MODIFIER GENE POLYMORPHISMS AND INFLUENCE ON DISEASE EXPRESSION IN HNPCC

MASTER THESIS IN PHARMACY

Ingvild Synnøve Brenne

2008/2009

Supervisors
Prof. Rodney J. Scott, University in Newcastle
Prof. Thrina Loennechen, University in Tromsø

Department of Pharmacology
The institute of Pharmacy
Faculty of Medicine
University of Tromsø

ACKNOWLEDGEMENT

First I would like to thank Professor Rodney J. Scott and Dr. Bente Talseth-Palmer for giving me the opportunity to do my Master thesis at the Hunter Medical Research Institute (HMRI) and for presenting me with an extremely interesting research project. Thank you Bente for helping me with everything, both work related and not, and for making my stay in Australia become a memorable time. Also, thank you Professor Thrina Loennoechen for allowing me to travel to the other side of the world to pursue my academic goals. A special thanks to Berit Stiberg for her help and support when needed the most.

I would also like to thank everyone at Information Based Medicine for making me feel welcome, for sharing laughter and for guidance throughout the year.

Eva Kristin Andreassen, my fellow researcher and friend- thank you for sharing this experience with me.

A special thanks to Dr. Katie Ashton for her very much appreciated friendship, support, for supervising me through this project and time spent proofreading the draft of this thesis. I will always be grateful for everything you have done for me.

To all my friends back home, your friendship and continues support means a lot for me.

My family, thank you for the love, support and encouragement you always give me. You thought me that there is nothing that holds me back when it comes to achieving what I want in life. However, without your love and encouragement throughout my studies this thesis would not have been completed. Mum and dad, you are the best. Thank you!

Newcastle May 2009

Ingvild Synnøve Brenne

ABSTRACT

Hereditary non-polyposis colorectal cancer (HNPCC) is an autosomal dominantly inherited predisposition to a variety of epithelial malignancies most notable colorectal and endometrial cancer. Unlike other genetic predispositions to colorectal cancer HNPCC does not present with a premalignant phenotype, which as such makes it difficult to predict when or if an affected person will present with disease. The genetic basis of HNPCC has been conclusively shown to be due to mutations in genes involved in DNA mismatch repair. Four DNA mismatch repair genes are associated with HNPCC; hMSH2, hMLH1, hMSH6 and PMS2, with the majority affecting hMLH1 and hMSH2.

Aim
Mutations in any of the DNA mismatch repair genes have not been shown to result in any obvious genotype/phenotype correlation such that it is not possible to predict with any accuracy the type or age of onset of disease in persons harbouring the same mutation within and between families. The most likely explanation for disease variation in persons harbouring the same mutation is either a genetic modifier of disease risk, environmental insult or a combination of both. Several reports suggest that genetic modifiers of disease risk are capable of influencing the age of disease onset in HNPCC and it is likely that many of them have not yet been identified. Recently several genome-wide association studies have revealed a number of colorectal cancer susceptibility loci on chromosomes 10p14, 8q23.3, 8q24, 11q23 and 18q21. These loci are of particular importance as they are associated with an increased risk of colorectal cancer and may therefore act as modifiers of disease risk in individuals diagnosed with HNPCC.

Materials and methods
373 Australian and 311 Polish HNPCC patients with a molecular diagnosis of HNPCC have been examined for nine polymorphisms in the five loci described above. All DNA samples were genotyped to determine the allele frequency in the nine polymorphisms investigated. A statistically evaluation of the exact nature of the effect on disease risk was assessed using the statistical software package SPSS Graduate Pack Version 12.0.

Results
The statistical analysis revealed a number of significant results indicating that there might be a correlation between the different polymorphisms and disease development. In this study, hMLH1 mutation carriers harbouring the variant genotype for polymorphism rs3802842 were associated to development of colorectal cancer (CRC) at an earlier age than hMLH1 carriers harbouring the heterozygous or wild type genotype. This suggests that the particular polymorphism might act as a modifier for disease development in hMLH1 mutation carriers. Interestingly, these findings have been observed in both the Australian and Polish population.

Conclusion

In this study, some of the polymorphisms investigated have been associated with an altered risk of disease in HNPCC patients. Still, to confirm a correlation between disease development and the presence of some polymorphisms, more HNPCC populations needs to be studied. This will provide more information to specifically assess the likelihood of disease risk in HNPCC and thereby providing better tools for patient assessment.

TABLE OF CONTENTS

ABBREVIATIONS

BMP	Bone morphogenetic protein
CRC	Colorectal cancer
DDR	DNA damage response
DNA	Deoxyribonucleic acid
EIF3H	Eukaryotic initiation factor 3 subunit H, gene recently associated with increased risk of developing CRC
GREM1	Gremlin 1 Cystein knot superfamily, gene recently associated with increased risk of developing CRC
hMLH1	human Mutant L Homolog 1, gene associated with HNPCC
hMSH2	human Muatant S Homolog 2, gene associated with HNPCC
hMSH6	human Muatant S Homolog 6, gene associated with HNPCC
HNPCC	Hereditary Non-Polyposis colorectal cancer
MMR	Mismatch repair genes
mRNA	Messenger Ribonucleic Acid
MSI	Microsatellite instability
NCBI	National Centre for Biotechnology Information
PMS2	Post meiotic segregation 2, gene associated with HNPCC
POU5F1P1	POU class 5 homeobox 1 pseudogene 1
RT-PCR	Real time-Polymerase chain reaction, machine to genotype DNA samples
SCG5	Secretogranin (SCG5) V (7B2 protein)
SMAD7	Smad family member 7
SNP	Single nucleotide polymorphism
SPSS	Statistical package for the social sciences
(+)	Affected with cancer
(-)	Unaffected with cancer

1 INTRODUCTION

1.1 Hereditary Non-Polyposis Colorectal Cancer (HNPCC)

Hereditary Non-Polyposis Colorectal Cancer (HNPCC), also known as Lynch syndrome is an autosomal dominantly inherited cancer syndrome which is associated with inherited defects in the DNA mismatch repair (MMR) pathway [1-5]. It is characterized by early age of disease onset, neoplastic lesions, microsatellite instability and increased incidence of extra-colonic cancers such as carcinoma of the endometrium, ovary, stomach, small bowel, ureter and renal pelvis [6-8]. Of these cancers, endometrial cancer is the most common cancer after colorectal cancer [9]. Interestingly, females with MMR mutation have a greater chance of developing endometrial cancer than colorectal cancer [9, 10]. HNPCC is the most common form of hereditary colon cancer, accounting for approximately 2-8% of all colorectal cancers (CRC), depending on the population studied [4]. HNPCC is not characterized as a cancer but a syndrome that increases a person's risk of developing cancer. A person with HNPCC might never develop cancer, but their risk of developing cancer is much higher than in the general population. Up until 2007 it was considered that 4% of the CRC population would develop CRC as a result of harbouring mutations in DNA mismatch repair genes [11]. Furthermore, disease penetrance estimates suggested that approximately 80% of men and 40% of women with HNPCC would develop CRC [6, 12]. However, more recent analysis suggests that 45% of men will develop CRC and 15% of females will develop endometrial cancer due to mutations in the mismatch repair pathway [11].

The development of CRC involves a number of steps, where environmental factors and endogenous carcinogens provoke initiation and proliferation of cells within the colon [13]. This causes the activation of oncogenes and the inactivation of tumour suppressor genes [14]. CRC is known as a multifactorial disease which suggests that many underlying processes influence cancer progression [12, 13] and these include genetic and environmental factors.

1.1.1 History of HNPCC

The predisposition of some families to develop cancer at a young age was first recognised by Alfred Warthin in 1895, when his seamstress, from her family history, predicted her own death from cancer [5]. By studying her family extensively he found that gynaecological, colonic and stomach cancer occurred frequently [5]. However it was not until the 1980s that the idea of cancer family syndromes became fully accepted and the term Lynch syndrome was first used [5]. Lynch syndrome was first divided into type I and type II, depending on whether individuals developed extra-colonic cancers [5]. Later when it was discovered that both of these syndromes were due to the inherited defects in the genes that regulate the excision of errors occurring during DNA replication (mismatch repair genes), it became clear that they were manifestations of the same disease [5]. Both syndromes are usually referred to as HNPCC, but the term Lynch syndrome is still used [5].

1.1.2 Clinical findings

The term Hereditary Non-Polyposis Colorectal Cancer indicates that there are no polyps present in those patients diagnosed with the syndrome. This is however not the case as a small number of polyps can be present in up to 30 per cent of affected patients [15-17].The polyps associated with HNPCC tend to occur at an early age and are larger in size than in the general bowel cancer population [5].

1.1.3 Diagnosis of HNPCC

Family history is the primary method used for identifying patients/families for HNPCC testing [9]. The diagnosis of HNPCC is based on the meeting of four criteria, known as the Amsterdam criteria (listed below), which originally derived as an aid in identifying the genetic basis of the disease.

