Modifier Genes in Lynch Syndrome: Functional Genomics and its Consequence on Disease Expression

Stuart Reeves
B.Sc (Hons)
The University of Newcastle

Doctor of Philosophy (Medical Genetics)
The University of Newcastle
October 2011

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying subject to the provisions of the Copyright Act 1968.

28-March-2012

Date Stuart Reeves

Acknowledgements

First and foremost I would like to thank my supervisor, Professor Rodney J Scott. I am sincerely appreciative for all the support, guidance and, confidence he has provided me throughout the duration of this project. To have a supervisor of his knowledge and understanding has been an enormous privilege and something I'm very thankful for.

I'd also like to thank Dr Cliff Meldrum for his technical support and guidance. The help he kindly provided throughout the experimental stages of this project have been invaluable.

Thank you to the members of the Information Based Medicine group in particular to David Mossman for his friendship along the journey of the project. I'd also like to thank Kim Colyvas for his statistical assistance in making sense of the mountains of data generated throughout this project. A special mention needs to be made to the members of our collaboration groups who kindly provided many of the precious samples that were utilised throughout this study.

I'd like to thank the staff in the Molecular Genetics laboratory at the Hunter Area Pathology Service for making me feel welcome and the sharing of their specialist lab knowledge with me.

Finally, a big thank you must go to all my family and friends. I could have not done without your tireless encouragement and support throughout this journey.

Contents

Common Abbreviations

CRC	Colorectal Cancer
DNA	Deoxyribonucleic Acid
DNMT3B	DNA Methyltransferase 3 Beta
HNPCC	Hereditary nonpolyposis Colorectal Cancer
hMLH1	Human MutL Homolog 1
hMSH2	Mut S Homolog 2
IGF-1	Insulin like Growth Factor 1
MMR	Mismatch Repair
MTHFR	Methylenetetrahydrofolate Reductase
PCR	Polymerase Chain Reaction
RNA	Ribonucleic Acid
SNP	Single Nucleotide Polymorphism

Publications

1. Reeves S, Meldrum C, and Scott, R.J., *IGF-1 gene polymorphism and risk for hereditary nonpolyposis colorectal cancer.* J Natl Cancer Inst 2006; Nov 15; 98 (22): 1664-5.
2. Reeves S.G, Rich D, Meldrum C, Colyvas K, Kurzawski G, Suchy J, Lubinski, & Scott, R.J., *IGF-1 is a modifier of disease risk in hereditary non-polyposis colorectal cancer.* Int J Cancer. 2008 Sep 15; 123(6):1339-43.
3. Reeves S.G, Meldrum C, Groombridge C, Spigelman A.D, Suchy J, Kurzawski G, Lubinski J, McElduff P, & Scott, R.J., *MTHFR 677 C>T and 1298 A>C polymorphisms and the age of onset of CRC in hereditary nonpolyposis CRC.* The European Journal of Human Genetics 2009 17, 629–635.
4. Reeves SG, Mossman D, Meldrum CJ, Kurzawski G, Suchy J, Lubinski J & Scott RJ., *The -149C>T SNP within the ΔDNMT3B gene, is not associated with early disease onset in hereditary non-polyposis colorectal cancer.* Cancer Lett 2008 June 28; 265(1):39-44.
5. Reeves S.G, Meldrum C, Groombridge C, Spigelman A.D, Suchy J, Kurzawski G, Lubinski J, & Scott, R.J., *DNA repair gene polymorphisms and risk of early onset colorectal cancer in Hereditary Nonpolyposis Colorectal Cancer* – Cancer Epidemiology (Article in press October 2011)

Abstract

Colorectal cancer (CRC) is globally a major cause of morbidity and mortality. Each year more than one million patients will be diagnosed with colorectal cancer, with about 15 - 20% of these patients having a family history or an inherited colorectal cancer syndrome. Somewhere between 1% and 7% (dependent on population under study) of these cases will have Lynch syndrome, which is the most common hereditary autosomal-dominant inherited cancer syndrome caused by germline mutations in deoxyribonucleic acid (DNA) mismatch repair genes.

Patients diagnosed with Lynch syndrome who harbour a confirmed germline mutation in DNA mismatch repair (MMR) genes have an 80% lifetime risk of developing an epithelial malignancy. Each patient belongs to a family that requires special medical attention including genetic counselling, DNA testing for mismatch repair genes (most frequently *hMLH1* or *hMSH2*) and screening for CRC.

There is, however, considerable variation in the age of disease onset which is explained by a combination of genetic and environmental factors. The studies described in this thesis are aimed to better understand the genetic modifying effects on disease expression and how they relate to the likely age of colorectal cancer onset.

Previous studies have identified a polymorphic CA repeat region in *IGF-1* and two specific single nucleotide polymorphisms in MTHFR that were thought to alter the age of disease onset in individuals with Lynch syndrome. The effects of these

polymorphisms were examined in larger multinational cohorts of patients and found to have significant effects on disease onset age. This is discussed in chapters 2, 3 and 4.

Another similar study had identified a single nucleotide polymorphism in DNMT3B which was reported to have a significant effect in colorectal cancer expression in Lynch syndrome. The effect of this polymorphism was examined in a large multinational cohort of patients however, it was found to have no effect on the age of disease onset. Several candidate polymorphisms were also identified in the DNA repair genes *BRCA2*, *hMSH3*, *Lig4*, *hOGG1*, *XRCC1*, *XRCC2* and *XRCC3* but no significant associations were identified. The results from these studies are discussed in chapters 5 and 6.

All data generated from these studies were extensively analysed by a combination of statistical tests that included Kaplan-Meier survival and Cox hazard regression analysis allowing data to be stratified by both single and multi variable factors. Allele frequencies were also tested for significant deviation from the Hardy-Weinberg equilibrium, while Pearson's Chi-square test was utilised to evaluate differences in the allele frequencies between the multinational cohort groups and distribution of genotypes.

The results described in this thesis contribute to a better understanding of disease expression in Lynch syndrome as it identifies genetic factors involved in the etiology of malignancy in this disease. The progress made in this area of medical research will aid in providing better predictive information of greater accuracy regarding the risks of colorectal cancer and enable the development of personalised cancer surveillance regimens.

Chapter 1

General Introduction

1.1 Background

Recent advances as a result of the human genome project continue to further our knowledge into genetic mechanisms that govern life on a molecular basis. In all living organisms the general process by which growth and development proceeds is overseen by an underlying series of molecular events. This has been summarised in the central dogma of molecular genetics. This states that genetic information is carried in deoxyribonucleic acid (DNA), and that this information is then transcribed into ribonucleic acid (RNA) before finally being translated for protein synthesis providing the basis for living tissue.[1]

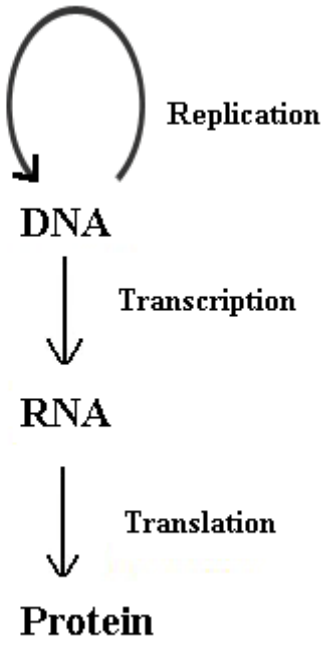

Figure 1 – The central dogma of Molecular Genetics.

The central dogma of molecular genetics forms the backbone of molecular biology. Three distinct steps are involved in this process: DNA replication, transcription of the genetic information into RNA, then the translation of the RNA in protein.

DNA is a large macromolecule which is susceptible to a range of endogenous and exogenous agents that can potentially damage its integrity. DNA mutations can arise during transcription and replication, occurring in either somatic cells (cells not involved in reproduction) or germcells (reproductive cells). A germline mutation occurs in germ cells which can then be passed onto offspring and inherited from generation to generation; whereas somatic mutations occur spontaneously after birth and cannot be passed onto offspring, often occurring in specific tissues as a result of environmental influence. *De novo* mutations however can also occur in germ cells which can then be inherited from generation to generation. Most *de novo* mutations reside in the germ cells of the individual and as such there is a lack of any phenotypic expression in the transmitting parent [2]. A mechanism by which *de novo* changes may occur is through genetic mosaicism. Mosaicism is the presence of two or more genetically different cell lines occurring in the same organism, which can occur in both somatic and germline tissue. Mosaicism of a genetic disease in either somatic or germline cells will not usually cause expression of the disease in that individual. Mutations occurring in a parent's germline however can cause *de novo* inherited disease in a child. When a *de novo* germline mutation is present in a person who harbours a large clone of mutant germline cells, a normal couple with no previous family history can produce more than one child with the same disease.

To ameliorate the effect of mutations, several DNA repair mechanisms have evolved to maintain genetic integrity and therefore ensure the functional activity of the respective protein. Figure 2 displays a summary of the role of DNA repair in the cell cycle.

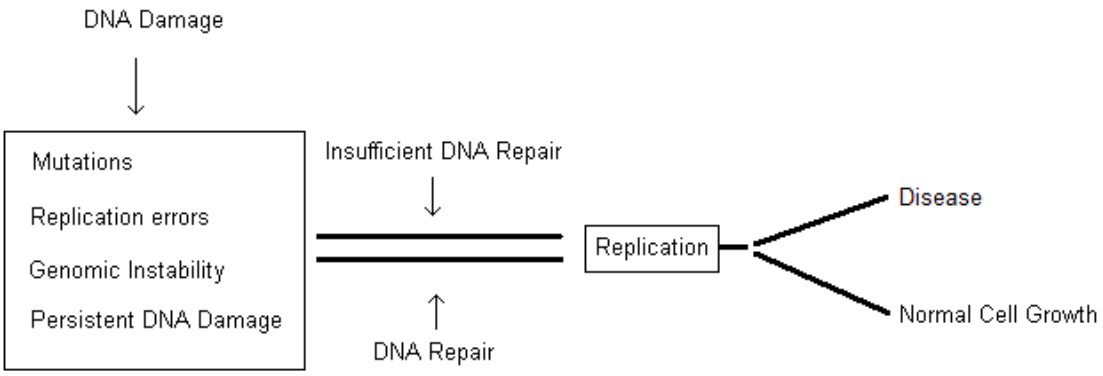

Figure 2 – DNA repair in response to DNA damage.

The role of DNA repair is to correct genetic errors which arise throughout the cell cycle. This figure shows a dual model of what occurs in the presence of and in the absence of a sufficient DNA repair system. Abnormalities in the repair process which are unchecked before entering cellular replication may lead to disease.

1.2 DNA Repair Pathways

In response to DNA damage a series of very efficient mechanisms which reduce the risk of DNA mutations and assist in the preservation of genomic integrity have evolved. The genes involved in genomic integrity are those whose function involves either DNA damage recognition or repair. There are several different types of DNA repair including: double strand (DS) break repair, single strand (SS) repair and excision repair. The term DNA excision repair includes further specific types of DNA repair that include; base excision repair (BER), nucleotide excision repair (NER) and mismatch repair (MMR). All three types of DNA excision repair are involved in the removal of either incorrect nucleic acid bases or bases that have been modified as a result of chemical interaction. They are then replaced with the appropriate sequence or base [3-9].

1.2.1 - DNA Mismatch Repair

The basic process of DNA MMR involves the detection of a single base mismatch which occurs during DNA replication. To begin with, the correct base must be identified and the erroneous one removed via an excision repair pathway. The major active components of the MMR system are described as the Mut proteins, including MutS, MutL and MutH. These three proteins are of particular importance in the detection and the recruitment of excision repair. The Mut proteins were originally discovered in bacteria; however homologous Mut proteins have since been identified in eukaryotes. In humans, the homologues of mutS (*hMSH2/hMSH3*) produce a heterodimeric complex which binds to the DNA mismatch. The mutL homologues (hMLH1 and hPMS1 and hPMS2) then form a heterodimeric complex which contributes to the overall MMR process. A MutH homologue has not been identified in humans; instead repair of the correct strand appears to be directed by nicks in the newly replicated strand [10-13]. A summary of the MMR pathway in eukaryotes is shown in figure 3.

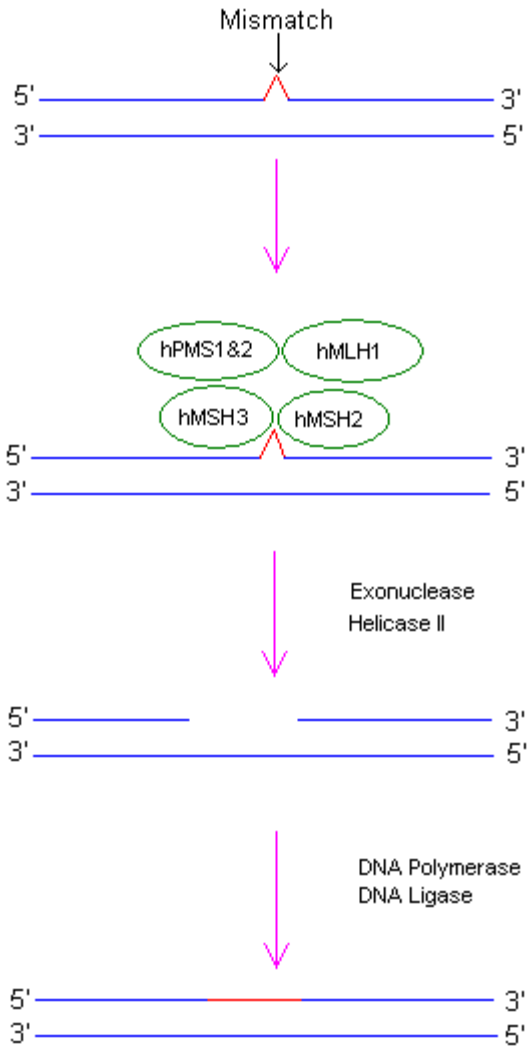

Figure 3 – Overview of the mismatch repair process.

A mispaired base is recognised by the hMSH2/hMSH3 complex. Mut L related proteins (hMLH1/hPMS1&2 complexes) then interact with the Mut S related proteins that are already bound to the mispaired base. Cleavage of the base is then assisted by DNA exonucleases and helicases which are guided by nicks in the newly replicated strand. The gap is then sealed by DNA polymerase and DNA ligase in the final step of the process. Adapted from Marti et al [14].

1.3 - The Pathway to Cancer

Three main classes of genes exist that may become inactivated resulting in an increased likelihood of malignancy. These are oncogenes, tumour suppressor genes, and the genes involved in the maintenance of genomic integrity such as those involved in mismatch repair [11, 15]. Oncogenes are genes which stimulate cellular proliferation, however when they acquire activating mutations, uncontrolled proliferation is often the result [15, 16]. Tumour suppressor genes slow down cell division, repair DNA errors and control cell death. Their role is to prevent or inhibit cells with irreparable DNA damage continuing further through the cell cycle, or by a programmed cell death process known as apoptosis [17]. Different types of tumour suppressor genes exist, some of which are similar to those involved in genomic integrity, however others are involved in different cell cycle check points including genes which control cell division and apoptosis. The *RB1* (retinoblastoma) gene is an example of a tumour suppressor gene that has an important role in controlling cellular proliferation [18]. In cases where DNA integrity is significantly compromised apoptotic signalling results via the cell cycle checkpoint control protein TP53 which results in the elimination of the damaged cell.

Gene loss of function occurs when tumour suppressor genes are inactivated either by mutation, deletion or both. Knudson's two hit theory states that a single inactivating mutation in either a tumour suppressor or similar gene associated with genomic integrity are not enough to trigger cancer development as long as an intact allele remains. A further inactivating "hit" is required in the second allele in the same cell for loss of function to occur. The two hit hypothesis provides insight into why germline mutations are more often observed in tumours in younger cancer patients compared to those with

sporadic disease. This is thought to be due to germline mutation carriers already having one allele inactivated at birth and therefore only require a second inactivating mutation for gene loss of function. In individuals who have no genetic predisposition (i.e. have both normal copies of the allele) two sporadic hits are required before gene inactivation occurs [19, 20]. Figure 4 gives an illustrated example of Knudson's two hit theory, highlighting the difference between inherited and sporadic forms of disease syndromes.

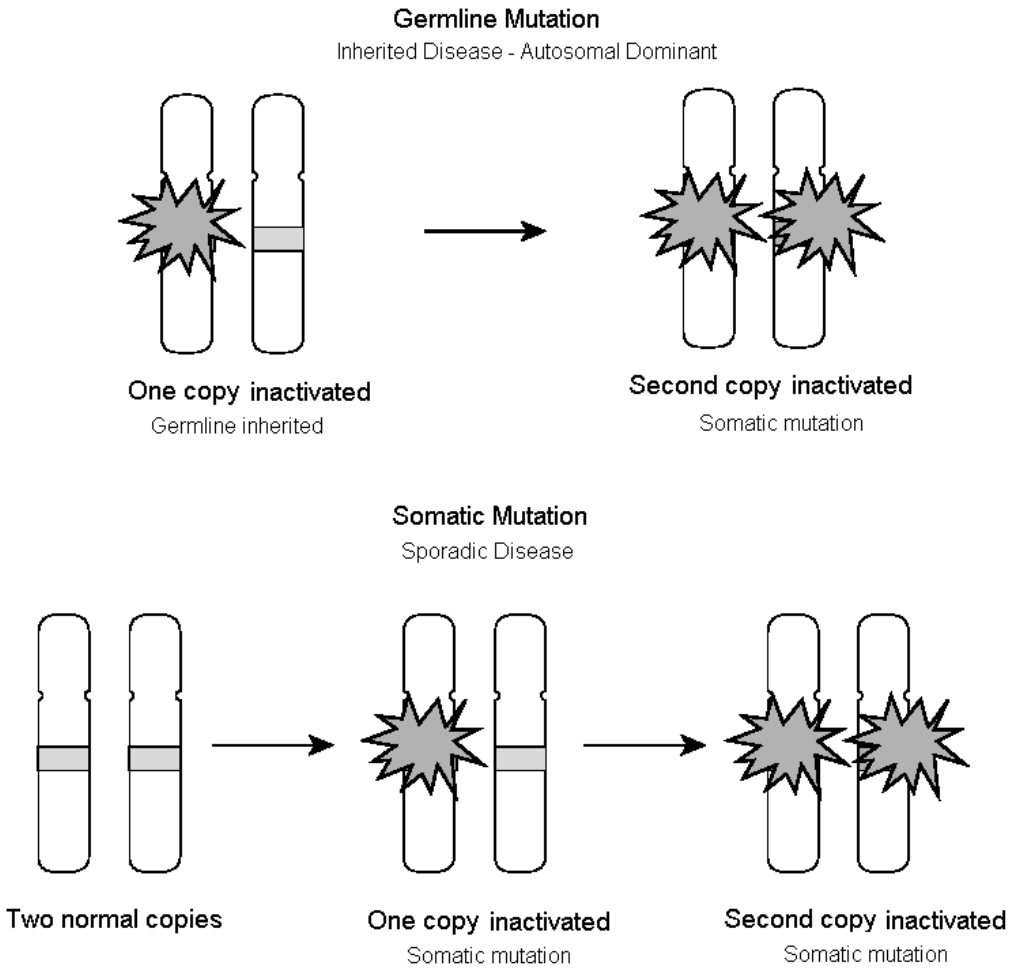

Figure 4 – Gene inactivation by Knutson's two hit hypothesis.

When an inactivating mutation is inherited via the germline (as in an autosomal dominant disorder) only one subsequent "hit" is required in the second allele for complete loss of function. When two functional copies are inherited however, two random somatic inactivating mutations are required for total gene inactivation. This is thought to be the foremost reason of why sporadic cancers usually occur later in life compared to inherited cancer syndromes.

Adapted from Richards M 2001 [21]

More recent evidence, however, suggests that this simplistic view is currently being challenged as it appears that a series of "driver" mutations are required for cancer to progress. A driver mutation provides a growth advantage to a cell and therefore may be implicated in oncogenesis as it has been specifically selected for during tumour development. Initial evidence suggested that the accumulation of somewhere between five and seven driver mutations were needed for malignancy to progress. Given new techniques that can interrogate entire genomes rapidly however, there may be as many as 20 driver mutations required for tumour development [22-25].

1.4 - The Genetics of Colorectal Cancer

Colorectal cancer (CRC) is the third most common cancer worldwide and second most frequent malignancy in some developed countries including the United States [26]. Despite being one of the more preventable cancers, it accounts for over 677,000 deaths annually according to the World Health Organisation [27]. There is a strong familial component associated with CRC which is thought to account for between 15% and 20% of all cases [28]. In terms of understanding genetic risk factors associated with CRC much of the familial component still remains largely elusive, however there are two well defined genetic syndromes, familial adenomatous polyposis (FAP) and Lynch syndrome/hereditary non-polyposis colorectal cancer (HNPCC), that have been well characterised since the identification of the genetic basis of the two diseases [28]. Most colorectal tumours including both sporadic and familial develop from adenomas, however the number of adenomas can vary considerably. The process of colorectal cancer development may take several decades to occur with the requirement of driver mutation accumulation in a series of tumour suppressor and oncogenes.

Two basic categories of genes are involved in CRC; including gatekeeper and caretaker genes. Gatekeeper genes are tumour suppressor genes and directly prevent malignancy by monitoring cellular growth or by promoting cell death by apoptosis [29]. When gatekeeper genes are functional, mutations in other tumour promoting genes will not result in cancer as they will be recognised and repaired. An inactivating mutation in a gatekeeper gene however may lead to the accumulation of genetic errors in a series of oncogenes resulting in uncontrolled cellular proliferation. In CRC the tumour suppressor gene adenomatosis polyposis coli (*APC*) is considered a gatekeeper that is frequently found to have undergone an inactivating mutation [30].

The second category comprises genes referred to as caretakers. Caretaker genes are involved in maintaining genomic integrity and may have an indirect effect on growth. These genes do not directly promote cancer; however their inactivation will result in genetic instability significantly altering the likelihood of mutation accumulation in all genes, including gatekeepers. Well recognised examples of caretaker genes are DNA repair genes that include those involved in the DNA mismatch repair pathway (*hMLH1*, *hMSH2, hMSH6*). When inactivated, the result is a failure to recognise or repair DNA mismatches (indicated by the presence of microsatellite instability) which can lead to the inactivation of tumour suppressor genes and the uncontrolled activation of oncogenes [31, 32]. In CRC inactivation of MMR genes can lead to accumulation of genetic errors in tumour suppressors and oncogenes including *p53, APC* and *K-ras* thereby altering the probability of malignant transformation [2]. As there is a requirement for several genetic mutations to occur before tumourigenesis, including the silencing of a tumour suppressor gene, individuals who inherit a germline mutation in a tumour suppressor gene are at

greater risk of developing a malignancy earlier in life than those who inherit a similar mutation in a gene involved in genomic integrity. This is observed in FAP where the average age of disease onset is earlier compared to Lynch syndrome.

There are two forms of genetic instability recognised thus far, chromosome instability (CIN) and microsatellite instability (MSI) that constitute the major molecular subtypes of CRC [33]. CIN and MSI lead to the accumulation of genomic instability, thereby increasing risk of tumour suppressor inactivation and uncontrolled oncogene activation [29]. As a result of CIN, structural rearrangements may occur such as deletions, amplifications or translocations arising from breaks in DNA. Alterations can also occur in the number of intact chromosomes, referred to as whole-chromosome missegregations that arise from errors in mitosis [32]. CIN tumours have been associated with poor patient prognosis, however to date it is still not entirely clear whether there is any significant difference in patient prognosis between CIN and MSI tumours [34]. MSI occurs through replication slippage in short tandem DNA repeat regions ranging from two to six base pairs long which may be repeated numerous times in length. The presence of MSI in CRC tumours is indicative of an impaired MMR system as a non functional MMR gene allows for the accumulation of genetic errors through DNA polymerase slippage during replication. MSI is frequently observed in HNPCC tumours whilst also present in approximately 15% of sporadic CRC cases [35].

1.5 - Lynch Syndrome

Lynch syndrome, also referred to as hereditary non polyposis colorectal cancer (HNPCC), is the most common form of hereditary CRC and is best understood as a hereditary predisposition to malignancy caused by a germline mutation in a DNA MMR gene. It is an autosomal dominant syndrome which includes a wide range of epithelial malignancies in addition to CRC. Malignancies commonly identified in Lynch syndrome include endometrial cancer, gastric tumours, brain tumours (most often glioblastoma multiform), ovarian and kidney tumours [36]. Initial reports were first described in 1895 by Aldred Warthin [37] however it was not until the 1960's when Henry Lynch revisited Warthins observations and began to more accurately define this form of hereditary cancer [38-40]. Initially, it was widely reported that there were two separate syndromes including; Lynch syndrome I (Hereditary Site-Specific Colon Cancer), which only included those families with colon cancer and Lynch syndrome II (Cancer Family Syndrome) that encompassed families presenting with a range of extra colonic cancers in addition to colorectal cancer. Soon afterwards however the term "hereditary non polyposis colorectal cancer" was introduced which covered both Lynch I and II syndromes [41, 42].

With the knowledge of this newly identified cancer syndrome, registries for hereditary cancers began to collect and describe families who had features similar to that described by Lynch. In an effort to standardise the collection of families and clinical reporting a set of guidelines were established by the International Collaborative Group on HNPCC in 1991 to aid in the identification of the genetic basis of the disease [43] known as the Amsterdam criteria. With the implementation of the Amsterdam criteria it was not long

before the genetic basis of Lynch syndrome/HNPCC was discovered. This was first described by Peltomaki and Lindblom who were able to isolate the chromosomal regions associated with cancer occurrences in several families by using microsatellite markers for linkage analysis [44, 45]. Soon afterwards, it was found that MSI was present in most CRC's occurring within the context of Lynch syndrome [44, 46] which also provided clues as to what genes were involved in this disease. DNA mismatch repair has been extensively studied in yeast and bacteria and indicated a high frequency of DNA mismatches in these organisms when the respective genes had been knocked out. [47]. In 1993 the first Lynch syndrome related gene (a yeast MutS homologue) *hMSH2* was identified [48, 49], closely followed by the MutL homologue *hMLH1* in early 1994 [50, 51]. A further two MMR genes (*hPMS2* and *hMSH6*) have since been discovered in Lynch syndrome families [52, 53], although these are observed less frequently in comparison to *hMLH1* and *hMSH2* mutations. Additional MutL, MutS and MYH homologues have been since discovered however their involvement has yet to be proven [52].

With the genetic mechanism of Lynch syndrome identified it soon became evident that the criteria listed in the first Amsterdam protocol were too restrictive to accurately define all families who carried a germline mutation in their MMR genes. The main limitation of the Amsterdam I criteria was the lack of scope for families where there were extracolonic malignancies. To help cover this shortfall, the second Amsterdam criteria II was implemented in 1999, allowing for the addition of several extra-colonic cancers into the guidelines including endometrial, small intestine and ureter malignancies [54, 55]. The second version of the Amsterdam criteria is still in current use, albeit with further

guidelines attached. The new guidelines, as listed in the Bethesda Criteria, were devised with a greater understanding of the genetics behind Lynch syndrome/HNPCC and in particular MSI. The additional guidelines have allowed for greater sensitivity by using the criteria set out in the Amsterdam II protocol but also integrating further details. The Bethesda Criteria has allowed for the inclusion of tumour pathology leading to an additional specific outline for accurate diagnosis. Additional malignancies were also adopted into this criteria that had not been previously covered in Amsterdam II including; biliary tract, stomach, ovary, pancreas and brain cancers [56, 57]. Despite the expansion offered by the Bethesda criteria, it is expected with evolving knowledge into Lynch syndrome/ HNPCC further adjustments to these guidelines will be required [58].

Table 1 – Summary of Amsterdam I and II criteria and Revised Bethesda Criteria

Name	**Criteria**
Amsterdam I	At least 3 relatives with CRC At least 2 successive generations affected, with at least one family member diagnosed before 50 years of age The exclusion of familial adenomatous polyposis
Amsterdam II	Same criteria as described in Amsterdam I plus CRC may be substituted for by other HNPCC related malignancies including endometrial, small bowel and pelviureter
Bethesda	Must meet criteria listed in the Amsterdam criteria CRC below 50 years of age Multiple CRC or HNPCC related cancers CRC with a MSI related histology CRC or HNPCC related tumour in 1 or more first degree relatives, with 1 of the cancers being diagnosed under the age of 50 CRC diagnosed in 2 or more first or second degree relatives with HNPCC related tumours, regardless of age

1.6 - Lynch Syndrome Risk

It has been recently defined that only families that adhere to the Amsterdam or Bethesda criteria are described as having HNPCC. However, not all families fulfilling these criteria have Lynch syndrome unless they are carrying germline DNA MMR mutations. Germline mutations in *hMLH1* and *hMSH2* account for approximately 60% of Lynch syndrome cases, with a smaller percentage accounted for in *hPMS2* and *hMSH6* [59, 60]. As mentioned previously, Lynch syndrome is the most common form of hereditary CRC with an incidence of approximately 1:4000, responsible for somewhere between 1% and 6% of all CRC cases [61]. The most accurate figures for the frequency of HNPCC are derived from the Danish cancer registry which indicates that 1.7% of all CRCs are due to Lynch Syndrome. If only CRC cases under the age of 50 years are considered Lynch syndrome is predicted to be responsible for 14.7% of these [62]. Nevertheless the frequency of Lynch syndrome cases can vary between different countries and populations largely due to genetic and environmental influences and the recruitment of families suspected of having this syndrome. It is widely reported that the penetrance is approximately 80% risk of developing a Lynch syndrome related malignancy by 70 years of age [63]. The overall lifetime risk of developing a Lynch syndrome related malignancy has been estimated at around 80%, with men having a significantly higher risk of developing CRC (74%) than women (30-52%) by 70 years of age [63-65]. On average, the age of CRC onset is 44 years, compared to 64 years for sporadic disease [28, 66]. In women, the risk of endometrial cancer is thought to be at least that of CRC if not greater, with estimates ranging from 42-54% by 70 years of age [63, 64]. There have been suggestions that cancer risk estimates in Lynch syndrome are overestimated due to the

selection criteria. This was supported recently in a study suggesting a significantly lower risk for CRC in 36 confirmed Lynch syndrome families located in France. The report claimed the risk of developing CRC in men and women was in the vicinity of 47% and 33% by 70 years of age, respectively [67]. Whether this is an accurate representation of the overall risk of endometrial and CRC within the Lynch syndrome population is still open to debate until additional studies are completed on larger population sizes and on different ethnic groups.

The presence of malignancies such as breast and rarer types of cancers within Lynch syndrome families also raises questions on which cancers should be included in the definition of this syndrome. Before the genetics of Lynch syndrome were revealed it was assumed that all types of cancers over-represented in families with this syndrome were a result of a deficient MMR system. Considerable variation exists between families with breast and rarer types of cancers which appear to be more prevalent in some populations compared to others [59, 68, 69]. MSI in tumours not thought to be part of the spectrum of disease encompassed within Lynch syndrome indicated that the mechanisms underlying these malignancies is due to an ineffective MMR activity. This leads to the suggestion that any type of tumour displaying MSI from a confirmed Lynch syndrome family member should be included as part of the syndrome, even if it is not part of the current Amsterdam/Bethesda criteria. Until further studies are completed in investigating tumours and their genetic composition in larger patient cohorts a more precise definition of this syndrome will await.

