyrograms

2.2.5.3 MLH1 promoter region methylation analysis

2.2.5.3.1 Interpreting MLH1 promoter methylation pyrograms

2.2.6 Data analysis

2.3 Results

2.3.1 Mismatch repair analysis of local sporadic tumour

2.3.2 Results of BRAF mutation and MLH1 promoter region methylation in MLH1 mutation carriers and sporadic MLH1 loss cancer

2.3.2.1 BRAF mutation analysis in MLH1 mutation carriers and sporadic MLH1 loss cancers

2.3.2.2 MLH1 promoter region methylation analysis in MLH1 mutation carriers and sporadic MLH1 loss cancers

2.3.2.3 BRAF mutation analysis OR MLH1 promoter region methylation in MLH1 mutation carriers and sporadic MLH1 loss cancers

2.3.2.4 BRAF mutation analysis AND MLH1 promoter region methylation in MLH1 mutation carriers and sporadic MLH1 loss cancers

2.3.2.5 Bayes calculations

2.3.2.6 Applying BRAF and MLH1 methylation testing to patients who fulfil Amsterdam Criteria

2.3.2.6.1 Wild-type BRAF as a predictor of germline MLH1 mutation

2.3.2.6.2 Normal MLH1 promoter region as a predictor of germline MLH1 mutation

2.3.2.6.3 Wild-type BRAF OR normal MLH1 promoter region as a predictor of germline MLH1 mutation

2.3.2.6.4 Wild-type BRAF AND normal MLH1 promoter region as a predictor of germline MLH1 mutation

2.3.2.7 Applying BRAF and MLH1 methylation testing to patients who fulfil revised Bethesda criteria and have an MLH1 loss tumour

2.3.2.7.1 Wild-type BRAF as a predictor of germline MLH1 mutation

2.3.2.7.2 Normal MLH1 promoter region as a predictor of germline MLH1 mutation
2.3.2.7.3 Wild-type BRAF OR normal MLH1 promoter region as a predictor of germline MLH1 mutation
2.3.2.7.4 Wild-type BRAF AND normal MLH1 promoter region as a predictor of germline MLH1 mutation

2.3.2.8 Applying BRAF and MLH1 methylation testing to patients from the general population who have an MLH1 loss tumour
   2.3.2.8.1 Wild-type BRAF as a predictor of germline MLH1 mutation
   2.3.2.8.2 Normal MLH1 promoter region as a predictor of germline MLH1 mutation
   2.3.2.8.3 Wild-type BRAF OR normal MLH1 promoter region as a predictor of germline MLH1 mutation
   2.3.2.8.4 Wild-type BRAF AND normal MLH1 promoter region as a predictor of germline MLH1 mutation

2.3.2.9 Summary of post-test probabilities of being an MLH1 mutation carrier according to known priori risk
   2.3.2.10 Applying BRAF and MLH1 analysis to priori risk

2.3.2.11 Cost

2.3.3 Results of KRAS mutation analysis

2.4 Discussion Study one
   2.4.1 Discussion
   2.4.2 Conclusion
   2.4.3 Further work

3 Study two: Genotype-phenotype correlation in colorectal polyposis

3.1 Introduction and background
   3.1.1 Introduction
   3.1.2 Aims of study two
   3.1.3 Hypothesis

3.2 Methods
   3.2.1 Participants
   3.2.2 Data analysis

3.3 Results
   3.3.1 Onset of polyposis
3.3.2 Colorectal cancer incidence 187
3.3.2 Survival 191

3.4 Discussion 202
3.4.1 Discussion 202
3.4.2 Conclusion 206
3.4.3 Further work 207

4 Study three: Metachronous colorectal cancer risk in patients with a moderate family history 208

4.1 Introduction 208
4.1.1 Introduction and background 208
4.1.2 Aims of study three 211
4.1.3 Hypothesis of study three 211

4.2 Methods 212
4.2.1 Participants 212
4.2.2 Data analysis 214

4.3 Results 215
4.3.1 Metachronous CRC 219
4.3.2 Age of onset of CRC 227
4.3.3 Survival from diagnosis of first colorectal cancer 229
4.3.4 Actuarial survival 233

4.4 Discussion 237
4.4.1 Discussion of metachronous colorectal cancer risk 237
4.4.1.1 Metachronous colorectal cancer risk 237
4.4.1.2 Discussion of statistical methodology 240
4.4.1.2.1 Discussion of Kaplan-Meier Survival analysis Methodology 242
4.4.1.2.2 Discussion of cumulative incidence function methodology 247
4.4.2 Discussion of survival 250
4.4.3 Conclusion 252
4.4.4 Further work 252
<table>
<thead>
<tr>
<th>References</th>
<th>252</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix one</td>
<td>264</td>
</tr>
<tr>
<td>Appendix two</td>
<td>267</td>
</tr>
<tr>
<td>Appendix three</td>
<td>272</td>
</tr>
<tr>
<td>Appendix four</td>
<td>278</td>
</tr>
<tr>
<td>Appendix five</td>
<td>289</td>
</tr>
<tr>
<td>Appendix six</td>
<td>297</td>
</tr>
<tr>
<td>Appendix seven</td>
<td>298</td>
</tr>
</tbody>
</table>

Word count: 55,724
Abstract

In the UK colorectal cancer is the second most common cancer in women and the third most common in men. Colorectal cancer is a heterogeneous disease. Improved knowledge of the molecular pathways involved in colorectal carcinogenesis has yet to translate into improved clinical outcomes. This thesis investigates three different aspects of colorectal cancer caused by a genetic susceptibility. Through these three separate studies this thesis aims to understand the relationship between molecular genetics and the screening for, clinical behaviour of and outcomes following colorectal cancer.

Study 1:

Background: Lynch syndrome patients have DNA mismatch repair deficiency and up to an 80% life-time risk of colorectal cancer. Screening of mutation carriers reduces colorectal cancer incidence and mortality. Selection for germline mutation testing relies on family history (Amsterdam and Bethesda Criteria) and tumour derived biomarkers. Initial tumour biomarker analysis uses mismatch repair protein immunohistochemistry and microsatellite instability. Abnormalities in either identify mismatch repair deficiency but do not differentiate sporadic epigenetic defects, due to MLH1 promoter region methylation (13% of all CRCs), and Lynch Syndrome (4% of all CRCs). A diagnostic biomarker capable of making this distinction would be valuable. This study compared two biomarkers in tumours with known mismatch repair; quantitative methylation of the MLH1 promoter region using a novel assay and BRAF V600E mutation in identification of germline mutations.

Methods: Tumour DNA was extracted from preserved (FFPE) tumour tissue and pyrosequencing used to test for MLH1 promoter methylation and BRAF p.V600E mutation in 71 CRCs from individuals with pathogenic MLH1 mutations and 73 CRCs with sporadic MLH1 loss. Specificity and sensitivity were compared. Results: Unmethylated MLH1 promoter: sensitivity 94.4% (95% CI 86.2-98.4%), specificity 87.7% (95% CI 77.9-94.2%), Wild-type BRAF p.V600E: sensitivity 65.8% (95% CI 53.7-76.5%), specificity 98.6% (95% CI 92.4-100.0%) for the identification of those with pathogenic germline MLH1 mutations.

Conclusions: Quantitative MLH1 promoter region methylation using pyrosequencing is superior to BRAF mutation in identifying germline mutations in mismatch repair deficient tumours.

Study 2:

Background: Familial adenomatous polyposis (FAP) is a colorectal cancer predisposition syndrome caused by autosomal dominantly inherited mutations in the Adenomatous Polyposis Coli (APC) gene. It has been divided into three clinical subtypes; mild, classical and severe FAP. MutYH associated polyposis is a recently described autosomal recessively inherited condition which causes mild polyposis and an increased risk of colorectal cancer. Prior to the discovery of the MutYH gene these patients were thought to have mild FAP. The aim of this study was to investigate for a correlation between genotype (site of the APC mutation, or MutYH mutation) and phenotype.
Methods: A retrospective longitudinal study of 492 patients on the Manchester Polyposis Registry was conducted. Patients were grouped according to genotype: 0; unknown mutation, 1; APC 0-178 (and 312-412 of exon 9), 2; APC>1550, 3; APC 179-1249, 4; APC 1250-1549, 5; MutYH. Date of onset of polyposis, incidence of colorectal cancer (CRC), actuarial survival and actuarial time to surgery were calculated.

Results: Median age of onset of polyposis: genotype 0; 20.3 years, genotype 1; 35.6 years, genotype 2; 32.2, genotype 3; 15.9 years, genotype 4; 14.8 years (p<0.0001). Age and onset of CRC was similar between genotypes. Median survival: genotype 0; 56.6, genotype 1; 74.9 years, genotype 2; 61.0, genotype 3; 63.0, genotype 4; 48.1, genotype 5; 69.7 (p=0.003). This survival difference was also seen when patients who underwent screening and those who did not were analysed separately. Patients with genotype 4 (APC 1249-1549) have a significantly worse survival despite screening and early prophylactic surgery.

Conclusions: This analysis supports a genotype-phenotype correlation. Patients with a mutation APC 1249-1549 develop polyposis younger and have a worse survival. Patients with a mutation APC 0-178 or 312-412 develop polyposis later and have an improved survival. This is the first study to demonstrate this survival difference.

Study 3:

Background: Life-time risk of metachronous colorectal cancer is 0.6%-3% following sporadic colorectal cancer and 15-26% in Lynch syndrome. The life-time incidence of incident colorectal cancer in individuals with moderate familial risk is 8-17%. Risk of metachronous colorectal cancer is unknown.

Methods: A retrospective longitudinal study of the Regional Familial colorectal cancer Registry was performed. Patients who had at least one colorectal cancer were categorised as follows: moderate risk (n=383), LS (n=528) and population risk (n=409). Kaplan-Meier estimate (1-KM) and cumulative incidence function (CI) were used to calculate the risk of metachronous colorectal cancer. 1-KM gives the risk for individuals remaining at risk (alive) at a given time point thus is useful for counselling. CI gives the risk for the whole population.

Results: 1-KM and CI demonstrated that the risk of metachronous colorectal cancer was significantly higher in moderate risk patients compared with population risk (1-KM p= 0.008, CI p= <0.005). Both were lower than Lynch Syndrome. Moderate risk 1-KM was 2.7%, 6.3% and 23.5% at five, ten and 20 years. Population risk 1-KM was 1.3%, 3.1% and 7.0% at 5, 10 and 20 years and CI was 0.3%, 0.6% and 2.4%.

Conclusion: These data indicate that the risk of metachronous colorectal cancer is significantly higher in patients with a moderate family history than in those at population risk. This justifies pro-active life-long surveillance.
Declaration

No portion of the work referred to in thesis has been submitted in support of an application for another degree or qualification at this or any other university or institute of learning.

Copyright statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and she has given the University of Manchester certain rights to use such copyright, including for administrative purposes.

ii. Copies of this thesis. Either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the university has from time to time. This page may form part of any such copies made.

iii. The ownership of certain Copyright, patients, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may not be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the Relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which the disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://documents.manchester.ac.uk/library/about us/regulations) and in the University’s policy on Presentation of Theses.
Acknowledgements

The histopathology work conducted as part of this thesis was performed in the pathology laboratory at Manchester Royal Infirmary. I would like to acknowledge Catherine Keeling and Lorraine Griffiths (Research Biomedical Scientists) for their assistance.

The molecular genetics work conducted as part of this thesis was conducted in the Molecular Genetics laboratories at St Mary’s hospital. I would like to acknowledge Dr Andrew Wallace (Senior Clinical Scientist) and Natalie Jorgenson (Genetic Technologist) for their assistance with assay development, and Natalie Jorgenson for her guidance conducting the laboratory work.
Publications and presentations relating to this research

Publications

1. **Metachronous colorectal cancer risk in patients with a moderate family history.** Newton K, Green K, Laloo F, Hill J, Evans DGR. Accepted for publication in Colorectal Disease August 2012.


Submitted

5. **Tumour MLH1 promoter region methylation testing is an effective pre-screen for Lynch Syndrome.** Newton K, Jorgensen N, Wallace AJ, Buchanan D, Laloo, F, McMahon RFT, Hill J, Evans DG. *Submitted to Gastroenterology March 2012*

Presentations

INTERNATIONAL:

- **Tumour MLH1 promoter region methylation testing is an effective pre-screen for Lynch Syndrome.** Newton K, Jorgensen N, Wallace AJ, Laloo, F, McMahon RFT, Hill J, Evans DG.

  Poster presentation at Association of Surgeons of Great Britain and Ireland International Congress. Liverpool, UK. May 2012

-**ORAL presentation** at the Tripartite Colorectal Meeting. Cairns, Australia. July 3rd-7th 2011. Presented by K Newton. **Awarded a Royal Society of Medicine, Section of Coloproctology Travelling Fellowship.**

-Poster presentation at biannual meeting of InSiGHT (International Society of hereditary gastrointestinal tumours. San Antonio Texas, USA. 31st March 2nd April 2011


-Poster presentation at biannual meeting of InSiGHT (International Society of hereditary gastrointestinal tumours. San Antonio Texas, USA. 31st March 2nd April 2011

**NATIONAL:**


-Poster presentation at the British Human Genetics Conference 5-7th September 2011

• Genotype-phenotype correlation in colorectal polyposis(updated data) Newton K, Mallinson E, Lalloo F, Bowen J, Hill J, Evans DGR.

- Poster presentation at the British Society of Human Genetics Conference 5-7th September 2011


- Poster presentation at the British Society of Human Genetics Conference 5-7th September 2011


- Poster presentation at the Association of Coloproctology of Great Britain and Northern Ireland 2010 Annual Meeting. 28-30th June 2010.

REGIONAL:

- Awarded Manchester Regional Association of Surgeons Trainee presentation prize, Manchester Royal Infirmary. June 2012


- Shortlisted for Manchester Medical Society, Section of Surgery- Trainees’ Prize, Manchester University. March 2012

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1:</td>
<td>Molecular and morphological features of Jass’ five subtypes of colorectal cancer</td>
<td>34</td>
</tr>
<tr>
<td>Table 2:</td>
<td>Relative risk of CRC according to family history</td>
<td>57</td>
</tr>
<tr>
<td>Table 3:</td>
<td>Moderate risk CRC categories</td>
<td>58</td>
</tr>
<tr>
<td>Table 4:</td>
<td>Amsterdam II Criteria</td>
<td>61</td>
</tr>
<tr>
<td>Table 5:</td>
<td>Revised Bethesda Criteria</td>
<td>62</td>
</tr>
<tr>
<td>Table 6:</td>
<td>Bisulphite Conversion of un-methylated cytosine residues</td>
<td>95</td>
</tr>
<tr>
<td>Table 7:</td>
<td><em>MSI Analysis Version 1.2 (Promega, USA)</em> Locus Information</td>
<td>108</td>
</tr>
<tr>
<td>Table 8:</td>
<td>Example of <em>KRAS</em> triplicate results and their interpretation or further action required</td>
<td>121</td>
</tr>
<tr>
<td>Table 9:</td>
<td>Example of <em>BRAF</em> triplicate results and their interpretation or further action required</td>
<td>123</td>
</tr>
<tr>
<td>Table 10:</td>
<td>Results of local sporadic tumour mismatch repair deficiency testing</td>
<td>134</td>
</tr>
<tr>
<td>Table 11:</td>
<td><em>BRAF</em> mutation and <em>MLH1</em> promoter region methylation in MLH1 mutation carriers and sporadic MLH1 loss cancers</td>
<td>135</td>
</tr>
<tr>
<td>Table 12:</td>
<td>Diagnostic Test (2 by 2 tables) analysis for the identification of MLH1 mutations carriers</td>
<td>136</td>
</tr>
<tr>
<td>Table 13:</td>
<td><em>BRAF</em> V600E mutation Diagnostic test analysis</td>
<td>137</td>
</tr>
<tr>
<td>Table 14:</td>
<td><em>MLH1</em> promoter region methylation Diagnostic test analysis</td>
<td>138</td>
</tr>
<tr>
<td>Table 15:</td>
<td><em>BRAF</em> V600E mutation and <em>MLH1</em> promoter region methylation Diagnostic test analysis where EITHER test being normal is a positive result</td>
<td>139</td>
</tr>
<tr>
<td>Table 16:</td>
<td><em>BRAF</em> V600E mutation and <em>MLH1</em> promoter region methylation Diagnostic test analysis where BOTH test being normal is a positive result</td>
<td>140</td>
</tr>
<tr>
<td>Table 17:</td>
<td>Variables required for calculating post-test probabilities using Bayes theorem</td>
<td>142</td>
</tr>
<tr>
<td>Table 18:</td>
<td>Post-test probabilities of positive and negative results irrespective of whether a mutation is present</td>
<td>142</td>
</tr>
<tr>
<td>Table 19:</td>
<td>Post-test probabilities of true positive and false negative results</td>
<td>142</td>
</tr>
<tr>
<td>Table 20:</td>
<td>Pre-test probabilities (priori risk) of being an MLH1 mutation carrier</td>
<td>143</td>
</tr>
<tr>
<td>Table 21:</td>
<td>Variables required for Bayes calculations</td>
<td>143</td>
</tr>
</tbody>
</table>
Table 22: Pre- and Post-test probabilities of being an $MLH1$ mutation carrier for three priori risk groups  156
Table 23: Performance characteristics of $BRAF$ and $MLH1$ methylation analysis according to varying priori risk  158
Table 24: Codon definitions of the phenotypes  178
Table 25: Population sample used for onset of polyposis analysis  184
Table 26: Table to show the probability of the onset of polyps by completion of each decade according to genotype group  186
Table 27: Colorectal cancer incidence according to genotype category  188
Table 28: Table to show the probability of the developing a first CRC by completion of each decade according to genotype group  190
Table 29: Survival according to genotype category  192
Table 30: Comparison of survival by individual genotype category  194
Table 31: Survival in patients who did not undergo pre-scymptomatic screening  195
Table 32: Survival in patients who underwent screening  198
Table 33: Age at surgery and median time to surgery according to genotype  200
Table 34: Baseline characteristics of patients included in this study  217
Table 35: Tumour site and stage  218
Table 36: Risk of developing a metachronous colorectal cancer according to familial risk category  219
Table 37: Cumulative incidence function of metachronous CRC according to familial risk  222
Table 38: Odds ratios of metachronous CRC risk (KM and Cumulative Incidence)  224
Table 39: Age of Onset of CRC  228
Table 40: Survival from cancer diagnosis (all stage)  230
Table 41: Survival following cancer diagnosis according to Dukes stage  232
Table 42: Actuarial survival of patients who have had at least one CRC  234
Table 43: Proportion of patients surviving at completion of each decade  235
Table 44: Example dataset 1- Kaplan-Meier cumulative survival  243
Table 45: Example dataset 2- Kaplan-Meier risk estimate 243
Table 46: Example dataset 2- Kaplan-Meier risk estimate (Competing risk of death) 246
Table 47: Cumulative risk of developing a mCRC in patients with moderate familial risk 250
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1:</td>
<td>Derivation of the molecular colorectal cancer groups 1-5 based on CpG island methylator status (CIMP: high, low or negative) and microsatellite instability (MSI: high, low or stable)</td>
<td>32</td>
</tr>
<tr>
<td>Figure 2:</td>
<td>Screening biomarker roadmap</td>
<td>46</td>
</tr>
<tr>
<td>Figure 3:</td>
<td>Diagnostic biomarker roadmap</td>
<td>47</td>
</tr>
<tr>
<td>Figure 4:</td>
<td>Pyrosequencing</td>
<td>94</td>
</tr>
<tr>
<td>Figure 5:</td>
<td>Electropherogram for a microsatellite stable tumour</td>
<td>110</td>
</tr>
<tr>
<td>Figure 6:</td>
<td>Electropherogram for a MSI-high tumour</td>
<td>111</td>
</tr>
<tr>
<td>Figure 7:</td>
<td>Electropherogram for a MSI-high tumour</td>
<td>112</td>
</tr>
<tr>
<td>Figure 8:</td>
<td>KRAS exon 1 codon 12/13. Primers and mutations</td>
<td>113</td>
</tr>
<tr>
<td>Figure 9:</td>
<td>BRAF Exon 15. Primers and mutation</td>
<td>114</td>
</tr>
<tr>
<td>Figure 10:</td>
<td>Pyrogram for wild-type codon 12/13 of KRAS gene</td>
<td>115</td>
</tr>
<tr>
<td>Figure 11:</td>
<td>Pyrogram of wild-type BRAF codon 600</td>
<td>116</td>
</tr>
<tr>
<td>Figure 12:</td>
<td>Pyrogram of KRAS 34 G&gt;T Gly12Cys</td>
<td>117</td>
</tr>
<tr>
<td>Figure 13:</td>
<td>Pyrogram of KRAS variant 35 G&gt;T Gly12Val</td>
<td>118</td>
</tr>
<tr>
<td>Figure 14:</td>
<td>Pyrogram of KRAS variant 35G&gt;AGly12Asp</td>
<td>119</td>
</tr>
<tr>
<td>Figure 15:</td>
<td>Pyrogram of BRAF variant V600E</td>
<td>122</td>
</tr>
<tr>
<td>Figure 16:</td>
<td>MLH1 gene promoter region</td>
<td>126</td>
</tr>
<tr>
<td>Figure 17:</td>
<td>Pyrogram of unmethylated MLH1 promoter region of tumour DNA</td>
<td>128</td>
</tr>
<tr>
<td>Figure 18:</td>
<td>Pyrogram of methylated MLH1 promoter region of tumour DNA</td>
<td>130</td>
</tr>
<tr>
<td>Figure 19:</td>
<td>Schematic of APC gene</td>
<td>179</td>
</tr>
<tr>
<td>Figure 20:</td>
<td>Consort diagram of population used to assess the impact of genotype on the onset of polyposis</td>
<td>181</td>
</tr>
<tr>
<td>Figure 21:</td>
<td>Consort diagram for population used to assess the impact of genotype on onset of CRC</td>
<td>182</td>
</tr>
<tr>
<td>Figure 22:</td>
<td>Consort diagram for the population used to assess the impact of genotype on survival</td>
<td>182</td>
</tr>
<tr>
<td>Figure 23:</td>
<td>Kaplan-Meier curve demonstrating probability of onset of polyposis according to mutation</td>
<td>185</td>
</tr>
</tbody>
</table>
Figure 24:  Kaplan-Meier curve demonstrating probability of developing first CRC according to genotype  189
Figure 25:  Kaplan-Meier curve demonstrating difference in survival according to genotype  193
Figure 26:  Kaplan-Meier curve demonstrating survival in patients who did not undergo pre-symptomatic screening  196
Figure 29:  Kaplan-Meier curve demonstrating survival in patients undergoing screening  199
Figure 28:  Kaplan-Meier curve demonstrating time to surgery according to genotype  201
Figure 29:  Consort diagram for population used to assess the incidence of metachronous colorectal cancer in patients with a moderate familial risk  216
Figure 30:  Kaplan-Meier curve demonstrating the probability of developing a metachronous colorectal cancer according to familial risk  221
Figure 31:  Cumulative incidence function of metachronous CRC (mCRC) according to familial risk  223
Figure 32:  Kaplan-Meier curve demonstrating the probability of developing a metachronous colorectal cancer according to Lynch Syndrome mutation (all)  225
Figure 33:  Cumulative incidence function of metachronous CRC according to Lynch Syndrome mutation  226
Figure 34:  Cumulative probability of overall survival from diagnosis with first colorectal cancer  231
Figure 35:  Actuarial survival of patients who have had at least one colorectal cancer according to familial risk  236
Figure 36:  Risk estimate for mCRC in patients with moderate familial risk- Kaplan-Meier and Cumulative Incidence function compared  249
Preface

The author of this thesis graduated from Manchester University in 2004. She undertook her basic surgical training in London and was then appointed as a core surgical trainee in General Surgery in the North West. Following nine months of core training she undertook the post of Clinical Research Fellow at Central Manchester University Hospitals Trust, where research was carried out into various aspects of genetic susceptibility to colorectal carcinoma. During her time as Clinical Research Fellow, the author was appointed to specialty training (StR3) in general surgery in the North Western Deanery. This appointment was deferred until completion of this research. The author intends to pursue a career as a general surgeon, with a sub-specialty interest in colorectal surgery.

During the period of research that has culminated in this thesis, the author has gained experience in project planning and execution, including successful ethics committee application. She has gained valuable experience in both histopathology and molecular genetics laboratory techniques which were obtained through research, course attendance and working alongside genetic technologists in the regional molecular genetics laboratory at St Mary’s Hospital. Data analysis and statistical methods were learned and applied. During the period of research, the author attended several international research meetings regarding the establishment of an International Mismatch Repair (IMMR) Consortium, as the representative of the Manchester Intestinal Genetics Group (Central Manchester based gastrointestinal genetics research group). This new consortium was formed as a group of clinical, academic and laboratory researchers with an interest in taking part in large international collaborative studies into Lynch Syndrome. These meetings resulted in a successful application to the International group, the Colon Cancer Family Registry, for use of biospecimens as part of this thesis. The author’s role as representative to the IMMR Consortium will continue. The author has published a review article and data from this research in peer reviewed journals, and has presented at both national and international meetings.

The author has gained knowledge about the molecular genetics of colorectal carcinoma and genetic susceptibility that will be invaluable throughout her career.
She hopes to continue her involvement in research with the Manchester Intestinal Genetics Group (established during her time as Clinical Research Fellow).

**Hypothesis and Aims**

Colorectal cancer is a heterogeneous disease due to differences at the molecular level. Inherited factors play a role in this heterogeneity. Patient management should be affected by knowledge of an individual’s tumour biology. The hypothesis of this thesis is that inherited genetic and epigenetic factors alter the molecular biomarker profile and natural history of colorectal cancer.

There are three studies within this thesis. These studies investigate three separate aspects of inherited predisposition to colorectal. The studies each aim to understand how a different aspect of molecular genetics or genetics affects the clinical behaviour and outcomes of colorectal cancer and in turn may influence screening, management and surveillance. The first study investigates the use of a tumour derived biomarker in the diagnosis of Lynch Syndrome. The aim was to demonstrate that this biomarker can effectively select patients for Lynch Syndrome germline mutation testing in order to improve the diagnosis of Lynch Syndrome. The second study investigates if the loci of germline APC gene mutation in patients with Familial Adenomatous Polyposis (and MutYH mutation in MutYH associated polyposis) may influence the clinical behaviour of and outcomes from polyposis, and whether this should influence screening and management. The aim was to demonstrate that genotype should be used as an aide to clinical decision making. The third study investigates the clinical behaviour of colorectal cancer (risk of a metachronous or second colorectal cancer) in patients with a moderately increased risk due to their family history. The aim was to demonstrate that patients with a moderate family history of colorectal have increased risk of a metachronous colorectal cancer.
1 Introduction and background to the studies

1.1 Introduction

Worldwide over one million people develop colorectal cancer (CRC) each year. In 2007, 38,600 new cases were diagnosed in the UK (http://info.cancerresearchuk.org/cancerstats/types/bowel/Bowel Accessed 2011). Of these around 50% can be expected to die of systemic or locally advanced disease within five years of diagnosis (Weitz, Koch et al. 2005). In the UK, colorectal cancer is the second most common cancer in women (after breast) and the third most common in men (after prostate and lung). The life-time incidence is one in 20 for women and one in 16 for men (http://info.cancerresearchuk.org/cancerstats/ Accessed 2010). The incidence has remained static over the last decade. Over the last 30 years, advances in screening programmes, surgical techniques, adjuvant/neoadjuvant therapy and surveillance programmes, have improved the five year survival. It is now twice what it was in the 1970s.

At the molecular level colorectal cancer is a heterogeneous disease. This heterogeneity translates into differences in disease progression, survival, and response to chemotherapeutic agents. Genomics and proteomics have allowed the recognition and improved understanding of this heterogeneity. Advances in the understanding of the molecular biology of CRC have led to the identification of many putative biomarkers. Some of these have been validated and are currently in clinical use. It is logical to hope that the widespread clinical use of biomarkers will herald the next major advance in the diagnosis and management of colorectal cancer and hence improve patient survival.

A biomarker can be defined as an objectively measurable substance that can be used as an indicator of a biological or pathological state. A biomarker must facilitate an improvement in life expectancy or quality of life in order to be of use in clinical practice. In colorectal cancer, biomarkers can be used to predict risk of developing CRC, to screen for CRC, to diagnose primary cancer or recurrence, to assess the
pharmacokinetics of drug therapies, to predict the response to adjuvant therapies, and to predict prognosis.

A large study of cancer incidence in nearly 45,000 twin pairs suggested that genetic susceptibility may play a role in up to 35% (95% confidence interval 10-48%) of all colorectal cancers.(Lichtenstein, Holm et al. 2000) High risk single gene mutations (e.g. Adenomatous polyposis coli gene mutation in Familial Adenomatous Polyposis) account for around 3-5% (Aaltonen, Johns et al. 2007; Cairns, Scholefield et al. 2010). The aetiology of the remainder is not fully understood but is thought to be due to inheritance of a combination of low risk, low penetrance cancer susceptibility genes.(Houlston, Webb et al. 2008) Individuals with a greatly increased risk of colorectal cancer can be identified by family history, phenotypic recognition of one of the polyposis or hamartomatous syndromes, molecular features of the cancer of the index patient, or a relative, suggestive of a familial syndrome or detection of a germline mutation. The single gene disorders are inherited in a Mendelian fashion thus are either autosomal dominant or recessive in nature, and all have high penetrance. The single gene disorders that are currently documented are: Familial Adenomatous Polyposis (FAP), Lynch syndrome (or Hereditary Non-Polyposis Colorectal Cancer: HNPCC), Peutz-Jeugers Syndrome (PJS), Juvenile Polyposis Syndrome (JPS) and MYH (MutY Homolog (E. Coli)) associated polyposis syndrome (MAP).

Appropriate surveillance of individuals with these high risk genetic disorders significantly reduces the incidence of CRC and improves life expectancy (Dunlop 2002).

### 1.2 Colorectal adenocarcinoma

#### 1.2.1 Adenocarcinoma

Over 90% of all colorectal cancer is colorectal adenocarcinoma. Colorectal adenocarcinoma is an epithelial malignancy that originates within the glandular
epithelium of the colorectal mucosa. Colorectal adenocarcinoma is an invasive cancer that develops as progressively dysplastic cells within a precursor lesion then invades the muscularis mucosae into the submucosa and beyond. These cells frequently develop the ability to invade local lymphatic vessels and blood vessels, and hence are able spread locally (transcoelemic) and metastasise to distant sites, for example to the liver and lungs.

1.2.2 Precursor lesions (polyps)

Colorectal polyps are mucosal lesions that project into the lumen of the colon or rectum. They can be classified according to their histological morphology, and may be sessile (flat) or pedunculated (on a stalk). The four types of colorectal polyps are adenomas, serrated polyps (which include hyperplastic polyps and sessile serrated adenomas), hamartomas and inflammatory polyps.

1.2.2.1 Adenomas

Adenomas are benign tumours of glandular epithelium. Proliferation within the adenoma is described as degrees of atypia or dysplasia. Dysplasia is characterized by the microscopic abnormalities of poikilocytosis (abnormally shaped cells; caused in part by nuclear elongation), anisocytosis (different sized cells), presence of mitotic figures (an abnormally high proportion of cells which are currently dividing) and hyperchromatism (an increased darkness of colour caused by increased nuclear density). The architecture of adenomas may be tubular, villous or tubulo-villous. Tubular adenomas have more than 75% of their epithelium arranged in tube like fashion which when cut across appear like transected gun barrels. Villous adenomas have more than 50% of their epithelium arranged in finger-like projections like that of small intestine epithelium. Tubulovillous adenomas have both tubular and villous components. The villous component makes up between 25-50% of the polyp and the remainder is tubular. Of all polyps that are resected, approximately 75% are found to be adenomas. Adenomas are the pre-cursor lesions described in the adenoma-carcinoma sequence of Fearon and Vogelstein’s “A genetic model for colorectal tumorigenesis” (Fearon and Vogelstein 1990). The lead time from an adenoma to adenocarcinoma (in the general population) is accepted as around 10
years, however risk of progression can be stratified according to the size and number of adenomas (Cairns, Scholefield et al. 2010). Endoscopic colonic examination and resection of adenomatous polyps is recommended. The timing of surveillance is determined by the size and number of lesions (Cairns, Scholefield et al. 2010).

1.2.2.2 Serrated lesions

Serrated polyps can be hyperplastic polyps or sessile serrated adenomas. Hyperplastic polyps are benign polyps that morphologically demonstrate serrated (saw-like) epithelium, architectural atypia (non-uniformity of crypt structure and orientation) and crypt dilatation. The majority of these are small, innocuous lesions occurring in the distal colon. However, over recent years, evidence has emerged for a less benign type of serrated polyp termed sessile serrated adenoma (SSA). SSAs are morphologically similar to hyperplastic polyps, but are larger, with more atypical architecture and increased proliferation, and are found more commonly in the right colon. These lesions have been found in association with CRCs (Jass 2007; Young, Jenkins et al. 2007). The development of a CRC from a SSA is thought to be through an intermediate dysplastic lesion (or mixed polyp i.e. serrated polyp with dysplasia within it). The dysplasia of these lesions is different to that of dysplastic adenomas. The dysplasia of serrated lesions demonstrates ovoid nuclei with preserved polarity, prominent nuclear membrane due to chromatin condensation, and reduced nuclear: cytoplasm ratio due to mucin production. SSAs are thought to develop into serrated carcinomas which are morphologically different (though not completely distinct) from typical colorectal adenocarcinomas (Jass 2007).

1.2.2.3 Hamartomas

Hamartomas are mucosal lesions that contain overgrowth of mesenchymal elements such as vascular tissue and fibrous stroma. They commonly also demonstrate glandular proliferation. These polyps are almost entirely benign. The autosomal dominant conditions of Peutz-Jeghers syndrome (PJS) and Juvenile Polyposis syndrome (JPS) cause gastrointestinal hamartomatous polyposis (amongst other phenotypic features). These hamartomas have malignant potential
and result in colorectal cancer in 70% of PJS patients with polyps (Gammon, Jasperson et al. 2009) and 39% of JPS patients with polyps (Brosens, van Hattem et al. 2007).

1.2.2.4 Inflammatory polyps

Inflammatory polyps are associated with conditions such as Crohn’s disease and ulcerative colitis. They are also found in infective conditions such as amoebiasis, schistosomiasis and bacillary dysentery. These polyps form as a result of the regenerative response to local inflammation. They are stable lesions which may or may not recede as the inflammation subsides. These polyps do not exhibit dysplasia and do not have a malignant potential.

1.3 Colorectal cancer molecular genetics

In health, cell division and differentiation are tightly regulated in order to maintain the normal function of systems within the body. Signalling pathways, both intracellular and extracellular, exert this control. Malignant cells have escaped this normal physiological control.

Multiple sequential genetic and epigenetic changes must take place within a single cell line in order for a successful cancer to develop. Some of these genetic alterations may be inherited, but the majority are somatic. Genomic and epigenomic instability causes these events to occur with increasing likelihood and frequency. “This accelerates the neoplastic evolutionary process” (Worthley and

1 Epigenetic refers to heritable changes in DNA (heritable within one cell line and occasionally across generations) which alter gene expression but do not alter the sequence of DNA (for example DNA hypermethylation).

2 Somatic refers to all the cells of the body that are not gametes and are therefore diploid (contain two copies of every chromosome). The term somatic mutation refers to mutations that occur within a somatic cell line rather than being inherited.
Leggett 2010). This instability results in mutations driving carcinogenesis to occur within one clonal lifespan.

1.3.1 Fearon and Vogelstein’s genetic model for tumourogenesis

For many years it was presumed that CRC was a homogenous disease. It was thought that cancers developed along a single molecular pathway, with the same mutations occurring in a fairly linear, predictable fashion. It was thought that \textit{APC} (\textit{adenomatous polyposis coli}) mutation was the early event that triggered neoplasia and adenoma formation, and that eventually more aggressive subclones were created by \textit{KRAS} (\textit{Kirsten rat sarcoma virus}) and \textit{TP53} (\textit{tumour protein 53}) mutations. In 1990 Fearon and Vogelstein described a genetic model of colorectal tumourogenesis. They described a step-wise progression from normal epithelium, to hyperproliferative epithelium, to early then late adenoma, and on to carcinoma then metastasis. It was suggested that the majority of CRCs develop within pre-existing benign adenomas as a progressive acquisition of genetic defects within a monoclonal cell line, on the background of polyclonal field changes (Fearon and Vogelstein 1990). This model included chromosomal instability, \textit{APC} gene loss or mutation, DNA global hypomethylation, \textit{KRAS} gene mutation, long arm of chromosome 18(q) loss resulting in deletion of the \textit{DCC} gene (\textit{deleted in colon cancer}) and \textit{TP53} gene mutation (Fearon and Vogelstein 1990).

1.3.2 Jass’ molecular classification of colorectal cancer

In the 20 years since Fearon and Vogelstein’s model, it has become clear that colorectal carcinoma develops through a multitude of pathways that may or may not be mutually exclusive. Advances in genomics, proteomics, transcriptomics and metabolomics have demonstrated that there are diverse pathways involved in the development of CRC. These translate into different molecular phenotypes. It is now widely accepted that CRC is a heterogeneous disease. The different molecular pathways that cause colorectal cancer overlap between inherited and acquired cancers. In 2007 Jeremy Jass, a British pathologist, outlined a far more detailed
model of the different pathways to colorectal carcinogenesis (Jass 2007). Jass suggested that there are five pathways. The main distinguishing features described are the type of genetic instability within the cancer and the presence or absence of DNA hypermethylation.

1.3.2.1 Genetic instability

Genetic instability can either be chromosomal or microsatellite in nature. Chromosomal instability (CIN) is the predominant form and results in aneuploidy (uneven number of chromosomes or amount of chromosomal material). This is thought to be initiated by alterations within genes involved in monitoring the cell cycle, chromosome condensation, microtubule formation and function, and chromatid adhesion (Nowak, Komarova et al. 2002). Microsatellites are repeating sequences of nucleotides. The repeating base pair sequence is usually no more than ten nucleotides long (e.g. CCA of a microsatellite sequence CCACCACCACCACCACCACCACCACCACCACCACCACCA). Microsatellites are frequently found in non-coding areas of DNA. The total length of the microsatellite repeats are hugely variable between individuals but should remain the same within the DNA of every cell of an individual. These nucleotide repeats are particularly susceptible to copy errors on the background of a poorly functioning DNA mismatch repair system. Microsatellite instability (MSI) is a change in length of the microsatellite alleles due to deletion or insertion of repeating units during DNA replication. For example (CCA)^n where n varies between alleles. Microsatellites frequently occur in the promoter regions of tumour suppressor genes, for example BAX and IGF2R (Jass 2007). When variants develop within these regions, the function of the promoter region may be changed. If so, this results in altered or reduced gene expression in much the same way as if the DNA sequence of the gene contained a mutation.

1.3.2.2 DNA methylation

DNA methylation is the result of the addition of a methyl group to the fifth nitrogen of a cytosine base. DNA methylation contributes to the regulation of gene transcription and is necessary for normal cell function and cellular differentiation.
Methylation is involved in normal genomic processes such as X-chromosome inactivation (where one X chromosome is inactivated in females) and imprinting (a phenomenon where certain genes are expressed only from one allele - the imprinted allele is silenced by methylation so only the non-imprinted allele is available for transcription). Hypermethylation of the promoter regions of tumour suppressor genes is a common occurrence in many cancers. Methylation occurs at cytosine residues, particularly in CpG dinucleotide (cytosine-guanine dinucleotide) enriched regions of DNA that are less than 500 base pairs in length. CpG islands are areas of this DNA that have a cytosine-guanine content of greater than 55%. These CpG islands are particularly found in regulatory regions of genes and can be found at the 5’ end (i.e. promoter region) of roughly 50% of all human genes (Toyota, Ahuja et al. 1999). When methylation occurs in the promoter region of tumour suppressor genes, it essentially causes the tumour suppressor gene to be switched off so the effect is the same as if the gene itself was mutated (Worthley and Leggett 2010). The cause of this methylation is unclear. Certainly advanced age has an influence, but there also seems to be other factors involved particular to carcinogenesis (Toyota, Ahuja et al. 1999). Cancers that display tumour suppressor gene methylation are described as CpG island methylator phenotype or CIMP. There are degrees of this methylation which are described as CIMP-high or CIMP-low. These methylation patterns seem not to be simply quantitatively different. CIMP-high cancers have global hypermethylation of many markers suggestive of an increase in de novo methylation. CIMP-low cancers have denser methylation of fewer gene promoter regions. This pattern is thought to be caused by the spread of methylation from an origin point, or methylation centre, due to some unknown epigenetic defect (Jass 2007).
1.3.2.3 Jass’ Classification

Jass described five subtypes of CRC. (See Figure 1)

Figure 1: Derivation of the molecular colorectal cancer groups 1-5 based on CpG island methylator status (CIMP: high, low or negative) and microsatellite instability (MSI: high, low or stable) (Jass 2007).

Type 1: MSI-high, chromosome stable, methylation of MLH1 promoter region, CIMP high, BRAF (V-RAF Murine sarcoma viral oncogene homolog B1) gene mutation. Originates from serrated polyps not adenomas. This pathway is thought to account for 13% of all CRC.

Type 2: MSI-stable or low, chromosome stable, CIMP-high, partial methylation of MLH1 promoter region, BRAF gene mutation. Originates from serrated polyps not adenomas. This pathway is thought to account for 8% of all CRC.

Type 3: MSI-stable or low, chromosomal instability, CIMP-low, KRAS mutation, MGMT (O-6-methylguanine-DNA methyltransferase) promoter region methylation.
Originates in adenomas or serrated polyps. This pathway is thought to account for 20% of CRC.

Type 4: MSI-stable, chromosomal instability, CIMP-negative. Originates in adenomas. This pathway is thought to account for 57% of all CRC which includes those of Familial adenomatous polyposis (FAP) and of MutYH associated polyposis (MAP) (See Section 1.3.3).

Type 5: MSI-high, chromosome stable, CIMP-negative, BRAF mutation negative. Originates in adenomas. This pathway is thought to account for all the CRC of Lynch Syndrome, so is thought to be roughly 3% of all CRC.
Table 1: Molecular and morphological features of Jass’ five subtypes of colorectal cancer (Jass 2007).

<table>
<thead>
<tr>
<th>Feature</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSI Status</td>
<td>H</td>
<td>S/L</td>
<td>S/L</td>
<td>S</td>
<td>H</td>
</tr>
<tr>
<td>Methylation</td>
<td>+++</td>
<td>+++</td>
<td>++</td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td>Ploidy</td>
<td>Dip&gt;An</td>
<td>Dip&gt;An</td>
<td>An&gt;Dip</td>
<td>An&gt;Dip</td>
<td>Dip&gt;An</td>
</tr>
<tr>
<td>APC</td>
<td>+/-</td>
<td>+/-</td>
<td>+</td>
<td>+++</td>
<td>++</td>
</tr>
<tr>
<td>KRAS</td>
<td>-</td>
<td>+</td>
<td>+++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>BRAF</td>
<td>+++</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TP53</td>
<td>-</td>
<td>+</td>
<td>++</td>
<td>+++</td>
<td>+</td>
</tr>
<tr>
<td>Location</td>
<td>R&gt;L</td>
<td>R&gt;L</td>
<td>L&gt;R</td>
<td>L&gt;R</td>
<td>R&gt;L</td>
</tr>
<tr>
<td>Gender</td>
<td>F&gt;M</td>
<td>F&gt;M</td>
<td>M&gt;F</td>
<td>M&gt;F</td>
<td>M&gt;F</td>
</tr>
<tr>
<td>Precursor lesion</td>
<td>SP</td>
<td>SP</td>
<td>SP/AD</td>
<td>AD</td>
<td>AD</td>
</tr>
<tr>
<td>Serration</td>
<td>+++</td>
<td>+++</td>
<td>+</td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td>Mucinous</td>
<td>+++</td>
<td>+++</td>
<td>+</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>Poor differentiation</td>
<td>+++</td>
<td>+++</td>
<td>+</td>
<td>+</td>
<td>++</td>
</tr>
</tbody>
</table>


It is current theory that three main molecular pathways to colorectal carcinogenesis exist. This theory is ever evolving as the understanding of molecular CRC carcinogenesis improves. This theory of three pathways is a progression from Fearon and Vogelstein’s original genetic model of colorectal tumourogenesis (Fearon and Vogelstein 1990) and a simplification of the five distinct groups described by Jass in 2007 (Jass 2007).
The three pathways outlined here are a simplification of Jass’ model but include all the fundamental genetic and epigenetic elements. The first pathway can be said to be the traditional adenoma-carcinoma pathway and is termed the chromosomal Instability (CIN) pathway. This is group four from Jass’ classification. The second pathway is that which causes CRC in Lynch Syndrome. Progression from adenomas, or flat adenomas, to CRC appears to be faster than in the CIN pathway. This is group 5 from Jass’ classification. The third pathway is the CpG Island methylator pathway. CRC develops within either serrated adenomas if MSI-high is present or adenomas if not. Methylation is the hallmark feature. This pathway is the most heterogeneous and includes Jass’ type 1, type 2 and type 3.

1.3.3 The Chromosomal Instability (CIN) Pathway

This pathway is thought to account for roughly 60-80% of all sporadic CRCs and all cancers caused by Familial Adenomatous Polyposis (Grady 2004) and MutYH associated polyposis. It represents the typical adenoma-carcinoma sequence of colorectal carcinogenesis (Fearon and Vogelstein 1990). The molecular aberrations of this pathway are largely through the development of structural chromosomal abnormalities called aneuploidy (Rajagopalan and Lengauer 2004). Thus the CIN pathway is associated with multiple gene mutations or deletions.

Mutation or loss of APC (adenomatous polyposis coli gene) located on the long arm of chromosome 5 (5q) is an early event in the CIN pathway. The APC gene product, the APC protein is closely involved with the Wnt-signalling pathway which is an important intracellular pathway involved in the regulation of cell survival, growth and differentiation. During the physiological cell cycle, a complex of proteins including APC, β-catenin, axin, GSK3β and CK1 modulate the degradation of stimulatory transcription factors. APC gene mutations frequently cause the APC protein to be truncated and thus alter or prevent binding to β-catenin. When the APC/β-catenin complex is reduced or absent, the usual suppression of the Wnt-signalling pathway is impaired and thus an abnormal survival advantage is conferred to that cell (Cadigan and Liu 2006; Worthley and Leggett 2010). APC mutations are found in up to 80% of all adenomas and around 60% and 82% of colonic and rectal cancers respectively (Worthley and Leggett 2010).
KRAS (Kirsten rat sarcoma virus) is an oncogene located on the short arm of chromosome 12 (12p) and is frequently involved with the CIN pathway. KRAS encodes a GTP-binding protein which functions as part of the RAF-MEK-ERK intracellular signalling pathway. When KRAS is mutated this signalling pathway which stimulates growth and differentiation is unchecked. KRAS mutations are not particular to the CIN pathway, or indeed CRC. KRAS mutations are found in 35-42% of all CRCs (Leslie, Carey et al. 2002).

Various genes located on the long arm of chromosome 18 (18q) including DCC (deleted in colorectal cancer), SMAD2 and SMAD4 (both code for SMAD family proteins) are frequently involved in the latter stages on CIN pathway carcinogenesis. Allelic loss at 18q is seen in up to 60% of all CRCs (Vogelstein, Fearon et al. 1988). Germline SMAD4 mutations are involved in a type of inherited colorectal cancer syndrome called juvenile polyposis syndrome.(Bevan, Woodford-Richens et al. 1999) The SMAD2 and SMAD4 proteins are both transcription modulators and signal transducers involved in the TGF-β signalling pathway; hence mutations in their genes also cause disturbed regulation of apoptosis and growth.

TP53 (tumour protein 53) is an important tumour suppressor gene involved in the latter stages of carcinogenesis of many cancers including CRC. TP53 is located on the short arm of chromosome 17 (17p). In carcinogenesis its loss of function is usually due to allelic loss as a late event as an advanced adenoma progresses into an invasive carcinoma. TP53 abnormalities are seen in between 4-26% of adenomas and 50-75% of carcinomas (Leslie, Carey et al. 2002). TP53 normally functions to control and slow progression through the cell cycle so that if DNA is damaged it has time to repair. Impaired TP53 gene function therefore promotes development of further genetic and epigenetic abnormalities (Worthley and Leggett 2010).
1.3.4 The Lynch Syndrome Pathway

This is the molecular pathway responsible for the CRCs of Lynch Syndrome, and represents group 5 of Jass’ classification. Lynch Syndrome causes 1-4% of all CRCs (Jass 2007). The initial genomic insult is an inherited (or de novo) germline mutation in one allele of the mismatch repair genes. A somatic insult of some sort then must occur to the second allele, which leads to a loss of function of the mismatch repair system and hence the accumulation of further genomic and epigenomic abnormalities. The precursor lesions are adenomas and flat adenomas which progress through carcinogenesis at an accelerated rate compared to those of the CIN pathway. These cancers are usually diploid.

The mismatch repair system functions to repair both nucleotide mismatches, and insertions and deletions. These can occur by chance during DNA replication, or can be caused by alkylation damage, ultraviolet light or oxidizing agents. In humans this system utilises at least seven proteins: MutL-homolog 1 (MLH1), MutL-homolog 3 (hMLH3), MutS-homolog 2 (hMSH2), MutS-homolog 6 (hMSH6), PMS-homolog 1 (hPMS1) and PMS-homolog 2 (hPMS2). These form heterodimers which are the functioning proteins. (MSH2:MSH3, MSH2:MSH6, MLH1:PMS1, MLH1:PMS2, MLH1:MLH3). MLH1 and MSH2 appear to be vital in maintaining the integrity of the MMR system. Germline mutations in MLH1 (3p21), MSH2 (2p22), MSH6 (2p16) and PMS2 (7p22) have been associated with Lynch Syndrome (Hoeijmakers 2001).

Microsatellite instability (MSI) is the hallmark genomic instability in this pathway and is caused by the defective DNA mismatch repair system.

1.3.5 The CpG Island Methylator Phenotype (CIMP) Pathway

This is the second most common pathway involved in sporadic CRC carcinogenesis. It is thought to account for around 15% of CRC (Weisenberger, Siegmund et al. 2006). The key fault in this pathway is epigenetic instability that allows
hypermethylation, and thus silencing, of the promoter regions of tumour suppressor genes.

As mentioned previously, CIMP may be high or low. CIMP-high cancers have methylation of many CpG islands, and have frequent \textit{BRAF} gene mutation. \textit{BRAF} gene (\textit{V}Raf \textit{mu}rine \textit{s}arcoma \textit{v}iral \textit{o}ncogene \textit{h}omolog \textit{B}1) is one of the isoforms of the \textit{RAF} gene, part of the RAS/RAF/ERK pathway. This pathway mediates proliferation, cellular differentiation and apoptosis, in response to extracellular signalling (Peyssonnaux and Eychene 2001). In contrast, CIMP-low cancers have more limited gene methylation and have frequent \textit{KRAS} gene mutation. It is thought that a synergy between either \textit{BRAF} or \textit{KRAS} mutation and a particular pattern of DNA methylation is required in order to stimulate the early stages of tumour development. \textit{BRAF} mutation is associated with widespread methylation of the promoter regions of multiple pro-apoptotic genes (e.g. \textit{RASSF1}, \textit{RASSF2}, \textit{MST1})(Esteller, Tortola et al. 2000). \textit{KRAS} mutation is associated with methylation of a smaller number of gene promoter regions but these are all genes involved in cell cycle arrest (e.g. \textit{p14} \textit{ARF} or \textit{TP53})(Esteller, Tortola et al. 2000).

CIMP-high tumours appear to be diploid and chromosome stable, and frequently the precursor lesion is a polyp with serrated morphology (Whitehall, Wynter et al. 2002; Jass 2007). There appears to be two subtypes of this pathway, dependant on the presence or absence of \textit{MLH1} gene promoter region methylation which results in microsatellite instability and a tumour phenotype similar to that of Lynch Syndrome tumours (Jass 2007; Worthley and Leggett 2010). CIMP-low tumours are frequently MSI-low or stable. The precursor lesion may be a serrated polyp or an adenoma.

\textbf{1.4 Biomarkers}

Sections 1.4-1.5 are quoted from published article (Newton, Newman et al. 2010).

A biomarker (BM) can be described as an objectively measurable substance that can be used as an indicator of a physiological or pathological state. When applied to
a disease state, a biomarker can be used to detect disease, predict prognosis, or predict response to a pharmacological agent. A biomarker must facilitate an improvement in life-expectancy or quality of life in order to be of use in clinical practice (Newton, Newman et al. 2010).