Amsterdam criteria for HNPCC patients:

1) Three or more relatives with histological verified colorectal cancer, one of whom is the first degree relative of the other two.
2) Colorectal cancer involving at least two successive generations.
3) At least one relative diagnosed with colorectal cancer under the age of 50.
4) Exclusion of familial adenomatous polyposis [4, 9, 18].

Since nearly 60 per cent of families failed to fit every criterion due to the appearance of extracolonic cancer, they were revised in 1999 (Amsterdam criteria II), to also include the cancers of the endometrium, ureter, renal pelvis and small bowel [19]. Families found to meet all criteria have a greater probability of being diagnosed with HNPCC syndrome. It is important to recognise these characteristics and offer genetic testing and counselling to those being affected with the syndrome [9]. The Bethesda guidelines, developed in 1996, are similar to the Amsterdam criteria. However, the major difference is that the Bethesda criteria aims to test patient's tumours for microsatellite instability before screening for mutations in the MMR genes [20, 21]. Both, the Amsterdam and Bethesda criteria are clinically used to diagnose possible HNPCC patients [20].

1.2 DNA Mismatch Repair Genes

HNPCC is caused by germline mutations in DNA mismatch repair (MMR) genes [22]. In HNPCC patients at least four DNA mismatch repair (MMR) genes have been associated with the disease: hMSH2, hMLH1, hMSH6 and PMS2, with the majority affecting hMLH1 and hMSH2 [1-4, 23]. Germline mutations in one of these 4 MMR genes, are detected in 70%-80% of families diagnosed with HNPCC [24]. More than 400 different predisposing MMR gene mutation are known, with approximately 50% affecting hMLH1, 40% hMSH2, 10% hMSH6 and less than 5% affecting PMS2 [24]. If a somatic mutation inactivates the remaining wild type allele, the affected cell will potentially accumulate new mutations at a very high rate [7]. This will enhance the potential for malignant transformation and the possibility to of developing cancer [7].

The names of the mismatch repair genes associated with HNPCC were a result of their structural similarity to the bacterial proteins, known as MutS, MutL and MutH and postmeiotic segregation (PMS2) [25-27]. The reason for this being the mismatch repair system was first studied in bacteria [27]. hMSH2 and hMSH6 are both similar to the bacteria protein MutS and is an abbreviation of human MutS Homolog, as well as the hMLH1 is an abbreviation of human MutL Homolog. The postmeiotic segregation (PMS2) gene was named before the function of the protein was elucidated [26].

1.2.1 Germline mutation

The principle behind inherited genetic risk is based on Knudson's 'two hit' hypothesis. All somatic cells contain two copies (or alleles) of any gene, therefore both alleles must be disrupted or 'hit' before genetic function is altered [14, 28]. Patients with an inherited risk of cancer due to a diagnosis of HNPCC are born with one defective copy of a gene, inherited from a parent. This is known as a germline mutation. Individuals at risk only require a somatic mutation in the remaining normal allele before gene function is lost and thereby alter the risk of disease development in the HNPCC population [14, 29], which explains the lower age of onset of cancer in the HNPCC population.

1.3 DNA damage and DNA repair

DNA is constantly exposed to external and internal mutagenic agents, for example free radicals, ionising agents, UV light and different kinds of toxins [30, 31]. These agents can among others potentially affect the integrity of the genome [30, 31]. If DNA damage fails to be repaired, it will result in DNA mutations [31]. It is therefore extremely important that the DNA repair systems are working efficiently so they can respond to DNA damage at any time.

DNA damage response (DDR) consists of numerous of signalling events crucial for sensing DNA damage [32]. The major role of the DDR, in response to DNA damage, is to activate cellular responses such as cell cycle arrest, DNA repair, senescence and apoptosis [32]. Since the human cell is exposed to many agents that can cause damage to the genetic code [33], it is necessary to have an efficient DNA repair system. In response to this, cells have developed five different pathways in which DNA damage can be detected and repaired [31]. One of these major pathways is mismatch repair (MMR). Failure to repair DNA lesions can lead to unregulated cell growth and ultimately cancer.

1.4 Human DNA Mismatch Repair

The primary function of the MMR system is to eliminate base-base mismatches and insertion-deletion loops which arise as a consequence of DNA polymerase slippage during DNA replication [4]. These lesions typically affect non repetitive DNA leading to single base substitutions (for example G→T) and it also involves gains or losses of short repeat units (for example CACA) [4]. In humans, at least six different MMR proteins are required to recognise mismatches [4]. Mismatch recognition is mediated by either one of the two heterodimers MutSα or MutSβ. The MutSα is composed of the MutS homologs MSH2 and MSH6, while the MutSβ is composed of the MutS homologs MSH2 and MSH3 [34]. The MSH2-MSH6 heteroduplex, known as MutSα, identifies single-base mispairs and initiates excision of the mismatch base [5, 34]. The MSH2-MSH3 heteroduplex, known as MutSβ, predominantly recognizes larger DNA insertion loops [5, 34].

The interplay between MLH1 and PMS2 (known as MutLα), coordinates which mismatch recognition complex (either MutSα or MutSβ) that is necessary for MMR [4]. In humans, the PMS2 subunit within the MutLα complex possesses endonuclease activity which enables MutLα to insert random nicks at sites spanning the mismatch [34]. This leads to the activation of exonuclease (EXO1) in the direction 5`-3`, which removes the incorrect DNA fragment [34]. The remaining single-stranded gap is filled with polymerase δ and its cofactors, proliferation cell nuclear antigen (PCNA) and replication factor C (RFC) [34]. The mismatch repair is finished when the nicks are sealed by DNA ligase I [34]. See figure 1.1 for the MMR pathway.

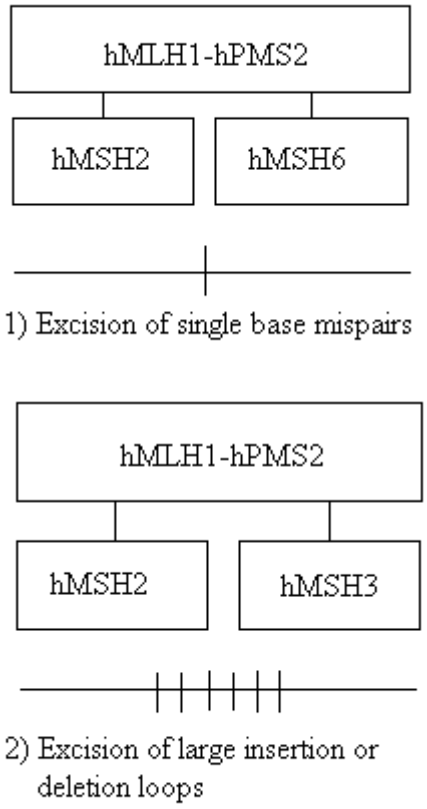


Figure 1.1: Mismatch repair pathway. The MSH2-MSH6 heteroduplex (MutSα) recognises single base mispairs and recruits heterodimer MLH1 and PMS2 to initiate repair of the DNA damage. The MSH2-MSH3 heteroduplex (MutSβ) recruits MLH1 and PMS2 to initiate repair of larger insertion deletion loops [5].

1.5 Microsatellite instability (MSI)

Nearly all HNPCC tumours display microsatellite instability (MSI) [21]. Microsatellite instability is a genetic signature of a tumour tissue that occurs as a result of DNA mismatch repair deficiency. Microsatellites are repetitive nucleotide sequences dispersed throughout the human genome [7]. These sequences do not code for any proteins and their function still remains largely unknown [5]. Mutations in MMR genes give rise to alterations in the number of repeat units in these sequences of DNA and thereby confer instability on the genome [9]. For this reason they are useful in detecting damage to DNA repair systems. MSI is defined as: "a change of any length due to either insertions or deletions of repeating units in a microsatellite within a tumour compared to normal tissue" and results from failure of the cell to repair

errors made during DNA replication [21]. Such errors are usually repaired by MMR proteins [24]. A defect in MMR genes allows errors to accumulate which will increase the risk of malignant transformation of the cells and MSI to develop [24]. If MSI is detected in a tumour sample it is likely that the individual will harbour a MMR gene mutation. MSI testing is often performed before screening for a mutation to decide which gene to test for as this is more efficient in terms of cost and time. However, MSI testing can not replace screening for mutations in MMR genes as not all HNPCC tumours display MSI. It should be noted as well that approximately 15% of all colorectal cancers display MSI.