The risk of developing a malignancy is also thought to vary depending on which MMR gene is inactivated. As most cases are diagnosed with either an *hMLH1* or *hMSH2* mutation disease risk, penetrance is more likely to be predicted between these two genes. To date it is still not conclusive as to whether a significant difference in disease risk exists between *hMLH1* and *hMSH2* mutation carriers; however a trend for increased disease penetrance in *hMSH2* has been reported [70]. Additionally, families harbouring *hMSH2* mutations have a tendency to present with more extra-colonic cancers compared to families with mutations in *hMLH1* [70, 71]. Families with *hMSH6* gene mutations have been described as having an over-representation of endometrial cancers, whilst also having a later age of CRC onset compared to *hMLH1/hMSH2* mutation carriers [71-74]. Initially it was unclear whether *hPMS2* was implicated in Lynch syndrome and difficulties still remain in determining the exact penetrance of *hPMS2* due to small numbers of patients carrying this mutation and adjacent pseudogenes that cause significant difficulties for mutation detection [75, 76]. Current estimates in the number of families accounted for by this gene range from somewhere between 5% and 7% [77, 78] however these figures are not accurate. As information has accumulated it is clear that mutations in *hPMS2* are more frequent than first thought as they display a range of associated tumours [78]. Whilst it is clear that a degree of variability exists in disease expression between MMR genes, taken as a whole there is still much to discover in relation to disease penetrance before accurate estimates can be made. This is only likely to be revealed once larger studies of combined Lynch syndrome cohorts are compiled and analysed.

1.7 - Lynch Syndrome Disease Expression

In addition to the inconsistency observed in MMR genes and their influence on disease expression, there still remains a clear difference in cancer onset age and overall disease expression within Lynch syndrome families, independent of either a germline MMR mutation or genetic variance within the gene. Despite having an approximate lifetime risk of 80% of developing CRC, considerable differences in symptoms and disease diagnosis age are frequently observed in and between families who carry the same dysfunctional MMR gene or mutation [79-82].

Since the discovery of the genetic basis of this syndrome, numerous studies have been aimed at trying to determine whether there are any genetic modifying factors that may provide a significant difference in disease risk. This is of particular importance in families with Lynch syndrome as there are no pre-malignant markers or symptoms associated with it. The presence of a tumour in a particular family member gives no accurate prediction of the age of diagnosis that another family member may have in developing the disease. Comprehensive pedigree analyses of Lynch syndrome families indicate a considerable range of disease onset ages and phenotype [83]. The difference in disease expression can be explained by either environmental, genetic or a combination of both factors. Whilst there is evidence that several environmental and lifestyle risk factors affect CRC risk [84-90], they remain difficult to precisely define, which may in part be due to the subjective manner in which many of these studies are undertaken. Individuals included in such studies are often required to provide information on their dietary and lifestyle habits which is affected by poor recall thereby adversely influencing the outcome measures [91]. Genetic studies on the other hand offer more definitive clues in

CRC risk as patient genotypes are more easily and accurately defined. This is indeed the case in assessing modifying effects on CRC onset age in Lynch syndrome families. As there are relatively fewer patients with Lynch syndrome compared to those with sporadic disease there have been many more studies into genetic risk factors in sporadic cases compared to hereditary ones. Genome wide association studies where thousands of DNA samples are examined have revealed numerous single nucleotide polymorphisms (SNPs) which have been identified in having significant influence on CRC risk.

1.8 - Modifying Effects by Single Nucleotide Polymorphisms (SNPs)

The majority of research focused on identifying key modifier genes in Lynch syndrome has utilised the candidate gene approach where genetic polymorphisms have been shown to have functional consequences in their respective genes. Genes identified by these studies are often those associated with the cell cycle and DNA repair genes. Since polymorphisms in these genes may influence cell cycle regulation through subtle changes to the DNA structure, they make excellent candidate modifier genes, in particular in an impaired DNA repair system as found in Lynch syndrome. Numerous studies have focused on these genes; however in many cases, inconsistencies have arisen. An example is a common polymorphism, *R72P*, found within the tumour suppressor gene *TP53*. Initially this polymorphism had been associated with lung cancer susceptibility [92-96], then later in gastric cancer [97]. In 2004 a report from the United States described a significant association between the *R72P* polymorphism and earlier disease onset in Lynch syndrome patients [98]. The report revealed that patients who carried the heterozygous form of the polymorphism were significantly more likely to develop CRC 13 years earlier than those carrying the wild type. This finding was subsequently

followed up by a further two studies in which cohorts from Australia, Poland and Finland were included; however neither study could confirm the initial claim [99, 100]. Whilst there is a possibility that the observed discrepancy between these studies may have been due to differences in allele frequencies between the different populations, a more likely explanation is the reporting of a type 1 statistical errors. The original report from the United States was based on a cohort of 92 Caucasian patients, whereas both the Finnish and Australian/Polish follow up reports included the considerably larger numbers of 193 and 220 patients, respectively. Taking into account both these larger population sizes it is likely these provide a more accurate representation, compared to the initial report thereby supporting the notion of a type 1 statistical error.

Another gene, *ATM*, is involved in the recognition DNA errors and repair involving double strand breaks, whilst also regulating several oncogenes [101]. One report claimed that the *ATM D1853N* polymorphism was associated with an increased risk of developing colorectal cancer within Lynch syndrome families [102]. Although not specifically followed up at this stage, another report has since emerged claiming *D1853N* has no effect on disease onset age which casts some doubt on increased penetrance of CRC [103] in Lynch syndrome patients harbouring the *D1853N ATM* SNP.

The *R72P* and *ATM* polymorphisms are not the only examples where controversy exists in the literature in regards to Lynch syndrome modifier genes. There have been several studies into the genes which are involved in xenobiotic metabolism, including *NAT1* and *NAT2* both of which have pivotal roles in the acetylation of aromatic amines. A report emerged in 1999 claiming an association between polymorphisms within *NAT2* and

Lynch syndrome. The authors claimed a protective effect between the *NAT2* phenotype and CRC development in both *hMLH1* and *hMSH2* carriers in a total of 78 patients [104]. A further study of *NAT2* in 86 Lynch syndrome patients varied in some detail, however overall it appeared to give weight to the initial report, indicating that *NAT2* genotypes may be an important factor in CRC and other cancers related to Lynch syndrome [105]. Follow-up studies including populations residing in Australia, Poland and Germany, however failed to substantiate the original report [106, 107]. As both more recent reports included larger population sizes it is likely that these provide a more accurate representation with no overall significant association in CRC risk within Lynch syndrome cases.

The xenobiotic enzyme, glutathione-S-transferase (*GST),* is a further example of where a modifying effect has yet to be validated in Lynch syndrome expression. Genetic polymorphic forms of *GST* have been associated with the risk of various malignancies including bladder, lung and hepatocellular carcinomas [108-110]. Inactivating mutations within different forms of *GST* can lead to enzyme deficiency resulting in reduced detoxification thereby increasing disease risk and possibly modifying effects on disease expression. Studies of the modifying effects caused by *GST* variants have also been inconsistent with some studies reporting a significant association whereas others have not. A small study from Korea consisting of 104 HNPCC family members found an association between null *GSTM1* alleles and disease risk [111]. This was consistent with a pilot study from 59 males in South Africa which described similar results in *GSTM1* null alleles [112]. When examined in larger Australian and European populations however, the effect disappeared [100, 104]. To date one of the main problems in

searching for modifying genes has been the small sample sizes used to identify a true effect. From this it is clear that studies focused on potential genetic modifying effects in hereditary cancer syndromes such as Lynch syndrome require larger patient cohorts that are capable of providing greater statistical robustness to support any association with disease.

The study of modifier genes can potentially identify more specific effects in defined subsets of Lynch syndrome patients. At first pass there does not appear to be much variance between *MSH2* and *MLH1* mutation carriers as the presenting disease phenotype is very similar. Nevertheless, a consideration of the role of causative gene in this disease does need to be considered. For example a polymorphism residing within the cyclin D1 gene was claimed to be associated with the age of onset of disease in a group of patients from the United States in 2000 [113]. Follow-up studies on larger Lynch syndrome cohorts could not substantiate this finding [114, 115] suggesting that the original study was symptomatic of a type 1 error. When examined in a combined Australian and Polish population, no significant association could be found in the complete set of patients, but a significant association was observed in *hMSH2* carriers when the analysis was divided by MMR mutation type [116]. Intriguingly, the two groups where no association was observed were predominantly composed of patients with *MLH1* mutation carriers [113, 114] Whether or not this is indicative of a true representation in *hMSH2* carriers, it highlights the need to look into the specifics of the study cohorts for associations that may not be clear initially. This emphasises the importance of sufficient numbers of patients from different population groups to be included which would not only give a more robust estimate of modifying effects with greater sample sizes but also provide

valuable information in potential variation in allele frequency between ethnic groups and the effect this may have on disease expression.

1.9 Aim of Project

Many reports in the literature have provided valuable insight into the potential role of modifier genes but clearly much more information is required before the role of modifier genes can be implemented clinically. As previously discussed, genes involved in DNA repair and integrity as well as xenobiotic metabolism have been prime modifier gene candidates some of which will be examined in this project. Additionally there are new target candidate genes that may influence Lynch syndrome penetrance which continues to be identified and will be examined in this thesis.

Given that there is much unknown in the role of modifying genes in Lynch syndrome expression, the aims of this study are targeted towards polymorphisms in several previously identified candidate genes including:

- To further investigate the role of the *IGF-1* CA repeat polymorphism in Lynch syndrome disease expression,
- To further examine the association of two key polymorphisms located within the important folate metabolism gene *MTHFR* and their combined influence on Lynch syndrome disease expression,
- To further investigate the role of an intronic polymorphism located near the promoter region of key gene involved DNA methylation known as *DNMT3B* and any effect on Lynch syndrome disease expression, and

- To identify several target polymorphisms in target DNA repair genes and investigate the association in Lynch syndrome expression.

The assembly of a diverse range of Lynch syndrome families through collaboration is of great benefit for this project as it allows for the consideration of differences in allele frequencies between specific populations of participants and ads to the veracity of the results. The combining of these different populations in statistical analysis where significant polymorphic variation may occur will provide valuable information on reported disease expression both within certain populations as well as providing greater statistical robustness in the reported results.

It is anticipated that the increased understanding and validation of the role of these modifier genes in Lynch syndrome expression will contribute to further understanding of this disease. It is then hoped that this will identify targets for new treatments and may eventually lead to a more personalised approach to the medical care for Lynch syndrome patients.

Chapter 2

IGF-1 Gene Polymorphism and Risk for Hereditary Nonpolyposis Colorectal Cancer

Insulin-like growth factor-1 (*IGF-1*) has emerged as a candidate gene due to its integral role in the cell cycle and reported association with a range of malignancies including cancers of the lung, breast, prostate and CRC [117-121]. The differential effects of IGF-1 expression have been shown to be associated with polymorphisms located within and adjacent to the coding regions of the gene [122-125]. Polymorphisms in genes such as *IGF-1* that are associated with altered gene expression are excellent candidate modifier genes when considering their potential role in mediating disease risk.

The *IGF-1* gene has been reported in having a potential modifying effect in Lynch syndrome expression owing to a CA repeat polymorphism located near the promoter region [125]. It has been described that this gene may have an association with the average age of disease onset depending upon the number of CA polymorphic repeats each Lynch syndrome individual carries. [125]. In the following two chapters the CA repeat region of *IGF-1* is further examined across two ethnically diverse populations of Lynch syndrome participants to provide greater insight into this intriguing association.

STATEMENT I

This statement explains the contribution of all authors in the article listed below:

Reeves S, Meldrum C, and Scott, R.J., *IGF-1 Gene Polymorphism and Risk for Hereditary Nonpolyposis Colorectal Cancer*. J of the National Cancer Institute 2006; November 15; 98 (22): 1664-5.

Table I: Author contribution Percentage and Description of Contribution to the article listed above.

Author	Contribution (%)	Description of Contribution to Article	Signature
Stuart G. Reeves	85%	Experimental design, executed the experiment, performed statistical analysis. Wrote the manuscript.	
Cliff J. Meldrum	5%	Provided samples and clinical information.	
Rodney J. Scott	10%	Designed the study, provided the concept and corrected the manuscript.	

CORRESPONDENCE

Re: IGF-1 Gene Polymorphism and Risk for Hereditary Nonpolyposis Colorectal Cancer

Recently, Zecevic et al. *(1)* reported an association between the size of the CA-repeat sequence residing in the 5′ untranslated promoter region upstream of the start site of the IGF-1 gene and age of disease onset in 121 hereditary nonpolyposis colorectal cancer patients who harbored germline mutations in the mismatch repair (MMR) genes hMLH1 or hMSH2 *(1)*. In their study, an association between the length of the polymorphism and age of disease onset in patients harboring hMSH2 germline mutations was observed.

To determine if this relationship was applicable to other populations, to only hMSH2 mutation carriers, or to only men or women, we investigated the IGF-1 CA-repeat polymorphism in a total of 220 MLH1 and MSH2 mutation–positive patients from 36 families, including 123 probands/single family members with confirmed hMLH1 or hMSH2 germline mutations. Polymerase chain reaction conditions and CA-repeat analyses were as previously described *(1,2)*. Allele sizes were categorized as reported by Zecevic et al. *(1)* such that patients were grouped as having one allele with 17 or fewer CA repeats (≤17 CA) or both alleles with 18 or more CA repeats (≥18 CA).

A clear relationship was observed for early-onset disease in the patient group with 17 or fewer CA repeats. Overall, patients with 17 or fewer CA repeats were more likely than patients with 18 or more repeats to have early onset of colorectal cancer using the log-rank (LR) test (LR [df = 1] = 4.71, $P = .03$). Kaplan–Meier analysis also revealed a 15-year difference in the age of colorectal cancer onset between patients in the two groups (≤17 CA, median age = 48 years, 95% confidence interval [CI] = 44.5 to 51.3 years, range 21–84 years versus ≥18 CA, median age = 63 years, 95% CI = 51.8 to 74.2 years, range 21–95 years). When patients with hMSH2 and hMLH1 mutations were analyzed separately, a statistically significant difference in the age of disease onset was observed only among the hMLH1 mutation carriers (≤17 CA versus ≥18 CA, LR [df = 1] = 5.05, $P = .025$).

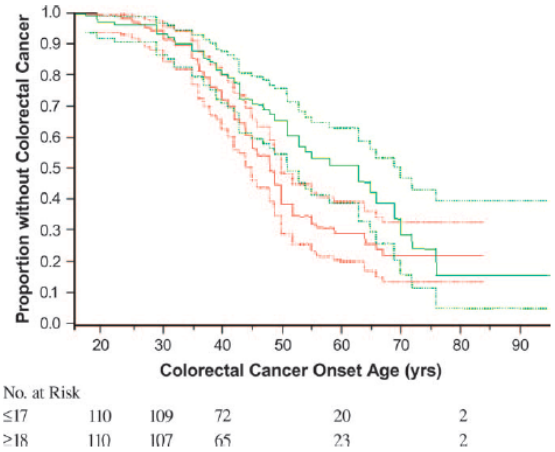

Fig. 1. Kaplan–Meier analysis of time to onset of colorectal cancer for patients with 17 or fewer IGF-1 gene CA repeats (**red**) and patients with 18 or more CA repeats (**green**). The **hatched** curves represent 95% confidence intervals. Kaplan–Meier analysis was used to conduct univariate analysis with mutation type and sex along with time to onset of colorectal cancer in regards to CA-repeat length. The age of diagnosis was used as the age of onset in all patients with colorectal cancer. For the unaffected mismatch repair gene mutation carriers, the age at last follow-up was used as the age of onset, and these subjects were censored in the analysis. All proportional hazard assumptions were verified using Schoenfeld residuals, and statistical tests were two-sided and performed using Intercool Stata 8.2 (Stata Corp, College Station, TX).

For proportional hazard Cox regression modeling, two models were tested including hMLH1–hMSH2 mutation group, IGF-1 CA-repeat group alone, and IGF-1 CA-repeat group plus sex with or without family clustering. No association between colorectal cancer risk and hMLH1–hMSH2 mutation or interaction between hMLH1–hMSH2 mutation group and the two CA-repeat IGF-1 groups was observed, which contradicted the Kaplan–Meier results based on MMR mutation type. The Cox model interpretation is preferred over the individual Kaplan–Meier survival curves, and it appears from the interaction test that it is the IGF-1 CA-repeat group alone, which is associated with risk of colorectal cancer and not the MMR mutation type. The final model from Cox regression analysis included the IGF-1 group plus sex and family as a cluster variable. The smaller number of IGF-1 CA repeats had the strongest association with colorectal cancer risk compared with the 18 or more CA group, (≤17 CA group, hazard ratio = 1.5, 95% CI = 1.02 to 2.16; $P = .044$). The association between sex and risk of colorectal cancer approached borderline statistical significance (compared with females, for males, hazard ratio = 1.40, 95% CI = 0.97 to 2.03; $P = .071$).

Our results indicate that the increased risk for colorectal cancer is equal in both hMLH1 and hMSH2 carriers. Although the association was not statistically significant in this study, the risk of colorectal cancer may be slightly more profound in males.

Stuart Reeves
Cliff Meldrum
Rodney J. Scott

References

(1) Zecevic M, Amos CI, Gu X, Campos IM, Jones JS, Lynch PM, et al. IGF-1 gene polymorphism and risk for hereditary nonpolyposis colorectal cancer. J Natl Cancer Inst 2006;98:139–43.

(2) Rosen CJ, Kurland ES, Vereault D, Adler RA, Rackoff PJ, Craig WY, et al. Association between serum insulin growth factor-I (IGF-I) and a simple sequence repeat in IGF-I gene: implications for genetic studies of bone mineral density. J Clin Endocrinol Metab 1998;83:2286–90.

Notes

Affiliations of authors: NBN Childhood Cancer Research Laboratory, University of Newcastle and the Hunter Medical Research Institute, Newcastle, New South Wales, Australia (SR, RJS); Division of Genetics, Hunter Area Pathology Service, John Hunter Hospital, Newcastle, New South Wales, Australia (CM, RJS).

Correspondence to: Rodney J. Scott, PhD, FRCPath, Division of Genetics, Hunter Area Pathology Service, John Hunter Hospital, Lookout Road,

New Lambton, Newcastle, NSW 2305, Australia (e-mail: rodney.scott@newcastle.edu.au).

DOI: 10.1093/jnci/djj452

RESPONSE

Hereditary nonpolyposis colorectal cancer (HNPCC) is an autosomal dominant disorder caused by DNA mismatch repair (MMR) gene mutations with hMLH1 and hMSH2 being the most frequently mutated *(1)*. Reeves et al. have genotyped a series of MMR gene mutation carriers for a CA-repeat polymorphism in the 5′ untranslated region of the insulin-like growth factor 1 (IGF-1) gene to determine whether findings from our recent study on this polymorphism were applicable to other populations. In our study, a group of 121 MMR gene mutation carriers for either MSH2 or MLH1 were genotyped for the IGF-1 gene polymorphism. We found a statistically significant association between shorter IGF-1 CA-repeat lengths and increased age-associated risk for HNPCC among the mutation carriers. When we stratified by MMR gene mutation, the association between HNPCC risk and IGF-1 CA-repeat length was statistically significant for MSH2 gene mutation carriers but not for MLH1 mutation carriers.

It is exciting that the work by Reeves et al. confirms the association between shorter IGF-1 gene CA-repeat lengths (≤17) and earlier disease onset in their study of 220 MMR gene mutation carriers. It will be interesting to find out whether shorter IGF-1 gene CA-repeat lengths influence risk for other cancer types as well.

When Reeves et al. stratified by MMR gene mutation, they observed the opposite from what we reported—a statistically significant association between HNPCC risk and IGF-1 CA-repeat length for MLH1 gene mutation carriers but not for MSH2 mutation carriers. Although it is possible that differences between the findings of the studies may have been due to differences in environmental and genetic backgrounds between the two study populations, it is also likely that they are due to the small sample sizes of the study groups created after stratification for the underlying MMR gene mutations. However, because the Cox regression modeling of Reeves et al. suggested that it is the IGF-1 CA-repeat group that was associated with the risk for HNPCC development and not the MMR mutation type per se, we tend to favor the latter hypothesis. Larger studies are needed to determine whether the type of MMR gene that is mutated influences the effect of the IGF-1 CA-repeat polymorphism.

Maja Zecevic
Christopher I. Amos
Marsha L. Frazier

Reference

(1) De la Chapelle A. Genetic predisposition to colorectal cancer. Nat Rev Cancer 2003;4:769–80.

Notes

Affiliations of authors: The Lancet, New York, NY (MZ); Department of Epidemiology, The University of Texas M. D. Anderson Cancer Center, Houston, TX (CIA, MLF).

Correspondence to: Marsha L. Frazier, PhD, Department of Epidemiology, The University of Texas M. D. Anderson Cancer Center, 1155 Pressler Blvd., Unit 1365, Houston, TX 77030 (e-mail: mlfrazier@mdanderson.org).

DOI: 10.1093/jnci/djj453

Chapter 3

IGF-1 is a Modifier of Disease Risk in Hereditary Nonpolyposis Colorectal Cancer (a more detailed analysis)

STATEMENT III

This statement explains the contribution of all authors in the article listed below:

Reeves S.G, Rich D, Meldrum C, Colyvas K, Kurzawski G, Suchy J, Lubinski, & Scott, R.J., *IGF-1 is a modifier of disease risk in hereditary non-polyposis colorectal cancer.* International Journal of Cancer. 2008 Sep 15; 123(6):1339-43.

Table III: Author contribution Percentage and Description of Contribution to the article listed above.

Author	Contribution (%)	Description of Contribution to Article	Signature
Stuart G. Reeves	65%	Experimental design, co-executed the experiment, co-performed statistical analysis. Wrote the manuscript.	
Dominique Rich	15%	Co-executed the experiment	
Kim Colyvas	5%	Co-performed statistical analysis	
Cliff J. Meldrum	2.5%	Provided samples and clinical information.	
Grzegorz Kurzawski	2.5%	Provided samples and clinical information.	
Janina Suchy	2.5%	Provided samples and clinical information.	
Jan Lubinski	2.5%	Provided samples and clinical information.	
Rodney J. Scott	5%	Designed the study, provided the concept and corrected the manuscript.	

Int. J. Cancer: **123,** 1339–1343 (2008)

IGF1 is a modifier of disease risk in hereditary non-polyposis colorectal cancer

Stuart G. Reeves[1], Dominique Rich[1], Cliff J. Meldrum[2], Kim Colyvas[3], Grzegorz Kurzawski[4], Janina Suchy[4], Jan Lubinski[4] and Rodney J. Scott[1,2]*

[1]*Discipline of Medical Genetics, Faculty of Health, University of Newcastle, and the Hunter Medical Research Institute, NSW, Australia*
[2]*Division of Genetics, Hunter Area Pathology Service, John Hunter Hospital, Newcastle, NSW, Australia*
[3]*Faculty of Science and Information Technology, School of Mathematical and Physical Sciences, University of Newcastle, NSW, Australia*
[4]*Department of Genetics and Pathology, International Hereditary Cancer Centre, Szczecin, Poland*

Patients diagnosed with HNPCC harbouring a confirmed germline mutation in DNA mismatch repair (MMR) genes have an 80% lifetime risk of developing an epithelial malignancy. There is, however, considerable variation in the age of disease onset in these patients. Insulin-like growth factor-I (IGFI) has been implicated in colorectal cancer (CRC), and elevated plasma IGFI levels are associated with both sporadic and hereditary CRC risk. In this study, we further investigate the cytosine-adenine (CA) dinucleotide repeat polymorphism located near the promoter region of IGF1 and its relation to early onset CRC risk in 443 Australian and Polish MMR gene mutation carriers using DNA sequencing, Kaplan-Meier survival curves and Cox proportional hazard regression analysis. A significantly smaller number of IGF1 CA repeats was observed in the Polish patient population, which was associated with an earlier age of disease onset compared to the Australian patients. The threshold for the observed modifying effect was again shown to be in patients with 17 or less CA repeats compared to those with 18 or more. Furthermore, when MMR mutation group (*i.e.*, MLH1 or MSH2), gender and family clustering were included in the final Cox model we observed a more robust trend for the role of the IGF1 CA repeat in predicting age of disease onset in HNPCC patients. In addition, this effect was shown to be equal in both MLH1 and MSH2 mutation carrier groups and not restricted to a particular MMR subgroup (p = 0.001). We conclude that the IGF1 CA repeat is an important modifier of disease onset in HNPCC and the first polymorphism to yield consistent results across different populations.

Key words: disease expression; HNPCC; IGF1 polymorphisms

Hereditary nonpoylposis colorectal cancer (HNPCC) is an autosomal dominantly inherited cancer predisposition characterized by the early onset of colorectal cancer (CRC).[1] HNPCC patients are also at an increased risk of developing a range of extracolonic epithelial malignancies such as endometrial, ovarian, stomach, biliary tract, small bowel, pancreas and brain tumours.[2] This syndrome is a result of inactivating germline mutations in DNA mismatch repair (MMR) genes, in particular MLH1, MSH2 that account for ~60% of HNPCC cases.[3] HNPCC is the most common form of hereditary CRC, with an incidence of ~1:2000; however, screening methods are less than 100% effective therefore these figures are likely to be underestimated.[4]

The age of CRC onset in HNPCC patients has been shown to vary quite considerably which is thought to be associated with both genetic and environmental influences. Previous studies have demonstrated polymorphisms in genes other than those associated with MMR may play a significant role in early onset disease risk in HNPCC but the results are controversial.[5,6]

Recently, a polymorphism in the IGF1 gene has been shown to be associated with the age of onset of disease in HNPCC.[7,8] Insulin-like growth factor 1 (IGF1) is a polypeptide that plays an important role in cell proliferation, differentiation and apoptosis and is essential for mammalian growth and development.[9] However, elevated IGF1 levels have been implicated in cell transformation, tumour growth and metastasis that are thought to be a result of the mitogenic and anti apoptotic effects of IGF1 on cells.[7,10] This has lead to several studies showing significant links between IGF1 and an increased risk of prostate cancer and CRC.[10–14]

Environmental and physiological factors including smoking, physical activity, stress and age can regularly cause fluctuations in IGF1 expression. However, genetic components have also been shown to alter IGF1 levels.[15,16] A region of particular interest has been the Cytosine–Adenine (CA) repeat polymorphism located in the untranslated region 969 bp upstream of the initiation start site of IGF1. The length of this CA repeat has been shown to influence transcription by altering promoter activity,[16] which is similar to the control of gene expression in other growth related genes such as EGFR (Epidermal Growth Factor Receptor) and acetyl-coenzyme carboxylase.[17,18]

The first report showing an association between age of disease onset and CA repeat length in the IGF1 gene indicated that only patients who harboured MSH2 gene mutations were significantly more are risk of developing malignancy if their IGF1 CA repeat length was less than or equal to 17 CA units. We have since shown that the IGF1 CA-repeat length, in 220 Australian MLH1 and MSH2 mutation carriers were associated with a more robust trend in predicting age of disease onset in HNPCC and that this effect was not restricted to patients harbouring MSH2 mutations.[8]

In this report, we have genotyped a further 223 HNPCC cases (all harbouring either an MLH1 or MSH2 mutations) of Polish origin for IGF1 CA-repeat polymorphisms. When combined with the previously reported data,[8] we find a significantly higher association between IGF1 CA-repeat length and the risk of early onset disease. Furthermore, we also have confirmed the cut off point of ≤17 CA repeats using Hazard ratio (HR) testing. We conclude that the IGF1 CA repeat is an important modifier of disease onset in HNPCC and is the first polymorphism in our knowledge to yield consistent results across different populations.

Material and methods

Patients

This study included a total of 443 HNPCC mutation carriers that comprised 249 Cases from 75 families and 194 unrelated individuals with confirmed causative MLH1 or MSH2 germline mutations. All missense mutations included in this study were deemed causative as there was evidence of their pathogenicity assessed by functional studies or segregation analysis in the literature. There was also no statistically significant difference in the average age of CRC diagnosis between patients harbouring missense mutations

Grant sponsor: NBN Children's Cancer Research Group, Hunter Medical Research Institute, The University of Newcastle and the Pomeranian Academy of Medicine.
***Correspondence to:** Division of Genetics, Hunter Area Pathology Service, John Hunter Hospital, Newcastle, NSW, Australia.
E-mail: rodney.scott@newcastle.edu.au or rodney.scott@hnehealth.nsw.gov.au
Received 27 November 2007; Accepted after revision 27 March 2008
DOI 10.1002/ijc.23668
Published online 11 July 2008 in Wiley InterScience (www.interscience.wiley.com).

Publication of the International Union Against Cancer

TABLE I – CLINICAL AND DEMOGRAPHIC CHARACTERISTICS OF THE PARTICIPANTS USED IN THIS STUDY

Characteristic	Colorectal cancer	No colorectal cancer	Total $N = 443$
Sex			
Female	119	142	261 (58.9%)
Male	100	82	182 (41.1%)
Age			
Median	–	41.5	
Mean (SD)		41.6 (16.2)	41.6
Range		16–95	16–95
Age of CRC			
Median	43	–	
Mean (SD)	43.3 (11.2)		43.3
Range	16–78		16–78
Proband			
Yes	135	59	194 (43.8%)
No	84	165	249 (56.2%)
MMR Mutation			
MLH1	122	136	258 (58.2%)
MSH2	97	88	185 (41.8%)
Mutation Type			
Truncation/ deletion	175	176	351 (79.2%)
Missense	44	48	92 (20.1%)

(41 years of age) compared with those with truncation/deletion mutations (43 years of age). Eighty-five healthy MMR proficient controls were also included from both Australian and Polish origins. The Institutional Ethics review boards of the Pomeranian Academy of Medicine and the Hunter New England Health Service approved the study. All participants gave written informed consent for the DNA samples to be used for research into HNPCC.

The clinical and demographic characteristics of the participants used in this study are shown in Table I.

PCR conditions and CA-repeat analysis

The CA repeat polymorphism 969 bp upstream of the transcription initiation site was genotyped using the forward 5′-GCTAGC CAGCTGGTGTTATT-3′ primer and the 5′-ACCACTCTGGGA GAAGGGTA-3′ reverse primer. Polymerase Chain Reaction (PCR) analysis was performed in 25 μl reactions using 50 ng of genomic DNA, 1× PCR buffer, 0.2 mM dNTP, 1 mM $MgCl_2$, 20 μM of each primer and 0.4 U of Platinum Taq Polymerase (Invitrogen). PCR reactions were performed in an Eppendorf PCR instrument (Eppendorf) using the following cycles: 94°C for 2 min, followed by 3 cycles of 94°C for 1 min, 61°C for 1 min and 72°C for 1 min then 29 cycles of 94°C for 30 sec, 61°C for 30 sec and 72°C for 2 min. This was followed by a final extension at 72°C for 6 min. IGF1 CA-repeat number was obtained through direct nucleotide sequencing of PCR products.

Statistical analysis

For the statistical analysis, 2 methods were employed for result interpretation. Kaplan-Meier (KM) analysis was used to conduct univariate analysis with mutation type and gender along with time to onset of CRC in regards to CA repeat length. Tests included in KM analysis were; the Log Rank Test (LR), which gives more weight to differences in the survival curves at higher ages relative to the Wilcoxon test (W) which give more weight to differences at lower ages, and finally the Tarone-Ware Test (TW) which gives an intermediate of the LR and W tests.