### 1.4.1 Biomarker Categories

The Cancer Research UK Biomarker Discovery and Development committee (BIDD) have defined the following categories of biomarkers (<http://science.cancerresearchuk.org/gapp/fundingcommittees/biomarker_imging_committee/> Accessed 2009; Accessed 2011).

1. **Risk assessment/predisposition BMs**: biomarkers that can be used to define individuals (or populations) that are at increased risk of developing a disease. These may be markers of a genetic predisposition or a measure of exposure to a substance that increases that individual’s risk. For example the *Adenomatous Polyposis Coli (APC)* gene and Familial adenomatous polyposis (FAP).

2. **Screening/early detection BMs**: biomarkers used to aid early diagnosis of a condition thus allowing detection at a more treatable stage. For example faecal occult blood testing in colorectal cancer.

3. **Diagnostic BMs**: biomarkers used to define the type of cancer or disease the patient has. These can be used alongside standard imaging techniques. For example PSA, CEA and CA-125. Diagnostic biomarkers can also be utilised to detect recurrent or relapsing disease.

4. **Pharmacological BMs**: biomarkers used to assess pharmokinetics, measure the effects of drug treatment on a specific target; e.g. enzyme inhibition, receptor blockade, induction of apoptosis or reduction in angiogenesis (proof of mechanism) or demonstrate clinical effects of that drug. These BMs can be used to aid dose optimisation and reduction in side-effects. For example *thiopurine methyltransferase (TMPT)* gene mutations and severe side-effects from azathioprine treatment in Crohn’s disease.
5. Predictive BMs: biomarkers that can define a subpopulation of patients who are likely to respond to a given therapy. For example wild-type KRAS and response to anti-EGFR therapy in metastatic colorectal cancer.

6. Prognostic BMs: biomarkers used to determine the likely course of a disease, independent of treatment. For example a biomarker that may reflect a cancer’s metastatic potential or potential growth rate. This may be used to estimate outcome and guide treatment and surveillance decisions.

1.4.2 Biomarker Discovery and Validation

1.4.2.1 Biomarker discovery

Biomarker discovery is the process through which putative biomarkers are investigated. This is either done in a hypothesis-driven manner following developments in the understanding of the molecular pathophysiology of a condition, or through a more random, hypothesis-free “fishing exercise” by investigating markers that occur differently in patients and in normal individuals.

A recent study has achieved, using a genetic technology called massively parallel sequencing, the mapping of the entire genome of a cell line derived from a small-cell lung cancer. The sequencing revealed 22,910 different genetic mutations (Pleasance, Stephens et al. 2009). This is a huge step towards demonstrating what today’s genetic technology is capable of achieving. Somewhere amongst all these mutations lie the key genetic changes that drive carcinogenesis. The next challenge is to understand the varying roles of such mutations, and in doing so, identify those that have clinical utility as biomarkers.

Advances in genetic technology have allowed progression from hypothesis-driven candidate gene research and family linkage analysis, to hypothesis-free genome-wide association studies (Van Limbergen, Russell et al. 2007). Candidate gene studies are population based association studies involving a small number of single nucleotide polymorphisms (SNPs). The selected genes are either candidate based
on functional pathophysiological knowledge of the disease, or a positional candidate based on previous family-based linkage studies. Genome-wide association studies use high-throughput technologies to investigate hundreds of thousands of SNPs throughout the genome to identify common causal variants (Garcia Rodriguez, Ruigomez et al. 2006; Van Limbergen, Russell et al. 2007). Genome-wide association studies in brief, compare the frequency of common SNPs in patients and healthy individuals, and allow identification of the common differences between the two. As noted by Vineis and Perera, these high throughput technologies “allow massive investigations not based on hypotheses” (Vineis and Perera 2007).

Molecular discovery techniques allow investigators to compare markers in patients and healthy individuals, but may not afford true biological understanding of the function of that marker. It is important that studies validating the role and limitations of putative biomarkers are performed before biomarkers are introduced into clinical practice. It is equally important that assay techniques are quality assured and standardised.

1.4.2.2 Biomarker validation

Biomarker validation can be divided into “analytical validation” and “clinical validation” (Brenner and Normolle 2007). Analytical validation relates to validation of the laboratory techniques used to isolate and measure the marker. There is a huge variation in the techniques used, and although many authors may claim validation, large multicentre studies are required to achieve this. Clinical validation relates to the use of the appropriate population for the study.

1.4.3 Biomarker discovery roadmaps

In order to prevent disorganised and uncoordinated biomarker research, Cancer Research UK’s Biomarker Discovery and Development Committee have suggested ‘roadmaps’ which define a succinct research pathway. The roadmaps are broadly divided into four chronological sections.

i) Rationale: Does the envisioned biomarker address an unmet clinical need?
ii) BM assay development: the assay must be accurate and reproducible. The assay should be simple and cost effective, and performed to GCLP standards (good clinical laboratory practice guidelines laid out by the British Association of Research Quality Assurance).

iii) BM discovery: the distribution of the BM in an appropriate sample population should be defined, and a retrospective analysis of the relationship between the BM and clinical outcome should be performed.

iv) BM Clinical qualification: firstly the relationship between the BM and clinical outcome should be assessed in a large retrospective analysis. Secondly, a large prospective randomised study should be performed to assess the impact of the BM on clinical outcome.

1.5 Biomarkers in Colorectal Cancer

There are numerous established and putative biomarkers related to colorectal cancer in the literature (Newton, Newman et al. 2010). Biomarkers that are relevant to work included in this thesis are discussed below.

1.5.1 Risk assessment (predisposition) Biomarkers

Potential risk assessment biomarkers have some specific requirements in order to demonstrate clinical utility. The sample from which the BM is measured must be easy to obtain and as minimally invasive as possible. The BM must have a good specificity defined using samples from the target population. The BM must be able to diagnose the disease earlier than the normal clinical presentation. Correlation of the BM and incidence must be validated in a large prospective study. This correlation must be statistically robust. A prospective randomised trial must demonstrate that the BM affords a significant reduction in mortality or morbidity.
Individuals with a greatly increased risk of colorectal cancer can be identified from family history, phenotypic recognition of one of the polyposis or hamartomatous syndromes, molecular features in the cancer of the consultand or a relative suggestive of a familial syndrome, or detection of a germline mutation. In this context, germline mutations can be considered as BMs. Appropriate surveillance of individuals with these high risk genetic disorders significantly reduces the incidence of CRC and improves life expectancy (Dunlop 2002).

The germline mutations responsible for familial adenomatous polyposis (FAP), Lynch Syndrome (HNPCC), Peutz-Jeghers Syndrome (due to mutations in STK11) and Juvenile Polyposis Syndrome (due to mutations in SMAD4) are well documented and are routinely analysed in genetics centres (See section 1.6.2). An interesting story in biomarker discovery is told by the recent delineation of MYH (MutY Homolog [E. Coli]) associated polyposis syndrome (MAP).

In 2002 a hypothesis-driven candidate gene study by Al-Tassan et al led to the identification of the first autosomal recessively inherited colonic polyposis and cancer susceptibility syndrome; MYH associated polyposis (Al-Tassan, Chmiel et al. 2002). A Welsh family in which three siblings were affected by multiple adenomas and colorectal cancers was investigated. An excess of somatic mutations within the APC gene within the siblings’ tumours was found to be associated with germline variants of the MYH gene (which encodes the base excision repair protein mutH). The three affected siblings were all compound heterozygotes (had one copy of each mutant gene and no wild-type gene). All phenotypically normal family members were either heterozygous for one of these variants or had two wild-type alleles. As the polyposis and cancers in this family only occurred in the compound heterozygotes, the condition was consistent with inheritance as an autosomal recessive trait (Al-Tassan, Chmiel et al. 2002). These findings were confirmed by Jones et al (Jones, Emmerson et al. 2002) and Sieber et al (Sieber, Lipton et al. 2003). Biallelic MYH mutations predispose to multiple polyps, colorectal cancers, and in some patients; extra-colonic manifestations (duodenal polyposis and duodenal malignancy). MYH mutations tend to lead to a milder disease than classical FAP, but more severe than simple multiple adenomas. More polyps develop in MYH polyposis than in HNPCC, but appear to progress more slowly into cancer (Sieber, Lipton et al. 2003).
1.5.1.1 Molecular biomarkers suggestive of Lynch syndrome

See section 1.7.4

1.5.1.2 Genome wide association (GWA) studies

The hypothesis underpinning the genome wide association studies in relation to CRC risk is that much of inherited risk is due to combinations of low risk, low penetrance gene mutations, or variants. GWA studies, which are a form of genetic epidemiological study, have recently confirmed this hypothesis and identified ten commonly mutated alleles (variants). It is thought that these alleles each contribute roughly 1% of all heritable cancers, and confer an increased risk of CRC of between 1.2 and 1.47 times that of the normal population (Houlston, Webb et al. 2008). GWA studies initially identified six loci of increased risk; these are located in the chromosome regions 8q24(Zanke, Greenwood et al. 2007), 10p14 and 8q23, (Tomlinson, Webb et al. 2008) and 15q13 (Jaeger, Webb et al. 2008). A recent meta-analysis of the GWA studies identified four further loci; located in chromosome regions 16q22.1 (part of the gene which encodes E-Cadherin), 19q13.1 (part of gene which encodes RHO GTPase binding protein 2 which plays a role in cell motility and the actin cytoskeleton), 14q22.2 (part of the gene which encodes bone morphogenetic protein 4 (BMP4) part of the transforming growth factor β cell signalling system) and 20p12.3 which appears to be located in a gene desert (Houlston, Webb et al. 2008). It remains unclear what clinical utility these relatively low risk mutations may have. General population testing in order to alter screening regimens is currently not feasible. As further loci of risk are discovered, and the role of the currently identified loci better understood, these findings may have a significant public health impact.

1.5.2 Screening Biomarkers

The study of screening BMs has some individual needs which are similar to those of risk assessment BM. It is important that the test be as minimally invasive as possible to afford good compliance and reduce the risk of complications. It must also be cost-effective. The modality of the screening tool is of upmost importance.
If compliance with the test is low due to patient dissatisfaction with the modality, then the specificity and sensitivity of the test become almost irrelevant.

It is accepted that screening for colorectal cancer reduces mortality through the early detection of cancers and the endoscopic removal of pre-cancerous polyps. However, there is no international consensus on the preferred strategy. In 2006 the UK CRC screening programme was implemented. Men and women aged between 60 and 69 were invited to undergo screening, and it is planned to repeat this invitation every two years. This utilises guaiac based faecal occult blood testing (gFOBT). The European council, however, recommends FOBT screening in average risk men and women between the age of 50 and 74. The uptake of screening is a major determinant of the effectiveness of the programme. It is known that compliance with stool based testing is low. A Cochrane review of faecal occult blood testing found compliance with at least one round of testing to be between 60 and 78% (Hewitson, Glasziou et al. 2008). There is an expectation that a new generation of blood based markers should improve compliance with CRC screening.

1.5.2.1 Currently available screening modalities

The currently utilised methods for early detection of colorectal cancer include FOBT and endoscopy. Four prospective randomised controlled trials have demonstrated that biennial gFOBT reduces CRC mortality by 15-33% (Mandel, Bond et al. 1993; Kewenter, Brevinge et al. 1994; Hardcastle, Chamberlain et al. 1996; Kronborg, Fenger et al. 1996). The main and significant pitfall of gFOBT is its relatively low sensitivity. A recently published systematic review on faecal-occult blood testing summarises the evidence used to inform the decision regarding which test to utilise in the UK screening program (Burch, Soares-Weiser et al. 2007). It can be seen that gFOBT meets all the requirements of the Cancer UK BM development roadmap (see Figure 2). Accurate, cost-effective assays have been developed. The BM has been shown to detect CRC earlier, in some patients, than the normal clinical presentation, and use of the BM affords a reduction in cancer mortality. The evidence for the use of gFOBT is robust. However, the relatively high false negative and false positive rates indicate a need for further BM development.
1.5.3 Diagnostic BM

The distribution of the BM value must be defined using samples from a representative population. This distribution must indicate a clinical utility. The BM must significantly improve diagnostic accuracy when compared to the gold standard current practice in a prospective study.

In the management of CRC, diagnostic BMs have a place in detection of recurrent disease. Currently carcinoembryonic antigen (CEA) is the only approved diagnostic BM in use (Wang, Tang et al. 1994). Circulating tumour cells and cytokeratins have
been examined as possible new BM of recurrence, though the studies are few and limited (Fernandes, Kim et al. 2006; Wang, Wu et al. 2006; Diehl, Schmidt et al. 2008).

**Figure 3: Diagnostic biomarker roadmap.**

Adapted from (<http://science.cancerresearchuk.org/gapp/fundingcommittees/biomarker_imagining_committee/> Accessed 2009)

**1.5.4 Pharmacological BM**

It is thought that variation in an individual’s germline DNA can affect how they respond to chemotherapeutic agents. These variations can alter drug metabolism and hence moderate the side-effect profile experienced by the patient. There is evidence that variants of genes coding for enzymes involved in the metabolism of chemotherapeutic agents affect the function and side-effect profile of those agents.
For example; variants of the gene which codes for dihydropyridimine dehydrogenase and the anti-metabolite agent 5-Flourouracil (Meinsma, Fernandez-Salguero et al. 1995; Yen and McLeod 2007), and the gene UGT1 and irinotecan which is a pro-drug topoisomerase inhibitor (Gupta, Lestingi et al. 1994; Ychou, Douillard et al. 2000; Cote, Kirzin et al. 2007).

1.5.5 Predictive BM

The distribution of the biomarker value must be defined using samples from a representative population. There must be a statistically robust correlation between the biomarker value and clinical outcome. This should be validated in a prospective study, or prospective analysis of a retrospective sample collection. The clinical outcome should be improved by the use of the biomarker. For example; a reduction in mortality by use of a treatment, or improved quality of life (reduced adverse events) secondary to avoidance of treatment that is predicted to be ineffective.

1.5.5.1 KRAS/BRAF and anti-EGFR therapy

1.5.5.1.1 KRAS

Cetuximab (ER-K0034, Erbitux, Merck-Serono KgaA, Darmstadt, Germany; ImClone Systems Inc, New York, NY) and panitumumab (ABX-EGF, Vectibix; Amgen Inc, Thousand Oaks, CA), are monoclonal antibodies (mABs) that target the epidermal growth factor receptor (EGFR) and have expanded the treatment options for metastatic CRC. Cetuximab is a chimeric mouse-human mAB and panitumumab is a fully human mAB. EGFR is part of the human epidermal growth factor (HER) erb family of tyrosine kinase receptors. Its activation plays a key role in tumour progression through proliferation, angiogenesis, invasion and metastasis.

Cetuximab and panitumumab appear to have similarly small but none the less, clinically important, response rates of around 10% in unselected chemotherapy refractory metastatic CRC patients. Immunostaining for EGFR was initially used to define entry criterion for trials. However, objective treatment responses to anti-EGFR mAB have been seen in tumours with low or absent, as well as high, expression of EGFR (Siena, Sartore-Bianchi et al. 2009).
**KRAS** (Kirsten rat sarcoma viral oncogene) encodes a g protein and **BRAF** (VRaf murine sarcoma viral oncogenes B1) encodes a protein kinase. Both these genes are part of the tightly regulated ras/raf/MAPK intracellular pathway and ultimately stimulate the expression of proteins involved in cell division and survival. Activating mutations in the KRAS gene lead to a gain of function and hence over-expression of RAS/RAF dependant proteins. KRAS mutations are considered an early event in colorectal carcinogenesis and are present in between 20 and 50% of all CRCs. It has been shown by many studies that in chemotherapy refractory metastatic CRC, KRAS mutation predicts a complete lack of response to anti-EGFR therapy. Progression free survival is approximately double that in the wild type than in the mutant type (Siena, Sartore-Bianchi et al. 2009). The first large study to confirm the negative predictive value of KRAS was the randomized phase III study of panitumumab monotherapy in the refractory/relapse setting. This study found that in those treated with panitumumab, progression free survival (PFS) was 12.3 in the wild-type KRAS subgroup but only 7.4 weeks in the mutant subgroup. This was statistically significant (Van Cutsem, Peeters et al. 2007).

**KRAS** data from two large randomized phase II-III studies examining the benefits of cetuximab in the first line adjuvant treatment of metastatic CRC have recently been published. The OPUS trial (Fluorouracil, leucovorin, oxaliplatin with and without cetuximab in the first line treatment of metastatic colorectal cancer) found that the addition of cetuximab to FOLFOX-4 regimen was only beneficial in the wild-type KRAS subgroup. (Bokemeyer, Bondarenko et al. 2009) The CRYSTAL trial (Cetuximab and chemotherapy as initial treatment in colorectal cancer) also found there was no benefit gained by adding cetuximab to standard chemotherapy (FOLFIRI; fluorouracil, leucovorin and irinotecan) in those patients with mutant type KRAS. With the addition of cetuximab progression free survival was 7.6 months and 9.9 months for the mutant and wild-type groups respectively, and overall survival was 17.5 months and 24.9 months. (Van Cutsem, Kohne et al. 2009)

**KRAS** testing should now be standard practice for CRC patients being considered for anti-EGFR therapy and is now supported Foods and Drugs Agency (FDA) and European Medicines Agency (EMEA) guidance. This is generally as second line treatment, but with the now available OPUS and CRYSTAL data, it may well soon become standard practice prior to first line chemotherapy. Of note, the evidence
suggests that the KRAS status of the primary tumour and any metastasis is equivalent (Siena, Sartore-Bianchi et al. 2009).

1.5.5.1.2 BRAF

Mutant KRAS only accounts for 30-40% of the 90% of patients who do not respond to anti-EGFR therapy. As BRAF is the major effector of KRAS, it has been hypothesised that BRAF mutation may have a predictive role in response to anti-EGFR therapy in wild-type KRAS cancers. Di Nicolantino et al retrospectively examined tumours from 113 patients who had received either cetuximab or panitumumab in a second or successive line chemotherapy regimen. They examined for KRAS mutations and BRAF mutations. They found, as predicted, that KRAS mutation was associated with non-response (p=0.011). They also found that none of those carrying a BRAF mutation responded, and that none of the responders had a BRAF mutation (p=0.029). The patients with mutant type BRAF had both shorter progression free survival (p=0.011) and shorter overall survival (p<0.001) (Di Nicolantonio, Martini et al. 2008). The CAIRO-2 study examined the effects of cetuximab plus capecitabine, oxaliplatin and bevacizumab in first line therapy. A retrospective analysis of these patients in relation to BRAF mutation status has revealed a similar picture. The progression free survival and overall survival were 6.6 months and 15.2 months for those with mutant BRAF, and 10.4 (p=0.01) and 21.5 months (p=0.001) for wild-type BRAF (Tol, Koopman et al. 2009). These two studies seem to suggest that BRAF analysis may well be of use, alongside KRAS analysis, in predicting those patients who may benefit from anti-EGFR therapy.

Examining the discovery of this biomarker against the Cancer Research UK roadmap it is evident that this biomarker is nearing clinical utility. It has addressed an unmet clinical need, and a statistically robust relationship between the biomarker assay level and clinical outcome has been established using a retrospective tissue bank. In order for BRAF analysis to be considered for routine clinical practice; a randomised controlled trial now needs to demonstrate an improvement in clinical outcome afforded by the use of BRAF mutation analysis prior to the use of anti-EGFR therapy.
1.5.5.2 Mismatch repair deficiency and 5-fluoruracil chemotherapy

The benefits of adjuvant 5-fluoruracil (5-FU) chemotherapy in the treatment of high risk Duke’s stage B and stage C CRC are well established. The intergroup study, the earliest large randomised controlled trial (RCT), of 1296 patients, demonstrated an overall reduction of CRC specific death of 33% and a reduction of recurrent disease of 41%. (Moertel, Fleming et al. 1990) These results have since been replicated by many other RCTs. Although other agents such as; oxaliplatin, irinotecan and Cetuximab are in use, 5-FU currently remains a component of all standard chemotherapeutic regimens for CRC.

There is a growing body of evidence that CRCs with a defective mismatch repair (MMR) system do not respond to 5-FU chemotherapy. The MMR system recognises and repairs both nucleotide mismatches and also alkylation damage, such as that caused by 5-FU. It has been hypothesised that in MMR deficient cancers, whilst 5-FU is still incorporated into the malignant cells’ DNA, the resulting alkylation damage is not recognised by the deficient MMR system. Therefore apoptosis is not triggered and the cancer cell survives. A systematic review and meta-analysis which examined the benefit of adjuvant 5-FU chemotherapy in MMR deficient and MMR proficient patients has recently been published (Des Guetz, Schischmanoff et al. 2009). The end points were relapse-free survival (RFS) and overall survival (OS). This study included seven studies, two of which were retrospective analyses of randomised controlled trials. For MMR deficient patients there was no statistically significant difference in RFS (HR RFS 0.96, 95%CI 0.62-1.49, p=0.86) or OS (HR for OS 0.70, 95% CI 0.44-1.09, p=0.12) whether or not they received 5-FU based chemotherapy (Des Guetz, Schischmanoff et al. 2009). A systematic review published by Popat et al also concluded that 5-FU conferred no benefit in MMR deficient cancers (Popat, Hubner et al. 2005). Combining the meta-analysis (Des Guetz, Schischmanoff et al. 2009; Des Guetz, Uzzan et al. 2009), systematic review (Popat, Hubner et al. 2005), and a recently published update by Jover et al (Jover, Zapater et al. 2009), a total of 545 mismatch repair deficient patients have, to date, been investigated. No significant effect on survival or disease free survival of adjuvant 5-FU chemotherapy has been found (Newton and Hill 2010).
However, there is emerging evidence that the molecular mechanism behind the mismatch repair deficiency may determine whether the tumour responds or not. A recent retrospective study analysed the mismatch repair status of 2141 tumours from patients included in a variety of adjuvant 5-fluorouracil based chemotherapy trials (none of which included irinotecan or oxaliplatin). In multi-variant analysis it was found that MMR deficiency was independently associated with delayed time to recurrence (HR 0.72 95% confidence interval = 0.56 to 0.91, p= 0.005), and improved disease free survival (p=0.035). These are consistent with previous findings. It was found that although in stage 3 MMR deficient cancers 5-FU based chemotherapy (versus surgery alone or no treatment) was associated with reduced distant recurrence rate (11% vs 29%; p=0.011), this effect was limited to cancers likely to be caused by germline mismatch repair mutations. MMR deficient cancers likely to be caused by germline mutations (as defined by age less than 55 on initiation of treatment, and/or MSH2 protein loss on IHC, and/or wild-type BRAF V600E) were found to have an improved disease free survival following 5-FU chemotherapy compared with the sporadic mismatch repair tumours where no benefit was seen (p=0.006) (Sinicrope, Foster et al. 2011).

In the UK, a small number of Oncology departments are using knowledge of mismatch repair status (usually according to typical histological features) to guide adjuvant therapy decisions. No Cancer network is yet performing routine IHC or MSI on all cancers being considered for chemotherapy. However, based on the recent data published by Sinicrope et al (Sinicrope, Foster et al. 2011), it seems that 5-FU based chemotherapy should not be with-held from patients with mismatch repair deficient tumours unless Lynch Syndrome has been confidently ruled out.

### 1.5.6 Prognostic biomarkers

Stage at diagnosis is currently the major determinant of whether a surgical cure will be achieved or not. This cure rate ranges from 90- 50% in stages I-III respectively, to <10% in stage IV disease (Nash, Gimbel et al. 2009). To date, KRAS mutation, BRAF mutation and mismatch repair deficiency have all been found to be independent predictors of prognosis. As mentioned previously, there is increasing evidence of at least three subsets of colorectal cancer defined by the molecular carcinogenic pathways. There does appear to be overlap between these groups. The
groups have been suggested according to chromosomal stability, microsatellite instability, CpG island methylation and whole gene mutations. Due to the differing molecular fingerprint of these groups they are likely to have differing clinical courses and hence prognosis.

As mentioned previously, it is well established that \textit{KRAS} mutation is a predictor of response to anti-EGFR therapy. It has also been well proven that mutant \textit{KRAS} is an independent predictor of poorer outcome regardless of stage or intervention (Andreyev, Norman et al. 1998; Andreyev, Norman et al. 2001; Nash, Gimbel et al. 2009; Richman SD 2009). Several large trials have demonstrated that \textit{BRAF} mutation is also a predictor of poor outcome, notably independent of MSI (Nosho, Kure et al. 2009; Richman SD 2009; Souglakos, Philips et al. 2009; Zlobec, Bihl et al. 2009).

\textbf{1.5.6.1 Mismatch repair deficiency}

A number of groups have investigated the relationship between microsatellite instability (MSI) and CRC prognosis. There is consensus that MSI predicts better prognosis despite being associated with more poorly differentiated cancers (Popat, Hubner et al. 2005; Sanchez, Krumroy et al. 2009).
1.6 Familial colorectal cancer syndromes

1.6.1 Familial risk

It is thought that inherited genetic susceptibility plays a role in up to 35% of all CRC (Houlston, Webb et al. 2008). In 2010, the British Society of Gastroenterologists published an updated set of guidelines regarding the management of patients at increased risk of developing CRC. In these guidelines familial risk is described as “a heterogeneous composite of high penetrance single gene disorders, multiple low penetrance genetic factors, and shared environmental exposure” (Cairns, Scholefield et al. 2010). It is known that the identification and appropriate surveillance of individuals at increased risk of developing CRC due to a genetic susceptibility reduces the incidence of CRC and improves life-expectancy (Gammon, Jasperson et al. 2009; Cairns, Scholefield et al. 2010; Mallinson, Newton et al. 2010).

1.6.2 Single gene disorders

There are five well recognised high risk single gene disorders which confer a lifetime risk of colorectal cancer of between 10 and 100%. These conditions account for 3-5% of all CRCs (Cairns, Scholefield et al. 2010). Patients may be identified following assessment of an appropriate family history, and/or phenotypic information and/or tumour derived biomarkers, and/or appropriate germline genetic sequencing. It must be noted that none of these conditions have complete penetrance. Individuals may carry the family mutation but never develop CRC. The genes responsible for these five syndromes have been identified and hundreds of pathogenic mutations have been characterised. A causative pathogenic mutation is not identified in every family in which the syndrome or condition has been diagnosed clinically.

Lynch syndrome is a dominantly inherited cancer disposition syndrome caused by a germline mutation in one of the mismatch repair genes. (See section 1.7)
Familial adenomatous polyposis (FAP) is an autosomal dominant condition, in which classically hundreds to thousands of precancerous adenomatous polyps develop throughout the colorectum. The life-time risk of CRC is nearly 100% if no screening is undertaken. The adenomatous polyposis coli gene (APC) is located on the long arm of chromosome 5 and encodes a protein consisting of 2843 amino acids. This is known to function as a tumour suppressor gene (Groden, Thliveris et al. 1991). The APC gene has 15 exons; up to 85% of mutations (germline and somatic) occur between codons 1250 and 1464, in the mutation cluster region (Friedl, Caspari et al. 2001). Over 1000 germline APC mutations have been described on the APC mutation database (Laurent-Puig, Beroud et al. 1998). Germline APC mutations are identified in between 30 and 85% of FAP families depending on the patient group and genetic analysis technique used (Friedl, Caspari et al. 2001). (See section 1.8)

*MutYH* associated polyposis (MAP) is an adenomatous polyposis syndrome with high penetrance caused by bialleic recessively inherited mutations in the base excision repair (BER) gene *MutYH*. This gene is located on chromosome 1p and functions to protect DNA from oxidative damage. MAP confers a nearly 100% life-time risk of developing CRC. MAP polyposis can mimic attenuated FAP (Jasperson, Tuohy et al. 2010). Adenomas typically develop by age 40 but can occur at any age (Lubbe, Di Bernardo et al. 2009).

Peutz-jeugers syndrome (PJS) is a dominantly inherited disorder with high penetrance which gives rise to gastrointestinal hamartomatous polyposis in association with abnormal skin and mucosal macular melanin deposits. Patients have a 15 times average risk of gastrointestinal cancers. Small bowel hamartomas are characteristic, but gastric and colonic hamartomas are also common, with an incidence of 25% and 30% respectively (McGarrity and Amos 2006). Complications of the polyposis are common in the first three decades of life and these include bleeding, anaemia, intussusceptions and bowel obstruction (Gammon, Jasperson et al. 2009). Peri-orbital and peri-oral mucocutaneous pigmentation are the typical extra-intestinal features and are usually seen in childhood (Houlston, Webb et al. 2008; Jasperson, Tuohy et al. 2010). PJS confers a lifetime risk of gastrointestinal cancers of around 70%. There is also a 50% life-time risk of breast cancer (Gammon, Jasperson et al. 2009). 20-63% of cases are caused by germline mutation in *STK1/LKB1* gene. There is also a further possible locus on chromosome
19q (*PTEN*) (Mehenni, Gehrig et al. 1998; Trojan, Raedle et al. 1998; Chen and Fang 2009). Screening colonoscopy is recommended at intervals of 1-2 years from age 15-18 years (Cairns, Scholefield et al. 2010). Upper gastrointestinal screening including oesophagogastroduodenoscopy and MRI enteroclysis (or alternative mode of small bowel visualisation) is recommended every 1-2 years also from age 25 (Cairns, Scholefield et al. 2010).

Juvenile polyposis (JPS) is a syndrome defined by multiple juvenile type hamartomatous polyps throughout the large and small bowel. The polyps have an apparently normal mucosa with a dense stroma and inflammatory infiltrate. Diagnostic criteria are more than five juvenile type polyps in the colon or rectum, juvenile polyps throughout the gastrointestinal tract, or any number of juvenile polyps in a person with a family history of juvenile polyposis. Patients frequently present with symptoms of the polyps including rectal bleeding, diarrhoea or pain. The life-time risk of developing CRC is 39% and (in those patients with multiple gastric polyps) the life-time risk of gastric cancer is 21%. Small bowel and pancreatic cancers have also been documented (Warthin 1913; Brosens, van Hattem et al. 2007). 20% of JPS cases are thought to be due to germline mutations within the *SMAD4* gene, and 20% due to *BMPR1A* mutation. The *DPC4* gene and a further locus on chromosome 6q have also been implicated (Bevan, Woodford-Richens et al. 1999). Screening colonoscopy is recommended at intervals of 1-2 years from age 25 years (Cairns, Scholefield et al. 2010). Screening oesophagoduodenoscopy is recommended every 1-2 years also from age 25 (Cairns, Scholefield et al. 2010).

**1.6.3 Moderate Familial Risk**

CRC is a common disease. Multiple cases may occur in a family by chance and not due to germline inheritance of susceptibility. Roughly 10% of the general population will have at least one relative who has, or has had, CRC. However, less than 1% of the general population will fulfil moderate risk criteria that warrant endoscopic screening (Cairns, Scholefield et al. 2010).
Johns and Houlston published the first systematic review and meta-analysis which collated the risk estimates for CRC depending on family history. This study included 27 case-control and cohort studies and pooled estimates for relative risk (Johns and Houlston 2001).

Table 2: Relative risk of CRC according to family history (Johns and Houlston 2001)

<table>
<thead>
<tr>
<th>Family history</th>
<th>Relative risk of CRC</th>
<th>Absolute risk by age 79</th>
</tr>
</thead>
<tbody>
<tr>
<td>No family history</td>
<td>1</td>
<td>4%</td>
</tr>
<tr>
<td>1 FDR* with CRC</td>
<td>2.25</td>
<td>9%</td>
</tr>
<tr>
<td>More than 1 FDR with CRC</td>
<td>4.25</td>
<td>17%</td>
</tr>
<tr>
<td>At least 1 FDR diagnosed with CRC below the age of 45</td>
<td>3.87</td>
<td>15.5%</td>
</tr>
</tbody>
</table>

*FDR: first degree relative

A further meta-analysis performed by Butterworth et al examined the effect of a family history of CRC on an individual’s personal risk. This analysis included 59 studies with good heterogeneity. Relative risk was found to be 2.24 (95% CI 2.06-2.43) with at least one affected relative, and 3.97 (95% CI 2.60-6.06) with at least two affected relatives. Absolute risk curves were also derived and demonstrated an increase in risk dependent on the time period/age of the consultand specified. The life-time risk is therefore greater the younger the patient (Butterworth, Higgins et al. 2006).

The British Society of Gastroenterology guidelines (for colorectal cancer screening in high risk groups) define two groups of patients at moderate risk of developing CRC due to their family history. These are patients whose family history of CRC is not suggestive of a high-risk single gene disorder, but is significant enough to warrant colonoscopy surveillance (table 3)(Cairns, Scholefield et al. 2010).
Table 3: Moderate risk CRC categories (Cairns, Scholefield et al. 2010)

<table>
<thead>
<tr>
<th>Family History</th>
<th>Life-time risk of CRC</th>
<th>Risk category</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRC in 3 FDRs* (1 in first degree kinship with consultand), none &lt; age 50 years</td>
<td>1: 6-10</td>
<td>High moderate</td>
</tr>
<tr>
<td>CRC in 2 FDRs (1 in first degree kinship with consultand) with a mean age of &lt;60 years</td>
<td>1:12</td>
<td>Low moderate</td>
</tr>
<tr>
<td>CRC in 2 FDRs (1 in first degree kinship with consultand) aged &gt;60 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRC in 1 FDR aged&lt;50 years</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*FDR: first degree relative

1.6.4 Low-penetrance, low-risk susceptibility genes (Genome Wide Association Studies)

Genome-wide association studies use high throughput laboratory technologies to investigate hundreds of thousands of single nucleotide polymorphisms (SNPs) throughout the genome to identify common possible disease causing variants (Rodriguez-Bigas, Boland et al. 1997). Genome-wide association studies compare the frequency of common SNPs in patients with the disease in question and healthy individuals. This allows identification of the common differences between the two and links these to the disease. Causation is not proven but an association identified.

As described in section 1.5.1.2, the hypothesis underpinning the genome wide association (GWA) studies in relation to CRC risk is that much of inherited risk is due to combinations of low risk, low penetrance gene mutations, or variants. GWA studies have recently identified ten commonly mutated alleles. It is thought that these alleles each contribute roughly 1% of all heritable cancers, and confer an increased risk of CRC of between 1.2 and 1.47 times that of the normal population (Umar, Boland et al. 2004; Houlston, Webb et al. 2008).
1.7 Lynch syndrome

1.7.1 Introduction and history

Lynch syndrome (or hereditary non-polyposis colorectal cancer: HNPCC) is a cancer predisposition syndrome, inherited in an autosomal dominant fashion. It is caused by a mutation in one of the mismatch repair genes (MLH1: mutL homolog, MSH2: mutS homolog 2, MSH6: mutS homolog 6, PSM2: post meiotic segregation increased 2) (Lynch and de la Chapelle 2003). The condition has variable penetrance but gives rise to up to an 80% life time risk of CRC if undetected and hence not screened. Lynch syndrome also predisposes to the development of extra-colonic cancers. The spectrum includes endometrial, ovarian, gastric, small bowel, pancreatic, hepatobiliary, brain (particularly glioma) and upper urothelial (renal pelvis and ureteric) malignancies. Lynch syndrome causes up to 4% of all CRCs and is the most prevalent form of inherited colorectal cancer. It has been demonstrated that effective lifelong large bowel surveillance can reduce the risk of developing CRC to 10%, attributed to the removal of premalignant adenomas. Families with Lynch syndrome need formal genetic assessment with counselling and predictive testing.

In 1913 Warthin described a family, termed Family G with a significant history of colorectal, gastric, and endometrial cancer (Warthin 1913). In 1966 Henry Lynch published details of two large American families with familial colorectal and endometrial cancer that did not fit with the phenotypic pattern of the already known familial adenomatous polyposis. Lynch described two syndromes: Lynch Syndrome 1; site specific colorectal cancer, and Lynch syndrome 2; familial cancer syndrome (Lynch, Shaw et al. 1966). Initially the Lynch syndromes were classified according to clinical criteria; the Amsterdam criteria (see section 1.7.3.1). In 1993 and 1994, genetic linkage analysis of Lynch syndrome families located the genetic loci of the genes responsible (Fishel, Lescoe et al. 1993; Lindblom, Tannergard et al. 1993; Peltomaki, Aaltonen et al. 1993; Papadopoulos, Nicolaides et al. 1994). Following this discovery, mutation analysis became possible and hence the delineation of Lynch syndrome.
1.7.2 Molecular genetics

The DNA mismatch repair system and the genes responsible for Lynch syndrome have been previously discussed. (See section 1.3.4)

1.7.3 Diagnosis

The diagnosis of Lynch syndrome is problematic. This is partly due to the phenotypic overlap with sporadic CRC. Other difficulties include frequently small family size and the relative prevalence of familial colorectal cancer without an identifiable genetic cause. The optimal way to identify patients with Lynch Syndrome remains a matter of debate. Currently, in the UK, patients are referred by their general practitioner or hospital doctor to a clinical genetics department on the basis of a suspicious family history (and/or personal history) of Lynch spectrum cancers. In the clinical genetics department, a family pedigree is taken and assessed against family history criteria; the Amsterdam 2 or Revised Bethesda Criteria. If the family fulfils Amsterdam II criteria, a living affected relative’s DNA (if available) is tested for germline MMR mutations. If the family or personal history fulfils Bethesda criteria tumour biomarker analysis (either MSI or MMR protein IHC) is performed on archived tumour tissue (if available) and if MMR deficiency is demonstrated mutation analysis is then performed on germline DNA. It has been suggested that a more effective strategy would be to assess every patient with newly diagnosed CRC for possible Lynch Syndrome, either using a prediction model or tumour testing. One of the most sensitive strategies would be to subject every individual with CRC (or other Lynch Spectrum cancers) to germline MMR mutation analysis. This is obviously too costly and time-consuming.

1.7.3.1 Amsterdam/Bethesda Clinical Criteria

In 1991 the International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer (ICG-HNPCC) developed the first clinical criteria termed the
Amsterdam Criteria (ACI) in an attempt to identify possible Lynch Syndrome patients for collaborative studies (Vasen, Mecklin et al. 1991). The ACI included only CRC. These criteria were revised in 1999 to include all Lynch spectrum cancers, termed the ACII (Vasen, Watson et al. 1999). If patients fulfil all the criteria listed in table 4 they are diagnosed with possible Lynch Syndrome and referred for gene testing and screening.

**Table 4: Amsterdam II Criteria (Vasen, Watson et al. 1999)**

<table>
<thead>
<tr>
<th><strong>Amsterdam II Criteria</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. At least three relatives with a Lynch Spectrum cancer (CRC, endometrial, upper urothelial, small bowel, ovary)</td>
</tr>
<tr>
<td>2. At least one of these is a first degree relative of the other two</td>
</tr>
<tr>
<td>3. At least two affected generations</td>
</tr>
<tr>
<td>4. At least one of the affected patients was under the age of 50 at the time of their cancer</td>
</tr>
<tr>
<td>5. FAP has been excluded</td>
</tr>
<tr>
<td>6. The cancers have all been verified from pathological records</td>
</tr>
</tbody>
</table>

In 1996, the US National Cancer Institute (NCI) International workshop on HNPCC developed a set of clinical criteria designed to pre-select patients whose tumour should be tested for MSI prior to formal germline mutation testing. (Rodriguez-Bigas, Boland et al. 1997) Only one of the criteria needs to be fulfilled to recommend tumour testing for MSI. The Bethesda criteria were revised in 2004 (see table 5). (Umar, Boland et al. 2004)
Table 5: Revised Bethesda Criteria (Umar, Boland et al. 2004)

<table>
<thead>
<tr>
<th>Revised Bethesda Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Any CRC diagnosed under the age of 50 years</td>
</tr>
<tr>
<td>2. *Synchronous or *metachronous CRC or any other Lynch spectrum cancer diagnosed at any age</td>
</tr>
<tr>
<td>3. CRC with typical MSI-H histology in a patient under the age of 60 years</td>
</tr>
<tr>
<td>4. CRC diagnosed in one or more first degree relatives with a Lynch spectrum cancer, with one of the cancers being diagnosed under the age of 60 years</td>
</tr>
<tr>
<td>5. CRC diagnosed in two or more first degree relatives or second degree relatives with a Lynch spectrum cancer regardless of age</td>
</tr>
</tbody>
</table>

*Synchronous cancer is a second CRC which is histologically distinct from but occurs at the same as, the primary CRC.

*Metachronous CRC is a second CRC which occurs more than 12 months after the primary CRC, is located in a different part of the colorectum, and was not present at the time of the primary CRC.

*i Metastatic CRC is a second CRC which is histologically distinct from but occurs at the same as, the primary CRC.

Neither the ACII nor the Revised BC are perfect in terms of sensitivity and specificity for selecting probable Lynch Syndrome patients. The ACII criteria are highly specific (98%) but have poor sensitivity (42%) (Barnetson, Tenesa et al. 2006). The revised BC are far more sensitive (95%) but less specific (38%) (Barnetson, Tenesa et al. 2006). The ACII criteria were designed to select those who could be referred directly for mutation analysis whilst the revised BC criteria were designed to be used in combination with MSI or IHC as pre-screening biomarkers.

1.7.3.2 Other predictive models

Several risk prediction models have been developed. These vary in the patient populations used to design them and the parameters included. The models can be used to calculate a probability of a given patient carrying a MMR mutation. These models include the MMRpredict (Barnetson, Tenesa et al. 2006), PREMM1,2 (Balmana, Stockwell et al. 2006) and MMRPro (Chen, Wang et al. 2006) models.
Barnetson et al designed the MMRpredict model following investigation of all patients presenting with a CRC below the age of 55 years in Scotland between 1999 and 2005. This two stage model uses tumour location, presence of either synchronous or metachronous CRC, family history of cancer, age of the youngest affected relative, presence of endometrial cancer to calculate the probability of an individual being a mutation carrier. The model allows adjustment depending on the presence of MMR deficiency (Barnetson, Tenesa et al. 2006). MMRpredict has an overall sensitivity of 62% and specificity of 99% (Barnetson, Tenesa et al. 2006). It improves slightly on the ACII criteria (Vasen, Watson et al. 1999). However, it relies on a detailed and accurate family history and thus may be less useful in smaller families and younger patients.

Balmana et al published the PREMM\(_{1,2}\) which predicts the likelihood of an individual carrying a pathogenic MLH1 or MSH2 mutation. A model development cohort of 1618 patients who had submitted blood for germline MLH1 and MSH2 testing at a commercial genetics laboratory in Salt Lake City, Utah, USA, was used initially to design the parameters of the model. Secondly, a validation cohort of 1016 patients recruited in the same way was used to test the model. This model is based on cancer history (age of onset/type/number of primary cancers) of the proband and their first and second degree relatives. The authors quote that when the model predicts a 5% chance of a mutation, this has a sensitivity of 94% and specificity of 29%, and when the model predicts a 10% likelihood sensitivity of the model is 85% and specificity 60% (Balmana, Stockwell et al. 2006).

Chen et al designed the MMRPro model in 2006. This model was designed using published mutation prevalence and penetrance data, and the specificity and sensitivity of tumour testing (IHC and MSI) and germline mutation analysis obtained from the published literature at that time. The model was validated using a cohort of 279 individuals who were obtained from a clinic-based population in Maryland, USA. This model has two tiers. It can initially assess the probability of a given individual harbouring a MMR mutation based on personal and family history of CRC and endometrial cancer. This probability can then be refined based on results of tumour testing. Using both tiers, the authors quote a sensitivity (concordance index) of 83% and a specificity of 94% (Chen, Wang et al. 2006).
Although these figures appear impressive, this model has only been validated in a clinic based population.

1.7.4 Tumour Biomarker Analysis

There are two analytical methods for detecting mismatch repair deficiency in Lynch spectrum tumour specimens. These are microsatellite instability (MSI) and mismatch repair protein immunohistochemistry. Currently these markers are used as pre-screening tools prior to germline mutation testing.

1.7.4.1 Microsatellite Instability

MSI is a somatic alteration in the number of repeats of short DNA sequences within a cell’s DNA. Tumour MSI analysis requires extraction of DNA from the tumour specimen and also of genomic DNA for comparison. Genomic DNA can be derived from blood lymphocytes or normal tissue adjacent to the tumour. The test is usually performed in a molecular genetics laboratory. A polymerase chain reaction (PCR) is used to amplify the area of DNA of interest and instability is detected by comparing germline and tumour DNA using a variety of techniques.

The National Cancer Institute (NCI) Workshop have recommended a standardised panel of microsatellites to be used for clinical testing (Moertel, Fleming et al. 1990). This panel includes BAT25 and BAT26 (mononucleotide microsatellites), and D5S346, D2S123 and D17S250 (dinucleotide microsatellites). If more than two of these sites are unstable, the tumour is classed microsatellite instability high (MSI-H), if just one site is unstable the tumour is classed microsatellite instability low (MSI-L) and if no instability is detected, the tumour is said to be microsatellite stable (MSS) (Jass 2007).

The EGAPP (Evaluation of Genomic Applications in Practice and Prevention) EGAPP working group is a North American group developed by the United States National Office of Public Health Genomics at the Center for Disease Control and Prevention in
order to vigorously examine the evidence base that supports the transition of genetic tests from research into clinical practice. In 2009 this working group reviewed the evidence regarding clinical testing for Lynch Syndrome (see section 1.10.4) and found that across studies MSI has the following performance characteristics: 80-91% sensitivity and 90% specificity for MLH1 and MSH2 mutation carriers, 55-77% sensitivity and 90% specificity for MSH6 and PMS2 mutation carriers (EGAPP 2009).

1.7.4.2 Mismatch repair protein Immunohistochemistry

Tumour immunohistochemistry can be reliably used to detect MMR deficiency. Immunohistochemistry (IHC) is a process of antigen detection that utilises relevant antibodies and a visualisation system. Monoclonal antibodies against the MMR proteins MLH1, MSH2, MSH6 and PMS2 are widely commercially available. IHC is performed on a fixed histological specimen and provides visual recognition of the presence or absence of the protein under investigation. (See section 1.10.7.1.1) Lack of immunoreactivity allows prediction of the most likely MMR gene involved. Mutations that result in an absent or truncated protein (e.g. the deleterious mutations) usually result in complete loss of immunostain. Certain mutations result in the production of a deficient protein, but one that retains antigenicity. These include the pathogenic frameshift mutations. This depends on the site of the particular antigen utilised by the monoclonal antibody system and results in a false-positive effect (Renshaw 2007).

An automated, semi-quantitative immunostaining technique for the MMR proteins has been established by Barrow et al (Barrow, Jagger et al. 2011). This study examined the performance characteristics of IHC by testing its ability to differentiate between cancers from patients with known MMR gene mutations and sporadic MMR proficient cancers. The MLH1 immunostain has a sensitivity of 100% (95% CI 84-100%) and a specificity 91.5% (95% CI 79.6- 97.6%) (for the detection of Lynch cancers). The MSH2 immunostain has a sensitivity 87.5% (95% CI 61.7- 98.4) and specificity of 88.5% (95% CI 76.5- 95.6%) (for the detection of Lynch cancers). This is a simple, reproducible, cost-effective test that can be performed in any pathology laboratory that uses immunohistochemistry (Barrow, Jagger et al. 2011).
1.7.5  Screening

Large bowel surveillance has been recommended in patients with Lynch Syndrome since the syndrome was first recognised. Identification of the causal genes has allowed predictive testing in families for whom a mutation has been identified. Updated guidelines from the British Society of Gastroenterologists were published in 2010. It is recommended that patients with confirmed or suspected Lynch syndrome should have total colonic screening in the form of colonoscopy at least every 2 years commencing at age 25 years (Cairns, Scholefield et al. 2010). It is also suggested that screening every 18 months may be more appropriate due to the incidence, in some series, of interval cancers (Vasen, Mecklin et al. 1991; Vasen, Nagengast et al. 1995; de Vos tot Nederveen Cappel, Nagengast et al. 2002).

It has been shown that appropriate screening confers a seven year survival advantage in Lynch patients and reduced cancer incidence, presumably due to the endoscopic removal of pre-cancerous adenomas (Vasen, van Ballegooijen et al. 1998).

1.7.6  Treatment

The treatment of a colorectal cancer in a patient with Lynch syndrome remains similar to that of sporadic CRC. A combination of surgery and adjuvant therapy is offered if the cancer is Dukes stage B and demonstrates high-risk histological features, or Dukes stage C. However, there is evidence that suggests that patients with a CRC due to Lynch Syndrome may benefit from different surgical and adjuvant therapy.
1.7.6.1 Surgery

If a known Lynch Syndrome patient suffers a colonic cancer, it is usual practice to offer a subtotal colectomy and ileo-rectal anastomosis (IRA) rather than a segmental resection. The rate of metachronous cancer following a segmental resection in Lynch Syndrome patients is 16%, and only 3% following a colectomy and IRA (Lynch, Smyrk et al. 1993). Parry et al recently published data on metachronous cancer (mCRC) risk in 382 MMR gene mutation carriers which demonstrated a independently significantly increased risk of mCRC in patients who underwent a segmental rather than an extensive resection (Parry, Win et al. 2011). No patients who had an extensive resection (total or subtotal colectomy) developed a mCRC and the incidence rate for those who had a segmental resection was 23.6 (95% confidence interval 18.8-29.7) per 1000 person years. In patients who had a segmental resection the effect of length of bowel removed on risk of metachronous cancer was estimated using Cox regression (length of bowel removed fitted as a continuous variable). The authors reported a risk ratio for mCRC for every 10cm of bowel removed was 0.69 (95% confidence interval 0.54-0.88) (in patients who have had a segmental resection). This reportedly equates to a mCRC relative risk reduction of 31% for every 10cm of bowel removed (Parry, Win et al. 2011).

Following subtotal colectomy and IRA, the retained rectum must continue to be screened due to the residual risk of rectal cancer (Cairns, Scholefield et al. 2010).

1.7.6.2 Adjuvant therapy

Currently in most centres, patients with Lynch Syndrome receive the same adjuvant (either neo-adjuvant or adjuvant therapy) as patients with sporadic cancers. There is evidence that mismatch repair deficient cancers don not respond to 5-flurouracil based chemotherapy, however recent evidence suggests this lack of response may be limited to sporadic mismatch repair deficient CRCs. (See section 1.5.5.1.3)
1.7.7 Lynch Syndrome Registries

It was suggested by the British Society of Gastroenterologists that all Lynch syndrome families should be referred to a regional genetics centre in order to facilitate their genetic assessment and screening (Cairns, Scholefield et al. 2010). This recommendation is based on expert opinion from both Europe (Vasen, Moslein et al. 2007) and North America (ASCO 1996) and thus is based on level IV evidence only. It is known that the use of a genetic registry to coordinate screening in patients with FAP significantly reduces the incidence of CRC (a reduction of 28.7% to 14.0%) and improves life expectancy (from 58.1 years to 69.6 years) (Mallinson, Newton et al. 2010). It is thought that a similar effect may be noted in the Lynch Syndrome population.

1.8 Familial Adenomatous polyposis

1.8.1 Introduction and history

Classical familial adenomatous polyposis (FAP) is an autosomal dominant condition, in which hundreds to thousands of precancerous adenomatous polyps develop throughout the colorectum. Extra-intestinal benign and malignant abnormalities are also well documented. These include duodenal polyposis and adenocarcinoma, gastric fundic gland polyposis, osteomas, congenital hypertrophy of the retinal pigment epithelium (CHRPEs), sebaceous cysts, and desmoid tumours. Phenotypic variation both amongst and within families is well recognised.

The prevalence of FAP is 1:14,000. Surgical prophylaxis means that only 0.07% of all CRC is accounted to FAP in modern practice. (Cairns, Scholefield et al. 2010)

The first histologically confirmed case of adenomatous polyposis was described in Russia by Sklifasowski in 1881. This was a case of a 51 year old merchant with a seven year history of abdominal pain and bloody diarrhoea. A colotomy was performed and polyps removed and examined (Bulow, Berk et al. 2006). During the
late 1880s several families were described with multiple cases of colonic polyposis (Cripps 1882; W 1884; T 1887). Doering conducted the first review of the literature in 1907. 44 cases were found. 31 out of 37 patients had died from cancer (Doering 1907). In 1925 Lockhart-Mummary noted that the hereditary factor in FAP is adenomas with a tendency to develop into malignancy, rather than the cancer itself (Lockhart-Mummary 1925). A case series of polyposis patients from St Marks Hospital was published and following this the St Marks Polyposis Registry was set up to coordinate the care of these patients (Lockhart-Mummary 1925). Cockayne first documented that FAP is inherited in a dominant fashion in 1927 (Cockayne 1927). In 1951 Gardner first described what came to be known as Gardner’s syndrome. He first described the extra-colonic manifestations of FAP which included colonic adenomas, desmoid tumours, bone tumours and subcutaneous cysts (Gardner 1951). Peri-ampullary adenocarcinoma was first described in association with FAP at Massachusetts General Hospital in 1935 (Bulow, Berk et al. 2006) and Halsted described the first case of endoscopically diagnosed gastric polyposis in 1950 (Halsted, Harris et al. 1950).