1.7 Single Nucleotide Polymorphisms (SNPs)

Approximately 90% of all human genetic variation are differences in single bases of DNA, called single nucleotide polymorphisms (SNPs) [35]. For a variation to be defined a SNP it must occur with a frequency of at least 1 % in a given population [36, 37]. A SNP (pronounced “snip”) [37] occurs within a DNA sequence and appears when a single nucleotide, for example A is being replaced by one of the three other nucleotides- T, G or C. SNPs are found within the coding regions (exon) of genes, in the non coding regions (intron) or in the intergenic regions between genes. Although SNPs in the coding region are more likely to cause functional changes than SNPs elsewhere, not every SNP will affect gene function [35]. In fact, certain SNPs do not cause an increased risk of cancer as the amino acid that the SNP encodes will be the same as the amino acid in the absence of the SNP [38]. Some SNPs cause a change in the amino acid (also known as missense mutations) but the functional significance needs to be investigated to determine whether a change in amino acid affects protein function. SNPs that encode stop codons are also known as nonsense mutations, usually lead to a severely altered and potentially non functional protein [38]. Insertions or deletions cause the frame of the sequence to be altered which can lead to a translation error and results in the loss of protein function [38].

The human genome is estimated to contain one single polymorphism (SNP) for every 300 base pairs [39]. Overall, this accounts for a total of several million SNPs [40]. SNPs are therefore extremely important in the study of structure and history of the human genome [40]. SNPs can in fact, be directly responsible for genetic diseases, as

they may alter the genetic sequence of a regulatory region [41]. Since most SNPs are inherited from one generation to the next; they also represent a powerful tool to study the evolution of our species [41]. With many SNPs yet to be detected, large scale databases of SNPs have been developed over the years growing rapidly every day [37]. The main purpose of gathering SNP data in an accessible database is to make it possible for researchers around the world to perform association studies.

1.8 Association studies

The aim of association studies is to find a connection between SNP alleles and the development of certain diseases by investigating polymorphisms in possible causative candidate genes (modifier genes) [41]. This is accomplished by comparing two populations which are different from one another in terms of phenotype [41]. By measuring the frequency of SNPs in both populations one can detect those SNPs that show significant difference in frequency [41]. If there is evidence to suggest that a polymorphism in a gene increases the risk of disease, that polymorphism should be found at a significantly different frequency in those individuals with the disease compared to healthy controls [37]. This can also be applied for non-genetic factors, such as smoking which is associated with lung cancer [37]. Since association studies can help us to better understand development of disease and disease onset, SNPs have received a considerable amount of attention as they are being widely used in studies that focus on the effect of modifier genes in disease [41].

1.9 Modifier genes

A gene existing in two or more different forms or alleles within a population is said to be polymorphic. A polymorphic gene can also be referred to as a modifier gene, as modifier genes have the ability to alter the function of a gene and ultimately change the phenotypic expression [13]. Modifier genes can influence the frequency of expression of an allele and the variation in allelic expression from one individual to another [13]. They are also responsible for a phenomenon where a single gene is responsible for a number of distinct and seemingly unrelated phenotypic diseases [42]. This is the case for HNPCC families with a known mutation, where the disease expression varies between individuals that have the same mutation [43]. Some

individuals will therefore develop disease earlier, while others might not develop cancer at all. Different types of cancers can also be seen in patients with the same mutation as well as the intensity of the disease. This indicates that other genes (modifier genes) are likely to influence disease expression [44].

1.9.1 The role of modifier genes in HNPCC

Any human disease cannot be explained ultimately by a single gene [45] [46] [5]. In HNPCC patients there is considerable variation in disease expression (such as age of diagnosis and tumour site) which cannot be entirely explained by the type and position of the mutation in MMR genes. Several reports have shown that genetic modifiers may contribute to disease in HNPCC [47-49].

Single nucleotide polymorphisms (SNP) in the TP53 gene have been associated with the age of onset of CRC in HNPCC patients [50]. TP53 is a tumour suppressor gene, which regulates the transcription of genes necessary to maintain genomic stability. The role of tumour suppressor genes is to initiate apoptosis in cancer cells and blocking cell proliferation and cell growth after DNA damage [14, 51]. HNPCC patients heterozygous for the wild type allele in the R72P SNP developed colorectal cancer at an average age of 13 years younger than those who where homozygote wild type for this particular allele [50]. However, controversial reports have been reported by Talseth et al. 2006 suggesting that the age of diagnosis of CRC in HNPCC is more complex than predicted by R72P polymorphisms in TP53 [44].

Two single nucleotide polymorphisms (SNPs) in genes involved in the cell-cycle, Aurora A and Cyclin D1, have also been associated with the age of onset of CRC in HNPCC patients [44, 52 Kong, 2000 #81] HNPCC patients homozygous for the wild type allele (TT) of the T91A SNP (F31I) in Aurora-A developed CRC approximately 7 years earlier than patients carrying the variant allele [52]. Aurora-A is involved in normal cell cycle, but is overexpressed in a variety of malignancies [53]. Aurora-A regulates the G_2-to-M phase of the cell cycle [54]. If DNA damage occurs the activation of Aurora-A is inhibited [54]. It is believed that DNA repair might be involved in cell cycle control as it has been suggested that MMR genes are necessary to activate G_2-M checkpoint in the presence of certain types of DNA damage [55].

For the 870 G>A SNP in Cyclin D1 patients with the variant allele were associated with an earlier age of disease diagnosis by an average of 11 years compared to patients homozygous for the wild type allele [56]. Cyclin D1 has an important role in the G_1-to-S phase in the cell cycle [57]. However, the relationship between Cyclin D1 and disease expression appears to be more complex than first predicted. While one study shows an association between the polymorphism and the age of disease onset [56], another study from Finland failed to show a similar relationship [58]. The major difference between the two studies was the predominance hMSH2 carriers in one population compared to hMLH1 carriers in the other. To better understand the relationship between disease phenotype and certain types of polymorphisms large groups of HNPCC patients are required.

1.9.2 Why search for modifier genes in HNPCC?

In the context of human disease modifier genes have been defined as inherited genetic variation which can change one person's phenotype either quantitatively or qualitatively [59]. The search for modifier genes is important because it makes it possible to provide more specific diagnosis of disease risk in HNPCC patients. Identification and removal of colorectal adenomas (as well as other types of polyps) will almost certainly reduce the incidence of developing cancer [60]. Therefore if modifier genes affecting your risk of developing CRC are found in HNPCC patients, early screening can be initiated to reduce the incidence of disease development and progression. In addition, the identification of modifier genes associated with disease will also help to optimize the response to drug treatment in individuals [61]. Currently there are no frequent, regular and effective screening methods for colorectal tumours in the general population [60]. The challenge is therefore, to determine whether particular SNPs in genes affect the function of specific proteins involved in HNPCC. This is of great interest as the discovery of modifier genes influencing disease expression in HNPCC can be useful for genetic testing.

1.9.3 Benefits of identifying modifier genes

There are two major benefits of identifying modifier genes in HNPCC patients at an early age [60]. Firstly, those patients carrying specific modifiers would initially be screened regularly for any abnormalities. Secondly, it is a unique way to understand

cancer biology by knowing how mutations in different genes interact in the genetic pathways of tumorgenesis [60].

1.10 Candidate modifier genes

The SNPs studied in this project were chosen since they have previously been linked to cancer development. The decision to look at these specific SNPs were based on several genome wide association studies that revealed a number of colorectal cancer susceptibility loci on chromosome 10p14, 8q23.3, 8q24, 11q23 and 15q13. These loci are of particular importance as they are associated with an increased risk of developing CRC. The SNPs elected may therefore act as modifiers of disease risk in individuals diagnosed with HNPCC. The SNPs are to be found within or close to a variety of genes: EIF3, SMAD7, GREM1, SCG5, POU5F1P1, BC031880, LOC38996, LOC120376 and FLJ45803.

1.10.1 Eukaryotic initiation factor 3 (EIF3)

SNP rs16892766 is located on chromosome 8q23.3 within the subunit H of the EIF3 gene [62]. Translation initiation is controlled by numerous different translation factors and the most important one is the eukaryotic translation initiation factor 3 (EIF3) [63]. Translation initiation is an important step in the regulation of gene expression in eukaryotes [63]. Deregulation at this step causes abnormal gene expression, leading to altered cell growth and possibly cancer [63]. Cells that contain high levels of the H subunit, which is a protein subunit of EIF3, have been associated with different types of cancers [64]. Overexpression of EIF3H leads to strongly translated mRNAs, affecting translation, proliferation and a number of malignant phenotypes [64]. The failure to down-regulate protein synthesis leads to an overproduction of oncogenic proteins, resulting in malignant transformation of cells [63].

1.10.2 SMAD family member 7 (SMAD7)

SMAD7 is located at chromosome 18q21 and a genome wide association study indicated that SNP rs4939827 and rs4464148 are associated with CRC [65]. The SMAD signalling pathway plays an important role in tumorigenesis and progression in cancer [66]. The protein level of SMAD is associated with growth, inhibition and metastasis in a variety of human cancers [66]. A previous study has shown that there

is overexpression of SMAD7 in gastric cancer tissue [66]. This overexpression of SMAD7 in colon cancer cells induces tumorigenicity by blocking transforming growth factor (TGF)-beta-induced growth inhibition and apoptosis [67, 68].