Cox proportional hazard regression models were then used to clarify KM results taking into account multiple variables including gender, MLH1/MSH2 mutation group and IGF1 repeat group with family clustering. This model also allowed for the testing of any interaction between MLH1 and MSH2 mutation groups. All statistical analysis was completed using Intercooled Stata 8.2 (Stata Corp, College Station, TX).

In the patients who had developed disease, age of CRC onset was defined as the patient's age of diagnosis. The age for the unaffected HNPCC carriers was determined by using their date of birth and disease free status at last follow-up, which was treated as censored in the analysis.

Results

Using sequencing analysis, the frequencies for both the total number of alleles and the shortest allele were determined in study participants and controls for the IGF1 repeat as shown in Tables II and III. In both the Australian and Polish populations the most common repeat length was 19 when both alleles were taken into consideration. No significant difference in the frequencies of the most common allele lengths (17, 18 and 19) were observed between participants and controls ($p = 0.8$). Our main focus, however, was on the shortest allele of IGF1 as it is in this repeat region where associations with CRC onset age have been previously reported.[7,8]

The frequency of the shortest IGF1 allele in the Polish population was determined as a percentage of a total of 223 alleles: 11 (1, 0.4%), 15 (4, 1.8%), 16 (52, 23.3%), 17 (126, 56.5%), 18 (37, 16.6%), and 19 (3, 1.3%) When combined with the Australian data, the combined frequencies in 443 alleles were as follows: 10 (0.7%), 11 (0.5%), 15 (0.9%), 16 (15.1%), 17 (49%), 18 (27.3%), 19 (6.3%) and 20 (0.2%).

Analysis of the shortest IGF1 allele within the population groups revealed differences between the two. Despite nearly equal numbers of participants in both groups (220 Australians compared to 223 Polish), the Australian participants overall had a higher representation of larger repeat lengths compared to the Polish in both participants and controls as seen in Tables II and III. The distribution of the IGF1 genotypes between the Australian and Polish groups clearly demonstrate population stratification when the polymorphism repeat numbers are considered independently with respect to the participants country of origin.

Nevertheless, in both population participant groups the most frequent shortest repeat number was 17, however, this accounted for more than half (56.5%) of the Polish participants studied compared to only 41.4% of Australian participants. In the Australian group, we also observed a similar number of participants harbouring the 18 CA repeat allele (38.2%) and a total of 50.1% of the Australian participants having 18 or more. In the Polish population, however, only 17.9% were in the 18 CA repeat allele group or greater and 79.8% of participants had either 17 or 16 CA repeat alleles. This was reflected in the Polish controls having a similar combined percentage for these alleles. Overall we observe a different distribution of IGF1 alleles in the Australian group compared to the Polish group. The difference in distribution of proportions for the IGF1 CA-repeat alleles between the 2 population groups was statistically significant ($p = <0.001$).

The previous 2 reports examining IGF1 CA repeat length and the risk of early onset colorectal cancer divided the participant groups into 17 or less compared with 18 or more CA repeats based on the majority of participants in these studies having either 17 or 18 repeats on the shortest allele and little variation outside of this range for statistical analysis to be performed.

We tested the cut off point of CA $\leq$ 17 or CA $\geq$ 18 to determine whether this arbitrary decision represented an associative difference in the control of IGF1 using a single variable Cox regression model with a 5 group version of IGF1 to determine hazard ratios and 95% confidence intervals (CI) on the combined population of participants. This analysis would reveal whether or not there was a differential relative risk of disease between different numbers of CA repeats. The results are shown in Table IV, where the repeat number of 19 is used as a reference point as it is reported to be the most common repeat length in the general population whilst also correlating with our data.[19,20]

TABLE II – IGF1 CA REPEAT PERCENTAGE DISTRIBUTIONS IN AUSTRALIAN AND POLISH STUDY PARTICIPANTS

	IGF-1 CA repeat number—Study participants									
Total Alleles	10	11	15	16	17	18	19	20	21	*n*
Australia	3 (0.7)	2 (0.4)	0	16 (3.6)	88 (20)	109 (24.8)	138 (31.4)	77 (17.5)	7 (1.6)	440
Poland	0	1 (0.2)	4 (0.9)	52 (11.7)	129 (28.9)	65 (14.6)	155 (34.8)	39 (8.7)	1 (0.2)	446
Combined	3 (0.3)	3 (0.3)	4 (0.4)	68 (7.7)	217 (24.5)	174 (19.6)	293 (33.1)	116 (13.1)	8 (0.9)	886
Shortest										
Australia	3 (1.4)	1 (0.5)	0	15 (6.8)	91 (41.4)	84 (38.2)	25 (11.4)	1 (0.5)	0	220
Poland	0	1 (0.4)	4 (1.8)	52 (23.3)	126 (56.5)	37 (16.6)	3 (1.3)	0	0	223
Combined	3 (0.7)	2 (0.5)	4 (0.9)	67 (15.1)	217 (49)	121 (27.3)	28 (6.3)	1 (0.2)	0	443

TABLE III – IGF1 CA REPEAT PERCENTAGE DISTRIBUTIONS IN AUSTRALIAN AND POLISH STUDY CONTROLS

	IGF-1 CA repeat number—Study controls									
Total Alleles	10	11	15	16	17	18	19	20	21	*n*
Australia	0	1 (0.6)	0	4 (2.4)	40 (23.5)	54 (31.8)	67 (39.4)	3 (1.8)	1 (0.6)	170
Poland	0	3 (1.8)	0	5 (2.9)	56 (32.9)	32 (18.8)	69 (40.6)	4 (2.4)	1 (0.6)	170
Combined	0	4 (1.2)	0	9 (2.6)	96 (28.2)	86 (25.3)	136 (40)	7 (2.1)	2 (0.6)	340
Shortest										
Australia	0	1 (1.2)	0	4 (4.7)	38 (44.7)	33 (38.8)	7 (8.2)	0	0	85
Poland	0	3 (3.5)	0	5 (5.9)	55 (64.7)	21 (24.7)	1 (1.2)	0	0	85
Combined	0	4 (2.4)	0	9 (5.3)	93 (54.7)	61 (35.9)	3 (1.8)	0	0	170

TABLE IV – HAZARD RATIOS AND 95% CONFIDENCE INTERVALS FOR IGF1 CA REPEAT LENGTHS IN RELATION TO EARLY ONSET CRC

Repeat length	Number of cases	Hazard ratio (HR)	95.0% confidence interval (CI)	
			Lower	Upper
≤15	9	1.85	0.586	5.830
16	67	1.90	0.965	3.718
17	217	1.60	0.861	2.976
18	121	0.98	0.503	1.888
≥19	29	1	–	–
≤17	293	1.70	1.25	2.31
≥18	150	1	–	–

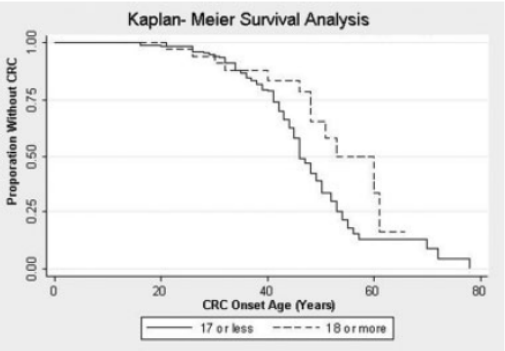

FIGURE 1 – Kaplan Meier Survival Analysis by IGF1 CA repeat genotype in Polish participants. The plots show the effect of the IGF1 genotypes on the risk of early onset cancer in the CA ≤ 17 group (lower unbroken line) compared to the CA ≥ 18 group (upper disjointed line).

Using 19 repeats as a reference it is observed that the 18 CA repeat allele group has a HR of nearly 1 (0.98), *i.e.*, nearly the same risk to that of the 19 CA repeat allele reference group. For the 17 CA repeat allele group, however, the HR increases to 1.6 and continues to increase for even lower numbers of repeats suggesting an increased risk for every decrease of 1 CA repeat but the greatest increase in risk was observed between 17 and 18 CA repeats. Combining the genotypes into the 2 groups of CA ≥ 18 and CA ≤ 17 gives a HR of 1 compared with 1.70, respectively. The change in the relative risk of disease between 17 and 18 CA repeat allele size provides statistically significant support that this is the most appropriate cut off point for the IGF1 effect. There was another observed small increase in risk between 17 and 16 repeats (HR 1.60–1.90), which stabilized in the ≤15 CA-repeat group that warrants further investigation in a larger population.

KM survival analysis revealed significant associations between the participants with CA repeat allele sizes ≤17 and ≥18 and an earlier onset of CRC. Breaking the results down into individual populations, a significant effect was observed in the Australian population for all tests (LR, $p = 0.03$, W $p = 0.03$, TW $p = 0.02$). In the Polish group KM analysis was also significant but the W test was marginally outside of the 0.05 significance value (LR $p = 0.03$, W $p = 0.08$, TW $p = 0.04$), see Figure 1. The reason for the less significant Wilcoxon test is due to smaller differences between the 2 CA repeat groups for early ages. Both population groups were then combined together in a final KM curve which proved to show a greater significant model overall. All 3 tests used were in agreement, including the (LR) test ($p = 0.001$), (W) test ($p = 0.009$) and finally the (TW) test ($p = 0.002$). This analysis additionally revealed a 12-year difference in the median age of CRC onset between patients in the CA ≤ 17 category (48 years) compared to the CA ≥ 18 category (60 years). A graphical representation of the KM curve for all 443 participants included in this study is shown in Figure 2.

Cox proportional hazard regression models were then used to test the strength of this association with multiple variables included and reported *p* values being based on the Wald test. When included in individual population based Cox models including all variable factors (*i.e.*, IGF1, gender and family) both Australian (HR 1.48, 95% CI 1.02–2.16, $p = 0.044$) and Polish (HR 1.95, 95% CI 1.01–3.75, $p = 0.046$) groups gave significant results overall, with both having a similar pattern indicating the CA ≤ 17 group has higher HR than participants in the CA ≥ 18 group.

Combining populations, the Cox model that included MMR mutation group (MLH1/MSH2) confirmed that this was not a significant factor as a main effect with HR = 0.98 for MSH2 relative to MLH1, 95% CI 74–1.30, $p = 0.88$. The interaction of MMR group with the 2 CA-repeat IGF1 groups was also not significant, $p = 0.86$. The final Cox model contained the IGF1 group plus gender and family as a cluster variable. An examination of Schoenfeld residuals indicated the proportional hazards assumption was not

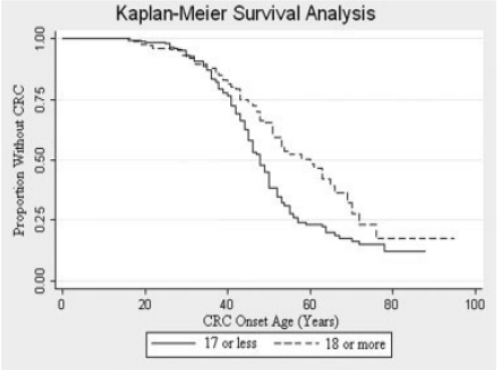

FIGURE 2 – Kaplan Meier Survival Analysis by IGF1 CA repeat genotype for the combined populations. The plots show the effect of the IGF1 genotypes on the risk of early onset cancer in the CA ≤ 17 group (lower unbroken line) compared to the CA ≥ 18 group (upper disjointed line).

broken. This model showed that the IGF1 group was the strongest predictor ($p = 0.001$) with the HR for the CA ≤ 17 group being 1.74 times greater than for the CA ≥ 18 group (95% CI 1.25–2.43). Interestingly gender was also shown to be significant in this final Cox model, which was independent of the IGF1 effect ($p = 0.010$, HR 1.41, 95% CI 1.09–1.82). This result suggested that men have a 1.41 times greater risk of developing earlier onset colorectal cancer compared to females in our combined population groups independent to any effect caused by IGF1.

Discussion

Overall our findings show a significant link between shorter IGF1 CA-repeat lengths and early onset CRC in patients who harbour MMR mutations. Both the Australian and Polish HNPCC cohorts displayed a clear association for early onset disease for people in the IGF1 CA ≤ 17 group. Interestingly, the Polish HNPCC population was found to have significantly shorter IGF1 repeat lengths compared to the Australian HNPCC population in this study. In addition, when both population groups were incorporated into a Cox regression model that included only country as a variable, a significant result was observed with an increased risk of earlier cancer onset in the Polish patients ($p = 0.01$, HR 0.70, 95% CI 0.53–0.92). When IGF1 was included into this model, however, the country effect lost its significance ($p = 0.10$, HR 0.79, 95% CI 0.60–1.05), suggesting IGF1 having an influence on Polish HNPCC cases to develop cancer at a younger age compared to the Australian group. This is an intriguing finding as it hints at the possibility of a greater or lesser disease risk between different populations as a result of population specific variations in IGF1 CA-repeat number.

KM survival analysis revealed a 12-year median difference in age of onset between the 2 general IGF1 groups when both populations were combined, resulting in significant *p*-values for the LR, W and TW tests. The Cox models are a preferred method of analysis as it provides a more robust form of analysis including all important variability factors. Utilizing Cox modeling, the combined power of 2 different population groups revealed a final significant value of $p = 0.001$ for the IGF1 effect which is to date the strongest association found between IGF1 CA repeat number and early onset HNPCC.

The gender effect was significant in the Polish group, $p = 0.05$ (HR 1.48) and approached significance in the Australian group $p = 0.07$ (HR 1.4). In each case, the HR for males was greater than the HR for females. This is independent of any IGF1 effect as the interaction between the 2 was not significant ($p = 0.93$), therefore, it must be attributed to some other factor. When IGF1 was tested in relation to gender no significant difference between the effect of this gene and disease risk was observed in either population or when they were combined. This indicates that any difference in the risk of colorectal cancer in males and females harbouring MMR mutations is not associated with variance in the IGF1 gene.

We confirmed by HR testing the cut of point of 17 IGF1 CA-repeats was most appropriate as there was the largest increase in the relative risk of disease between the CA repeat sizes of 17 and 18. There was another observed small increase in risk between 17 and 16 repeats stabilizing in the ≤15 CA-repeat group. This result is suggestive of a possible functional difference in the expression of IGF1 between these 2 groups with further investigations required to define disease risk into 3 CA repeat groups (≥18, 17, ≤16). To observe differences in overall relative risk between these groups, sufficient patients need to be included into a study to accurately characterize disease risk. Further studies may also include more separate population groups in order to determine any trends in IGF1 from country to country and perhaps the combined influence of other candidate polymorphisms such as cyclin D1, IGFBP3 and IRS1.[7,20,21]

In this study, we report the strongest association thus far found of the IGF1 CA-repeat effect on early onset colorectal cancer in HNPCC patients. We show that this trend is not limited to an individual population and may play a significant part in the differences in cancer risk between separate countries or populations depending on their IGF1 genotype.

Acknowledgements

The authors would like to thank Prof. Allan Spigelman and Ms. Claire Groombridge for their contribution to this study.

References

1. De La Chapelle A. Genetic predisposition to colorectal cancer. Nat Rev Cancer 2003;4:769–80.
2. DeFrancisco J, Grady WM. Diagnosis and management of hereditary non-polyposis colon cancer. Gastrointest Endosc 2003;58:390–408.
3. Scott RJ, McPhillips M, Meldrum CJ, Fitzgerald PE, Adams K, Spigelman A, du Sart D, Tucker K, Kirk J. Hereditary nonpolyposis colorectal cancer in 95 families: differences and similarities between mutation-positive and mutation-negative families. Am J Hum Genet 2001;68:118–27.
4. De La Chapelle A. The incidence of Lynch syndrome. Fam Cancer 2005;4:233–7.
5. Talseth BA, Meldrum C, Suchy J, Kurzawski G, Lubinski J, Scott RJ. Age of diagnosis of colorectal cancer in HNPCC patients is more complex than that predicted by R72P polymorphism in TP53. Int J Cancer 2006;118:2479–84.
6. Jones JS, Chi X, Gu X, Lynch PM, Amos CI, Frazier ML. p53 polymorphism and age of onset of hereditary nonpolyposis colorectal cancer in a Caucasian population. Clin Cancer Res 2004;10:5845–9.
7. Zecevic M, Amos CI, Gu X, Campos IM, Jones JS, Lynch PM, Rodriguez-Bigas MA, Frazier ML. IGF1 gene polymorphism and risk for hereditary nonpolyposis colorectal cancer. J Natl Cancer Inst 2006;98:139–43.
8. Reeves S, Meldrum C, Scott RJ. Re: IGF1 gene polymorphisms and risk for hereditary nonpolyposis colorectal cancer. J Natl Cancer Inst 2006;98:1664–5.
9. LeRoith D, Werner H, Beitner-Johnson D, Roberts CT. Molecular and cellular aspects of the insulin-like growth factor 1 receptor. Endocr Rev 1995;16:143–55.
10. Giovannucci E. Insulin, insulin-like growth factors and colon cancer: a review of the evidence. J Nutr 2001;131:109S–120S.
11. Pollak MN, Schernhammer ES, Hankinson SE. Insulin-like growth factors and neoplasia. Nat Rev Cancer 2004;4:505–18.
12. Chan JM, Stampfer MJ, Giovannucci E, Gann PH, Ma J, Wilkinson P, Hennekens CH, Pollak M. Plasma insulin-like growth factor-I and prostate cancer risk: a prospective study. Science 1998;279:563–6.
13. Kaaks R. Nutrition, energy balance and colon cancer risk: the role of insulin and insulin like growth factor-I. IARC Sci Publ 2002;156:289–93.

14. Sandhu MS, Dunger DB, Giovannucci EL. Insulin, insulin-like growth factor-I (IGF-I). IGF binding proteins, their biologic interactions, and colorectal cancer. J Natl Cancer Inst 2002;94:972–80.
15. Rotwein P, Pollock KM, Didier DK, Krivi GG. Organization and sequence of the human insulin-like growth factor I gene. J Biol Chem 1986;261:4828–32.
16. Rosen CJ, Kurland ES, Vereault D, Adler RA, Rackoff PJ, Craig WY, Witte S, Rogers J, Bilezikian JP. Association between serum insulin growth factor-I (IGF-I) and a simple sequence repeat in IGF-I gene: implications for genetic studies of bone mineral density. J Clin Endocrinol Metab 1998;83:2286–90.
17. Tae HJ, Luo X, Kim KH. Roles of CCAAT/enhancer-binding protein and its binding site on repression and derepression of acetyl-CoA carboxylase gene. J Biol Chem 1994;269:10475–84.
18. Gebhardt F, Zanker KS, Brandt B. Modulation of epidermal growth factor receptor gene transcription by a polymorphic dinucleotide repeat in intron 1. J Biol Chem 1999;274:13176–80.
19. Jernstrom H, Chu W, Vesprini D, Tao Y, Majeed N, Deal C, Pollak M, Narod SA. Genetic factors related to racial variation in plasma levels of insulin-like growth factor-1: implications for premenopausal breast cancer risk. Mol Genet Metab 2001;72:144–54.
20. Slattery ML, Samowitz W, Curtin K, Ma KN, Hoffman M, Caan B, Neuhausen S. Associations among IRS1. IRS2, IGF1, and. IGFBP3 genetic polymorphisms and colorectal cancer. Cancer Epidemiol Biomarkers Prev 2004;13:1206–14.
21. Wagner M, Hemminki K, Israelsson E, Grzybowska E, Soderberg M, Pamula J, Pekala W, Zientek H, Mielzynska D, Siwinska E, Försti A. Polymorphisms in the IGF1 and IGFBP3 promoter and the risk of breast cancer. Breast Cancer Res Treat 2005;92:133–40.

Chapter 4

MTHFR 677 C>T and 1298 A>C Polymorphisms and the Age of Onset of Colorectal Cancer in Hereditary Nonpolyposis Colorectal Cancer

Another reported modifier in Lynch syndrome expression is the Methylenetetrahydrofolate reductase (*MTHFR*) gene. MTHFR is a key folate-metabolizing enzyme involved in both DNA methylation and DNA synthesis. This enzyme is required for important steps in purine and thymidine synthesis which is necessary for methionine production. If thymidine synthesis is inadequate this can lead to uracil misincorporation during DNA synthesis, resulting in increased incidence of DNA damage and genetic instability. The pivotal role played by MTHFR has lead to polymorphisms associated with MTHFR gene function prime candidates as factors that may be involved in disease risk.

Two polymorphic regions within the *MTHFR* gene, *C677T* and *A1298C*, have been shown to influence the resultant enzyme activity to reduce folate utilization. Both *C677T* and *A1298C* have also been associated with a marked influence in folate metabolism and are reported to influence the risk of a range of malignancies including CRC in Lynch syndrome patients [126-129]. In this chapter we examine the role of these two important polymorphisms in diverse populations of Lynch syndrome participants to further clarify their role in overall disease expression.

STATEMENT IV

This statement explains the contribution of all authors in the article listed below:

Reeves S.G, Meldrum C, Groombridge C, Spigelman A.D, Suchy J, Kurzawski G, Lubinski J, McElduff P, & Scott, R.J., *MTHFR 677 C>T and 1298 A>C polymorphisms and the age of onset of colorectal cancer in hereditary nonpolyposis colorectal cancer.* The European Journal of Human Genetics 2009 17, 629–635.

Table IV: Author contribution Percentage and Description of Contribution to the article listed above.

Author	Contribution (%)	Description of Contribution to Article	Signature
Stuart G. Reeves	80%	Experimental design, co-executed the experiment, co-performed statistical analysis. Wrote the manuscript.	
Cliff J. Meldrum	2%	Provided samples and clinical information.	
Claire Groombridge	2%	Provided samples and clinical information.	
Allan Spigelman	2%	Provided samples and clinical information.	
Janina Suchy	2%	Provided samples and clinical information.	
Grzegorz Kurzawski	2%	Provided samples and clinical information.	
Jan Lubinski	2%	Provided samples and clinical information.	
Patrick McElduff	2%	Provided statistical analysis support	
Rodney J. Scott	6%	Designed the study, provided the concept and corrected the manuscript.	

European Journal of Human Genetics (2009) 17, 629–635

www.nature.com/ejhg

ARTICLE

MTHFR 677 C>T and 1298 A>C polymorphisms and the age of onset of colorectal cancer in hereditary nonpolyposis colorectal cancer

Stuart G Reeves[1], Cliff Meldrum[2], Claire Groombridge[3], Allan D Spigelman[3], Janina Suchy[4], Grzegorz Kurzawski[4], Jan Lubinski[4], Patrick McElduff[5] and Rodney J Scott[1,2]

[1]*Discipline of Medical Genetic, Faculty of Health, University of Newcastle, and the Hunter Medical Research Institute, New-Lambton Heights, Newcastle, NSW, Australia;* [2]*Division of Genetics, Hunter Area Pathology Service, John Hunter Hospital, Newcastle, NSW, Australia;* [3]*Hunter Family Cancer Service, Hunter New England Health Service, Newcastle, NSW, Australia;* [4]*International Hereditary Cancer Center, Department of Genetics and Pathology, Szczecin, Poland;* [5]*Faculty of Health, School of Medicine and Public Health, University of Newcastle, and the Hunter Medical Research Institute, New-Lambton Heights, Newcastle, NSW, Australia*

Hereditary non-polyposis colorectal cancer (HNPCC) or Lynch syndrome is characterized by inactivating germline mutations in DNA mismatch repair genes resulting in an increased risk of developing an epithelial malignancy. There is considerable variability in disease expression observed in this syndrome, which is thought to be due to a combination of genetic and environmental factors. Alterations in the kinetics of methylene tetrahydrofolate reductase (MTHFR) due to the presence of polymorphisms in the *MTHFR* gene have been associated with an increased risk of colorectal cancer (CRC). Two common single nucleotide polymorphisms (SNPs) located within the *MTHFR* gene, 677 C>T and 1298 A>C, that alter the function of the encoded protein have been the focus of many studies on CRC risk outside the context of an inherited predisposition to disease. In this report, a total of 417 HNPCC participants were genotyped for the 677 C>T and 1298 A>C SNPs to determine whether there exists an association with the age of disease onset of CRC. Genotyping of both SNPs was performed by TaqMan® assay technology. Associations in disease risk were further investigated using Kaplan–Meier survival analysis and Cox hazard regression. The average ages of disease diagnosis were found to be different between individuals harbouring either one of the *MTHFR* polymorphisms. Both Kaplan–Meier and Cox hazard regression analyses revealed a more complex relationship between the two polymorphisms and the age of CRC onset. The Kaplan–Meier survival analysis revealed that compound heterozygotes for the two SNPs developed CRC 10 years later compared with those carrying only wild-type alleles.
European Journal of Human Genetics (2009) 17, 629–635; doi:10.1038/ejhg.2008.239; published online 21 January 2009

Keywords: HNPCC; colorectal cancer; *MTHFR*; polymorphisms

Correspondence: Professor RJ Scott, Discipline of Medical Genetics, Faculty of Health, School of Biomedical Sciences, University of Newcastle, and the Hunter Medical Research Institute, Hunter Area Pathology Service, Lookout Road, New Lambton Heights, NSW 2305, Australia. Tel: +61 2 4921 4974; Fax: +61 2 4921 4253;
E-mail: rodney.scott@newcastle.edu.au
Received 28 February 2008; revised 22 October 2008; accepted 13 November 2008; published online 21 January 2009

Introduction

Hereditary non-polyposis colorectal cancer (HNPCC) is characterised by an increased risk of early onset of colorectal cancer (CRC).[1] HNPCC or Lynch syndrome is an autosomal-dominant disorder caused by deleterious germline mutations in DNA mismatch repair (MMR) genes.[1] Deficiencies in MMR activity can result in reduced

recognition and correction of mismatched bases during DNA replication. This results in an increase in genomic instability, which can be visualised in tumour DNA, where there is a loss of MMR activity. This is known as microsatellite instability (MSI), a hallmark of HNPCC tumours.[2] Mutations in the MMR genes *hMLH1* and *hMSH2* account for approximately 60% of HNPCC cases.[3] It has been commonly reported that individuals with HNPCC have an 80% lifetime risk of developing CRC by 70 years of age and this predisposition accounts for somewhere between 2 and 7% of all diagnosed cases. Analyses that are more recent suggest, however, that CRC penetrance has been significantly overestimated, being 47% and 34% for males and females, respectively.[4] The average age of onset of CRC is 44 years of age (as assessed from high-risk families) compared with 64 years in individuals who do not have this genetic predisposition.[5,6] In addition to CRC, there is an increased risk of other epithelial malignancies that include cancers of the endometrium, stomach, ovaries, uroepithelial and biliary tracts, small intestine and brain.[7]

Despite the presence of a germline mutation in a MMR gene being a strong predictor of disease, there is considerable variation in the phenotypic expression in HNPCC patients, in particular the age of CRC onset.[3] This appears to be largely independent of the type or location of MMR mutation, suggesting that genetic or environmental modifying effects influence the age of disease onset.

Methylene tetrahydrofolate reductase (MTHFR) is an essential enzyme in folate metabolism and subsequently plays a key role in DNA synthesis and methylation.[8] The role of this enzyme is to catalyse the irreversible reaction of 5,10-methyl-tetrahydrofolate (MTHF) to 5-MTHF, which is an integral part of the folate metabolism pathway. 5,10-MTHF is required for DNA synthesis and is in particular involved in uracil incorporation, whereas its product 5-MTHF is the methyl donor for regeneration of methionine from homocysteine for methylation reactions.[9] MTHFR activity can therefore affect levels of both 5,10-MTHF and 5-MTHF, both of which may influence the initiation and growth of tumour cells. Fluctuating amounts of 5,10-MTHF may lead to uracil misincorporation during DNA synthesis resulting in double-strand breaks,[10] whereas inconsistent amounts of 5-MTHF can affect methylation, therefore potentially influencing tumour suppressor or oncogene expression.[8,11]

Two common polymorphisms found within the *MTHFR* gene have recently been the focus of numerous studies on CRC risk.[8,9,12–18] The nucleotide polymorphism 677 C>T (rs1801133) is located within the region coding for the catalytic domain of MTHFR and results in an amino acid substitution from an alanine to a valine at codon position 222 (exon 4).[9,19] The 677 C>T variant has been associated with a reduced enzyme activity.[20,21] This single nucleotide polymorphism (SNP) has been implicated in CRC risk in several CRC patient populations;[17] however, conflicting results remain.[8] The second polymorphism, 1298 A>C (rs1801131), results in an amino acid change from a glutamine to alanine at codon position 429 (exon 7) and is found in a regulatory region of the MTHFR enzyme.[12] This polymorphism is also thought to cause a reduction in MTHFR activity, although its effect is considered to be less than that conferred by the 677 C>T change.[15] Further studies indicate that individuals heterozygous for both SNPs have a 50–60% decrease in MTHFR enzyme activity compared with their wild-type counterparts.[14]

Despite numerous studies examining associations of these two SNPs and CRC risk, there has only been one report that has specifically focused on the potential association of the *MTHFR* variants, 677 C>T and 1298 A>C, with the age of diagnosis of CRC in HNPCC.[18] In a small study by Pande *et al*[18] among 185 *hMSH2* or *hMLH1* mutation carriers, an ~4-year difference in the age of CRC diagnosis was observed in patients harbouring the 677 C>T polymorphism, whereas no effect was observed for the 1298 A>C SNP.

In this study, we investigated whether the 677 C>T and 1298 A>C SNPs in *MTHFR* influence the age of CRC diagnosis in a large group of 417 HNPCC cases, all of which harboured causative mutations in either *hMLH1* or *hMSH2*.

Methods

The sample population consisted of 220 (53%) Australian and 197 (47%) Polish participants. All participants harboured causative mutations in *MLH1* or *MSH2*. Both the Australian and Polish participants were all of North-Western European origin. The Australian patients enrolled in this study were clinic-based and recruited from family cancer clinics from the State of New South Wales. The Polish patients were recruited from the hereditary cancer clinics in and around the city of Szczecin. There were 194 (47%) *MLH1* and 193 (46%) *MSH2* cases with truncating or exon splice site mutations and 30 (7%) missense cases (deemed causative as there was pathogenicity evidence determined by functional studies or segregation analysis reported in the literature). There was no difference in the average age of disease diagnosis in the missense mutation carriers with disease compared with the truncating or exon splice site mutation carriers. Average age of missense mutation carriers was 41 years, whereas that of the nonsense mutation carriers was 43 years. To certify that the missense mutations were deleterious, expression analyses of MSH2, MLH1, PMS2 or MSH6 were performed by immunohistochemistry, and in cases with an ambiguous result, DNA microsatellite testing was undertaken. All tumours associated with missense mutations failed to express the respective protein or showed MSI (data not shown).

Of the 417 participants, 206 (49%) had been diagnosed with CRC as their first tumour. Only patients presenting with CRC as their first tumour were included in this study. The median age of the participants in this study with CRC was 43 years compared with 41 years for those without CRC. The Institutional Ethics Review Boards of the Pomeranian Academy of Medicine and the Hunter New England Health Service approved the study. All participants gave written informed consent for the DNA samples to be used for research into HNPCC. The clinical and demographic characteristics of the study participants are shown in Table 1.