In 1986 Herrera et al first noted the association between polyposis and a deletion in chromosome 5q. (Herrera, Kakati et al. 1986) In 1987 Bodmer et al (Bodmer, Bailey et al. 1987) and Leppert et al (Leppert, Dobbs et al. 1987) both independently documented the location of the APC gene. The adenomatous polyposis coli gene (APC) is located on the long arm of chromosome 5 and encodes a protein consisting of 2843 amino acids. This is known to function as a tumour suppressor gene (Groden, Thliveris et al. 1991). The APC gene has 15 exons; up to 85% of mutations (germline and somatic) occur between codons 1250 and 1464, in the mutation cluster region (Friedl, Caspari et al. 2001). Over 1000 germline APC mutations have been described on the APC mutation database (Laurent-Puig, Beroud et al. 1998). Until recently germline APC mutations were identified in between 30 and 85% of FAP families depending on the patient group and genetic analysis technique used (Friedl, Caspari et al. 2001). It is now possible to detect the causative mutation in around 95% of patients/families due to current techniques for whole gene sequence analysis and assays that are able to detect larger defect mutations (Cairns, Scholefield et al. 2010). Around 25% of all cases of FAP are due to de novo mutations (i.e. new mutation in proband; no family history) (Cairns, Scholefield et al. 2010).
1.8.2 Diagnosis

The clinical diagnosis of FAP typically requires the observation of diffuse adenomatous polyps in the large bowel at any age, or fewer polyps in the colorectum of a symptomatic child. Although colonic polyps may develop in the colorectum of the normal population, FAP sufferers develop multiple polyps usually beginning in childhood and adolescence.

In order to identify a causative APC mutation in a new family, germline DNA analysis on an affected individual is performed on DNA derived from saliva or blood borne lymphocytes. Multiplex ligation-probe dependant amplification (MLPA) of the mutation cluster region on exon 15 is typically performed initially. MLPA is a type of multiplex polymerase chain reaction (PCR) assay. This technique allows detection of small copy number changes so is able to identify even single nucleotide deletions or duplications. If no mutation is found the whole gene is then sequenced. Once the family mutation is identified, predictive testing can be offered to at risk family members.

Genetic analysis is not recommended until the child is able to understand the consequences of the result, which is generally agreed to be around the age of 11 years (Hyer and Fell 2001; Reyes Moreno, Ginard Vicens et al. 2007). This is also the time at which genetic diagnosis would alter the management of the child. Some individuals present symptomatically either due to new mutations or if they are not known to a screening programme, and are subsequently diagnosed with FAP. These individuals have a worse prognosis (Bulow 2003).

1.8.3 Screening

“There is an interval of approximately 10 years between the appearance of polyps and the onset of symptoms” (Berk, Cohen et al. 1981), and thus large bowel endoscopic surveillance (screening) aims to monitor patients in this latent phase.
Endoscopy allows quantification of polyp load and thus an assessment of CRC risk, and appropriate planning of prophylactic surgery.

It is recommended that colonic surveillance is performed annually with flexible sigmoidoscopy and colonoscopy alternating (Cairns, Scholefield et al. 2010). In families in whom no mutation has been found it is recommended that family members at 50% risk have annual surveillance from age 13-15 years until age 30 years, and every 3-5 years until age 60 years. In typical FAP, it is recommended that this commences at age 10-12 (Dunlop 2002; Vasen, Moslein et al. 2008). After prophylactic surgery, the retained bowel (rectum following colectomy and ileo-rectal anastomosis, and anal cuff after restorative proctocolectomy) requires annual surveillance because of the continued risk of cancer (Cairns, Scholefield et al. 2010). The risk of rectal cancer after colectomy is 12-29% (Bulow 1989; De Cosse, Bulow et al. 1992; Nugent and Phillips 1992). There is increasing evidence supporting the existence of an attenuated FAP phenotype determined by genotype. These patients tend to have fewer, more proximal polyps and develop CRC 10-15 years later than those with typical FAP (Burt, Leppert et al. 2004; Nielsen, Hes et al. 2007). It is thought that biennial colonoscopy commencing at age 18-20 years is appropriate for these patients (Lockhart-Mummery, Dukes et al. 1956; Vasen, Moslein et al. 2008). This avoids unnecessary procedures in a proportion of the paediatric FAP population.

Bulow et al calculated that 94% of their call-up patients (patients undergoing pre-symptomatic screening) survived 10 years post diagnosis in comparison to only 41% of probands (patients diagnosed with symptoms) (Bulow, Bulow et al. 1995). A further smaller study has confirmed these findings (Reyes Moreno, Ginard Vicens et al. 2007). Heiskanen et al also observed a statistically significant improved survival in screen detected patients following colectomy (Heiskanen, Luostarinen et al. 2000). Pre-symptomatic screening of FAP patients improves the long term survival as suggested by several studies (Vasen, Griffioen et al. 1990; Morton, Macdonald et al. 1993; Heiskanen, Luostarinen et al. 2000). Mallinson et al found that FAP patients who underwent pre-symptomatic screening had improved survival from 57.8 years to 70.4 years (p<0.001) (Mallinson, Newton et al. 2010). Fewer patients diagnosed on screening subsequently develop CRC (Vasen, Griffioen et al. 1990; Bulow, Bulow et al. 1995; Goldberg, Madden et al. 1995; Bulow, Bulow et al. 1996). Morton et al investigated patients notified to the West Midlands Polyposis
Registry and found the rate of CRC to be 6% in the screen detected population and 64% in the symptomatic population \( (p < 0.001) \) \( \text{(Morton, Macdonald et al. 1993)} \). Rhodes et al found a similar ten-fold reduction amongst patients notified to the Northern region Polyposis Register \( \text{(Rhodes, Chapman et al. 1991)} \). Mallinson et al found that the incidence of CRC was reduced from 43.5% to 3.8% by screening \( \text{(Mallinson, Newton et al. 2010)} \).

1.8.4 Prophylactic Surgery

The risk of CRC exceeds 90% by age 70 in classical FAP if prophylactic surgery is not undertaken. Lockhart-Mummery at St Mark’s Hospital was the first to advocate prophylactic colectomy. In 1956 he reviewed the literature and suggested that although proctocolectomy with ileo-anal anastomosis left the patient with poor function, colectomy with ileo-rectal anastomosis (IRA) gave good functional results with an acceptable risk of rectal cancer if screening was continued \( \text{(Lockhart-Mummery, Dukes et al. 1956)} \). Colectomy and IRA was the only alternative to a permanent stoma until 1980 when Parks and Utsunomiya both described the operation of restorative proctocolectomy with mucosectomy and ileal pouch-anal anastomosis \( \text{(Parks, Nicholls et al. 1980)} \). This operation removes the vast majority of colorectal epithelium that has neoplastic potential. The ileo-anal cuff still requires annual follow-up but this can be performed by simple manual examination rather than endoscopy. The pouch formed from ileum creates a reservoir for stool thus functionally acts as a neo-rectum.

Surgery is performed after discussion between the patient (and their family) and the multi-disciplinary team of healthcare professionals. Generally surgery is considered when the endoscopist feels that it is becoming difficult to safely monitor the polyp load.
1.8.5 Registries

In 1925 Lockhart-Mummery demonstrated that pre-symptomatic examination of at risk family members, and prophylactic colectomy reduced the incidence of CRC and hence improved survival of FAP patients. Following this, the St Mark’s Hospital Polyposis Registry was established as the first polyposis registry in the world (Lockhart-Mummery 1925). Polyposis Registries have now been established across the world, both regionally and nationally. However, despite the British Society of Gastroenterologists (grade B) recommendation, published in 2002, that all FAP families should be notified to a regional registry (Dunlop 2002), there are currently only two formal polyposis registries in the UK (Manchester Regional Genetics Centre and the North West Thames Regional Genetics Service: St Mark’s Polyposis Registry). The Manchester Regional Genetics Centre serves a population of 4.7 million. The St Marks Polyposis Registry serves a regional population of 3.5 million, although as this Registry receives many referrals from other centres, both in London and nationwide, the exact population served is difficult to define. One other centre (Oxford regional genetics centre; serves a population of 3.6 million) has a more informal local database which essentially functions as a registry. Patients are regularly followed up and screening outcomes are documented. There are 23 regional genetics centres in the UK.

The aims of a polyposis registry include: registration of polyposis patients and their family members, counselling and genetic testing, initiation and co-ordination of screening of family members at risk, distribution of information and advances in scientific knowledge and research (Bulow, Burn et al. 1993). A further benefit of an established register is the maintenance of post surgical follow-up, which is of particular importance after ileo-rectal anastomosis surgery, to ensure monitoring of the retained rectum (Bulow, Bulow et al. 1995). The continuity of care provided and annual contact maintained with consistent departmental staff including genetic counsellors also benefits the psychological and social aspects of patients’ health.

Several European studies have demonstrated the effect a Registry has on the FAP populations it monitors. An improvement in survival, an increase in detection rate of FAP and a reduction in the incidence of CRC have been demonstrated (Vasen,
Griffioen et al. 1990; Reyes Moreno, Ginard Vicens et al. 2007). Two UK groups have reported on the outcomes of regional registries. Rhodes et al reported on the early results of the Northern Region Polyposis Registry in 1991 (this registry is no longer in existence). They found that their registry increased the identification of gene carriers and hence screening of at-risk relatives. Mean age of diagnosis was reduced from 36.6 years to 24.7 years. At that time the registry had only been established for 36 months (Rhodes, Chapman et al. 1991). Morton et al reported similar results following the establishment of the West Midlands Polyposis Register (again this registry is no longer in existence). Median age of diagnosis was reduced from 32 years to 23 years (p=0.0004), and incidence of colorectal cancer was reduced from 35% to 14% (p<0.05) (Morton, Macdonald et al. 1993).

1.9 St Mary’s Hospital department of Medical Genetics

The department of Genetic Medicine at St Mary’s Hospital, Central Manchester University Hospitals NHS Trust serves a population of 6.5 million. It is a tertiary referral centre and has specialist services for those with a genetic predisposition to cancer. General practitioners, physicians, surgeons and oncologists from across the North West of England refer patients for assessment, and screening and treatment advice.

With regard to increased risk of colorectal cancer; patients are referred for a variety of reasons: moderate risk family history of CRC, Amsterdam or Bethesda criteria fulfilled, young-onset CRC (with or without a significant family history), multiple primary cancers, or a phenotype suggestive of one of the single-gene disorders (i.e. FAP, MutYH associated polyposis, Juvenile polyposis, Peutz-Jeghers). Following assessment patients and their family members are entered onto either the Manchester Familial Colorectal Cancer Registry or the Manchester Polyposis Registry. Patients diagnosed with FAP or MAP are entered onto the Polyposis registry. All other patients; including those who are felt to have no increased risk are entered onto the Familial Colorectal Cancer Registry. These registries are databases that are used to coordinate the clinical care of these patients.
The Manchester Familial Colorectal Cancer Registry was established in 1990. Patients were entered retrospectively initially and prospectively since the inception of the registry. Patients’ details are contained on a Filemaker pro® database. All patients referred to the regional department of Genetic Medicine due to a family history of colorectal cancer are entered onto the database. All relatives within the pedigree are listed. The database contains demographic and pedigree information, as well as cancer incidence and pathology. Information on genetic diagnostic investigations including tumour mismatch repair immunohistochemistry and microsatellite instability, and mismatch repair gene sequencing is also documented. Mismatch Repair Gene mutation status is classified as positive, positive obligate, positive putative, negative or untested. Dates of cancer diagnosis and dates of last follow up or death, are verified using the North West Cancer Intelligence Service (NWCIS), family genetic records and the NHS Summary care records.

The Manchester Polyposis Registry was established in 1989. Patient details are contained upon a Filemaker Pro database. FAP and MAP patients are confirmed from medical records or cancer registration and entered on the database along with all available information on endoscopy screening. Dates of FAP/MAP diagnosis, dates of cancer diagnosis and dates of last follow up or death, were obtained from the North West Cancer Intelligence Service (NWCIS), family genetic records and also from the NHS Summary care records. Patients are entered onto Filemaker pro® database if they were found to have over 100 colonic adenomas and/or an APC mutation.
1.10 Introduction to studies one, two and three

In the introduction to this thesis the molecular genetics of colorectal cancer and genetic disorders which cause an increased risk of colorectal cancer have been discussed. The three studies within this thesis attempt to understand how differences in molecular genetics and genetics affect the natural history and outcomes from colorectal cancer in patients with a genetic predisposition, and how this might be used to alter screening, management and surveillance.

Study One:
The diagnosis of Lynch Syndrome is difficult. Lynch Syndrome is under-diagnosed. Diagnosis is important as screening reduces morbidity and mortality. Study one investigates how differences in molecular genetics could be used to improve our ability to identify potential Lynch Syndrome probands from both population based screening, and from medium and high risk populations.

Germline Lynch Syndrome mutation testing is expensive and time-consuming. Mismatch repair immunohistochemistry identifies tumours with mismatch repair deficiency. Loss of MSH2, MSH6 or PMS2 protein expression indicates high risk of Lynch Syndrome so germline mutation testing is justified. Loss of MLH1 protein is caused by Lynch Syndrome but is also caused by a sporadic (somatic) epigenetic defect in MLH1 gene, which is seen in around 13-15% of sporadic CRCs. These patients do not have Lynch Syndrome and therefore they and their family members are not at increased risk of further cancer. In order for screening for Lynch Syndrome in low and medium risk groups to be feasible it is important to be able to differentiate between the two groups of patients who have MLH1 mismatch repair deficient tumours. It is hypothesised that cancers caused by germline MLH1 mutations in Lynch Syndrome have a different tumour biomarker profile to cancers caused by a sporadic defect in MLH1. Lynch Syndrome is associated with somatic unmethylated MLH1 promoter region and somatic wild-type BRAF. The sporadic MLH1 loss cancers are associated with methylated MLH1 promoter region and mutant BRAF. In order to test the sensitivity and specificity of each of these biomarkers; cancers from patients with known pathogenic germline MLH1 mutations and sporadic cancers with MLH1 loss...
will be tested and compared. The application of these biomarker tests to patients from different risk groups (low, medium and high) will be assessed using Bayes theorem.

Study Two:
Familial adenomatous polyposis is caused by mutations within the APC gene, located on the long arm of chromosome 5. Three clinical subtypes are recognised: attenuated or mild FAP, classical FAP and severe FAP. Previous studies have suggested that these subtypes are associated with the loci of APC mutation. These were descriptive studies of small numbers of large FAP families. It remains unclear whether knowledge of a patient’s genotype should affect their management in terms of timing of screening, timing and type of prophylactic surgery and surveillance. Current management guidelines do not recommend a consideration of genotype. Study two investigates genotype-phenotype correlation in polyposis in a large polyposis population by using all patients on the Manchester Polyposis Registry. This registry also includes patients who were later found to have a MutYH mutation as they were originally diagnosed with mild FAP prior to the discovery of the MYH gene. Patients are classified according to the loci of their mutation. Age at diagnosis of polyposis, time of onset of CRC and overall survival are used to investigate the relationship between the clinical behaviour and the mutation loci. Comprehensible genotype information should be relayed to the clinicians looking after these patients to be used as an aide to clinical decision making.

Study Three:
Moderate familial risk colorectal cancer is the most frequently encountered heritable CRC by surgeons and physicians. Moderate familial risk is thought to be caused by inheritance of a combination of low-risk, low penetrance gene variants. Little is known about the clinical behaviour of cancers within this large group. It is recognised that these individuals with a moderate family history have an increased risk of CRC, with an earlier age of onset, compared to the general population. The British Society of Gastroenterologists and the Association of Coloproctology recommend large bowel screening with either a one off colonoscopy at age 55, or five-yearly colonoscopy from age 50 depending on particulars of the family history. In order to plan surveillance, it is important to know the risk of a metachronous colorectal cancer. Life-time risk of a metachronous colorectal cancer (CRC) is
0.6%-3% following sporadic CRC and 15-26% in Lynch syndrome. The risk of metachronous CRC is unknown in the moderate risk group. In order to investigate the risk of metachronous colorectal cancer in patients with a moderate family history, three groups of patients who have had at least one CRC are identified from the regional familial colorectal cancer database. These groups are patients with background general population CRC risk, patients with moderate familial CRC risk and patients with Lynch Syndrome. Data on incidence of metachronous CRC is gathered and Kaplan-Meier and cumulative incidence function used to estimate the risk of metachronous colorectal cancer in these three groups. It is hypothesised that moderate risk patients with have an increased risk of metachronous CRC compared to the general population, but less than that of Lynch Syndrome.
2 Study One:
DNA mismatch gene promoter region methylation analysis and 
BRAF gene mutation analysis- an alternative pre-screening 
strategy in Lynch Syndrome (hereditary nonpolyposis 
colorectal cancer).

2.1 Introduction

2.1.1 Background

Despite the greatly elevated lifetime risk of colorectal and other cancers, patients 
with Lynch syndrome do not have a distinct phenotype. Thus diagnosis is 
problematic. Studies suggest that up to 4% of all CRCs are due to an undiagnosed 
Lynch syndrome (Salovaara, Loukola et al. 2000). Those individuals will also have 
at-risk relatives. It has been demonstrated that effective lifelong screening (large 
bowel endoscopy) can reduce the risk of developing CRC to 10%. It is known that 
colonoscopy surveillance reduces mortality. The risk of a second CRC in a patient 
with Lynch syndrome is estimated at 16% within ten years. Without screening the 
risk of a new cancer in a first or second degree relative is around 45% for males 
and 35% for females by age 70 years (EGAPP 2009). Thus families with Lynch 
syndrome need formal genetic assessment with counselling and predictive testing.

2.1.2 Family history as a method of ascertainment

Vasen et al conducted a “survey to evaluate the strategies currently used to identify 
individuals at high risk for CRC in 14 Western European countries”(Vasen, Moslein 
et al. 2007). In all countries surveyed (via a European Collaborative group of 
experts on hereditary cancers), the main method of ascertainment was following 
obtaining a family history and referral to a clinical genetics department. The family 
history is usually taken by a general practitioner or a hospital doctor involved in 
that patient’s cancer care (Bulow, Burn et al. 1993). It is known, both objectively
from this survey, and subjectively from other European studies and a direct assessment of GP records and hospital notes in a UK study, that the quality of documented family history is poor (Rose, Murphy et al. 2004; Trano, Wasmuth et al. 2009; Singh, Schiesser et al. 2010).

The use of pre-screening biomarkers on sporadic CRCs has been suggested as a better way of ascertaining families who may have Lynch Syndrome. As described in section 1.3.4 the carcinogenic pathway to Lynch syndrome CRCs is unique. However, there are some overlaps in the hallmark molecular events with the CIMP pathway. Over 90% of Lynch syndrome CRCs have mismatch repair deficiency demonstrable either by lack of mismatch repair proteins seen on immunohistochemistry or microsatellite instability of their DNA. Around 13% of all sporadic colorectal cancers also display mismatch repair deficiency (these are cancers that develop through the serrated pathway due to CpG island methylator phenotype as described in section 1.3.5, hence are Jass type 1 cancers) (Jass 2007). It would be unfeasible (and unnecessary) to test 13% of patients who have a colorectal cancer for germline Lynch Syndrome mutations. Hence it is clearly important to be able to distinguish these two patient groups.

The consensus criteria for the diagnostic strategy for Lynch Syndrome suggest that tumour testing for mismatch repair deficiency (either MSI or MMR protein immunohistochemistry) should be conducted on all those fulfilling Amsterdam or revised Bethesda criteria (Umar, Boland et al. 2004). Patients demonstrating MSI-H or loss of MMR protein on immunohistochemistry should be subjected to germline mutation testing. However, less than a third of these will harbour a mutation (Pinol, Castells et al. 2005), the remainder are associated with sporadic MLH1 hypermethylation.

**2.1.3 BRAF gene mutation and MLH1 promoter region methylation**

Two independent molecular pathways which lead to MSI CRC are now well recognised as described in section 1.3 (Jass 2007; Young, Jenkins et al. 2007). Lynch Syndrome cancers are typically characterised by MSI-H, a normal
(unmethylated) \textit{MLH1} promoter region, and wild-type \textit{BRAF}. The MSI-H phenotype also occurs as a result of epigenetic silencing of \textit{MLH1} through hypermethylation of its promoter region. It is associated with the \textit{BRAF} p.V600E mutation and found in around 13% of all sporadic CRC (Weisenberger, Siegmund et al. 2006).

\textit{BRAF} gene (VRaf murine sarcoma viral oncogenes B1) is one of the isoforms of the \textit{RAF} gene, part of the RAS/RAF/ERK pathway. This pathway mediates proliferation, cellular differentiation and apoptosis, in response to extracellular signalling (Perez-Carbonell, Alenda et al. 2010). \textit{BRAF} mutations can be found in around 10% of all CRC, and the V600E mutation accounts for over 80% of these (Muller, Burgart et al. 2001). \textit{BRAF} mutations occur more frequently in MSI-High CRC than MSI-stable CRCs (37% and 7% respectively) (Rajagopalan, Bardelli et al. 2002).

In 2004 Kambara et al examined a selection of hyperplastic polyps, serrated polyps, classical adenomas, sporadic CRCs (both MSI-H and MSI-S) and cancers from patients with Lynch Syndrome for \textit{BRAF} gene mutations. They found a higher incidence in serrated adenomas (75%), hyperplastic polyps (19%) and sporadic MSI-H CRCs (76%), than in classical adenomas (0%), sporadic MSI-S CRCs (9%) and Lynch Syndrome CRCs (0%) (p<0.001) (Kambara, Simms et al. 2004). These findings help support the theory discussed previously of distinct molecular carcinogenesis pathways to microsatellite unstable sporadic cancer and Lynch Syndrome (Jass 2007). This study also examined for CpG Island methylation, and \textit{MLH1} promoter methylation was found in 66% of those tumours with \textit{BRAF} mutation.

In 2004 Nagasaka et al and Deng et al, both demonstrated through experimental studies, that \textit{BRAF} mutation is associated with \textit{MLH1} promoter region hypermethylation in sporadic cancers (Deng, Bell et al. 2004; Nagasaka, Sasamoto et al. 2004).

Nagasaka et al retrospectively examined 264 sporadic CRCs for \textit{BRAF} mutation, \textit{KRAS} mutation and hypermethylation at multiple loci. They found that \textit{BRAF} and \textit{KRAS} mutation never occurred in the same tumour. They also found that all tumours with hypermethylation of the 3’promoter region of the \textit{MLH1} gene
demonstrated loss of *MLH1* protein expression on IHC and *BRAF* gene mutation (Nagasaka, Sasamoto et al. 2004).

Deng at al examined 80 sporadic CRCs and 20 from patients with Lynch syndrome (fulfilled Amsterdam II, and had at least 2 first degree relatives with CRC one of whom was aged <50 years) for *MLH1* hypermethylation, and *BRAF* gene mutation. Of the sporadic CRCs, *BRAF* mutations were found in 42% of all the microsatellite unstable cancers (p<0.0001), and 100% of cancers with *MLH1* hypermethylation. Of the sporadic cancers with *MLH1* hypermethylation 87% were found to have *BRAF* gene mutation. Also of interest, the group examined for *KRAS* mutations and found that these and *BRAF* mutations were mutually exclusive. Of the 20 Lynch syndrome cancers examined; none were found to have mutated *BRAF*, although one case had *MLH1* hypermethylation (and a known *MLH1* gene mutation, raising the suggestion that methylation may be the second hit event (according to Knudson’s two hit hypothesis) in some Lynch cancers) (Deng, Bell et al. 2004).

*MLH1* promoter region methylation and/or *BRAF* p.V600E mutation testing of MMR deficient tumours, has the potential to enable better selection of patients who should undergo genetic testing. The methodology for *BRAF* mutation testing is well established. Whilst a number of methods for *MLH1* methylation analysis have been developed, most are technically difficult (particularly in FFPE tissue) and expensive.

### 2.1.4 The use of population based biomarkers

It has been widely suggested that the use of population based tumour biomarker testing (testing of all patients diagnosed with colorectal cancer) may be a more effective way of identifying Lynch families than the currently used family history criteria. Population based risk assessment testing must have high sensitivity and specificity, be cost and time effective and use technologies that are or can be made widely available, and reproducible.
In 2009, a set of recommendations was published by the EGAPP (Evaluation of Genomic Applications in Practice and Prevention) Working Group. EGAPP is a North American group developed by the United States National Office of Public Health Genomics at the Center for Disease Control and Prevention in order to vigorously examine the evidence base that supports the transition of genetic tests from research into clinical practice. This report examined the evidence for genetic testing strategies in patients with newly diagnosed CRC aimed at reducing morbidity and mortality from Lynch syndrome, by identification of mutation carriers and their at risk relatives. The group assessed the evidence for tests or series of tests that could be offered to individuals with newly diagnosed CRC in order to identify those with sufficient risk of Lynch syndrome to warrant mutation testing. The tests discussed are MSI, MMR protein IHC and *BRAF* gene mutation testing. The recommendations are intended to apply to all newly diagnosed CRCs irrespective of family history. In a population setting, the use of family history criteria (Amsterdam and Bethesda criteria) as a pre-screening strategy was felt to result in poor identification of Lynch syndrome patients (EGAPP 2009).

In this report, the clinical performance of MSI, MMR IHC and *BRAF* testing are summarised as follows:

- **MSI**: 80-91% sensitivity and 90% specificity for *MLH1* and *MSH2* mutation carriers. 55-77% sensitivity and 90% specificity for *MSH6* and *PMS2* mutation carriers
- **MMR IHC**: 83% sensitivity and 89% specificity for all genes
- **BRAF** mutation testing: (wild-type *BRAF* for the detection of Lynch Syndrome) sensitivity of 69% and specificity of nearly 100% when confined to cases with demonstrable loss of *MLH1* on IHC.

The authors of the report recommend that sufficient evidence exists to support population based genomic testing following incident CRC. However, areas lacking in evidence are highlighted:

1. Analytical validity- laboratory assays still vary widely and need to be properly validated prior to transition into clinical practice.
2. Clinical validity- better quality studies comparing the tests utility in a population based setting are required.

3. Clinical utility- a large multi-centre population based trial that begins with the genomic testing and follows through to clinical outcomes is required. Does the genomic testing improve outcomes for the Lynch families identified?

4. Cost-effectiveness analyses are required.

Since this report, further studies have been published which go some way to answering the research gaps highlighted.

Jensen et al retrospectively examined 287 unselected cases of CRC to test an algorithm of MMR IHC followed by \textit{BRAF} mutation testing in those with \textit{MLH1} loss, followed by \textit{MLH1} methylation analysis in those with wild type \textit{BRAF} in order to identify Lynch syndrome cases from incident CRC. The hypothesis was that \textit{BRAF} mutation would initially identify those with sporadic \textit{MLH1} loss, and \textit{MLH1} methylation testing would then identify further cases with likely \textit{MLH1} loss. Those with demonstrable loss of \textit{MSH2} (n=6), or \textit{MSH6} (n=1) were presumed to be possible Lynch syndrome cases. Those with loss of \textit{MLH1} (n=29) could be either sporadic loss or Lynch syndrome cases. These were then examined for \textit{BRAF} mutation. Of the 27 that had DNA of usable quality, 14 were found to have \textit{BRAF} V600E mutations and were presumed sporadic cases. Those with wild-type \textit{BRAF} (n=13) and had usable DNA (n=11) were subjected to \textit{MLH1} methylation testing. 10 were found to be hypermethylated and were therefore also presumed sporadic cases. From this study, the authors concluded that they had identified, from a typical risk population, six cases of possible \textit{MSH2} mutation, one possible \textit{MSH6} mutation and five cases of \textit{MLH1} loss on IHC that could be due to mutation, either due to lack of \textit{BRAF} mutation and lack of \textit{MLH1} methylation or because there was no usable DNA to rule out \textit{BRAF} mutation and \textit{MLH1} methylation. From family history review and follow up data, the authors concluded that they had diagnosed four definite, three probable and five possible Lynch syndrome families (Jensen, Dysager et al. 2010). Due to the retrospective nature of the study, the authors were unable to perform mutation testing on all the suspected cases. Although this study demonstrates the clinical utility of the algorithm in a population based
setting, the lack of complete MMR mutation test data resulted in an inability to calculate specificity and sensitivity values for the pre-screening tests utilised.

Bouzerene et al attempted to develop an algorithm for investigating patients with incident CRC with demonstrable MLH1 loss on IHC for possible MLH1 gene mutation. They examined 11 sporadic cases of MLH1 loss and 16 known MLH1 gene mutation carriers. Of the sporadic cases, 100% had MLH1 methylation, and 73% had BRAF V600E mutation. Of the Lynch syndrome cases, none had BRAF mutation, and 6% (only one case) had MLH1 methylation (Bouzourene, Hutter et al. 2010). Although these numbers are small, the results are consistent with the hypothesis of the current study.

Perez-Carbonell et al aimed to compare the ability of MLH1 promoter methylation analysis with BRAF mutation to select patients likely to have an MLH1 mutation. They examined 73 CRCs with MLH1 loss on IHC for germline MLH1 mutations, and found 10. The specimens were examined for MLH1 methylation and BRAF mutation. Specificities for the detection of MLH1 mutation carriers of 75% and 78% for the two methods of methylation analysis used were found. Specificity for the detection of mutation carriers of 40% was found for BRAF mutation. Of interest the authors found the methylation analysis to be 40% cheaper than BRAF mutation analysis (Perez-Carbonell, Alenda et al. 2010).

2.1.5 Applying pre-screening strategies to patients with different pre-test risks

The American Society of Clinical Oncology published a statement in 1996 regarding genetic testing for cancer susceptibility (ASCO 1996). These guidelines recommend that germline gene mutation testing be carried out if there is a 10% or greater chance of a mutation being detected. More recent guidelines from ASCO suggest that even this may be lowered based on clinical opinion (ASCO 2003).
Individuals who fulfil Amsterdam criteria have a 50-60% chance of having a *MLH1* or *MSH2* mutation (Park, Vasen et al. 2002).

Julie et al investigated 214 consecutive CRCs from patients with a mean age of 64.9 (range 30-93 years). 90 patients fulfilled Bethesda criteria, and 12 of these were found to have loss of MLH1 on IHC. Two germline *MLH1* mutations were found (Julie, Tresallet et al. 2008). From this study, it can be said that an individual who fulfils BC, and has loss of MLH1 in their tumour has a pre-test likelihood of a germline *MLH1* mutation of 2/12; 16.7%. Wright et al also recently published the results of seven years of routine MMR IHC in all patients who had a CRC aged 50 or younger (they thus fulfilled the revised Bethesda criteria). Of 243 patients, 214 were eligible. 12/214 had loss of MLH1, and 2/11 tested had a germline *MLH1* mutation. In this group of early onset cancers, the likelihood of having an *MLH1* mutation was 2/11, 18.1% (Wright, Arnold et al. 2011). Perez-Carbonell investigated 2093 consecutive CRCs. 486 fulfilled the revised Bethesda guidelines. Of these 53 demonstrated loss of MLH1 on IHC and four were found to have *MLH1* mutations (Perez-Carbonell, Ruiz-Ponte et al. 2011). This data suggests a likelihood of 4/53, 7.5%. These studies are the only population based studies in the literature that have examined the incidence of MMR gene mutations in patients fulfilling Bethesda criteria. From these three studies, 8/76 patients who fulfilled the revised Bethesda criteria and had loss of MLH1 were found to have an *MLH1* mutation, resulting in a likelihood of 10.5%.

Hampel et al investigated 1066 consecutive CRCs, from patients with a mean age of 62.9. 115 patients had loss of MLH1 in their tumour and five (4.3%) were found to have a germline *MLH1* mutation (Hampel, Frankel et al. 2005). Pinol et al investigated 1222 consecutive CRCs. 60 were found to have loss of MLH1, and four (6.7%) were found to have a germline *MLH1* mutation (Pinol, Castells et al. 2005). Aaltonen et al investigated 509 consecutive cases of CRC. 63 were found to be MSI-H, nine had germline *MLH1* mutations, and one had a germline *MSH2* mutation. Thus 9/62 (14.5%) cases with MLH1 loss were due to a *MLH1* mutation (Aaltonen, Salovaara et al. 1998). Perez-Carbonell investigated 2093 CRCs. Loss of MLH1 was found in 115 cases, of these four (3.4%) were found to have a germline *MLH1* mutation (Perez-Carbonell, Ruiz-Ponte et al. 2011). These studies suggest that the risk of having a pathogenic *MLH1* mutation for an individual from the general population who has a CRC with MLH1 loss is 4.3-14.5%. Several of the patients
examined in these studies, although they did not have a known familial mutation, did fulfil the Bethesda criteria. It is likely, therefore, that the risk of having an *MLH1* mutation in this risk group is towards the lower bound of 4%.

Applying the 10% rule as suggested by the American Society of Clinical Oncology, currently all patients who fulfil Amsterdam II criteria and those who fulfil the Revised Bethesda criteria and who have loss of MLH1 in their tumour would be tested for germline *MLH1* mutations. The consensus criteria for the diagnostic strategy for Lynch Syndrome suggest that tumour testing for mismatch repair deficiency (either MSI or MMR protein immunohistochemistry) should be conducted on all those fulfilling Amsterdam or revised Bethesda criteria (Umar, Boland et al. 2004). Patients demonstrating MSI-H or loss of MMR protein on immunohistochemistry should be subjected to germline mutation testing. However, less than a third of these will harbour a mutation (Pinol, Castells et al. 2005). Neither guideline advises testing patients from the general population with loss of MLH1 in their tumour.

By applying *BRAF* V600 mutation and *MLH1* promoter region methylation analysis it is hoped to further clarify which patients should be tested for germline MLH1 mutations.

### 2.1.6 Case and Control selection

In order to calculate the sensitivity and specificity of tumour DNA *BRAF* mutation and *MLH1* promoter region methylation testing for the identification of germline mutation carriers- a group of known pathogenic germline *MLH1* mutation carriers will be identified and tested, and a group of sporadic MLH1 loss cancers will be indentified and tested. The presence of *BRAF* mutation and *MLH1* promoter methylation will be compared between the two groups.

Germline *MLH1* gene mutation carriers will be identified using the Northwest Genetics Database. Patients with sporadic epigenetic silencing of MLHI will be identified (by MMR immunohistochemistry and tumour DNA microsatellite instability
analysis) from those who have had a colorectal cancer resected at the Manchester Royal Infirmary or Trafford General Hospital (aged over 50 and without a family history of CRC). A control group aged over 50, with no family history of colorectal cancer and no evidence of epigenetic silencing of MLHI will also be used. Inclusion criteria: patients aged 18 years of age of over. Exclusion criteria: those patients aged under the age of 18 or have previously placed a restriction on the use of their tissue. Patients with familial adenomatous polyposis coli and patients with inflammatory bowel disease will also be excluded.

2.1.7 Critique of laboratory methods

Formalin-fixed paraffin embedded (FFPE) tissue blocks of tumour and normal colonic tissue will identified and anonymised. For the purpose of immunohistochemistry, 4 x 4µm thick sections will be cut and prepared onto slides. For the purpose of DNA retrieval for MSI analysis, MLH1 promoter region hypermethylation and BRAF (codon 600) and KRAS (codon 12/13) mutation analysis 10-20 x 10 micron thick sections will be cut and stored at -20 °C.

1. Immunohistochemistry will be used to determine the mismatch repair status of the control patients. A validated semi-quantitative scoring system will be used to assess the presence of mismatch repair proteins MLH1, MSH2, and MSH6. This system is based on the intensity of the staining and the percentage of cells stained. Normal colonic tissue from each patient will be used as an internal control.

2. Microsatellite instability will be assessed using tumour DNA from all patients in all groups to further verify MMR deficiency.

3. Pyrosequencing will be used to analyse MLH1 promoter region hypermethylation, and BRAF and KRAS mutations. This allows real-time quantitative high throughput analysis. To our knowledge, this technique has not been used in this context before. DNA retrieved from the FFPE tissue blocks, and amplified using polymerase chain reaction amplification will be used. Stored DNA with known KRAS, and BRAF mutations and known universally hypermethylated DNA will be used as positive controls.
2.1.7.1  Analysis of Mismatch repair status

There are currently two available techniques to determine the mismatch repair status of a given tissue sample. As previously described, these are MMR protein IHC and MSI. These techniques have comparable sensitivity and specificity for the detection of mismatch repair deficiency due to MLH1 loss. As summarised in the EGAPP report discussed in section 1.10.4; MSI: 80-91% sensitivity and 90% specificity for MLH1 and MSH2 mutation carriers and 55-77% sensitivity and 90% specificity for MSH6 and PMS2 mutation carriers, MMR IHC: 83% sensitivity and 89% specificity for all genes.

MMR IHC costs £45 per patient, and can be performed as part routine histopathological assessment. Using an automated Dako Techmate Plus (Dako, Cambridgeshire, UK), the test is performed in less than 24 hours. The interpretation of the test is according to the previously described scoring system devised by the ICG-HNPCC. (Muller, Burgart et al. 2001)

MSI analysis requires extraction of tumour and normal tissue DNA, amplification by polymerase chain reaction, and comparison using a variety of sequencing or visualisation techniques.

In this project, MMR IHC and MSI will be used to determine the mismatch repair status of the local sporadic CRC patients. This will allow identification of sporadic MMR deficient CRCs. It would be impracticable to consider routine MSI as a population based risk assessment biomarker in an NHS setting. It is also thought that, in the clinical setting, IHC is now preferred due to the poor sensitivity of MSI for MSH6 mutations.

2.1.7.1.1  Immunohistochemistry

Immunohistochemistry is a pathology laboratory technique used to identify and visualise target proteins (antigens) in a biological tissue or cytological specimen. The specimen may be formalin-fixed, paraffin embedded tissue, fresh or frozen tissue, or a cytological specimen. This technique utilises antibodies which bind to the target antigen, and a visualisation or detection system.
Antibodies are glycoproteins which belong to a class of proteins known as immunoglobulins. The five major classes of antibodies are IgA, IgM, IgD, IgE and IgG. These are Y shaped proteins, the tip of which is termed a paratope and is able to bind specifically to an epitope which is an area of the target antigen. Physiologically, antibodies act to recognise antigens on foreign material such as viruses and bacteria and form the major part of the humoral immune system.

In immunohistochemistry, antibodies are utilised to bind to a specific antigen of interest. These antibodies can be polyclonal or monoclonal. Polyclonal antibodies are produced from multiple cell lines, typically through the use of an immunized animal. Monoclonal antibodies are produced from a single cell line (clones of a plasma cell).

Primary antibodies are used to bind directly to an epitope. The epitope may be visualised directly by the use of a detection system that binds to the primary antibody, or indirectly by the use of a secondary antibody which binds to the primary antibody. In indirect immunohistochemistry, a detection system is used that binds to the secondary antibody. Use of direct immunostaining is quicker and limits the amount of non-specific reactivity, but can produce weak visualisation. Indirect staining allows amplification of visualisation through the use of a secondary antibody, and also allows the detection of multiple antigens (by the use of multiple primary antibodies) if appropriate.

The tissue used for immunohistochemistry will have been fixed and preserved by one of a variety of methods so must first undergo steps to make the antigen available for binding. For example deparaffinisation and antigen retrieval in the case of formalin fixed paraffin embedded (FFPE) tissue.

Detection systems can be enzymatic or fluorescent. Enzymatic systems utilise conjugated antibodies (either primary or secondary depending on whether the staining method is direct or indirect) which are then exposed to a chromogen substrate which is colourless. An enzymatic reaction converts this substrate to
stable coloured precipitate. As this is bound to the antigen of interest, the now coloured antigen-antibody-polymer complex can essentially be visualised using microscopy. Fluorescent detection systems utilise primary or secondary antibodies that are bound to fluorochromes (fluorescent molecules) and the antigen-antibody complex can then be visualised using a fluorescence microscope.

Immunohistochemistry can be performed manually or using an automated system. Automated systems, which can process multiple slides concurrently, are now in use in most clinical laboratories. Their use allows rapid optimisation of assays and improves throughput time and reproducibility (Renshaw 2007).

**2.1.7.1.2 Microsatellite Instability Analysis:**

Microsatellite instability analysis requires both tumour DNA and normal DNA for comparison. Several methods of MSI analysis are available all of which rely on polymerase chain amplification of the tumour and normal DNA. The PCR products may be visualised on a polyacrylamide gel using silver or ethidium bromide, or radio-isotope labelled oligonucleotide primers may be employed in the PCR step which then can be visualised using gamma camera technology. Automated, fluorescence based DNA sequencers can also be used. These systems allow automated comparison of the amplified PCR products. Both methods are interpreted by comparing the size (in base pairs) of the amplified microsatellite alleles (either visually or using software within an automated system), and when the tumour allele is larger than the normal allele, the tumour is said to be unstable at that particular marker (Bocker, Diermann et al. 1997).

The MSI Analysis version 2.1 (*Promega*), is a fluorescent multiplex polymerase chain reaction based system. It utilises two mononucleotide repeats and three dinucleotides repeats (D2S123, DS5346 and D175250). These markers meet the requirements of the NCI (US National Cancer Institute) 2002 workshop on MSI. The kit also encompasses two pentanucleotide markers (Penta C and Penta D) which are used to ensure that the matched normal and tumour DNA are truly from the same patient. These markers are said to more sensitive than the original five marker panel laid out at the Bethesda meeting.
2.1.7.2 KRAS and BRAF mutation testing

Many techniques are available to test for gene mutations. The technique suitable for use in a particular test depends upon the type of mutation being investigated. Techniques such as denaturing gradient gel electrophoresis, denaturing high performance liquid chromatography and single strand conformational analysis are suitable for detecting small insertions and deletions, and single nucleotide polymorphisms. Multiplex ligation-dependant probe amplification, which is a type of polymerase chain reaction, is capable of analysing for insertion and deletions that encompass whole exons. Gene sequencing, either by Sanger sequencing (a chain termination technique) or by high throughput automated sequencing platforms, can be utilised to sequence larger sections of genes, or whole genes. These include technologies such as Big Dye Terminator cycle sequencing kit (Applied Biosystems, Foster City, CA) and ABI Prism 3900 DNA Analyzer automated sequencer (Applied Biosystems).

All these techniques rely on polymerase chain reaction amplification of the DNA to be sequenced. A polymerase chain reaction is a technique used in molecular biology to selectively amplify small amounts of DNA to allow sequencing. This technique generates thousands of copies of a particular DNA sequence through an exponential DNA cloning chain reaction. Primers (short sequences of bases which are complimentary to the 3’ ends of the sense and antisense DNA) are specifically designed to initiate the reaction. DNA polymerase enzymes and thermal cycling are employed to facilitate and catalyse the reaction. The necessary components for a PCR reaction are DNA template, heat stable DNA polymerase (e.g Taq DNA polymerase), deoxynucleotide diphosphates (dNTPs- utilised by the DNA polymerase to synthesise the new complimentary strand of DNA), buffer solution which optimises the enzyme used, divalent Mg\(^{2+}\) cations and monovalent K\(^{+}\) cations.

The PCR reaction typically utilises 20-40 cycles of temperature changes. The exact temperature will depend on a variety of parameters. The steps within the cycles are:

1. Initialisation: The reaction is heated to around 90°C to heat activate the DNA polymerase. This step is only required once.
2. Denaturation: This step heats the reaction to around 94-98 °C and the heat disrupts the hydrogen bonds between the complementary bases so denatures the DNA to single stranded DNA.

3. Annealing: This step is at a lower temperature and allows the primers to anneal to their complementary sequence of DNA. The DNA polymerase binds to the primer-DNA template complex and DNA synthesis commences.

4. Extension: This step occurs at 70-80 °C (or whatever is the optimum temperature of the DNA polymerase). The DNA polymerase adds dNTPs to the template DNA in the 5’ to 3’ direction.

5. Final Extension: This is the final step and occurs at around 70 °C. This ensures that all single stranded DNA is fully extended.

2.1.7.2.1 Pyrosequencing
Pyrosequencing is a method of DNA sequencing that was designed by Ronaghi et al in 1996 (Ronaghi, Karamohamed et al. 1996). This technique relies on the sequencing by synthesis principle, and the detection of the activity of DNA polymerase enzyme by a chemoluminescent enzyme as the DNA is synthesised. The region of a gene of interest is amplified using a polymerase chain reaction, and denatured to produce a single strand. A mix of the appropriate nucleotide sequencing primer and the specified concentration of nucleotides (dNTPs) are added with a number of enzymes. The sequencing primer binds to the single stranded DNA under analysis and initiates the sequencing reaction. As the nucleotides are incorporated into the complementary strand of DNA, pyrophosphate is released which in turn fuels a reaction that produces a quantifiable amount of light per nucleotide added. (See Figure 4) The sequencer software produces a pyrogram: a diagrammatical representation of the amount, and sequence in which, each nucleotide was incorporated.

This technique allows rapid, high throughput analysis. Each run can analyse 96 samples and takes only 90 minutes. Pyrosequencing is limited to the sequencing of 300-500 nucleotides only, whereas Sanger sequencing and other chain termination techniques are capable of sequencing 800-1000 nucleotides. Pyrosequencing is suitable for sequencing of DNA to investigate for known short mutations (e.g. point mutations in two adjacent codons as in the KRAS 12/13 mutations, or small insertion or deletions).
Figure 4: Pyrosequencing
2.1.7.3  **MLH1 gene promoter hypermethylation analysis**

There are several techniques that are widely used to detect DNA methylation patterns. These include methylation-specific PCR, methylation-specific MLPA (multiplex ligation-dependant probe amplification), real time PCR, the ChIP GLAS (Chromatin Immuno Precipitation and Guided Ligation And Selection) procedure, and pyrosequencing. These techniques all rely on bisulphite assisted modification of the DNA. This process allows differentiation between the methylated cytosines and the unmethylated cytosines in a given DNA sequence. During the bisulphite modification, the DNA is denatured to become single stranded, and the unmethylated cytosine residues are converted (deaminated) to thiamine. The methylated cytosine residues are resistant to this deamination thus remain unchanged within the sequence. The DNA can then be sequenced and the methylation pattern identified.

<table>
<thead>
<tr>
<th></th>
<th>Original Sequence</th>
<th>After Bisulphite treatment</th>
</tr>
</thead>
</table>

2.1.7.3.1  **Methylation specific PCR**

Methylation specific PCR, or MSP, was first described by Herman et al in 1996 ([Herman, Graff et al. 1996](#)), and is widely used to detect DNA methylation patterns within the epigenetic regions of tumour suppressor genes. MSP is able to accurately detect methylation at any CpG site. It has a sensitivity of up to 99%. MSP relies on an initial bisulphite treatment stage. The DNA is amplified using a PCR with nucleotide primers which are designed to amplify a region of DNA which includes the CpG sites in question. After PCR the uracil from the unmethylated DNA is converted to thymine, and the methylated cytosine remains unchanged. MSP can be used to analyse small volumes of DNA such as those derived from FFPE tissues. MSP is able to detect methylation in any CpG island even within very large genes (anything greater than 0.1% methylated alleles in a particular CpG locus). This
technique is quick and high throughput so is ideal for clinical samples. The major disadvantage with MSP is the qualitative not quantitative nature of its results. The technique is able to determine if there is significant methylation at a particular CpG island but is not able to quantify this. The level/percentage of methylation, when considering tumour suppressor genes is very important. It has previously been demonstrated that promoter methylation of 10% or greater silences causes loss of function of the MLH1 gene (Vaughn, Wilson et al. 2010).

2.1.7.3.2 Methylation specific-MLPA

Methylation specific multiplex ligation-dependant probe amplification utilises DNA probes designed for the site of interest. These are annealed to denatured tumour DNA and then ligated using a methylation-sensitive end nuclease that digests unmethylated sites. Gel electrophoresis is used to visualise the products of the reaction and a percentage of methylation can be estimated. This technique requires a larger concentration of DNA than MSP so is less applicable for use with FFPE samples. As with MSP, this technique is qualitative rather than quantitative. Although the percentage of methylation can be estimated, this is purely a visual estimation. Therefore this method does not precisely quantify the level of methylation.

2.1.7.3.3 PCR

Non-methylation specific PCR utilises direct sequencing of the area of interest.

2.1.7.3.4 ChIP-GLAS Procedure

This technique allows whole gene methylation profiling. As I am only investigating methylation at CpG islands within the MLH1 gene promoter region, this technique is not applicable.

2.1.7.3.5 Pyrosequencing for methylation analysis

This technique allows real time sequencing by synthesis, as described in section 2.1.7.2.1. When applied to methylation analysis this technique uses bisulphite modified DNA that has been amplified using a polymerase chain reaction. The
pyrosequencing process is able to accurately quantify the percentage of methylation at the sites of interest. Pyrosequencing allows high throughput, real time assessment. The concentrations of DNA required to get interpretable results is small thus this is highly suitable for use on FFPE clinical samples. The main benefits of this platform are that it allows quantification of the level of methylation and that multiple tests can be performed on multiple patients at one time. Thus, \textit{BRAF} V600E mutation, \textit{KRAS} codon 12 and codon 13 mutations, and methylation of \textit{MLH1} gene promoter region can all be performed at the same time. And the results will be produced at the same time.

In the North West Regional Molecular Genetics laboratory, the pyrosequencing assay for \textit{KRAS} and \textit{BRAF} mutation has been developed and validated as a clinical diagnostic test. This test is being utilised by the regional oncology department to assess patient suitability for anti-EGFR antibody therapy in metastatic CRC. Anti-EGFR therapy is only efficacious in patients with wild-type \textit{KRAS} and \textit{BRAF} genes. Pyrosequencing, as an assay for this clinical application, was chosen due to its ease, speed, reproducibility, and high though put. Currently only chemotherapy refractory metastatic CRCs are considered for anti-EGFR therapy. However, a large multicentre trial is currently examining the efficacy of these therapies as first line treatment. It is reasonable to expect that a higher proportion of CRCs patients will be offered anti-EGFR therapy in the future, thus will require \textit{BRAF} and \textit{KRAS} genotyping. It is feasible to think that future research, including this study, will provide evidence for routine \textit{MLH1} methylation analysis in mismatch repair deficient CRCs. As \textit{BRAF} and \textit{KRAS} analysis for assessment for anti-EGFR treatment also becomes more widely utilised it is sensible to validate assays for all three genetic/epigenetic variants using the same platform.

2.1.8 Aims of study one

The aim of this study was to: 1) develop a simple, cheap, reproducible method for quantitative \textit{MLH1} promoter methylation analysis in FFPE tissue and, 2) compare it with \textit{BRAF} p.V600E mutation testing in patients whose CRCs demonstrate loss of MLH1 protein expression to select patients for germline \textit{MLH1} mutation testing.
In order for a risk assessment biomarker panel such as this to have clinical utility, there are several important factors to consider. Firstly, the specimens used must be readily clinically available. Currently the vast majority of CRC specimens in the UK are fixed in formalin immediately following surgical resection and then embedded in paraffin after macroscopic pathological examination. Thus formalin fixed, paraffin embedded (FFPE) samples were used. Secondly, the tests must be cost effective, time effective and easily reproducible. The laboratory methods that have been chosen fulfil these criteria. The panel of tests (KRAS, BRAF, MLH1 methylation) are going to be performed using the same platform. This technique should be simple, repeatable and reproducible, as well as cost and time effective.

2.1.9 Hypothesis of study 1

Mismatch repair deficient colorectal cancers can be attributed either to epigenetic alterations leading to inactivation of MLH1 or germline mutations in mismatch repair genes. These cancers should have mismatch repair deficiency demonstrable by deficient staining of MMR proteins on immunohistochemistry. Somatic MMR deficient cancers display BRAF gene mutation and MLH1 epigenetic gene silencing via MLH1 promoter region hypermethylation, whereas Lynch syndrome cancers display wild type BRAF gene and lack of MLH1 hypermethylation. KRAS and BRAF gene mutation appear to be mutually exclusive in colorectal cancers, so KRAS analysis may further help to select the Lynch syndrome cancers. Formal sensitivity and specificity testing of these biomarkers in a cohort of patients with Lynch syndrome and sporadic colorectal cancers has not previously been performed. The hypothesis is that this panel of biomarkers will accurately determine which patients (with an MMR deficient CRC) have a 10% or greater risk of harbouring an MLH1 mutation and therefore should undergo germline analysis.
2.2 Methods Study one: DNA mismatch gene promoter region methylation analysis and *BRAF* gene mutation analysis—an alternative pre-screening strategy in Lynch Syndrome.

2.2.1 Ethical Approval

Ethical approval was obtained from South Manchester Research Ethics Committee. Project reference number: 10/H1003/11

As per the human tissue act 2004, residual diagnostic tissue may be stored for future use and be used for research without prior patient or relative consent if the study is ethically approved and the tissue is anonymised. In this study clinical data was used to identify the archived tissue samples. Following this the samples and information was anonymised.