1.10.3 Gremlin 1 (GREM1), cysteine knot superfamily

SNP rs10318 is located within the GREM1 gene which encodes a secreted bone morphogenetic protein (BMP) antagonist. GREM1 has shown to be overexpressed in several of human tumours, as for instance carcinomas of the colon [69]. The TGF-beta/BMP pathway is also known to have an important role in colorectal tumorigenesis [70]. It is therefore believed that GREM1 may increase tumour proliferation and thereby influence cancer development [71].

1.10.4 Secretogranin (SCG5) V (7B2 protein)

SNP rs4779584 lies between GREM1 and SCG5. SCG5 is a protein coding gene involved in neuroendocrine signalling which is thought to influence cellular proliferation [71]. A genome wide association study indicated that genetic variants at the 15q3.3 influence on colorectal cancer risk and SNPs near GREM1 and SCG5 were strongly associated with increased CRC risk [71]. SNP rs10318 and rs4779584 are both located on chromosome 15q13.3 [71].

1.10.5 POU class 5 homeobox 1 pseudogene (1 POU5F1P1)

SNP rs6983267 is located near the POU5F1P1 pseudogene on chromosome 8q24.21. This gene acts as a transcriptional activator and has been shown to encode a protein. This SNP is associated with increased risk of colon and prostate cancer [27].

SNP rs7014346 is also located on chromosome 8q24 but does not code for a protein. The SNP is located near the POU5F1P1 pseudogene, however its function is not well understood [72]. A genome wide association study has suggested an association between the particular loci and CRC [72].

1.10.6 BC031880 and LOC38996 gene

SNP rs10795668 is not found within any gene, but is thought to be near the BC031880 and the LOC38996 gene [62]. This SNP is located at chromosome 10q14 and a recent study has shown that this SNP is associated with CRC [62].

1.10.6 LOC120376 and FLJ45803 gene

SNP rs3802842 is found within the LOC120376 and FLJ45803 genes on chromosome 11q23 [72]. This SNP has recently been reported to show population differences in risk for developing CRC [72].

1.11 Aims and hypothesis of the Study

Mutations in the four mismatch repair (MMR) genes; hMLH1, hMSH2, hMSH6 and PMS2 are already known to be associated with HNPCC. The identification of predisposing mutations in these genes has demonstrated some genotype-phenotype correlation. However, there is still significant phenotypic variation among individuals with HNPCC. In fact, the age of diagnosis, severity of disease and the appearance of extracolonic cancers vary within families and between families that harbour the same mutation. The explanation for this disease variation is thought to be due to polymorphisms in modifier genes.

Recently, several genome wide association studies have revealed a number of colorectal cancer susceptibility loci on four chromosomes; 10p14, 8q23.3, 8q24, 11q23 and 15q13. Since these loci have been associated with an increased risk of sporadic CRC, they may also influence the cancer development modifiers in HNPCC patients.

The study hypothesis is that polymorphisms in modifier genes influence disease expression in HNPCC patients carrying a mutation in hMLH1, hMSH6 or hMSH2. The identification of polymorphisms with modifying effect can help to predict with more accuracy the type of cancer and age of disease onset in individuals harbouring the syndrome.

The aim of this project is to examining 373 Australian and 311 Polish HNPCC patients with a molecular diagnosis of HNPCC (mutation positive) for nine different polymorphisms in the five loci described above to determine if these polymorphisms have modifying effect on disease expression in HNPCC.

2 MATERIALS AND METHODS

2.1 Materials

The reagents and equipment used in this study are stated in the appendix 1.

2.2 HNPCC Participant Information

Patients diagnosed with HNPCC were selected for this study. The selection criteria were based on the molecular diagnosis of HNPCC. All patients harboured a mutation in one of the mismatch repair (MMR) genes, hMLH1, hMSH2 or hMSH6. All the patients were of Caucasian origin, but divided into two subpopulations according to the country in which the samples were collected, Australia or Poland.

2.2.1 Australian Population

373 samples were collected in the state of New South Wales, Australia. Of the 373 individuals, 165 (44%) had been diagnosed with colorectal cancer while 202 (54%) were not affected with colorectal cancer. In 6 (2%) of the cases it was unknown if the participant were affected with colorectal cancer or not. There were 25 (7%) affected with endometrial cancer, 7 (1.9%) with ovarian cancer, 7 (1.9%) with breast cancer, 5 (1.3%) with kidney cancer, 4 (1.1%) with melanoma, 4 (1.1%) with stomach, 3 with bladder and 2 with pancreatic cancer, which together account for less than 1%. Among the 165 affected with CRC, 22 (13%) had a recurrent CRC. There were a total of 221 (59%) females and 150 (40%) males. For 2 (1%) of the participant gender was unknown. There were 184 (49%) cases with germline hMLH1 mutation, 164 (44%) with hMSH2 and 25 (7%) with hMSH6. Of these mutation carriers there were 355 (95%) nonsense insertion, deletion or splice mutation (leading to a truncated protein) and 18 (5%) with missense mutation. All of the missense mutations were in the hMLH1 mutation carriers.

2.2.2 Polish Population

311 samples were collected from the Pomerian Academy of Medicine, Poland. Of the 311 individuals, 121 (39%) were affected with CRC while 190 (61%) were unaffected with CRC. Among the 121 affected with CRC, 10 (8%) had a recurrent CRC. Other types of cancer that could be seen within the Polish population were; 38 (12.2%)

individuals with endometrial cancer, 6 (1.9%) with ovarian, 3 with bladder, 2 with breast, 2 with kidney, 2 with stomach, 1 with lung and 1 with cervical cancer, in which all accounts for less than 1% of the cases. There were a total of 189 (61%) females and 122 (39%) males. In the Polish samples there were 172 (55%) with hMLH1 mutation, 114 (37%) with hMSH2 and 25 (8%) with hMSH6. There were no cases with missense mutation.

2.2.3 Ethical Consideration

All participants had given written consent for their DNA to be used for further ethics approved research into the cause of their condition at the time of counselling. HNPCC samples were supplied from the state centre for colorectal testing. Approval for this study was obtained from Hunter New England Health Research Ethics Committee (Australia), the Human Research Ethics Committee at the University of Newcastle (Australia) and the Ethics Committee of the Pomeranian Academy of Medicine (Poland).

2.2.4 Sample Groups

Altogether, 684 samples were genotyped in this study. To determine any association between the disease characteristics and polymorphisms investigated, the samples were subdivided into different subgroups according to:

I. Their gene mutation status (hMLH1, hMSH2 or hMSH6)
II. Disease expression (affected with CRC, unaffected with CRC and affected with gynaecological cancer *)
III. Gender (female or male)

* Females that were affected with endometrial cancer and ovarian cancer were placed into the same group called gynaecological cancer.

Each individual had previously contributed blood from which DNA was extracted using the salt precipitation method. Each DNA sample was diluted to a concentration of 50 approximately ηg/μL [73].

2.3 Genotyping

All DNA samples were genotyped to determine the genotype frequency in the nine single nucleotide polymorphisms (SNPs) investigated. Three different genotypes were obtained for each SNP: Homozygous wild type, where the sample did not contain the variant allele; heterozygous, where one allele had the variant allele and homozygous mutant, where both alleles harboured the variant allele. For example, SNP rs16892766 harbours a change of A to C, where the homozygous wild type is AA, the heterozygous is AC and the homozygous variant is CC. The letters A and C refers to the nucleotide present at the polymorphic site.

2.4 The Principle behind Real time PCR

Real time PCR is a high throughput technique that detects and quantitates fluorescence released from probes bound to nucleic acid sequence [74] [75]. The ABI PRISM® 7900HT (Applied Biosystems) machine distributes light from an argonic laser that excites samples in each well. The resulting fluorescence is captured by a charge couple device camera where fluorescence is detected between 500nm and 600nm [76]. This allows the CCD camera to capture fluorescence of multiple fluorophores [76, 77].

The real time PCR method used in this study uses two TaqMan probes that differ at the polymorphic site and carry a different 5' fluorophore. The probes are approximately 20 base pair (bp) sequences that are designed over the region of the SNP. One probe has the complementary sequence of the wild type allele and the other has the complementary sequence of the variant allele. A 5' reporter dye and a 3' quencher dye are covalently linked to the probes [78]. During the PCR annealing step, the probe binds specifically to the target polymorphic site [78]. If a probe binds to the polymorphic site, the 5' reporter is cleaved by the nuclease activity of the Taq polymerase which increases the characteristic fluorescence of the reporter dye [79, #99]. When a probe does not bind to the target site, the 3' quencher decreases the fluorescence of the 5' reporter, thereby minimizing background fluorescence[78].