SNP genotyping

Genotyping of the *MTHFR* 677 C>T and 1298 A>C SNPs was performed on the ABI PRISM® 7500 Real-Time (RT) PCR System (PE Applied Biosystems, Foster City, CA, USA), using primers and probes from Assay-by-Demand (Applied Biosystems) for 677 C>T (rs1801133, assay ID: 526886) and 1298 A>C (rs1801131, assay ID: 526686). The assay was performed under universal conditions, with each reaction containing 50 ng DNA, 0.125 µl 40× Assay Mix and 2.5 µl TaqMan® Universal PCR master mix made to a final volume of 5 µl with sterile water. Thermal cycling conditions were as follows: 50°C for 2 min, 95°C for 10 min, and 50 cycles of 92°C for 15 s and 60°C for 1 min. After the PCR reaction, plates were scanned by the ABI PRISM® 7500 PCR system to determine genotypes by allelic discrimination. Genotyping accuracy was assessed by repeating 10% of the DNA samples, which were randomly selected from both the control and affected patient populations. Concordance of 100% was observed between both genotypes for all samples.

Table 1 Demographics and genetic traits of HNPCC study participants according CRC status

	CRC (n=206)	*No CRC (n=211)*	*Total (n=417)*
Gender			
Male	83	73	156 (37.4%)
Female	123	138	261 (62.6%)
Age of onset (years)			
Mean (SD)	42.7 (11.4)	41.1 (15.2)	41.9 (13.3)
Median	43	41	42
Range	16–78	17–95	16–95
MMR mutation			
Truncation	188	199	387 (92.8%)
Missense	18	12	30 (7.2%)
MMR gene mutation			
MLH1	104	120	224 (53.7%)
MSH2	102	91	193 (46.3%)
MTHFR 677 C>T			
CC	105	101	206 (49.4%)
CT	83	91	174 (41.7%)
TT	18	19	37 (8.9%)
MTHFR 1298 A>C			
AA	92	86	178 (42.7%)
AC	89	98	187 (44.8%)
CC	25	27	52 (12.5%)

CRC, colorectal cancer; HNPCC, hereditary non-polyposis colorectal cancer; MMR, mismatch repair.

Statistical analysis

Differences in the average age of disease onset in the polymorphism carriers compared with patients wild type for both polymorphisms were determined using the Student's *t*-test.

The assessment of any association between *MTHFR* genotypes and the age of CRC onset was carried out by the survival analysis methods of Kaplan–Meier and Cox hazard regression modelling. Kaplan–Meier survival curves were used to plot the proportion of participants who were cancer free *versus* the patient age of diagnosis of CRC in relation to *MTHFR* genotype. This univariate survival analysis method used the Wilcoxon test, which emphasises observations from early diagnosis, the log-rank test, which gives more weight to later ages, and finally the Tarone–Ware test, which is an intermediate of the two previous tests. In cases where nonsignificant results were found for all three tests, only the log rank test was stated. Cox proportional hazard regression models were used to test significant findings found by Kaplan–Meier and to generate hazard ratios (HRs) and 95% confidence intervals (CIs) in a multivariate analysis taking into account MMR mutation, family member status and gender. Age of diagnosis was defined as patient age at the time of CRC diagnosis. For unaffected participants, age was based on the date of birth and disease-free status at last follow-up, being treated as censored in the analysis. All statistics were set at a significance level of $P \leq 0.05$. The statistical analysis in this study was carried out using Intercooled Stata 8.0 (Stata Corp., College Station, TX, USA).

Additional survival analysis was performed using haplotype pairs rather than SNPs as the predictor variable of interest. The haplotype analysis was implemented using a stochastic EM algorithm in a Cox proportion hazards regression framework.[22] Additional predictor variables included in the model were gender, MMR mutation type (missense or truncation/deletion) and MMR gene mutation. The model was fitted using THESIAS software.

Results

The *MTHFR* genotypes were determined for 677 C>T and 1298 A>C by RT-PCR. Both SNPs were found to be in the Hardy–Weinberg equilibrium in both the Polish and Australian participants. There was no significant difference in allele and genotype frequencies between the Australian

and Polish participants, which allowed for pooling of the genotyping results. The allele and genotype distributions by CRC status, mutation and gender are shown in Table 2. Using Lewontin statistics, the two SNPs were found to be in linkage disequilibrium with a *D* score of 0.96.

MTHFR genotypes and the average age of disease onset

A comparison of the average age of disease diagnosis revealed a significant difference ($P=0.008$) in the age of disease diagnosis in individuals who harboured one or more polymorphisms in *MTHFR* (43.72 years) compared with those CRC cases who did not harbour either SNP (37.62 years).

Association of the 677 C>T and 1298 A>C polymorphisms with CRC penetrance

The Kaplan–Meier survival analysis was undertaken to determine more precisely the relationship between the two polymorphisms and disease expression. No statistically significant association between the 677 C>T genotypes and the age of CRC onset (log rank, $P=0.57$) was detected (see Figure 1a). Also when the heterozygotes and variant homozygotes were compared with wild-type homozygotes no association was observed (test, $P=0.41$) (Figure 1b). Analysis of the *hMLH1* and *hMSH2* mutation carriers alone failed to reveal any association (log rank, $P=0.65$ and log rank, $P=0.86$, respectively). Subdivision of the group by gender also revealed no associations (males: Log rank test, $P=0.76$; females: Log rank test, $P=0.52$).

Genotyping of the *MTHFR* 1298 A>C SNP revealed similar results (Figure 2a). No significant difference in the age of onset was found for the individual 1298 A>C genotypes (log rank, $P=0.13$), *hMLH1* (log rank, $P=0.17$), *hMSH2* (log rank, $P=0.1$), nor by subdividing by gender, males (log rank, $P=0.79$) or females (log rank, $P=0.11$). When the heterozygotes and variant homozygotes were compared with wild-type homozygotes, however, a protective trend for 1298 A>C polymorphism was observed (Figure 2b). A significant association in two of the three tests was found (log rank, $P=0.045$ and the Tarone–Ware, $P=0.044$). The Wilcoxon test, which emphasises cases at an early age of diagnosis, showed a trend in the same direction ($P=0.053$). Overall, the Kaplan–Meier analysis suggested that the median age of diagnosis was 50 years of age for the heterozygous/homozygous variant 1298 A>C allele carriers compared with 47 years of age for the wild-type carriers. Multivariate Cox hazard regression modelling, including the variables MMR mutation, gender and

Table 2 *MTHFR* allele frequencies for 677 C>T and 1298 A>C for the study cohort

MTHFR 677 C>T	*CC*	*CT*	*TT*	*Any T allele*
All participants (n = *417*)	*206 (49.4%)*	*174 (41.7%)*	*37 (8.8%)*	*211 (50.6%)*
Allele frequency	*0.703*		*0.297*	
CRC status				
CRC+ (*n*=206)	105 (51%)	83 (40.3%)	18 (8.7%)	101 (49%)
CRC− (*n*=211)	101(47.9%)	91 (43.1%)	19 (9.0%)	110 (52.1%)
Mutation				
MLH1 (*n*=224)	105(46.7%)	95 (42.4%)	24 (10.7%)	119 (53.1%)
MSH2 (*n*=193)	101(52.6%)	79 (40.7%)	13 (6.7%)	92 (47.6%)
Gender				
Male (*n*=156)	72 (46.2%)	70 (44.9%)	14 (8.9%)	84 (53.8%)
Female (*n*=261)	134(51.3%)	104 (40%)	23 (8.7%)	127 (48.7%)
MTHFR 1298 A>C	*AA*	*AC*	*CC*	*Any C allele*
All participants (n = *417*)	*178(42.5%)*	*187 (44.9%)*	*52 (12.6%)*	*239 (57.5%)*
Allele frequency	*0.650*		*0.350*	
CRC status				
CRC+ (*n*=206)	92 (44.7%)	89 (43.2%)	25 (12.1%)	114 (55.3%)
CRC− (*n*=211)	86 (40.8%)	98 (46.4%)	27 (12.8%)	125 (59.2%)
Mutation				
MLH1 (*n*=224)	94 (41.8%)	103 (46.2%)	27 (12%)	131 (58.2%)
MSH2 (*n*=193)	84 (43.5%)	84 (43.5%)	25 (13.0%)	109 (56.5%)
Gender				
Male (*n*=156)	69 (44.2%)	67 (43%)	20 (12.8%)	87 (55.8%)
Female (*n*=261)	109 (41.8%)	120 (46%)	32 (12.2%)	152 (58.2%)

CRC, colorectal cancer.

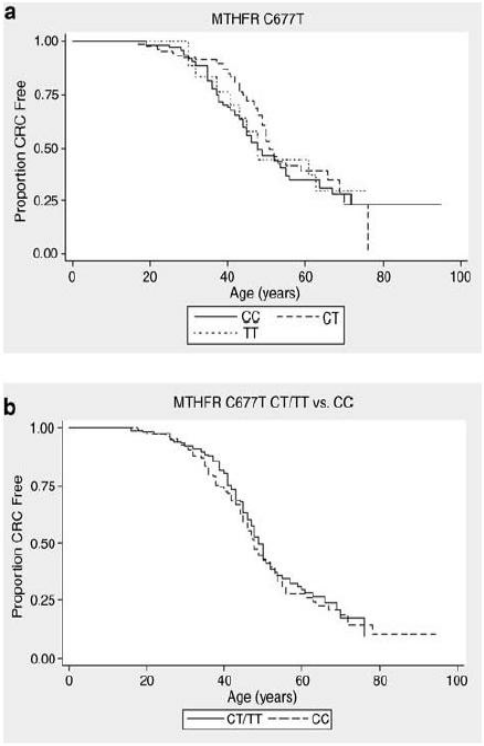

Figure 1 Kaplan–Meier survival curves for *MTHFR* 677 C>T representing time to age of CRC by genotype. (**a**) By wild type (CC), heterozygote (CT) and homozygote (TT) genotypes. (**b**) By combined heterozygote/homozygote variant (CT/TT) *versus* wild type (CC).

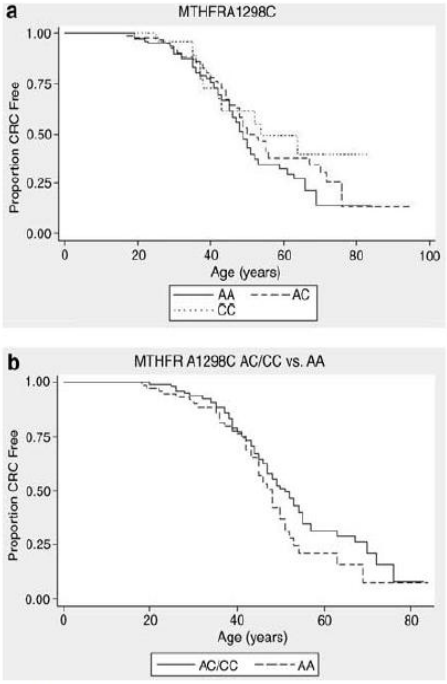

Figure 2 Kaplan–Meier survival curves for *MTHFR* 1298 A>C representing time to age of CRC by genotype. (**a**) By wild type (AA), heterozygote (AC) and homozygote (CC) genotypes. (**b**) By combined heterozygote/homozygote variant (AC/CC) *versus* wild type (AA).

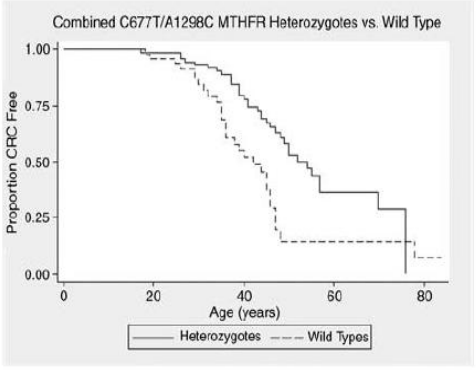

Figure 3 Combined Kaplan–Meier survival analysis for *MTHFR* 677 C>T and 1298 A>C. Plots comparing time with CRC onset by combined heterozygote (CT/AC) *versus* combined wild type (CC/AA).

family, was then used to determine the significance of the Kaplan–Meier analysis. Without the family variable, the association retained its borderline significance (HR: 0.79, 95% CI: 0.6–1.0, $P=0.073$), but when the family variable was included into this model the final outcome was non-significant ($P=0.11$). Despite this statistically nonsignificant result, the trend for wild-type carriers of 1298 A>C to develop CRC at an earlier age still remained.

Next, a joint analysis was performed between heterozygote forms of *MTHFR* 677 C>T (CT/TT) and 1298 A>C (AC/CC) compared with the wild types (CC/AA) for both SNPs. A strong association between the age of CRC onset and the combined *MTHFR* genotypes was detected (Figure 3). The log rank ($P=0.001$), Wilcoxon ($P=0.0008$), and Tarone–Ware ($P=0.0005$) tests were all highly significant with the median age of CRC onset 10 years later (52 years of age) in the participants heterozygous for both SNPs compared with those carrying both wild-type genotypes (42 years of age). This result was substantiated using Cox modelling where the variables of MMR mutation, gender and family were included. A significant *P*-value was obtained ($P=0.007$; HR: 0.47; 95% CI: 0.27–0.81) among the 133 patients who carried both heterozygous or both wild-type genotypes. Owing to the absence of patients carrying both homozygous variant genotypes, these were not assessed.

The haplotype analysis verified the results of the heterozygote forms of *MTHFR* 677 and 1298 compared with the wild types (CC/AA). The frequency of the AC, CC, AT and CT haplotypes were 36, 34, 29 and 1%, respectively. Compared with the AC haplotype, the HRs for the AT and CC haplotypes were similar and statistically significant with estimates of 0.70 (0.55, 0.90) for the AT haplotype and 0.66 (0.52, 0.84) for the CC haplotype. The HR for the CT haplotype was 1.06 (0.44–2.56).

Analysis of *hMSH2* mutation carriers alone carrying the combined 677 C>T/1298 A>C (CT/AC) genotype compared with those with only the wild-type alleles revealed a significant difference in the age of disease onset similar to that identified in the combined study population (log rank: $P=0.05$; Wilcoxon test: $P=0.037$; and Tarone–Ware test: $P=0.032$). Similarly, differences in the age of CRC onset were also observed for *hMLH1* mutation carriers (log rank: $P=0.005$; Wilcoxon test: $P=0.007$; and Tarone–Ware: $P=0.005$).

Discussion

MTHFR polymorphisms have been the focus of many studies and in particular for investigations into CRC where fluctuations in folate levels caused by 677 C>T and 1298 A>C variants potentially lead to an altered risk of cancer by subsequent variation of the deoxynucleotide pool.[8,9,12–18]

The results from this study are highly suggestive of a protective effect against CRC development in HNPCC individuals who harbour the heterozygote genotypes of 677 C>T and 1298 A>C in *MTHFR* compared with those carrying only wild-type alleles. Overall, a significant effect was observed for the patients who carry 677 C>T SNP (CT or TT) regarding the average age of disease onset compared with those who did not carry this SNP ($P=0.0279$). These findings are consistent with several studies that have shown the 677 C>T change to be associated with a reduced risk of bowel cancer in unselected CRC populations and in the study by Pande *et al*,[18] which reported a similar effect with the TT genotype being associated with later ages of CRC onset.

The differences in the average ages of disease onset indicate that there is a protective effect conferred by the presence of one or more polymorphisms in the *MTHFR* gene. This type of analysis does not, however, provide any information regarding age-specific differences that may become stronger or weaker as patients get older. To address this issue, a more specific measure of the effects of the polymorphism(s) is required. This has been achieved by examining where the greatest protective effects of the polymorphisms exist.

The *MTHFR* 1298 A>C SNP has also been found to have a similar protective effect in sporadic cases of CRC and our findings also indicate a protective effect in HNPCC as homozygote and heterozygote carriers together had an average age of disease onset of 44.32 years compared with non-carriers whose average age was 37.62 years ($P=0.009$). This is in contrast to Pande *et al*,[18] who did not observe any significant association in the age of disease onset in HNPCC patient with this SNP.

MTHFR has a fundamental position in folate metabolism catalysing the irreversible reaction of 5,10-MTHF to 5-MTHF. A decrease in 5-MTHF may lead to a reduction in DNA methylation, whereas 5,10-MTHF is required for thymidine synthesis,[10] emphasising the significance of MTHFR enzyme activity for these two important biochemical pathways. It has been shown earlier that individuals who carry the heterozygote forms 677 C>T and 1298 A>C have a 50–60% decrease in MTHFR activity compared with those with only wild-type alleles.[14] A reduction in MTHFR activity leads to greater quantities of its substrate 5,10-MTHF, required for DNA synthesis, and thereby reduces the availability of uracil. Misincorporation of uracil during DNA synthesis may result in double-strand breaks during DNA excision repair.[10] The increased pool of 5,10-MTHF pushes the folate metabolism towards DNA synthesis, in turn reducing the pool of uracil. Reduced amounts of uracil may also reduce the overall risk of uracil misincorporation owing to its limited availability. For individuals with MMR deficiency, the effect of reduced MTHFR enzyme activity may be a substantial advantage as uracil misincorporation could be particularly deleterious in conjunction with an impaired DNA repair pathway. The subsequent lower levels of 5-MTHF may also be beneficial due to a potential reduction in DNA methylation. Hypermethylation of the promoter of tumour suppressor or MMR genes may lead to gene silencing, therefore a reduction in methylation through decreased MTHFR activity could lead to less likelihood of this silencing occurring. Despite the risk of oncogene activation, a decreased level of DNA methylation may help maintain the integrity of the active allele in MMR-deficient individuals as seen in HNPCC cases.

This notion of methylation levels and uracil misincorporation being linked to age of CRC onset in HNPCC patients is shared by Pande *et al*;[18] however, our findings differ quite considerably. Our results are based on a larger sample size (419 *versus* 185), thereby providing more robust statistical support. The claim by Pande *et al*[18] of an earlier disease onset age for *hMLH1* mutation carriers harbouring only the wild-type allele *MTHFR* (C677C) could not be substantiated in our study. In the study by Pande *et al*,[18] the association observed was based on relatively few observations and it remains likely that this represents a type 1 statistical error rather than a true association.

The main limitation of our study is that we were unable to include any dietary or lifestyle factors, in particular levels of folate. Folate has been shown to be strongly linked to CRC in many studies,[23] with low levels of folate found to

increase risk by hypomethylation and the misincorporation of uracil. In contrast to this, however, folate deficiency in rapidly dividing cells as found in a tumour could lead to inefficient DNA synthesis and eventual apoptosis of the developing cancer. In this case, increased folate levels would promote malignant growths.[15] Evidently, folate status also influences MTHFR enzyme activity and would therefore be extremely useful information in future studies.

In conclusion, our findings indicate significant evidence for the combined *MTHFR* 677 C>T and 1298 A>C effect in CRC onset age for HNPCC cases. We speculate that this effect is because of the substantial influence of both these SNPs in folate metabolism, with both heterozygote forms required to make a significant impact on disease expression.

Acknowledgements

We acknowledge the support to the University of Newcastle, the Hunter Medical Research Institute and the Centre for Information Based Medicine in funding this study. We also thank the participants in the study for allowing their DNA to be used for further research into this disease.

References

1 de la Chapelle A: Genetic predisposition to colorectal cancer. *Nat Rev Cancer* 2004; 4: 769–780.
2 Peltomaki P: Deficient DNA mismatch repair: a common etiologic factor for colon cancer. *Hum Mol Genet* 2001; 10: 735–740.
3 Scott RJ, McPhillips M, Meldrum C *et al*: Hereditary nonpolyposis colorectal cancer in 95 families: differences and similarities between mutation-positive and mutation-negative kindreds. *Am J Hum Genet* 2001; 8: 118–127.
4 Alarcon F, Lasset C, Carayol J *et al*: Estimating cancer risk in HNPCC by the GRL method. *Eur J Hum Genet* 2007; 15: 831–836.
5 Aaltonen LA, Salovaara R, Kristo P *et al*: Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *N Engl J Med* 1998; 338: 1481–1487.
6 Samowitz WS, Curtin K, Lin HH *et al*: The colon cancer burden of genetically defined hereditary nonpolyposis colon cancer. *Gastroenterology* 2001; 121: 830–838.
7 DeFrancisco J, Grady WM: Diagnosis and management of hereditary non-polyposis colon cancer. *Gastrointest Endosc* 2003; 58: 390–408.
8 Plaschke J, Schwanebeck U, Pistorius S, Saeger HD, Schackert HK: Methylenetetrahydrofolate reductase polymorphisms and risk of sporadic and hereditary colorectal cancer with or without microsatellite instability. *Cancer Lett* 2003; 191: 179–185.
9 Brockton NT: Localized depletion: the key to colorectal cancer risk mediated by MTHFR genotype and folate? *Cancer Causes Control* 2006; 17: 1005–1016.
10 Blount BC, Mack MM, Wehr CM *et al*: Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: implications for cancer and neuronal damage. *Proc Natl Acad Sci USA* 1997; 94: 3290–3295.
11 Stern LL, Mason JB, Selhub J, Choi SW: Genomic DNA hypomethylation, a characteristic of most cancers, is present in peripheral leukocytes of individuals who are homozygous for the 677C>T polymorphism in the methylenetetrahydrofolate reductase gene. *Cancer Epidemiol Biomarkers Prev* 2000; 9: 849–853.
12 van der Put NM, Gabreels F, Stevens EMB *et al*: A second common mutation in the methylenetetrahydrofolate reductase gene: an additional risk factor for neural-tube defects? *Am J Hum Genet* 1998; 62: 1044–1051.
13 Weisberg I, Tran P, Christensen B, Sibani R, Rozen R: A second genetic polymorphism in methylenetetrahydrofolate reductase (MTHFR) associated with decreased enzyme activity. *Mol Genet Metab* 1998; 64: 169–172.
14 Weisberg IS, Jacques PF, Selhub J *et al*: The 1298A→C polymorphism in methylenetetrahydrofolate reductase (MTHFR): *in vitro* expression and association with homocysteine. *Atherosclerosis* 2001; 156: 409–415.
15 Le Marchand L, Donlon T, Hankin JH, Kolonel LN, Wilkens LR, Seifried A: B-vitamin intake, metabolic genes, and colorectal cancer risk (United States). *Cancer Causes Control* 2002; 13: 239–248.
16 Mitrou PN, Watson MA, Loktionov AS *et al*: MTHFR (677C>T and 1298A>C) polymorphisms and risk of sporadic distal colorectal adenoma in the UK Flexible Sigmoidoscopy Screening Trial (United Kingdom). *Cancer Causes Control* 2006; 17: 793–801.
17 Hubner RA, Houlston RS: MTHFR 677C>T and colorectal cancer risk: a meta-analysis of 25 populations. *Int J Cancer* 2007; 120: 1027–1035.
18 Pande M, Chen J, Amos CI, Lynch PM, Broaddus R, Frazier ML: Influence of methylenetetrahydrofolate reductase gene polymorphisms 677C>T and 1298A>C on age-associated risk for colorectal cancer in a Caucasian lynch syndrome population. *Cancer Epidemiol Biomarkers Prev* 2007; 16: 1753–1759.
19 Frosst P, Blom HJ, Milos R *et al*: A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat Genet* 1995; 10: 111–113.
20 Kang SS, Zhou J, Wong PWK, Kowalisyn J, Strokaosch G: Intermediate homocysteinemia: a thermolabile variant of methylenetetrahydrofolate reductase. *Am J Hum Genet* 1988; 43: 414–421.
21 Shan X, Wang L, Hoffmaster R, Kruger WD: Functional characterization of human methylenetetrahydrofolate reductase in *Saccharomyces cerevisiae*. *J Biol Chem* 1999; 274: 32613–32618.
22 Tregouet D-A, Tiret L: Cox proportional hazards survival regression in haplotype-based association analysis using the Stochastic-EM algorithm. *Eur J Hum Genet* 2004; 12: 971–974.
23 Kim YI: Folate and colorectal cancer: an evidence-based critical review. *Mol Nutr Food Res* 2007; 51: 267–292.

Chapter 5

The -149C>T SNP Within the ΔDNMT3B Gene is Not Associated with Early Disease Onset in Hereditary Nonpolyposis Colorectal Cancer

DNA methylation is a crucial part of the cell cycle and plays important roles in cellular development and genomic integrity. DNA methylation is regulated by a family of DNA methyltransferases (DNMTs), of which three active forms (DNMT1, DNMT3A and DNMT3B) have been identified in mammalian cells. Aberrant DNA methylation has been associated with a large number of human malignancies where hypermethylation of tumour suppressor genes and/or global hypomethylation may occur. Increased DNA enzyme activity of the DNA methyltransferases DNMT1, DNMT3A and DNMT3B, has been shown to be elevated in several types of disease including leukaemia, prostate, lung, breast and endometrial cancers which has lead to these becoming areas of interest for modifying effects in disease expression. A polymorphism located within DNMT3B drew particular attention as it had been suggested that in *in vitro* assays the C>T variant could alter promoter activity, leading to an increase of activity of up to 30%. Not long afterwards this polymorphism was reported in a small cohort of Lynch syndrome participants in the United States as significantly altering disease expression. In this chapter the DNMT3B C>T polymorphism is studied in greater detail in a diverse combined population of Lynch syndrome participants. The results described provide valuable information into the previously reported results of this polymorphism.

STATEMENT V

This statement explains the contribution of all authors in the article listed below:

Reeves, S.G., Mossman, D., Meldrum, C.J., Kurzawski, G., Lubinski, J., and Scott, R.J., (2008). *The -149C>T SNP within the ΔDNMT3B gene is not associated with early disease onset in hereditary non-polyposis Colorectal Cancer*. Cancer Letters, 265(1):39-44.

Table V: Author contribution Percentage and Description of Contribution to the article listed above.

Author	Contribution (%)	Description of Contribution to Article	Signature
Stuart G. Reeves	45%	Experimental design, executed the experiment, co-performed statistical analysis. Co-wrote the manuscript.	
David Mossman	40%	Experimental design, executed the experiment, co-performed statistical analysis. Co-wrote the manuscript.	
Cliff J. Meldrum	2.5%	Provided samples and clinical information.	
Grzegorz Kurzawski	2.5%	Provided samples and clinical information.	
Janina Suchy	2.5%	Provided samples and clinical information.	
Jan Lubinski	2.5%	Provided samples and clinical information.	
Rodney J. Scott	5%	Designed the study, provided the concept and corrected the manuscript.	

Available online at www.sciencedirect.com
ScienceDirect
Cancer Letters 265 (2008) 39–44
CANCER Letters
www.elsevier.com/locate/canlet

The −149C>T SNP within the ΔDNMT3B gene, is not associated with early disease onset in hereditary non-polyposis colorectal cancer

S.G. Reeves [a], D. Mossman [a], C.J. Meldrum [b], G. Kurzawski [c], J. Suchy [c], J. Lubinski [c], R.J. Scott [a,b,*]

[a] *Discipline of Medical Genetics, Faculty of Health, University of Newcastle, The Hunter Medical Research Institute, NSW, Australia*
[b] *Division of Genetics, Hunter Area Pathology Service, John Hunter Hospital, Newcastle, NSW, Australia*
[c] *International Hereditary Cancer Centre, Department of Genetics and Pathology, Szczecin, Poland*

Received 12 September 2007; received in revised form 31 January 2008; accepted 3 February 2008

Abstract

Hereditary non-polyposis colorectal cancer (HNPCC) is an autosomal dominantly inherited syndrome caused by germline mutations in mismatch repair (MMR) genes. HNPCC patients have a lifetime risk of 80% of developing colorectal cancer (CRC); however the likely age of onset is difficult to predict. A single C>T polymorphism located within the promoter region of the ΔDNMT3B gene has recently been reported to be associated with a significant increase to the risk of early onset CRC. In this study we determined the ΔDNMT3B genotype in 404 confirmed HNPCC participants (total of 194 CRC cases) from Australia (203) and Poland (201). From the total number of participants there were 194 diagnosed cases of CRC and 210 healthy MMR gene mutation carriers. The study was undertaken to assess whether the reported effect observed in a previous study of 146 HNPCC patients is consistent in a larger separate and unrelated participant cohort. Through the statistical tests of Kaplan–Meier survival analysis and Cox hazard regression models we did not observe any significant association between the ΔDNMT3B C>T SNP and early onset CRC in HNPCC patients.

Keywords: DNMT3B; Hereditary non-polyposis colorectal cancer; Disease expression; Epidemiology

1. Introduction

Hereditary non-polyposis colorectal cancer (HNPCC) is the most common form of hereditary colorectal cancer (CRC), accounting for approximately 2–5% of all CRC cases [1]. The disease is due to either a loss or reduced function of one of the DNA mismatch repair (MMR) genes MLH1, MSH2, MSH6 or PMS2 [2]. A reduction in the fidelity of MMR results in an increased likelihood of colorectal tumours and a heightened risk of other epithelial malignancies [3]. The MMR genes MLH1 and MSH2 are the most frequently mutated;

* Corresponding author. Address: Discipline of Medical Genetics, Faculty of Health, University of Newcastle, The Hunter Medical Research Institute, NSW, Australia.
E-mail address: rodney.scott@newcastle.edu.au (R.J. Scott).

0304-3835/$ - see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.canlet.2008.02.005

however, insult to these genes is not limited to HNPCC. Approximately 15–20% of sporadic cancers display features that are suggestive of MLH1 silencing [4] likely to have arisen from aberrant DNA methylation.

Three methyltransferase enzymes, namely DNMT1, 3A and 3B catalyse the DNA methylation process in humans. The DNMT1 form is largely responsible for duplicating the methylation pattern on the newly synthesized DNA strand following replication. The *de novo* methyltransferases, DNMT3A and 3B, however are thought to establish and co-ordinate methylation during times of development and in imprinting [5]. Due to their mechanism of action, it is thought that the *de novo* methyltransferases are responsible for the aberrant methylation of promoters. Expression of DNMT3A and 3B is elevated in multiple forms of cancer [6–8], and the DNMT3B methyltransferase levels have been shown to be 3.7-fold greater in colon tumour tissue compared to normal surrounding tissue [9].

Four enzymatically active variants of the DNMT3B gene have been identified namely, DNMT3B1, 3B2, 3B3 and 3B6. In 2006, a new isoform, ΔDNMT3B, has been identified and has been shown to be the predominant form of DNMT3B expressed in non-small cell lung cancer cell lines [10]. It is stated that the ΔDNMT3B version uses an alternate promoter located within intron 4 and exon 5 of the DNMT3B gene.

In 2004 it incorrectly reported that a C>T single nucleotide polymorphism (SNP) was present within the promoter of DNMT3B6 [11], as the SNP does not fall within the promoter region of DNMT3B6 (Accession No. NM_175850.1). In 2006 the same SNP was again incorrectly reported to fall within the promoter of DNMT3B [12] but this does not match the Genbank record for DNMT3B (Accession No. NM_006892.3). Both of these reports are referring to the same SNP that is within the promoter of the ΔDNMT3B isoform (located at position 1570351, Accession No. NT_028392), −149 bp from the transcription start site of ΔDNMT3B, which falls on intron 4 of the regular DNMT3B gene. The T variant of this SNP has been associated with a 2-fold increase in promoter activity [13], and the ΔDNMT3B C>T SNP has been reported to play a role in the age of onset of colorectal cancer [12], where it has been proposed that the T variant allele results in increased ΔDNMT3B expression, and aberrant *de novo* methylation, which is expressed as an earlier age of cancer onset.

In this study we have further investigated the role of the ΔDNMT3B SNP and its association with the age of colorectal cancer onset in 404 MSH2 and MLH1 mutation positive participants. Using restriction fragment length polymorphism analysis, Kaplan–Meier statistics and Cox hazard regression models we re-assessed the correlation between the ΔDNMT3B C>T SNP and early onset CRC in HNPCC patients.