2.2.2 Participants

2.2.2.1 Cases

Lynch Syndrome *MLH1* mutation carriers who have had a colorectal cancer since 2005 were identified from the Manchester Familial Colorectal Cancer Registry. Only cancers that have occurred since 2005 were used, as older FFPE specimens were unlikely to yield an adequate concentration or quality of DNA for methylation analysis (oral communication Dr Andrew Wallace PhD, Senior Clinical Scientist, Saint Mary’s Hospital Molecular Genetics Laboratory, September 2009). Mutation carriers are defined as those who have tested positive for a pathogenic germline mutation, and those who are obligate mutation carriers as they have had a Lynch spectrum cancer and due to their position in a mutation positive family pedigree.
Families from the North West with a family or personal history suggestive of Lynch Syndrome were referred to the department of Genetic Medicine at St Mary’s Hospital. Full pedigree information was taken. Dates of birth, dates of death, and cancer incidence were confirmed from birth and death certificates, medical records and the North West Cancer Intelligence Service (NWCIS; cancer registration data) and NHS Summary care records. After 1996 families who fulfilled Amsterdam criteria or Bethesda criteria were tested directly for germline mutations in the mismatch repair genes (if there was an available living affected relative). Multiplex Ligation Probe Dependant Amplification (MLPA) was used to screen all exons of the MSH1, MSH2 and latterly MSH6 genes. For all other patients tumour samples were tested (if available) for evidence of mismatch repair deficiency. Prior to 2008 microsatellite testing was conducted but due to concerns regarding high false negative rates for MSH6 mutations, and poor cost and time efficiency, mismatch repair immunohistochemistry has been used since 2008 as the pre-screening test. If the tumours demonstrated mismatch repair deficiency, and there was an available living affected relative, germline mutation testing was then performed.

The INSiGHT website (www.insight-group.org Accessed 2011) was used to assess the pathogenicity of all mutations found.

Genetic counselling and appropriate screening is offered to appropriate family members (mutation carriers and those at 50% risk) known to the Manchester Familial Colorectal Cancer Database in accordance with the British Society of Gastroenterologist guidelines for colonic and upper gastrointestinal screening (Cairns, Scholefield et al. 2010). Females are offered gynaecological screening (annual transvaginal ultrasound, hysteroscopy and endometrial aspiration biopsy from age 35). The International Collaborative Group for Hereditary Non-polyposis CRC reviewed the literature in 2007 and at that time found that the effectiveness of endometrial screening is unknown. This group reported that the literature suggests that screening may lead to the detection of pre-malignant lesions or endometrial cancers at an earlier stage (Vasen, Moslein et al. 2007). Since that report a further retrospective study has been published (Stuckless, Green et al. 2012). Only two of the four available studies now suggest that screening may be beneficial (Dove-Edwin, Boks et al. 2002; Rijcken, Mourits et al. 2003; Renkonen-Sinisalo, Butzow et al. 2007; Stuckless, Green et al. 2012). Although the benefit of endometrial screening remains unclear, it is current practice at Saint Mary’s Hospital to offer
screening to all mutation carriers and patients at 50% risk from age 30 to 35 (verbal communication, Dr Tara Clancy PhD; Consultant Genetic Counsellor July 2012).

The American Society of Clinical Oncology published a statement in 1996 regarding genetic testing for cancer susceptibility (ASCO 1996). These guidelines recommend that gene mutation testing be carried out if there is a 10% or greater chance of a mutation being detected (as determined by personal history, family history, pre-screening tests or risk calculation models). Applying the Amsterdam Criteria to families will only detect families with a 50-60% chance of having a Lynch syndrome mutation. If we apply Bayes theorem to the cut-off of 10% chance of carrying a mutation, a sensitivity of at least 86% for a pre-screen test is necessary to identify individuals who should undergo germline testing (Evans, Laloo et al. 2006).

I consulted Dr Stephen Roberts PhD (Applied Biostatistician and Senior Lecturer in Biostatistics at the University of Manchester) for advice regarding sample size. He advised that as these assays are new, a formal sample size calculation was not possible. However, we decided that 60 per group was a practical number. Dr Roberts advised that a sample size of 60 would give a reasonably accurate first estimate of the sensitivity/specificity as described by the width of the confidence interval. Dr Roberts estimated that with an expected sensitivity of 95% of the two assays being investigated (based on the literature and an expert opinion from Dr Andrew Wallace, Senior Clinical Scientist) a sample size of 60 would be adequate to demonstrate a sensitivity of at least 86% (see paragraph above).

In order to ensure the inclusion of at least 60 cases, I applied to the biospecimens committee of the Colon Cancer Family Registry (C-CFR) for use of biospecimens. The Colon Cancer Family Registry is an international consortium of six centres in North America and Australasia. “The Colon Cancer Family Registry is an international research infrastructure for investigators interested in conducting population and clinic-based interdisciplinary studies on the genetic and molecular epidemiology of colon cancer and its behavioural implications. A central goal of the C-CFR is the translation of this research to the clinical and prevention setting for the benefit of Registry participants and the general public. The C-CFR has information and biospecimens contributed by greater than 11,300 families across
the spectrum of risk for colon cancers and from population-based or relative controls” (http://epi.grants.cancer.gov/CFR/about_colon.html Accessed 2011). C-CFR application number C-EX-0710-01. This application was successful. The biospecimens committee granted me use of enough FFPE tissue to extract DNA for BRAF gene mutation and MLH1 gene promoter region hypermethylation analysis from 40-50 MLH1 mutation carriers and 40-50 sporadic mismatch repair deficient cancers.

The additional cases were provided by The Jeremy Jass Memorial Pathology Bank, Australasian Colorectal Cancer Family Registry (U01 CA097735)(http://epi.grants.cancer.gov/CFR/about_colon.html Accessed 2011). MMR IHC was performed to identify CRCs with loss of MLH1 protein expression as previously described (Walsh, Cummings et al. 2008). Screening for germline mutations in MLH1, MSH2, MSH6 and PMS2 was performed for all probands recruited from high-risk clinics (clinic-based) and for population-based probands who had a CRC displaying evidence of MSI or loss of MMR protein expression by IHC. Mutation testing was performed by Sanger sequencing or denaturing high performance liquid chromatography (dHPLC), followed by confirmatory DNA sequencing. Large duplication and deletion mutations were detected by Multiplex Ligation Dependent Probe Amplification (MLPA) according to the manufacturer’s instructions (MRC Holland, Amsterdam, The Netherlands) (Southey, Jenkins et al. 2005; Senter, Clendenning et al. 2008). A pathogenic mutation was defined as a variant that was predicted to result in a stop codon, a frameshift mutation, a large duplication or deletion, or a missense mutation previously reported within scientific literature and databases to be pathogenic.

2.2.2.2 Controls

Consecutive patients over the age of 60 with a histologically proven right-sided CRC since 2005 were identified from the histopathology database at the Manchester Royal Infirmary, Central Manchester University Hospitals Trust and Trafford General Hospital. Right-sided cancers were selected as it is known that sporadic mismatch repair deficient cancers occur more frequently in the right colon than the left colon or rectum (Jass 2007). The medical case notes of those patients whose cancers were found to have loss of MLH1 protein were retrospectively examined to ensure that the patient did not have a family history that fulfilled either the Amsterdam or the revised Bethesda criteria. All patients in the sporadic mismatch repair group
had a documented family history in an outpatient clinic letter, or a hospital admission clerking and this was either no family history or CRC, or a minor family history which does not fulfil moderate risk criteria or Bethesda criteria. This family history is not a true pedigree as would be documented and verified during a genetic medicine consultation. However, it was not possible (due to restraints of ethics committee approval) to contact these patients in order to verify their family history of cancer.

The additional sporadic MLH1 loss cases were provided by The Jeremy Jass Memorial Pathology Bank, Australasian Colorectal Cancer Family Registry (U01 CA097735)(http://epi.grants.cancer.gov/CFR/about_colon.html Accessed 2011). CRC cases with MLH1 loss were classified as sporadic based on the presence of the BRAF p.V600E mutation and/or methylation of the MLH1 gene promoter and did not harbour a pathogenic mutation in the MLH1 gene. Detection of the BRAF p.V600E mutation was determined on CRC tissue DNA using an allele-specific PCR assay as has been previously described (Buchanan, Sweet et al. 2010). Methylation of the MLH1 gene promoter region was assayed using MethyLight qPCR on sodium bisulfite converted tissue DNA where samples with a percent of methylated reference (PMR) greater than or equal to 10 were classified as positive for MLH1 methylation (Weisenberger, Siegmund et al. 2006; Poynter, Siegmund et al. 2008).

2.2.3 Tissue Specimens

Case archived formalin-fixed paraffin-embedded (FFPE) tissue specimens were requested from the hospital at which the cancer was resected. Three blocks were requested for each case; one block of mixed tumour and normal tissue for mismatch repair protein immunohistochemistry, one block of at least 80% tumour cells for tumour DNA extraction and one block of normal tissue for normal DNA extraction. On receipt of these blocks they were anonymised. Control archived FFPE tissue specimens located at Manchester Royal Infirmary, Central Manchester University Hospitals Trust and Trafford General Hospital were retrieved and anonymised. For all cases and controls diagnosis was confirmed using the histopathology reports and a haematoxylin and eosin (H&E) slide of each block.
The FFPE tissue samples obtained from The Jeremy Jass Memorial Pathology Bank, Australasian Colorectal Cancer Family Registry based at Familial Cancer Laboratory, Queensland Institute of Medical Research (QIMR) were sent as 2-5x 4 µm thick tumour sections on microscope slides (unmounted) per sample. This amount of tissue was adequate to conduct MLH1 methylation and BRAF mutation analysis only.

2.2.3.1 Tissue specimen preparation

3.2.3.1.1 Slide preparation (Manchester samples only)

All FFPE blocks were sectioned using a microtome by an experienced technician. 4µm sections were required for the H&E stain and the mismatch repair immunohistochemistry. These sections were mounted on Dako REALcapillary gap microscope slides (Dako, Cambridgeshire, UK) and placed in a dryer overnight at 37°C overnight. The slides were then heated at 70°C for 30 minutes to liquify the wax.

2.2.3.1.2 Tissue preparation for DNA extraction

As mentioned H&E slides were performed to confirm the contents of each FFPE tissue block from all local cases and controls. One block containing at least 80% tumour cells (visual estimation performed by the author and verified by Dr R McMahon, Senior Lecturer and Honorary Consultant Gastrointestinal Histopathologist) and one block of normal tissue, for each case and control were sectioned for DNA extraction. This visual estimation method is an accepted method of selecting FFPE tumour samples for molecular genetics assays in clinical practice (oral communication Dr Andrew Wallace PhD, Senior Clinical Scientist, Saint Mary’s Hospital Molecular Genetics Laboratory, September 2009). 10 x 10µm thick sections per block were placed in a sterile labelled 1.5 ml ependorff container. These were stored at -30°C.

The ACCFR samples were provided as 2-5 x 4µm thick FFPE slices on microscope slides (unmounted). To enable DNA extraction, the tissue was scraped off the slides with a sterile No.11 blade scalpel and placed into a 1.5ml ependorff for storage.
2.2.4 Histopathology (Manchester samples only)

Two histopathological techniques were used.

1. Haemotoxylin and eosin stain was used to verify the cellular contents of the FFPE tissue blocks

2. Mismatch repair immunohistochemistry was used to determine the mismatch repair status of all the sporadic tumour and normal samples

See appendix one for full methods.

2.2.4.1 Mismatch repair protein immunohistochemistry interpretation

The three MMR IHC slides for each case and control were assessed and scored according to a previously validated semi-quantitative scoring system (Barrow, Jagger et al. 2011). The slides were scored separately by the author and an experienced Consultant Gastrointestinal Pathologist. Both were blinded to the others score and the patient’s mutation status.

The presence of normal cells within the specimen was used as an internal positive control (benign colonocytes, lymphocytes). Nuclear staining of the malignant colonocytes was given a score for intensity and a score for percentage positivity. The intensity score is on a visual scale of 0-3. This is a comparison between the intensity of the positive internal control and the malignant cells. Score 0: no stain present in tumour cells, score 1: nuclear stain present in malignant cells but much paler than the internal positive control cells, score 2: nuclear stain present in malignant cells but a little paler than the internal positive control cells, and score 3: malignant cells stain equivalent to stain of normal cells (Barrow, Jagger et al. 2011). The percentage positivity scoring was described by the ICG-HNPCC study (Muller, Burgart et al. 2001). A score 0 indicates no nuclear stain within the malignant cells. Score 1= 1-10% of malignant cells stained, score 2=11-50% of malignant cells stained, score 3=51-80% of malignant cells stained, score 4= >80% of malignant cells stained. The intensity score and percentage positivity score were multiplied together to give a score from 0-12.
Slides where there was a disagreement in score between the two observers, or where there was an incongruity between expected and observed results, were restained twice and reassessed in a blinded manner.

A score of ≤4 was interpreted as loss of that specific protein. Barrow et al used CRCs from a group of known pathogenic Lynch syndrome mutation carriers and a further group of patients with sporadic CRC to test the semi-quantitative scoring system for mismatch repair protein immunohistochemistry (Barrow, Jagger et al. 2011). The investigators used receiver–operating characteristic (ROC) curves of the staining performance of each specific antibody and estimated the area under each ROC curve was using a nonparametric approach. The sensitivity and specificity of each observed score for each antibody were plotted, and the authors concluded that a score of ≤4 can be used to identify carriers of Lynch syndrome mutations. For the MLH1 stain this has a sensitivity of 100% (95% confidence interval 84.0-100%) and a specificity of 91.5% (95% confidence interval 79.6-97.6%) (Barrow, Jagger et al. 2011).

### 2.2.5 Molecular genetics

**Manchester samples:**

10 x 10µm sections of FFPE tissue per sample were booked into the regional molecular genetics laboratory to use for:

1. Microsatellite instability to verify mismatch repair status
2. *BRAF* V600E mutation and and *KRAS* codon 12/13 gene mutations
3. *MLH1* promoter region hypermethylation

**ACCFR samples:**

2-5 x 4µm sections of FFPE tissue per samples were booked into the regional molecular genetics laboratory for use for:

1. *BRAF* V600E mutation analysis
2. *MLH1* promoter region hypermethylation
Each sample was assigned a laboratory reference identification number in order that they be processed according to the same stringent checks as clinical samples. These checks include; two person check each time a sample is moved between containers, and two person check when tissue samples are loaded onto an automated DNA extraction robot. The FFPE samples were kept at 2 °C in the fridge untill needed.

The assays used to analyse for MSI, and BRAF and KRAS mutations are currently used as clinical standard (GCLP) assays in the Saintt Mary’s Hospital regional molecular genetics laboratory. Therefore these assays have previously been optimised and validated. The assay for the quantification of hypermethylation of the MLH1 gene promoter region was a new assay thus all stages were optimised, and validated.

2.1.5.1 Microsatellite Instability

The microsatellite (MSI) assay used is a fluorescent polymerase chain reaction-based assay. The MSI Analysis Version 1.2 (Promega, USA) was used in conjunction with ABI PRISM® 3100 genetic analyser. The assay involves comparison of alleic profiles of microsatellite markers which are generated by PCR amplification of matched normal and tumour DNA. The large alleles that are present in the tumour derived DNA but not in the matched normal DNA indicate MSI. The MSI alleles are larger than the normal counterparts due to the multiple nucleotide repeats. The size of the DNA fragments are detected as fluorescence. The assay kit includes an Internal Lane Standard (ILS 600) which contains 22 fragments of DNA of differing but known lengths. These fragments are labelled with carboxy-X-rhodamine and are detected separately by the software in the analyser. This allow quantification of the size of the PCR products. The kit also contains a positive control of genomic DNA (K562) that is MSI at all the loci tested.

The kit enables analysis of five microsatellite loci. Instability of these markers has high specificity and sensitivity for tumours with mismatch repair defects. Almost all individuals are homozygous for the same common allele for each marker (monomorphic). The kit also encompasses two pentanucleotide markers (Penta C and Penta D) which are used to ensure that the matched normal and tumour DNA
are truly from the same patient (i.e. to identify sample mix-ups). These two markers are highly polymorphic (therefore variable between individuals). Promega who produce the kit quote a matching probability (the number of randomly selected individuals that you would need to test in order to find two with the same allele) of between $1 \times 10^{10}$ and $1 \times 10^{16}$ for Penta C and $1 \times 10^{18}$ and $1 \times 10^{33}$ for Penta D (based on allele frequency data) (Technical Manuel MSI Analysis System, Version 1.2, Instructions for use of product MD1641). From this, promega state that sample mix-ups should be detected 99% of the time.

Table 7: *MSI Analysis Version 1.2 (Promega, USA) Locus Information*

<table>
<thead>
<tr>
<th>Marker Name</th>
<th>GenBank Number</th>
<th>Major Repeat Sequence</th>
<th>Size range (bp)</th>
<th>K562 alleles (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR-21</td>
<td>XM_033393</td>
<td>(A)21</td>
<td>94-101</td>
<td>101</td>
</tr>
<tr>
<td>BAT-26</td>
<td>U41210</td>
<td>(A)26</td>
<td>103-115</td>
<td>113</td>
</tr>
<tr>
<td>BAT-25</td>
<td>L04143</td>
<td>(A)25</td>
<td>114-124</td>
<td>122</td>
</tr>
<tr>
<td>NR-24</td>
<td>X60152</td>
<td>(A)24</td>
<td>130-133</td>
<td>130</td>
</tr>
<tr>
<td>MONO-27</td>
<td>Ac007684</td>
<td>(A)27</td>
<td>142-154</td>
<td>150</td>
</tr>
<tr>
<td>Penta C</td>
<td>AL138752</td>
<td>(AAAAG)3-15</td>
<td>143-194</td>
<td>164, 174</td>
</tr>
<tr>
<td>Penta D</td>
<td>AC000014</td>
<td>(AAAAG)2-17</td>
<td>135-201</td>
<td>168, 187</td>
</tr>
</tbody>
</table>

This assay has three main steps.

1. DNA extraction
2. DNA amplification
3. Detection of amplified fragments using the ABI PRISM® 3100 genetic analyser

Each kit performs 100 reactions (50 patients).

See Appendix two for protocol.
2.2.5.1.1 Interpretation of electropherograms produced by the ABI PRISM® 3100 genetic analyser

Two electropherograms were produced per tumour. One from tumour derived DNA and from DNA derived from normal tissue from the same patient. Interpretation of the microsatellite status of the tumour tissue required comparison between these two electropherograms. Firstly the peaks which represent the Penta C and D pentanucleotides were compared to ensure that the tumour DNA and normal DNA samples were truly matching. The five mononucleotide markers were then compared. The appearance of novel alleles (larger signal) in the tumour DNA signifies microsatellite instability. Instability at two or more of the markers was considered microsatellite instability-high. Instability at one of the markers was considered microsatellite instability-low. No instability was considered microsatellite stable (EGAPP 2009).

Figure 5 is an electropherogram from a tumour DNA sample that is microsatellite stable. The peaks for each marker are the same between the tumour and the normal DNA.
Figure 5: Electropherogram for a microsatellite stable tumour sample
Figure 6 is an electropherogram from a tumour DNA sample that demonstrates microsatellite instability at all five markers. The peaks for each marker are larger in the tumour DNA than the normal DNA.

Figure 6: Electropherogram for an MSI-high tumour
Figure 7 is an electropherogram from a tumour that demonstrates microsatellite instability at two markers; Bat25 and Bat26.

2.2.5.2 *KRAS* and *BRAF* gene mutation analysis

The samples were screened to detect mutations at codons 12/13 of the *KRAS* gene using the *Pyromark Q96 KRAS v2.0 kit* (*Qiagen, UK*). The samples were also screened for mutations in codon 600 of *BRAF* gene using the same kit with different primers. This kit and its protocol encompasses the PCR and the pyrosequencing.

DNA for the PCR reactions was extracted from FFPE tissue using the *Qiagen EZ1 robot* in conjunction with the *Tissue Extraction kit* (*Qiagen, UK*), as described in the MSI protocol in appendix two.
The assays for analysis of mutations in the *KRAS* and *BRAF* genes have three steps. The *BRAF* and *KRAS* assays were run in parallel.

1. Extraction of DNA from FFPE tissue
2. Amplification of the DNA region of interest using a PCR
3. Sequencing of the PCR amplicon using pyrosequencing

See Appendix three for protocols.

**Figure 8: KRAS exon 1 codon 12/13 primers and mutations**

<table>
<thead>
<tr>
<th>Primer Type</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Custom made Forward PCR primer</td>
<td>GCATTTATGCATTTTTCTTTAAGCGTGAGGGAGTTTGTAAAATGAAGTACAGTTCATT</td>
</tr>
<tr>
<td>Custom made Reverse (biotinylated) PCR primer</td>
<td>ACGA TACAGTCCTGCAGTCAACTGGAATTTTCTATGGAATTTTGTAAAGGTATTGGAGTAATGTTAATGGCTGGCAGGAAAG</td>
</tr>
<tr>
<td>Sequencing Primers from the kit</td>
<td>AGTGCCTTTGACAGATACAGCTAATTCAGAATCATTTTGTGGAGAATAT</td>
</tr>
</tbody>
</table>

**Codon 12/13 of the *KRAS* gene (position 34-39)**

**Common mutations seen in CRC**

1. **Codon 12**
   - G at position 34. G>A (GGT>AGT) Glycine to Serine
     - G>C (GGT>CGT) Glycine to Arginine
     - G>T (GGT>TGT) Glycine to Cystine
   - C at position 35. G>T (GGT>GTT) Glycine to Valine
     - G>A (GGT>GAT) Glycine to Aspartic acid
     - G>C (GGT>GCT) Glycine to Alanine

2. **Codon 13**
   - G at position 38. G>A (GGG>GAC) Glycine to Aspartic acid
**Figure 9: BRAF Exon 15. Primers and mutation**

5’
GGTTATAGGCTAAATAGAAACCTAATTTTGGACATACATATTGGACTCTAAGAGGAAAGATGAAAGTACTATGTTTTAAAAAGAATATTATATTACAGAATTATAGAAATTTAGAATCTCTACAAACCTTCTTGATTTGCTGATAGGAAAATGAGATCTACTGTTTTCTTACTTACTACACCTCAG
EXON 15
TAAATATTCTTCATGAAGACCTCACATATAAAATAGGATTTTTGGCTAGCTACAATGAAATCTCGATGGAGTGTTGATCCCATCTGTTTTGGAACAGTGTCTGGATCCATTTCG
TGGATG
GTAAGAATTGAGGCTATTTTCTCCACTGATTAAATTTTGGCCCTGAGATGCTGCTAGTTACTAGAAGAGCTTTGAAGTCTCAACTATAGTATTTTCATGTTGTCCAGTATTCAC
AAAATCTGATTTCTATTTTATGTAATATAGATTATTAACTTTTCTTTACCCCTAAACGAATATTGGAACCAGTTTCAGTGTATTTCACAAAATAATATATATATATAGTCTTATAAA
ACAGTGGTTTATATTTTATTTTTCAATAAAATATGAAACCCTTAAACGAATATTTTGGAAACCATTT

Custom made Forward PCR primer

Custom made Reverse (biotinylated) PCR primer

Custom made sequencing primer

Codon 600 of BRAF gene

A/T common allelic variations. Not known to alter function.

Common mutations seen in CRC (V600E)

T at position 1799 (in exon 15) T>A (GTG>GAG) Valine to glutamine

---

**2.2.5.2.1 Interpreting KRAS and BRAF Pyrograms (genotyping)**

The pyrosequencer and its software produce a diagramatic output. These pyrograms are charts which allow genotyping of the DNA sequence. The pyrograms demonstrate the percentage of each nucleotide at each particular position from all the DNA within that sample.

The peaks of the pyrograms are proportional to the amount of a nucleotide at that position. The absolute values are not read from the y axis, but the peak heights
compared to each other. Thus a small peak of A next to a large peak (e.g. four times that) of T will correspond to a DNA sequence ATTTT.

Wild-type codon 12/13 of KRAS is sequence: **GGT GGC** (forward). The common mutations occur at positions 34/35/38; which are highlighted here in blue:  

**GGT GGC** (see Figure 8).

The pyrogram below (Figure 10) can be genotyped as **GGT GGC GTA GG** so is wild-type KRAS codon 12/13. The three highlighted guanine nucleotides are positions 34/35/38 the common sites for mutations.

**Figure 10: Pyrogram for wild-type codon 12/13 of KRAS gene**
Wild-type *BRAF* codon 600 is **GTG** (forward). The V600E mutation is at position 1799 (highlighted in blue). (see figure 9).

The pyrogram seen in Figure 11 can be genotyped as **GTG AAA TC** so is wild-type for *BRAF* codon 600.

**Figure 11: Pyrogram of wild-type *BRAF* codon 600**
The pyrogram seen in Figure 12 is genotyped as **TGT GGC** for codons 12 and 13 of *KRAS* for 18.1% of the DNA in the sample, so is the *KRAS* variant 34G>T Gly12Cys.

**Figure 12: Pyrogram of KRAS 34 G>T Gly12Cys**

Peak indicating G>T at position 34 in 18.1% of DNA in the sample. The G peak is smaller than in wild-type and the adjacent T peak is larger.
The pyrogram seen in Figure 13 is genotyped as G\text{T} GGC for codons 12 and 13 of KRAS for 31.3% of the DNA in the sample, so is the KRAS variant 35G>T Gly12Val.

**Figure 13: Pyrogram of KRAS variant 35 G>T Gly12Val**

Peak indicating G>T at position 35 in 31.3% of DNA in the sample. The G peak is smaller than in wild-type and the adjacent T peak is larger.
The pyrogram seen in Figure 14 is genotyped as GAT GGC for codons 12 and 13 of KRAS for 39.8% of the DNA in the sample, so is the variant KRAS 35G>AGly12Asp.

Figure 14: Pyrogram of KRAS variant 35G>AGly12Asp

Each DNA sample was tested in triplicate from the primary PCR stage. When genotyping each sample, the three results were interpreted together. Each pyrogram gives a percentage of the particular point mutation (exchange of one base for another). For example the pyrogram in figure 14 represents a 39.8% point mutation of G>A of the DNA within that sample.

If a mutation was detected at a percentage value of >10% in at least two of the triplicates, that test was passed and genotyped to have the mutation present. If the mutation was found to be present at a low level (i.e. in <10% in at least two of the triplicates, the PCR and pyrosequencing was re-run). If the mutation was found to be at very low level (i.e. less than 10% in two of the triplicates and absent in the third triplicate), or the second run of PCR and pyrosequencing was again low level,
an ARMS assay was run which is able to positively detect lower level mutations. This is illustrated in Table 8.
Table 8: Example of KRAS triplicate results and their interpretation or further action required

<table>
<thead>
<tr>
<th>Lab no.</th>
<th>Run #1 (% of point mutation)</th>
<th>Action or result</th>
<th>Run #2 (% of point mutation)</th>
<th>Action or result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T2</td>
<td>T3</td>
<td></td>
</tr>
<tr>
<td>10003500</td>
<td>8.3</td>
<td>8.0</td>
<td>6.3</td>
<td>Repeat PCR &amp; pyrosequencing</td>
</tr>
<tr>
<td>10003507</td>
<td>18.1</td>
<td>8.2</td>
<td>15.2</td>
<td>c.34G&gt;T p.Gly12Cys</td>
</tr>
<tr>
<td>10003512</td>
<td>5.5</td>
<td>5.1</td>
<td>Normal</td>
<td>ARMS: Normal</td>
</tr>
</tbody>
</table>
The pyrogram seen in Figure 15 is genotyped as $\text{GAG AAA TC}$ for 22.4% of the DNA in the sample so is the variant V600E at $\text{BRAF}$ codon 600.

**Figure 15: Pyrogram of $\text{BRAF}$ variant V600E**

The $\text{BRAF}$ assay triplicate results were assessed, and repeated as necessary in the same fashion as the $\text{KRAS}$ assay triplicates. This is illustrated in table 9.
Table 9: Example of BRAF triplicate results and interpretation/further action required

<table>
<thead>
<tr>
<th>Lab no.</th>
<th>Run #1 (% of point mutation)</th>
<th>Action or result</th>
<th>Run #2 (% of point mutation)</th>
<th>Action or result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T2</td>
<td>T3</td>
<td></td>
</tr>
<tr>
<td>10003502</td>
<td>25.3</td>
<td>Fail</td>
<td>23.4</td>
<td>c.1799T&gt;A p.Val600Glu</td>
</tr>
<tr>
<td>10003535</td>
<td>9.2</td>
<td>8.7</td>
<td>10.1</td>
<td>Repeat PCR &amp; pyrosequencing</td>
</tr>
<tr>
<td>10003567</td>
<td>29.9</td>
<td>30.0</td>
<td>29.5</td>
<td>c.1799T&gt;A p.Val600Glu</td>
</tr>
</tbody>
</table>
2.1.5.3  

**MLH1 promoter region methylation analysis**

The *MLH1* promoter region lies at bases -711 to +15 relative to the gene transcription start site, which is located on the short arm of chromosome 3. This region contains a large number of CpG islands. Deng et al investigated the functional significance of hypermethylation by correlating the methylation status of all CpG sites with protein expression and function. They found that hypermethylation of a small proximal region -248 to -178 bases relative to the gene transcription start site correlated well with a lack of gene expression (Deng, Chen et al. 1999). This region is examined in the assay described below.

The assay for the quantification of *MLH1* promoter region methylation has three steps:

1. Extraction of DNA from FFPE tissue, bisulphite assisted conversion of unmethylated cytosines, and purification of converted DNA.

2. Amplification of the DNA region of interest using a polymerase chain reaction.

3. Quantification of methylation of the PCR amplicon using pyrosequencing.

Pyrosequencing has not previously been used to assess the methylation status of the *MLH1* promoter region of DNA derived from FFPE tissue. The optimisation process of this assay was extensive. The early experiments resulted in either complete failed pyrograms or in inconsistent and irreproducible results. The steps which were altered to reach the final assay are briefly described below.

Initially the *EZ DNA Methylation™ Kit* (Zymo Research) was used for DNA extraction and bisulphite treatment. The *Pyromark™ Q24 MLH1 kit* (Qiagen, UK) was used for the PCR and pyrosequencing steps. These two kits were used in combination on trial tissue exactly as the protocol for each kit described, however the results were unsatisfactory. The first alteration made was to the PCR primers. It was thought that the primers provided by the kit may be too large and distant from the area of DNA of interest. New PCR primers were designed which were smaller
and closer to the region of interest (see Figure 16). This alteration did not result in consistent pyrograms results. The PCR protocol was then altered. Different annealing temperatures were tested. The concentration of products from the PCR was not satisfactory. A nested PCR (PCR with two steps) was tested and the pyrograms began to be more reproducible but still were not 100% consistent. The next alteration made was the use of a different DNA extraction and bisulphite modification kit. EpiTect plus FFPE Bisulphite kit® (Qiagen, UK) was trialled. This kit is especially designed for use with FFPE tissue. Methylation analysis is especially problematic with DNA derived from FFPE tissue. The formalin fixation used to preserve the tissue can cause excessive DNA crosslinking. If this crosslinking affects the region of DNA of interest, it prohibits PCR amplification and hence sequencing. Two further steps were altered as part of the DNA extraction and bisulphite treatment protocol. An overnight incubation of the FFPE tissue with proteinase K was used because the hour incubation described in the protocol clearly left visible unlysed tissue. An additional step of incubation at 90 °C for 60 minutes on a shaking heated block was also added to further facilitate separation of the formalin induced DNA crosslinks. Initially the pyrosequencing primer was used from the Pyromark™ Q24 MLH1 kit (Qiagen, UK). A custom sequencing primer was designed during the optimisation process that was felt to be more robust.

Following alteration of all these steps, consistent, reproducible pyrograms were being produced from the trial tissue. As part of this study the assay was tested and validated to Good Clinical Laboratory standards.

The PCR reaction requires forward and reverse primers in order for the DNA polymerase enzyme to synthesis the appropriate forward and reverse portions of new DNA. These primers, as described in section 2.1.7.2, need to be nuceotide sequences that are located 3’ and 5’ to the region of DNA to be synthesised. If the primers are too long, they may form bonds with themselves rather than the target DNA, and if they are too short, they may not be specific enough to stimulate synthesis of the correct portion of DNA. The pyrosequencing assay also requires a sequencing primer to stimulate the process of sequencing by synthesis.
A critical step for the determination of the methylation status of DNA is the complete conversion of unmethylated cytosines. This is achieved by incubating the DNA in high bisulphite salt concentrations at high temperature and low pH. This results in the conversion of unmethylated cytosines to thiamine leaving methylated cytosines unchanged. Therefore this step results in different DNA sequences for methylated and unmethylated DNA which can then be interpreted when sequencing.
the DNA. The target DNA region also contains a cytosine base which acts as a control for complete bisulphite conversion. This cytosine is not within a CpG island so will always be unmethylated. This cytosine should always be converted to thiamine by incubation with bisulphite.

The **Pyromark™ Q24 MLH1 kit (Qiagen, UK)** was used for both the PCR and pyrosequencing steps of the assay. Custom designed primers were used for the PCR and sequencing.

See Appendix four for the assay protocol.

### 2.2.5.3.1 Interpreting *MLH1* promoter region methylation pyrograms

The pyrosequencer and its software produces a diagramatic output. These pyrograms are charts which allow genotyping of a particular DNA sequence. The charts demonstrate the percentage of each nucleotide at each particular position from all the DNA within that sample. In the context of methylation analysis this allows quantification of methylation.

The peaks of the pyrograms are proportional to the amount of a nucleotide at that position. The absolute values are not read from the y axis, but the peak heights compared to each other. Thus a small peak of A next to a large peak (e.g. 4 times that) of T will correspond to a DNA sequence ATTTT.

The pyrogram below (Figure 17) for bisulphite treated DNA can be genotyped as:

![Pyrogram](image)

The ⫷ represents unmethylated cytosines (from CpG islands) that were converted to thiamines during bisulphite conversion. The ⫸ is the control for complete bisulphite modification. This was a cytosine that was not within a CpG islands therefore will never be methylated.
The genotype of the tumour DNA prior to bisulphite treatment would therefore be:

CGGACAGCGATTITTAACGCG unmethylated.

This pyrogram of the \textit{MLH1} promoter region is interpreted as:

a) appropriate bisulphite conversion as the control cytosine (C) has been converted to Thiamine (T)

b) all cytosines within the area of interest are unmethylated

\textbf{Figure 17: Pyrogram of unmethylated \textit{MLH1} promoter region of tumour DNA}
The pyrogram below (Figure 18) for bisulpite treated DNA is genotyped as:

CGGATACGATTTTTAACCGCG

The C represents methylated cytosines (from CpG islands) that were resistant to bisulphite conversion to thiamine. The T is the control for complete bisulphite modification.

The genotype of this tumour DNA prior to bisulphite treatment would therefore be:

CmGGACAGCmGATTTTTAACmGCmG (methylated CpG islands).

This pyrogram of the MLH1 promoter region can be interpreted as:

a) appropriate bisulphite conversion as the control C has been converted to T

b) all cytosines within the area of interest are methylated

- 91.4% methylation of cytosines at position 3
- 91.5% methylation of cytosines at position 13
- 94.5% methylation of cytosines at position 21
- 85.6% methylation of cytosines at position 25
2.2.6 Data analysis

The mismatch repair status of the local sporadic tumours was examined and the tumours that were mismatch deficient due to a loss of MLH1 protein were identified.

The performance characteristics of wild-type BRAF V600E and non-methylated MLH1 promoter region for the identification of cancers from the individuals with an MLH1 mutation from the sporadic MLH1 loss cancers were analysed using Diagnostic test two by two tables using Stats direct (StatsDirect Ltd, Cheshire UK) software. Each test was analysed separately and also in conjunction by applying the ‘either positive’ and ‘both positive’ rules. In this setting the either positive rule indicates that EITHER wild-type BRAF OR normal (unmethylated) MLH1 promoter region is a positive result (and BRAF V600E AND MLH1 methylation must be present for a negative result) when predicting an MLH1 mutation. The both positive rule
states that BOTH wild-type \textit{BRAF} AND normal \textit{MLH1} promoter region are a positive result (and either \textit{BRAF} V600E OR \textit{MLH1} methylation is a negative result).

The sensitivity and specificity of these pre-screening tests were applied to different \textit{a priori} risk groups using Bayes theorem. These different \textit{a priori} risk groups were patients who fulfil Amsterdam criteria, patients who fulfil Bethesda criteria and patients from the general population regardless of family history.

Bayes theorem describes the relationship that exists between simple and conditional probabilities. It allows the interpretation of a diagnostic test’s sensitivity and specificity by incorporating knowledge of a given individual’s pre-test probability of having the disease that the test is being used to diagnose. If an individual has a very low pre-test probability of having that disease, then according to Bayes theorem, a positive result of the test (that has a sensitivity of less than 100\%) will result in a lower post-test probability compared to the same positive result in an individual with a high pre-test probability.

This is important in interpreting the results of study one. \textit{BRAF} V600E mutation and \textit{MLH1} promoter region methylation analysis could be used to pre-screen for germline \textit{MLH1} mutations in individuals with differing pre-test risk (\textit{a priori} risk). An individual who fulfils Amsterdam II criteria will have a higher prior risk than an individual who fulfils Bethesda criteria, or an individual who has had a mismatch repair deficient CRC but who has no relevant family history. Interpreting the results of these assays must be done with knowledge of the individual’s prior risk.

In order to apply Bayes theorem to the results of study 1, pre-test probabilities of having an \textit{MLH1} mutation were identified from the studies discussed in section 1.10.15. Patients who fulfil Amsterdam II criteria have a pre-test probability of at least 60\% or 0.6 (Park, Vasen et al. 2002). Patients who fulfil the revised Bethesda criteria and have loss of \textit{MLH1} have a pre-test probability of at least 10.5\% or 0.105 (Julie, Tresallet et al. 2008; Perez-Carbonell, Ruiz-Ponte et al. 2011; Wright, Arnold et al. 2011). Patients from the general population (without a known familial mutation) who have \textit{MLH1} loss (MMR deficient [dMMR] tumour) have a pre-test probability of at least 4.0\% or 0.04 (Aaltonen, Salovaara et al. 1998; Hampel,
Frankel et al. 2005; Pinol, Castells et al. 2005; Perez-Carbonell, Ruiz-Ponte et al. 2011).

Bayes theorem was used to calculate post-test probability of a true positive result and a false negative result. In order for the pre-screening strategy to be effective in a given population, the post-test probability of a true positive must be greater than 10%, and the post-test probability of a false negative must be less than 10%, as 10% risk is the threshold above which guidelines state that an individual should be tested for a germline mutation which causes cancer susceptibility (ASCO 1996).
2.3 Results Study one: DNA mismatch gene promoter region methylation analysis and BRAF gene mutation analysis- an alternative prescreening strategy in Lynch Syndrome

See Appendices five, six and seven.

In total, tumour sections from 71 carriers of pathogenic MLH1 mutations, 73 tumours demonstrating sporadic mismatch repair deficiency, and 52 sporadic mismatch repair proficient tumours were identified.

Of these, tumour sections from 22 carriers of pathogenic MLH1 mutations, 33 tumours demonstrating sporadic mismatch repair deficiency and 52 sporadic mismatch repair proficient tumours were obtained locally. 49 tumours from carriers of pathogenic MLH1 mutations and 40 tumours with sporadic mismatch repair deficiency (of which 34 were known to be wild-type for MLH1) were obtained from the Australian branch of the Colon Cancer Family Registry (ACCFR) based at the Familial Cancer Laboratory, Queensland Institute of Medical Research (QIMR).
2.3.1 Mismatch repair analysis of local sporadic tumours

86 right-sided colorectal cancers were identified from patients aged over 50 years with no family history fulfilling Amsterdam or revised Bethesda criteria. 34 were found to be mismatch repair deficient. 33 of these were MSI-High with loss of MLH1 protein on immunohistochemistry. 1 case which was MSI-High was also found to have loss of MSH2 on immunohistochemistry. This suggests likely Lynch Syndrome. It is outside of the bounds of the ethics approval of this study, to relay this information to the patient. However this patient is undergoing clinic follow-up for his CRC so review of his case (with possible referral to Genetic Medicine if felt appropriate by his clinician) has been suggested.

Table 10: Results of local sporadic tumour mismatch repair deficiency testing

<table>
<thead>
<tr>
<th>Total</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sporadic mismatch repair deficient CRC due to loss of MLH1</strong></td>
<td>33</td>
</tr>
<tr>
<td>32 Loss of MLH1 and MSI-High.</td>
<td></td>
</tr>
<tr>
<td>1 All proteins present, MSI-High, found to have BRAF V600 and methylated MLH1.</td>
<td></td>
</tr>
<tr>
<td><strong>Sporadic mismatch repair proficient</strong></td>
<td>52</td>
</tr>
<tr>
<td>48 All proteins present and MSI-stable</td>
<td></td>
</tr>
<tr>
<td>1 All proteins present and MSI-low</td>
<td></td>
</tr>
<tr>
<td>1 All proteins present and MSI-High</td>
<td></td>
</tr>
<tr>
<td>2 Loss of MLH-1 and MSI-stable</td>
<td></td>
</tr>
<tr>
<td><strong>Possible MSH2 mutation</strong></td>
<td>1</td>
</tr>
<tr>
<td>1 Loss of MSH2 and MSI-High</td>
<td></td>
</tr>
</tbody>
</table>
### 2.3.2 Results of *BRAF* mutation and *MLH1* promoter region methylation in MLH1 mutation carriers and sporadic MLH1 loss cancers

Table 11: *BRAF* mutation and *MLH1* promoter region methylation in MLH1 mutation carriers and sporadic MLH1 loss cancers

<table>
<thead>
<tr>
<th>Tumour Status</th>
<th><em>BRAF</em> V600E variant AND methylated <em>MLH1</em> promoter region</th>
<th>Wild-type <em>BRAF</em> negative AND normal <em>MLH1</em> promoter region</th>
<th><em>BRAF</em> V600E variant AND normal <em>MLH1</em> promoter region</th>
<th>Wild-type <em>BRAF</em> AND methylated <em>MLH1</em> promoter region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total MLH1 mutation carrier n=71 (percentage in parentheses is the percentage of total MLH1 mutation carriers)</td>
<td>0</td>
<td>66 (92.9%)</td>
<td>1 (1.4%)</td>
<td>4 (5.6%)</td>
</tr>
<tr>
<td>Sporadic mutation untested <em>dMMR</em> tumours n=40 (percentage in parentheses is the percentage of total untested dMMR tumours)</td>
<td>30 (75.0%)</td>
<td>4 (10%)</td>
<td>0</td>
<td>6 (15.0%)</td>
</tr>
<tr>
<td>Sporadic mutation negative <em>dMMR</em> tumours n=33 (percentage in parentheses is the percentage of total sporadic mutation negative dMMR tumours)</td>
<td>16 (48.5%)</td>
<td>3 (9.1%)</td>
<td>2 (6.0%)</td>
<td>12 (36.4%)</td>
</tr>
<tr>
<td>Total <em>dMMR</em> sporadic tumours n=73 (percentage in parentheses is the percentage of total dMMR sporadic tumours)</td>
<td>46 (63.0%)</td>
<td>7 (9.6%)</td>
<td>2 (2.7%)</td>
<td>18 (24.7%)</td>
</tr>
</tbody>
</table>

*dMMR= mismatch repair deficient*
Table 12: Diagnostic Test (2 by 2 tables) analysis for the identification of MLH1 mutations carriers

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wild-type BRAF</strong></td>
<td>98.59% (92.4-99.96%)</td>
<td>65.75% (53.72-76.47%)</td>
</tr>
<tr>
<td><strong>Normal MLH1 promoter region</strong></td>
<td>94.37% (86.2-98.44%)</td>
<td>87.67% (77.88-94.2%)</td>
</tr>
<tr>
<td><strong>Wild-type BRAF OR Normal MLH1 promoter region</strong></td>
<td>100% (94.94% to 100%)</td>
<td>63.01% (50.91% to 74.03%)</td>
</tr>
<tr>
<td><strong>Wild-type BRAF AND Normal MLH1 promoter region</strong></td>
<td>92.96% (84.33% to 97.67%)</td>
<td>90.41% (81.24% to 96.06%)</td>
</tr>
</tbody>
</table>

This analysis has demonstrated that wild-type BRAF has a similar sensitivity (98.59% [92.40-99.96%]) to normal (unmethylated) MLH1 promoter region (94.37% [95% CI 86.20-98.44%]) for the identification of MLH1 mutation carriers, but normal MLH1 promoter region has a significantly higher specificity (87.67% [95% CI 77.88-94.2%]) than wild type BRAF 65.75% [95% CI 53.72-76.47%]).

When used in combination, the result of wild-type BRAF OR normal MLH1 promoter region has the highest sensitivity for the identification of MLH1 mutation carriers (100% [95% CI 94.94% to 100%]), but has the lowest specificity (63.01% [50.91% to 74.03%]). The result of wild-type BRAF AND normal MLH1 promoter region has the lowest sensitivity (92.96% [95% CI 84.33% to 97.67%]) but the highest specificity (90.41% [81.24% to 96.06%]).
2.3.2.1 **BRAF** mutation analysis in **MLH1** mutation carriers and sporadic MLH1 loss cancers

### Table 13: **BRAF** V600E mutation Diagnostic test analysis

<table>
<thead>
<tr>
<th></th>
<th><strong>MLH1</strong> mutation carriers</th>
<th><em>Sporadic dMMR</em></th>
<th><strong>Total</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wt BRAF V600E</strong></td>
<td>70</td>
<td>25</td>
<td>95</td>
</tr>
<tr>
<td><strong>Mutated BRAF V600E</strong></td>
<td>1</td>
<td>48</td>
<td>49</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>71</td>
<td>73</td>
<td>144</td>
</tr>
</tbody>
</table>

*sporadic dMMR = sporadic mismatch repair deficient

Including 95% confidence intervals:

**Prevalence** (pre-test likelihood of being **MLH1** mutation carrier in test population)
0.49 (0.41 to 0.58), 49.31% (40.88% to 57.76%)

**Predictive value of positive test** (post-test likelihood of being **MLH1** mutation carrier)
0.74 (0.63 to 0.82), 73.68% (63.65% to 82.19%), {change = 25%}

**Predictive values of negative test**
(post-test likelihood of sporadic dMMR)
0.98 (0.89 to 1.0), 97.96% (89.15% to 99.95%), {change = 47%}
(post-test being **MLH1** mutation carrier likelihood despite negative test)
0.02 (0.00 to 0.11), 2.04% (0.05% to 10.85%), {change = -47%}

**Sensitivity** (true positive rate)
0.99 (0.92 to 1.0),
98.59% (92.4% to 99.96%)

**Specificity** (true negative rate)
0.66 (0.54 to 0.76),
65.75% (53.72% to 76.47%)

**Likelihood Ratio**
LR (positive test) = 2.88 (2.15 to 4.05)
LR (negative test) = 0.02 (0.00 to 0.12)

Diagnostic Odds Ratio
Odds ratio = 134.4 (20.02 to 5491.57)

2.3.2.2 MLH1 promoter region methylation analysis in MLH1 mutation carriers and sporadic MLH1 loss cancers

Table 14: MLH1 promoter region methylation Diagnostic test analysis

<table>
<thead>
<tr>
<th></th>
<th>MLH1 mutation carriers</th>
<th>Sporadic dMMR</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal MLH1 prom</td>
<td>67</td>
<td>9</td>
<td>76</td>
</tr>
<tr>
<td>Methylated MLH1 prom</td>
<td>4</td>
<td>64</td>
<td>68</td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>73</td>
<td>144</td>
</tr>
</tbody>
</table>

Including 95% confidence intervals:

Prevalence (pre-test likelihood of being MLH1 mutation carrier in test population)
0.493056 (0.408823 to 0.577579), 49.31% (40.88% to 57.76%)

Predictive value of positive test (post-test likelihood of being MLH1 mutation carrier in test population)
0.88 (0.79 to 0.94), 88.16% (78.71% to 94.44%), {change = 39%}

Predictive values of negative test
(post-test likelihood of sporadic dMMR)
0.94 (0.86 to 0.98), 94.12% (85.62% to 98.37%),
(post-test MLH1 mutation carrier likelihood despite negative test)
0.06 (0.02 to 0.14), 5.88% (1.63% to 14.38%),

Sensitivity (true positive rate)
0.94 (0.86 to 0.98),
94.37% (86.2% to 98.44%)

Specificity (true negative rate)
0.88 (0.78 to 0.94),
Likelihood Ratio
LR (positive test) = 7.65 (4.31 to 14.28)
LR (negative test) = 0.06 (0.03 to 0.16)

Diagnostic Odds Ratio
Odds ratio = 119.11 (31.63 to 523.0)

2.3.2.3 BRAF mutation analysis OR MLH1 promoter region methylation in MLH1 mutation carriers and sporadic MLH1 loss cancers

Table 15: BRAF V600E mutation and MLH1 promoter region methylation Diagnostic test analysis where EITHER test being normal is a positive result.

<table>
<thead>
<tr>
<th></th>
<th>MLH1 mutation carriers</th>
<th>Sporadic dMMR</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wt BRAF OR Normal MLH1 promoter region</td>
<td>71</td>
<td>27</td>
<td>98</td>
</tr>
<tr>
<td>Mutated BRAF V600E AND methylated MLH1 promoter region</td>
<td>0</td>
<td>46</td>
<td>46</td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>73</td>
<td>144</td>
</tr>
</tbody>
</table>

Including 95% confidence intervals:

Prevalence (pre-test likelihood of disease)
0.49 (0.41 to 0.58), 49.31% (40.88% to 57.76%)

Predictive value of positive test (post-test likelihood of disease)
0.72 (0.63 to 0.81), 72.45% (62.5% to 80.99%), {change = 23%}

Predictive values of negative test
(post-test likelihood of no disease)
1 (0.92 to 1 [97.5% one-sided CI]), 100% (92.29% to 100%), {change = 49%}
(post-test disease likelihood despite negative test)
0 (0 to 0.08 [97.5% one-sided CI]), 0% (0% to 7.71%), {change = -49%}

Sensitivity (true positive rate)
1 (0.94 to 1 [97.5% one-sided CI]),
100% (94.94% to 100%)

87.67% (77.88% to 94.2%)
Specificity (true negative rate)
0.63 (0.51 to 0.74),
63.01% (50.91% to 74.03%)

Likelihood Ratio
LR (positive test) = 2.70 (2.05 to 3.70)
LR (negative test) = * (0 to 0.08)

Diagnostic Odds Ratio
Odds ratio = * (28.05 to infinity)

2.3.2.4 BRAF mutation analysis AND MLH1 promoter region methylation

Table 16: BRAF V600E variant and MLH1 promoter region methylation Diagnostic test analysis where BOTH test being normal is a positive result.