The amount of amplified PCR product increases exponentially during PCR because after each successive cycle the probe is able to bind more DNA copies, which then increases the reporter fluorescence (Figure 2.2) [79]. Sequence detection system (SDS) software quantifies and compares the fluorescence signal which makes it possible to determine the allelic content of each sample on the plate. The values of the fluorescence in each well are plotted in a graph and it can then be determined whether the samples are homozygous wild type, heterozygous or homozygous variant.

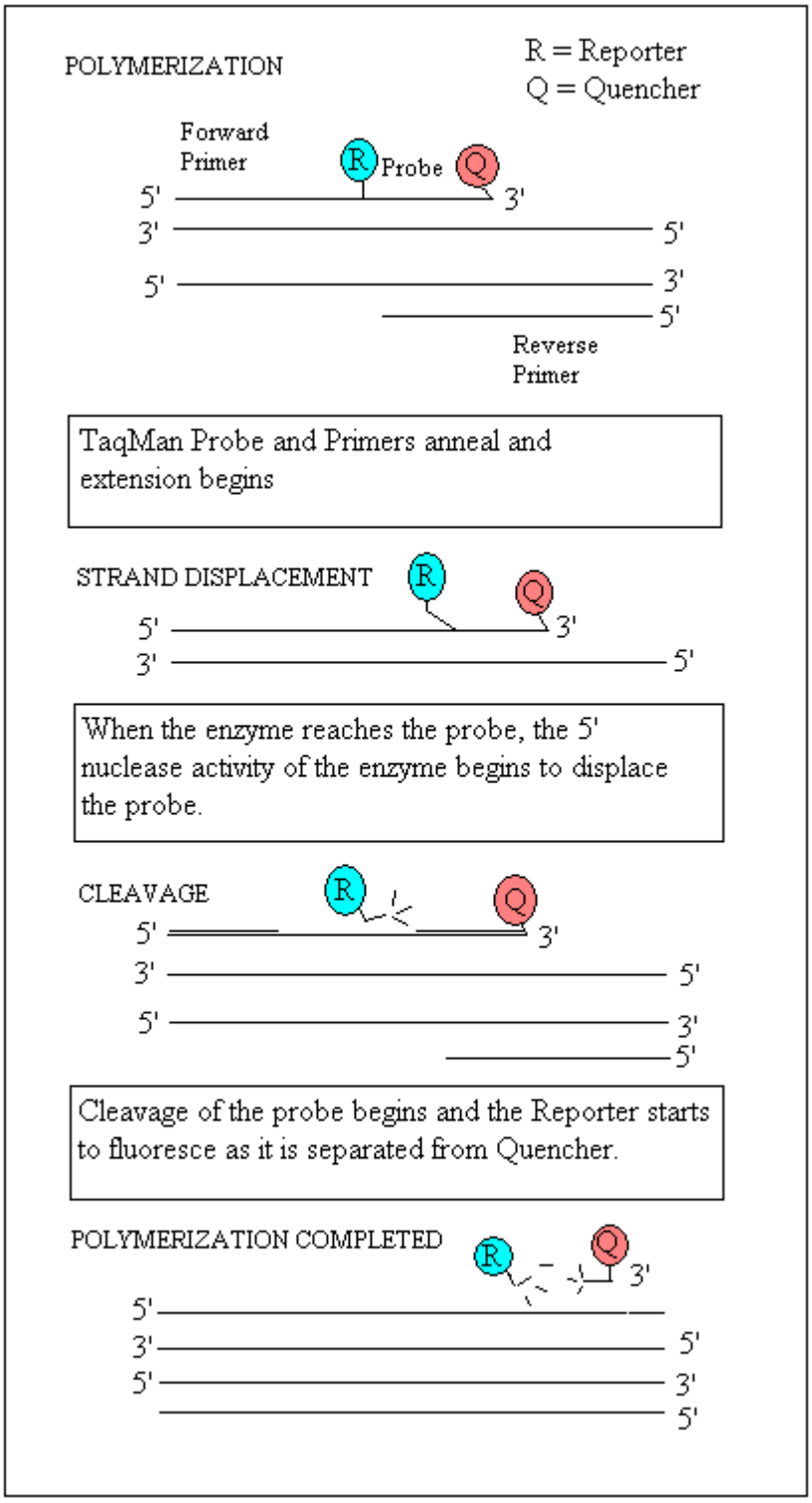


Figure 2.1: Chemistry of TaqMan real time PCR reactions. When the probe successfully binds to the target sequence, fluorescence is released from the TaqMan reporter probe. Reproduced with permission of Applied Biosystems, see appendix 2 [79].

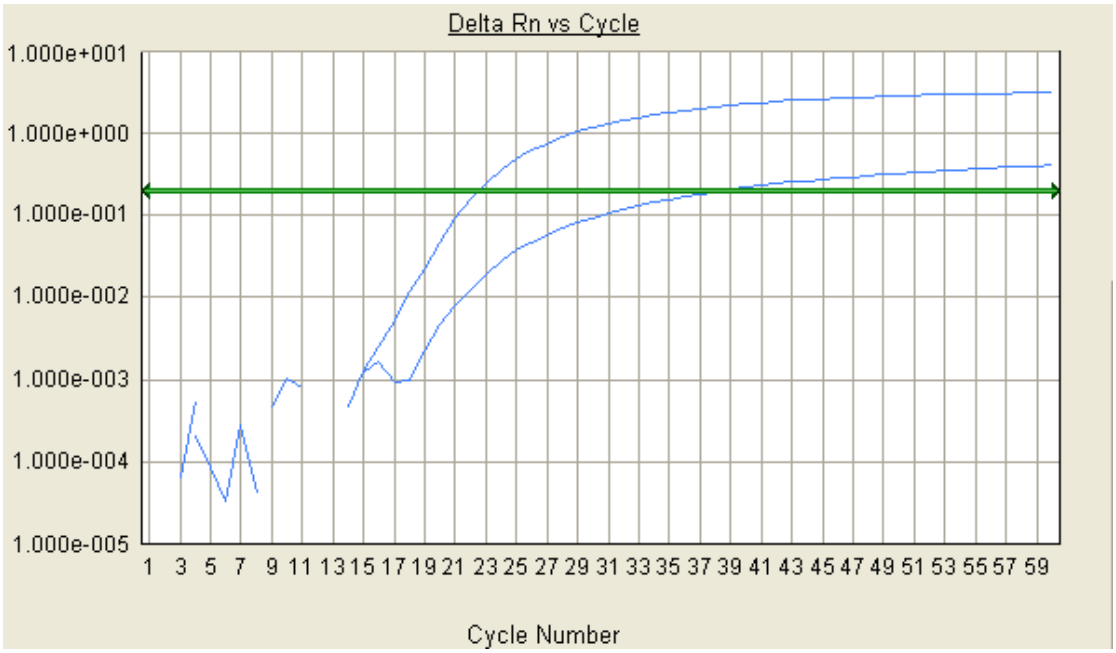


Figure 2.2: Amplification plot. The amplification of DNA can be viewed using real time PCR. The upper blue line shows the amplification of the wild type allele in the DNA while the lower blue line shows the amplification of the variant allele.

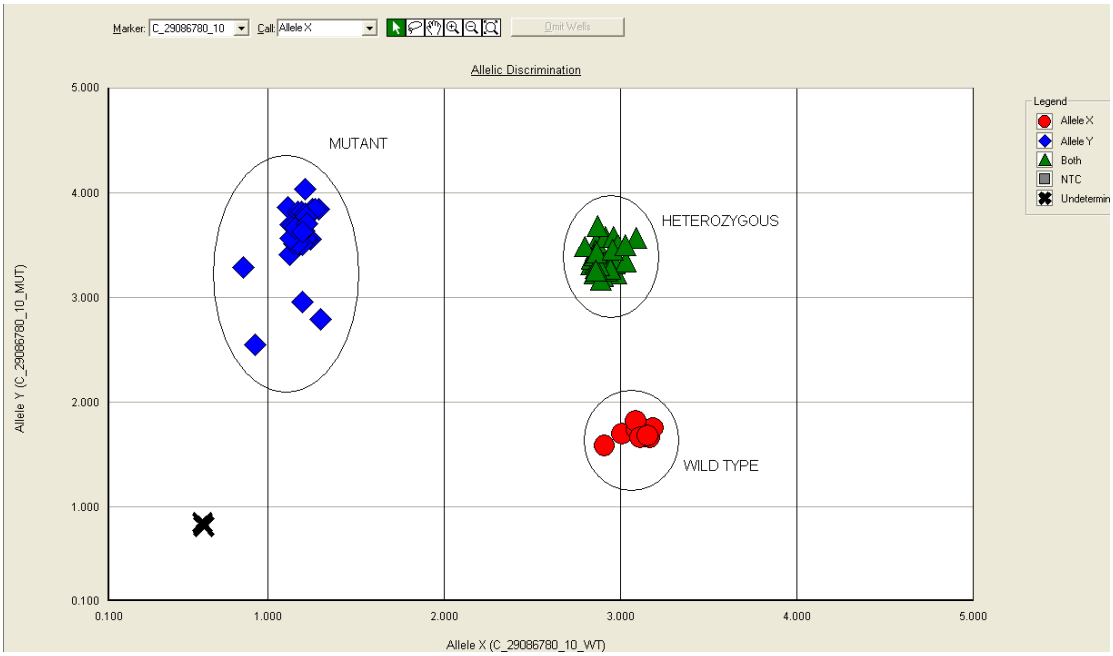


Figure 2.3: Allelic discrimination graph. The allelic discrimination graph shows difference in fluorescence between the wild type, heterozygous and variant which makes it possible to determine the genotype in each sample.