2. Methods

2.1. Patients

All participants included in this study were enrolled after they had been evaluated at a family cancer clinic and were diagnosed with HNPCC. The selection criteria used for enrolment into this study was strictly defined on the basis of a molecular diagnosis of this syndrome due to the presence of a causative germline mutation in a DNA mismatch repair gene. Our study included a total of 404 participants (194 CRC cases) including 203 Australian and 201 Polish cases of which there were 143 unrelated probands and 261 patients from 69 families all with confirmed hMLH1 or hMSH2 germline mutations. The Institutional Ethics review boards of the Pomeranian Academy of Medicine and the Hunter New England Health Service approved the study. All participants gave written informed consent for the DNA samples to be used for research into HNPCC.

The clinical and demographic characteristics of the participants used in this study are shown in Table 1.

2.2. PCR conditions and C>T SNP analysis

The region containing the C>T SNP located −149 bp from the transcription start site of ΔDNMT3B were genotyped using the same primers as described previously by Jones et al. [12]. A 376 bp fragment spanning the region surrounding the SNP was generated using Polymerase Chain Reaction (PCR). Reactions were performed using 50 ng of genomic DNA, 1× PCR buffer, 0.2 mM dNTP, 1 mM $MgCl_2$, 1.2 μM of each primer and 0.2 U of Platinum Taq Polymerase (Invitrogen). Thermal cycling conditions involved a denaturation step at 94 °C for 5 min, 14 cycles of 94 °C for 30 s, 63 °C for 45 s (decreasing by 0.5 °C per cycle to 56.5 °C), 72 °C for 1 min then 20 cycles at the above conditions with an annealing temperature of 56 °C. This was followed by a final extension step at 72 °C for 10 min.

PCR products were digested for 16 h at 37 °C with the restriction enzyme AvrII (New England Biolabs) before being run on a DNA electrophoresis gel to determine results. Digestions were performed using 3 μl PCR product, 1× buffer (50 mM NaCl, 10 mM Tris–HCl, 10 mM $MgCl_2$, 1 mM dithiothritol), and 0.6 U of AvrII. This

Table 1
Clinical and demographic characteristics of study participants

Characteristic	Colorectal cancer	No colorectal cancer	Total $n = 404$
Population			
Australian	98	105	203
Polish	96	105	201
Gender			
Female	113	139	252
Male	81	71	152
Age			
Mean	–	46.5	46.5
Range	–	18–95	18–95
Age of CRC			
Mean	42.7	–	42.7
Range	16–78	–	16–78
Proband			
Yes	87	56	143
No	109	152	261
MMR mutation			
MLH1	108	128	236
MSH2	86	82	168
Mutation type			
Truncation/ deletion	171	173	344
Missense	23	37	60

*All missense mutations included were deemed to be pathogenic through previous functional assay studies. Age for participants without colorectal cancer is defined as age at last follow up.

cut the 376 bp PCR product only when the C>T SNP was present, producing two bands (136 bp and 240 bp) for a homozygote C>T change or three bands (376 bp, 240 bp and 136 bp) if a heterozygote change was present. For wild-type C/C genotypes the 376 bp product remained intact. Genotyping was repeated in a total of 5% of participants to confirm the original analysis; re-genotyping revealed a concordance rate of 100%.

2.3. Statistical analysis

All statistical analyses were completed using Intercool Stata 8.2 (Stata Corp, College Station, TX).

Kaplan–Meier (KM) analysis was initially used to conduct univariate analysis with models including MMR mutation type, and gender compared to the age of clinical onset of CRC in relation to the ΔDNMT3b C>T SNP genotype. Through several statistical tests including the long rank test, KM was able to identify the proportion of the patient population that are cancer free at the age of evaluation compared to those patients who have developed disease.

Cox proportional hazard regression models were employed to determine the association of early onset colorectal cancer risk with multiple variables. These models allowed us to take into consideration the variables of: gender, MLH1/MSH2 mutation group, ethnicity and ΔDNMT3b genotype in regards to family clustering (i.e. groups of participants belonging to the same family). *p*-Values generated through Cox modelling were supported with hazard ratios (HR) and 95% confidence intervals (CI) giving an overall more robust model.

The age of diagnosis was defined as the patient's age at the time of colorectal cancer diagnosis, whereas the age for unaffected MMR gene carriers was determined by using their date of birth and disease free status at last follow-up, which was treated as censored in the analysis.

3. Results

Utilising the technique of PCR a 376 bp fragment was generated. Restriction enzyme digestion followed by agarose gel electrophoresis allowed three different ΔDNMT3B genotypes (CC, CT, TT) to be identified as shown in Fig. 1.

Participants were genotyped for ΔDNMT3B by this method before results were tabulated. Table 2 shows a summary of the patient demographics included in the study, including CRC status, gender, mismatch repair mutation and type, ethnicity and allele frequency by ΔDMNT3B genotype. Both populations were assessed separately and found to be in Hardy–Weinberg Equilibrium (HWE). When analysed as a single population, the allele frequencies remained in HWE ($p = 0.18$).

KM analysis was performed on all genotypes generated to determine whether any significant differences between the ΔDNMT3B C>T SNP and early onset disease were present. P values were determined using the Log Rank (LR) test, which gives equal weight to all failures, Wilcoxons (W) test, which emphasises observations from early onset patients, and finally the Taron–Ware (TW) test that gives an intermediate of the LR and W tests. Table. 3.

The three different ΔDNMT3B genotypes (CC, CT, TT) were first individually tested for all participants however no significant result was found overall (LR, $p = .83$), (W, $p = .82$) and (TW, $p = .77$). We then divided the patient data into two more general groups namely wild-type homozygous (CC) and combined heterozygous/homozygous (CT or TT) SNP. This again however

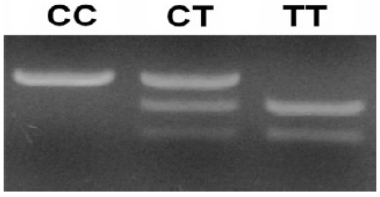

Fig. 1. DNMT3b genotypes (CC, CT, TT) generated by AvrII restriction enzyme digestion.

Table 2
Patient demographics by DNMT3B genotype

	CC, *n* (%)	CT, *n* (%)	TT, *n* (%)	Total
Colorectal cancer				
Yes	57(29.4)	91(47.0)	46(23.6)	194
No	63(30.0)	97(46.2)	50(23.8)	210
Gender				
Male	48(31.6)	70(46.0)	34(22.4)	152
Female	72(28.6)	119(47.2)	61(24.2)	252
MMR gene				
MLH1	68(28.8)	111(47.0)	57(24.2)	236
MSH2	52(31.0)	77(45.8)	39(23.2)	168
MMR mutation type				
Truncation/ deletion	106(30.8)	157(45.6)	81(23.5)	344
Missense	14(23.3)	31(51.7)	15(25.0)	60
Ethnicity				
Australian	61(30.0)	93(45.8)	49(24.2)	203
Polish	59(29.3)	95(47.3)	47(23.4)	201
Allele frequency				
C allele				0.5347
T allele				0.4653

Table 3
Hazard ratios and 95% confidence intervals by ΔDNMT3B genotype

Genotype	HR	95% CI	*p*-value
CT	0.94	[0.66, 1.31]	0.70
TT	0.94	[0.63, 1.42]	0.78
CT + TT	0.94	0.68, 1.29]	0.70
Per allele dose model	0.97	[0.79, 1.12]	0.77

Reference group for CT, TT and CT + TT is CC.
Per allele dose model refers to the individual CC, CT and TT genotypes.

showed no significant *p*-values for early onset disease associations (LR, $p = .54$), (W, $p = .63$), (TW, $p = .51$). Kaplan–Meier plots of these results are shown below in Fig. 2.

No significant difference was also observed in the separate population groups (Australian vs. Polish) or in the specific MMR mutation groups (MLH1 vs. MSH2), data not shown.

Cox regression models were then used to verify results shown by KM as well as to take into account any variables that may prove to be significant. Cox analysis on the individual genotypes (CC, CT, TT) along with the heterozygote, homozygote and combined heterozygote/homozygote forms were tested using the wild-type allele as a reference. All models included in this analysis contained gender and family clustering as additional variables resulting in non-significant *p*-values with hazard ratios pointing in the opposite direction as opposed to Jones et al. [12]. Additionally, dividing participants into groups of ⩽40 vs. >40 did not alter this non-significant association for either individual (CC, CT, TT) or combined (CC vs. CT/TT) genotypes. Individual gender affects along with separate populations and MMR mutation groups were also included in additional Cox models. These also confirmed the initial KM results showing no significant association overall.

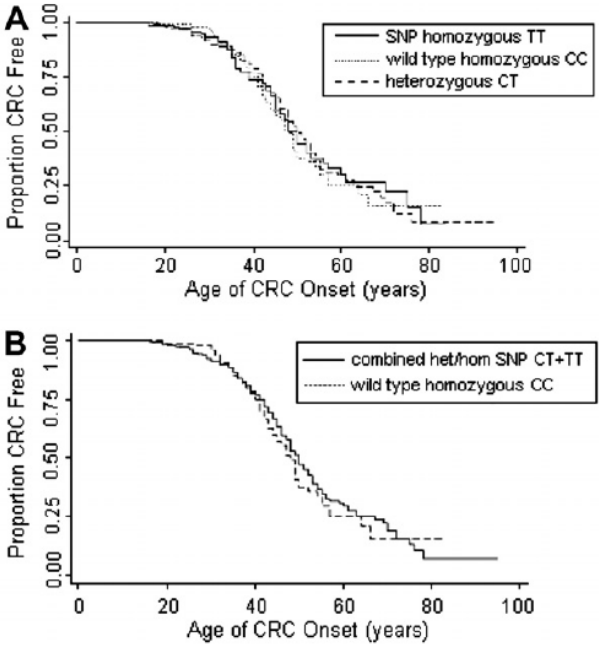

Fig. 2. Kaplan–Meier survival analysis by ΔDNMT3B genotypes The plots show the non-significant effect of the ΔDNMT3B genotype on age of onset for CRC in HNPCC patients. (A) By specific genotype and (B) by CT and TT genotypes compared to CC wild-type.

4. Discussion

We have analysed the ΔDNMT3B C>T SNP in 404 HNPCC participants to determine if the genotype correlates with the age of onset of colorectal cancer. Our results suggest no association between this SNP and age of CRC onset in our study populations.

This is inconsistent with the results published by Jones et al. [12] where a significant difference in age of onset between wild-type, heterozygous and homozygous mutant carriers in the ΔDNMT3B gene was reported. We utilised KM survival analysis and Cox hazard regression models to assess for differences amongst the three genotypes and additionally to compare wild-type against the combined heterozygotes and homozygote forms. Our results show the ΔDNMT3B C>T SNP did not alter age

of onset in either the Australian or Polish population, and no effect was observed when patients were grouped based on the type of MMR mutation they harboured (i.e. MLH1 or MSH2). The LR, W and TW tests were all used to assess for significant differences at different time points of the KM survival curves. The results of these tests however suggested no significant association for earlier disease onset age.

Cox regression modelling that included gender, MMR group and family clustering as variables also confirmed the KM results for no significant association for each individual genotype or for the homozygous/heterozygous C>T SNP versus wild-type. This was found to be the case for individual population groups (Australia and Poland) and additionally when both groups were combined together.

The ΔDNMT3B promoter SNP lies within an intronic region between the fourth and fifth transcribed exons of the longest isoform of the DNMT3B gene (DNMT3B1). It has been demonstrated that this SNP causes an increased level of expression of the ΔDNMT3B gene [13], and whilst it has been proposed that this SNP causes an earlier age of disease onset in HNPCC patients, our results provide evidence to suggest this is not the case. It is plausible however, that basic expression of ΔDNMT3B, rather than the presence of the ΔDNMT3B C>T SNP, is related to earlier age of disease onset. Wang et al. [10] reported that ΔDNMT3B is the predominant form of DNMT3B expressed in non-small cell lung carcinoma, and that a smaller percentage of matched healthy tissues express the ΔDNMT3B form of the gene. In this case, a patient who carries the ΔDNMT3B T allele may not develop disease any earlier than a patient with the C allele, as the ΔDNMT3B form is not expressed. This may also explain the discrepancies between our results and that of Jones et al. [12].

Considering the 3-fold larger participant size in this study, it is likely that a type 1 error may account for the discrepancy of results between these two studies. Jones et al. included only 12 confirmed CRC patients who carried the CC wild-type allele, 45 with a heterozygote CT and 17 patients with the homozygote TT allele. Statistical tests on such few patients in each category are therefore more likely to give an erroneous result due to the small sample size. Our cohort however contained 57 CC, 91 CT and 46 TT genotypes in patients with CRC; therefore the larger sample size supports greater strength to our statistical analysis.

Potential limitations of the current study may include population stratification however; we believe this should not affect overall outcomes as a true modifying polymorphism will affect disease expression in all HNPCC patients independent of population group. Additionally, no significant effect was observed for this SNP in either the Australian or Polish population by Cox analysis (results not shown) providing support of limited population stratification confounding effects. Different environmental influences on CRC onset age between the two countries, could potentially affect the results, but again this is unlikely as the average age of onset for the Australian and Polish groups were 42.8 and 43.5 years, respectively. The potential for individual environmental effects however cannot be entirely ruled out. Another limiting factor in this study is the reduced power to detect any small moderate effects that may be occurring. Despite being one of the larger HNPCC participant cohort studies, it would be beneficial to examine larger HNPCC populations thereby providing greater statistical rigour.

In conclusion, we have found no significant association for earlier cancer onset age in the Australian and Polish HNPCC populations or in a combined population cohort of 404 participants. Our findings lead us to conclude that there is no distinct correlation between the C>T SNP in the ΔDNMT3B promoter and early onset CRC in HNPCC patients and believe previous results reporting a positive association may be due to a type 1 error.

Acknowledgments

This work was support in part by the NBN Children's Cancer Research Group, Hunter Medical Research Institute, The University of Newcastle and the Pomeranian Academy of Medicine.

References

[1] H.T. Lynch, T.C. Smyrk, P. Watson, S.J. Lanspa, J.F. Lynch, P.M. Lynch, R.J. Cavalieri, C.R. Boland, Genetics, natural history, tumour spectrum, and pathology of hereditary nonpolyposis colorectal cancer: and updated review, Gastroenterology 104 (5) (1993) 1535–1549.

[2] H.T. Lynch, A. de la Chapelle, Genetic susceptibility to nonpolyposis colorectal cancer, Journal of Medical Genetics 36 (1999) 801–818.

[3] M. Aarnio, R. Sankila, E. Pukkala, R. Salovaara, L.A. Aaltonen, A. de la Chapelle, P. Peltomaki, J.P. Mecklin, H.J. Jarvinen, Cancer risk in mutation carriers of DNA-mismatch repair genes, International Journal of Cancer 81 (2) (1999) 214–218.

[4] S.N. Thibodeau, G. Bren, D. Schaid, Microsatellite instability in cancer of the proximal colon, Science 260 (1993) 816–819.
[5] H.D. Morgan, F. Santos, K. Green, W. Dean, W. Reik, Epigenetic reprogramming in mammals, Human Molecular Genetics 1 (14 review issue) (2005) r47–r58.
[6] J.R. Melki, P. Warnecke, P.C. Vincent, S.J. Clark, Increased DNA methyltransferase expression in leukemia, Leukemia 12 (3) (1998) 311–316.
[7] F. Jin, S.C. Dowdy, Y. Xiong, N.L. Eberhardt, K.C. Podratz, S.W. Jiang, Up-regulation of DNA methyltransferase 3B expression in endometrial cancers, Gynaecological Oncology 96 (2) (2005) 531–538.
[8] S.A. Belinsky, K.J. Nikula, S.B. Baylin, J.P. Issa, Increased cytosine DNA-methyltransferase activity is target-cell-specific and an early event in lung cancer, Proceedings of the National Academy of Science, USA 93 (9) (1996) 4045–4050.
[9] C. Schmutte, A.S. Yang, T.T. Nguyen, R.W. Beart, P.A. Jones, Mechanisms for the involvement of the DNA methylation in colon carcinogenesis, Cancer Research 56 (1996) 2375–2381.
[10] L. Wang, J. Wang, S. Sun, M. Rodriguez, P. Yue, S.J. Jang, L. Mao, A Novel DNMT3B subfamily, ΔDNMT3B, is the predominant form of DNMT3B in non-small cell lung cancer, International Journal of Oncology 29 (2006) 201–207.
[11] L. Wang, M. Rodriguez, E.S. Kim, Y. Xu, N. Bekele, A.K. El-Naggar, W.K. Hong, L. Mao, Y.W. Oh, A Novel C/T polymorphism in the core promoter of human de novo cytosine DNA methyltransferase 3B6 is associated with prognosis in head and neck cancer, International Journal of Oncology 25 (4) (2004) 993–999.
[12] J.S. Jones, C.I. Amos, M. Pande, Z. Gu, J. Chen, I.M. Campos, Q. Wei, M. Rodriguez-Bigas, P.M. Lynch, M.L. Frazier, DNMT3b polymorphism and hereditary nonpolyposis colorectal cancer age of onset, Cancer Epidemiology Biomarkers and Prevention 15 (5) (2006) 886–889.
[13] H. Shen, L. Wang, M.R. Spitz, W.K. Hong, L. Mao, Q. Wei, A novel polymorphism in human cytosine DNA-methyltransferase-3B promoter is associated with an increased risk of lung cancer, Cancer Research 62 (2002) 4992–4995.

Chapter 6

DNA Repair Polymorphisms and Risk of Early Onset Colorectal Cancer in Hereditary Nonpolyposis Colorectal Cancer

As previously mentioned the greater part of research into the search for modifying genes in Lynch syndrome expression has been focused on the candidate gene approach. Genetic polymorphisms similar to those described in *IGF-1* and *MTHFR* have been prime candidates as they have been shown to have functional consequences in their respective genes, and are often associated with the cell cycle or DNA repair. Polymorphisms which are capable of altering gene function in these important domains may influence cell cycle regulation through subtle changes to the DNA structure and make excellent candidate modifier genes, in particular in an impaired DNA repair system as found in Lynch syndrome. In this chapter a selection of polymorphisms located within the DNA repair pathway genes *BRCA2, hMSH3, Lig4*, *OGG1* and *XRCC* 1, 2 and 3 were chosen as these genes have been previously implicated in a wide range of malignancies. None of the selected polymorphisms had been previously been assessed for disease risk in Lynch syndrome, however were considered as prime candidates owing to their location within these genes. In this chapter the selected 8 polymorphisms *BRCA2* (rs11571653), *MSH3* (rs26279), *Lig4* (rs1805386), *OGG1* (rs1052133), *XRCC1* (rs25487), *XRCC2* (rs3218536 and rs1799793) and *XRCC3* (rs861539) were genotyped across the cohort of both Australian and Polish participants. With the utilization of multivariable statistical analysis the association of these DNA repair polymorphisms on CRC expression was investigated.

STATEMENT VI

This statement explains the contribution of all authors in the article listed below:

Reeves S.G, Meldrum C, Groombridge C, Spigelman A.D, Suchy J, Kurzawski G, Lubinski J, & Scott, R.J., *DNA repair polymorphisms and risk of early onset colorectal cancer in hereditary nonpolyposis colorectal cancer*. Cancer Epidemiology, September 2011 (Article in Press)

Table VI: Author contribution Percentage and Description of Contribution to the article listed above.

Author	Contribution (%)	Description of Contribution to Article	Signature
Stuart G. Reeves	85%	Experimental design, co-executed the experiment, co-performed statistical analysis. Wrote the manuscript.	
Cliff J. Meldrum	2%	Provided samples and clinical information.	
Claire Groombridge	2%	Provided samples and clinical information.	
Allan Spigelman	2%	Provided samples and clinical information.	
Janina Suchy	2%	Provided samples and clinical information.	
Grzegorz Kurzawski	2%	Provided samples and clinical information.	
Rodney J. Scott	5%	Designed the study, provided the concept and corrected the manuscript.	

G Model
CANEP 355 1–7
ARTICLE IN PRESS

Cancer Epidemiology xxx (2011) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

DNA repair gene polymorphisms and risk of early onset colorectal cancer in hereditary nonpolyposis colorectal cancer

Q1 Stuart G. Reeves[a], Cliff Meldrum[b], Claire Groombridge[c], Allan Spigelman[c], Janina Suchy[d], Grzegorz Kurzawski[d], Jan Lubinski[d], Rodney J. Scott[a,e,*]

[a] Discipline of Medical Genetic, Faculty of Health, University of Newcastle, and the Hunter Medical Research Institute, NSW, Australia
[b] Peter Macallum Cancer Institute, Melbourne, Victoria, Australia
[c] Hunter Family Cancer Service, John Hunter Hospital, Newcastle, NSW, Australia
[d] International Hereditary Cancer Center, Department of Genetics and Pathology, Szczecin, Poland
[e] Division of Genetics, Hunter Area Pathology Service, John Hunter Hospital, Newcastle, NSW, Australia

ARTICLE INFO

Article history:
Received 22 June 2011
Received in revised form 1 September 2011
Accepted 4 September 2011
Available online xxx

Keywords:
Lynch syndrome
DNA repair polymorphisms
Modifier genes
SNPs

ABSTRACT

DNA repair plays a pivotal role in maintaining genomic integrity with over 130 genes involved in various Q2 repair pathways that include base excision repair, nucleotide excision repair, double strand break repair and DNA mismatch repair. Polymorphisms within genes that are involved in these processes have been widely reported to be associated with cancer susceptibility in an extensive range of malignancies that include colorectal cancer (CRC). Lynch syndrome is caused by inherited germline mutations in DNA mismatch repair genes, predominantly in MLH1 and MSH2, that predispose to a variety of epithelial malignancies, most notably CRC. Despite being a relatively well understood hereditary cancer syndrome there remain several questions in relation to genetic influences on disease expression. Since Lynch syndrome is associated with a breakdown in DNA mismatch repair variation in other DNA repair genes may influence disease expression. In this report we have genotyped 424 Australian and Polish Lynch syndrome participants for eight common DNA repair gene polymorphisms to assess any association with the age of CRC onset. The DNA repair gene SNPs included in the study were: *BRCA2* (rs11571653), *MSH3* (rs26279), *Lig4* (rs1805386), *OGG1* (rs1052133), *XRCC1* (rs25487), *XRCC2* (rs3218536 and rs1799793) and *XRCC3* (rs861539). Cox multi-variant regression modelling failed to provide any convincing evidence of an effect in any of the polymorphisms analysed. The data suggest that polymorphisms in DNA repair genes do not contribute to cancer risk in a population of CRC patients who are at increased risk of disease as a result in a deficiency of DNA mismatch repair.

1. Introduction

Lynch syndrome is an autosomal dominant disorder characterized by an increased risk of epithelial malignancies, in particular colorectal cancer (CRC) [1,2]. This syndrome is associated with inactivating germline mutations in genes involved in DNA mismatch repair (MMR), with the genes *MLH1* and *MSH2* accounting for approximately 60% of all diagnosed Lynch syndrome cases [3]. It is the most common form of hereditary CRC and is estimated to account for somewhere between 2% and 7% of all CRC cases [4]. Initial reports on the penetrance of Lynch syndrome indicated an 80% lifetime risk of developing CRC [4,5]. More recent analyses suggest, however, that colorectal cancer penetrance has been significantly over-estimated and is more likely to be in the vicinity of 47% and 34% for males and females, respectively [6].

Despite being one of the better understood hereditary cancer syndromes, there is considerable variation in disease expression and no obvious gene-specific genotype/phenotype correlations have been demonstrated nor does there appear to be any relationship between the location of a mutation and type of disease. This variability in disease expression is especially evident in regards to the onset age of CRC which varies considerably between family members who have inherited the same causative DNA MMR gene mutation [3]. The best explanation for such differences in disease expression is likely to be either a result of genetic, environmental and/or a combination of both factors.

DNA repair is essential for the maintenance of genomic integrity which has evolved to ameliorate the effects of environment insult on

* Corresponding author at: Hunter Area Pathology Service, John Hunter Hospital, Lookout Road, New Lambton Heights, 2305 NSW, Australia. Tel.: +61 02 4921 4974; fax: +61 02 4921 4253.
E-mail addresses: rodney.scott@newcastle.edu.au, rodney.scott@hnehealth.nsw.gov.au (R.J. Scott).

1877-7821/$ – see front matter © 2011 Published by Elsevier Ltd.
doi:10.1016/j.canep.2011.09.003

G Model
CANEP 355 1–7
ARTICLE IN PRESS

DNA as well as correcting errors that occur during DNA synthesis. There are at least 130 genes involved in a variety of different DNA repair pathways that all play significant roles in maintaining the veracity of the genome [7]. The DNA repair pathways of mismatch repair and base excision repair are both involved in the identification, removal and repair of replication induced DNA errors. The mismatch repair system involves correcting mismatched bases that occur during DNA replication [8] whereas base excision repair (BER) is highly specific for the repair of oxidative DNA damage [9]. Loss of DNA MMR activity can be observed as changes in the size of microsatellite DNA, which is observed in tumours from Lynch syndrome patients and is known as microsatellite instability [10]. Several variants in DNA MMR genes have been implicated in cancer susceptibility and are not restricted to Lynch syndrome [11–13].

A series of enzymes are involved in BER including oxoguanine glycosylase 1 (*OGG1*) and X-ray repair cross-complementing 1 (*XRCC1*). Both of these genes harbour polymorphisms that have been associated with a change in the risk of malignancy. A codon 326 polymorphism in *OGG1* has been associated with the risk of lung, stomach and bladder cancer [14–16] and the codon 399 polymorphism in *XRCC1* has been reported to be associated with a decreased risk of bladder cancer but an increased risk of lung cancer [16,17].

Double-strand breaks (DSBs) in DNA can be repaired by either non-homologous end joining (NHEJ) or homologous recombination (HR). HR requires the use of the sister chromosome for the correct repair of DNA double strand breaks. *XRCC2* and *XRCC3* are two key components in HR that repair DNA double-strand breaks whilst also having essential roles in the maintenance of chromosome stability during cell division [8]. Polymorphisms in *XRCC2* and *XRCC3* have been linked to susceptibility of smoking related pancreatic and oral/pharyngeal tumours [18,19]. *BRCA2* is also involved in the repair of DSBs involving homologous recombination [20]. It is recognized that truncating mutations *BRCA2* are associated with breast cancer however polymorphisms in this gene have also been linked to an increased risk of not only breast cancer but also prostate cancer [21]. NHEJ requires DNA ligase 4 (*Lig4*) to form a complex in order orchestrate this repair process. Polymorphic variations in *Lig4* have also been associated with cancer susceptibility [22–24]. The recognition that many polymorphisms in DNA repair genes have been associated with cancer susceptibility suggests that altered repair function may explain some of the phenotypic differences observed in Lynch syndrome.

In this study we have genotyped a total of 424 Lynch syndrome participants all with causative *MLH1* or *MSH2* germline mutations for common polymorphisms in; *MSH3*, *OGG1*, *XRCC1*, *XRCC2*, *XRCC3*, *BRCA2* and *Lig4* to determine if one or more of the variants could be associated with the diversity of disease expression observed in Lynch syndrome.

2. Methods

2.1. Study subjects

A total of 424 participants, including 220 Australian and 204 Polish cases, were included in this study. The selection criteria for inclusion in this study was based on the molecular diagnosis of Lynch syndrome with the presence of either a *MLH1* (54.2%) or *MSH2* (45.8%) germline causative mutation. Within these cases 384 (90.6%) harboured a truncating or splice site mutation, whilst 40 (9.4%) were missense mutations. All missense cases were deemed causative as a result of functional studies, segregation analysis, and *in silico* analysis as reported in the literature. There was no difference in the average age of onset of patients with missense mutations compared to nonsense/truncating mutations. Out of the 424 participants 209 (49.3%) had been diagnosed with CRC. The median age in cases with CRC was 43 years compared to 40 years in those who were CRC free. The Institutional Ethics review boards of the Pomeranian Academy of Medicine and the Hunter New England Health Service approved the study. All included participants had given written informed consent for their DNA samples to be used for research into Lynch syndrome. The clinical and demographic characteristics of participants included are shown in Table 1.

Table 1
Demographics of HNPCC study participants according CRC status.

	Colorectal cancer ($n=209$)	No colorectal cancer ($n=215$)	Total ($n=424$)
Gender			
Male	82	76	158
Female	127	139	266
Age of onset			
Mean (SD)	42.8 (11.3)	40.5 (15.1)	41.9 (13.3)
Median	43	40	42
Range	16–78	17–95	16–95
MMR mutation			
Truncation	192	192	384
Missense	17	23	40
Gene mutation			
MLH1	108	122	230
MSH2	101	93	194
Ethnicity			
Australian	106	114	220
Polish	103	101	204

2.2. SNP genotyping

The SNPs used in this study were chosen due to their functional significance or tight linkage disequilibrium with function variants located in the respective gene. Genotyping for all DNA repair SNPs in participants was performed on the ABI PRISM® 7500 Real-Time (RT) PCR System (PE Applied Biosystems, Foster City, CA), using primers and probes from Assay-by-Demand (Applied Biosystems) for *BRCA2* A>G (rs11571653), *MSH3* A>G (rs26279), *Lig4* T>C (rs1805386), *OGG1* C>G (rs1052133), *XRCC1* G>A (rs25487), *XRCC2* G>A (rs3218536), *XRCC2* G>A (rs1799793) and *XRCC3* C>T (rs861539). The assay was performed under universal conditions with each reaction containing: 50 ng DNA, 0.125 μl 40× assay mix and 2.5 μl Taqman® Universal PCR master mix made to a final volume of 5 μl with Milli-Q water. Thermal cycling conditions were as follows: 50 °C for 2 min, 95 °C for 10 min, then 50 cycles of 92 °C for 15 s and 60 °C for 1 min. Following the PCR reaction, plates were then scanned by the ABI PRISM® 7500 to determine genotypes by allelic discrimination.

2.3. Statistical analysis

A power calculation was undertaken to determine whether there was a sufficient number of samples to reveal any association. Using the power calculator from the University of British Columbia (http://www.stat.ubc.ca/~rollin/stats/ssize/caco.html) the population used for this study had on average enough power to detect a RR of 1.5 with a p value of 0.05.

Using the standard χ^2 statistics, allele frequencies were tested for significant deviation from the Hardy–Weinberg equilibrium. Pearson's Chi-square test was also used to evaluate differences in the allele frequencies between cohort groups and distribution of genotypes. The assessment of the association between the DNA repair SNP genotypes and age of CRC onset was initially performed using Kaplan–Meier (KM) survival analysis. The KM survival curves were used to plot the proportion of participants who were cancer free against the patient age of CRC diagnosis in

Please cite this article in press as: Reeves SG, et al. DNA repair gene polymorphisms and risk of early onset colorectal cancer in hereditary nonpolyposis colorectal cancer. Cancer Epidemiology (2011), doi:10.1016/j.canep.2011.09.003

G Model
CANEP 355 1–7
ARTICLE IN PRESS

relation to specific genotypes. This univariate survival analysis method used three statistical tests including: the Wilcoxon test, emphasising observations from early diagnosis, the log-rank test, which gives greater weight to later onset ages and the Tarone–Ware test which is an intermediate of both the Wilcoxon and log rank tests. In cases where a significant value was obtained in KM testing Cox proportional hazard regression modelling was used to test the significant findings of KM analysis more rigorously. Through these Cox models hazard ratios (HR) and 95% confidence intervals (CI) could be generated in a multivariate analysis taking into account population group, *MLH1*/*MSH2* mutation, family clustering and gender. The age of diagnosis was defined as patient age at the time of CRC diagnosis. In the unaffected participants, age was based on date of birth and disease free status at last follow-up which was treated as censored in the analysis. Bonferroni corrections were applied to all statistical results to account for multiple testing. All statistics were set at a significance level of $p \leq 0.00625$. The statistical analysis in this study was carried out using Intercooled Stata 8.0 (Stata Corp., College Station, TX).