<table>
<thead>
<tr>
<th></th>
<th>MLH1 mutation carriers</th>
<th>Sporadic dMMR</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wt BRAF AND Normal MLH1 promoter region</td>
<td>66</td>
<td>7</td>
<td>73</td>
</tr>
<tr>
<td>Mutated BRAF V600E OR methylated MLH1 promoter region</td>
<td>5</td>
<td>66</td>
<td>71</td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>73</td>
<td>144</td>
</tr>
</tbody>
</table>

Including 95% confidence intervals:

Prevalence (pre-test likelihood of disease)
0.49 (0.41 to 0.587), 49.31% (40.88% to 57.76%)

Predictive value of positive test (post-test likelihood of disease)
0.90 (0.81 to 0.96), 90.41% (81.24% to 96.06%), {change = 41%}

Predictive values of negative test
(post-test likelihood of no disease)
0.93 (0.84 to 0.98), 92.96% (84.33% to 97.67%), {change = 42%}
(post-test disease likelihood despite negative test)
0.07 (0.02 to 0.16), 7.04% (2.33% to 15.67%), {change = -42%}

Sensitivity (true positive rate)
Specificity (true negative rate)
0.93 (0.84 to 0.98),
92.96% (84.33% to 97.67%)

Likelihood Ratio
LR (positive test) = 9.69 (5.01 to 19.73)
LR (negative test) = 0.08 (0.03 to 0.17)

Diagnostic Odds Ratio
Odds ratio = 124.46 (33.63 to 500.9)
2.3.2.5 Bayes calculations

Table 17: Variables required for calculating post-test probabilities using Bayes theorem

| P(A)     | the probability that an MLH1 mutation will be present in any particular person |
| P(~A)    | the probability that an MLH1 mutation will not be present in any particular person |
| P(B|A)    | the probability that the test will yield a positive result [B] if an MLH1 mutation is present [A] |
| P(~B|A)   | the probability that the test will yield a negative result [~B] if an MLH1 mutation is present [A] |
| P(B|~A)   | the probability that the test will yield a positive result [B] if an MLH1 mutation is not present [~A] |
| P(~B|~A)  | the probability that the test will yield a negative result [~B] if an MLH1 mutation is not present [~A] |

Table 18: Post-test probabilities of positive and negative results irrespective of whether a mutation is present

| P(B)     | the probability of a positive test result [B], irrespective of whether an MLH1 mutation is present [A] or not present [~A] |
| P(~B)    | the probability of a negative test result [~B], irrespective of whether an MLH1 mutation is present [A] or not present [~A] |

Table 19: Post-test probabilities of true positive and false negative results.

| P(A|B) = [P(B|A) x P(A)] / P(B) | the probability that an MLH1 mutation is present [A] if the test result is positive [B] (i.e., the probability that a positive test result will be a true positive) |
| P(A|~B) = [P(~B|A) x P(A)] / P(~B) | the probability that an MLH1 mutation is present [A] if the test result is negative [~B] (i.e., the probability that a negative test result will be a false negative) |
Table 20: Pre-test probabilities (a priori risk)

<table>
<thead>
<tr>
<th></th>
<th>Amsterdam criteria</th>
<th>Revised Bethesda criteria + loss of MLH1</th>
<th>General population with loss of MLH1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(A)$ =</td>
<td>0.6</td>
<td>0.105</td>
<td>0.073</td>
</tr>
<tr>
<td>$P(\sim A)$ =</td>
<td>1-0.6=0.4</td>
<td>1-0.105=0.895</td>
<td>1-0.073=0.927</td>
</tr>
</tbody>
</table>

Table 21: Variables required for Bayes calculations

(The lowest bound of the 95% confidence intervals of both sensitivity and specificity were used in the Bayes calculations.)

<table>
<thead>
<tr>
<th></th>
<th>Wt $BRAF$</th>
<th>Normal $MLH1$ promoter</th>
<th>Wt $BRAF$ OR Normal $MLH1$ promoter</th>
<th>Wt $BRAF$ AND Normal $MLH1$ promoter</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_B(A)$ =</td>
<td>0.924</td>
<td>0.862</td>
<td>0.949</td>
<td>0.843</td>
</tr>
<tr>
<td>$P(\sim B</td>
<td>A)$ =</td>
<td>1-0.924</td>
<td>1-0.862</td>
<td>1-0.949=0.051</td>
</tr>
<tr>
<td>$P_B(\sim A)$ =</td>
<td>1-0.537</td>
<td>1-0.779</td>
<td>1-0.509=0.491</td>
<td>1-0.812=0.188</td>
</tr>
<tr>
<td>$P(\sim B</td>
<td>\sim A)$ =</td>
<td>0.537</td>
<td>0.779</td>
<td>0.509</td>
</tr>
</tbody>
</table>
2.3.2.6 Applying *BRAF* testing and *MLH1* promoter region methylation testing to patients who fulfil Amsterdam Criteria

### 2.3.2.6.1 Wild-type *BRAF* as a predictor of germline *MLH1* mutation

\[
P(B) = [P(B|A) \times P(A)] + [P(B|\sim A) \times P(\sim A)] \quad \text{(probability of wt *BRAF* irrespective of whether a *MLH1* mutation is present or not)}
\]

\[
= [0.924 \times 0.6] + [0.463 \times 0.4] = 0.5544 + 0.1852 = 0.740
\]

\[
P(\sim B) = [P(\sim B|A) \times P(A)] + [P(\sim B|\sim A) \times P(\sim A)] \quad \text{(probability of mutant *BRAF* irrespective of whether a *MLH1* mutation is present or not)}
\]

\[
= [0.076 \times 0.6] + [0.537 \times 0.4] = 0.0456 + 0.2148 = 0.260
\]

Post-test likelihood of true positive (i.e. *MLH1* mutation being present if *BRAF* is wt)

\[
P(A|B) = \frac{[P(B|A) \times P(A)]}{P(B)} = \frac{[0.924 \times 0.6]}{0.74} = 0.5544/0.74 = 0.749 = 74.9\%
\]

Post-test likelihood of false negative (i.e. *MLH1* mutation being present despite mutant *BRAF*)

\[
P(A|\sim B) = \frac{[P(\sim B|A) \times P(A)]}{P(\sim B)} = \frac{[0.076 \times 0.6]}{0.260} = 0.0456/0.260 = 0.175 = 17.5\%
\]
3.2.6.2 Normal $MLH1$ promoter region as a predictor of germline $MLH1$ mutation

$P_B = [P_{(B|A)} \times P_A] + [P_{(B|\sim A)} \times P_{\sim A}]$ (probability of normal $MLH1$ promoter region irrespective of whether a $MLH1$ mutation is present or not)

$= [0.862 \times 0.6] + [0.221 \times 0.4]$

$= 0.517 + 0.0884$

$= 0.605$

$P_{\sim B} = [P_{(\sim B|A)} \times P_A] + [P_{(\sim B|\sim A)} \times P_{\sim A}]$ (probability of methylated $MLH1$ promoter region irrespective of whether a $MLH1$ mutation is present or not)

$= [0.138 \times 0.6] + [0.779 \times 0.4]$

$= 0.0828 + 0.3116$

$= 0.394$

Post-test likelihood of true positive (i.e. $MLH1$ mutation being present if $MLH1$ promoter region is normal)

$P_{(A|B)} = \frac{P_{(B|A)} \times P_A}{P_B}$

$= \frac{[0.862 \times 0.6]}{0.605}$

$= 0.517/0.605$

$= 0.855 = 85.5\%$

Post-test likelihood of false negative (i.e. $MLH1$ mutation being present despite methylated $MLH1$ promoter region)

$P_{(A|\sim B)} = \frac{P_{(\sim B|A)} \times P_A}{P_{\sim B}}$

$= \frac{[0.0828 \times 0.6]}{0.394}$

$= 0.050/0.395$

$= 0.126 = 12.6\%$
2.3.2.6.3 Wild-type BRAF OR normal MLH1 promoter region as a predictor of germline MLH1 mutation

\[ P(B) = [P(B|A) \times P(A)] + [P(B|\sim A) \times P(\sim A)] \] (probability of wt BRAF OR normal MLH1 promoter region irrespective of whether a MLH1 mutation is present or not)

\[ = [0.949\times0.6] + [0.491\times0.4] \]
\[ = 0.594 + 0.1964 \]
\[ = 0.7958 \]

\[ P(\sim B) = [P(\sim B|A) \times P(A)] + [P(\sim B|\sim A) \times P(\sim A)] \] (probability of mutant BRAF or methylated MLH1 promoter region irrespective of whether a MLH1 mutation is present or not)

\[ = [0.051\times0.6] + [0.509\times0.4] \]
\[ = 0.0306 + 0.204 \]
\[ = 0.2346 \]

Post-test likelihood of true positive (i.e. MLH1 mutation being present if BRAF is wt or MLH1 promoter region is normal)

\[ P(A|B) = [P(B|A) \times P(A)] / P(B) \]
\[ = [0.949\times0.6] / 0.7958 \]
\[ = 0.5694/0.7958 \]
\[ = 0.7155 = 71.6\% \]

Post-test likelihood of false negative (i.e. MLH1 mutation being present despite mutant BRAF AND methylated MLH1 promoter region)

\[ P(A|\sim B) = [P(\sim B|A) \times P(A)] / P(\sim B) \]
\[ = [0.051\times0.6] / 0.234 \]
\[ = 0.030 / 0.234 \]
\[ = 0.1307 = 13.10\% \]

146
2.3.2.6.4 Wild-type *BRAF* AND normal *MLH1* promoter region as a predictor of germline *MLH1* mutation

\[ P(B) = [P(B|A) \times P(A)] + [P(B|\sim A) \times P(\sim A)] \] (probability of wild-type *BRAF* AND normal *MLH1* promoter region irrespective of whether a *MLH1* mutation is present or not)

\[ = [0.84 \times 0.6] + [0.19 \times 0.4] \]
\[ = 0.504 + 0.076 \]
\[ = 0.58 \]

\[ P(\sim B) = [P(\sim B|A) \times P(A)] + [P(\sim B|\sim A) \times P(\sim A)] \] (probability of mutant *BRAF* AND methylated *MLH1* promoter region irrespective of whether a *MLH1* mutation is present or not)

\[ = [0.16 \times 0.6] + [0.81 \times 0.4] \]
\[ = 0.096 + 0.324 \]
\[ = 0.42 \]

Post-test likelihood of true positive (i.e. *MLH1* mutation being present if *BRAF* is wild-type AND *MLH1* promoter region is normal)

\[ P(A|B) = \frac{[P(B|A) \times P(A)]}{P(B)} \]
\[ = \frac{[0.84 \times 0.6]}{0.58} \]
\[ = 0.504/0.58 \]
\[ = 0.868 = 86.9\% \]

Post-test likelihood of false negative (i.e. *MLH1* mutation being present despite mutant *BRAF* OR methylated *MLH1* promoter region)

\[ P(A|\sim B) = \frac{[P(\sim B|A) \times P(A)]}{P(\sim B)} \]
\[ = \frac{[0.16 \times 0.6]}{0.42} \]
\[ = 0.096/0.42 \]
\[ = 0.229 = 22.9\% \]
2.3.2.7 Applying BRAF testing and MLH1 promoter region methylation testing to patients who fulfil revised Bethesda criteria and have an MLH1 loss tumour

2.3.2.7.1 Wild-type BRAF as a predictor of germline MLH1 mutation

\[ P(B) = [P(B|A) \times P(A)] + [P(B|\sim A) \times P(\sim A)] \]

\[ = [0.924 \times 0.105] + [0.463 \times 0.895] \]

\[ = 0.097 + 0.414 \]

\[ = 0.511 \]

\[ P(\sim B) = [P(\sim B|A) \times P(A)] + [P(\sim B|\sim A) \times P(\sim A)] \]

\[ = [0.076 \times 0.105] + [0.537 \times 0.895] \]

\[ = 0.00798 + 0.480 \]

\[ = 0.489 \]

Post-test likelihood of true positive (i.e. MLH1 mutation being present if BRAF is wt)

\[ P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \]

\[ = \frac{0.924 \times 0.105}{0.511} \]

\[ = 0.190 = 19.0\% \]

Post-test likelihood of false negative (i.e. MLH1 mutation being present despite mutant BRAF)

\[ P(A|\sim B) = \frac{P(\sim B|A) \times P(A)}{P(\sim B)} \]

\[ = \frac{0.076 \times 0.105}{0.489} \]

\[ = 0.0163 = 1.6\% \]
2.3.2.7.2 Normal MLH1 promoter region as a predictor of germline MLH1 mutation

\[ P(\text{B}) = [P(\text{B|A}) \times P(A)] + [P(\text{B|~A}) \times P(\text{~A})] \] (probability of normal MLH1 promoter region irrespective of whether a MLH1 mutation is present or not)

\[ = [0.862 \times 0.105] + [0.221 \times 0.895] \]
\[ = 0.0905 + 0.1978 \]
\[ = 0.288 \]

\[ P(\text{~B}) = [P(\text{~B|A}) \times P(A)] + [P(\text{~B|~A}) \times P(\text{~A})] \] (probability of methylated MLH1 promoter region irrespective of whether a MLH1 mutation is present or not)

\[ = [0.138 \times 0.105] + [0.779 \times 0.895] \]
\[ = 0.0145 + 0.697 \]
\[ = 0.7117 \]

Post-test likelihood of true positive (i.e. MLH1 mutation being present if MLH1 promoter region is normal)

\[ P(\text{A|B}) = [P(\text{B|A}) \times P(A)] / P(\text{B}) \]
\[ = [0.862 \times 0.105] / 0.288 \]
\[ = 0.0905 / 0.288 \]
\[ = 0.314 = 31.4\% \]

Post-test likelihood of false negative (i.e. MLH1 mutation being present despite methylated MLH1 promoter region)

\[ P(\text{A|~B}) = [P(\text{~B|A}) \times P(A)] / P(\text{~B}) \]
\[ = [0.0828 \times 0.105] / 0.7117 \]
\[ = 0.0087 / 0.7117 \]
\[ = 0.0122 = 1.2\% \]
2.3.2.7.3 Wild-type \textit{BRAF} OR normal \textit{MLH1} promoter region as a predictor of germline \textit{MLH1} mutation

\[
P(B) = [P(B|A) \times P(A)] + [P(B|\sim A) \times P(\sim A)] \quad \text{(probability of wt \textit{BRAF} OR normal \textit{MLH1} promoter region irrespective of whether a \textit{MLH1} mutation is present or not)}
\]

\[
= [0.949 \times 0.105] + [0.491 \times 0.895]
\]

\[
= 0.0997 + 0.439
\]

\[
= 0.539
\]

\[
P(\sim B) = [P(\sim B|A) \times P(A)] + [P(\sim B|\sim A) \times P(\sim A)] \quad \text{(probability of mutant \textit{BRAF} or methylated \textit{MLH1} promoter region irrespective of whether a \textit{MLH1} mutation is present or not)}
\]

\[
= [0.051 \times 0.105] + [0.509 \times 0.895]
\]

\[
= 0.0054 + 0.456
\]

\[
= 0.41
\]

Post-test likelihood of true positive (i.e. \textit{MLH1} mutation being present if \textit{BRAF} is wt OR \textit{MLH1} promoter region is normal)

\[
P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}
\]

\[
= \frac{0.949 \times 0.105}{0.539}
\]

\[
= 0.099/0.539
\]

\[
= 0.185 = \text{18.5%}
\]

Post-test likelihood of false negative (i.e. \textit{MLH1} mutation being present despite mutant \textit{BRAF} OR methylated \textit{MLH1} promoter region)

\[
P(A|\sim B) = \frac{P(\sim B|A) \times P(A)}{P(\sim B)}
\]

\[
= \frac{0.051 \times 0.105}{0.41}
\]

\[
= 0.00534/0.41
\]

\[
= 0.0131 = \text{1.3%}
\]
2.3.2.7.4 Wild-type **BRAF** AND normal **MLH1** promoter region as a predictor of germline **MLH1** mutation

\[
P(_B) = [P(_B|A) \times P(A)] + [P(_B|~A) \times P(~A)] \quad \text{(probability of wild-type **BRAF** AND normal **MLH1** promoter region irrespective of whether a **MLH1** mutation is present or not)}
\]

\[
= [0.84 \times 0.105] + [0.19 \times 0.895]
\]

\[
= 0.0882 + 0.175
\]

\[
= 0.263
\]

\[
P(~_B) = [P(~_B|A) \times P(A)] + [P(~_B|~A) \times P(~_A)] \quad \text{(probability of mutant **BRAF** AND methylated **MLH1** promoter region irrespective of whether a **MLH1** mutation is present or not)}
\]

\[
= [0.16 \times 0.105] + [0.81 \times 0.895]
\]

\[
= 0.0168 + 0.725
\]

\[
= 0.742
\]

Post-test likelihood of true positive (i.e. **MLH1** mutation being present if **BRAF** is wild-type AND **MLH1** promoter region is normal)

\[
P(A|_B) = \frac{[P(_B|A) \times P(A)]}{P(_B)}
\]

\[
= \frac{[0.84 \times 0.105]}{0.263}
\]

\[
= 0.335 = 33.5\%
\]

Post-test likelihood of false negative (i.e. **MLH1** mutation being present despite mutant **BRAF** AND methylated **MLH1** promoter region)

\[
P(A|~_B) = \frac{[P(~_B|A) \times P(A)]}{P(~_B)}
\]

\[
= \frac{[0.16 \times 0.105]}{0.742}
\]

\[
= 0.0226 = 2.26\%
\]
2.3.2.8 Applying BRAF testing and MLH1 promoter region methylation testing to patients from the general population who have an MLH1 loss tumour

2.3.2.8.1 Wild-type BRAF as a predictor of germline MLH1 mutation

\[ P(B) = \left[ P_{(B|A)} \times P(A) \right] + \left[ P_{(B|\sim A)} \times P(\sim A) \right] \] (probability of wt BRAF irrespective of whether a MLH1 mutation is present or not)

\[ = \left[ 0.924 \times 0.04 \right] + \left[ 0.463 \times 0.96 \right] \]

\[ = 0.03696 + 0.44448 \]

\[ = 0.48144 \]

\[ P(\sim B) = \left[ P_{(\sim B|A)} \times P(A) \right] + \left[ P_{(\sim B|\sim A)} \times P(\sim A) \right] \] (probability of mutant BRAF irrespective of whether a MLH1 mutation is present or not)

\[ = \left[ 0.076 \times 0.04 \right] + \left[ 0.537 \times 0.96 \right] \]

\[ = 0.00304 + 0.51552 \]

\[ = 0.51856 \]

Post-test likelihood of true positive (i.e. MLH1 mutation being present if BRAF is wt)

\[ P(A|B) = \frac{P_{(B|A)} \times P(A)}{P(B)} \]

\[ = \frac{0.924 \times 0.04}{0.48144} \]

\[ = 0.03696/0.48144 \]

\[ = 0.0767 = 7.7\% \]

Post-test likelihood of false negative (i.e. MLH1 mutation being present despite mutant BRAF)

\[ P(A|\sim B) = \frac{P_{(\sim B|A)} \times P(A)}{P(\sim B)} \]

\[ = \frac{0.076 \times 0.04}{0.51856} \]

\[ = 0.00304/0.51856 \]

\[ = 0.00586 = 0.6\% \]
2.3.2.8.2 Normal MLH1 promoter region as a predictor of germline MLH1 mutation

\[ P(B) = [P(B|A) \times P(A)] + [P(B|\sim A) \times P(\sim A)] \] (probability of normal MLH1 promoter region irrespective of whether a MLH1 mutation is present or not)

\[ = [0.862 \times 0.04] + [0.221 \times 0.96] \]
\[ = 0.03448 + 0.21216 \]
\[ = 0.24664 \]

\[ P(\sim B) = [P(\sim B|A) \times P(A)] + [P(\sim B|\sim A) \times P(\sim A)] \] (probability of methylated MLH1 promoter region irrespective of whether a MLH1 mutation is present or not)

\[ = [0.138 \times 0.04] + [0.779 \times 0.96] \]
\[ = 0.00552 + 0.74784 \]
\[ = 0.75336 \]

Post-test likelihood of true positive (i.e. MLH1 mutation being present if MLH1 promoter region is normal)

\[ P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \]
\[ = \frac{[0.862 \times 0.04]}{0.24664} \]
\[ = 0.03448/0.24664 \]
\[ = 0.1397989 = 14.0\% \]

Post-test likelihood of false negative (i.e. MLH1 mutation being present despite methylated MLH1 promoter region)

\[ P(A|\sim B) = \frac{P(\sim B|A) \times P(A)}{P(\sim B)} \]
\[ = \frac{[0.0828 \times 0.04]}{0.75336} \]
\[ = 0.003312/0.75336 \]
\[ = 0.0044 = 0.4\% \]
2.3.2.8.3 Wild-type BRAF OR normal MLH1 promoter region as a predictor of germline MLH1 mutation

\[ P(B) = [P(B|A) \times P(A)] + [P(B|\sim A) \times P(\sim A)] \] (probability of wt BRAF OR normal MLH1 promoter region irrespective of whether a MLH1 mutation is present or not)

\[ = [0.949 \times 0.04] + [0.491 \times 0.96] \]
\[ = 0.3796 + 0.471 \]
\[ = 0.509 \]

\[ P(\sim B) = [P(\sim B|A) \times P(A)] + [P(\sim B|\sim A) \times P(\sim A)] \] (probability of mutant BRAF or methylated MLH1 promoter region irrespective of whether a MLH1 mutation is present or not)

\[ = [0.051 \times 0.04] + [0.509 \times 0.96] \]
\[ = 0.002 + 0.489 \]
\[ = 0.491 \]

Post-test likelihood of true positive (i.e. MLH1 mutation being present if BRAF is wt OR MLH1 promoter region is normal)

\[ P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \]
\[ = \frac{[0.949 \times 0.04]}{0.509} \]
\[ = 0.03796/0.509 \]
\[ = 0.0746 = 7.5\% \]

Post-test likelihood of false negative (i.e. MLH1 mutation being present despite mutant BRAF OR methylated MLH1 promoter region)

\[ P(A|\sim B) = \frac{P(\sim B|A) \times P(A)}{P(\sim B)} \]
\[ = \frac{[0.051 \times 0.04]}{0.491} \]
\[ = 0.002/0.491 \]
\[ = 0.0042 = 0.04\% \]
2.3.2.8.4 Wild-type BRAF AND normal MLH1 promoter region as a predictor of germline MLH1 mutation

\[ P(B) = [P(B|A) \times P(A)] + [P(B|\sim A) \times P(\sim A)] \] (probability of wild-type BRAF AND normal MLH1 promoter region irrespective of whether a MLH1 mutation is present or not)

\[ = [0.84 \times 0.04] + [0.19 \times 0.96] \]
\[ = 0.0336 + 0.1824 \]
\[ = 0.216 \]

\[ P(\sim B) = [P(\sim B|A) \times P(A)] + [P(\sim B|\sim A) \times P(\sim A)] \] (probability of mutant BRAF AND methylated MLH1 irrespective of whether a MLH1 mutation is present or not)

\[ = [0.16 \times 0.04] + [0.81 \times 0.96] \]
\[ = 0.0064 + 0.776 \]
\[ = 0.784 \]

Post-test likelihood of true positive (i.e. MLH1 mutation being present if BRAF is wild-type AND MLH1 promoter region is normal)

\[ P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \]
\[ = \frac{[0.84 \times 0.04]}{0.216} \]
\[ = 0.0336/0.216 \]
\[ = 0.1555 = 15.5\% \]

Post-test likelihood of false negative (i.e. MLH1 mutation being present despite mutant BRAF OR methylated MLH1 promoter region)

\[ P(A|\sim B) = \frac{P(\sim B|A) \times P(A)}{P(\sim B)} \]
\[ = \frac{[0.16 \times 0.04]}{0.784} \]
\[ = 0.00816 \]
\[ = 0.82\% \]
### 2.3.2.9 Summary of post-test probabilities of being an $MLH1$ mutation carrier (heterozygote) according to known $a$ priori risk

Table 22: Pre- and Post-test probabilities of being an $MLH1$ mutation carrier for three priori risk groups

<table>
<thead>
<tr>
<th>Pre-test probability</th>
<th>$^1$ACII &amp; $^2$dMMR tumour</th>
<th>$^2$BC &amp; dMMR tumour</th>
<th>General population &amp; dMMR tumour</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>60%</td>
<td>10.5%</td>
<td>4.0%</td>
</tr>
<tr>
<td>Post-test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wt $BRAF$</td>
<td>True positive</td>
<td>74.9%</td>
<td>19.0%</td>
</tr>
<tr>
<td></td>
<td>False negative</td>
<td>17.5%</td>
<td>1.6%</td>
</tr>
<tr>
<td>Post-test normal $MLH1$ promoter region</td>
<td>True positive</td>
<td>85.5%</td>
<td>31.4%</td>
</tr>
<tr>
<td></td>
<td>False negative</td>
<td>12.6%</td>
<td>1.2%</td>
</tr>
<tr>
<td>Post-test Wt $BRAF$ OR normal $MLH1$ promoter region</td>
<td>True positive</td>
<td>71.6</td>
<td>18.5</td>
</tr>
<tr>
<td></td>
<td>False negative</td>
<td>13.1</td>
<td>1.31</td>
</tr>
<tr>
<td>Post-test Wt $BRAF$ AND normal $MLH1$ promoter region</td>
<td>True positive</td>
<td>86.9</td>
<td>33.5</td>
</tr>
<tr>
<td></td>
<td>False negative</td>
<td>22.9</td>
<td>2.3</td>
</tr>
</tbody>
</table>

$^1$Amsterdam II criteria

$^2$Revised Bethesda criteria

$^3$dMMR= MMR deficient

Applying $BRAF$ mutation and $MLH1$ promoter region methylation analysis to tumour DNA from patients who fulfil the Amsterdam II criteria is uninformative. Although the identification of wild-type $BRAF$ and/or normal $MLH1$ promoter region highly suggests that a germline $MLH1$ mutation is present, mutant $BRAF$ V600E and/or $MLH1$ promoter region methylation does not bring the (post-test) risk of a mutation to below 10%. This is true whether one or other of the tests are performed or if they are performed in combination.

Applying $BRAF$ mutation and $MLH1$ promoter region methylation analysis to tumour DNA from patients who fulfil the revised Bethesda criteria and have loss of $MLH1$ on
immunohistochemistry is informative. The identification of wild-type BRAF suggests a post-test risk of 19.0%. Normal MLH1 promoter region suggests a post-test risk of a germline MLH1 mutation of 31.4%. The finding of mutant BRAF V600E in this setting brings the post-test risk to 1.6%. The finding of MLH1 promoter region methylation brings the post-test risk to 1.2%. Application of the two tests in combination results in a post-test probability of a true positive risk of 18.5% if either is wild-type/normal, and 33.5% if both are wild-type/normal. Post-test false negative risk is 1.3% if either is wild-type/normal and 2.3% if both are.

Applying BRAF mutation analysis to tumour DNA from patients from the general population who have loss of MLH1 on IHC is uninformative. Wild-type BRAF only indicates a post-test risk of 7.7%, so germline MLH1 mutation testing would not be indicated. This is due to the lower specificity of wild-type BRAF (65.75% [95% CI 53.72-76.47%]) for the identification of MLH1 mutation carriers. Applying MLH1 promoter region methylation analysis to tumour DNA from patients from the general population who have loss of MLH1 on IHC is informative. The finding of normal MLH1 promoter region suggests a post-test risk of a germline MLH1 mutation of 14.0%. The finding of methylated MLH1 promoter region suggests a post-test risk of 0.4%. Applying both BRAF and MLH1 analysis is only informative if both are wild-type/normal. The post-test risk is 15.5% of a true positive and 0.8% risk of a false negative.
2.3.2.10 Applying BRAF and MLH1 analysis to *a priori* risk

Table 23: Performance characteristics of *BRAF* and *MLH1* methylation analysis according to varying *a priori* risk

<table>
<thead>
<tr>
<th>Prior risk</th>
<th>Wild-type <em>BRAF</em></th>
<th>Normal <em>MLH1</em> promoter region</th>
<th>Wild-type <em>BRAF</em> OR Normal <em>MLH1</em> promoter region</th>
<th>Wild-type <em>BRAF</em> AND Normal <em>MLH1</em> promoter region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True positive</td>
<td>False negative</td>
<td>True positive</td>
<td>False negative</td>
</tr>
<tr>
<td>2.5%</td>
<td>2.4%</td>
<td>0.4%</td>
<td>9.1%</td>
<td>0.3%</td>
</tr>
<tr>
<td>5%</td>
<td>9.6%</td>
<td>0.7%</td>
<td>17.0%</td>
<td>0.6%</td>
</tr>
<tr>
<td>10%</td>
<td>19.0%</td>
<td>1.6%</td>
<td>31.4%</td>
<td>1.2%</td>
</tr>
<tr>
<td>20%</td>
<td>33.3%</td>
<td>4.3%</td>
<td>49.2%</td>
<td>2.6%</td>
</tr>
<tr>
<td>30%</td>
<td>46.1%</td>
<td>5.7%</td>
<td>63.1%</td>
<td>4.3%</td>
</tr>
<tr>
<td>40%</td>
<td>57.1%</td>
<td>8.6%</td>
<td>71.8%</td>
<td>6.2%</td>
</tr>
<tr>
<td>50%</td>
<td>66.9%</td>
<td>12.2%</td>
<td>79.5%</td>
<td>9.0%</td>
</tr>
<tr>
<td>60%</td>
<td>74.9%</td>
<td>17.5%</td>
<td>85.5%</td>
<td>12.6%</td>
</tr>
</tbody>
</table>
From table 23, it can be deduced that wild-type \textit{BRAF} is able to identify patients who should undergo \textit{MLH1} gene testing if that patient has 10-40\% \textit{a priori} risk. It can also be deduced that normal (unmethylated) \textit{MLH1} promoter region is able to identify patients who should undergo \textit{MLH1} gene testing if that patient has 5-50\% \textit{a priori} risk. When used in combination, wild-type \textit{BRAF} and normal \textit{MLH1} promoter region are able to predict patients who should undergo \textit{MLH1} gene testing if that patient has 2.5-30\% \textit{a priori} risk. If either test is wild-type/normal, patients with \textit{a priori} risk of 10-50\% can be identified.

2.3.2.11 Cost

The total laboratory cost of germline \textit{MLH1} mutation testing is £483 per individual. The tumour \textit{MLH1} promoter region methylation assay costs £138 per individual and tumour V600E \textit{BRAF} testing is £69 (oral communication Dr Andrew Wallace PhD, Senior Clinical Scientist, Saint Mary’s Hospital Molecular Genetics Laboratory, March 2012). These costs are laboratory costs only and do not take into account the clinical time needed for patient counselling and sample collection.

Overall cost per mutation carrier identified is (for example if 100 individuals are tested):

\[
\frac{\text{[Cost of pre-screen test for 100 individuals + cost of germline mutation testing for all those testing positive in the pre-screen test (i.e. all true positives and false positives)]}}{\text{[number of germline mutation carriers identified]}}
\]

For the calculations below, the true positive and false positives rates are used from the Bayes calculations in sections 2.3.2.7 and 2.3.2.8. The \textit{a priori risk} is used as number of mutation carriers identified.

Patients who fulfil Bethesda criteria and have loss of MLH1 in their tumour have a 10.5\% risk of harbouring a germline \textit{MLH1} mutation. If 100 individuals are investigated:

- No pre-screen test is used: All individuals are tested for germline \textit{MLH1} mutation to identify 10.5 patients with a germline mutation.
  - Cost = 100 \times 483 = £48,300
  - Cost per germline mutation identified = 48300/10.5 = £4,600
- MLH1 methylation testing is used: All those who have a normal (unmethylated) promoter region (true positives and false positives) are tested for germline MLH1 mutation to identify 10.5 patients with a germline mutation.
  - Cost of methylation testing = 100 $\times$ 138 = £13,800
  - Cost of germline mutation testing = 28.8 $\times$ 483 = £13,668.9
  - Total cost = £27,468.9
  - Cost per germline mutation identified = £27,468.9/10.5 = £2,616.1

- Tumour BRAF mutation testing is used. All those who have a normal wild-type BRAF (true positives and false positives) are tested for germline MLH1 mutation to identify 10.5 patients with a germline mutation.
  
  Cost of BRAF testing = 100 $\times$ 69 = £6,900
  Cost MLH1 germline testing = 51.1 $\times$ 483 = £24,681.3
  Total cost = £6,900 + 24,681.3
  Cost per germline MLH1 mutation identified = 31,581/10.5 = £3,007.7

Patients from the general population who have loss of MLH1 in their tumour have a 7.3% risk of harbouring a germline MLH1 mutation. If 100 individuals are investigated:
- No pre-screen test is used therefore all individuals are tested for germline MLH1 mutation to identify 7.3 patients with a germline mutation.
  
  Cost = 100 $\times$ 483 = £48,300
  Cost per germline mutation identified = 48,300/7.3 = £6,616.4

- MLH1 methylation testing is used. All those who have a normal (unmethylated) promoter region (true positives and false positives) are tested for germline MLH1 mutation to identify 7.3 patients with a germline mutation.
  - Cost of methylation testing = 100 $\times$ 138 = £13,800
  - Cost of germline mutation testing = 24.7 $\times$ 483 = £11,930.1
  - Total cost = £25,730.1
  - Cost per germline mutation identified = £25,730.1/7.3
- Tumour *BRAF* mutation testing is used. All those who have a normal wild-type *BRAF* (true positives and false positives) are tested for germline *MLH1* mutation to identify 7.3 patients with a germline mutation.

  Cost of BRAF testing = 100 x 69 = £6900  
  Cost MLH1 germline testing = 48.1 x 483 = £23232.3  
  Total cost = £6900 + 23232.3 = £30,132.3  
  Cost per germline *MLH1* mutation identified = 30,132.3/7.3  
  = £4127.7

For patients who fulfil the Bethesda criteria and have a loss of MLH1 protein in their tumour, the cost to identify one individual with a germline *MLH1* mutation is £4600 if no pre-screen test is used, £2616.1 if tumour *MLH1* methylation testing is used, and £30007.7 if *BRAF* testing is used.

For patients from the general population who have loss of MLH1 protein in their tumour, the cost to identify one individual with a germline *MLH1* mutation is £6616.4 if no pre-screen test is used, £3524.7 if tumour *MLH1* methylation testing is used, and £4127.7 if *BRAF* testing is used.

### 2.3.3 Results of *KRAS* mutation analysis

*KRAS* mutation analysis was only performed on the tumours sourced locally due to the small volume of tissue from the C-CFR samples.

*KRAS* mutations (codons 12/13) were found in 28/49 (57.0%) local sporadic mismatch proficient colorectal cancers.

*KRAS* mutations were found in tumours from 4/22 (9.0%) local *MLH1* mutation carriers and in 2/31 (6.5%) local sporadic MLH1 loss tumours. *KRAS* mutations were only found in tumours with wild-type *BRAF*. It can be said that *KRAS* mutations are found less frequently in mismatch repair deficient cancers, however *KRAS* mutation analysis is not able to differentiate between sporadic MLH1 loss tumours and tumours from *MLH1* mutation carriers.
2.4 Discussion Study one: DNA mismatch gene promoter region methylation analysis and *BRAF* gene mutation analysis—an alternative prescreening strategy in Lynch Syndrome

### 2.4.1 Discussion

This study has demonstrated the utility of pyrosequencing colorectal cancer tumour DNA for *MLH1* promoter region methylation in clinical practice in a large cohort of CRCs from 71 pathogenic *MLH1* mutation carriers and 73 sporadic MLH1 loss CRCs. A novel *MLH1* promoter region methylation assay has been developed to GLP (Good Laboratory Practice) standards and clinical utility demonstrated in the assessment of Bethesda criteria patients and in population based pre-screening for Lynch Syndrome. A single assay is time and cost effective and may encourage the introduction of pre-screening into routine clinical practice. Unmethylated *MLH1* promoter region has a sensitivity of 94.4% (95% CI 86.2-98.4%) and a specificity of 87.7% (95% CI 77.9-94.2%) for the identification of *MLH1* mutation carriers from a group of cancers with MLH1 loss mismatch repair deficiency. This is the first large scale assessment of pre-screening mismatch repair deficient CRCs using tumour *MLH1* promoter methylation and *BRAF* mutation for the presence of a germline *MLH1* mutation.

Normal (un-methylated) *MLH1* promoter region has a sensitivity of 94.37% (95% CI 86.2-98.44%) and a specificity of 87.67% (95% CI 77.88-94.2%), and wild-type *BRAF* V600E has a sensitivity of 98.59% (95% CI 92.4-99.96%) and a specificity of 65.75% (95% CI 53.72-76.47%) for the detection of *MLH1* mutation carriers from a group of cancers with MLH1 loss mismatch repair deficiency. When used in combination wild-type *BRAF* OR normal *MLH1* promoter region as has a sensitivity of 100% (95% CI 94.94% to 100%) and a specificity of 63.01% (95% CI 50.91% to 74.03%). Wild-type *BRAF* AND normal *MLH1* promoter region has a sensitivity of 92.96% (95% CI 84.33% to 97.67%) and a specificity of 90.41% (81.24% to 96.06%).
MLH1 methylation analysis is 4.2% less sensitive than BRAF mutation analysis, but is 21.9% more specific. In the sample of 71 MLH1 mutation carriers, 67/71 (94.4%) were unmethylated and 4/71 (5.6%) were methylated. In the sample of 73 sporadic MLH1 loss cancers, 64/73 (87.7%) were methylated and 9/73 (12.3%) were unmethylated. In the sample of 71 mutation carriers 70/71 (98.6%) were wild-type BRAF and 1/71 (1.4%) had the V600E mutation. In the sample of 73 sporadic MLH1 loss cancers, 48/73 (65.8%) had the BRAF V600E mutation and 25/73 (34.2%) were wild-type.

Application of Bayes theorem to the sensitivity and specificity of these tests, with knowledge of an individual’s a priori (pre-test) risk of harbouring a germline MLH1 mutation, allows refinement of that individual’s risk. This enables more efficient identification (pre-screening) of those individuals who should undergo testing for a germline MLH1 mutation (greater than 10% risk) and those that should not (less than 10%).

Individuals who fulfil Amsterdam II criteria have a priori risk of carrying an MLH1 mutation of 60%. If their tumour demonstrates a normal (unmethylated) MLH1 promoter region, they have an 85.5% chance of an MLH1 mutation. If a methylated MLH1 promoter region is demonstrated, they have a 12.6% chance of having a MLH1 mutation, therefore should still undergo germline mutation testing. If wild-type BRAF is found, their risk is increased to 74.9%. If their tumour has the BRAF V600E mutation, the risk is 17.5%, thus the individual should still undergo germline MLH1 mutation testing. Using the two tests in combination results in a post-test risk of 71.6% if either is wild-type/normal, and 86.9% risk if both are wild-type/normal. If either is mutated/methylated the risk is 13.1%, and if both are mutated/methylated the risk is 22.9%. The application of BRAF mutation and MLH1 promoter region methylation testing to individuals who fulfil Amsterdam II criteria (whether individually or in combination) does not alter who should be tested for germline mutations. This is an ineffective pre-screening strategy in this group.

Individuals that fulfil the revised Bethesda criteria and have a tumour with loss of MLH1, have a priori risk of at least 10.5% of having a germline MLH1 mutation. If the tumour is found to have normal MLH1 promoter region the risk is increased to 31.4%, and if MLH1 promoter region is methylated the risk is reduced to 1.2%. If their tumour DNA demonstrates wild-type BRAF, the risk of an MLH1 mutation is increased to 19.0%. If their tumour is found to have the BRAF V600E mutation, this
risk is reduced to 1.6%. Using the two tests in combination results in a post-test risk of 18.5% if either is wild-type/normal, and 33.3% risk if both are wild-type/normal. If either is mutated/methylated the risk is 1.3%, and if both are mutated/methylated the risk is 2.3%. The application of either *BRAF* V600E or *MLH1* promoter region methylation analysis, or in combination, to MLH1 loss CRCs from patients who fulfil the revised Bethesda criteria allows effective identification of patients who are at, at least, 10% risk of having an *MLH1* mutation and should therefore undergo germline mutation analysis. The presence of *BRAF* V600E mutation and/or methylation of the *MLH1* promoter region suggest a less than 10% risk, so these patients do not need mutation analysis.

In this setting, an assay with high sensitivity is required. The pre-test risk in this patient group is higher than the 10% cut-off for *MLH1* gene testing. The aim of conducting pre-screening tests is to reduce the number of patients who are tested for germline mutations, without excluding any patients who may have germline mutations. The combination of wild-type *BRAF* OR normal *MLH1* promoter region methylation has the highest sensitivity of 100% (95% CI 94.94% to 100%), and good specificity 90.41% (95% CI 81.24% to 96.06%). There is a spectrum of risk within the Bethesda risk group. A patient with a single CRC in their late 40s and no family history will have a lower risk than a patient who has had a CRC in their 20s and a metachronous CRC in their 30s with no family history, or a patient who had a CRC in their 40s whose parent also had a CRC in their 40s (lack of a third relative with cancer rules out the Amsterdam II criteria). There is little published data to allow calculation of the specific *a priori* risk of Lynch Syndrome according to age at primary cancer, or other details within the family pedigree. However, some Bethesda patients (with MLH1 loss cancers) will have a priori risk (for an *MLH1* mutation) much higher than 10.5%. This study has demonstrated, using Bayes theorem, that the use of *BRAF* and *MLH1* promoter region methylation analysis (where either wild-type *BRAF* or normal *MLH1* promoter region suggests screening for *MLH1* germline mutations) as pre-screening is valid in patients with a priori risk of between 10 and 50%. This is likely to encompass all patients who fulfil the revised Bethesda criteria. When MLH1 methylation analysis is used alone, normal MLH1 promoter region predicts a post-test risk of greater than 10% and a false negative rate of less than 10% for patients with a priori risk of between 5 and 50%. Using *MLH1* promoter region methylation analysis alone is likely to be more cost-effective than use in combination with *BRAF* mutation analysis.
It can be said that individuals from the general population (with no known familial mutation) who have a CRC which demonstrates MLH1 loss on immunohistochemistry have at least a 4.0% a priori risk of having a pathogenic germline MLH1 mutation. If their tumour is found to have normal MLH1 promoter region the risk is increased to 14.0%, and if MLH1 is methylated the risk is reduced to 0.4%. If their tumour is found to have wild-type BRAF, the risk of having an MLH1 mutation is 7.7%. If their tumour is found to have the BRAF V600E mutation, this risk is 0.6%. This data suggests that at a 4.0% level of a priori risk, MLH1 methylation analysis (of tumours with demonstrable loss of MLH1) is able to differentiate between those who should undergo testing for a germline MLH1 mutation and those whose risk is sufficiently low for germline testing not to be indicated. An assay with the highest specificity is required in this setting. The pretest risk of carrying an MLH1 mutation is low. If population based biomarker testing is to be put into clinical practice it needs to be cost and time efficient. The combination of wt BRAF and normal MLH1 has the highest specificity (90.41% [95% CI 81.24% to 96.06%]) and retains good sensitivity (92.96% [95% CI 84.33% to 97.67%]). However, normal MLH1 promoter region has just slightly lower specificity but as a single assay will be more time and cost effective in the population based setting (sensitivity 94.37% [895% CI 6.2-98.44%] and specificity 87.67% [95% CI 77.88-94.2%]).

It was previously unknown whether BRAF mutation or MLH1 promoter region methylation analysis or both in combination, is better able to distinguish between sporadic MLH1 loss colorectal cancers and cancers in patients with germline MLH1 gene mutation. Jensen et al demonstrated that although MLH1 methylation and BRAF mutation are associated, MLH1 methylation also occurs in a proportion of sporadic MLH1 loss colorectal cancers with wild-type BRAF. They examined 287 consecutive CRC cases with mismatch repair immunohistochemistry. In those with MLH1 loss (n=29), tumour DNA was extracted and sequenced for BRAF mutation. Those that were wild-type for BRAF were then investigated for MLH1 promoter region methylation. However, they did not test any of the patients for germline MLH1 mutations (Jensen, Dysager et al. 2010). Bouzourene et al found all MLH1 deficient cancers had MLH1 methylation (11/11) and 73% (8/11) had BRAF mutation. Of their 16 Lynch cases 6% (only 1 case) had MLH1 methylation and none had BRAF mutation (Bouzourene, Hutter et al. 2010). Perez-Carbonell et al investigated 10 MLH1 lynch cancers, and 63 sporadic MLH1 loss cancers. They calculated a specificity of 40% for BRAF mutation for the identification of mutation...
carriers, and a specificity of 75-78% (two methods used) of MLH1 methylation (Perez-Carbonell, Alenda et al. 2010). Due to the low number of Lynch cancers, sensitivity could not be calculated accurately.

Sporadic mismatch repair deficient cancers that demonstrate loss of MLH1 protein on IHC and are MSI-high are associated with MLH1 gene silencing through the epigenetic effect of promoter region methylation. In this study, MLH1 promoter region methylation was found in 64/73 (87.7%) sporadic mismatch repair deficient colorectal cancers. 33/73 (45.0%) of the sporadic mismatch repair deficient CRCs are known to be negative for germline MLH1 mutations. Of these 28/33 (84.8%) were found to have MLH1 methylation. Methylation was consistent across all four cytosine residues examined within the functional area of the promoter region as described by Deng et al (Deng, Bell et al. 2004). Pyrosequencing allows accurate quantification of the methylation at each residue. Each sample was tested in triplicate which adds certainty to the result. Each cytosine across all triplicates was methylated to a level of at least 50% and the majority were above 80%. The vast majority of the unmethylated samples demonstrated 100% non-methylation at the cytosine residues interpreted by 100% thiamine (converted from unmethylated cytosine by bisulphite conversion) and 0% cytosine. Low level methylation (<5%) was found in around 15% of samples, but this is not considered to be functionally significant (Deng, Bell et al. 2004). MLH1 promoter region methylation was not found in any mismatch repair proficient CRCs nor in any normal tissue. Samples with lower methylation levels also had lower levels of BRAF mutation. This may indicate that these samples contained proportionally less tumour to normal tissue. 9/73 (12.3%) sporadic MMR deficient CRCs were found to have normal MLH1 promoter region. Two were found to have BRAF mutation. Of the seven that were wild-type BRAF, four had been tested for germline MLH1 mutations by the ACCFR and were found to be negative. The remaining three had been classified as sporadic MMR deficient due to the patient’s age (over 50 years) and lack of family history. The aetiology of these cancers without promoter region hypermethylation is unclear but possible factors include loss of protein expression, somatic mutation of MLH1 and loss of heterozygosity (Kuismanen, Holmberg et al. 2000). It is feasible that the MLH1 promoter displayed mosaic or heterogeneous patterns of methylation for the CpGs dinucleotides captured in the pyrosequencing amplicon but enough of the surrounding CpGs dinucleotides were methylated to result in loss of MLH1 protein expression. Alternatively, the untested patients may be carriers of germline MLH1 mutations.
It has been thought that MLH1 promoter methylation is found exclusively in sporadic MMR deficient CRCs (Clendenning, Buchanan et al.; Southey, Jenkins et al. 2005; Chen, Wang et al. 2006; Lindor, Petersen et al. 2006; Jass 2007; Jensen, Dysager et al. 2010). Recent population-based studies that have investigated screening for LS mutations have used MLH1 methylation to exclude patients from germline testing (Lindor, Petersen et al. 2006). One group has reported one case of MLH1 promoter methylation in a pathogenic MLH1 mutation carrier. This was a CRC of the transverse colon in a patient aged 59 with a MLH1, c739T>C (pSer247Pro) mutation (missense variant so could be a variant with uncertain pathogenicity) and was reported at 66% methylation (SNapshot methylation analysis). (Rahner, Friedrichs et al. 2008). The current study is the largest dataset of MLH1 mutation carriers tested for MLH1 promoter region methylation. MLH1 methylation was found in 4/71 (5.6%) mutation carriers. All these individual tumours displayed significant methylation at all four of the cytosine residues (within CpG islands) examined. Sample 11005300 was methylated at a level of 29-38% across the four cytosine residues and each triplicate assay. This tumour was a carcinoma in situ within an advanced tubulovillous adenoma from a patient aged 40 with an MLH1 mutation 405 insertion A in Exon 5 (frame shift mutation which results in a truncated protein). A further CRC from an individual within the same family (a first degree relative) and with the same mutation was also tested as part of this study (sample 11005240) and did not show methylation. Sample 11007868 was methylated at a level of 19-60% across the four cytosine residues and each triplicate. Sample 11007938 was methylated at a level 70-80%; sample 11007944 was methylated at a level 8-82%. These three samples were obtained from the Colon Cancer Family Registry so further demographic and genetic data is unknown. It has previously been suggested that sporadic inactivation of the second normal MLH1 allele by methylation may be the ‘second hit’ event in mutation carriers (Rahner, Friedrichs et al.).

This low frequency of MLH1 methylation in mutation carriers is supported by a recent literature review and meta-analysis conducted by Parsons and colleagues (Parsons M 2012). This study reported low frequency of MLH1 methylation in MLH1 mutation carriers (5.56%) finding eight positively methylated tumours in MLH1 mutation carriers taken from 12 studies. Whilst MLH1 promoter region methylation is an infrequent event (5.6% in this study and also the review) in carriers of
germline mutations, this data demonstrates that it is not rare and supports the hypothesis that it may be the second hit event. These findings also suggest that the discovery of MLH1 methylation does not exclude the diagnosis of Lynch Syndrome. Whilst the sensitivity of MLH1 methylation testing is adequate for low risk individuals, it is not for those with a high prior risk as in those who fulfil Amsterdam criteria.

*BRADF* gene mutations are found in 5-15% of all colorectal cancers (Davies, Bignell et al. 2002; Rajagopalan, Bardelli et al. 2002). They are found more frequently in cancers from Jass’s subtypes 1 and 2 (see section 1.3.2). Subtype 1 is MSI-high, chromosome stable, CIMP high, has methylation of the MLH1 promoter region and is thought to account for 13% of CRCs. Subtype 2 is MSI-low or stable, chromosome stable, CIMP-high, has partial methylation of the MLH1 promoter and is thought to account for 8% of all CRCs. Both these subtypes are said to originate in serrated lesions. *BRADF* mutation is thought to be an unequivocal marker of the serrated neoplasia pathway. Therefore, the discovery of a *BRADF* mutation within a tumour is thought to rule out Lynch Syndrome (Walsh, Buchanan et al. 2009). In the current study *BRADF* V600E mutation was found in 4/52 (8%) mismatch repair proficient CRCs and 48/73 (66%) sporadic MLH1 loss CRCs. This is consistent with previous studies (Wang, Cunningham et al. 2003; Woods, Younghusband et al. 2010). Interestingly a *BRADF* mutation was detected in 1/71 (1.4%) CRC from a pathogenic MLH1 mutation carrier. This particular CRC had been obtained through the Colon Cancer Family Registry, who had previously conducted a *BRADF* V600E allele specific PCR assay as part of their ongoing investigations. This confirmed the presence of the somatic *BRADF* mutation (Buchanan). *BRADF* mutations have previously been reported as a rare finding in patients with Lynch Syndrome (Vandrovcova, Lagerstedt-Robinsson et al. 2006; Lagerstedt Robinson, Liu et al. 2007), and are thought to represent a mixed lineage of cancer predisposition (multiple independently inherited mutations/variants). Walsh et al have reported on two families with evidence of Lynch Syndrome and probable additional germline factors. These family pedigrees demonstrate Lynch Syndrome with germline pathogenic mutations, and a seemingly independently segregating serrated neoplasia phenotype. *BRADF* mutations were found both within serrated polyps and CRCs (Walsh, Buchanan et al.). Senter et al investigated 99 probands with Lynch spectrum cancers that demonstrated isolated loss of PMS2 on IHC. Germline *PMS2* mutations were detected in 62%, and three (one exon 10 deletion, two c.736_741del6ins11) of these were found to have tumour p.*BRADF* V600E mutation
It has previously been suggested that in mismatch repair deficient cancers, \textit{BRAF} mutation was a surrogate marker for \textit{MLH1} promoter methylation. However, there is now gathering evidence that \textit{BRAF} mutation occurs in only 50-75\% of sporadic MLH1 loss cancers. In a series of 270 CRCs, Wang et al found \textit{BRAF} mutations in 21/170 (12\%) mismatch repair proficient cases and 42/123 (34\%) mismatch repair deficient cases. \textit{BRAF} was closely associated with \textit{MLH1} promoter hypermethylation (30/36 [83.3\%] hypermethylated cases also had a \textit{BRAF} mutation) (Wang, Cunningham et al.).

In a large population based study, Woods et al examined 68 MSI-high CRCs for both \textit{BRAF} and \textit{MLH1} methylation. \textit{BRAF} mutation was closely, but not exclusively associated with \textit{MLH1} methylation. 31/40 (78\%) of the hypermethylated tumours also had \textit{BRAF} mutations, where as only 2/28 (7.1\%) of the non-methylated tumours had \textit{BRAF}. In the current study 46/73 (63.0\%) sporadic MLH1 loss cancers had both \textit{BRAF} mutation and \textit{MLH1} methylation, but 18/73 (24.7\%) had only \textit{MLH1} methylation. This is consistent with previous studies (Wang, Cunningham et al. 2003; Woods, Younghusband et al. 2010).

There are some limitations to the current study. A proportion of the sporadic samples did not undergo germline testing for \textit{MLH1} mutations. Even full sequencing and a dosage test of \textit{MLH1} may miss mutations such as deep intronic splicing mutations and sensitivity may therefore be reduced to 90-95\%. Clendenning et al reported the discovery of an intronic \textit{MSH2} mutation, 478bp upstream from exon 2 causing LS (Clendenning, Buchanan et al. 2011). As such some of the ‘sporadic’ non germline \textit{MLH1} loss CRCs may have had an undiscovered germline mutation. Having said this, it is unlikely that more than one of the 73 samples would have been falsely called as \textit{MLH1} mutation negative. This would be very unlikely to affect the results or conclusions to any great extent. Also the rate of non-methylated \textit{MLH1} and wt \textit{BRAF} in the 40 (4/40, 10\%) sporadic tumours with mutation testing was not different to that in the 33 (3/33, 9.1\%) untested sporadic cases.
Germline MLH1 promoter region methylation has been described as a rare (32 reported cases) cause for CRC (van Roon, van Puijenbroek et al.; Gazzoli, Loda et al. 2002; Hitchins, Williams et al. 2005; Hitchins, Wong et al. 2007; Valle, Carbonell et al. 2007; Morak, Schackert et al. 2008; Niessen, Hofstra et al. 2009; Goel, Nguyen et al. 2011). Whilst it is thought that this epimutation is usually erased in the gametes, inheritance has been demonstrated in three cases (Hitchins, Wong et al. 2007; Morak, Schackert et al. 2008; Crepin, Dieu et al. 2011). This is likely a rare cause for tumour DNA MLH1 promoter region methylation, and an exceedingly rare cause for Lynch syndrome. However, in a young onset MLH1 loss CRC with MLH1 promoter region methylation, it may be prudent to examine the MLH1 promoter of germline DNA.