2.5.1 Design of primers and probes for RT-PCR

Primers and probes for the SNPs rs4939827, rs4464148, rs6983267, rs16892766, rs10795668, rs3802842, rs7014346, rs4779584 and rs10318 were designed by Assay-on-Demand which is a service offered by Applied Biosystems (PE Applied Biosystems). The nine SNPs and their respective NCBI reference, assay ID, allele, chromosome and gene are shown in table 2.1.

Table 2.1: SNP information. Listed below are the nine SNPs and their NCBI reference, assay ID, alleles, chromosome and gene.

NCBI SNP Reference	Assay ID	Alleles	Chromosome	Gene
rs16892766	C_32670283_10	A>C	Chr 8	EIF3H
rs3802842	C_27503482_10	A>C	Chr 11	LOC120376;FLJ45803
rs10318	C_12070332_20	C>T	Chr 15	GREM1
rs4939827	C_27913406_10	C>T	Chr 18	SMAD 7
rs4464148	C_27989234_10	T>C	Chr 18	SMAD 7
rs6983267	C_29086771_20	T>G	Chr 8	POU5F1P1
rs7014346	C_29086780_10	G>A	Chr 8	POU5F1P1
rs4779584	C_28019826_10	C>T	Chr 15	SCG5
rs10795668	C_1779559_10	G>A	Chr 10	BC031880; LOC38996

Approximately 250bp of the DNA sequence either side of the SNP was sent to Applied Biosystems which was used to design the appropriate primers and probes, which are also tested and validated by the company. The context sequence for each SNP is listed in table 2.2.

Table 2.2: Context sequence. Context sequence from Assay-on-Demand Service for the nine polymorphisms

rs16892766 Reverse Sequence	5' AGACGCAAACAGTTTCAAGACTATT**[A/C]**GCTGTTAAAGGTTATGCCTTATGTC
rs3802842 Forward Sequence	5' GCCCTTGCAGACCCATAGAAAATCT**[A/C]**TCCCAGAAATTCACCTCATTTTAGG
rs10318 Forward Sequence	5' AAGATATTTGTGGTCTTGATCATAC**[C/T]**TATTAAAATAATGCCAAACACCAAA
rs4939827 Forward Sequence	5' TCACAGCCTCATCCAAAAGAGGAAA**[C/T]**AGGACCCCAGAGCTCCCTCAGACTC
rs4464148 Reverse Sequence	5' GGGGGAACAGACAGAGAAGGATGAA**[C/T]**GTGAAAAGGAAACACCCTGGTAACT
rs6983267 Forward Sequence	5' GTCCTTTGAGCTCAGCAGATGAAAG**[G/T]**CACTGAGAAAAGTACAAAGAATTTT
rs7014346 Forward Sequence	5' TCAAGATGGCTTCTGGAGTGCTACC**[A/G]**TTACATCCATGTTGTAGGCTAGAAG
rs4779584 Reverse Sequence	5' AGAACTTGTTGATAAGCCATTCTTC**[C/T]**GAACAGAAACCATAACTATACACAC
rs10795668 Reverse Sequence	5' AGAAAGAGAAAAAGTTAGATTCTTA**[A/G]**ATTCCATGATTTTATATTTCCCACC

2.5.2 Initial run of RT-PCR

All Assays-by-Demand reactions work under the same conditions and a test reaction of each SNP was performed prior to sample analysis to confirm this. Each reaction contained: 1ηg DNA, 0.125μL Assay (Applied Biosystems) and 2.5μL TaqMan Universal PCR master mix (Applied Biosystems). The reaction was made up to a final volume of 5μL using MilliQ water (Millipore North Ryde, Australia). The PCR reaction was conducted using the ABI PRISM® 7900HT sequencing detection system set to the following conditions: 50ºC for 2 minutes, 95ºC for 10 minutes and 40 cycles of 92ºC for 15 seconds and 60ºC for 1 minute. After the PCR amplification, the plate was scanned to detect fluorescence in each well to generate a graph. The graph is converted to a scatterplot which displays wild type reporter fluorescence versus mutant reporter fluorescence. Low fluorescence was detected during optimisation of some SNPs and this was adjusted by increasing the number of cycles from 40 to 60. For each SNP the volume of the reaction was set to 5μL.

2.5.3 Qualitative analysis of SNPs

After successful optimisation of reaction conditions the samples were genotyped in 96-well optical reaction plates (Applied Biosystems and Axygen Scientific (Union City, CA, USA)). The reagents and volumes used for each assay are shown in table 2.3. Each plate contained three controls: homozygous wild type, heterozygous and homozygous variant. Each plate also contained two no-template controls where sterile water was used in place of DNA (negative control). In addition there were three controls with known genotype to make sure that the run worked properly each time. The remaining wells contained samples with unknown genotype. If the DNA did not produce any fluorescence, genotyping was performed a second time. If the reaction did not work after a second attempt, the DNA sample was taken out of the study for the particular SNP. Therefore different sample numbers can be seen between the same groups for different SNPs. ABI PRISM® 7900 HT sequencing detection system (Applied Biosystems) was used for thermal cycling and the allelic discrimination (AD) for each plate.

Table 2.3: Reaction components of a genotyping reaction. Listed below are the volumes of each reagent in one reaction.

Reagents	Volume in 1 reaction
TaqMan Universal PCR Master Mix	2,5µL
Assay-by-design (40X) Primers and Probes	0.125µL
MilliQ water	1.375µL
DNA (50ηg/µL)	1µL
Reaction Volume per Sample	5µL

2.6 Statistical Analysis

Statistical analysis was performed to determine whether particular single nucleotide polymorphisms are associated with specific types of disease expression or age of survival in HNPCC patients. The genotype frequency of all polymorphisms studied were analysed using the statistical software package SPSS Graduate Pack Version 12.0 (SPSS Incorporated, Chicago, IL USA). The chi-squared test was used to see if differences within the subgroups described in section 2.2.4 are statistically significant, while the Kaplan Meier survival analysis was used to compare genotype and age of diagnosis of CRC in HNPCC patients. The significance level was set at $p<0.05$ for all

tests performed and the confidence interval (CI) was set to 95%. This test measures the probability value (p value), which is the likelihood that a certain range data is found within 95% of the population. Odds ratio was calculated for significantly different results (for 2x2 tables).

2.6.1 Determining allele frequency distribution

After determining the genotype of the nine polymorphisms in each sample, the genotype frequency was examined in subgroup I, II and III, described in section 2.3.4. For the three groups, the genotype frequency was assessed in all HNPCC individuals in both the Australian and Polish population separately. Genotypes from the Australian and Polish populations were then combined to perform the exact same analysis. Pearson's Chi-squared test was used to assess the distribution of the genotypes if the genotype frequency was >5 and Fisher's exact test was used for <5. Three different genotypes were obtained for each SNP examined. In addition to compare the three different genotypes to one another, a combination of two genotypes were also performed. Therefore, for each analysis performed three different results (p-values) can be obtained. The three genotypes homozygous wild type (AA), heterozygous (Aa) and homozygous mutant (aa) was grouped as followed:

- 1. analysis: AA versus Aa versus aa
- 2. analysis: Aa + aa versus AA
- 3. analysis: Aa + AA versus aa

2.6.2 Genotype and colorectal cancer risk

The genotype frequency of individuals affected with colorectal cancer (CRC) was compared to individuals unaffected with CRC to see if the genotype frequency varied between the groups.

2.6.3 Genotype and MMR gene

HNPCC individuals were subdivided into groups depending on MMR gene mutation, harbouring a mutation in hMLH1, hMSH2 or hMSH6. In the hMSH6 group there were only 50 patients and therefore not enough statistical power to analyse this group to hMLH1 and hMSH2 mutation carriers. The genotype frequency of hMLH1 mutation carriers was compared to hMSH2 mutation carriers to see if there were any differences between patients with mutations in different MMR genes.