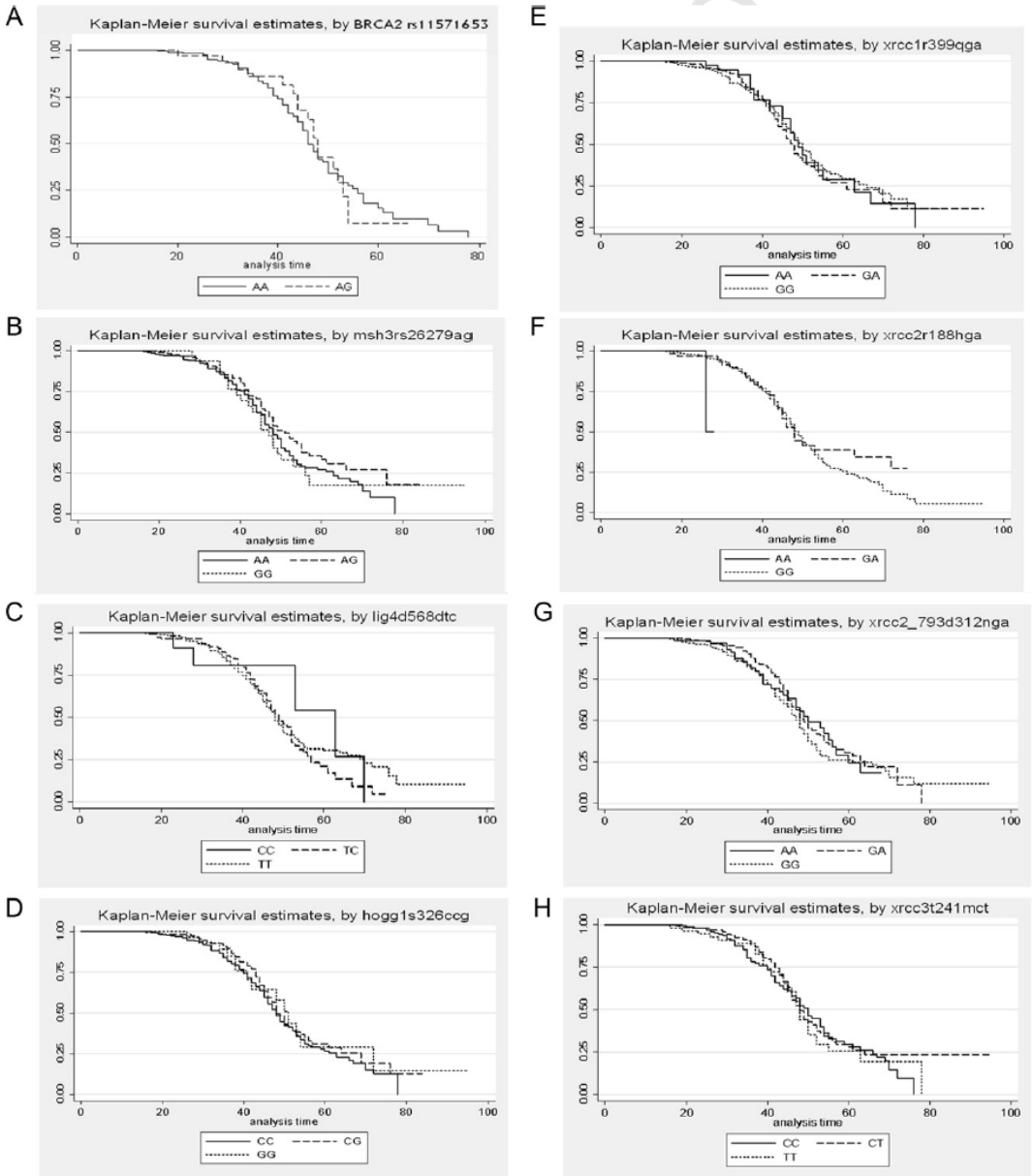

Fig. 1.

Please cite this article in press as: Reeves SG, et al. DNA repair gene polymorphisms and risk of early onset colorectal cancer in hereditary nonpolyposis colorectal cancer. Cancer Epidemiology (2011), doi:10.1016/j.canep.2011.09.003

G Model
CANEP 355 1–7
ARTICLE IN PRESS

3. Results

The genotypes for all DNA repair gene polymorphisms were successfully determined for all participants by RT-PCR. The distribution of alleles within both ethnic groups was consistent with Hardy–Weinberg equilibrium. The minor allele frequency distributions differed significantly by ethnicity for three of the eight SNPs included in this study, however when the two populations were combined the genotype frequency was similar to that recorded in the National Center for Biotechnology Information database. The minor allele frequencies reported in the overall population under study and for the separate Australian and Polish populations included in the analysis are shown in Table 2.

Kaplan–Meier survival analysis failed to show any significant association of the DNA repair SNPs examined when combined in an overall analysis of all 424 participants (see Fig. 1). This was also the case when the analysis was repeated using wild type vs. heterozygote/homozygote cases for the SNPs studied for either Australian or Polish cohorts or in either *MLH1* or *MSH2* gene mutation carriers alone (see Fig. 2a and b, respectively). Data output for KM analysis for the combined and individual Australian and Polish groups are displayed in Table 3.

Examination of the individual ethnic groups revealed no overall significant difference by KM analysis for the Australian participants for all SNPs studied. In the Polish group however, differences were observed in the minor allele frequencies of the *MSH3* A>G SNP (rs26279) and *XRCC2* G>A (rs1799793) SNPs. The presence of the *MSH3* SNP as either a heterozygote or homozygote was compared against wild-type carriers resulting in a trend in age of disease diagnosis attributed to the presence of the SNP. This was, however, not significant at all age groups studied as determined by the Log rank ($p = 0.028$), Wilcoxon ($p = 0.009$) and Tarone–Ware ($p = 0.009$) tests when correction for multiple testing was included. Similarly, the results for the *XRCC2* G>A SNP (rs1799793) did not reach the significance threshold required for multiple testing but did display a trend towards an association.

4. Discussion

In this study we explored the association of several common DNA repair pathway polymorphisms and risk of early onset CRC in Lynch syndrome. The frequency distributions of three of the minor alleles in MSH3, Lig4 and OGG1 varied significantly by ethnicity in population specific cases but when considered within the context of the entire study cohort remained similar to the frequencies listed in the National Center for Biotechnology Information (NCBI) data base. The difference in frequencies between the two populations is most likely a result of specific genetic differences between the Anglo-Celtic and the Slavic populations. There are several reports indicating that there are founder mutations in Poland, most notably the 5382insC mutation in *BRCA1* which is the most common mutation identified in that population whereas it is not frequently observed in the Anglo-Celtic population [25].

The association between the majority of the polymorphisms investigated and risk of an earlier age of cancer diagnosis was not significant in all eight polymorphisms. No trend at all for *BRCA2* A>G (rs11571653), *Lig4* T>C (rs1805386), *OGG1* C>G (rs1052133), *XRCC1* G>A (rs25487), *XRCC2* G>A (rs3218536), and *XRCC3* C>T (rs861539) was observed, suggesting that in the context of Lynch Syndrome these SNPs do not modify disease risk.

Two SNPs in the Polish population did, however, demonstrate a trend in altering the age of disease diagnosis in Lynch Syndrome. The SNP rs26279 located in MSH3 and rs1799793 in *XRCC2* both were associated with a decreased age of disease diagnosis. These two SNPs were not found to be significant in the Australian population. Both *MSH3* and *XRCC2* have pivotal roles in the DNA mismatch repair and double strand break repair, respectively. The failure to detect an effect in the Australian population may be

Table 2
Frequency distribution of minor alleles among participant cohort by characteristic for combined, Australian and Polish HNPCC participants. *p*-Value refers to difference in allele frequency between Australian and Polish populations.

SNPs	Minor allele	NCBI listed frequency	Combined SNP frequency	Australian SNP frequency	Polish SNP frequency	Frequency of SNP in CRC +ve patients	Frequency of SNP in CRC −ve patients	Frequency of SNP in MLH1 mutation carriers	Frequency of SNP in MSH2 mutation carriers	*p*-Value
MSH3 (rs26279)[a]	G	0.312	0.241	0.279	0.201	0.246	0.235	0.228	0.255	0.006
OGG1 (rs1052133)[a]	G	0.224	0.213	0.257	0.168	0.211	0.214	0.189	0.204	0.001
XRCC1 (rs25487)	A	0.303	0.325	0.323	0.329	0.328	0.323	0.296	0.322	0.859
XRCC2 (rs3218536)	A	0.092	0.081	0.101	0.061	0.074	0.088	0.076	0.091	0.039
XRCC2 (rs1799793)	A	0.314	0.376	0.341	0.415	0.352	0.401	0.335	0.387	0.028
XRCC3 (rs861539)	T	0.417	0.361	0.359	0.363	0.371	0.351	0.339	0.397	0.912
Lig4 (rs1805386)[a]	C	0.164	0.171	0.134	0.212	0.171	0.172	0.174	0.168	0.003
BRCA2 (rs11571653)	G	0.006	0.068	0.057	0.081	0.072	0.065	0.072	0.064	0.265

[a] Allele frequency was significantly different between the Australian and Polish populations.

Table 3
Statistical analysis of the combined and individual populations against the respective single nucleotide polymorphisms (SNP).

Gene SNP (rs number)	Kaplan–Meier statistical test								
	Combined			Australian			Polish		
	Log rank (*p*)	Wilcoxon (*p*)	Tarone–Ware (*p*)	Log rank (*p*)	Wilcoxon (*p*)	Tarone–Ware (*p*)	Log rank (*p*)	Wilcoxon (*p*)	Tarone–Ware (*p*)
MSH3 A>G (rs26279)	0.238	0.408	0.311	0.647	0.188	0.284	0.085	0.066	0.074
OGG1 C>G (rs1052133)	0.565	0.484	0.544	0.537	0.387	0.462	0.541	0.479	0.101
XRCC1 G>A (rs25487)	0.767	0.834	0.804	0.957	0.637	0.858	0.561	0.553	0.537
XRCC2 G>A (rs1799793)	0.358	0.519	0.315	0.761	0.966	0.961	0.145	0.587	0.452
XRCC2 G>A (rs3218536)	0.477	0.155	0.222	0.187	0.184	0.456	0.126	0.321	0.151
XRCC3 C>T (rs861539)	0.665	0.636	0.737	0.233	0.397	0.361	0.426	0.743	0.687
Lig4 T>C (rs1805386)	0.685	0.776	0.769	0.346	0.386	0.427	0.186	0.083	0.082
BRCA2 A>G (rs11571653)	0.866	0.716	0.804	0.671	0.835	0.775	0.991	0.571	0.711

Please cite this article in press as: Reeves SG, et al. DNA repair gene polymorphisms and risk of early onset colorectal cancer in hereditary nonpolyposis colorectal cancer. Cancer Epidemiology (2011), doi:10.1016/j.canep.2011.09.003

G Model
CANEP 355 1–7
ARTICLE IN PRESS
S.G. Reeves et al. / Cancer Epidemiology xxx (2011) xxx–xxx

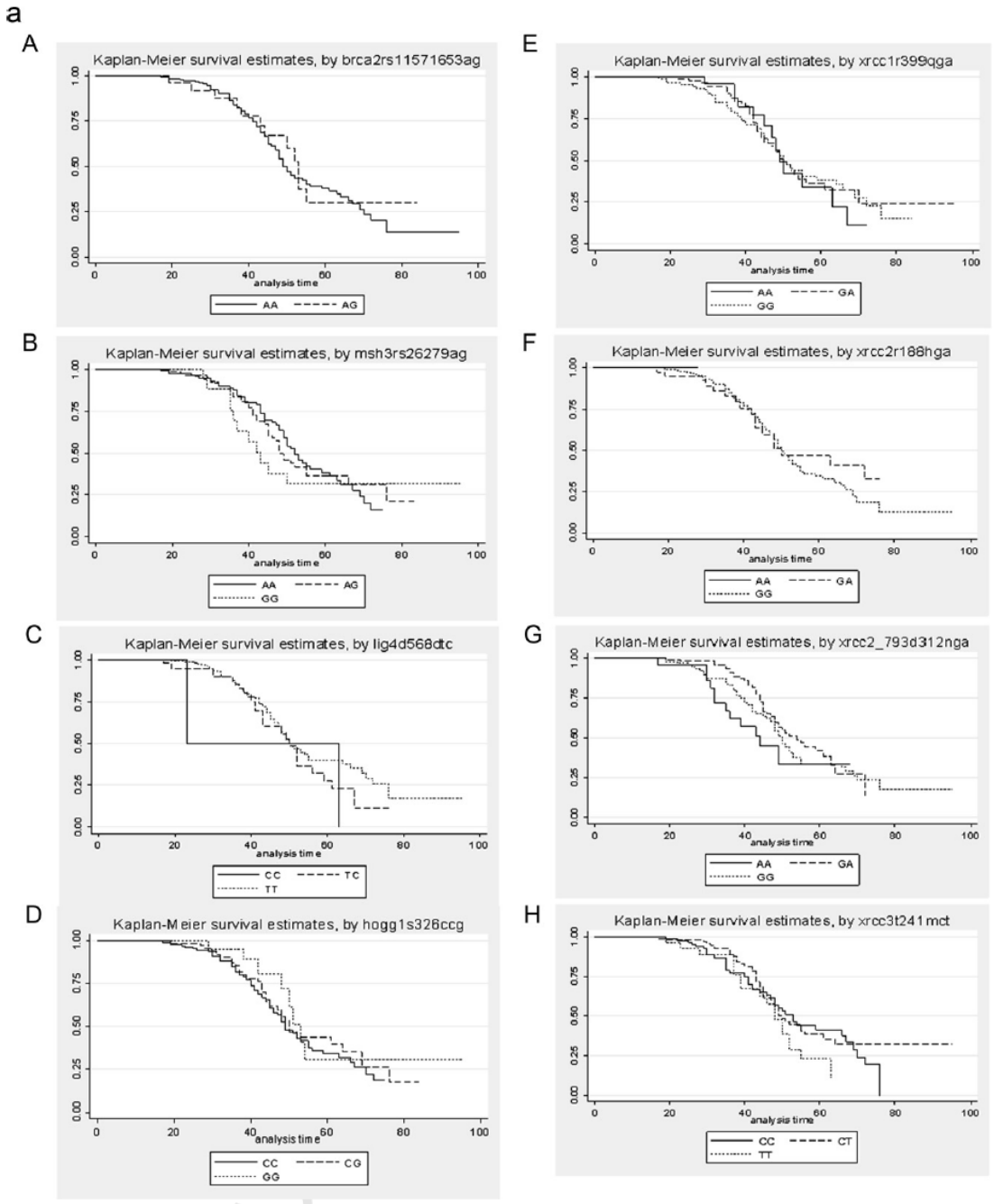

Fig. 2.

associated with population differences in which case the size of the Polish population studied is too small to allow for the identification and confirmation of any modifier effect. Since these two genes are important regulators of different aspects of DNA repair the failure to identify an association in the Australian population does not rule out their possible involvement in modifying disease expression in Lynch syndrome.

The failure to observe any association with SNPs associated with DNA repair does not negate the findings of others in terms of their effects on disease risk in the general population. The focus of this report is on a highly selected population who all carry deleterious germline changes in genes involved in DNA mismatch repair and as such this may outweigh the effects conferred by other more subtle alterations in DNA repair.

Please cite this article in press as: Reeves SG, et al. DNA repair gene polymorphisms and risk of early onset colorectal cancer in hereditary nonpolyposis colorectal cancer. Cancer Epidemiology (2011), doi:10.1016/j.canep.2011.09.003

G Model
CANEP 355 1–7
ARTICLE IN PRESS

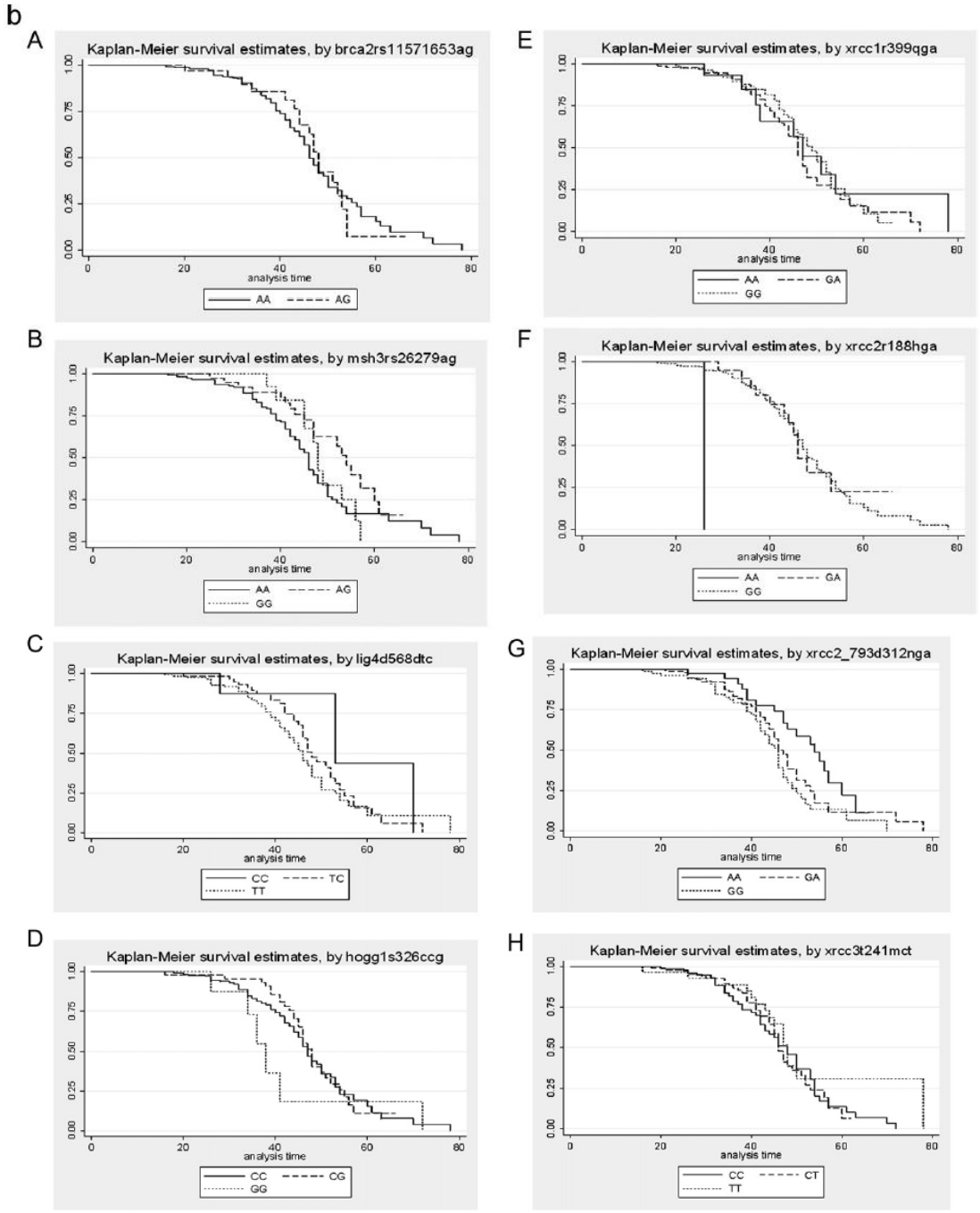

Fig. 2. (*Continued*).

In conclusion the results in this study indicate that common polymorphisms in DNA repair genes do not appear to be associated with the risk of developing colorectal cancer in a susceptible population. These results of this report do not rule out the possibility that the SNPs chosen for this study may be associated with disease progression and/or overall survival.

Finally, this study does not rule out the possibility of a small influence on disease risk in Lynch syndrome as it did not have the power to identify small effect sizes and as such larger studies are

Please cite this article in press as: Reeves SG, et al. DNA repair gene polymorphisms and risk of early onset colorectal cancer in hereditary nonpolyposis colorectal cancer. Cancer Epidemiology (2011), doi:10.1016/j.canep.2011.09.003

G Model
CANEP 355 1–7
ARTICLE IN PRESS

required to delineate the exact effects of functional polymorphisms in DNA repair genes.

Conflict of interest

No conflicts of interest.

Q4 Uncited reference

[26].

Q5 References

[1] de la Chapelle A. Genetic predisposition to colorectal cancer. Nat Rev Cancer 2004;4(10):769–80.
[2] Lynch HT, de la Chapelle A. Genetic susceptibility to non-polyposis colorectal cancer. J Med Genet 1999;36:p.801–18.
[3] Scott RJ, McPhillips M, Meldrum CJ, Fitzgerald PE, Adams K, Spigelman AD, et al. Hereditary nonpolyposis colorectal cancer in 95 families: differences and similarities between mutation-positive and mutation-negative kindreds. Am J Hum Genet 2001;68(1):118–27.
Q6 [4] Aaltonen LA, Salovaara R, Kristo P, et al. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. N Engl J Med 1998;338(21):1481–7.
[5] DeFrancisco J, Grady WM. Diagnosis and management of hereditary non-polyposis colon cancer. Gastrointest Endosc 2003;58(3):390–408.
[6] Alarcon F, Lasset C, Carayol J, Bonadona V, Perdry H, Desseigne F, et al. Estimating cancer risk in HNPCC by the GRL method. Eur J Hum Genet 2007;15(8):831–6.
[7] Wood RD, Mitchell M, Sgouros M, Lindahl T. Human DNA repair genes. Science 2001;291(0):1284–9.
[8] Kunkel TA, Erie DA. DNA mismatch repair. Annu Rev Biochem 2005;74:681–710.
[9] David SS, O'Shea VL, Kundu S. Base-excision repair of oxidative DNA damage. Nature 2007;447:941–50.
[10] Peltomaki P. Deficient DNA mismatch repair: a common etiologic factor for colon cancer. Hum Mol Genet 2001;10(7):735–40.
[11] Koessler T, Oestergaard MZ, Song H, Tyrer J, Perkins B, Dunning A, et al. Common variants in mismatch repair genes and risk of colorectal cancer. Gut 2008;1–9.
[12] An Y, Jin G, Wang H, Wang Y, Liu H, Li R, et al. Polymorphisms in hMLH1 and risk of early-onset lung cancer in a southeast Chinese population. Lung Cancer 2007;59(2):164–70.
[13] Hirata H, Hinoda Y, Kawamoto K, Kikuno N, Suehiro Y, Okayama N, et al. Mismatch repair gene MSH3 polymorphism is associated with the risk of sporadic prostate cancer. J Urol 2008;179(5):2020–4.
[14] Marchand L, Donlon T, Lum-Jones A, Seifried A, Wilkens LR. Association of the OGG1 Ser326Cys polymorphism with lung cancer risk. Cancer Epidemiol Biomarkers Prev 2002;11:409–12.
[15] Tsukino H, Hanaoka T, Otani T, Iwasaki M, Kobayashi M, Hara M, et al. OGG1 Ser326Cys polymorphism, interaction with environmental exposures, and gastric cancer risk in Japanese populations. Cancer Sci 2004;95(12):977–83.
[16] Arizono K, Osada Y, Kuroda Y. DNA repair gene OGG1 codon 326 and XRCC1 codon 399 polymorphisms and bladder cancer risk in a Japanese population. Jpn J Clin Oncol 2008;38(3):186–91.
[17] Park JY, Lee SY, Jeon HS, Bae NC, Sang CC, Joo S, et al. Polymorphism of the DNA repair gene XRCC1 and risk of primary lung cancer. Cancer Epidemiol Biomarkers Prev 2002;11:23–7.
[18] Jiao L, Hasson MM, Bondy ML, Wolff RA, Evans MD, Abbruzzese JL, et al. XRCC2 and XRCC3 gene polymorphism and risk of pancreatic cancer. Am J Gastroenterol 2008;103(2):360–7.
[19] Benhamou S, Tuimala J, Bouchardy C, Dayer P, Sarasin A, Hirvonen A. DNA repair gene XRCC2 and XRCC3 polymorphisms and susceptibility to cancers of the upper aerodigestive tract. Int J Cancer 2004;112(5):901–4.
[20] Abbott DW, Freeman ML, Holt JT. Double-strand break repair deficiency and radiation sensitivity in BRCA2 mutant cancer cells. J Natl Cancer Inst 1998;90(13):978–85.
[21] Douglas JA, Levin AM, Zuhlke KA, Ray AM, Johnson GR, Lange EM, et al. A common variation in the BRCA1 gene and prostate cancer risk. Cancer Epidemiol Biomarkers Prev 2007;16(7):1510–6.
[22] Sakiyama T, Kohno T, Mimaki S, Ohta T, Yanagitani N, Sobue T, et al. Association of amino acid substitution polymorphisms in DNA repair genes TP53, POLI, REV1 and LIG4 with lung cancer risk. Int J Cancer 2005;114(5):730–7.
[23] Kuschel B, Auranen A, McBride S, Novik KL, Antoniou A, Lipscombe JM, et al. Variants in DNA double-strand break repair genes and breast cancer susceptibility. Hum Mol Genet 2002;11:1399–407.
[24] Li D, Suzuki H, Liu B, Morris J, Liu J, et al. DNA repair gene polymorphisms and risk of pancreatic cancer. Clin Cancer Res 2009;15:740–6.
[25] Gorski B, Byrski T, Huzarski T, et al. Founder mutations in the BRCA1 gene in Polish families with breast-ovarian cancer. Am J Hum Genet 2000;66:1963–8.
[26] Broberg K, Bjork J, Paulsson K, Hoglund M, Albin M. Constitutional short telomeres are strong genetic susceptibility markers for bladder cancer. Carcinogenesis 2005;126(7):1263–71.

Please cite this article in press as: Reeves SG, et al. DNA repair gene polymorphisms and risk of early onset colorectal cancer in hereditary nonpolyposis colorectal cancer. Cancer Epidemiology (2011), doi:10.1016/j.canep.2011.09.003

Chapter 7

General Discussion

Factors associated with hereditary cancer syndromes such as Lynch syndrome continue to prove elusive in regards to the prediction of individual specific disease risk. Patients diagnosed with Lynch syndrome have a significantly increased life time risk of developing a malignancy especially CRC. Despite this well known likelihood of developing disease, accurately defining individual risk and at what age a malignancy is likely to develop is a much more challenging task that requires the assessment of multiple factors, both genetic and environmental.

Risk factors for CRC have traditionally included obesity, a diet low in fruits and vegetables, physical inactivity, and smoking [130, 131]. More recently however, the list of dietary factors has been severely challenged with the reported protection against CRC by fruits and vegetables to be at best modest [132]. Nevertheless CRC has traditionally been a disease characteristic of developed nations that include the United States, Europe and the countries of Oceania whose populations were more significantly influenced by inappropriate nutrition [133]. In recent years however, CRC rates have been reported to be increasing in more advanced developing countries where disease risk was once considered low [134, 135]. The increasing prevalence of obesity and decreasing physical activity in many parts of the world, resulting from the influence of industrialization will likely continue to contribute to the growing international CRC incidence rates as the trend in affluence increases [136]. This is likely to be compounded in countries where aging populations are the norm.

Disease risk is not only a function of environmental influence. How an individual responds biologically to an environmental trigger is influenced by their genome. This suggests that there will be individuals who harbour genetic variation that either promotes disease development or inhibits it given a particular environmental insult. At present there is an increasing body of evidence to suggest that disease risk is influenced by genetic polymorphisms that are linked to important processes such as cellular metabolism or DNA repair pathways. The examination of a model disease like Lynch syndrome is important because not only does it better define disease risk in this hereditary condition, it also serves as a mechanism to identify genetic variation that influences disease risk in the general population.

In the context of Lynch syndrome, genetic variation combined with lifestyle factors are likely to modulate the effect of predisposing mutations between individuals particularly in an impaired DNA repair environment. A significant outcome in the identification of modifying genes affecting disease expression in Lynch syndrome is that this will result in a more accurate assessment of disease risk in these individuals so that prophylactic measures could be implemented patient specifically thereby reducing the risks of screening for disease.

7.1 The IGF-1 Gene

The *IGF-1* gene was first reported as a potential modifying gene in Lynch syndrome expression in 2006. The CA repeat polymorphism located near the *IGF-1* promoter region was described as having an association with the age of disease onset in a cohort of 121 Lynch syndrome participants originating from the United States [125]. Certainly this is not the first time that a repeat region has been implicated in disease; with numerous studies reporting a link between DNA repeat regions significantly altering risk of prostate, [137-171] breast, squamous cell carcinoma, bladder and lung cancers [172-175]. DNA microsatellite repeat regions are also strongly associated with Lynch Syndrome and are considered a genetic tumour phenotype [35].

IGF-1 is important for cellular proliferation and differentiation, however, elevated levels of IGF-1 have been reported to have significant links to disease such as CRC which is thought to be a result of the mitogenic and anti-apoptotic effects elicited by this protein [125, 176]. Several environmental and physiological reasons have been proposed that influence IGF-1 expression; however it has been only recently that evidence has accumulated suggesting a genetic role. Rosen *et al.* was the first to report that the length of the CA repeat region in *IGF-1* may be associated with circulating IGF-1 levels [177]. In a similar growth factor related gene, Epidermal Growth Factor Receptor (EGFR), a CA repeat region is located in intron 1. A study of this *EGFR* polymorphic repeat region revealed lower transcriptional activity with increasing numbers of polymorphic CA repeats coinciding with lower levels of gene expression [178]. In 2007, a similar result was reported for the *IGF-1* gene in swine where the length of the CA repeat region was clearly associated with circulating levels of IGF-1 [179]. More recently additional human

data has been published which supports the notion that this polymorphism is linked to serum levels of IGF-1 [180]. From this data a trend is emerging that CA repeat polymorphisms in growth factor related genes, such as *IGF-1*, are related to gene expression differences which is reflected in the serum levels of these important genes. If, as is the case for *EGFR*, longer CA repeats in *IGF-1* are predicted to result in a decrease in circulating serum levels which in turn potentially lead to a reduced likelihood of disease in Lynch syndrome patients. Accumulating evidence suggests that serum IGF-1 levels appear to be linked to disease with recent reports indicating that elevated levels of IGF-1 are observed in breast, prostate and CRC [181-184]. There have even been some estimates that higher circulating levels of IGF-1 result in a 15% increase in the risk of developing disease, insinuating the importance of circulating IGF-1 in disease progression [185].

As CRC involves the accumulation of a number of specific molecular alterations [186, 187], consistently high IGF-1 serum levels may increase cellular proliferation, thereby enhancing the rate by which genetic alterations accumulate. Both normal colonic epithelial and transformed cells are IGF-1 responsive; thus, IGF-1 can influence not only the likelihood of disease initiation but also disease progression. This overall process provides some insight into how intracellular serum levels of IGF-1 may have a significant influence in accelerating the accumulation of genetic errors leading to disease, especially in persons who have inherited a predisposition to develop malignancy characterized by a mutator phenotype as observed in Lynch syndrome.