This study has demonstrated the utility of a novel MLH1 promoter region methylation assay. This assay allowed quantification of CpG island methylation, which when found was consistent, reproducible and high (>50%). This assay requires 10µm FFPE tissue with at least 80% tumour cells. Inconsistency in results was only seen when there was a much higher proportion of normal tissue within the sample, or if the amount of tissue available for extraction was much smaller. In this case the concentration of DNA eluted was low (e.g. 5ng/µl). The selection of DNA from this small sample for each of the triplicate PCRs may have had disproportionate amounts of tumour and normal DNA resulting in inconsistent pyrograms. When this occurred in this study, the assay was repeated with a larger volume of tissue for extraction and consistent results were always achieved. Reproducibility was confirmed during the assay validation process. The assay was validated to Good Laboratory Practice diagnostic assay standards on CRC specimens archived in the North-West molecular genetics laboratory. This process tests that the assay protocol can be followed multiple times on the same sample by different (appropriately trained) individuals and the same result achieved (data not included in this thesis).

Identification of families with LS is vital to enable reduction in morbidity and mortality with surveillance. The Revised Bethesda criteria have been criticised for being overly complicated and are known to be little used in clinical practice (Overbeek, Hoogerbrugge et al. 2008; Van Lier, De Wilt et al. 2009; Singh, Schiesser et al. 2010). The Amsterdam II criteria are problematic due to low
sensitivity (Lindor, Petersen et al. 2006). The use of population based pre-screening has been hampered by a lack of evidence for the specificity and sensitivity of MLH1 promoter region methylation analysis for the detection of mutation carriers. It is hoped that this current study provides that evidence.

As previously mentioned (see section 1.7.6.2) there is gathering evidence that mismatch repair deficient cancers do not respond to 5-fluorouracil based chemotherapy which is the mainstay of most CRC chemotherapy regimens (Jover, Zapater et al. 2009). Recent data suggests that the origin of the mismatch repair is important. Cancers in patients from Lynch Syndrome mutations may respond whilst sporadic mismatch repair deficient cancers may not (Sinicrope, Foster et al. 2011). Whilst conducting this research an application was made to the cancer care commissioning body of Greater Manchester (Oldham primary care trust) to fund cancer network wide mismatch repair immunohistochemistry. This application is ongoing. If this application is successful, and all CRCs being considered for chemotherapy are subjected to mismatch repair immunohistochemistry, it will be vital to be able to differentiate which cancers are sporadic and which are caused by Lynch Syndrome. Not only for the future screening of the proband and their family members but also for the adjuvant therapy of the incident CRC. The MLH1 promoter region methylation assay may prove vital in this assessment.

2.4.2 Conclusion

MLH1 promoter region methylation analysis is simple, reproducible and cost-effective (see section 2.3.2.11). It can be used in conjunction with mismatch repair protein (MMR) IHC in the pre-screening of low and moderate risk patients for Lynch Syndrome mutation testing. Patients who fulfil Amsterdam criteria be tested for germline mutations regardless of pre-screening status.

2.4.3 Future work

MLH1 promoter region methylation can be reliably detected using the novel pyrosequencing assay. MLH1 promoter region methylation analysis is an effective
way to distinguish between sporadic MLH1 loss cancers and cancers from *MLH1* mutation carriers in the general population, and amongst patients who fulfil Bethesda criteria. This study has demonstrated the clinical utility of this risk assessment biomarker. The biomarker is detected in an easy to obtain sample (archived FFPE tissue) and the assay is reliable and reproducible, and is time and cost effective. This study has demonstrated a robust relationship between the biomarker and the disease in question. In order to complete the biomarker roadmap as described by Cancer research UK(<http://science.cancerresearchuk.org/gapp/fundingcommittees/biomarker_imaging_committee/> Accessed 2009), prospective trial evidence which examines the use of the biomarker in the target population (either Bethesda patients or general population) is needed.

*MLH1* promoter region methylation and *BRAF* V600E mutation testing are now being used in clinical practice in the Department of Genetic Medicine, Saint Mary’s Hospital, Central Manchester University Hospital Foundation Trust. All families who are referred with a moderate family history of CRC, or who fulfil Bethesda criteria have tumour testing conducted on the index case if tissue is available. MMR protein IHC is initially conducted. If MSH2, MSH6 or PMS2 loss is demonstrated, germline mutation testing is conducted. If MLH1 loss is demonstrated, tumour DNA is extracted and *MLH1* promoter region methylation and *BRAF* mutation analysis conducted as described in this thesis. Germline *MLH1* analysis is conducted on all patients who fulfil Bethesda criteria and have MLH1 loss. This is a prospective trial in Bethesda criteria patients.
3 Study two: Genotype-phenotype correlation in colorectal polyposis

3.1 Background for Study two

Study two has been published (Newton, Mallinson et al. 2011).

3.1.1 Introduction

The discovery of the APC gene, and latterly the MutYH gene, represent a major breakthrough for the genetic management of polyposis patients in terms of predictive testing for at risk patients. Phenotypic variation both amongst and within families is well recognised. Since the discovery of the APC gene there is evidence that suggests that the site of the inherited APC mutation may correlate with the patient’s particular FAP phenotype. It is remains unclear whether genotyping of FAP patients is relevant for clinical management in terms of screening endoscopy protocols and, timing and nature of prophylactic surgery.

In 1988, Lynch and colleagues first described a colorectal cancer (CRC) prone family in which many members had mild polyposis (less than 100 polyps) (Lynch, Smyrk et al. 1988). The existence of a spectrum of polyposis subtypes caused by mutations in the APC gene was suggested by Leppert and colleagues in 1991. This followed the investigation of a large kindred with a history of CRC. Two individuals within this family were found to have classic polyposis (greater than 100 polyps) but eight individuals were found to have much milder polyposis (between two and 40 polyps). Genetic linkage analysis of the affected individuals suggested a causal APC mutation (Leppert, Burt et al. 1990). Spirio and colleagues latterly discovered four different mutations within the APC gene in seven families with an attenuated phenotype similar to that of Leppart’s prototype family (Spirio, Olschwang et al. 1993).

Three FAP phenotypes have been suggested. There is no consensus on the exact boundaries of the mutation loci responsible. Mild or attenuated FAP (AFAP or attenuated adenomatous polyposis coli; AAPC) is characterised by a later age of
onset of polyposis and fewer polyps, yet patients sustain a greatly elevated risk of colorectal cancer. AFAP has been associated with mutations before codon 157, after codon 1595, or in the alternatively spliced fragment of exon 9 (Giardiello, Brensinger et al. 1997; Walon, Kartheuser et al. 1997). Severe, or profuse polyposis, with earlier age of onset of polyposis and much greater numbers of polyps, is thought to be associated with mutations from codons 1250 to 1464 (especially codon 1309) (Nagase, Miyoshi et al. 1992; Friedl, Caspari et al. 2001). Classical FAP mutations are thought to lie within the remainder of the APC gene, in particular the 5’ end between codon 168 and 1550/1580 excluding codon 1309 (Friedl, Caspari et al. 2001; Nieuwenhuis and Vasen 2007).

It is known that desmoid disease is associated with mutations 3’ to codon 1500. Mutations in this region are associated with the mild polyposis phenotype. Patients with a mutation in this region have an 85% cumulative risk of desmoid disease if they undergo surgery and a 10% cumulative risk if they do not (Speake, Evans et al. 2007).

Mutations from codons 1250 to 1464 are thought to be associated with severe or profuse polyposis. The evidence is strongest for patients with mutations in codon 1309 (Friedl, Caspari et al. 2001; Nieuwenhuis, Bulow et al. 2009). Studies have suggested that patients with these mutations who undergo a colectomy and IRA have a significantly increased risk of progressive rectal disease and therefore the need for a later completion proctectomy (Bertario, Russo et al. 2000; Sinha, Tekkis et al. 2010). In clinical practice, patients with mutations in this region who develop severe polyposis are usually advised to undergo a proctocolectomy (with or without an ileal pouch) as their primary operation.

There is currently no evidence that the site of APC mutation is associated with a difference in overall survival.

In the UK most patients with FAP are cared for in centres without in-house genetic medicine departments. The majority of FAP patients will have been investigated in a regional department of genetic medicine, however it is not usual for comprehensible
genotype-phenotype information to be relayed to the surgeons and gastrointestinal physicians who conduct the screening and surgery for these patients.

3.1.2 Aim of Study two

The aim of this study is to investigate whether the site of mutation (genotype: loci of APC mutation in FAP, and MutYH mutation in MutYH associated polyposis) could be used as a predictive and prognostic biomarker in familial polyposis as an aide in clinical management. This will be done by conducting a retrospective longitudinal study of patients on the Manchester Polyposis Registry. Patients will be grouped according to genotype (site of mutation) and a correlation between age at diagnosis of polyposis, age of onset of CRC and survival will be tested. No previous study has demonstrated a codon-specific survival difference. Comprehensible genotype information should be used to plan screening and surgery in patients with polyposis.

3.1.3 Hypothesis study two

It is hypothesised that genotype is associated with significant differences in age at diagnosis of polyposis, age of onset of CRC and overall survival.
3.2 Methods Study two: Genotype-phenotype correlation in colorectal polyposis

3.2.1 Participants

The Manchester Polyposis Registry was interrogated. Patients on the database are classified according to the loci of their APC or MutYH mutation (See table 24). Patients were entered onto the Filemaker Pro database if they were found to have over 100 colonic adenomas and/or an APC mutation. Thus the database and this study include families who were found to have MutYH Associated Polyposis (See table 24).

To investigate the correlation between genotype and age at diagnosis of polyposis, patients undergoing pre-symptomatic screening endoscopy were included (screen detected patients). Screening recommendations from the department of genetic medicine at Saint Mary’s Hospital is that annual screening (alternating colonoscopy and sigmoidoscopy) is commenced at age 12-14 years (in known APC mutation carriers and those at 50% risk). Biannual colonoscopy from age 25 years is recommended in homozygous carriers of MutYH mutations. The actual age at which screening starts will depend on the age of the patient at the time of family ascertainment. The screening intervals may vary depending on patient compliance and local availability of endoscopy services. At the time of this study, the polyposis registry did not contain details of the dates of endoscopies at which the finding was a normal colorectum. So compliance to screening is unknown. Although this may affect the findings and outcomes of this study it reflects current clinical practice. Therefore the findings of this study will reflect the outcomes in patients in an NHS service setting (Newton, Mallinson et al. 2011). Those with no known age at diagnosis of polyposis were excluded from the age at diagnosis part of the study. Those patients who remained free of polyps at the time of the study but had no known date of last follow up were also excluded, as polyp free time could not be predicted. To investigate the correlation between genotype and colorectal cancer incidence only patients living within the boundaries of the regional genetics centre were included because the accuracy of cancer incidence data on non-regional patients could not be verified using the North West Cancer Intelligence Service.
(NWCIS). Those with no known date of last follow-up, or death, or no mutation group classification were excluded. To investigate the correlation between genotype and survival all patients with a known date of death or date of last follow up were included. Those with no mutation group classification were excluded.

A further analysis was conducted to establish if any survival difference is diminished by the effects of colonic screening. Patients who underwent pre-symptomatic screening and those that did not were analysed separately. It has previously been shown, in this population, that pre-symptomatic screening increases actuarial survival in FAP from 57.8 years to 70.4 years (codon-specific survival was not investigated) (Mallinson, Newton et al. 2010). A further analysis was also conducted to establish if any survival difference is diminished by prophylactic surgery. Patients who underwent pre-symptomatic screening and had a known date of last follow up or death were included.

There is currently no consensus on the exact sites responsible for the different FAP phenotypes. The current evidence for genotype-phenotype correlation is from the APC gene mutation screening of polyposis kindreds, with defined phenotypes, known to regional or national registries. The number of kindreds investigated by each group is small (Nagase, Miyoshi et al. 1992; Spirio, Olschwang et al. 1993; Giardiello, Brensinger et al. 1997; Soravia, Berk et al. 1998). The following codon definitions were formulated:
<table>
<thead>
<tr>
<th>Groups</th>
<th>Genotype classification</th>
<th>Codons</th>
<th>Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unclassified mutation</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mild FAP (3’ mutations)</td>
<td>2</td>
<td>APC &gt;1550: mild (desmoids)</td>
<td></td>
</tr>
<tr>
<td>MutYH associated polyposis</td>
<td>5</td>
<td>MutYH: Chromosome 1 codons 34.3-32.1</td>
<td>Al-Tassan et al, (2002)</td>
</tr>
</tbody>
</table>
I have defined the upper bound for the severe phenotype as codon 1549. I have classified mutations at codon 1550 and within the region 3’ to this as the mild (desmoids) genotype. This decision was influenced by a previously described large very mild polyposis family with a codon 1557 mutation within the Manchester Registry population (Evans, Guy et al. 1993; Davies, Armstrong et al. 1995) and by the data which suggests that mutations in the region 3’ to codon 1550 have been shown to be associated with mild polyposis and desmoid disease (Evans, Hill et al. 1997; Speake, Evans et al. 2007). Although patients with mutations in this region have a later onset of polyposis in accordance with the attenuated phenotype, it is thought that the early onset of desmoids impacts on their overall survival. For the purpose of this study, the two mild genotypes were assessed separately.

### 3.2.2 Data analysis

The filemaker pro® database was examined. Missing data, for example causes of death, were indentified. Medical notes (St Mary’s Hospital genetic medicine notes) were reviewed and additional information sought where possible from local hospitals. The database was updated.
Data was downloaded from the *Filemaker Pro*® database into separate Excel spreadsheets for each analysis to be conducted. The data was coded numerically to allow analysis. For the age at diagnosis of polyps analysis, time to polyps (defined using date of endoscopy at which polyps were first found) was calculated from birth. In patients who have not yet had polyps found, date of last follow-up (or death) was used and the patient censored. Time to onset of colorectal cancer was calculated in the same way using date of diagnosis of cancer (defined either as date of endoscopic or pathological diagnosis) and date of birth. If the patient had not had a cancer at the time of the study, date of last follow-up or death was used and the patient censored. Overall survival was calculated from birth. If the patient remained alive at the time of the study, date of last follow-up was used and the patient was censored. Cause of death data was also included.

The data was then transferred into SPSS 13.0 (*IBM, UK*) and Kaplan-Meier analysis and log-rank (Mantel-Cox) tests were completed. Kaplan-Meier life tables calculated the mean and median survival estimates and from these I derived 1-KM which is the cumulative incidence (of diagnosis of polyposis, or CRC). SPSS estimates 95% confidence intervals of both the mean and the median times as a normal approximation (1.96*standard error) (Baker 2009).
3.3 Results Study two: Genotype-phenotype correlation in polyposis

At the time of the study the Manchester Polyposis Registry contained information on 492 individuals from 127 families. There were 230 patients undergoing presymptomatic endoscopic screening. There were not any suitable individuals in the MutYH mutation group (no MutYH mutation carriers known to the database were screened prior to the onset of symptoms); thus this group was eliminated from the age at diagnosis of polyps analysis. In order to assess the effect of genotype on the age at diagnosis of polyps, a sample size of 175 was used. (see Figure 20) In order to assess the effect of genotype on colorectal cancer incidence, a sample size of 378 was used. (see Figure 21) In order to assess the effect of genotype on survival, a sample size of 428 was used. (Figure 22)

Figure 20: Consort diagram of population used to assess the impact of genotype on the age at diagnosis of polyps
Figure 21: Consort diagram for population used to assess the impact of genotype on onset of CRC

- Assessed for eligibility n=492
  - Regional patients n=392
    - Excluded patients n=14
      - n=7 no known date of last follow up
      - n=7 no mutation group classification
  - Eligible patients n=378

Figure 22: Consort diagram for the population used to assess the impact of genotype on survival

- Assessed for eligibility n=492
  - Excluded patients n=64
    - n=59 no know date of last follow up or death
    - n=5 no mutation group classification
  - Eligible patients n=428
3.3.1. Onset of polyposis

Genotype 4 (APC, 1250-1549) had the youngest age at diagnosis of polyps (14.8 years), followed by genotype 3 (APC, 179-1249, excluding 312-412 of exon 9) (15.9 years) and genotype 0 (20.3 years). Genotypes 1 and 2 were diagnosed with polyps much later at 35.6 years (APC, 0-178, and 312-412) and 32.2 years (APC, >1550) (see Table 25).

10.9% of individuals did not have a recognised mutation.

There is a significant difference in the age at diagnosis of polyposis between the genotypes (log rank Mantel-cox p=0.00002). Genotypes 1 and 2 were diagnosed with polyps at a much later age. The age at diagnosis of polyps overlaps within genotypes 3, 4, and 0. The distinction between these genotypes is not as clear (see Table 25).

The classical genotype is the most common mutation group amongst this FAP population (41.71%).
Table 25: Population sample used for analysis of age at diagnosis of polyps

<table>
<thead>
<tr>
<th>Genotype 0</th>
<th>Genotype 1</th>
<th>Genotype 2</th>
<th>Genotype 3</th>
<th>Genotype 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown mutation site</td>
<td>APC 0-178, exon 9 (312-412)</td>
<td>APC &gt;1550</td>
<td>APC 179-1249 (excluding 312-412)</td>
<td>APC 1250-1549</td>
</tr>
<tr>
<td>Total Sample Number (n=175)</td>
<td>19 (10.86%)</td>
<td>26 (14.86%)</td>
<td>26 (14.86%)</td>
<td>73 (41.71%)</td>
</tr>
<tr>
<td>Median age at diagnosis of polyps (years)</td>
<td>20.32 (12.43-28.21)</td>
<td>35.61 (34.61-36.60)</td>
<td>32.19 (18.96-45.42)</td>
<td>15.91 (13.78-18.04)</td>
</tr>
<tr>
<td>Mean age at diagnosis of polyps (years)</td>
<td>19.58 (16.34-22.82)</td>
<td>37.42 (31.64-43.20)</td>
<td>29.13 (23.64-34.63)</td>
<td>21.52 (18.58-24.45)</td>
</tr>
</tbody>
</table>

*95% CI in parentheses

- *SPSS estimates 95% confidence intervals of both the mean and the median times as a normal approximation (1.96*standard error)(Baker 2009)
Table 26 demonstrates the probability of diagnosis of polyps by various ages for the different genotype groups. This includes the censored data. Genotypes 1 and 2 have a significant rise in the probability of diagnosis of polyps between the ages of 30 and 40. Genotypes 3 and 4 both have a greater than 60% probability of diagnosis of polyps by age 20. In the genotype 4 group the increase in probability after age 20 is rapid. There is a 100% probability of diagnosis of polyps by age 40. In the genotype 3 group the rate of increase is such that the probability of diagnosis of polyps is 100% by age 60.
Table 26: Table to show the probability of diagnosis of polyps by completion of each decade according to genotype group

*(derived from Kaplan Meier)*

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Genotype 0 Unknown mutation site</th>
<th>Genotype 1 APC 0-178, exon 9 (312-412)</th>
<th>Genotype 2 APC &gt;1550</th>
<th>Genotype 3 APC 179-1249 (excluding 312-438)</th>
<th>Genotype 4 APC 1250-1549</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>20</td>
<td>0.56</td>
<td>0.16</td>
<td>0.34</td>
<td>0.64</td>
<td>0.61</td>
</tr>
<tr>
<td>30</td>
<td>0.83</td>
<td>0.29</td>
<td>0.48</td>
<td>0.79</td>
<td>0.78</td>
</tr>
<tr>
<td>40</td>
<td>1.0</td>
<td>0.622</td>
<td>0.86</td>
<td>0.89</td>
<td>1.00</td>
</tr>
<tr>
<td>50</td>
<td>1.0</td>
<td>0.788</td>
<td>0.91</td>
<td>0.95</td>
<td>1.00</td>
</tr>
<tr>
<td>60</td>
<td>1.0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Proportion of censored data</strong></td>
<td><strong>26.3%</strong></td>
<td><strong>26.9%</strong></td>
<td><strong>15.4%</strong></td>
<td><strong>8.2%</strong></td>
<td><strong>3.2%</strong></td>
</tr>
</tbody>
</table>
3.3.2 Colorectal cancer Incidence

Table 27 describes the incidence of CRC according to genotype, and the median and mean age of onset of first CRC according to genotype. The Kaplan-Meier graph below (see Figure 24) demonstrates the similarity of CRC incidence between the genotype groups at various ages (Log rank Mantel-cox p=0.215). No significant difference was found in CRC cancer incidence between males and females, either when comparing all groups (p=0.285) or when looking at the effect of gender on CRC incidence within each genotype (p=0.088 for unknown mutation genotype, p=0.432 for mild early genotype, p=0.226 for mild desmoids genotype, p=0.303 for classical genotype, p=0.226 for severe genotype, and p=0.670 for MutYH associated polyposis).
Table 27: Colorectal cancer incidence according to genotype category

<table>
<thead>
<tr>
<th></th>
<th>Genotype 0 Unknown mutation site</th>
<th>Genotype 1 APC 0-178, exon 9 (312-412)</th>
<th>Genotype 2 APC &gt;1550</th>
<th>Genotype 3 APC 179-1249 (excluding 312-412)</th>
<th>Genotype 4 APC 1250-1549</th>
<th>MutYH associated polyposis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Sample Number</strong> n=378</td>
<td>73 (19.3%)</td>
<td>45 (11.9%)</td>
<td>44 (11.6%)</td>
<td>151 (40.0%)</td>
<td>55 (14.6%)</td>
<td>10 (2.6%)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td>F 27 M 45</td>
<td>F 23 M 22</td>
<td>F 24 M 20</td>
<td>F 88 M63</td>
<td>F 24 M 41</td>
<td>F 2 M 8</td>
</tr>
<tr>
<td><strong>Number of Dead Patients</strong></td>
<td>30 (41.1%)</td>
<td>10 (22.2%)</td>
<td>19 (43.2%)</td>
<td>37 (24.5%)</td>
<td>19 (34.6%)</td>
<td>3 (30.0%)</td>
</tr>
<tr>
<td><strong>CRC Incidence</strong></td>
<td>29/73 (39.7%)</td>
<td>10/45 (22.2%)</td>
<td>13/44 (29.6%)</td>
<td>36/151 (23.8%)</td>
<td>13/55 (23.6%)</td>
<td>5/10 (50.0%)</td>
</tr>
<tr>
<td><strong>Female: male ratio</strong></td>
<td>1:1.4</td>
<td>1:0.8</td>
<td>1:0.7</td>
<td>1:1.9</td>
<td>1:1.4</td>
<td>1:1.1</td>
</tr>
<tr>
<td><strong>Age of onset of earliest cancer within group</strong> (non-censored data)</td>
<td>21.1 years</td>
<td>32.4 years</td>
<td>25.5 years</td>
<td>23.5 years</td>
<td>25.5 years</td>
<td>34. years</td>
</tr>
<tr>
<td><strong>Median age of onset of first CRC</strong> (includes censored data)</td>
<td>53.7 years</td>
<td>65.0 years</td>
<td>63.3 years</td>
<td>61.3 years</td>
<td>52.9 years</td>
<td>55.2 years</td>
</tr>
<tr>
<td><strong>Mean age of onset of first CRC</strong> (includes censored data)</td>
<td>53.0 years</td>
<td>62.5 years</td>
<td>61.0 years</td>
<td>60.0 years</td>
<td>52.9 years</td>
<td>53.3 years</td>
</tr>
</tbody>
</table>
Figure 24: Kaplan-Meier curve demonstrating probability of developing first CRC according to genotype

Log-rank (Mantel-cox)
Chi square= 7.083
Df= 5
P=0.215
Table 28: Table to show the probability of the developing a first CRC by completion of each decade according to genotype group (derived from Kaplan Meier thus includes censored data)

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Genotype 0 Unknown mutation site</th>
<th>Genotype 1 APC 0-178, exon 9 (312-412)</th>
<th>Genotype 2 APC &gt;1550</th>
<th>Genotype 3 APC 179-1249 (excluding 312-412)</th>
<th>Genotype 4 APC 1250-1549</th>
<th>MutYH associated polyposis</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>0.125</td>
<td>0</td>
<td>0.029</td>
<td>0.025</td>
<td>0.048</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>0.3</td>
<td>0.098</td>
<td>0.063</td>
<td>0.207</td>
<td>0.269</td>
<td>0.125</td>
</tr>
<tr>
<td>50</td>
<td>0.404</td>
<td>0.178</td>
<td>0.225</td>
<td>0.288</td>
<td>0.269</td>
<td>0.4</td>
</tr>
<tr>
<td>60</td>
<td>0.598</td>
<td>0.281</td>
<td>0.389</td>
<td>0.433</td>
<td>0.415</td>
<td>0.8</td>
</tr>
<tr>
<td>70</td>
<td>0.598</td>
<td>0.743</td>
<td>0.782</td>
<td>0.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of censored data</td>
<td>59.7%</td>
<td>75.6</td>
<td>70.5%</td>
<td>76.3%</td>
<td>80.0%</td>
<td>50%</td>
</tr>
</tbody>
</table>
This data suggests a similar incidence and age of onset of CRC between the six genotype groups. However the occurrence of, and age at which prophylactic surgery was performed may have been different between the groups so may have affected CRC incidence.

### 3.3.3 Survival

This study has demonstrated a significant difference in overall survival between all the FAP genotypes (Log rank Mantelcox p=0.003) (see figure 25). When comparing the genotypes separately there was a significant survival difference between genotype 1 (APC 0-178, and 312-412), and genotype 4 (APC 1250-1549) and between the 3 genotype 3 (APC 179-1249, excluding 312-412) and genotype 4 (APC 1250-1549). Median survival in patients with 5’ APC mutations from codons 0-178 and 312-412 was 26.8 years greater than patients with APC mutations from codons 1250-1549. (p=0.0003). Median survival in patients with APC mutations from codons 179-1249 (excluding 312-412) was 14.9 years greater than those with APC mutations from codons 1250-1549. (p=0.001). Median survival in genotype 1 (APC 0-178 and 312-412) is 13.1 years greater than of genotype 2 (APC >1550) group which is approaching significance (p=0.088). There were 21 deaths in the genotype 2 group (APC, >1550) (40.4%), and 4 (19.0% of the deaths) were attributed to desmoid disease. (see table 29)

The proportion of deaths due to CRC, FAP related tumours, non-FAP related causes, and of unknown cause are similar between the genotypes. (see table 29)
Table 29: Survival according to genotype category

<table>
<thead>
<tr>
<th>Genotype 0</th>
<th>Genotype 1</th>
<th>Genotype 2</th>
<th>Genotype 3</th>
<th>Genotype 4</th>
<th>MutYH associated polyposis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown mutation site</td>
<td>APC 0-178, exon 9 (312-412)</td>
<td>APC &gt;1550</td>
<td>APC 179-1249 (excluding 312-412)</td>
<td>APC 1250-1549</td>
<td></td>
</tr>
<tr>
<td>Total sample number (n=428)</td>
<td>80 (18.69%)</td>
<td>49 (11.45%)</td>
<td>52 (12.15%)</td>
<td>175 (40.89%)</td>
<td>60 (14.02%)</td>
</tr>
<tr>
<td>Number of dead patients</td>
<td>31 (38.75%)</td>
<td>12 (24.49%)</td>
<td>21 (40.38%)</td>
<td>54 (30.86%)</td>
<td>22 (36.67%)</td>
</tr>
<tr>
<td>Censored data</td>
<td>49 (61.25%)</td>
<td>37 (75.5%)</td>
<td>31 (59.62%)</td>
<td>121 (69.14%)</td>
<td>38 (45.0%)</td>
</tr>
<tr>
<td>Cause of death CRC</td>
<td>20 (64.51%)</td>
<td>6 (50%)</td>
<td>9 (42.90%)</td>
<td>34 (62.96%)</td>
<td>12 (54.54%)</td>
</tr>
<tr>
<td>Cause of death extra-colonic FAP related tumour</td>
<td>2 (6.45%)</td>
<td>3 (25%)</td>
<td>5 (23.81%)</td>
<td>5 (9.26%)</td>
<td>5 (22.73%)</td>
</tr>
<tr>
<td>Cause of death unrelated</td>
<td>3 (9.68%)</td>
<td>3 (25%)</td>
<td>5 (23.81%)</td>
<td>9 (16.67%)</td>
<td>3 (13.67%)</td>
</tr>
<tr>
<td>Cause of death unknown</td>
<td>5 (16.13%)</td>
<td>0</td>
<td>2 (9.52%)</td>
<td>6 (11.11%)</td>
<td>2 (9.09%)</td>
</tr>
<tr>
<td>Median survival (years) 95% CI in parentheses</td>
<td>56.61 (44.26-68.97)</td>
<td>74.90 (59.02-90.78)</td>
<td>60.98 (49.95-72.01)</td>
<td>63.04 (56.33-69.76)</td>
<td>48.14 (40.63-55.64)</td>
</tr>
<tr>
<td>Mean survival (years) 95% CI in parentheses</td>
<td>57.10 (51.90-62.30)</td>
<td>66.46 (60.66-72.25)</td>
<td>58.37 (52.37-64.37)</td>
<td>61.06 (57.20-64.92)</td>
<td>58.37 (44.48-52.51)</td>
</tr>
</tbody>
</table>

3 FAP related tumour here includes all extra-colonic tumours known to be associated with FAP: these are all upper gastrointestinal cancers and desmoid tumours
Figure 25: Kaplan-Meier curve demonstrating difference in survival according to genotype.
Table 30: Comparison of survival by individual genotype category

<table>
<thead>
<tr>
<th>Group Comparison</th>
<th>Median survival (years)</th>
<th>Survival difference (years)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype 1 and Genotype 3</td>
<td>74.1 and 63.0</td>
<td>13.8</td>
<td>p=0.156</td>
</tr>
<tr>
<td>Genotype 1 and Genotype 4</td>
<td>74.1 and 48.1</td>
<td>26.8</td>
<td>p=0.0003</td>
</tr>
<tr>
<td>Genotype 3 and Genotype 4</td>
<td>63.0 and 48.1</td>
<td>14.9</td>
<td>p=0.001</td>
</tr>
<tr>
<td>Genotype 1 and Genotype 2</td>
<td>74.1 and 61.0</td>
<td>13.1</td>
<td>P=0.088</td>
</tr>
</tbody>
</table>

The Bonferroni correction describes that significance should be tested at $p < \frac{1}{n} \times 0.05$ where $n$ is the number of hypotheses tested from the one dataset. Here six hypotheses are tested (1 compared to 2, 1 compared to 3, 1 compared to 4, 2 compared to 3, 2 compared to 4, 3 compared to 4) therefore a $p$ value of $1/6 \times 0.05= 0.0083$.

Analysis of survival data taking into account the effect of surveillance demonstrates that the difference in overall survival (both median and mean as estimated by Kaplan-meier) remains significant when the effect of pre-symptomatic screening is removed. (log rank Mantel-cox $p=0.004$). Median survival of individuals with an APC mutation 1250-1549 is 19.8 years less than those with an APC mutation 0-178 and 312-412, and 11.6 years less those with a mutation 179-1249 (excluding 312-314) when not screened before the onset of symptoms. (Table 31)
Table 31: Survival in patients who did not undergo pre-symptomatic screening  n= 211 (95% confidence intervals in parentheses)

<table>
<thead>
<tr>
<th>Genotype 0</th>
<th>Genotype 1</th>
<th>Genotype 2</th>
<th>Genotype 3</th>
<th>Genotype 4</th>
<th>5: MutYH associated polyposis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown mutation site</td>
<td>APC 0-178, exon 9 (312-412)</td>
<td>APC &gt;1550</td>
<td>APC 179-1249 (excluding 312-412)</td>
<td>APC 1250-1549</td>
<td></td>
</tr>
<tr>
<td>Deaths</td>
<td>28/53 (52.8%)</td>
<td>10/15 (66.7%)</td>
<td>17/23 (73.9%)</td>
<td>47/81 (58.0%)</td>
<td>19/28 (67.9%)</td>
</tr>
<tr>
<td>Causes</td>
<td>20 CRC</td>
<td>4 CRC</td>
<td>10 CRC</td>
<td>25 CRC</td>
<td>7CRC</td>
</tr>
<tr>
<td>1 non-FAP</td>
<td>1 duodenal ca</td>
<td>1 Desmoid</td>
<td>2 duodenal ca</td>
<td>1 pancreas ca</td>
<td>2 gastric ca</td>
</tr>
<tr>
<td>6 Unknown</td>
<td>1 pancreas ca</td>
<td>1 hepatic</td>
<td>2 gastric</td>
<td>1 duodenal ca</td>
<td>1 unknown</td>
</tr>
<tr>
<td>1 non FAP</td>
<td>4 non FAP</td>
<td>7 non FAP</td>
<td>1 bony tumour</td>
<td>4 Non FAP</td>
<td></td>
</tr>
<tr>
<td>3 unknown</td>
<td>1 unknown</td>
<td>9 Unknown</td>
<td>4 unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median survival in years derived from Kaplan-Meier</td>
<td>56.6 (46.3-66.8)</td>
<td>66.8 (43.4-90.0)</td>
<td>60.9 (46.9-90.0)</td>
<td>58.6 (53.8-63.4)</td>
<td>47.0 (40.2-53.9)</td>
</tr>
</tbody>
</table>
Figure 26: Kaplan-Meier curve demonstrating survival in patients who did not undergo pre-symptomatic screening.
Median overall survival of individuals with an APC mutation 1250-1549 is 18.8 years less than those with an APC mutation 0-178 and 312-412, and 19.7 years less those with a mutation 179-1249 (excluding 312-314) when screened before the onset of symptoms. (See Table 32, Figure 27) The trend is for worse overall survival in genotype 4 compared to all other groups. This did not reach significance (p=0.132). Interestingly there were two deaths at a young age in the screening population. These were both caused by large mesenteric desmoids which developed following prophylactic colectomy (patient A aged 20.7 with genotype 2, and patient B aged 24.8 with genotype 3). There have been no deaths in screened patients with MutYH mutations.
Table 32: Survival in patients who underwent screening n= 197

<table>
<thead>
<tr>
<th></th>
<th>Genotype 1</th>
<th>Genotype 2</th>
<th>Genotype 3</th>
<th>Genotype 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>*APC 0-178, exon 9 (312-412)</td>
<td>*APC &gt;1550</td>
<td>*APC 179-1249 (excluding 312-412)</td>
<td>*APC 1250-1549</td>
</tr>
<tr>
<td>Deaths</td>
<td>2/34 (5.9%)</td>
<td>4/38 (10.5%)</td>
<td>3/92 (3.3%)</td>
<td>4/33 (12.1%)</td>
</tr>
<tr>
<td>Cause (age in years)</td>
<td>Medullary thyroid ca (44.2)</td>
<td>Desmoid (20.7)</td>
<td>Desmoid (24.8)</td>
<td>Rectal ca (41.7)</td>
</tr>
<tr>
<td></td>
<td>Non FAP (49.9)</td>
<td>Desmoid (42.4)</td>
<td>CRC (37.0)</td>
<td>Rectal ca (45.8)</td>
</tr>
<tr>
<td></td>
<td>CRC (48.5)</td>
<td></td>
<td>Rectal ca (54.6)</td>
<td>CRC (27.1)</td>
</tr>
<tr>
<td></td>
<td>*1PE (57.9)</td>
<td></td>
<td></td>
<td>Non FAP (48.1)</td>
</tr>
<tr>
<td>Mean survival (years)</td>
<td>72.0 *(66.7-78.3)</td>
<td>67.4 *(56.8-78.0)</td>
<td>72.9 *(67.6-78.1)</td>
<td>53.2 *(48.2-58.1)</td>
</tr>
<tr>
<td></td>
<td>derived from Kaplan Meier</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* 95% confidence intervals

*1Pulmonary embolus
When analysing survival in FAP patients undergoing screening, age at surgery and time to surgery, I excluded the group with unknown mutation loci as I was only interested in comparing those patient groups with known genotype.
Table 33: Age at surgery and median time to surgery according to genotype

<table>
<thead>
<tr>
<th>Genotype 1</th>
<th>Genotype 2</th>
<th>Genotype 3</th>
<th>Genotype 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>APC 0-178, exon 9 (312-412)</em></td>
<td><em>APC &gt;1550</em></td>
<td><em>APC 179-1249 (excluding 312-412)</em></td>
<td><em>APC 1250-1549</em></td>
</tr>
<tr>
<td>Operated 15/34 (44.2%)</td>
<td>Operated 11/38 (28.9%)</td>
<td>Operated 48/93(51.6%)</td>
<td>Operated 24/33 (72.7%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operation</th>
<th>Col*¹</th>
<th>IRA*²</th>
<th>PPC*³</th>
<th>Col</th>
<th>IRA</th>
<th>PPC</th>
<th>Col</th>
<th>IRA</th>
<th>PPC</th>
<th>Col</th>
<th>IRA</th>
<th>PPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>1</td>
<td>9</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>19</td>
<td>23</td>
<td>6</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>(2.9%)</td>
<td>(26.5%)</td>
<td>(14.7%)</td>
<td>(5.3%)</td>
<td>(5.3%)</td>
<td>(18.4%)</td>
<td>(6.5%)</td>
<td>(20.4%)</td>
<td>(24.7%)</td>
<td>(9.1%)</td>
<td>(18.2%)</td>
<td>(45.5%)</td>
</tr>
<tr>
<td>Median age at surgery (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non censored data only</td>
<td>36.0</td>
<td>37.0</td>
<td>38.0</td>
<td>41.5</td>
<td>40.0</td>
<td>22.5</td>
<td>16.0</td>
<td>19.3</td>
<td>21.5</td>
<td>35.0</td>
<td>16.5</td>
<td>17.0</td>
</tr>
<tr>
<td>Median time to surgery derived from KM</td>
<td>40.0 years</td>
<td>43.0 years</td>
<td>27.1 years</td>
<td>21.0 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*⁴ (34.4-45.7)</td>
<td>*⁴ (39.4-45.7)</td>
<td>*⁴ (20.4-33.6)</td>
<td>*⁴ (17.2-24.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*¹ Col= colectomy (unspecified) *² IRA= total colectomy and ileorectal anastomosis. *³ PPC= panproctocolectomy
*⁴ 95% confidence intervals
In the screened population, time from birth to first surgery was significantly reduced in genotypes 4 and 3, compared to genotypes 1 and 2. (p<0.0001) All surgery was prophylactic except two panproctocolectomies performed for CRC diagnosed on first screen (one patient genotype 3, one patient genotype 4). When screened, FAP patients with an APC mutation 1250-1549 undergo colorectal resection 19.0 years earlier than patients with an APC mutation 0-178 and 312-412, and 6.1 years earlier than patients with an APC mutation 179-1249 (excluding 312-412). (see Figure 28)

Figure 28: Kaplan-Meier curve demonstrating time (from birth) to surgery according to genotype

![Kaplan-Meier curve](image)
3.4 Discussion Study two: Genotype-phenotype correlation in polyposis

3.4.1 Discussion

APC germline mutation analysis is being used effectively for predictive testing in FAP families. When the causative mutation is known within a family, predictive testing can accurately define the mutation carriers who need regular endoscopic screening. Genetic analysis is not recommended in children until the child is able to understand the consequences of the result. This is generally agreed to be around the age of 11 years (Hyer and Fell 2001).

Knowledge of a patient’s genotype-phenotype should routinely be taken into consideration when making decisions regarding timing and nature of prophylactic surgery. Comprehensible genotype information should always be provided to the clinicians performing the endoscopy and surgery, and these clinicians need to understand the implication of genotype-phenotype correlations.

This study has demonstrated that patients with genotype 4, the severe genotype, have a significantly worse overall survival than the other genotypes. A similar proportion of deaths amongst the different genotypes were due to CRC or another FAP related tumour (genotype 1: 75% of deaths, genotype 2: 66.67%, genotype 3: 72%, genotype 4: 77%). Thus it can be said that this severe genotype group are dying prematurely from FAP related causes. These deaths should be preventable. This group should be more pro-actively followed up to ensure that timely screening and prophylactic surgery occur.

It has been consistently demonstrated that there is a correlation between the site of the APC mutation and the severity of colonic polyposis in FAP (Spirio, Otterud et al. 1992; Giardiello, Petersen et al. 1997; Soravia, Berk et al. 1998; Friedl, Caspari et al. 2001). This study provides further evidence for the correlation between the attenuated or mild FAP phenotype and mutations from codon 0-178, exon 9 (312-438), and 3’ (distal) to1550. The mean age at onset of polyps was 34.25 years in
this mild group, genotype 1 (95% CI 31.29-37.25). This finding is in agreement with previous studies (Spirio, Otterud et al. 1992; Evans, Hill et al. 1997; Giardiello, Petersen et al. 1997; Soravia, Berk et al. 1998). Patients with a mutation located at codon 1557 (mild group) have been noted to remain free of polyps as late as "33, 36, 39 and 54 years" (Evans, Hill et al. 1997). Most other studies have included patients that presented with symptoms as well as those that were diagnosed on screening to assess the correlation between genotype and phenotype. Patients that present symptomatically will usually present later (Nagase, Miyoshi et al. 1992; Spirio, Otterud et al. 1992; Lynch, Smyrk et al. 1995; Soravia, Berk et al. 1998; Friedl, Caspari et al. 2001; Kanter-Smoler, Fritzell et al. 2008). This may result in ascertainment bias as patients with attenuated FAP tend to develop symptoms later, therefore being ascertained at an older age. As the family ascertainment is later their children will tend to have their first screen at an older age also creating bias. This study only included screen detected patients thus the likelihood of ascertainment bias has been reduced.

This study has demonstrated that those with mutations in the APC gene codon region 1250-1549 (genotype 4) have a severe phenotype in terms of earlier onset of polyposis than the mild (genotypes 1 and 2) genotypes. These individuals also have a significantly reduced overall survival compared to both the mild (genotype 1: 26.76 years difference) and classical genotypes (genotype 3:14.90 years difference). Kanter-Smoler showed that mutations between codons 1250 and 1464 predicted a severe phenotype with a median age of onset of symptoms of 21 compared to that of 34 for other mutation loci (Kanter-Smoler, Fritzell et al. 2008). Further studies have demonstrated an association between APC mutations between codons 1250 and 1311 and the development of over 5000 colonic polyps (Nagase, Miyoshi et al. 1992; Enomoto, Konishi et al. 2000; Friedl, Caspari et al. 2001).

Studies have also demonstrated that mutations at codon 1309 of the APC gene are associated with a severe FAP phenotype, with an age of onset of symptoms of ten years younger than those with mutations at other sites (Gebert, Dupon et al. 1999; Ficari, Cama et al. 2000; Friedl, Caspari et al. 2001). Bertario et al and Sinha et al both found APC mutations between codon 1250 and 1464 to be an independent predictor of progressive rectal disease following a colectomy and ileorectal anastomosis (IRA) (Bertario, Russo et al. 2000; Sinha, Tekkis et al. 2010). Sinha et al found a hazard ratio of 3.91 (p=0.007) for the development of rectal cancer or severe rectal polyposis necessitating proctectomy (Sinha, Tekkis et al. 2010).
Bertario et al found a hazard ratio of 6.2 (p=0.003) for the development of rectal cancer (Bertario, Russo et al. 2000).

Although there are now well documented statistically significant associations between FAP genotype and phenotype, it remains unclear as to whether this information should guide clinical decision making. It has previously been suggested that genotype could guide screening protocols, and timing and nature of prophylactic surgery. However, the most recently published recommendations from the British Society of Gastroenterologists do not included any alteration in screening regimen or surgical treatment according to genotype (Cairns, Scholefield et al. 2010). Vasen et al published guidelines in 2008 regarding the management of FAP. For classical FAP it is recommended that biannual sigmoidoscopy be commenced at age 10-12 until polyps develop and at that point biannual or annual colonoscopy should be performed dependant on the polyp load. In these guidelines, Vasen et al recommend a different endoscopic screening protocol for patients from attenuated FAP families. Biannual colonoscopy commencing at age 18-20 is suggested (Vasen, Moslein et al. 2008). The department of genetic medicine at St Mary’s Hospital Manchester recommends an initial screen in all at risk patients, or known mutation carriers, at age 10-12, or slightly later at 12-14 years in an attenuated FAP family. The regularity is then reduced if the patient is known to have an attenuated/mild genotype and is thought to have an attenuated/mild phenotype. It is felt that this is prudent given the known risk of intra-familial phenotypic variation, and preserved cancer risk in AFAP. This practice also enables the child to become confident with undergoing colonic examination and aids the smooth transition to the adult service. If patients within an attenuated FAP family are screened at age 18-20 as recommended in the 2008 guidelines, it is likely that some patients will be diagnosed with significant polyposis necessitating surgery at their first endoscopy. This is likely to engender a state of fear and concern in these AFAP families and may damage relationships between the family and clinicians.

Vasen et al noted in 1996 that patients with mutations 3’ to codon 1250 had a higher rate of rectal cancer than those with mutations 5’ to codon 1250. This led to the recommendation that genotyping should inform the decision regarding whether to undertake a restorative proctocolectomy or colectomy and ileorectal anastomosis (IRA) in these patients (Vasen, van der Luijt et al. 1996). This recommendation was latterly withdrawn by the authors as some patients with mutations 3’ to 1250 were
also found to have a mild polyposis phenotype (Vasen, van der Luijt et al. 1996; Evans, Hill et al. 1997).

Intra-familial phenotypic variation is well known. This may be due to other modifier genes, epigenetic factors, other endogenous or exogenous factors, and/or the varying “second hit” mutation. In this study 52 patients were classified as mild genotype (genotype 1) according to the site of their germline APC gene mutation. As a group, these patients were found to have a significantly later onset of polyposis, by greater than ten years, compared to the classic (genotype 3 and severe polyposis (genotype 4) groups. However, within this attenuated FAP group, there are some individuals (three patients in the group used in the onset of polyposis study) who developed polyposis at a much younger age and required a prophylactic colectomy in their late teens or early twenties more characteristic of classical FAP.

Patient A developed osteomas and multiple sebaceous cysts as a child. Endoscopic screening revealed multiple colonic polyps at age 18 and a colectomy and ileorectal anastomosis was performed at age 19. Further members of this kindred, with the same APC mutation (codon 1557) have the attenuated phenotype. Patient B, again with a mutation at codon 1557 was screened as a child following the symptomatic diagnosis of FAP in a parent aged 32. Significant polyposis was noted during the late teens and a restorative proctocolectomy was performed at age 20. A further kindred with 12 known mutation carriers, has three members who developed significant polyposis during their late teens. Patient C underwent a colectomy and ileorectal anastomosis aged 17. The other mutation carriers in this family have the attenuated phenotype. The family mutation is at codon 172.

Mutations in the region 3’ to codon 1550 are known to be associated with increased incidence of desmoid disease. Desmoid disease also occurs on patients with mutations outside this region. This study has demonstrated that these patients have a later median age of onset of polyposis than classical FAP (32.19 years compared to 15.91 years). However, there is a suggestion that they might have a reduced survival compared to the mild early genotype (genotype 1) (60.98 years compared to 74.09 years, this is nearing significance p=0.088). This may be attributed to young onset desmoid disease. Within this group there were four deaths caused by
desmoids. However, death from desmoids has not been specifically investigated within this study, so no real conclusion can be made. It has previously been shown that patients with mutations in the desmoid region of the APC gene have an 85% cumulative risk of developing desmoids if they undergo surgery, and 10% cumulative risk if they do not. Surgery alone was the most useful predictive factor for the development of desmoids (p<0.001) (Speake, Evans et al. 2007).

Knowledge of genotype, may guide surgical decision making in that colectomy may be delayed and aggressive endoscopic polypectomy used as the treatment of choice in patients with mild to moderate polyposis. If these patients present with cancer, heavy polyp burden or polyp progression despite endoscopic management, surgery is indicated. The knowledge of potential post-operative desmoid development should accompany this surgery, and the risks thoroughly discussed with the patient.

### 3.4.2 Conclusion

It is known that pre-symptomatic screening and the use of a polyposis registry improves prognosis in FAP patients. It has been demonstrated that when screened before the onset of symptoms, FAP patients have a reduced incidence of CRC and improved survival. It has also been demonstrated that the use of a polyposis registry to coordinate patient management improves survival. Although genotyping may be used to inform decisions regarding screening and surgery, timely pre-symptomatic screening remains likely to be the most important factor effecting outcome.

The decision to perform prophylactic surgery is most influenced by the confidence with which the endoscopist can plan further screening without the risk of an interval cancer. This decision is also guided by the age of the patient, family history of cancer, and personal and family disease experience in terms of stomas, desmoids and operative complications. Genotype should be used as an adjunct to these other factors.
3.4.3 Further work

This study has demonstrated that patients with mutations in the region \textit{APC} 1249-1549 have a reduced life-expectancy compared to other FAP patients. Previous studies have demonstrated that these patients have an increased risk of rectal failure (polyp load or cancer requiring proctectomy). It has been suggested that these patients should undergo a total proctocolectomy (with or without an ileal pouch formation) as their primary prophylactic operation. A large study investigating post-operative outcomes in these patients is required to definitively inform clinicians regarding the appropriateness of more radical primary surgery. It is hoped that a multi-centre European collaboration will be undertaken to answer this question.
4 Study three: Metachronous colorectal cancer risk in patients with a moderate family history

4.1 Introduction

4.1.1 Background

Metachronous colorectal cancer is defined as a new occurrence of cancer in a different anatomical part of the colorectum (therefore not local recurrence), more than 12 months after the first and assumes adequate visualisation of the whole large bowel at the time of diagnosis of the first cancer.

The incidence of metachronous sporadic colorectal cancer has been reported at between 0.6% and 3% with a follow-up period of between 22 months and 22 years (Nelson, Persky et al. 1999; Balleste, Bessa et al. 2007; Pramateftakis, Hatzigianni et al. 2010). A persistently elevated risk of second CRC in the over 75 age group has been reported (Kahi, Azzouz et al. 2007). The US Cancer Society and the US multi-society task force on colorectal cancer published guidelines on colonoscopy surveillance after colorectal resection. The group reviewed a large number of studies, 23 of which included only patients who had adequate peri-operative removal of neoplastic disease. This group found the life-time incidence of a metachronous cancer to be 1.52% (Rex, Kahi et al. 2006). Hollington et al conducted a retrospective audit of 569 patients undergoing surveillance colonoscopy following the curative resection of a CRC. Patients with familial syndromes were excluded. The surveillance regimen was generally one colonoscopy at 12 months following the resection, and three yearly thereafter. This group found the rate of sporadic mCRC to be 2.5% (14/569) with a median time to detection of 60 months (Hollington, Tiong et al. 2011).

It is known from case series and case-controlled series evidence that patients with Lynch syndrome have a significant risk of developing metachronous colorectal cancer (de Vos tot Nederveen Cappel, Nagengast et al. 2002; Kalady, McGannon et al. 2010; Natarajan, Watson et al. 2010). This risk is substantially higher after a
segmental resection (life-time incidence 15-26%) than a subtotal colectomy and ileorectal anastomosis (life-time incidence 3.4%) (de Vos tot Nederveen Cappel, Nagengast et al. 2002; Kalady, McGannon et al. 2010; Natarajan, Watson et al. 2010). Endoscopic surveillance of any remaining colon is recommended following a CRC resection. In 2011 Parry et al published a retrospective cohort study of 382 carriers of pathogenic MLH1, MSH2, MSH6 and PMS2 mutations who had a CRC resected. Using Kaplan-Meier risk estimate this group found that in those 332 patients who had a segmental resection, the risk of a metachronous CRC was 16% (95% confidence interval 10-25%) at 10 years, 41% (95% CI 30-52%) at 20 years and 62% (95% CI 50-77%) at 30 years (Parry, Win et al. 2011).