2.6.4 Genotype and gender

Male HNPCC patients were compared to female HNPCC patients to observe if there were any differences in genotype frequency between genders.

2.6.5 Genotype and gynaecological cancer risk

Female HNPCC patients were subdivided into groups depending on whether they were unaffected or affected with either endometrial or ovarian cancer. Females affected with endometrial or ovarian cancer were placed into the same group called gynaecological cancer.

2.6.6 Kaplan- Meier survival analysis

Kaplan–Meier survival curves were used to plot the participants (cancer or not) versus the patient age of diagnosis of CRC/age of unaffected in relation to the genotype for each SNP examined. The Kaplan–Meier survival curves for each genotype were tested to see if there were differences in age of diagnosis of CRC by genotype. Comparison between the three different genotypes obtained from each SNP was performed by the Log Rank test. Two other non-parametric linear rank tests were also added to the analysis, the Breslow and Tarone-Ware tests. All tests were used to assess the equality of the survivor function across the three genotypes. The Breslow (generalized Wilcoxon) test was used to determine the significance of observation from early ages of diagnosis while the Log Rank test give more weight to later ages of diagnosis. The Tarone-Ware test, which is an intermediate of the two other tests, was also used to examine the homogeneity of the survival curve. Age of diagnosis was defined as patient age at the time of CRC diagnosis. For unaffected participant, age was based on the date of birth and disease free status at last consultation. For polymorphisms that showed a statistically significant difference between the genotypes and the age of survival, the odds ratio (OD) was calculated using unconditional logistic regression.

3 RESULTS

The analysis of the relationship between single nucleotide polymorphisms (SNPs) and their association with disease development were undertaken with the following nine SNPs: rs4939827, rs4464148, rs6983267, rs16892766, rs10795668, rs3802842, rs7014346, rs4779584 and rs10318. All HNPCC samples were subdivided into the following groups and compared to each other:

- Affected with CRC vs Unaffected with CRC
- Affected with hMLH1 vs affected with hMSH2
- Female vs male
- Females affected with gynaecological cancer vs females unaffected with gynaecological cancer

3.1 Genotype frequencies

The genotype frequency distribution for every SNP was compared to determine if any significant difference existed between the various groups. We first analysed the Australian and Polish population separately (See appendix 3 and 4 for tables and figures). As there was no major significant difference in genotype frequencies between the Australian and Polish HNPCC participants, this allowed for pooling of the genotype results from the two populations. If the SNPs examined in this study are true modifiers, they are likely to influence the chance of developing disease HNPCC populations regardless of which country the samples are from. Therefore we combined the Australian and Polish data which presented us with excellent statistical power. The genotypes for the nine SNPs were determined by RT-PCR.

3.2 Genotype distribution in the Australian HNPCC population

The distribution of allele frequency of SNP rs4464148, rs7014346 and rs10795668 within the different groups did not reveal any statistically association. See appendix 3, tables 5, 7 and 9.

The distribution of the genotypes for SNP rs16892766 among individuals affected with CRC differed from individuals unaffected with CRC (p=0.032), see appendix 3, table 1. No samples in the Australian population were homozygous variant for the rs16892766 polymorphism. A significant difference (p=0.044) in genotype frequency can be see for SNP rs10318 among individuals affected with CRC versus individuals unaffected with CRC when the homozygous variant genotype (TT) was compared to combination of heterozygous and homozygous wild type genotype (CT+CC), see appendix 3, table 3.

The genotype frequency for SNP rs6983267 revealed a significant difference between individuals that harboured a mutation in the hMLH1 gene compared to individuals that harboured a mutation in the hMSH2 gene (p=0.012). A significantly different result (p=0.004) can also be seen for the same SNP when homozygous variant genotype (GG) was compared to heterozygous and wild type genotypes (TG+TT), see appendix 3, table 6.

The genotype frequency differed significantly among females and males for three SNPs: rs4939827 (p=0.031), rs6983267 (p=0.030) and rs4779584 (p=0.043), see appendix 3, table 4, 6 and 8 respectively. For SNP rs4939827 and rs6983267 the difference can still be seen when the combination of heterozygous and homozygous variant genotypes are compared to homozygous wild type, p=0.010 and p=0.009 respectively.

SNP rs3802842 revealed a statistically different result (p=0.017) in genotype distribution in females affected with gynaecological cancer versus females not affected with the disease when the homozygous variant genotype (CC) was compared to combination of heterozygous and wild type genotypes (AC+AA), see appendix 3, table 2.

3.3 Kaplan-Meier survival analysis in the Australian HNPCC population

The Kaplan-Meier survival analysis estimated by rs16892766 (SNP1) revealed a significant difference in the age of diagnosis of CRC between homozygous wild type genotypes (AA, 52-years) compared to heterozygous genotype (AC, 44-years), log-rank test: p=0.004, Breslow test: p=0.014 and Tarone-Ware test: p=0.006. SNP1 also showed a significantly different result between homozygous wild type genotype (AA-52 years) compared to heterozygous genotype (AC-41 years) when comparing age of diagnosis of CRC in hMLH1 mutation carriers (log-rank test: p=0.001, Breslow test: p=0.001 and Tarone-Ware test: p=0.0008) but not in hMLH2 mutation carriers, see appendix 3, figure 1 and 2.

A significantly difference can also be seen for rs3802842 (SNP2) in hMLH1 mutation carriers (log-rank test: p=0.003, Breslow test: p=0.013 and Tarone-Ware test: p=0.006), see figure 3.3 (AA-54 years, AC-50 years and CC-42 years).

The Kaplan-Meier curves for rs4779584 (SNP8) show a significant result (log-rank test: p=0.021) when the wild type genotype (CC-50 years) was compared to the combination of heterozygous and variant genotypes (CT+TT-62 years). This finding can also be observed in hMSH2 mutation carriers; homozygous wild type genotype (CC, 49 years) compared to combination of heterozygous and variant genotype (CC+TT-64 years, see appendix 3 figure 5 and 6).

Kaplan-Meier survival analysis estimated by rs7014346 (SNP7) genotype in hMSH2 mutation carriers revealed a statistical significant difference between the genotypes (GG-45 years, GA-59 years and AA-56 years) and the age of diagnosis of CRC for the log-rank test: p=0.031 and when the wild type genotype (GG-45 years) was compared to combination of heterozygous and variant genotypes (GA+AA-59 years), see appendix 3, figure 3 and 4.

3.4 Genotype distribution in Polish HNPCC population

The distribution of genotype frequency of SNP rs3802842 and rs4779584 within the different groups did not reveal any statistically associations, see appendix 4, tables 2 and 8.

The distribution of allele frequency for SNP rs16892766 among individuals affected with CRC differed from individuals unaffected with CRC (p=0.018) when the homozygous wild type (AA) was compared to combination of heterozygous (AC) and homozygous variant genotype (CC). See appendix 4, table 1.

The genotype frequency differed among hMLH1 and hMSH2 mutation carriers for SNP rs49398727 (p=0.014) and rs7014346 (p=0.039), see appendix 3, table 4 and 7 respectively. For SNP rs7014346 the homozygous wild type (GG) compared to the combination to combination of heterozygous (AC) and homozygous variant genotype (CC) was significantly (p=0.039) different, see appendix 4, table 1.

The genotype frequency distribution of the rs10318 SNP differed among females and males, p=0.016. The difference can still be observed when homozygous wild type (CC) was compared to combination of heterozygous and homozygous variant genotypes (CT+TT) (p=0.012), see appendix 4, table 3.

The genotype frequency of the rs10795668 polymorphism revealed a significant different result (p=0.036) among females and males when the homozygous variant genotype (AA) was compared to combination heterozygous and wild type genotypes (GA+GG). See appendix 4, table 9.

The distribution of genotype frequency for SNP rs6983267 within the group of females affected with gynaecological cancer differed from females not affected with the disease revealed statistically significant results p=0.039. A significantly different result (p=0.034) can also be observed when examining wild type (TT) compared to combination of heterozygous and homozygous variant genotypes (TG+GG), and for homozygous variant (GG) genotype compared to combination of heterozygous and wild type genotypes (TG+TT), p=0.039, see appendix 4, table 6.

3.5 Kaplan-Meier survival analysis in the Polish HNPCC

Kaplan-Meier survival analysis estimated by rs16892766 (SNP1) revealed a statistical significant difference between the genotypes (AA-53 years, the two other genotypes did not reach 50 %) and age of diagnosis of CRC (log-rank test: p=0.031) when age of diagnosis of CRC was examined, see appendix 4, figure 1. This is also true for hMLH1 mutation carriers who showed a statistical significant difference between the genotypes and age of diagnosis of CRC (log-rank test: p=0.023), see appendix 4, figure 2.