An equally important facet to disease risk as a result of increased levels of IGF-1 is its link with obesity. Obesity and physical inactivity are strong independent determinants of insulin resistance and hyperinsulinaemia [188-193] and is associated with an increased risk of CRC [194, 195]. Increased blood insulin lowers IGF-1 binding protein levels, which often result in an increase of free IGF-1 [196]. As IGF-1 is associated with both percentage body fat and general overall obesity levels [197], an increased level of IGF-1 expression as a result of shorter CA repeat lengths may have an enhanced effect in persons who are obese where IGF-1 serum levels are already elevated.

In addition to the IGF-1 effect, CRC risk is also increased in obese patients through oxidative stress in fat. This is caused by increased lipid peroxidation leading to the production of reactive oxygen species. In regards to cancer, reactive oxygen species can damage DNA by several methods including DNA base modification, deletions, frame shifts, strand breaks, DNA-protein cross-links, and chromosomal rearrangements [198]. Both these components may promote tumour development by generating reactive oxygen species, increasing hormone production/bioavailability of IGF-1 and providing an energy-rich environment. This combined mechanism is potentially a risk factor for all types of CRC, however in Lynch syndrome this may be of greater significance in a deficient DNA repair environment where enhanced levels of IGF-1 inhibit cell death whilst encouraging cellular proliferation. Together, the relationship between obesity and *IGF-1* CA repeat length may be of particular importance in obese Lynch syndrome cases as these may be at greatest risk of developing disease at a younger age.

The role of inherited factors in circulating IGF1 serum levels is likely to be substantial with estimates of the proportion of variance in IGF-1 that is genetically determined varying somewhere between 38% to over 80% [199]. Considerable data exists revealing differences in IGF levels across ethnic groups [200-202], however more recent data is suggestive of dietary and lifestyle factors having a more modifiable effect in serum levels when combined with genetic ancestry. One such study has shown that the impact of several nutritional factors such as calcium, dairy products and vegetables on IGF levels is quite different in race stratified models between African-American and European American males [203]. This is strongly suggestive of there being ethnic differences that differentially modify the effect of several nutrients on IGF levels. Together this information is suggestive that environmental factors such as dietary intake, lifestyle and demographic factors are probably playing a substantial role in ethnic variation in disease risk in regards to serum IGF levels. This is intriguing as it may also be contributing to the differences in relative disease risk observed between the Polish and Australian cohorts as described previously.

The data presented in chapters 2 and 3 combined with others [125] indicate a significant interaction between the CA repeat polymorphism length and disease expression in Lynch syndrome which is likely to be linked to circulating levels of IGF-1. The data indicated a significant correlation for earlier onset CRC in participants who carry 17 or less IGF-1 CA repeats in over 400 Lynch syndrome patients of multinational decent. An encouraging aspect of these described results is the fact that significance is retained across different population ethnicities where differences in IGF-1 allele frequencies occur. A limitation however in defining the exact relationship between IGF-1 expression

and cancer incidence in Lynch syndrome patients is the genotype–phenotype correlation between the *IGF-1* CA-repeat number and the corresponding serum levels in MMR mutation carriers. Assessment of serum IGF-1 concentration however, have for the most part had the inherent problem of serum IGF-I measurements being done at one time point whilst, what is required for accurate assessment is multiple tests performed on any given patient. Whether it would be feasible to monitor IGF-1 serum levels in families with Lynch syndrome is an area which would need to be further investigated. If more evidence is found to support the role of IGF-1 in disease risk monitoring may become a worthwhile practice, and it is reasonable to suggest that investigations into the *IGF-1* CA repeat polymorphism in other hereditary CRC cancer syndromes as well as sporadic disease may well be warranted. Future work should also include additional candidate polymorphisms located within *IGF-1* or *IGFBP-3* that interact with the IGF-1 pathway and may provide further insight into the overall IGF-1 effect. At present, however the IGF-1 pathway remains largely under-investigated, and there is now a requirement for further work to develop a more thorough understanding of the relationship between *IGF-1* genotype, expression and its implication in disease risk.

7.2 - The Methylenetetrahydrofolate Reductase (MTHFR) Gene

There have been tantalizing reports in the literature that polymorphisms in the *MTHFR* gene are associated with altered CRC risk. These polymorphisms occur in relatively high frequency in the general population and the two that promote special attention are both associated with altered enzyme function. MTHFR is a key folate-metabolizing enzyme involved in both DNA methylation and DNA synthesis. The enzyme catalyses the irreversible conversion of 5,10-methylenetetrahydrofolate (5,10-MTHF), needed for purine and thymidine synthesis, to 5-methyltetrahydrofolate (5-MTHF), which is necessary for methionine production. Insufficient thymidine results in uracil misincorporation into DNA, leading to single-strand and double-strand breaks. This can increase the incidence of DNA damage, thereby increasing the risk of genetic instability. The understanding that folate metabolism can both equally influence DNA synthesis and methylation has made the study of environmental and genetic variants associated with this metabolic pathway particularly attractive candidates as factors that influence cancer susceptibility. Two common polymorphisms, *C677T* and *A1298C* are located within the *MTHFR* gene and have been linked in altering the function of the encoded protein. This has lead to these variants becoming the focus of numerous studies into CRC risk outside the context of an inherited predisposition to disease. Both polymorphisms result in a substitution of an amino acid and have been previously shown to significantly influence MTHFR enzyme activity [204]. *C677T* is located within the coding region for the catalytic domain, resulting in an amino acid substitution from alanine to valine that is associated with a reduction of enzyme activity. The *A1298C* polymorphism, located in the regulatory region of MTHFR, substitutes an amino acid change from glutamine to alanine. Evidence suggests that *A1298C* also reduces MTHFR activity, however it is

reported to be less influential than *C677T* [204]. This modifying effect incurred by the presence of one or both polymorphisms in a pivotal folate metabolism pathway and its association with sporadic disease suggests that these polymorphisms are of particular interest with respect to modifying disease risk in Lynch syndrome.

Both *A1298C* and *C677T* are in strong linkage disequilibrium with no evidence of the existence of a *MTHFR* allele that carries both homozygote genotypes (*C1298C/T677T*) [205-207]. Owing to this linkage disequilibrium, no participants in this study were found to inherit both homozygote forms of these polymorphisms when screened throughout the study cohort.

The data indicated that in this cohort of Australian and Polish Lynch syndrome participants both heterozygote forms of the *MTHFR* variants were required for a significant protective effect to occur. The Kaplan-Meier survival estimates predicted a median age gap of 10 years later for CRC onset in patients carrying the combined heterozygote *MTHFR* genotype which was supported by multi variable regression modelling statistics. The data also suggested this effect was significant in both *hMLH1* and *hMSH2* carriers, where previously only a significant association had been described in *hMLH1* for *C677T* only [129]. The most plausible cause for this discrepancy between our reported results and those by Pande *et al* is likely due to a type 1 error as the reported association in *hMLH1* carriers were in a considerable smaller sample size, although differences in the ethnicity of Lynch syndrome cohorts cannot be ruled out as a contributing factor.

The mechanism by which *C677T* and *A1298C* appears to influence disease risk can be explained by the functional effects that these polymorphisms have on MTHFR and consequently folate metabolism. Previous reports have demonstrated a reduction of up to 60% in the activity of MTHFR when both *C677T* and *A1298C* heterozygote alleles were present in the gene. The reduction of *MTHFR* activity leads to an increased concentration of its substrate 5,10-MTHF. The increased pool of 5,10-MTHF pushes folate metabolism towards DNA synthesis, in turn reducing the pool of uracil. A reduced quantity of uracil potentially reduces the overall risk of uracil misincorporation as a result of its limited availability. For individuals with a MMR deficiency, the effect of reduced MTHFR enzyme activity may be advantageous since uracil misincorporation could be particularly deleterious in conjunction with an impaired DNA repair pathway. The subsequent lower levels of 5-MTHF may also be beneficial due to a potential reduction in DNA methylation. Hypermethylation of the promoter of tumour suppressor or MMR genes may lead to gene silencing, therefore a reduction in methylation through decreased MTHFR activity could lead to lower probability of this type of gene silencing occurring [205].

Numerous case control and cohort studies have investigated the relationship between folate intake and CRC risk with the majority reporting a reduction in CRC incidence with higher levels of folate [208]. The outcome of one meta-analysis suggested that CRC risk could be reduced by up to 25% with a high level of dietary folate compared to a low level one [209]. Further studies are required to clarify to what extent total folate has on disease risk; however it is generally accepted that there is an association and that a number of common genetic variants alter either the cellular levels or functioning of folate

metabolism enzymes and are likely to have an important role in determining an individual's response to changes in dietary folate. With this in mind further studies into functional polymorphisms in the folate metabolism pathway would benefit significantly by including total folate levels so that a more exact assessment of the role of folate could be made. Using this approach a more precise view of the relationship between folate intake and disease risk may become apparent where Lynch syndrome patients may be able to be stratified by *MTHFR* genotype whilst including important disease-folate associations in the subgroup of the cohort as defined by these genetic variants. Accurately estimating dietary folate intake however may prove difficult and therefore the analysis of plasma folate levels may be a more viable alternative. This may provide a more accurate measure of folate status, however be more susceptible to short-term fluctuations in dietary folate intake which could lead in the requirement of several measurements over time for an accurate baseline to be calculated. Future studies would also benefit to include other dietary factors including alcohol, choline, and methionine intake which are known to effect folate metabolism besides folate and folic acid [210]. An accurate level of folate plasma combined with *MTHFR C677T* and *A1298C* genotypes is an interesting prospect and may provide a more accurate indicator of an individual's risk of developing a Lynch syndrome related CRC.

The identification of *MTHFR* polymorphisms being associated with divergence in disease risk in Lynch syndrome provides the basis for targeted intervention measures that could be used to influence the risk of disease development. Dietary supplementation of folate/folic acid in Lynch syndrome families may prove to be beneficial in decreasing disease risk or prolonging the time before the diagnosis of malignancy. Dietary

supplementation and a change in disease risk however, are more complex than previously thought. Folic acid supplementation has been proven to be beneficial in decreasing neural tube defects (NTD's), [211] and was the catalyst for the United States and Canada introducing the compulsory supplementation of folic acid in flour in 1996 in an aim to reduce the incidence of NTD's. Despite proving successful for this purpose an unexpected trend was observed in both countries as described by Mason *et al.* who investigated the relationship between the onset of folic acid fortification and rises in the incidence of CRC. This analysis indicated that in the early part of the 1990's the age-adjusted incidence of CRC had declined gradually in both countries. Between 1995 and 1996 however, the incidence rate in the United States showed a slight increase followed by more marked increases in 1997 and 1998. A similar finding was observed in the Canadian population which also corresponded to the mandatory supplementation of folic acid. In both populations the increase in CRC incidence was highly significant when compared to pre-existing trends in both men and women. These observations have lead to a hypothesis that mandatory folic acid supplementation was responsible for the spike in CRC rates which after peaking approximately 2-3 years after its introduction have begun to decline once again [212].

The association of increased CRC incidence with folate supplementation has been supported by the results of two large scale studies which have recently emerged from both the United States and United Kingdom. In both these phase III studies a common trend was observed in participants who supplemented their diets for three years with a daily dose of 1000ug and 500ug folic acid respectively, and an increased risk of developing a colorectal adenoma, with the greater risk in those participants consuming

the higher 1000ug dose [213, 214]. Studies in mismatch repair or tumour suppressor gene deficient mice have demonstrated that the timing of folate supplementation is important in the association it may have on disease risk. In the first few months of folate supplementation a threefold decrease in colorectal adenomas has been observed when compared to mice with a moderately folate deficient diet. Dietary folate treatment after the development of carcinomas had the opposite effect however, with folate deficiency significantly decreasing the number of adenomas compared with supplementation [215, 216]. Together, this evidence suggests that as long as an individual is healthy, folate supplementation is protective whereas if a tumour has been initiated folate restriction is more important. This dual modulator role of folate may be of even greater influence in an impaired DNA mismatch repair pathway as found in Lynch syndrome patients. In this case folate supplementation may be particularly beneficial or deleterious depending upon any early tumour development.

7.3 - Candidate Polymorphisms Not Associated With Disease Risk

Not all polymorphisms which have been associated with hereditary disease have remained consistently significant across cohorts. An example is the delta *DNMT3b* SNP which was reported to have a significant association in a cohort of participants in the United States [217]. *DNMT3B* has been identified as a candidate in disease modifying risk due to its role in methylation. DNA methylation is regulated by a family of DNA methyltransferases (DNMTs), of which three active forms (DNMT1, DNMT3A and DNMT3B) have been identified in mammalian cells [218]. It has been reported that an increase in DNA methyltransferase enzyme activity of the DNA methyltransferases DNMT1, and DNMT3A and DNMT3B, is elevated in several types of disease including

leukaemia, prostate, lung, breast and endometrial cancers [219-222]. A polymorphism located within *DNMT3b* has been reported to influence enzyme expression through altering promoter activity. It has been suggested that in *in vitro* assays the C>T variant could lead to an increase of promoter activity of up to 30% [218]. Using a study group of over 400 individuals, no association was observed between age of onset and *DNMT3b* genotype in an Australian and Polish Lynch syndrome cohort. The failure to confirm the potential modifying influence of a polymorphism in one population compared to another could be simply due to insufficient numbers of test subjects. If a polymorphism is a true modifier however its response should be similar no matter what population is examined even though it may not reach statistical significance. In the case of the delta *DNMT3b* SNP no such trend was observed suggesting type 1 or type 2 statistical errors may be playing a role. The Australian/Polish study group was approximately three times larger than the participants of a previous study [217] reducing the likelihood of a type 1 error in this cohort. Notwithstanding, it is worth noting that it does not rule out the possibility that *DNMT3b* expression may be associated with Lynch syndrome disease expression. Different isoforms of DNMT3b exist therefore expression levels of these may vary influencing disease risk. Numerous other polymorphisms have also been reported in the functional domains of DNMT3b which could also alter methylation status and thereby influence disease risk.

Genes involved in DNA repair have also been prime candidates in the search for modifying effects due to their important role in the cell cycle. Polymorphisms located within genes involved in this process have been widely reported to be associated with cancer susceptibility in an extensive range of malignancies that include CRC. For our

combined cohorts, eight common polymorphisms were selected across several genes involved in the DNA repair pathway including *BRCA2, hMSH3, Lig4, hOGG1* and *XRCC* 1, 2 and 3 which had not previously been assessed for disease risk in Lynch syndrome. The data from this work resulted in some interesting and somewhat surprising results in that two of the eight polymorphisms indicated a significant association in the Polish cohort however was not observed in the Australian counterparts. Cox regression modelling indicated a significant protective effect in Polish participants for both polymorphisms *hMSH3* A>G (rs26279) and *XRCC2* G>A (rs1799793). This finding was somewhat contradictory as the homozygote form of both rs26279 and rs1799793 have been previously weakly associated with an increased risk of CRC and bladder cancer respectively [223, 224]. In both cases rs26279 and rs1799793 had significantly different minor allele frequencies compared to the Australian cohort providing a possible explanation to the lack of significance in either the Australian or combined population analysis. Most notably however were the hazard ratios in the Australian cohort which pointed in the opposite direction to those in the Polish cohort. If a true modifying effect were present it is reasonable to assume that hazard ratios should at least point in the same direction, even if not being significant. The previously published results before this study is therefore suggestive of a type 1 error which may be caused by differences in allele frequencies between groups. For this reason it has been beneficial to have access to two separate Lynch syndrome cohorts where background ethnicity provides variation in allele frequencies. If a modifying effect is then observed it is more likely to be that of a true effect if remaining consistent in both populations.

7.4 – Overall Conclusions

In conclusion, the research conducted in this project has further discovered and defined several new interesting associations in disease expression in Lynch syndrome patients including:

i) The CA repeat region in *IGF-1* has now shown a significant effect in disease onset age in three separate cohorts of Lynch syndrome participants of multinational descent. The studies performed in this thesis have also further defined the specific number of 17 or less CA repeats to be the greatest risk of developing earlier onset disease and that population specific risk may occur.

ii) A significant result was further established for the target functional polymorphisms in the *MTHFR* gene. This was also particularly exciting as strong significance was retained across multinational study cohorts, despite some variation in allele frequencies. This combined with *IGF-1* is likely to prove to be of particular importance in the understanding of the development of CRC in Lynch syndrome and at the very least aid in providing a more accurate estimate in CRC onset age.

iii) The study into the polymorphism located within *DNMT3B* was found to have no effect on the age of disease onset in Lynch syndrome. This was in disagreement with a previous smaller study which hypothesized a mechanism by which this polymorphism may regulate the epigenetic status of a gene at a separate locus, however this could not be confirmed in a larger cohort of Lynch syndrome participants.

iv) An investigation into the target polymorphisms located within the DNA repair genes *BRCA2, hMSH3, Lig4, hOGG1* and *XRCC* did not reveal a substantiated

significant association in disease onset age in Lynch syndrome. The polymorphisms found in *MSH3* (rs26279) and *XRCC2* (rs1799793) in the Polish cohort did demonstrate a trend. However, the size of the population did not allow for the confirmation of any modifier effect with multiple testing rendering the trend insignificant.

7.5 – Future Directions

With increased knowledge into hereditary genetic diseases such as Lynch syndrome, more questions often arise as we gain insight into the modifying influences driving disease expression. Although the results published throughout this study have been of significant value more questions arise that will no doubt be the target of more defined research to come.

Future studies into the combination of the polymorphisms in *IGF-1* and *MTHFR* could benefit substantially by gaining more information into additional factors such as Body Mass Index and dietary intake of foods and nutrients of specific interest such as folate. This may then be able to be stratified by genotype and ethnicity providing potentially powerful information into these "modifiable" modifiers in overall disease risk. It is plausible to suggest that when a certain combination of both *IGF-1* and *MTHFR* polymorphisms are present, a more profound effect in disease risk may be observed which opens up an exciting new avenue of research. The collective influence on disease risk by a combination of these gene polymorphisms may be amplified in effect by a number of lifestyle and dietary factors. This is likely to be of a greater significance in westernised societies providing partial explanation into why CRC incidence is higher in

these regions, whilst ethnicity may also be modifying the effect on not only genetic disease risk, but also on how great an influence some of these factors may have.

The candidate gene approach for the identification of modifier genes has proven to be a useful approach in better defining the role of additional genetic factors that are associated with disease onset in inherited susceptibilities to cancer. The major drawback in using the candidate approach is the reliance on the identification of functional polymorphisms thereby limiting the choice of gene to study. Furthermore, additional changes occurring in the genome that may confer novel alterations in function and or point to new genes cannot be investigated using this approach. With the advent of genome wide association studies and the unbiased identification of new polymorphisms associated with disease risk it is inevitable that hitherto unexplored genes and polymorphisms will contribute further to our knowledge about modifier genes and their roles in dictating disease risk in inherited forms of malignancy.

7.6 - Summary

The data gathered in this project has been of significant value in the search and clarification of genetic modifying effects in Lynch syndrome. In addition to specific genotypes, these genes may also have a wider implication in disease risk in concert with other environmental factors that may not only be restricted to Lynch syndrome. Individuals with obesity are already at an increased risk of developing CRC by an imbalance of hormones, reactive oxygen species, and energy availability which promotes cell transformation, angiogenesis, migration, and proliferation, as well as inhibition of apoptosis. These patients could be of an even greater risk in cases where the *IGF-1* CA polymorphism was present resulting in a growth advantage in any early stage colonic tumours. In a similar instance in cases where an individual has wild type forms of the polymorphisms located within *MTHFR* and a high folate diet, this may lead to a selective growth advantage for neoplastic growth.

The results described for *DNMT3b* and DNA repair genes (*BRCA2, hMSH3*, *Lig4*, *hOGG1* and *XRCC*) despite being insignificant in disease association, are still useful in our overall knowledge in Lynch syndrome disease expression. These results will prove beneficial for future studies as it may allow for researchers to focus on other genome areas of interest. When searching for modifying effects it is important to report those polymorphisms where no significant association is found. The majority of literature reports are of significant Lynch syndrome modifying associations; however there is a requirement to publish non-significant findings so that some potential modifying polymorphisms may be ruled out from future studies.

As with some of the cases previously illustrated however, care still needs to be taken in ruling out SNPs where no modifying effect is observed in small numbers of participants where type 2 errors may occur. Given the numbers and the diverse population groups included in this study however, there is an increased likelihood that these reported results give a more accurate representation.

The biological consequence of the variants reported herein on disease risk has the potential for a more personalized approach to medicine dependent on the genetic constitution of each individual. This will allow for a more accurate risk assessment of predicted age of disease onset in families diagnosed with Lynch syndrome along with measures that can be incorporated to help prevent the onset of disease.

Chapter 8

Bibliography

1. Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**(5258): p. 561-3.
2. Fearon, E.R. and B. Vogelstein, *A genetic model for colorectal tumorigenesis.* Cell, 1990. **61**(5): p. 759-67.
3. Doetsch, P.W., *What's old is new: an alternative DNA excision repair pathway.* Trends Biochem Sci, 1995. **20**(10): p. 384-6.
4. Friedberg, E.C., *Out of the shadows and into the light: the emergence of DNA repair.* Trends Biochem Sci, 1995. **20**(10): p. 381.
5. Jackson, S.P. and P.A. Jeggo, *DNA double-strand break repair and V(D)J recombination: involvement of DNA-PK.* Trends Biochem Sci, 1995. **20**(10): p. 412-5.
6. Kolodner, R.D., *Mismatch repair: mechanisms and relationship to cancer susceptibility.* Trends Biochem Sci, 1995. **20**(10): p. 397-401.
7. Lehmann, A.R., *Nucleotide excision repair and the link with transcription.* Trends Biochem Sci, 1995. **20**(10): p. 402-5.
8. Seeberg, E., L. Eide, and M. Bjoras, *The base excision repair pathway.* Trends Biochem Sci, 1995. **20**(10): p. 391-7.
9. Shinohara, A. and T. Ogawa, *Homologous recombination and the roles of double-strand breaks.* Trends Biochem Sci, 1995. **20**(10): p. 387-91.
10. Rhyu, M.S., *Molecular mechanisms underlying hereditary nonpolyposis colorectal carcinoma.* J Natl Cancer Inst, 1996. **88**(5): p. 240-51.
11. Chung, D.C. and A.K. Rustgi, *DNA mismatch repair and cancer.* Gastroenterology, 1995. **109**(5): p. 1685-99.
12. Modrich, G.M.L.a.P., *Restoration of mismatch genes to nuclear extracts of H6 colorectal tumor cells by a heterodimer of MUTL homologues.* Proc Natl Acad Sci 1995. **92**: p. 1950–1954.
13. Lynch, H.T. and T. Smyrk, *Hereditary nonpolyposis colorectal cancer (Lynch syndrome). An updated review.* Cancer, 1996. **78**(6): p. 1149-67.
14. Marti, T.M., C. Kunz, and O. Fleck, *DNA mismatch repair and mutation avoidance pathways.* J Cell Physiol, 2002. **191**(1): p. 28-41.
15. Weinberg, R.A., *Oncogenes and tumor suppressor genes.* CA Cancer J Clin, 1994. **44**(3): p. 160-70.
16. Sherr, C.J., *Cancer cell cycles.* Science, 1996. **274**(5293): p. 1672-7.
17. Pietsch, E.C., et al., *The p53 family and programmed cell death.* Oncogene, 2008. **27**(50): p. 6507-21.
18. Murphree, A.L. and W.F. Benedict, *Retinoblastoma: clues to human oncogenesis.* Science, 1984. **223**(4640): p. 1028-33.
19. Knudson, A.G., Jr., *Hereditary cancer, oncogenes, and antioncogenes.* Cancer Res, 1985. **45**(4): p. 1437-43.
20. Knudson, A.G., Jr., *Mutation and cancer: statistical study of retinoblastoma.* Proc Natl Acad Sci U S A, 1971. **68**(4): p. 820-3.
21. F.M, R., *Molecular pathology of von Hippel–Lindau disease and the VHL tumour suppressor gene.* Expert reviews in molecular medicine, 2001.
22. Stratton, M.R., P.J. Campbell, and P.A. Futreal, *The cancer genome.* Nature, 2009. **458**(7239): p. 719-24.
23. Miller, D.G., *On the nature of susceptibility to cancer. The presidential address.* Cancer, 1980. **46**(6): p. 1307-18.

24. Schinzel, A.C. and W.C. Hahn, *Oncogenic transformation and experimental models of human cancer.* Front Biosci, 2008. **13**: p. 71-84.
25. Beerenwinkel, N., et al., *Genetic progression and the waiting time to cancer.* PLoS Comput Biol, 2007. **3**(11): p. e225.
26. Toribara, N.W. and M.H. Sleisenger, *Screening for colorectal cancer.* N Engl J Med, 1995. **332**(13): p. 861-7.
27. Organisation, T.W.H., *Colorectal cancer fact sheet.* 2008. **297**.
28. Lynch, H.T. and A. de la Chapelle, *Genetic susceptibility to non-polyposis colorectal cancer.* J Med Genet, 1999. **36**(11): p. 801-18.
29. Grady, W.M. and J.M. Carethers, *Genomic and epigenetic instability in colorectal cancer pathogenesis.* Gastroenterology, 2008. **135**(4): p. 1079-99.
30. Su, L.K., et al., *Genomic rearrangements of the APC tumor-suppressor gene in familial adenomatous polyposis.* Hum Genet, 2000. **106**(1): p. 101-7.
31. Levitt, N.C. and I.D. Hickson, *Caretaker tumour suppressor genes that defend genome integrity.* Trends Mol Med, 2002. **8**(4): p. 179-86.
32. Muleris, M., et al., *Chromosomal instability in near-diploid colorectal cancer: a link between numbers and structure.* PLoS One, 2008. **3**(2): p. e1632.
33. Fodde, R., et al., *Mutations in the APC tumour suppressor gene cause chromosomal instability.* Nat Cell Biol, 2001. **3**(4): p. 433-8.
34. Walther, A., R. Houlston, and I. Tomlinson, *Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis.* Gut, 2008. **57**(7): p. 941-50.
35. Peltomaki, P., *Deficient DNA mismatch repair: a common etiologic factor for colon cancer.* Hum Mol Genet, 2001. **10**(7): p. 735-40.
36. DeFrancisco, J. and W.M. Grady, *Diagnosis and management of hereditary non-polyposis colon cancer.* Gastrointest Endosc, 2003. **58**(3): p. 390-408.
37. *Classics in oncology. Heredity with reference to carcinoma as shown by the study of the cases examined in the pathological laboratory of the University of Michigan, 1895-1913. By Aldred Scott Warthin. 1913.* CA Cancer J Clin, 1985. **35**(6): p. 348-59.
38. Lynch, H.T., et al., *Hereditary factors in cancer. Study of two large midwestern kindreds.* Arch Intern Med, 1966. **117**(2): p. 206-12.
39. Lynch, H.T. and A.J. Krush, *Cancer family "G" revisited: 1895-1970.* Cancer, 1971. **27**(6): p. 1505-11.
40. Lynch, H.T., et al., *The cancer family syndrome: a status report.* Dis Colon Rectum, 1981. **24**(4): p. 311-22.
41. Lynch, H.T., et al., *Hereditary nonpolyposis colorectal cancer (Lynch syndromes I and II). II. Biomarker studies.* Cancer, 1985. **56**(4): p. 939-51.
42. Lynch, H.T., et al., *Hereditary nonpolyposis colorectal cancer (Lynch syndromes I & II). Genetics, pathology, natural history, and cancer control, Part I.* Cancer Genet Cytogenet, 1991. **53**(2): p. 143-60.
43. Vasen, H.F., et al., *The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC).* Dis Colon Rectum, 1991. **34**(5): p. 424-5.
44. Peltomaki, P., et al., *Microsatellite instability is associated with tumors that characterize the hereditary non-polyposis colorectal carcinoma syndrome.* Cancer Res, 1993. **53**(24): p. 5853-5.

45. Lindblom, A., et al., *Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer.* Nat Genet, 1993. **5**(3): p. 279-82.
46. Aaltonen, L.A., et al., *Clues to the pathogenesis of familial colorectal cancer.* Science, 1993. **260**(5109): p. 812-6.
47. Strand, M., et al., *Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair.* Nature, 1993. **365**(6443): p. 274-6.
48. Fishel, R., et al., *The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer.* Cell, 1993. **75**(5): p. 1027-38.
49. Leach, F.S., et al., *Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer.* Cell, 1993. **75**(6): p. 1215-25.
50. Bronner, C.E., et al., *Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer.* Nature, 1994. **368**(6468): p. 258-61.
51. Papadopoulos, N., et al., *Mutation of a mutL homolog in hereditary colon cancer.* Science, 1994. **263**(5153): p. 1625-9.
52. Nicolaides, N.C., et al., *Mutations of two PMS homologues in hereditary nonpolyposis colon cancer.* Nature, 1994. **371**(6492): p. 75-80.
53. Miyaki, M., et al., *Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer.* Nat Genet, 1997. **17**(3): p. 271-2.
54. Vasen, H.F., et al., *New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC.* Gastroenterology, 1999. **116**(6): p. 1453-6.
55. Jass, J.R., *Hereditary Non-Polyposis Colorectal Cancer: the rise and fall of a confusing term.* World J Gastroenterol, 2006. **12**(31): p. 4943-50.
56. Rodriguez-Bigas, M.A., et al., *A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines.* J Natl Cancer Inst, 1997. **89**(23): p. 1758-62.
57. Umar, A., et al., *Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability.* J Natl Cancer Inst, 2004. **96**(4): p. 261-8.
58. Urso, E., et al., *Proximal colon cancer in patients aged 51-60 years of age should be tested for microsatellites instability. A comment on the Revised Bethesda Guidelines.* Int J Colorectal Dis, 2008. **23**(8): p. 801-6.
59. Scott, R.J., et al., *Hereditary nonpolyposis colorectal cancer in 95 families: differences and similarities between mutation-positive and mutation-negative kindreds.* Am J Hum Genet, 2001. **68**(1): p. 118-127.
60. Peltomaki, P. and H.F. Vasen, *Mutations predisposing to hereditary nonpolyposis colorectal cancer: database and results of a collaborative study. The International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer.* Gastroenterology, 1997. **113**(4): p. 1146-58.
61. Lynch, H.T. and A. de la Chapelle, *Hereditary colorectal cancer.* N Engl J Med, 2003. **348**(10): p. 919-32.
62. Katballe, N., et al., *Frequency of hereditary non-polyposis colorectal cancer in Danish colorectal cancer patients.* Gut, 2002. **50**(1): p. 43-51.
63. Dunlop, M.G., et al., *Cancer risk associated with germline DNA mismatch repair gene mutations.* Hum Mol Genet, 1997. **6**(1): p. 105-10.