The British Society of Gastroenterologists describe familial risk as “a heterogeneous composite of high penetrance single gene disorders, multiple low penetrance genetic factors, and shared environmental exposure”(Cairns, Scholefield et al. 2010). Two groups of patients at moderate risk of developing CRC due to their family history are described, as mentioned in section 1.6.3 (see table 3 which is repeated below). These are patients whose family history is not suggestive of a high-risk single gene disorder, but is significant enough to warrant colonoscopy surveillance (see Table 3, life-time risk of CRC of 1:6-10 to 1:12). This moderately elevated inherited risk is thought to be due to combinations of low risk, low penetrance gene mutations, or variants. In the general population genome wide association studies (GWAS) have thus far identified ten common variant alleles. It is thought that these alleles each contribute roughly 1% of all heritable cancers, and confer an increased risk of CRC of between 1.2 and 1.47 times that of the normal population (Houlston, Webb et al. 2008).
Table 3: Moderate risk CRC categories (Cairns, Scholefield et al. 2010)

<table>
<thead>
<tr>
<th>Family History</th>
<th>Life-time risk of CRC</th>
<th>Risk category</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRC in 3 FDRs*s (1 in first degree kinship with consultand), none &lt; age 50 years</td>
<td>1:6-10</td>
<td>High moderate</td>
</tr>
<tr>
<td>CRC in 2 FDRs (1 in first degree kinship with consultand) with a mean age of &lt;60 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRC in 2 FDRs (1 in first degree kinship with consultand) aged &gt;60 years</td>
<td>1:12</td>
<td>Low moderate</td>
</tr>
<tr>
<td>CRC in 1 FDR aged &lt;50 years</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*FDR: first degree relative

In the UK follow up is standard practice for all patients who have had a resection for a colorectal cancer (CRC). The purpose of this follow-up is two-fold. Firstly to detect potentially curable locally recurrent and/or metastatic disease and secondly to detect treatable metachronous CRC (Cairns, Scholefield et al. 2010). The evidence supports clinical follow-up but is not able to clarify whether minimal follow-up or intensive follow-up is better (Edelman, Meyers et al. 1997; Richard and McLeod 1997; Jeffery, Hickey et al. 2002; Renehan, Egger et al. 2002). The 2010 British Society of Gastroenterologists (BSG) guidelines on the surveillance of high risk groups recommended that a colonoscopy is performed five years after surgery and every five years thereafter until the risk outweighs the benefit (grade B recommendation) (Cairns, Scholefield et al. 2010). The evidence suggests that although continued post-operative colonic surveillance does not improve survival, some treatable metachronous cancers are found. In order to plan appropriate follow-up for an individual, it is important to know that individual’s personal risk of developing a metachronous CRC.

It is unclear what influence a moderate risk family history of CRC has on the incidence of metachronous CRC. Studies investigating metachronous sporadic cancer have excluded patients with FAP or Lynch Syndrome/HNPCC, but do not describe whether those patients with a moderate family history were included. The
risk of metachronous colorectal cancer in this group is unknown. There is no data to inform post-operative colonic surveillance.

### 4.1.2 Aim of study three

The aim of study three is to investigate the incidence of metachronous CRC in a group of patients with a moderate family history of CRC in order to inform screening guidelines. This has not been done previously. The incidence of metachronous CRC in patients with a moderate familial risk of CRC, those with Lynch Syndrome, and those with population risk will be compared.

### 4.1.3 Hypothesis of study three

It is hypothesised that patients with a moderate familial risk of colorectal cancer, are at increased risk of a metachronous cancer compared to the normal population. This risk is likely to be less than in patients with Lynch syndrome.
4.2 Methods Study three: Metachronous colorectal cancer risk in patients with a moderate family history

4.2.1 Participants

The Manchester Familial Colorectal Cancer Registry was used to identify two groups of patients at increased risk of CRC.

- Moderate familial risk (Cairns, Scholefield et al. 2010)
- Lynch Syndrome

Patients on the database with a family history not indicative of increased risk were also identified (any family history less than that required for moderate risk). This group of patients was considered to be at background population risk.

The database was interrogated for all patients who have had at least one colorectal cancer. In order to accurately define a group of patients with moderate familial risk; patients were categorised according to cancer mismatch repair status as well as family history. Patients were defined as having moderate familial risk if they had a mismatch repair intact first primary CRC and a family history suggestive of low or high moderate risk (see Table 3, section 1.6.3 and repeated above). Both high moderate and low moderate risk groups were investigated together for this study. Cancers were defined as mismatch repair intact if all mismatch repair proteins were present on immunohistochemistry (method and validation previously described in section 2.1.4.2 and in the literature) (Barrow, Jagger et al. 2011), or if tumour derived DNA was found to be microsatellite stable (on a panel of between 1 and 5 microsatellite markers; again method described in section 2.1.5.1). Patients were defined as having Lynch syndrome if they fulfilled one of the categories below:

1. Proven pathogenic mismatch repair gene (MMR) mutation carriers (MLH1, MSH2, MSH6, PMS2)
2. Obligate pathogenic MMR gene mutation carriers (due to their position in the pedigree in relation to mutation positive relatives)
3. First degree relatives of a proven pathogenic MMR gene carrier who has been affected with a Lynch spectrum cancer

4. Individuals who have had a MMR deficient CRC and fulfilled Amsterdam Criteria (I or II) from mutation untested families or from families in whom no mutation has been found

At the time of study, the data held on the database regarding many non-Lynch patients was incomplete. Missing data was identified. Medical case notes were reviewed, additional pathology reports sought where necessary, and enquiries made to local hospitals where appropriate in order to make the data as complete as possible. Some patients were contacted directly (by their genetic counsellors). Cancer incidence was verified with the North West Cancer Intelligence Service (NWCIS).

All patients from a small local district general hospital who had a CRC between 1981 and 1997 were investigated as part of a previous epidemiological study (Evans, Walsh et al. 1997). They all have a thorough verified and documented family history. All patients in this large cohort who were classified as average risk (i.e. not moderate risk and not Amsterdam I or II) were identified as an additional cohort of general population risk patients.

For all patients data was gathered on date of birth, gender, date of first and subsequent CRCs, date of death or date of last follow-up, family risk category, family diagnosis, germline mutation if found, and mismatch repair status. Site and Dukes’ stage of CRC was gathered if available. Metachronous cancers were defined as cancers that occurred at least 12 months after the first and were in broadly a different part of the colorectum. For example, if a patient had a left sided cancer resected and had a second cancer five years later in the right side of the colon this was considered to be a metachronous cancer. If the same patient had had a second cancer 14 months later at the site of the anastomosis, this was considered to be recurrence of the first cancer and therefore was not a metachronous cancer. Time till second CRC was calculated using date of first CRC and date of second CRC. If the patient had not had a metachronous CRC at the time of the study, date of last follow or death was used and metachronous cancer-free time predicted using Kaplan-Meier curves. A cumulative risk analysis was also performed in order to
estimate cumulative incidence of a metachronous cancer within each group, and test equality between groups. This analysis takes into account the competing risk of death before occurrence of a metachronous cancer. Overall survival and five year survival following cancer diagnosis were calculated. Actuarial survival was calculated from date of birth till date of death. If the patient was still alive at the time of the study, date of last follow-up was used and survival times predicted using Kaplan-Meier curves.

4.2.2 Data analysis

The Manchester Familial Colorectal Cancer filemaker pro® database was interrogated. Data was downloaded into Excel software and coded numerically.

Two statistical methods were used to estimate the life-time risk of metachronous colorectal cancer. The Kaplan-Meier risk estimate (1-KM) is the risk for individuals who remain alive (i.e. at risk) at a given time point and the cumulative incidence function is the risk for the population as a whole. The Kaplan-Meier risk estimate (1-KM) was estimated using SPSS 13.0 (IBM, UK) software, and Mantel-cox log rank used to test for equality. To estimate the cumulative incidence of metachronous CRC, a cumulative incidence function was conducted using Stata version 11 (Stata Corps, College Station, Texas, USA) and R version 2.12.2 (http://CRAN.R-project.org/doc/FAQ/R-FAQ.html Accessed 2011). This was calculated as presented by Coviello and Boggess (Coviello V 2004) and R version 2.12.2 was used to conduct specific K-sample tests (Gray RJ 1988) to test the null hypotheses that there is no difference in risk of metachronous colorectal cancer between the different familial risk categories.

To estimate overall survival and survival from cancer onset Kaplan-Meier survival function was used. The data was entered into SPSS 13.0 (IBM, UK) software. Kaplan-Meier analysis and log-rank (Mantel-Cox) tests were completed.
4.3 Results Study three: Metachronous colorectal cancer risk in patients with a moderate familial risk

At the time of the study the Manchester Familial Colorectal Cancer database contained information on 6421 individuals. 1879 have had at least one CRC. 154 patients with an unverified family history were excluded from the study. 565 patients with unknown tumour mismatch repair status were also excluded as Lynch syndrome could not be confidently ruled out. 383 patients were included in the moderate familial risk group, 528 in the Lynch Syndrome group and 249 in the average population risk group. An additional 160 patients were identified from the small local district general hospital population based dataset who have had a colorectal cancer and whose family history suggests normal population risk (see figure 29).

At the time of the study 65/528 (12.3%) Lynch Syndrome patients, 18/383 (4.7%) moderate risk patients, and 6/409 (1.5%) population risk patients developed a metachronous CRC (mCRC) and 301/528 (57.0%) of the Lynch group, 186/383 (48.6%) of the moderate risk group and 278/409 (68.0%) of the population risk group had died.

The incident CRC in the Lynch syndrome group was diagnosed between 01/01/1911 and 25/02/2010. The median follow-up was 54.3 years (range 16.9-93.3 years). In the moderate risk group the incident CRC was diagnosed between 01/01/1949 and 08/02/2010. The median follow-up was 60.6 years (range 16.9-93.4 years). In the population risk group the incident CRC was diagnosed between 19/11/1964 and 26/11/2008. The median follow-up was 76.2 years (range 27.6-102.2 years). Data on 83180.3 person-years were obtained.
Figure 29: Consort diagram used to assess the incidence of metachronous colorectal cancer in patients with a moderate familial risk.
Table 34: Baseline characteristics of patients included in this study

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Lynch n=528</th>
<th>Moderate risk n=383</th>
<th>Population risk n=409</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>*¹NO mCRC</td>
<td>*²mCRC</td>
<td>No mCRC</td>
</tr>
<tr>
<td>Total</td>
<td>463</td>
<td>65</td>
<td>365</td>
</tr>
<tr>
<td>Sex (female %)</td>
<td>205 (44.3%)</td>
<td>31 (47.7%)</td>
<td>180 (40.3%)</td>
</tr>
<tr>
<td>Median age 1st CRC (years)</td>
<td>45.8</td>
<td>44.1</td>
<td>55.1</td>
</tr>
<tr>
<td>Median actuarial survival (years) derived from KM 95% CI in parentheses</td>
<td>61.9 (60.2-63.5)</td>
<td>69.8 (66.1-73.5)</td>
<td>70.4 (67.0-73.8)</td>
</tr>
</tbody>
</table>

*¹No mCRC = patients who did not have a metachronous colorectal cancer

*²mCRC = patients who had a metachronous colorectal cancer
Table 35: Tumour site and stage. *1 mCRC = patients who did not have a metachronous colorectal cancer, *2 mCRC = patients who had a metachronous colorectal cancer

<table>
<thead>
<tr>
<th>Site of 1st CRC</th>
<th>Lynch *1 no mCRC</th>
<th>Lynch *2 mCRC</th>
<th>Moderate no mCRC</th>
<th>Moderate mCRC</th>
<th>Population risk no mCRC</th>
<th>Population risk mCRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right colon</td>
<td>155 (33.5%)</td>
<td>26 (40.0%)</td>
<td>100 (27.4%)</td>
<td>4 (22.2%)</td>
<td>94 (31.3%)</td>
<td>2 (33.3%)</td>
</tr>
<tr>
<td>Left colon/rectum</td>
<td>110 (23.8%)</td>
<td>23 (35.4%)</td>
<td>205 (51.2%)</td>
<td>11 (55.6%)</td>
<td>294 (73.0%)</td>
<td>4 (66.7%)</td>
</tr>
<tr>
<td>Synchronous</td>
<td>9 (1.4%)</td>
<td>0</td>
<td>5 (1.4%)</td>
<td>0</td>
<td>1 (0.2%)</td>
<td>0</td>
</tr>
<tr>
<td>Unknown</td>
<td>188 (40.6%)</td>
<td>16 (24.6%)</td>
<td>55 (15.1%)</td>
<td>3 (16.7%)</td>
<td>14 (3.5%)</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stage of 1st CRC</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>11 (2.4%)</td>
<td>45 (9.7%)</td>
<td>3 (0.6%)</td>
<td>3 (0.6%)</td>
<td>368 (79.5%)</td>
</tr>
<tr>
<td>B</td>
<td>5 (7.7%)</td>
<td>15 (23.0%)</td>
<td>5 (7.7%)</td>
<td>0</td>
<td>43 (75.4%)</td>
</tr>
<tr>
<td>C</td>
<td>17 (4.7%)</td>
<td>38 (10.4%)</td>
<td>50 (13.7%)</td>
<td>0</td>
<td>255 (69.9%)</td>
</tr>
<tr>
<td>D</td>
<td>1 (5.6%)</td>
<td>4 (22.2%)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unknown</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>403 (100%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site of 2nd CRC</th>
<th>Right colon</th>
<th>Left colon/rectum</th>
<th>Synchronous</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-</td>
<td>24 (36.9%)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B</td>
<td>-</td>
<td>25 (38.5%)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D</td>
<td>-</td>
<td>18 (27.7%)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stage of 2nd CRC</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1 (1.5%)</td>
<td>8 (12.3%)</td>
<td>5 (7.7%)</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1 (5.5%)</td>
<td>3 (16.7%)</td>
<td>1 (5.5%)</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
### 4.3.1 Metachronous CRC

Table 36: Risk of developing a metachronous colorectal cancer (mCRC) according to familial risk category (*derived from Kaplan-Meier 95% confidence intervals in parentheses*)

<table>
<thead>
<tr>
<th>Familial Risk</th>
<th>5 year *mCRC risk</th>
<th>10 year mCRC risk</th>
<th>20 year mCRC risk</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lynch All</strong> (n=528)</td>
<td>5.1% (3.1-6.4%)</td>
<td>15.1% (14.1-18.9%)</td>
<td>31.6% (27.7-35.5%)</td>
</tr>
<tr>
<td><strong>MLH1</strong> (n=224)</td>
<td>8.6% (3.4-13.9)%</td>
<td>22.2% (18.1-26.3%)</td>
<td>47.1% (40.7-53.5%)</td>
</tr>
<tr>
<td><strong>MSH2</strong> (n=219)</td>
<td>2.5% (1.0-4.0%)</td>
<td>8.5% (5.5-11.5%)</td>
<td>23.5% (17.6-29.4%)</td>
</tr>
<tr>
<td><strong>MSH6</strong> (n=28)</td>
<td>No cases of mCRC</td>
<td>No cases of mCRC</td>
<td>25.0% (22.8-27.2%)</td>
</tr>
<tr>
<td><strong>Unknown</strong> (n=57)</td>
<td>5.3% (1.2-9.5%)</td>
<td>22.3% (14.6-30.0%)</td>
<td>38.6% (37.2-40.0%)</td>
</tr>
<tr>
<td><strong>Moderate risk</strong> (n=383)</td>
<td>2.7% (1.6-3.7%)</td>
<td>6.3% (4.3-8.3%)</td>
<td>23.5% (15.8-31.2%)</td>
</tr>
<tr>
<td><strong>Population risk</strong> (n=409)</td>
<td>1.3% (0.5-2.0%)</td>
<td>3.1% (4.5-7.3%)</td>
<td>7.0% (3.0-11.0%)</td>
</tr>
</tbody>
</table>

*mCRC = metachronous colorectal cancer

Using the Kaplan-Meier estimate the risk of developing a second, or metachronous, CRC is higher in the Lynch Syndrome group than the moderate familial risk group. The risk estimate for mCRC is higher in the moderate risk group than the population risk group. This risk is 31.6% (95% CI 27.7-35.5%) for Lynch syndrome, 23.5% (95% CI 15.8-31.2%) for patients with moderate familial risk and 7.0% (95% CI 3.0-11.0%) for patients at population risk at 20 years (p=0.0037). (See Figure 30)

Life time risk of developing a mCRC is 41.1% (95% CI 31.3-50.7%) in patients with Lynch Syndrome, 24.0% (95% CI 16.6-31.4%) in patients with a moderate
familial risk and 3.1% (95% CI 1.6-4.6) in patients with population risk (p<0.0001).

Figure 30: Kaplan-Meier curve demonstrating the probability of developing a metachronous colorectal cancer according to familial risk
Table 37: Cumulative incidence function of metachronous CRC according to familial risk

<table>
<thead>
<tr>
<th>Familial Risk</th>
<th>5 year *mCRC risk</th>
<th>10 year mCRC risk</th>
<th>20 year mCRC risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lynch All (n=528)</td>
<td>1.3%</td>
<td>3.3%</td>
<td>14.0%</td>
</tr>
<tr>
<td>MLH1 (n=224)</td>
<td>2.4%</td>
<td>4.7%</td>
<td>19.0%</td>
</tr>
<tr>
<td>MSH2 (n=219)</td>
<td>0.5</td>
<td>2.0%</td>
<td>10.1%</td>
</tr>
<tr>
<td>MSH6 (n=28)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unknown (n=57)</td>
<td>0</td>
<td>3.0%</td>
<td>15.8%</td>
</tr>
<tr>
<td>Moderate risk (n=383)</td>
<td>1.2%</td>
<td>2.5%</td>
<td>9.6%</td>
</tr>
<tr>
<td>Population risk (n=409)</td>
<td>0.3%</td>
<td>0.6%</td>
<td>2.4%</td>
</tr>
</tbody>
</table>

*mCRC = moderate colorectal cancer risk

The cumulative incidence function estimates the cumulative incidence of mCRC to be greatest in the Lynch syndrome group (3.3% at 10 years and 14.0% at 20 years), followed by the moderate risk group (2.5% at 10 years and 9.6% at 20 years) and lastly the population risk group (0.6% at 10 years and 2.4% at 20 years). These risks are roughly half that predicted using the Kaplan-Meier risk estimate. This can be accounted for by the large proportion of deaths in these groups.
Figure 31: Cumulative incidence function of metachronous CRC (mCRC) according to familial risk
Table 38: Odds ratios of metachronous CRC risk (KM and Cumulative Incidence)  *mCRC = metachronous colorectal cancer

<table>
<thead>
<tr>
<th></th>
<th>KM risk estimate</th>
<th>Cumulative incidence function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odds ratio and Log rank (mantel-Cox)</td>
<td>Lynch Syndrome vs. Moderate Risk</td>
<td>Moderate Risk vs. Population Risk</td>
</tr>
<tr>
<td>OR of *mCRC at 5 years</td>
<td>1.9</td>
<td>2.1</td>
</tr>
<tr>
<td>OR of mCRC at 10yrs</td>
<td>2.4</td>
<td>2.0</td>
</tr>
<tr>
<td>OR of mCRC at 20 yrs</td>
<td>1.3</td>
<td>3.4</td>
</tr>
<tr>
<td>P Value</td>
<td>p=0.006</td>
<td>p=0.008</td>
</tr>
<tr>
<td>Chi square</td>
<td>7.513</td>
<td>7.12</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
The Bonferroni correction describes that significance should be tested at \( p < \frac{1}{n} \times 0.05 \) where \( n \) is the number of hypotheses tested from one dataset. Here 2 null hypotheses are tested (the mCRC incidence curve is the same in Lynch and moderate risk, and the mCRC incidence curve is the same in moderate risk as population risk) therefore a \( p \) value of \( \frac{1}{2} \times 0.05 = 0.025 \).

Using the Kaplan-Meier risk estimate, the cumulative risk of developing a second CRC is higher for \( MLH1 \) mutation carriers than \( MSH2 \) or \( MSH6 \) mutation carriers. (See Figure 32) This was significant when comparing the whole Lynch Syndrome group (\( p=0.02 \)) and when directly comparing \( MLH1 \) and \( MSH2 \) carriers (\( p=0.008 \)). (Bonferroni correction \( p=\frac{1}{2} \times 0.05=0.025 \)).

Figure 32: Kaplan-Meier curve demonstrating the probability of developing a metachronous colorectal cancer according to Lynch Syndrome mutation (all)
Using the cumulative incidence function the significant difference in the cumulative incidence of mCRC between the Lynch Syndrome mutations is preserved (p=0.047). The risk is significantly higher in those with an MLH1 mutation than those with an MSH2 mutation (p=0.037). (see Figure 33)

Figure 33: Cumulative incidence function(Fine JP 1999) of metachronous CRC (mCRC) according to Lynch Syndrome mutation
The ten year risk of developing a mCRC in patients with an *MLH1* mutation is 22.2% according the Kaplan-Meier estimate (1-KM) and 4.7% according to the cumulative incidence function. The ten year risk of developing a mCRC in patients with an *MSH2* mutation is 8.5% according the Kaplan-Meier estimate (1-KM) and 2.6% according to the cumulative incidence function.

The 20 year risk of developing a mCRC in patients with an *MLH1* mutation is 47.1% according the Kaplan-Meier estimate (1-KM) and 19.0% according to the cumulative incidence function. The 20 year risk of developing a mCRC in patients with an *MSH2* mutation is 23.5% according the Kaplan-Meier estimate (1-KM) and 10.1% according to the cumulative incidence function.

### 4.3.2 Age of onset of CRC

The age of onset of first CRC is significantly different between all groups (Anova test of variance p<0.0001). The age of onset of second CRC is significantly different between patients with Lynch Syndrome and patients with moderate familial risk (Logrank Mantel-Cox p=0.012), and between patients with moderate familial risk and population risk (Logrank Mantel-Cox p<0.0001). The age of onset of second cancer was estimated using the Kaplan-meier risk estimate (1-KM). This method censors patients who have died in the same way as those who have not had the event of interest (metachronous colorectal cancer in this case). Because of this there is an upward shift beyond the age at which many patients will have died, particularly in the population risk group who are older at the time of their first cancer. This is why the estimated age of onset of metachronous cancer is 100.9 years in the population risk group. However it is still possible to say that the age of onset of second cancer is significantly greater in the population risk group than the moderate risk group, and the Lynch Syndrome group respectively.
Table 39: Age of Onset of CRC

<table>
<thead>
<tr>
<th>Risk Category</th>
<th>Mean (years)</th>
<th>Median (years)</th>
<th>Mean (years)</th>
<th>Median (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lynch Syndrome n=528</td>
<td>46.8</td>
<td>45.5</td>
<td>57.4 (54.3-60.4)</td>
<td>56.8 (51.2-62.5)</td>
</tr>
<tr>
<td>Moderate familial risk n=383</td>
<td>55.6</td>
<td>55.5</td>
<td>66.8 (58.9-74.6)</td>
<td>66.8 (59.6-74.1)</td>
</tr>
<tr>
<td>Population risk n=409</td>
<td>65.5</td>
<td>70.3</td>
<td>100.9 (99.7-102.2)</td>
<td>Too much censored data</td>
</tr>
</tbody>
</table>
4.3.3 Survival from diagnosis of first colorectal cancer

Survival from diagnosis of first CRC is significantly different between all groups (Log rank Mantel-Cox \( p=0.003 \)). This is due to a significant difference in overall survival between those with a moderate familial risk and those with population risk (\( p=0.004 \)). This is likely due to the greater age of onset of first CRC in patients with population risk. Their risk of dying from other causes is therefore greater. There is no significant difference between survival in those patients with Lynch syndrome and those with a moderate familial risk (\( p=0.55 \)) (See Figure 34). The Bonferroni correction here would test significance at a \( p \) value of \( 1/3 \times 0.05 = 0.017 \). Survival from first cancer diagnosis is similar between the different mutations in Lynch syndrome (\( p=0.351 \)) and 5 year survival from cancer diagnosis is also similar (\( p=0.481 \)).
Table 40: Survival from cancer diagnosis (all stage) *(derived from Kaplan-Meier 95% confidence intervals in parentheses)*

<table>
<thead>
<tr>
<th>Familial Risk</th>
<th>5 year survival (95% CI)</th>
<th>Median overall survival (years) &amp; (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lynch syndrome</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>58.0% (55.8-60.2%)</td>
<td>10.5 (7.3-13.6)</td>
</tr>
<tr>
<td>MLH1</td>
<td>58.3% (57.9-61.7%)</td>
<td>11.8 (7.3-16.3)</td>
</tr>
<tr>
<td>MSH2</td>
<td>57.0% (53.6-60.4%)</td>
<td>9.5 (5.4-13.6)</td>
</tr>
<tr>
<td>MSH6</td>
<td>51.8% (42.1-61.58%)</td>
<td>2.9 (0.0-13.9)</td>
</tr>
<tr>
<td>Unknown</td>
<td>62.8% (61.9-69.7%)</td>
<td>14.1 (3.3-24.9)</td>
</tr>
<tr>
<td>Moderate family Risk</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58.7% (56.0-61.4%)</td>
<td>9.1 (5.3-12.9)</td>
<td></td>
</tr>
<tr>
<td>Population Risk</td>
<td>55.4% (52.9-57.9%)</td>
<td>6.8 (5.3-8.4)</td>
</tr>
</tbody>
</table>
Figure 34: Cumulative probability of overall survival from diagnosis with first colorectal cancer
Table 41: Survival following cancer diagnosis according to Dukes stage (derived from Kaplan-Meier, 95% confidence intervals in parentheses)

<table>
<thead>
<tr>
<th>Familial Risk</th>
<th>Proportion (of patients for whom Dukes stage was known)</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lynch syndrome</td>
<td></td>
<td>21.0%</td>
<td>64.8%</td>
<td>10.3%</td>
<td>3.9%</td>
</tr>
<tr>
<td>5 year survival (95% confidence interval)</td>
<td></td>
<td>93.3% (86.9-99.7%)</td>
<td>87.4% (82.5-92.3%)</td>
<td>71.0% (63.6-78.4%)</td>
<td>33.3% (6.1-60.5%)</td>
</tr>
<tr>
<td>Median survival (95% confidence interval)</td>
<td></td>
<td>Too much censored data</td>
<td>19.1 years (no CI as censored data)</td>
<td>17.7 years (11.6-23.8)</td>
<td>2.5 years (0.08-6.2)</td>
</tr>
<tr>
<td>Moderate family Risk</td>
<td></td>
<td>15.8%</td>
<td>36.8%</td>
<td>43.9%</td>
<td>3.5%</td>
</tr>
<tr>
<td>5 year survival (95% confidence interval)</td>
<td></td>
<td>100%</td>
<td>75.2% (67.9-82.5%)</td>
<td>46.7% (38.5-54.9%)</td>
<td>0%</td>
</tr>
<tr>
<td>Median survival (95% confidence interval)</td>
<td></td>
<td>23.1 years (20.1-26.1)</td>
<td>19.5 years (13.2-25.8)</td>
<td>10.1 years (7.5-12.6)</td>
<td>1.2 years (0.2-2.3)</td>
</tr>
<tr>
<td>Population Risk</td>
<td></td>
<td>13.2%</td>
<td>36.8%</td>
<td>35.9%</td>
<td>9.4%</td>
</tr>
<tr>
<td>5 year survival from CRUK(<a href="http://info.cancerresearchuk.org/cancerstats/types/bowel/Bowel">http://info.cancerresearchuk.org/cancerstats/types/bowel/Bowel</a> Accessed 2011)</td>
<td></td>
<td>93.2%</td>
<td>77.0%</td>
<td>48.0%</td>
<td>6.6%</td>
</tr>
</tbody>
</table>
Completeness of data regarding Dukes stage of individuals at population risk was poor. For comparison table 41 includes figures quoted by Cancer Research UK (http://info.cancerresearchuk.org/cancerstats/types/bowel/Bowel Accessed 2011). 5 year survival in the moderate familial risk group is similar to the population figures at 100% for Dukes A, 75.2% for Dukes B, and 46.7% for Dukes C. 5 year survival in those patients with Lynch syndrome is better than the population figures at 93.3% for Dukes A, 87.4% for Dukes B, 71.0% for Dukes C and 33.3% for Dukes D. However, in this study population, the Dukes stage was unknown for 77.8% of patients with Lynch Syndrome and 69.4% of patients with moderate familial risk so statistical analysis would be uninformative.

4.3.4 Actuarial Survival

The actuarial survival from birth of patients who have had at least one CRC associated with a familial risk of colorectal cancer is significantly different between those with Lynch Syndrome and those with a moderate familial risk. The median overall survival of Lynch syndrome patients is 8 years less than patients with moderate family risk (p<0.0001). There is no significant difference in overall survival between the Lynch Syndrome mutations (p=0.318). The overall survival of patients whom have had at least one CRC associated with a moderate family risk is significantly lower than that of those with population risk (Log rank Mantel-Cox p<0.0001). The median survival is 7 years less in patients with a moderate family history than those with population risk. (see table 42, 43 and Figure 35) This is likely due to difference in the age of onset.
Table 42: Actuarial survival of patients who have had at least one CRC (*derived from Kaplan-Meier 95% confidence intervals in parentheses*)

<table>
<thead>
<tr>
<th>Familial Risk</th>
<th>Median overall survival (years)</th>
<th>Mean overall survival (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lynch syndrome</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>63.0 (61.0-65.0)</td>
<td>62.6 (61.0-71.8)</td>
</tr>
<tr>
<td>MLH1</td>
<td>62.1 (56.7-67.5)</td>
<td>62.4 (59.7-65.1)</td>
</tr>
<tr>
<td>MSH2</td>
<td>63.0 (60.9-65.1)</td>
<td>61.7 (59.3-64.1)</td>
</tr>
<tr>
<td>MSH6</td>
<td>71.1 (55.9-86.4)</td>
<td>67.9 (62.1-73.7)</td>
</tr>
<tr>
<td>Unknown</td>
<td>66.7 (61.4-72.0)</td>
<td>65.3 (61.2-69.4)</td>
</tr>
<tr>
<td><strong>Moderate Familial Risk</strong></td>
<td>70.5 (67.7-73.3)</td>
<td>70.0 (68.3-71.8)</td>
</tr>
<tr>
<td><strong>Population Risk</strong></td>
<td>77.9 (75.5-80.3)</td>
<td>76.4 (75.0-77.9)</td>
</tr>
</tbody>
</table>
Table 43: Proportion of patients surviving at completion of each decade (derived from Kaplan-Meier, 95% confidence intervals in parentheses)

<table>
<thead>
<tr>
<th>Familial Risk</th>
<th>&lt;30 years</th>
<th>&lt;40 years</th>
<th>&lt;50 years</th>
<th>&lt;60 years</th>
<th>&lt;70 years</th>
<th>&lt;80 years</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lynch Syndrome</strong></td>
<td>98.1%</td>
<td>92.2%</td>
<td>73.8%</td>
<td>57.7%</td>
<td>33.2%</td>
<td>16.7%</td>
</tr>
<tr>
<td></td>
<td>(97.5-98.7%)</td>
<td>(91.0-93.4%)</td>
<td>(71.8-75.8%)</td>
<td>(55.3-60.1%)</td>
<td>(30.5-35.9%)</td>
<td>(14.0-19.4%)</td>
</tr>
<tr>
<td><strong>Moderate Family Risk</strong></td>
<td>98.9%</td>
<td>97.4%</td>
<td>88.6%</td>
<td>72.7%</td>
<td>51.8%</td>
<td>32.5%</td>
</tr>
<tr>
<td></td>
<td>(98.4099.4%)</td>
<td>(96.6-97.8%)</td>
<td>(86.9-92.0%)</td>
<td>(70.3-75.1%)</td>
<td>(48.8-54.8%)</td>
<td>(28.9-36.1%)</td>
</tr>
<tr>
<td><strong>Population Risk</strong></td>
<td>99.3%</td>
<td>98.0%</td>
<td>95.3%</td>
<td>87.7%</td>
<td>70.8%</td>
<td>45.9%</td>
</tr>
<tr>
<td></td>
<td>(99.0-99.6%)</td>
<td>(97.3-98.7%)</td>
<td>(94.2-96.4%)</td>
<td>(86.1-89.3%)</td>
<td>(68.5-73.1%)</td>
<td>(43.2-48.6%)</td>
</tr>
</tbody>
</table>
Figure 35: Actuarial survival of patients who have had at least one colorectal cancer according to familial risk.
4.4 Discussion Study Three: Metachronous colorectal cancer risk in patients with a moderate familial risk

4.4.1 Discussion of metachronous CRC risk

4.4.1.1 Metachronous CRC risk

This dataset represents the largest reported cohort of individuals with moderate familial risk colorectal cancer. The risk of metachronous CRC in such a population is previously unknown. Knowledge of risk of metachronous CRC is vital to inform screening guidelines.

This study has demonstrated that the risk of a metachronous CRC is significantly higher for patients with a moderate risk than that for patients at normal population risk. Odds ratio of risk at 10 years is 2.0 according to 1-Kaplan-Meier (p=0.008) and 4.2 according to cumulative incidence function (p<0.001). Odds ratio of risk at 20 years is 3.4 according to 1-Kaplan-Meier (p<0.001) and 4.0 according to cumulative incidence function (p<0.001). In patients with moderate familial risk; 1-Kaplan-Meier risk estimate found the risk to be 2.7% (95% CI 1.6-3.7%) at 5 years, 6.3% (95% CI 4.3-8.3%) at 10 years, and 23.5% (95% CI 15.8-31.2%) at 20 years. The cumulative incidence estimate is 1.2% at 5 years, 2.5% at 10 years, and 9.6% at 20 years. This risk is linear and does not reduce after 5-10 years.

Currently there is no guidance regarding post-operative surveillance of remaining bowel in patients with a moderate family history of colorectal cancer. The British Society of Gastroenterology guidelines recommend surveillance regimens for those with a high moderate or low moderate family risk, but these guidelines do not specify how to proceed following bowel resection for a cancer. In the sporadic setting, usual practice is an initial post-operative colonoscopy at three years followed thereafter either at five year intervals, at varying intervals according to the adenoma guidance, or is discontinued. This depends on individual clinician practice.
This study has found that for patients at population risk the risk of a metachronous CRC is; 1-Kaplan-Meier risk estimate 1.3% (95% CI 0.5-2.0%) at 5 years, 3.1% (95% CI 4.5-7.3%) at 10 years, and 7.0% (95% CI 3.0-11.0%) at 20 years. The cumulative incidence estimate 0.3% at 5 years, 0.6% at 10 years, and 2.4% at 20 years. This is consistent with published data. A meta-analysis performed by the American Cancer Society and the US multi-society task force on colorectal cancer found the life-time incidence of a metachronous cancer to be 1.517% in sporadic cases (Rex, Kahi et al. 2006).

This study has found that the risk of a metachronous CRC for patients with Lynch Syndrome is; 1-Kaplan-Meier risk estimate 5.1% (95% CI 3.1-6.4) at five years, 15.1% (95% CI 14.1-18.9%) at 10 years and 31.6% (95% CI 27.7-35.5%) at 20 years. The cumulative risk estimate is 1.3% at five years, 3.3% at 10 years and 14.0% at 20 years. Parry et al recently reported a comparison of metachronous CRC risk in 382 patients with Lynch syndrome (ascertained through the Colon Cancer Family Registry based in North America and Australasia) according to the extent of the surgical resection for their first CRC (3545 person years follow up, mean follow up 9 years, range 1-40 years). This group found the Kaplan-Meier risk estimate for a mCRC to be 16% (95% CI 10-25%) at 10 years and 41% (95% CI 30-52%) at 20 years in 332 patients who had a segmental resection. There were no cases of mCRC in the 50 patients who had an extensive resection (Parry, Win et al. 2011). Nataranjan et al also recently reported data on mCRC risk in Lynch syndrome mutation carriers and obligate carriers (from the Creighton University Lynch Syndrome Registry). Over a median follow-up of 12 years they found a mCRC rate of 26% in 69 patients who had a segmental resection and a rate of 6% in 37 patients who had an extended resection (Natarajan, Watson et al. 2010). De vos tot Nederveen Cappel et al reported data from the Dutch and Finnish Lynch Syndrome Registries and found the rate of mCRC at 10 years to be 15.7% in 68 patients who had a segmental resection (mean follow-up 7.1 years) and 3.4% at 10 years in 29 patients who had an extended resection (mean follow-up 5 years) (de Vos tot Nederveen Cappel, Nagengast et al. 2002). The vast majority of patients in the current study underwent a segmental resection for their first CRC so the findings are comparable. In comparison with Lynch syndrome, the risk of metachronous colorectal cancer is low in moderate risk patients and this risk does not justify extended resection for the first colorectal cancer.
The rate of metachronous CRC was significantly different between patients with *MLH1* and *MSH2* mutations. The *MSH6* group was small and there was only one case of a second CRC 30 years after the first. 1-KM risk estimation found the mCRC risk to be 15.1% (95% CI 14.1-18.9) for *MLH1* mutation carriers and 8.5% (95% CI 5.5-11.5%) for *MSH2* carriers at 10 years, and 31.6% (95% CI 27.7-35.5%) for *MLH1* and 23.5% (95% CI 17.6-29.4%) for *MSH2* at 20 years. (Logrank Mantel-Cox \( p=0.008 \)). Cumulative incidence function found the 10 year risk to be 4.7% in *MLH1* and 2.0% in *MSH2*, and the 20 years risk was 19.0% in *MLH1* and 10.1% in *MSH2*. (Specific K sample test of equality \( p=0.037 \)). Although the absolute values of risk are different between these two statistical methods, the pattern remains the same. The increase in risk for mCRC with time is linear. The increase in risk is parallel between *MLH1* and *MSH2* mutation carriers with the risk for *MLH1* carriers being roughly double that of *MSH2* carriers. The reason for this is unclear. Previous studies have found no difference in the incidence of mCRC between *MLH1* and *MSH2* mutation carriers. (Parry, Win et al. 2011) This difference in risk is unlikely to justify different screening protocol or surgical management.

This study was a retrospective observational study of patients listed on a database within the regional department of medical genetics. Although the ideal study would be a prospective observational study this would not be possible in the present day as many patients would have mCRCs prevented by screening. The study was not strictly a population based study. This may have led to an ascertainment bias. Patients who have had metachronous CRC are more likely to be referred for genetic consultation and thus to have been entered onto the Familial Colorectal Cancer Registry and included in this study (this consideration is relevant for both the Lynch syndrome group and the moderate familial risk group). In an attempt to investigate for possible ascertainment bias, I conducted a small retrospective population based study of all patients who had a colorectal cancer at Trafford General Hospital between January 1995 and December 1997. Every patient had a complete and verified family history documented. The regional cancer registry (North West Cancer Intelligence Service, NWCIS) was interrogated for subsequent CRCs. In this 36 month period, there were 233 patients who were diagnosed with CRC. There was a total of 1058 person-years follow-up data. One patient was diagnosed with a pathogenic *MLH1* mutation (patient was Bethesda criteria positive so tumour MSI testing and then germline mutation testing was conducted). 228 patients were
classified as at population risk and three were classified as having moderate familial risk. Lynch Syndrome was excluded in all the moderate risk patients. 1/3 (33.3%) of the moderate familial risk patients and 9/228 (3.9%) of the population risk patients went on to have a mCRC. Although, the numbers are too small to produce meaningful incidence curves, both the cumulative incidence function (specific k-sample test; p=0.0274) and the 1-Kaplan-Meier risk estimate (Log rank mantel-cox p=0.009) demonstrated a significantly greater risk of a mCRC in the patients with moderate familial risk. This population based cohort is small but the data demonstrates that the increased risk of mCRC in the main cohort of patients with moderate familial risk may not be limited to patients who have been referred to the genetic medicine service (i.e. demonstrates that the effect of ascertainment bias may be small).

A further limitation to this study is the lack of data regarding extent of resection as treatment for the first colorectal cancer. This is an important consideration as patients who have had a segmental colectomy are at a greater risk of a metachronous colorectal cancer compared to those patients who have had a subtotal or total colectomy. Data regarding the type of surgery performed is documented on the Manchester Familial Colorectal Cancer database for patients with Lynch Syndrome but not for patients with moderate risk or population risk so it was not possible to adjust for the type of surgery performed. However, the numbers of patients within the moderate risk group and the population risk group who were likely to have had an extensive resection (e.g. those with synchronous first cancers) was very small (12/383 (3.1%) of the moderate risk group, and 15/409 (3.7%) of the population risk group). Consequently this would be unlikely to affect the results or conclusions to any great extent.

4.4.1.2 Discussion of statistical methodology

This study utilised two statistical methods for assessing the risk of a metachronous CRC. Although both methods have demonstrated a statistically significantly increased risk of mCRC in patients with moderate familial risk compared to those at population risk, the absolute values of risk differ. In order to be able to make an informed decision about post-operative screening of remaining large bowel in these patients, it is imperative to have an understanding of how these values were
derived and to be able to assess which absolute value is the most accurate within this particular dataset. The first method; the Kaplan-Meier risk estimate censors patients at death. The Kaplan-meier risk estimate therefore calculates risk at a given time-point that applies to those who are still alive at that time-point. This method does not allow for the consideration of competing risk events, whereas the latter method; cumulative incidence function does. Competing risk is defined as “an event whose occurrence precludes or alters the probability of occurrence of a main event under examination” (Coviello V 2004) In this study death (either due to the first CRC or another cause) is a competing risk for the occurrence of a mCRC.

Kaplan-Meier is frequently used to estimate survival curves. It allows for the calculation of the probability of survival at given time points using both complete and censored data. The Kaplan-Meier risk estimate (1-KM) calculates the probability of an event occurring at given time points. Time to event (in this case metachronous CRC) is used where available (i.e. when the event has happened), and when the event has not occurred the patient is censored at the time of last known contact. Censoring allows for that patient to be incorporated into the risk estimator up to the point at which the status of the patient was last known. Kaplan-Meier calculates the risk of an event occurring during a time interval as a proportion of the patients still at risk at the end of that interval. Cumulative survival or risk is the product of the survival or risk calculated for each subsequent time interval. This method has been widely used in the literature for both survival analysis and cumulative risk estimation. Indeed it was used recently by Parry et al (Parry, Win et al. 2011) in a study of a large Lynch Syndrome cohort to calculate the risk of metachronous CRC following different types of colorectal resection. In this method of estimating cumulative risk, patients who die before the event of interest has happened are censored in the same way that patients who have not experienced the event of interest but are still alive at the end of the study. This relies on the assumption that the censoring is uninformative (i.e. that the patients who died would have been as likely to get a second cancer, had they not died, than the remaining alive patients) The Kaplan-Meier estimate gives the risk for the patients that remain at risk at any given time point, rather than the population as a whole. It can be argued that this is the most clinically relevant method. Thus for a person with a Duke’s A CRC Kaplan-Meier estimate is likely to be very relevant. However, if undertaking a health economics analysis of effectiveness deaths are very relevant as many individuals at time of diagnosis will not survive to develop a predicted mCRC and would not benefit from surveillance to prevent it. Nonetheless in genetic
counselling many cases of CRC are seen years after the original diagnosis (of Moderate risk or Lynch Syndrome) and their survival chances are thus much improved. As such the 1-KM figure may again be more relevant.

4.4.1.2.1 Discussion of Kaplan-Meier Survival Analysis Methodology

Kaplan-Meier survival analysis calculates the risk of a patient surviving a given time interval as a proportion of the number of patients at risk of death at the end of each interval. Although in the example dataset in table 45 the intervals are linear, in a real KM calculation each interval is bounded by a death or deaths. For example, if a patient dies at t=1.3, the proportion surviving to t=1.3 will be calculated. If the next death occurs at t=3.7, the proportion surviving to t=3.7 will then be calculated, and cumulative survival up to t=3.7 will be the product of these.
Table 44: Example dataset 1 - Kaplan-Meier cumulative survival

<table>
<thead>
<tr>
<th>Interval</th>
<th>No. at risk at start of interval</th>
<th>No. censored during interval</th>
<th>No. at risk at the end of interval</th>
<th>No. who died during interval</th>
<th>Proportion surviving</th>
<th>Cumulative survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1</td>
<td>100</td>
<td>10</td>
<td>85</td>
<td>5</td>
<td>85/90 = 0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>1-2</td>
<td>85</td>
<td>15</td>
<td>67</td>
<td>3</td>
<td>67/70 = 0.96</td>
<td>0.94*0.96 = 0.90</td>
</tr>
<tr>
<td>2-3</td>
<td>67</td>
<td>12</td>
<td>53</td>
<td>4</td>
<td>53/55 = 0.96</td>
<td>0.90*0.96 = 0.86</td>
</tr>
<tr>
<td>3-4</td>
<td>53</td>
<td>10</td>
<td>41</td>
<td>2</td>
<td>41/43 = 0.95</td>
<td>0.86*0.95 = 0.82</td>
</tr>
</tbody>
</table>

Table 45: Example dataset 2 - Kaplan-Meier risk estimate (No competing risk)

<table>
<thead>
<tr>
<th>Interval</th>
<th>No. at risk at start of interval</th>
<th>No. censored during interval</th>
<th>No. at risk at the end of the interval</th>
<th>No. who had event during interval</th>
<th>Proportion experiencing event</th>
<th>Cumulative risk estimate (1-KM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1</td>
<td>100</td>
<td>10</td>
<td>85</td>
<td>5</td>
<td>1- (85/90) = 1- 0.94 = 0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>1-2</td>
<td>85</td>
<td>15</td>
<td>67</td>
<td>3</td>
<td>1- (67/70) = 1- 0.96 = 0.04</td>
<td>1- (0.94*0.96) = 0.10</td>
</tr>
<tr>
<td>2-3</td>
<td>67</td>
<td>12</td>
<td>53</td>
<td>4</td>
<td>1- (53/55) = 1- 0.96 = 0.04</td>
<td>1- (0.90*0.96) = 0.14</td>
</tr>
<tr>
<td>3-4</td>
<td>53</td>
<td>10</td>
<td>41</td>
<td>2</td>
<td>1- (41/43) = 1- 0.95 = 0.05</td>
<td>1- (0.86*0.95) = 0.18</td>
</tr>
</tbody>
</table>
When a patient is censored, Kaplan-Meier estimates a potential contribution to the proportion surviving for that patient and redistributes it across those patients still at risk. It does this by removing censored patients from both the numerator and denominator of the proportion surviving calculation. This works on the assumption that the censored patients are as likely to die or survive as the patients who have not been censored. For example- using the data in table 48 the proportion surviving at t=1 is calculated as:

\[
\text{Proportion surviving} = \frac{(\text{No. alive at end of interval} - \text{no. censored patients})}{(\text{no. alive at start of the interval} - \text{no. censored patients})}
\]

\[
\text{Proportion surviving} = \frac{(100-5-10)}{(100-10)}
\]

\[
= \frac{85}{90}
\]

\[
= 0.94 \text{ or } 94%
\]

By removing the number of censored patients from the number alive at the start of the interval rather than just the number alive at the end, the proportion surviving is inflated to account for the fact that a proportion of censored patients are also likely to still be alive. If the number of censored patients was only removed from the number of patients alive at the end of the interval, the calculation would be inferring that all censored patients had died and so the proportion surviving each interval, and in turn the cumulative survival, would be reduced. For example, for the interval t=0 to t=1: (see table 49)

\[
\text{Proportion surviving} = \frac{(\text{No. alive at end of interval} - \text{no. censored patients})}{(\text{no. alive at start of the interval})}
\]

\[
\text{Proportion surviving} = \frac{[(100-5-10)/100]}
\]

\[
= \frac{85}{100}
\]

\[
= 0.85 \text{ or } 85%
\]

5.3.2.2 Discussion of Kaplan-Meier risk estimate methodology

The Kaplan-Meier risk estimate works principally in exactly the same way as the Kaplan-Meier survival function except the cumulative risk estimate is calculated by subtracting the proportion experiencing the event of interest in a given time period
from 1. So using the data from table 50 the proportion experiencing event during the interval t=0 to t=1 would be calculated as:

\[
\text{Proportion experiencing event} = 1 - \frac{(\text{No. at risk of event at end of interval} - \text{no. censored patients})}{(\text{No. at risk of event at start of the interval} - \text{no. censored patients})}
\]

\[
= 1 - \frac{(100-5-10)}{(100-10)}
\]

\[
= 1 - \frac{85}{90}
\]

\[
= 1 - 0.94
\]

\[
= 0.06 \text{ or } 6\%
\]

Again, by subtracting the censored patients from both the denominator and the numerator, the proportion of patients experiencing the event of interest is inflated to take into account the fact that a proportion of the censored patients are likely to experience the event of interest at some point in time after the point of censoring but before the end of the interval. This again works on the assumption that the censored patients are as likely to experience the event of interest as the patients who have not been censored.
Table 46: Example dataset 2 - Kaplan-Meier risk estimate (Competing risk of death)

<table>
<thead>
<tr>
<th>Interval</th>
<th>No. at risk at start of interval</th>
<th>No. censored during interval</th>
<th>No. at risk at the end of the interval</th>
<th>No. who had event during interval</th>
<th>Proportion experiencing event</th>
<th>Cumulative risk estimate (1-KM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1</td>
<td>100</td>
<td>10 total</td>
<td>85</td>
<td>5</td>
<td>85/90 = 0.94</td>
<td>1-0.94 = 0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 censored</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 deaths</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2</td>
<td>85</td>
<td>15 total</td>
<td>67</td>
<td>3</td>
<td>67/70 = 0.96</td>
<td>1- (0.94*0.96) = 1- 0.90 = 0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8 censored</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 deaths</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-3</td>
<td>67</td>
<td>12 total</td>
<td>53</td>
<td>4</td>
<td>53/55 = 0.96</td>
<td>1- (0.90*0.96) = 1- 0.86 = 0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6 censored</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6 deaths</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td>53</td>
<td>10 total</td>
<td>41</td>
<td>2</td>
<td>41/43 = 0.95</td>
<td>1- (0.86*0.95) = 1- 0.82 = 0.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6 censored</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 deaths</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
When there is one competing risk present in a study, there are three possible outcomes at each time point (e.g. t=x):

- patient experiences outcome of interest (e.g. mCRC) at t=x
- patient suffers competing risk event (e.g. death) without outcome of interest at t=x
- patient has suffered neither competing risk event nor the event of interest but has follow-up only to t=x

The second and third outcomes are distinct as the second outcome excludes the possibility of the first outcome occurring at some future time point. Censoring implies that the event of interest is still possible beyond the time point at which the censoring occurred. Patients who have suffered the competing risk event are no longer at risk of the event of interest so their contribution to the risk estimate should be 0. However, as explained above, when a patient is censored, Kaplan-Meier estimates a potential contribution to the risk estimate for that patient and redistributes it across those patients still at risk (by removing censored patients from both the numerator and denominator of the proportion experiencing event calculation).

In the example data in table 47, the total number of censored patients includes both patients who have suffered the competing risk and are therefore precluded from experiencing the event of interest in the future, and those whose are truly censored. Kaplan-Meier does not allow distinction between these two groups. By using the Kaplan-Meier risk estimate to calculate cumulative risk in a population that has been subject to competing risk, the cumulative risk is incorrectly inflated to account for both truly censored patients and patients censored due to competing risk. In this example, the cumulative risk estimate would be calculated exactly the same as for the example data in table 46. This cannot be accurate, as the dead patients in table 47 are no longer at risk of the event of interest and the cumulative risk should therefore be less. The greater the proportion of patients who have suffered the competing risk, the greater the overestimation of cumulative risk by the Kaplan-Meier risk estimate.

**4.4.1.2.2 Discussion of cumulative incidence function methodology**

Cumulative incidence risk analysis, as described by Gray et al (Gray RJ 1988) and by Gooley et al (Gooley TA 1999), allows the estimation of cumulative incidence
taking into account competing risk(s). The cumulative incidence estimates a proportion of all at risk individuals who actually develop a mCRC i.e. at the starting time-point. In the current study, the competing risks of death due to the initial colorectal cancer and death due to other causes are both significant. 57.0% of the Lynch Syndrome group, 48.6% of the moderate familial risk group and 68.0% of the population risk group have died during the period of follow-up.

As Gooley et al describe, most commercially available statistical software packages include a function to calculate the Kaplan-Meier risk estimate but not a function that allows for consideration of competing risk (Gooley TA 1999). This probably contributes to why Kaplan-Meier risk estimate is often inappropriately used. Coviello and Bogess describe a program that can be written in Stata software. In this study Stata version 11 (Stata Corps, College Station, Texas, USA) was used. The cumulative incidence function was written in syntax as presented by Coviello and Boggess (Coviello V 2004). This programme adjusts the cumulative risk to include a potential contribution from patients who have been truly censored, yet removes those patients who have suffered from the competing risk (in this case death) from that contribution. This programme is able to produce both a Kaplan-Meier risk estimate and a cumulative incidence function on the same graph for direct comparison (see figure 36).
As time goes on, and a greater proportion of patients die, so the Kaplan-Meier risk estimate curve and the cumulative incidence function curve diverge. Figure 36 demonstrates the difference between the Kaplan-Meier risk estimate and the cumulative incidence estimate in patients with moderate familial risk in this study.

The Kaplan-Meier risk estimate calculates the probability of an event occurring at given time points. Cumulative risk is the product of risk calculated for each subsequent time interval. The Kaplan-Meier estimate is appropriate if there is no competing risk(s), or if the occurrence is small. Thus for a person with a Duke’s A or B CRC Kaplan-Meier estimate is likely to be very relevant. Also in genetic counselling many cases of CRC are seen years after their original diagnosis and their survival chances are thus much improved. As such the Kaplan-Meier figure is most relevant. If undertaking a health economics analysis of effectiveness, the cumulative incidence function is most useful. Many individuals at time of diagnosis will not survive to develop a predicted mCRC and so would not benefit from surveillance to prevent it. Therefore as the cumulative incidence function predicted risk is elevated, a surveillance programme may still be beneficial.
Table 47: Cumulative risk of developing a mCRC in patients with moderate familial risk

<table>
<thead>
<tr>
<th>Risk estimate</th>
<th>5 years</th>
<th>10 years</th>
<th>20 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaplan-Meier risk estimate</td>
<td>2.7%</td>
<td>6.3%</td>
<td>23.5%</td>
</tr>
<tr>
<td>OR moderate risk vs population risk (1-KM)</td>
<td>2.1</td>
<td>2.0</td>
<td>3.4</td>
</tr>
<tr>
<td>Cumulative Incidence estimate</td>
<td>1.2%</td>
<td>2.5%</td>
<td>9.6%</td>
</tr>
<tr>
<td>OR moderate risk vs population risk (CI)</td>
<td>4</td>
<td>4.2</td>
<td>4</td>
</tr>
</tbody>
</table>

In any screening programme it is important to consider the risks of the screening procedure against the benefit. Colonoscopy, and the bowel preparation medication required before the procedure carry not inconsiderable risks. Bowel perforation rate after polypectomy is 22 (95% CI 13.85-33.3) per 10,000, post-polypectomy bleeding occurs in 89 (95% CI 71.5-109.5) per 10,000 cases and mortality (in specialist centres) is 0.83 (95% CI 0.025-3.69) per 10,000 colonoscopies. (Cairns, Scholefield et al. 2010) A cross-sectional study of endoscopy units within the UK revealed a much higher rate of death of 6 per 10,000 that could be directly attributed to colonoscopy. (Bowles, Leicester et al. 2004) The BSG screening guidelines quote a complication rate of 0.3% perforation, 0.3% bleeding and 0.02% death caused by colonoscopy.

### 4.4.2 Discussion of survival

This study has demonstrated that the actuarial survival of patients with a moderate risk colorectal cancer is 7.5 years greater than those with a Lynch Syndrome CRC. The data has also demonstrated that actuarial survival is 7.4 years greater in patients with a CRC associated with background population risk than those with a moderate risk CRC. That is 63.0 years (95% CI 61.0-65.0) in patients with Lynch Syndrome, 70.5 years (95% CI 67.7-73.7) in patients with moderate familial risk and 77.9 years (95% CI 75.5-80.3) in patients with population risk. These differences are significant (Logrank Mantel-Cox p<0.001). This is clearly due to the
differences in the ages of onset. The median age of onset of first CRC was 15 years younger in the moderate risk group compared to the population risk group and 9.8 years younger in the Lynch Syndrome group compared to the moderate familial risk group (p<0.001, Lynch syndrome 45.5 years, moderate familial risk 55.3 years, population risk 70.3 years).

The five year survival from first cancer was similar between all the groups (58.0% for Lynch Syndrome, 58.7% for moderate risk, and 55.4% for population risk) but median survival (in years) from first cancer was less in the population risk group indicating, as expected, higher risk of death from other causes. Median survival from first cancer was 10.5 years in the Lynch Syndrome group, 9.1 years in the moderate familial risk group, and 6.8 years in the population risk group. The distribution of Dukes staging in the Lynch group demonstrated a shift towards better prognosis tumours as expected for a population in whom screening occurs (A 21.9%, B 64.9%, C 10.3%, D 3.9%). The distribution of Dukes stage was similar in the moderate familial risk group to that quoted by Cancer Research UK for the general population. 15.8% were Dukes A (13.2% in the general population), 36.8% were Dukes B (3.8% in the general population), 43.9% were Dukes C (35.9%) and 3.5% were Dukes D (9.4% in the general population).

It would be expected for both the overall survival and median survival from diagnosis of first cancer in the Lynch Syndrome group to be better/greater than has in fact been demonstrated. This assumption is partly because of the greater proportion of lower stage/better prognosis (i.e, Dukes stage A and B) cancers normally associated with Lynch syndrome (due to screening) and also because cancers with mismatch repair deficiency (either due to Lynch Syndrome or sporadic loss of MLH1) are thought to have a better prognosis than mismatch repair proficient cancers. Sinicrope et al investigated 2041 patients from a variety of adjuvant chemotherapy trials and found the hazard ratio for overall survival to be 0.79 (95% CI 0.4-0.99, p=0.03) for MMR proficient cancers compared to deficient cancers. (Sinicrope, Foster et al. 2011) This anomaly can be explained by the historical distribution of first cancer cases. In this study 169/528 (32.0%) of first cancers in the Lynch Syndrome group occurred before 1980 whereas only 12/383 (3.1%) of the moderate familial risk group and 6/409 (1.5%) of the population risk group were diagnosed prior to 1980. Since 1980s the prognosis for all CRCs has improved due to developments in operative and adjuvant management. In additio
to this the screening practice for patients with Lynch Syndrome has changed considerably over this period of time. The first Amsterdam Criteria were published in 1991, predictive testing for the gene mutations became possible after 1993, and in 2002 national guidelines were published.

4.4.3 Conclusion

This study has quantified the risk of metachronous colorectal cancer in patients with a moderate family history of colorectal cancer. This risk is sufficiently high to justify ongoing colonoscopy screening, with five yearly screening likely to be adequate, but does not justify a change in practice from standard segmental resection surgery at presentation with the first colorectal cancer.

4.4.4 Further work

An audit of screening outcomes in patients with a moderate familial risk of colorectal cancer is needed to assess the benefit of surveillance in this group. This data is now being collected in the department of genetic medicine at St Mary’s Hospital.
References:


Newton, K. and J. Hill (2010). "5-FU and mismatch repair deficient colorectal cancer: is it time to consider a change in practice?" *Colorectal Dis*.


Appendix one: Immunohistochemistry Methods.

Haematoxylin and Eosin Slides

1 Materials

Xylene

Industrial methylated spirits (IMS)

Haematoxylin

Eosin

Xam mounting media

Coverslips (surgipath, Peterborough, UK)

2 The mounted 4µm sections were dewaxed through 3 sequential xylene washes for 5 minutes each.

2 The sections were rehydrated through a series of 4 IMS washes for 2 minutes each and then rinsed under running water for 5 minutes.

3 The sections were placed in haematoxylin stain for 5 minutes then rinsed under running water for 5 minutes.

4 The sections were placed in Eosin stain for 1 minute.

5 The sections were dehydrated through a series of IMS washes (4 x 2 minutes) and cleared in xylene for 3 x 5 minutes.

6 Coverslips were mounted with Xam and left to dry.

DAB automated mismatch repair protein Immunohistochemistry

Mismatch repair protein immunohistochemistry method

A block containing both tumour tissue and normal tissue (to be used as an internal positive control) for each case and control were stained with antibodies against the mismatch repair proteins MLH1, MSH2 and MSH6. A known negative control was used in each run. This was performed on the automated platform Roche Benchmark XT IHC/ISH staining machine. This platform is a fully automated system. The
detection kit used for this platform is the XT ultraView DAB v3 (Roche). This system uses a peroxidase linked polymer and DAB substrate. The protocol is used in the Manchester Royal Infirmary histopathology laboratory as a clinical test thus has been previously optimised.

1 Materials

Provided in XT ultraView DAB (Roche) kit- preloaded in cartridges

- EZPrep
- Depar volume adjust
- Cell conditioner #1
- Reaction buffer
- UV Inhibitor
- Amplifier B

Not provided in kit

- Thermoscientific superfrost plus slides
- Cell path coverslips
- Pertex mounting medium
- Xylene
- 99% IMS

Mouse anti-human primary antibody diluted in antibody diluent (Dako, Cambridgeshire, UK)

Mouse anti-human MLH1 antibody (NCL/L/MLH1 Lieca/Novocastra Newcastle, UK)

Mouse anti-human MSH2 antibody (NCL/L/MSH2 Lieca/Novocastra Newcastle, UK)

Mouse anti-human MSH6 antibody (NCL/L/MSH6 Lieca/Novocastra Newcastle, UK)
2 4µm thick tumour sections were mounted onto slides and labelled and barcoded.

3 The slides were heated at 60°C for 30 minutes.

4 Deparafinization was performed in xylene.

5 The slides were then loaded into the Benchmark XT IHC/ISH staining machine. The machine is pre-loaded with pre-filled reagent cartridges.

6 The appropriate protocol was selected on the automated platform, and the protocol initiated. Stains for MLH1, MSH2 and MSH6 were performed separately for each patient.

7 Antigen retrieval was performed on the machine- heat treatment for 30 mins in a high pH antigen retrieval solution called cell conditioning 1 (CC1).
   • MLH1 1:50 for 1hr20 mins
   • MSH2 1:40 for 1hr 20mins
   • MSH6 1:500 for 48 mins

8 Incubation with primary antibodies and detection system, and counter staining with haematoxylin were performed by the platform.

9 The slides were removed from the platform and rinsed with water.

10 The slides were dehydrated with 99% IMS.

11 The slides were passed through three washes of 100% xylene and mounted.
Appendix two: Protocol for microsatellite instability assay

DNA Extraction

DNA was extracted from FFPE tissue using the *Qiagen EZ1 robot* in conjunction with the *Tissue Extraction kit (Qiagen, UK)*. DNA was extracted in this manner to use for MSI, *BRAF* and *KRAS* analysis.

1 Materials  BioRobot EZ1

    *Provided in Qiagen Tissue Extraction Kit (Qiagen, UK)*

    G2 Buffer

    2ml samples tubes

    EZ1 paraffin section card

    *Not provided in the kit*

    Proteinase K solution

    1.5ml screw cap tubes

    Elution tubes

2  8-9 x 10µm FFPE tissue sections per sample (one sample of tumour tissue and one of normal tissue per patient) were transferred into labelled 1.5ml screw cap tubes.

3  190µl of G2 Buffer was added to each tube to submerge the tissue.

4  The tubes were incubated at 75°C on a shaking hot block for 5 minutes.

5  The hot block temperature was lowered to 56 °C and the samples were allowed to cool to 56 °C.

6  10µl of proteinase K was added to each sample and left incubate on the shaking hotblock overnight at 56 °C.

7  The samples were briefly mixed by centrifugation and transferred to 2ml sample tubes (leaving any wax behind).
If there was any undigested tissue remaining the Proteinase K digestion step was repeated. 

The BioRobot EZ1 was initiated and the EZ1 paraffin section card inserted into the card slot of the machine. 

Instructions on the screen of the robot were followed. 

Proper set up of the samples and elution tubes was checked and signed for by another person. 

The protocol was initiated and after 20 minutes the eluted DNA was ready. 

A NanoDrop® Spectrophotometer reading was taken to check the concentration of the eluted DNA using 2µl of the product. The DNA concentration ranged from 8-40ng/µl. 

DNA was stored at -20°C until required for PCR amplification. 

**DNA Amplification** 

1 Materials Provided in *MSI Analysis Version 1.2 (Promega, USA)*

MSI 10X primer pair mix (100µl) 

Gold STAR 10X Buffer (300µl) 

Nuclease-free water (1.25ml) 

3µg K562 High Molecular Weight DNA (10ng/µl) 

Not provided in Kit 

FFPE derived DNA extracted using the *Qiagen Tissue Extraction Kit (Qiagen, UK)* and BioRobot EZ1 as described in section 2.2.5.1.1 

Thermal cycler 

96 well reaction plates 

1.5ml microtrefuge tubes 

*AmpliTaq Gold®* DNA polymerase (*Applied Biosystems)*
The Gold STAR 10X buffer and MSI system analysis were thawed to room temperature.

These were mixed by vortexing for 10 seconds.

Reagents were added to each well needed of a 96 well plate:
- 5.85µl Nuclease-free water
- 1µl Gold STAR 10X Buffer
- 1µl MSI Analysis System 10X Primer Pair Mix
- 0.15µl AmpliTaq Gold® DNA polymerase (5u/µl)

2µl of template (sample) DNA was added to the relevant well.

The plate was mixed by vortexing and centrifuged.

A positive control reaction was created using K562 genomic DNA from the MSI Analysis Version 1.2 (Promega, USA) kit.

The thermal cycler was programmed as in table below.
PCR conditions for MSI assay

<table>
<thead>
<tr>
<th>STEP</th>
<th>TIME</th>
<th>TEMPERATURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initialization</td>
<td>11 minutes</td>
<td>95°C</td>
</tr>
<tr>
<td>Initialization</td>
<td>1 minute</td>
<td>96°C</td>
</tr>
<tr>
<td>Annealing</td>
<td>30 seconds</td>
<td>94°C</td>
</tr>
<tr>
<td>Extension</td>
<td>30 seconds</td>
<td>58°C</td>
</tr>
<tr>
<td>Final extension</td>
<td>1 minute</td>
<td>70°C</td>
</tr>
<tr>
<td>Annealing</td>
<td>30 seconds</td>
<td>90°C</td>
</tr>
<tr>
<td>Extension</td>
<td>30 seconds</td>
<td>58°C</td>
</tr>
<tr>
<td>Final Extension</td>
<td>1 minute</td>
<td>70°C</td>
</tr>
<tr>
<td></td>
<td>30 minutes</td>
<td>60°C</td>
</tr>
<tr>
<td>Soak</td>
<td>Indefinite</td>
<td>4°C</td>
</tr>
</tbody>
</table>

9 The samples were stored at -20°C until needed.

Detection of amplified fragments using the ABI PRISM® 3100 genetic analyser

10 Materials Provided in MSI Analysis Version 1.2 (Promega, USA)

- Internal Lane Standard (ILS 600)
- Not provided in the kit
- Thermal cycler
  
  Performance optimised polymer 4 (POP-4™) (Applied Biosystems)

  10X genetic analyzer buffer with EDTA (Applied Biosystems)

  Hi-Di™ Formamide (Applied Biosystems)

  PowerPlex® Matrix Standards 310
A mastermix of ILS600 and formamide was prepared:

- 1.0µl ILS600 and 24 µl per sample to be tested

- Vortex for 15 seconds

25 µl of mastermix and 1 µl of PCR product was added to the wells of a clean 96 well plate.

The samples were denatured by heating them to 95°C and then immediately cooled on ice.

The plate was then placed into the *ABI PRISM® 310* and the instrument closed.

The pre-programmed GeneMapper software was employed.
Appendix three: Protocol for KRAS and BRAF mutation assay

DNA extraction

See appendix two

Polymerase chain reaction

PCR reactions were set up up for each sample in triplicate for each mutation being analysed. Therefore for each sample there was three PCR reactions set up for KRAS codons 12/13, and three PCR reactions for BRAF codon 600.

For each sample reactions were set up in triplicate for each assay, in a 25µl reaction. Buffers and primers were supplied in the Pyromark Q96 KRAS v2.0 kit (Qiagen, UK).

1. Materials

Supplied in the Pyromark Q96 KRAS v2.0 kit (Qiagen, UK)

CM129 Buffer

Codon 12/13 Forward Primer

Codon 12/13 Biotinylated Reverse Primer

Not supplied in the kit

Distilled water

96 well PCR plate

TBE dye

2% agarose gel

A test ladder of 100bp
2  *Codon 12/13*: A master mix of buffer and primers was prepared. For each reaction the following volume was required:

- 12.5µl CM129 buffer
- 0.5µl 12/13 Forward primer
- 0.5µl 12/13 Biotinylated reverse primer
- 9µl distilled water

3  *BRAF V600e*: A master mix of buffer and primers was prepared. For each reaction the following volume was required:

- 22.5µl of the appropriate master mix was added to a well in a clean 96 well PCR plate.

  For each sample 2.5µl of the DNA extracted from the FFPE tissue (at approximately 30ng/µl) was added to three wells.

3  For codon *KRAS* 12/13 a positive control with a known mutation and a normal control was also added in triplicate

4  For *BRAF V600E* a positive control with a known mutation and a normal control was also added in triplicate

5  For each codon specific assay a water blank was added.

6  The PCR was performed on a thermal cycler (*Veriti 96 well thermal cycler, AB Applied Biosystems*) pre-programmed with the cycle below.
PCR conditions for KRAS/BRAF mutation assays

<table>
<thead>
<tr>
<th>STEP</th>
<th>TIME</th>
<th>TEMPERATURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initialization</td>
<td>15 mins</td>
<td>95°C</td>
</tr>
<tr>
<td>Denaturation</td>
<td>20 seconds</td>
<td>95°C</td>
</tr>
<tr>
<td>Annealing</td>
<td>30 seconds</td>
<td>53°C</td>
</tr>
<tr>
<td>Extension</td>
<td>20 seconds</td>
<td>72°C</td>
</tr>
<tr>
<td>Final extension</td>
<td>5 minutes</td>
<td>72°C</td>
</tr>
<tr>
<td>Hold</td>
<td>Overnight or as required</td>
<td>8°C</td>
</tr>
</tbody>
</table>

Repeated for 38 cycles

The PCR products were monitored using electrophoresis on a 2% agarose gel.

4.5µl of PCR product and 2µl of TBE dye (for each PCR reaction) was added to the wells in the 2% agarose gel. This was run at 200volts for 3 minutes. The amplicons were examined visually and with the UV camera to ensure visible bands of DNA. The PCR product for codon 12/13 is 122bp. This was compared to the 100bp ladder.

If the amplicons were visible and the water control was blank - I proceeded to pyrosequencing.

The PCR products were stored at -20°C as required.

2.2.5.2.3 Pyrosequencing

The KRAS codon 12/13 and BRAF V600 pyrosequencing assays were run in parallel. Usually three patients were analysed together in each run of the assays.

1. Materials
   - Pyrosequencer (PSQ 96MA) and connected computer
   - Pyrosequencing cassette
   - ABgene 96 well plate
   - Pyrosequencing plate
Dyad incubating block
70% ethanol
0.2M NaOH
MilliQ Water
Wash buffer
Supplied in the *Pyromark Q96 KRAS v2.0 kit* (Qiagen, UK)
Streptavidin beads
Binding buffer
Enzyme Mix
Substrate Mix
cNTP, aNTP, tNTP, uNTP nucleotides

2 The reagents were taken out of the fridge to equilibrate to room temperature.

3 The pyrosequencing cassette was examined and washed with MilliQ water to ensure that each injection hole within the cassette was patent and any blockages cleared.

4 The prosequencer and computer were turned on.

5 On the computer the *PSQ 96MA* icon was opened and the pyrosequencing programme initiated. “New SNP run” was selected. Details of the sample numbers were entered. Instrument parameters 0002 Pyrogold were selected. The appropriate primer kit numbers were selected from a drop down list

   ‘KRAS_Bio_kit1’ for codon 12/13

   ‘BRAF_custom_primers_codon600’ for *BRAF* codon 600

6 ‘Run view’ was selected on the programme and the volumes of enzyme mix, substrate mix and of each nucleotide from the *Pyromark Q96 KRAS v2.0 kit* (Qiagen, UK) were calculated by the programme, and were noted down.

7 A mastermix of streptavidin beads and binding buffer was made. 3μl of beads and 37μl of buffer per reaction. This was mixed by inversion.
8. 40µl of mastermix was aliquoted into each well, as required, of a ABgene 96 well plate.

9. 20µl (all the remaining PCR product) was added to the aliquoted mastermix. This was sealed with a sticky lid and put on the shaking mixing plate for at least 5 minutes.

10. 40µl of the appropriate sequencing primers were added to the wells of the pyrosequencing plate.

11. The Pyrosequencer was prepared by adding 70% ethanaol, wash buffer, MilliQ water and 0.2M NaOH to the trays next to the prep tool as labelled. The pump was turned on.

12. The prep tool was washed by placing it in the tray of MilliQ water for 20 seconds (the pump is attached to the prep tool so liquid is pulled through the tool).

13. The tool was then placed into the plate containing the PCR products and bead mix and left until all the liquid was aspirated up onto the prongs of the tool. A white coating was visible on the end of the prongs at this point.

14. Immediately the tool was washed in ethanol for at least 5 seconds.

15. The tool was placed in the wash buffer for another 5 seconds and the tool was elevated to ensure all liquid had drained through.

16. The pump was turned off

17. The prep tool was placed carefully into the plate containing the sequencing primers (in the correct orientation) and the tool was wiggled to ensure that all the beads (and hence attached DNA) were released into the primer solution.

18. The prep tool was removed from the primer plate and was washed with water thoroughly.

19. The pyrosequencing plate was placed on the heated Dyad block at 80 °C for 2 minutes in order to denature the DNA.

20. The pyrosequencing cassette was filled with the reagents as specified in step 8 (enzyme mix, substrate mix and of each nucleotide from the Pyromark Q96 KRAS v2.0 kit (Qiagen, UK)).
The pyrosequencing plate was removed from the Dyad block and placed in
the appropriate place on the pyrosequencer. The cassette was also added.
The clamp was pulled down.

‘Run’ was clicked in the software.

When the run was completed the plate was disposed of and the cassette
rinsed out and left to dry. The trays next to the prep tool were washed with
water and left to dry.

Data analysis: The PSQ 96MA programme was used to analyse the data.
Pyrograms were created and saved.
Appendix four: Protocol for MLH1 promoter region methylation assay

Extraction of DNA from FFPE tissue, Bisulphite assisted conversion of unmethylated cytosines in DNA and purification of converted DNA.

Sample lysis (DNA extraction) and bisulphite assisted conversion of unmethylated cytosines was performed using the EpiTect plus FFPE Bisulfite kit® (Qiagen, UK). The EpiTect Plus bisulfite conversion procedure has seven steps which result in purified, bisulphite converted DNA ready for a PCR reaction. These steps are preparation of DNA from the sample (in this case FFPE tissue), bisulphite-mediated conversion of unmethylated cytosine residues, binding of the converted single-stranded DNA to the membrane of an MinElute spin column, washing, desulphonation of the membrane-bound DNA, washing of the membrane-bound DNA to remove the excess desulphonation agent, and elution of the pure, clean DNA from the spin column.

Sample lysis (DNA extraction)

1. **Materials**
   Provided in the EpiTect plus FFPE Bisulphite kit® (Qiagen, UK):
   - Deparaffinization solution
   - Lysis buffer FTB
   - Proteinase K

   *Not provided in kit:*
   - Distilled water
   - 1.ml ependorffs

2. Two 1.5ml ependorffs (tubes) were labelled for each patient: 1 normal and 1 tumour.

3. One 10µm thick FFPE tissue section was placed in each ependorff.

4. 150µl of Deparaffinization solution was added to each ependorff.
Each ependorff was vortexed to ensure all the paraffin was dissolved.

20µl distilled water, 15µl Lysis Buffer FTB and 5µl of proteinase K was added to each ependorff.

Each ependorff was vortexed and centrifuged.

The ependorffs were incubated overnight at 56°C on a shaking heated block.

This step was altered from the kit protocol. The protocol describes a 30-60 minute incubation step. When initially tested, this short incubation period did not result in complete lysis of the tissue on visual inspection. A trial run of overnight was performed and this resulted in adequate lysis.

The ependorffs were incubated at 90°C for 60 minutes on a shaking heated block to facilitate decrosslinking of DNA that occurs due to formalin fixation.

**Bisulphite assisted conversion of unmethylated cytosines and purification of converted DNA**

1 Materials  *Provided in the EpiTect plus FFPE Bisulphite kit® (Qiagen, UK).*

Bisulphite mix (800µl of RNAase-free water was added to each aliquot of bisulphite mix to make up the solution)

DNA protect buffer (designed to prevent the fragmentation of DNA which is usually associated with bisulphite treatment of DNA which requires high temperatures and a low pH)

*Not provided in the kit*

CpGenome Universal Methylated DNA control (*Millipore™, cat no. S7821*)

200µl PCR tubes

2 The deparaffinization solution was removed from the ependorffs containing the lysis reactions. This solution had formed a layer above the lysis reaction.

3 The lysis reactions (volume approx 40µl) were transferred to 200µl PCR tubes.
85µl Bisulphite mix was added to each tube.

15µl DNA protect buffer was added to each tube

A positive control reaction was added at this point. This was made up of 5µl universally methylated DNA (CpGenome Universal Methylated DNA control Millipore™, cat no. S7821), 35 µl RNAase-free water, 85 µl bisulphite mix, and 35 µl DNA protect buffer.

The reactions were mixed by vortexing and brief centrifugation (the DNA protect buffer turned from green to blue which indicated sufficient mixing and the correct pH for the bisulphite reaction).

The DNA bisulphite conversion was performed using a thermal cycler. The complete cycle took approximately 5 hours (this was performed overnight) and the reaction was then maintained at 20°C.

**Thermal cycling conditions for bisulphite conversion**

<table>
<thead>
<tr>
<th>STEP</th>
<th>TIME</th>
<th>TEMPERATURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denaturation</td>
<td>5 mins</td>
<td>95°C</td>
</tr>
<tr>
<td>Incubation</td>
<td>25 mins</td>
<td>60°C</td>
</tr>
<tr>
<td>Denaturation</td>
<td>5 mins</td>
<td>95°C</td>
</tr>
<tr>
<td>Incubation</td>
<td>85 mins</td>
<td>60°C</td>
</tr>
<tr>
<td>Denaturation</td>
<td>5 mins</td>
<td>95°C</td>
</tr>
<tr>
<td>Incubation</td>
<td>175 mins</td>
<td>60°C</td>
</tr>
<tr>
<td>Denaturation</td>
<td>5 mins</td>
<td>95°C</td>
</tr>
<tr>
<td>Incubation</td>
<td>120 mins</td>
<td>60°C</td>
</tr>
<tr>
<td>Hold</td>
<td>Indefinite (usually overnight)</td>
<td>20°C</td>
</tr>
</tbody>
</table>

**DNA purification and elution**

Materials Provided in the EpiTect plus FFPE Bisulphite kit® (Qiagen, UK).

RNAase-free water
MinElute spin columns (tubes with a membrane designed for catching DNA during the wash process)

Collection tubes (2ml)

Buffer BL (promotes binding of the single-stranded DNA to the membrane in the MinElute spin columns)

Buffer BD (desulphonation buffer)

Buffer BW (a wash buffer which is designed to remove any residual sodium bisulphite from the DNA after conversion)

Buffer EB (a buffer designed to elute the single-stranded DNA from the membrane of the spin columns)

Carrier RNA (used to enhance binding of small fragments of DNA to the membrane on the spin columns is added to the Buffer BL)

Following the bisulphite assisted conversion reaction, the samples were briefly mixed by centrifugation and the reactions transferred to clean 1.5ml tubes.

310µl of Buffer BL and 3.5µl of carrier RNA was added to each sample. This was mixed for 15 seconds using vortex and centrifugation.

250µl of 100% ethanol was added to each sample. This was mixed for 15 seconds using a vortex and centrifugation.

The samples were then transferred into MinElute DNA spin columns which were held in 2ml collection tubes.

The samples were mixed by centrifugation in the spin columns at maximum speed (>1300G) for 1 minute and the flow through was discarded.

500µl Buffer BW (wash buffer) was added to each spin column and mixed by centrifugation at maximum speed for 1 minute. The flow through was discarded.

500µl Buffer BD (desulphonation buffer) was added to each spin column and incubated at room temperature for 15 minutes, then mixed by centrifugation maximum speed for 1 minute. The flow through was discarded.
500µl Buffer BW (wash buffer) was added to each spin column and mixed by centrifugation at maximum speed for 1 minute. The flow through was discarded. This step was repeated once.

250µl of 100% ethanol was added to each spin column and mixed by centrifugation at maximum speed for 1 minute.

The spin columns were placed into clean 2ml collection tubes and mixed by centrifugation at maximum speed for 1 minute to remove any excess liquid.

The DNA was then eluted from the spin columns. The spin columns were placed into clean 1.5ml centrifuge tubes (ependorffs).

15µl Buffer EB (elution buffer) was added directly onto the centre of the membrane in each spin column.

The spin columns were incubated at room temperature for 1 minute.

The spin columns were mixed by centrifugation at maximum speed for 1 minute to elute the DNA.

The purified converted DNA was stored at -20°C until needed for amplification in the PCR reaction.

Polymerase chain reaction

A polymerase chain reaction is used to amplify the single-stranded bisulphite converted DNA. The Pyromark™ Q24 MLH1 kit (Qiagen, UK) was used. Custom made primers were used (see figure 16). The assay is designed to amplify the target region of -209 to -181 from the transcription start site of the MLH1 gene, but using the custom made PCR primers, the assay amplified -248 to 178. During assay optimisation, it was found that a single PCR resulted in too dilute DNA for reliable sequencing. A nested, or two stage, PCR was tested and proved to result in an better concentration of DNA for reproducible pyrosequencing. Both stages of the nested PCR use 25 µl reactions.

The pyrosequencing assay tests each sample in triplicate. This allows confident analysis. At this stage, the PCR reactions are performed in triplicate per sample. A triplicate reaction containing universally methylated DNA was used as a positive
control. A single reaction containing only mastermix and water was used as a negative control.

1 Materials

Provided in the Pyromark™ Q24 MLH1 kit (Qiagen, UK)

Forward MLH1 PCR primer at 5µM concentration: used for primary and secondary PCR

Reverse MLH1 PCR primer at 5µM concentration (biotinylated; which is necessary for pyrosequencing): used for secondary PCR

CM-129 Annealing buffer (Qiagen, UK)

Distilled water (dH₂O)

Not provided in the kit

Reverse MLH1 PCR custom prom ext primer at 5µM concentration: used for primary PCR

Purified DNA

96 well PCR plate

2 Primary PCR: 12.5µl of CM-129 annealing buffer, 7.5µl dH₂O and 2.5µl of a mixture of the forward and reverse primers (each at 5µM concentrations) was added to the required number of wells on a 96 well PCR plate. (Forward MLH1 PCR primer + Reverse MLH1 PCR custom prom ext primer).

3 2.5µl of purified bisulphite converted DNA per sample was added to three wells. Three reactions containing purified converted universally methylated DNA were created. A reaction containing only mastermix and water was created as a negative control.

3 The primary PCR was performed on a thermal cycler machine which was pre-programmed to the cycle below.

4 Secondary PCR: 12.5µl of CM-129 annealing buffer, 7.5µl dH₂O and 2.5µl of a mixture of the forward and reverse primers (each at 5µM concentrations) was added to the required number of wells within a 96 well plate. (Forward MLH1 PCR primer + Reverse MLH1 PCR primer).

5 2.5µl of primary PCR product was added to each corresponding well.
6 The secondary PCR was performed on a thermal cycler machine which was pre-programmed to the cycle in table below.

**Primary and secondary PCR conditions for MLH1 promoter region methylation assay**

<table>
<thead>
<tr>
<th>STEP</th>
<th>TIME</th>
<th>TEMPERATURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>15 minutes</td>
<td>95°C</td>
</tr>
<tr>
<td>The cycle below was repeated 25 times during the primary PCR and 35 times during the secondary PCR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denaturation</td>
<td>30 seconds</td>
<td>95°C</td>
</tr>
<tr>
<td>Annealing</td>
<td>30 seconds</td>
<td>55°C</td>
</tr>
<tr>
<td>Extension</td>
<td>60 seconds</td>
<td>72°C</td>
</tr>
<tr>
<td>Final extension</td>
<td>5 minutes</td>
<td>72°C</td>
</tr>
<tr>
<td>At the end of each PCR reaction the products are held at 4°C as necessary.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Following secondary PCR, the products were tested by gel electrophoresis on a 2% agarose gel to test for the presence of amplified DNA.

4.5µl of PCR product and 2µl of TBE dye (for each PCR reaction) was added to the wells in a 2% agarose gel. This was run at 200volts for 3 minutes. The amplicons were examined visually and with the UV camera to ensure visible bands of DNA. The PCR product for region of interest of the MLH1 promoter region is 181bp. This was compared to the 100bp ladder.

If the amplicons were visible and the water control was blank - I proceeded to pyrosequencing.

The PCR products were stored at -20°C as required.
**Pyrosequencing**

The pyrosequencing assay sequences from -211 to -192 using a custom designed sequencing primer.

1. **Materials**
   - Pyrosequencer (PSQ 96MA) and connected computer
   - Pyrosequencing cassette
   - ABgene 96 well plate
   - Pyrosequencing plate
   - Dyad incubating block
   - 70% ethanol
   - 0.2M NaOH
   - MilliQ Water
   - 0.4Mm MLH1 promoter region
   - Wash buffer
   - Supplied in the *Pyromark™ Q24 MLH1 kit (Qiagen, UK)*
   - Streptavidin beads
   - Binding buffer
   - Enzyme Mix
   - Substrate Mix
   - cNTP, aNTP, tNTP, uNTP nucleotides

7. The reagents were taken out of the fridge to equilibrate to room temperature

8. The pyrosequencing cassette was examined and washed with MilliQ water to ensure that each injection hole within the kit was patent and any blockages cleared

9. The pyrosequencer and computer were turned on

10. On the computer the *PSQ 96MA* icon was opened and the pyrosequencing programme initiated. “New SNP run” was selected. Details of the sample
numbers were entered. Instrument parameters 0002 Pyrogold were selected. The appropriate primer kit numbers were selected from a drop down list

11 ‘Run view’ was selected on the programme and the volumes of enzyme mix, substrate mix and of each nucleotide from the Pyromark™ Q24 MLH1 kit (Qiagen, UK) were calculated by the programme, and were noted down.

12 A mastermix of streptavidin beads and binding buffer was made. For each reaction 3µl of beads and 37µl of buffer was added. This was mixed by inversion.

13 40µl of this mastermix was aliquoted into each well as required of a ABgene 96 well plate.

14 20µl (all the remaining secondary PCR product) was added to the aliquoted mastermix. This was sealed with a sticky lid and put on a shaking mixing plate for at least 5 minutes.

15 40µl of the appropriate sequencing primers were added to the wells of the pyrosequencing plate.

16 The Pyrosequencer was prepared by adding 70% ethanaol, wash buffer, MilliQ water and 0.2M NaOH to the trays next to the prep tool as labelled. The pump was turned on.

17 The preparation tool was washed by placing it in the tray of MilliQ water for 20 seconds (the pump is attached to the preparation tool so liquid is pulled through the tool).

18 The tool was then placed into the plate containing the PCR products and bead mix and left until all the liquid was sucked up onto the prongs of the tool. A white coating was visible on the end of the prongs at this point.

19 Immediately the tool was washed in ethanol for at least 5 seconds.

20 The tool was placed in the wash buffer for another 5 seconds and the tool was elevated to ensure all liquid had drained through.

21 The pump was turned off.

22 The preparation tool was placed carefully into the plate containing the sequencing primers (in the correct orientation) and the tool was wiggled to ensure that all the beads (and hence attached DNA) were released into the primer solution.
The prep tool was removed from the primer plate and was washed with water thoroughly.

The pyrosequencing plate was placed on the heated Dyad block at 80 °C for 2 minutes in order to denature the DNA.

The pyrosequencing cassette was filled with the reagents as specified in step 5 (enzyme mix, substrate mix and of each nucleotide from the Pyromark Q94 MLH1 kit (Qiagen, UK)).

The pyrosequencing plate was removed from the Dyad block and placed in the appropriate place on the pyrosequencer. The cassette was also added. The lid was then pulled down and secured.

‘Run’ was clicked in the software.

When the run was completed the plate was disposed of and the cassette rinsed out and left to dry. The trays next to the prep tool were washed with water and left to dry.

Data analysis: The PSQ 96MA programme was used to analyse the data. Pyrograms were created and saved for interpretation and genotyping.
### Appendix five: Data of Manchester sporadic colorectal cancer samples

<table>
<thead>
<tr>
<th>Lab no</th>
<th>Tumour Origin</th>
<th>IHC RESULT</th>
<th>KRAS Result</th>
<th>BRAF Result</th>
<th>MLH1 Methylation Result</th>
<th>MSI Result</th>
<th>Classified as</th>
</tr>
</thead>
<tbody>
<tr>
<td>10003506</td>
<td>sporadic</td>
<td>MSH2 loss</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in 5 markers</td>
<td>? Lynch Syndrome</td>
</tr>
<tr>
<td></td>
<td>untested</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10003562</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in 1 marker</td>
<td>MMR proficient</td>
</tr>
<tr>
<td></td>
<td>untested</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10003564</td>
<td>sporadic</td>
<td>All present</td>
<td>c.35G&gt;A p.Gly12Asp</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in 5 markers</td>
<td>MMR proficient</td>
</tr>
<tr>
<td></td>
<td>untested</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10003644</td>
<td>sporadic</td>
<td>All present</td>
<td>c.35G&gt;T p.Gly12Val</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in Penta D</td>
<td>MMR proficient</td>
</tr>
<tr>
<td></td>
<td>untested</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10003500</td>
<td>sporadic</td>
<td>All present</td>
<td>c.35G&gt;T p.Gly12Val (ARMS)</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td></td>
<td>untested</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10003501</td>
<td>sporadic</td>
<td>All present</td>
<td>c.35G&gt;T p.Gly12Val</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td></td>
<td>untested</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10003504</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td></td>
<td>untested</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10003505</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td></td>
<td>untested</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10003507</td>
<td>sporadic</td>
<td>All present</td>
<td>c.34G&gt;T p.Gly12Cys</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td></td>
<td>untested</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lab no</td>
<td>Tumour Origin</td>
<td>IHC RESULT</td>
<td>KRAS Result</td>
<td>BRAF Result</td>
<td>MLH1 Methylation Result</td>
<td>MSI Result</td>
<td>Classified as</td>
</tr>
<tr>
<td>---------</td>
<td>---------------</td>
<td>------------</td>
<td>---------------------</td>
<td>-------------</td>
<td>--------------------------</td>
<td>------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>10003508</td>
<td>sporadic</td>
<td>All present</td>
<td>c.35G&gt;T p.Gly12Val</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003509</td>
<td>sporadic</td>
<td>All present</td>
<td>c.38G&gt;A p.Gly13Asp</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003514</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003521</td>
<td>sporadic</td>
<td>All present</td>
<td>c.38G&gt;A p.Gly13Asp</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003523</td>
<td>sporadic</td>
<td>All present</td>
<td>c.38G&gt;A p.Gly13Asp</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003525</td>
<td>sporadic</td>
<td>All present</td>
<td>c.34G&gt;A p.Gly12Ser</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003538</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003539</td>
<td>sporadic</td>
<td>All present</td>
<td>c.35G&gt;A p.Gly12Asp</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003541</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003543</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003558</td>
<td>sporadic</td>
<td>All present</td>
<td>c.35G&gt;A p.Gly12Asp</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003559</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>Lab no</td>
<td>Tumour Origin</td>
<td>IHC RESULT</td>
<td>KRAS Result</td>
<td>BRAF Result</td>
<td>MLH1 Methylation Result</td>
<td>MSI Result</td>
<td>Classified as</td>
</tr>
<tr>
<td>--------</td>
<td>---------------</td>
<td>------------</td>
<td>-------------</td>
<td>-------------</td>
<td>--------------------------</td>
<td>------------</td>
<td>---------------</td>
</tr>
<tr>
<td>10003560</td>
<td>sporadic</td>
<td>All present</td>
<td>c.35G&gt;A p.Gly12Asp</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003561</td>
<td>sporadic</td>
<td>All present</td>
<td>c.34G&gt;T p.Gly12Cys</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003565</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003567</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003596</td>
<td>sporadic</td>
<td>All present</td>
<td>c.35G&gt;T p.Gly12Val</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003597</td>
<td>sporadic</td>
<td>All present</td>
<td>c.34G&gt;A p.Gly12Ser</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003598</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003599</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003600</td>
<td>sporadic</td>
<td>All present</td>
<td>c.35G&gt;A p.Gly12Asp</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003601</td>
<td>sporadic</td>
<td>All present</td>
<td>c.35G&gt;T p.Gly12Val</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003605</td>
<td>sporadic</td>
<td>All present</td>
<td>c.34G&gt;A p.Gly12Ser</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>Lab no</td>
<td>Tumour Origin</td>
<td>IHC RESULT</td>
<td><strong>KRAS Result</strong></td>
<td><strong>BRAF Result</strong></td>
<td>MLH1 Methylation Result</td>
<td>MSI Result</td>
<td>Classified as</td>
</tr>
<tr>
<td>--------</td>
<td>---------------</td>
<td>------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>--------------------------</td>
<td>------------</td>
<td>---------------</td>
</tr>
<tr>
<td>10003607</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003608</td>
<td>sporadic</td>
<td>All present</td>
<td>c.34G&gt;A p.Gly12Ser</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003609</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003610</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003611</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003613</td>
<td>sporadic</td>
<td>All present</td>
<td>c.35G&gt;A p.Gly12Asp</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003614</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003615</td>
<td>sporadic</td>
<td>All present</td>
<td>c.35G&gt;A p.Gly12Asp</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003631</td>
<td>sporadic</td>
<td>All present</td>
<td>c.35G&gt;T p.Gly12Val</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003632</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003633</td>
<td>sporadic</td>
<td>All present</td>
<td>c.35G&gt;T p.Gly12Val</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003634</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>Lab no</td>
<td>Tumour Origin</td>
<td>IHC RESULT</td>
<td>KRAS Result</td>
<td>BRAF Result</td>
<td>MLH1 Methylation Result</td>
<td>MSI Result</td>
<td>Classified as</td>
</tr>
<tr>
<td>--------</td>
<td>---------------</td>
<td>------------</td>
<td>---------------------</td>
<td>-------------</td>
<td>-------------------------</td>
<td>------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>10003635</td>
<td>sporadic</td>
<td>all present</td>
<td>c.35G&gt;A p.Gly12Asp</td>
<td>normal</td>
<td>normal</td>
<td>normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003636</td>
<td>sporadic</td>
<td>all present</td>
<td>c.38G&gt;A p.Gly13Asp</td>
<td>normal</td>
<td>normal</td>
<td>normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003637</td>
<td>sporadic</td>
<td>all present</td>
<td>c.38G&gt;A p.Gly13Asp</td>
<td>normal</td>
<td>normal</td>
<td>normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003639</td>
<td>sporadic</td>
<td>all present</td>
<td>normal</td>
<td>c.1799T&gt;A</td>
<td>p.Val600Glu</td>
<td>normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003640</td>
<td>sporadic</td>
<td>all present</td>
<td>normal</td>
<td>normal</td>
<td>normal</td>
<td>normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003641</td>
<td>sporadic</td>
<td>all present</td>
<td>normal</td>
<td>normal</td>
<td>normal</td>
<td>normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003642</td>
<td>sporadic</td>
<td>all present</td>
<td>normal</td>
<td>normal</td>
<td>normal</td>
<td>normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>10003643</td>
<td>sporadic</td>
<td>all present</td>
<td>normal</td>
<td>c.1799T&gt;A</td>
<td>p.Val600Glu</td>
<td>normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>11005267</td>
<td>sporadic</td>
<td>mlh1 loss</td>
<td>normal</td>
<td>normal</td>
<td>normal</td>
<td>normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>11005301</td>
<td>sporadic</td>
<td>all present</td>
<td>normal</td>
<td>normal</td>
<td>normal</td>
<td>normal</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>11005014</td>
<td>sporadic</td>
<td>all present</td>
<td>c.35G&gt;T p.Gly12Val</td>
<td>normal</td>
<td>normal</td>
<td>Possible MSI in 1 marker /possible</td>
<td>MMR proficient</td>
</tr>
<tr>
<td>Lab no</td>
<td>Tumour Origin</td>
<td>IHC RESULT</td>
<td>KRAS Result</td>
<td>BRAF Result</td>
<td>MLH1 Methylation Result</td>
<td>MSI Result</td>
<td>Classified as</td>
</tr>
<tr>
<td>--------</td>
<td>---------------</td>
<td>------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------------------</td>
<td>-----------</td>
<td>----------------</td>
</tr>
<tr>
<td>10003512</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal (ARMS)</td>
<td>?c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>10003535</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>?c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>11005013</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>Fail x3</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>10003566</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 2 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>11005294</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 3 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>11005007</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 4 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>10003503</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>10003511</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>10003563</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>10003606</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>10003616</td>
<td>sporadic</td>
<td>All present</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>Lab no</td>
<td>Tumour Origin</td>
<td>IHC RESULT</td>
<td>KRAS Result</td>
<td>BRAF Result</td>
<td>MLH1 Methylation Result</td>
<td>MSI Result</td>
<td>Classified as</td>
</tr>
<tr>
<td>---------</td>
<td>---------------</td>
<td>------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------------------</td>
<td>---------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>10003629</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>11005005</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>11005009</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>11005011</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>11005015</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>11005281</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>11005423</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>11005428</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>10003502</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>11005269</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>c.38G&gt;A p.Gly13Asp</td>
<td>Normal</td>
<td>Hypermethylation</td>
<td>MSI in 2 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>10003630</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>Normal</td>
<td>Hypermethylation</td>
<td>MSI in 4 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>10003603</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>Normal</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>Lab no</td>
<td>Tumour Origin</td>
<td>IHC RESULT</td>
<td>KRAS Result</td>
<td>BRAF Result</td>
<td>MLH1 Methylation Result</td>
<td>MSI Result</td>
<td>Classified as</td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
<td>------------</td>
<td>-------------</td>
<td>-------------</td>
<td>--------------------------</td>
<td>------------</td>
<td>---------------</td>
</tr>
<tr>
<td>11005279</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>Normal</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>11005275</td>
<td>untested</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>Normal</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>11008758</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>Normal</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>11002139</td>
<td>mut neg</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>Normal</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>10003604</td>
<td>untested</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>Normal</td>
<td>Hypermethylation</td>
<td>Normal</td>
<td>MSI in 2 markers</td>
</tr>
<tr>
<td>10003602</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>10003612</td>
<td>untested</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>Normal</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>10003638</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>11005427</td>
<td>sporadic</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in 5 markers</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>10008758</td>
<td>mut neg</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>Normal</td>
<td>Hypermethylation</td>
<td>Not done</td>
<td>sporadic MLH1 loss</td>
</tr>
<tr>
<td>11002139</td>
<td>mut neg</td>
<td>MLH1 LOSS</td>
<td>Normal</td>
<td>Normal</td>
<td>Hypermethylation</td>
<td>Not done</td>
<td>sporadic MLH1 loss</td>
</tr>
</tbody>
</table>
### Appendix Six: Data Colon Cancer Family Registry

**Sporadic Mismatch repair deficient colorectal cancers**

<table>
<thead>
<tr>
<th>Lab No.</th>
<th>Tumour Origin</th>
<th><strong>BRAF Result</strong></th>
<th><strong>MLH1 Result</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>11007877</td>
<td>sporadic mut neg</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007930</td>
<td>sporadic mut neg</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11008038</td>
<td>sporadic mut neg</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007968</td>
<td>sporadic mut neg</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007966</td>
<td>sporadic mut neg</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007961</td>
<td>sporadic mut neg</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007828</td>
<td>sporadic mut neg</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007942</td>
<td>sporadic mut neg</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007927</td>
<td>sporadic mut neg</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007940</td>
<td>sporadic mut neg</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007886</td>
<td>sporadic mut neg</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007821</td>
<td>sporadic mut neg</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007823</td>
<td>sporadic mut neg</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007934</td>
<td>sporadic mut neg</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007876</td>
<td>sporadic mut neg</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007805</td>
<td>sporadic mut neg</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007889</td>
<td>sporadic mut neg</td>
<td>Normal</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007924</td>
<td>sporadic mut neg</td>
<td>Normal</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007937</td>
<td>sporadic mut neg</td>
<td>Normal</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007928</td>
<td>sporadic mut neg</td>
<td>Normal</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007914</td>
<td>sporadic mut neg</td>
<td>Normal</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007814</td>
<td>sporadic mut neg</td>
<td>Normal</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007822</td>
<td>sporadic mut neg</td>
<td>Normal</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007812</td>
<td>sporadic mut neg</td>
<td>Normal</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007888</td>
<td>sporadic mut neg</td>
<td>Normal</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007887</td>
<td>sporadic mut neg</td>
<td>Normal</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007965</td>
<td>sporadic mut neg</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Normal</td>
</tr>
<tr>
<td>11007962</td>
<td>sporadic mut neg</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Normal</td>
</tr>
<tr>
<td>11007964</td>
<td>sporadic mut neg</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>11007963</td>
<td>sporadic mut neg</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>11007872</td>
<td>sporadic mut neg</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>11007967</td>
<td>sporadic untested</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007913</td>
<td>sporadic untested</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007829</td>
<td>sporadic untested</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007819</td>
<td>sporadic untested</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007874</td>
<td>sporadic untested</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007808</td>
<td>sporadic untested</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007915</td>
<td>sporadic untested</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007869</td>
<td>sporadic untested</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007945</td>
<td>sporadic untested</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
<tr>
<td>11007825</td>
<td>sporadic untested</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Hypermethylation</td>
</tr>
</tbody>
</table>
### Appendix seven: Data Manchester and Colon Cancer Family Registry proven germline

#### MLH1 mutation carriers

<table>
<thead>
<tr>
<th>Lab No.</th>
<th>Tumour Origin</th>
<th>KRAS Result</th>
<th>BRAF Result</th>
<th>MLH1 Result</th>
<th>MSI Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>11007883</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>c.1799T&gt;A p.Val600Glu</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007910</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007868</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Hypermethylation</td>
<td>N/A</td>
</tr>
<tr>
<td>11008037</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11005300</td>
<td>MLH1 carrier</td>
<td>Normal</td>
<td>Normal</td>
<td>Hypermethylation</td>
<td>MSI in 5 markers</td>
</tr>
<tr>
<td>11007938</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007944</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Hypermethylation</td>
<td>N/A</td>
</tr>
<tr>
<td>11005224</td>
<td>MLH1 carrier</td>
<td>Fail x2</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>11005226</td>
<td>MLH1 carrier</td>
<td>c.38G&gt;A p.Gly13Asp</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in 1 marker</td>
</tr>
<tr>
<td>11005228</td>
<td>MLH1 carrier</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in 5 markers</td>
</tr>
<tr>
<td>11005230</td>
<td>MLH1 carrier</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in 5 markers</td>
</tr>
<tr>
<td>11005232</td>
<td>MLH1 carrier</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in 5 markers</td>
</tr>
<tr>
<td>11005234</td>
<td>MLH1 carrier</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>11005236</td>
<td>MLH1 carrier</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in 5 markers</td>
</tr>
<tr>
<td>11005238</td>
<td>MLH1 carrier</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in 5 markers</td>
</tr>
<tr>
<td>11005240</td>
<td>MLH1 carrier</td>
<td>c.35G&gt;T p.Gly12Val</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in 5 markers</td>
</tr>
<tr>
<td>11005241</td>
<td>MLH1 carrier</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in 4/4 markers</td>
</tr>
<tr>
<td>11005265</td>
<td>MLH1 carrier</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>11005271</td>
<td>MLH1 carrier</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in 5 markers</td>
</tr>
<tr>
<td>11005283</td>
<td>MLH1 carrier</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in 5 markers</td>
</tr>
<tr>
<td>11005287</td>
<td>MLH1 carrier</td>
<td>c.34G&gt;C p.Gly12Asp</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in 5 markers</td>
</tr>
<tr>
<td>11005291</td>
<td>MLH1 carrier</td>
<td>RT PCR</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>11005293</td>
<td>MLH1 carrier</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>11005295</td>
<td>MLH1 carrier</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>MSI in 5 markers</td>
</tr>
<tr>
<td>Lab No.</td>
<td>Tumour Origin</td>
<td>KRAS Result</td>
<td>BRAF Result</td>
<td>MLH1 Result</td>
<td>MSI Result</td>
</tr>
<tr>
<td>---------</td>
<td>---------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>11005297</td>
<td>MLH1 carrier</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>11005298</td>
<td>MLH1 carrier</td>
<td>c.35G&gt;A p.Gly12Asp</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>11005302</td>
<td>MLH1 carrier</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>11005303</td>
<td>MLH1 carrier</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>11007806</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007807</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007810</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007811</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007813</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007815</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007816</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007817</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007818</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007820</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007824</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007826</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007870</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007873</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007878</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007879</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007881</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007882</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007884</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007885</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007909</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007911</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007912</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007916</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007917</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>11007918</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007919</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007925</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007926</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007929</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007931</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007932</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007933</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007935</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007936</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007939</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007941</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007943</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007946</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007908</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11007920</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
<tr>
<td>11008036</td>
<td>MLH1 carrier</td>
<td>N/A</td>
<td>Normal</td>
<td>Normal</td>
<td>N/A</td>
</tr>
</tbody>
</table>