64. Hampel, H., et al., *Cancer risk in hereditary nonpolyposis colorectal cancer syndrome: later age of onset.* Gastroenterology, 2005. **129**(2): p. 415-21.
65. Aarnio, M., et al., *Cancer risk in mutation carriers of DNA-mismatch-repair genes.* Int J Cancer, 1999. **81**(2): p. 214-8.
66. Samowitz, W.S., et al., *The colon cancer burden of genetically defined hereditary nonpolyposis colon cancer.* Gastroenterology, 2001. **121**(4): p. 830-8.
67. Alarcon, F., et al., *Estimating cancer risk in HNPCC by the GRL method.* Eur J Hum Genet, 2007. **15**(8): p. 831-6.
68. Oliveira Ferreira, F., et al., *Frequency of extra-colonic tumors in hereditary nonpolyposis colorectal cancer (HNPCC) and familial colorectal cancer (FCC) Brazilian families: An analysis by a Brazilian Hereditary Colorectal Cancer Institutional Registry.* Fam Cancer, 2004. **3**(1): p. 41-7.
69. de Leeuw, W.J., et al., *Correspondence re: A. Muller et al., Exclusion of breast cancer as an integral tumor of hereditary nonpolyposis colorectal cancer. Cancer Res., 62: 1014-1019, 2002.* Cancer Res, 2003. **63**(5): p. 1148-9.
70. Vasen, H.F., et al., *MSH2 mutation carriers are at higher risk of cancer than MLH1 mutation carriers: a study of hereditary nonpolyposis colorectal cancer families.* J Clin Oncol, 2001. **19**(20): p. 4074-80.
71. Bandipalliam, P., et al., *Clinical presentation correlates with the type of mismatch repair gene involved in hereditary nonpolyposis colon cancer.* Gastroenterology, 2004. **126**(3): p. 936-7.
72. Hendriks, Y.M., et al., *Cancer risk in hereditary nonpolyposis colorectal cancer due to MSH6 mutations: impact on counseling and surveillance.* Gastroenterology, 2004. **127**(1): p. 17-25.
73. Plaschke, J., et al., *Lower incidence of colorectal cancer and later age of disease onset in 27 families with pathogenic MSH6 germline mutations compared with families with MLH1 or MSH2 mutations: the German Hereditary Nonpolyposis Colorectal Cancer Consortium.* J Clin Oncol, 2004. **22**(22): p. 4486-94.
74. Kolodner, R.D., et al., *Germ-line msh6 mutations in colorectal cancer families.* Cancer Res, 1999. **59**(20): p. 5068-74.
75. Hayward, B.E., et al., *PMS2 mutations in HNPCC.* Clin Genet, 2004. **66**(6): p. 566-7; author reply 568.
76. De Vos, M., et al., *Novel PMS2 pseudogenes can conceal recessive mutations causing a distinctive childhood cancer syndrome.* Am J Hum Genet, 2004. **74**(5): p. 954-64.
77. Thompson, E., et al., *Hereditary non-polyposis colorectal cancer and the role of hPMS2 and hEXO1 mutations.* Clin Genet, 2004. **65**(3): p. 215-25.
78. Niessen, R.C., et al., *PMS2 involvement in patients suspected of Lynch syndrome.* Genes Chromosomes Cancer, 2009. **48**(4): p. 322-9.
79. Vasen, H.F., et al., *Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis.* Gastroenterology, 1996. **110**(4): p. 1020-7.
80. Mary, J.L., et al., *Mutational analysis of the hMSH2 gene reveals a three base pair deletion in a family predisposed to colorectal cancer development.* Hum Mol Genet, 1994. **3**(11): p. 2067-9.

81. Moisio, A.L., et al., *Genetic polymorphisms in carcinogen metabolism and their association to hereditary nonpolyposis colon cancer.* Gastroenterology, 1998. **115**(6): p. 1387-94.
82. Brown, S.R., P.J. Finan, and D.T. Bishop, *Are relatives of patients with multiple HNPCC spectrum tumours at increased risk of cancer?* Gut, 1998. **43**(5): p. 664-8.
83. Aarnio, M., et al., *Life-time risk of different cancers in hereditary non-polyposis colorectal cancer (HNPCC) syndrome.* Int J Cancer, 1995. **64**(6): p. 430-3.
84. MacLennan, R., *Diet and colorectal cancer.* Int J Cancer, 1997. **Suppl 10**: p. 10-2.
85. Heavey, P.M., D. McKenna, and I.R. Rowland, *Colorectal cancer and the relationship between genes and the environment.* Nutr Cancer, 2004. **48**(2): p. 124-41.
86. Howe, G.R., et al., *The relationship between dietary fat intake and risk of colorectal cancer: evidence from the combined analysis of 13 case-control studies.* Cancer Causes Control, 1997. **8**(2): p. 215-28.
87. Boutron-Ruault, M.C., et al., *Energy intake, body mass index, physical activity, and the colorectal adenoma-carcinoma sequence.* Nutr Cancer, 2001. **39**(1): p. 50-7.
88. Norat, T. and E. Riboli, *Meat consumption and colorectal cancer: a review of epidemiologic evidence.* Nutr Rev, 2001. **59**(2): p. 37-47.
89. Trock, B., E. Lanza, and P. Greenwald, *Dietary fiber, vegetables, and colon cancer: critical review and meta-analyses of the epidemiologic evidence.* J Natl Cancer Inst, 1990. **82**(8): p. 650-61.
90. Terry, P., et al., *Prospective study of major dietary patterns and colorectal cancer risk in women.* Am J Epidemiol, 2001. **154**(12): p. 1143-9.
91. Kenneth A. Rasinskia, P.S.V., Maria Zagatskyb and Edith M. Rickettb, *Using implicit goal priming to improve the quality of self-report data* Journal of Experimental Social Psychology, 2004. **41**(3): p. 321-327.
92. Fan, R., et al., *The p53 codon 72 polymorphism and lung cancer risk.* Cancer Epidemiol Biomarkers Prev, 2000. **9**(10): p. 1037-42.
93. Birgander, R., et al., *p53 polymorphisms and haplotypes in nasopharyngeal cancer.* Hum Hered, 1996. **46**(1): p. 49-54.
94. Birgander, R., et al., *P53 polymorphisms and haplotypes in lung cancer.* Carcinogenesis, 1995. **16**(9): p. 2233-6.
95. Murata, M., et al., *Analysis of a germ line polymorphism of the p53 gene in lung cancer patients; discrete results with smoking history.* Carcinogenesis, 1996. **17**(2): p. 261-4.
96. Wang, Y.C., et al., *Prognostic significance of p53 codon 72 polymorphism in lung carcinomas.* Eur J Cancer, 1999. **35**(2): p. 226-30.
97. Zhang, Z.W., et al., *Age-associated increase of codon 72 Arginine p53 frequency in gastric cardia and non-cardia adenocarcinoma.* Clin Cancer Res, 2003. **9**(6): p. 2151-6.
98. Jones, J.S., et al., *p53 polymorphism and age of onset of hereditary nonpolyposis colorectal cancer in a Caucasian population.* Clin Cancer Res, 2004. **10**(17): p. 5845-9.

99. Sotamaa, K., et al., *p53 codon 72 and MDM2 SNP309 polymorphisms and age of colorectal cancer onset in Lynch syndrome.* Clin Cancer Res, 2005. **11**(19 Pt 1): p. 6840-4.
100. Talseth, B.A., et al., *Age of diagnosis of colorectal cancer in HNPCC patients is more complex than that predicted by R72P polymorphism in TP53.* Int J Cancer, 2006. **118**(10): p. 2479-84.
101. Khanna, K.K., et al., *ATM, a central controller of cellular responses to DNA damage.* Cell Death Differ, 2001. **8**(11): p. 1052-65.
102. Maillet, P., et al., *A polymorphism in the ATM gene modulates the penetrance of hereditary non-polyposis colorectal cancer.* Int J Cancer, 2000. **88**(6): p. 928-31.
103. Jones, J.S., et al., *ATM polymorphism and hereditary nonpolyposis colorectal cancer (HNPCC) age of onset (United States).* Cancer Causes Control, 2005. **16**(6): p. 749-53.
104. Heinimann, K., et al., *N-acetyltransferase 2 influences cancer prevalence in hMLH1/hMSH2 mutation carriers.* Cancer Res, 1999. **59**(13): p. 3038-40.
105. Frazier, M.L., et al., *Age-associated risk of cancer among individuals with N-acetyltransferase 2 (NAT2) mutations and mutations in DNA mismatch repair genes.* Cancer Res, 2001. **61**(4): p. 1269-71.
106. Talseth, B.A., et al., *Genetic polymorphisms in xenobiotic clearance genes and their influence on disease expression in hereditary nonpolyposis colorectal cancer patients.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(11): p. 2307-10.
107. Pistorius, S., et al., *N-Acetyltransferase (NAT) 2 acetylator status and age of tumour onset in patients with sporadic and familial, microsatellite stable (MSS) colorectal cancer.* Int J Colorectal Dis, 2007. **22**(2): p. 137-43.
108. Garcia-Closas, M., et al., *NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses.* Lancet, 2005. **366**(9486): p. 649-59.
109. Rebbeck, T.R., *Molecular epidemiology of the human glutathione S-transferase genotypes GSTM1 and GSTT1 in cancer susceptibility.* Cancer Epidemiol Biomarkers Prev, 1997. **6**(9): p. 733-43.
110. Kiran, M., Y.K. Chawla, and J. Kaur, *Glutathione-S-transferase and microsomal epoxide hydrolase polymorphism and viral-related hepatocellular carcinoma risk in India.* DNA Cell Biol, 2008. **27**(12): p. 687-94.
111. Shin, J.H., et al., *Glutathione S-transferase M1 associated with cancer occurrence in Korean HNPCC families carrying the hMLH1/hMSH2 mutation.* Oncol Rep, 2003. **10**(2): p. 483-6.
112. Felix, R., et al., *GSTM1 and GSTT1 polymorphisms as modifiers of age at diagnosis of hereditary nonpolyposis colorectal cancer (HNPCC) in a homogeneous cohort of individuals carrying a single predisposing mutation.* Mutat Res, 2006. **602**(1-2): p. 175-81.
113. Kong, S., et al., *Effects of cyclin D1 polymorphism on age of onset of hereditary nonpolyposis colorectal cancer.* Cancer Res, 2000. **60**(2): p. 249-52.
114. Bala, S. and P. Peltomaki, *CYCLIN D1 as a genetic modifier in hereditary nonpolyposis colorectal cancer.* Cancer Res, 2001. **61**(16): p. 6042-5.
115. Kruger, S., et al., *Absence of association between cyclin D1 (CCND1) G870A polymorphism and age of onset in hereditary nonpolyposis colorectal cancer.* Cancer Lett, 2006. **236**(2): p. 191-7.

116. Talseth, B.A., et al., *Aurora-A and Cyclin D1 polymorphisms and the age of onset of colorectal cancer in hereditary nonpolyposis colorectal cancer.* Int J Cancer, 2008. **122**(6): p. 1273-7.
117. Toniolo, P., et al., *Serum insulin-like growth factor-I and breast cancer.* Int J Cancer, 2000. **88**(5): p. 828-32.
118. Ma, J., et al., *Prospective study of colorectal cancer risk in men and plasma levels of insulin-like growth factor (IGF)-I and IGF-binding protein-3.* J Natl Cancer Inst, 1999. **91**(7): p. 620-5.
119. Yu, H., et al., *Plasma levels of insulin-like growth factor-I and lung cancer risk: a case-control analysis.* J Natl Cancer Inst, 1999. **91**(2): p. 151-6.
120. Chan, J.M., et al., *Plasma insulin-like growth factor-I and prostate cancer risk: a prospective study.* Science, 1998. **279**(5350): p. 563-6.
121. Cheng, I., et al., *Common genetic variation in IGF1 and prostate cancer risk in the Multiethnic Cohort.* J Natl Cancer Inst, 2006. **98**(2): p. 123-34.
122. Wong, H.L., et al., *Insulin-like growth factor-1 promoter polymorphisms and colorectal cancer: a functional genomics approach.* Gut, 2008. **57**(8): p. 1090-6.
123. Sarma, A.V., et al., *Genetic polymorphisms in CYP17, CYP3A4, CYP19A1, SRD5A2, IGF-1, and IGFBP-3 and prostate cancer risk in African-American men: the Flint Men's Health Study.* Prostate, 2008. **68**(3): p. 296-305.
124. Schildkraut, J.M., et al., *IGF1 (CA)19 repeat and IGFBP3 -202 A/C genotypes and the risk of prostate cancer in Black and White men.* Cancer Epidemiol Biomarkers Prev, 2005. **14**(2): p. 403-8.
125. Zecevic, M., et al., *IGF1 gene polymorphism and risk for hereditary nonpolyposis colorectal cancer.* J Natl Cancer Inst, 2006. **98**(2): p. 139-43.
126. Cao, H.X., et al., *Genetic polymorphisms of methylenetetrahydrofolate reductase and susceptibility to colorectal cancer.* Asian Pac J Cancer Prev, 2008. **9**(2): p. 203-8.
127. Sun, L., et al., *Methylenetetrahydrofolate reductase polymorphisms and susceptibility to gastric cancer in Chinese populations: a meta-analysis.* Eur J Cancer Prev, 2008. **17**(5): p. 446-52.
128. Kamel, A.M., et al., *Synergistic effect of methyltetrahydrofolate reductase (MTHFR) C677T and A1298C polymorphism as risk modifiers of pediatric acute lymphoblastic leukemia.* J Egypt Natl Canc Inst, 2007. **19**(2): p. 96-105.
129. Pande, M., et al., *Influence of methylenetetrahydrofolate reductase gene polymorphisms C677T and A1298C on age-associated risk for colorectal cancer in a caucasian lynch syndrome population.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(9): p. 1753-9.
130. Botteri, E., et al., *Smoking and colorectal cancer: a meta-analysis.* Jama, 2008. **300**(23): p. 2765-78.
131. Giovannucci, E., *Modifiable risk factors for colon cancer.* Gastroenterol Clin North Am, 2002. **31**(4): p. 925-43.
132. Boffetta, P., et al., *Fruit and vegetable intake and overall cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC).* J Natl Cancer Inst, 2010. **102**(8): p. 529-37.
133. Popkin, B.M., *The nutrition transition: an overview of world patterns of change.* Nutr Rev, 2004. **62**(7 Pt 2): p. S140-3.

134. Parkin, D.M., et al., *Global cancer statistics, 2002.* CA Cancer J Clin, 2005. **55**(2): p. 74-108.
135. Popkin, B.M., *The nutrition transition in low-income countries: an emerging crisis.* Nutr Rev, 1994. **52**(9): p. 285-98.
136. Center, M.M., et al., *Worldwide variations in colorectal cancer.* CA Cancer J Clin, 2009. **59**(6): p. 366-78.
137. Balic, I., et al., *Androgen receptor length polymorphism associated with prostate cancer risk in Hispanic men.* J Urol, 2002. **168**(5): p. 2245-8.
138. Beilin, J., et al., *A case-control study of the androgen receptor gene CAG repeat polymorphism in Australian prostate carcinoma subjects.* Cancer, 2001. **92**(4): p. 941-9.
139. Bennett, C.L., et al., *Racial variation in CAG repeat lengths within the androgen receptor gene among prostate cancer patients of lower socioeconomic status.* J Clin Oncol, 2002. **20**(17): p. 3599-604.
140. Bratt, O., et al., *CAG repeat length in the androgen receptor gene is related to age at diagnosis of prostate cancer and response to endocrine therapy, but not to prostate cancer risk.* Br J Cancer, 1999. **81**(4): p. 672-6.
141. Chang, B.L., et al., *Polymorphic GGC repeats in the androgen receptor gene are associated with hereditary and sporadic prostate cancer risk.* Hum Genet, 2002. **110**(2): p. 122-9.
142. Chen, C., et al., *Androgen receptor polymorphisms and the incidence of prostate cancer.* Cancer Epidemiol Biomarkers Prev, 2002. **11**(10 Pt 1): p. 1033-40.
143. Correa-Cerro, L., et al., *(CAG)nCAA and GGN repeats in the human androgen receptor gene are not associated with prostate cancer in a French-German population.* Eur J Hum Genet, 1999. **7**(3): p. 357-62.
144. Cude, K.J., et al., *The role of an androgen receptor polymorphism in the clinical outcome of patients with metastatic prostate cancer.* Urol Int, 2002. **68**(1): p. 16-23.
145. Edwards, S.M., et al., *Androgen receptor polymorphisms: association with prostate cancer risk, relapse and overall survival.* Int J Cancer, 1999. **84**(5): p. 458-65.
146. Ekman, P., et al., *Links between genetic and environmental factors and prostate cancer risk.* Prostate, 1999. **39**(4): p. 262-8.
147. Giovannucci, E., et al., *The CAG repeat within the androgen receptor gene and its relationship to prostate cancer.* Proc Natl Acad Sci U S A, 1997. **94**(7): p. 3320-3.
148. Gsur, A., et al., *Polymorphic CAG repeats in the androgen receptor gene, prostate-specific antigen polymorphism and prostate cancer risk.* Carcinogenesis, 2002. **23**(10): p. 1647-51.
149. Hakimi, J.M., et al., *Androgen receptor variants with short glutamine or glycine repeats may identify unique subpopulations of men with prostate cancer.* Clin Cancer Res, 1997. **3**(9): p. 1599-608.
150. Hardy, D.O., et al., *Androgen receptor CAG repeat lengths in prostate cancer: correlation with age of onset.* J Clin Endocrinol Metab, 1996. **81**(12): p. 4400-5.
151. Ho, G.Y., et al., *Transmission/disequilibrium tests of androgen receptor and glutathione S-transferase pi variants in prostate cancer families.* Int J Cancer, 2002. **98**(6): p. 938-42.

152. Hsing, A.W., et al., *Polymorphic CAG and GGN repeat lengths in the androgen receptor gene and prostate cancer risk: a population-based case-control study in China.* Cancer Res, 2000. **60**(18): p. 5111-6.
153. Ingles, S.A., et al., *Association of prostate cancer risk with genetic polymorphisms in vitamin D receptor and androgen receptor.* J Natl Cancer Inst, 1997. **89**(2): p. 166-70.
154. Irvine, R.A., et al., *The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer.* Cancer Res, 1995. **55**(9): p. 1937-40.
155. Lange, E.M., et al., *The polymorphic exon 1 androgen receptor CAG repeat in men with a potential inherited predisposition to prostate cancer.* Cancer Epidemiol Biomarkers Prev, 2000. **9**(4): p. 439-42.
156. Latil, A.G., et al., *Prostate carcinoma risk and allelic variants of genes involved in androgen biosynthesis and metabolism pathways.* Cancer, 2001. **92**(5): p. 1130-7.
157. Miller, E.A., et al., *Polymorphic repeats in the androgen receptor gene in high-risk sibships.* Prostate, 2001. **48**(3): p. 200-5.
158. Mir, K., et al., *The CAG trinucleotide repeat length in the androgen receptor does not predict the early onset of prostate cancer.* BJU Int, 2002. **90**(6): p. 573-8.
159. Modugno, F., et al., *Allelic variants of aromatase and the androgen and estrogen receptors: toward a multigenic model of prostate cancer risk.* Clin Cancer Res, 2001. **7**(10): p. 3092-6.
160. Mononen, N., et al., *Androgen receptor CAG polymorphism and prostate cancer risk.* Hum Genet, 2002. **111**(2): p. 166-71.
161. Nam, R.K., et al., *Significance of the CAG repeat polymorphism of the androgen receptor gene in prostate cancer progression.* J Urol, 2000. **164**(2): p. 567-72.
162. Panz, V.R., et al., *Tandem CAG repeats of the androgen receptor gene and prostate cancer risk in black and white men.* Endocrine, 2001. **15**(2): p. 213-6.
163. Platz, E.A., et al., *The androgen receptor gene GGN microsatellite and prostate cancer risk.* Cancer Epidemiol Biomarkers Prev, 1998. **7**(5): p. 379-84.
164. Platz, E.A., et al., *Racial variation in prostate cancer incidence and in hormonal system markers among male health professionals.* J Natl Cancer Inst, 2000. **92**(24): p. 2009-17.
165. Santos, M.L., et al., *Androgen receptor CAG repeat polymorphism in prostate cancer from a Brazilian population.* Cancer Detect Prev, 2003. **27**(5): p. 321-6.
166. Shibata, A., et al., *Polymorphisms in the androgen receptor and type II 5 alpha-reductase genes and prostate cancer prognosis.* Prostate, 2002. **52**(4): p. 269-78.
167. Stanford, J.L., et al., *Polymorphic repeats in the androgen receptor gene: molecular markers of prostate cancer risk.* Cancer Res, 1997. **57**(6): p. 1194-8.
168. Suzuki, H., et al., *CAG polymorphic repeat lengths in androgen receptor gene among Japanese prostate cancer patients: potential predictor of prognosis after endocrine therapy.* Prostate, 2002. **51**(3): p. 219-24.
169. Visvanathan, K., et al., *Association among an ornithine decarboxylase polymorphism, androgen receptor gene (CAG) repeat length and prostate cancer risk.* J Urol, 2004. **171**(2 Pt 1): p. 652-5.
170. Watanabe, M., et al., *Genetic alterations of androgen receptor gene in Japanese human prostate cancer.* Jpn J Clin Oncol, 1997. **27**(6): p. 389-93.

171. Xue, W., et al., *Susceptibility to prostate cancer: interaction between genotypes at the androgen receptor and prostate-specific antigen loci.* Cancer Res, 2000. **60**(4): p. 839-41.
172. Nowacka-Zawisza, M., et al., *Dinucleotide repeat polymorphisms of RAD51, BRCA1, BRCA2 gene regions in breast cancer.* Pathol Int, 2008. **58**(5): p. 275-81.
173. Vashist, Y.K., et al., *Microsatellite GTn-repeat polymorphism in the promoter of heme oxygenase-1 gene is an independent predictor of tumor recurrence in male oral squamous cell carcinoma patients.* J Oral Pathol Med, 2008. **37**(8): p. 480-4.
174. Wang, L., et al., *Association of a functional tandem repeats in the downstream of human telomerase gene and lung cancer.* Oncogene, 2003. **22**(46): p. 7123-9.
175. Wang, S., et al., *A novel variable number of tandem repeats (VNTR) polymorphism containing Sp1 binding elements in the promoter of XRCC5 is a risk factor for human bladder cancer.* Mutat Res, 2008. **638**(1-2): p. 26-36.
176. Giovannucci, E., *Insulin, insulin-like growth factors and colon cancer: a review of the evidence.* J Nutr, 2001. **131**(11 Suppl): p. 3109S-20S.
177. Rosen, C.J., et al., *Association between serum insulin growth factor-I (IGF-I) and a simple sequence repeat in IGF-I gene: implications for genetic studies of bone mineral density.* J Clin Endocrinol Metab, 1998. **83**(7): p. 2286-90.
178. Gebhardt, F., K.S. Zanker, and B. Brandt, *Modulation of epidermal growth factor receptor gene transcription by a polymorphic dinucleotide repeat in intron 1.* J Biol Chem, 1999. **274**(19): p. 13176-80.
179. Estany, J., et al., *Association of CA repeat polymorphism at intron 1 of insulin-like growth factor (IGF-I) gene with circulating IGF-I concentration, growth, and fatness in swine.* Physiol Genomics, 2007. **31**(2): p. 236-43.
180. Hoyo, C., et al., *Predictors of variation in serum IGF1 and IGFBP3 levels in healthy African American and white men.* J Natl Med Assoc, 2009. **101**(7): p. 711-6.
181. Chen, W., et al., *Phenotypes and genotypes of insulin-like growth factor 1, IGF-binding protein-3 and cancer risk: evidence from 96 studies.* Eur J Hum Genet, 2009. **17**(12): p. 1668-75.
182. Espelund, U., et al., *Elevated free IGF2 levels in localized, early-stage breast cancer in women.* Eur J Endocrinol, 2008. **159**(5): p. 595-601.
183. Renehan, A.G., et al., *Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis.* Lancet, 2004. **363**(9418): p. 1346-53.
184. Shi, R., et al., *IGF-I and breast cancer: a meta-analysis.* Int J Cancer, 2004. **111**(3): p. 418-23.
185. Warren, R.S., et al., *Induction of vascular endothelial growth factor by insulin-like growth factor 1 in colorectal carcinoma.* J Biol Chem, 1996. **271**(46): p. 29483-8.
186. Baserga, R., *The insulin-like growth factor I receptor: a key to tumor growth?* Cancer Res, 1995. **55**(2): p. 249-52.
187. Kaulfuss, S., et al., *Dual silencing of insulin-like growth factor-I receptor and epidermal growth factor receptor in colorectal cancer cells is associated with decreased proliferation and enhanced apoptosis.* Mol Cancer Ther, 2009. **8**(4): p. 821-33.

188. Bjorntorp, P., *Metabolic implications of body fat distribution.* Diabetes Care, 1991. **14**(12): p. 1132-43.
189. Donahue, R.P. and R.D. Abbott, *Central obesity and coronary heart disease in men.* Lancet, 1987. **2**(8569): p. 1215.
190. Kissebah, A.H., et al., *Relation of body fat distribution to metabolic complications of obesity.* J Clin Endocrinol Metab, 1982. **54**(2): p. 254-60.
191. Koivisto, V.A., H. Yki-Jarvinen, and R.A. DeFronzo, *Physical training and insulin sensitivity.* Diabetes Metab Rev, 1986. **1**(4): p. 445-81.
192. Krotkiewski, M., et al., *Impact of obesity on metabolism in men and women. Importance of regional adipose tissue distribution.* J Clin Invest, 1983. **72**(3): p. 1150-62.
193. Regensteiner, J.G., et al., *Relationship between habitual physical activity and insulin levels among nondiabetic men and women. San Luis Valley Diabetes Study.* Diabetes Care, 1991. **14**(11): p. 1066-74.
194. Potter, J.D., et al., *Colon cancer: a review of the epidemiology.* Epidemiol Rev, 1993. **15**(2): p. 499-545.
195. Riccardi, G. and A.A. Rivellese, *Effects of dietary fiber and carbohydrate on glucose and lipoprotein metabolism in diabetic patients.* Diabetes Care, 1991. **14**(12): p. 1115-25.
196. Powell, D.R., et al., *Insulin inhibits transcription of the human gene for insulin-like growth factor-binding protein-1.* J Biol Chem, 1991. **266**(28): p. 18868-76.
197. Kajantie, E., et al., *Serum insulin-like growth factor (IGF)-I and IGF-binding protein-1 in elderly people: relationships with cardiovascular risk factors, body composition, size at birth, and childhood growth.* J Clin Endocrinol Metab, 2003. **88**(3): p. 1059-65.
198. Valko, M., et al., *Role of oxygen radicals in DNA damage and cancer incidence.* Mol Cell Biochem, 2004. **266**(1-2): p. 37-56.
199. Palles, C., et al., *Identification of genetic variants that influence circulating IGF1 levels: a targeted search strategy.* Hum Mol Genet, 2008. **17**(10): p. 1457-64.
200. Colangelo, L.A., et al., *IGF-1, IGFBP-3, and nutritional factors in young black and white men: the CARDIA Male Hormone Study.* Nutr Cancer, 2005. **53**(1): p. 57-64.
201. Cruickshank, J.K., et al., *Epidemiology of the insulin-like growth factor system in three ethnic groups.* Am J Epidemiol, 2001. **154**(6): p. 504-13.
202. Platz, E.A., et al., *Racial variation in insulin-like growth factor-1 and binding protein-3 concentrations in middle-aged men.* Cancer Epidemiol Biomarkers Prev, 1999. **8**(12): p. 1107-10.
203. McGreevy, K.M., et al., *Impact of nutrients on insulin-like growth factor-I, insulin-like growth factor binding protein-3 and their ratio in African American and white males.* Public Health Nutr, 2007. **10**(1): p. 97-105.
204. Weisberg, I., et al., *A second genetic polymorphism in methylenetetrahydrofolate reductase (MTHFR) associated with decreased enzyme activity.* Mol Genet Metab, 1998. **64**(3): p. 169-72.
205. Chen, J., et al., *Linkage disequilibrium between the 677C>T and 1298A>C polymorphisms in human methylenetetrahydrofolate reductase gene and their contributions to risk of colorectal cancer.* Pharmacogenetics, 2002. **12**(4): p. 339-42.

206. Reeves, S.G., et al., *MTHFR 677 C>T and 1298 A>C polymorphisms and the age of onset of colorectal cancer in hereditary nonpolyposis colorectal cancer.* Eur J Hum Genet, 2009. **17**(5): p. 629-35.
207. Yin, G., et al., *Methylenetetrahydrofolate reductase C677T and A1298C polymorphisms and colorectal cancer: the Fukuoka Colorectal Cancer Study.* Cancer Sci, 2004. **95**(11): p. 908-13.
208. Sharp, L. and J. Little, *Polymorphisms in genes involved in folate metabolism and colorectal neoplasia: a HuGE review.* Am J Epidemiol, 2004. **159**(5): p. 423-43.
209. Giovannucci, E., *Epidemiologic studies of folate and colorectal neoplasia: a review.* J Nutr, 2002. **132**(8 Suppl): p. 2350S-2355S.
210. Sanjoaquin, M.A., et al., *Folate intake and colorectal cancer risk: a meta-analytical approach.* Int J Cancer, 2005. **113**(5): p. 825-8.
211. Hubner, R.A. and R.S. Houlston, *Folate and colorectal cancer prevention.* Br J Cancer, 2009. **100**(2): p. 233-9.
212. Mason, J.B., et al., *A temporal association between folic acid fortification and an increase in colorectal cancer rates may be illuminating important biological principles: a hypothesis.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(7): p. 1325-9.
213. Cole, B.F., et al., *Folic acid for the prevention of colorectal adenomas: a randomized clinical trial.* Jama, 2007. **297**(21): p. 2351-9.
214. Logan, R.F., et al., *Aspirin and folic acid for the prevention of recurrent colorectal adenomas.* Gastroenterology, 2008. **134**(1): p. 29-38.
215. Song, J., et al., *Effects of dietary folate on intestinal tumorigenesis in the apcMin mouse.* Cancer Res, 2000. **60**(19): p. 5434-40.
216. Song, J., et al., *Chemopreventive effects of dietary folate on intestinal polyps in Apc+/-Msh2-/- mice.* Cancer Res, 2000. **60**(12): p. 3191-9.
217. Jones, J.S., et al., *DNMT3b polymorphism and hereditary nonpolyposis colorectal cancer age of onset.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(5): p. 886-91.
218. Shen, H., et al., *A novel polymorphism in human cytosine DNA-methyltransferase-3B promoter is associated with an increased risk of lung cancer.* Cancer Res, 2002. **62**(17): p. 4992-5.
219. Jin, F., et al., *Up-regulation of DNA methyltransferase 3B expression in endometrial cancers.* Gynecol Oncol, 2005. **96**(2): p. 531-8.
220. Mizuno, S., et al., *Expression of DNA methyltransferases DNMT1, 3A, and 3B in normal hematopoiesis and in acute and chronic myelogenous leukemia.* Blood, 2001. **97**(5): p. 1172-9.
221. Montgomery, K.G., et al., *The DNMT3B C-->T promoter polymorphism and risk of breast cancer in a British population: a case-control study.* Breast Cancer Res, 2004. **6**(4): p. R390-4.
222. Patra, S.K., et al., *DNA methyltransferase and demethylase in human prostate cancer.* Mol Carcinog, 2002. **33**(3): p. 163-71.
223. Chang, C.H., et al., *Significant association of XPD codon 312 single nucleotide polymorphism with bladder cancer susceptibility in Taiwan.* Anticancer Res, 2009. **29**(10): p. 3903-7.
224. Koessler, T., et al., *Common variants in mismatch repair genes and risk of colorectal cancer.* Gut, 2008. **57**(8): p. 1097-101.