A significantly different result can also be obtained for SNP2 rs3802842 between the genotypes (AA-53 years, AC-60 yeras and CC-44 years) and age of diagnosis of CRC (log-rank test: p=0.031, Breslow test: p=0.034 and the Tarone-Ware tests: p=0.030), see figure 3.4.

3.6 Genotype distribution in the Australian and Polish HNPCC population combined

The combined genotype distribution can be seen in table 3.1-3.9. When combining the Australian and Polish HNPCC population only one significant result can be observed when genotype frequency between all the different subgroups are compared. The rs4779584 SNP revealed a significant different result (p=0.038) in genotype distribution among females and males, see table 3.8. None of the other polymorphism investigated revealed any statistically association of allele frequency distribution (See table 3.1-3.7 and table 3.9).

3.6 Combined Kaplan-Meier survival analysis

Kaplan-Meier survival analysis estimated by rs3802842 (SNP2) revealed a highly statistical significant difference between the genotypes (AA-54 years, AC-53 years and CC-42 years) and age of diagnosis of CRC (log-rank test: p=0.0002, Breslow test: p=0.0005 and Tarone-Ware test: p=0.0003) when it comes to age of diagnosis of CRC, see figure 3.1. When homozygous variant genotype (CC-42 years) was compared to combination of heterozygous and wild type genotypes (AC+AA,-53

years), a significantly result was also observed. Log-rank test: p=0.0001, Breslow test: p=0.0001 and the Tarone-Ware tests: p=0.0001), see figure 3.2.

TABLE 3.1- ALLELE FREQUNCY DISTRIBUTION OF THE rs16892766 POLYMORPHISM IN THE AUSTRALIAN AND POLISH POPULATION COMBINED

SNP 1 Rs16892766	AA (%)	AC (%)	CC (%)	p-value[1]	AC+CC (%)	p-value[2]	AC+AA (%)	p-value[3]
Allele frequency								
Subject group (n=676)								
CRC+ (n=285)	241 (85)	44 (15)	0	p=0.69	44 (15)	p=0.95	285 (100)	p=1.0*
CRC- (n=391)	330 (84.5)	60 (15)	1 (0.5)		61 (15.5)		389 (99.5)	
Subject group (n=630)								
hMLH1 (n=351)	299 (85)	52 (15)	0	p=0.53	52 (15)	p=0.97	350 (99.5)	p=0.44*
hMSH2 (n=279)	238 (85)	40 (14.5)	1 (0.5)		41 (15)		278 (99.5)	
Subject group (n=679)								
Female (n=407)	351 (86)	56 (14)	0	p=0.18	56 (14)	p=0.13	406 (100)	p=0.40*
Male (n=272)	223 (82)	48 (17.5)	1 (0.5)		49 (17.5)		271 (99.5)	
Subject group (n=406)								
Gynaecological+ (n=80)	67 (84)	13 (16)	0	p=0.48	13 (16)	p=0.48	80 (100)	
Gynaecological- (n=326)	283 (87)	43 (13)	0		43 (13)		325 (100)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (AA) compared to combination of heterozygous and homozygous variant genotypes (AC+CC)
[3]p-value: Homozygous variant genotype (CC) compared to combination of wild type and heterozygous genotypes (AC+AA)
* p-value was calculated using Fisher's exact test

TABLE 3.2- ALLELE FREQUNCY DISTRIBUTION OF THE rs3802842 POLYMORPHISM IN THE AUSTRALIAN AND POLISH POPULATION COMBINED

SNP2 Rs3802842	AA (%)	AC (%)	CC (%)	p-value[1]	AC+CC (%)	p-value[2]	AC+AA (%)	p-value[3]
Allele frequency								
Subject group (n=670)								
CRC+ (n=283)	156 (55)	102 (36)	25 (9)	p=0.87	127 (45)	p=0.96	258 (91)	p=0.62
CRC- (n=387)	214 (55)	143 (37)	30 (8)		173 (45)		357 (92)	
Subject group (n=624)								
hMLH1 (n=345)	186 (54)	133 (39)	26 (7)	p=0.57	159 (46)	p=0.56	319 (93)	p=0.52
hMSH2 (n=279)	157 (56)	97 (35)	25 (9)		122 (44)		254 (91)	
Subject group (n=673)								
Female (n=402)	218 (54)	148 (37)	36 (9)	p=0.62	184 (46)	p=0.51	366 (91)	p=0.37
Male (n=271)	154 (57)	98 (36)	19 (7)		117 (43)		252 (93)	
Subject group (n=401)								
Gynaecological+ (n=79)	47 (60)	27 (34)	5 (6)	p=0.49	32 (40)	p=0.31	74 (94)	p=0.35
Gynaecological- (n=322)	171 (53)	120 (37)	31 (10)		151 (47)		291 (90)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (AA) compared to combination of heterozygous and homozygous variant genotypes (AC+CC)
[3]p-value: Homozygous variant genotype (CC) compared to combination of heterozygous and wild type genotypes (AC+AA)

TABLE 3.3- ALLELE FREQUNCY DISTRIBUTION OF THE rs10318 POLYMORPHISM IN THE AUSTRALIAN AND POLISH POPULATION COMBINED

SNP 3 Rs10318	CC (%)	CT (%)	TT (%)	p-value[1]	CT+TT (%)	p-value[2]	CT+CC (%)	p-value[3]
Allele frequency								
Subject group (n=662)								
CRC+ (n=278)	184 (66)	83 (30)	11 (4)	p=0.95	94 (34)	p=0.99	267 (96)	p=0.77
CRC- (n=384)	254 (66)	113 (30)	17 (4)		130 (34)		367 (96)	
Subject group (n=616)								
hMLH1 (n=340)	222 (65)	107 (32)	11 (3)	p=0.17	118 (35)	p=0.87	329 (97)	p=0.083
hMSH2 (n=276)	182 (66)	77 (28)	17 (6)		94 (34)		259 (94)	
Subject group (n=665)								
Female (n=398)	256 (64)	129 (33)	13 (3)	p=0.056	142 (36)	p=0.26	385 (97)	p=0.092
Male (n=267)	183 (69)	68 (25)	16 (6)		84 (31)		251 (94)	
Subject group (n=397)								
Gynaecological+ (n=78)	50 (64)	25 (32)	3 (4)	p=0.95	28 (36)	p=0.98	75 (96)	p=0.725*
Gynaecological- (n=319)	206 (65)	103 (32)	10 (3)		113 (35)		309 (97)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (CC) compared to combination of heterozygous and homozygous variant genotypes (CT+TT)
[3]p-value: Homozygous variant genotype (TT) compared to combination heterozygous and wild type genotypes (CT+CC)
*p-value was calculated using Fisher's exact test

TABLE 3.4- ALLELE FREQUNCY DISTRIBUTION OF THE rs4939827 POLYMORPHISM IN THE AUSTRALIAN AND POLISH POPULATION COMBINED

SNP 4 Rs4939827	CC (%)	CT (%)	TT (%)	p-value[1]	CT+TT (%)	p-value[2]	CT+CC (%)	p-value[3]
Allele frequency								
Subject group (n=674)								
CRC+ (n=285)	87 (30)	145 (51)	53 (19)	p=0.68	198 (70)	p=0.39	232 (81)	p=0.64
CRC- (n=389)	107 (28)	204 (52)	78 (20)		282 (72)		311 (80)	
Subject group (n=628)								
hMLH1 (n=349)	105 (30)	171 (49)	73 (21)	p=0.30	244 (70)	p=0.28	276 (79)	p=0.48
hMSH2 (n=279)	73 (26)	154 (55)	52 (19)		206 (74)		227 (81)	
Subject group (n=677)								
Female (n=406)	110 (27)	217 (53)	79 (20)	p=0.42	296 (73)	p=0.19	327 (80)	p=0.84
Male (n=271)	86 (32)	134 (49)	51 (19)		185 (68)		220 (81)	
Subject group (n=405)								
Gynaecological+ (n=80)	27 (34)	38 (47)	15 (19)	p=0.30	53 (66)	p=0.12	65 (81)	p=0.75
Gynaecological- (n=325)	82 (25)	177 (55)	66 (20)		243 (75)		259 (80)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (CC) compared to combination of heterozygous and homozygous variant genotypes (CT+TT)
[3]p-value: Homozygous variant genotype (TT) compared to combination heterozygous and wild type genotypes (CT+CC)

TABLE 3.5- ALLELE FREQUNCY DISTRIBUTION OF THE rs4464148 POLYMORPHISM IN THE AUSTRALIAN AND POLISH POPULATION COMBINED

SNP 5 Rs4464148	TT (%)	TC (%)	CC (%)	p-value[1]	TC+CC (%)	p-value[2]	TC+TT (%)	p-value[3]
Allele frequency								
Subject group (n=678)								
CRC+ (n=286)	131 (46)	117 (41)	38 (13)	p=0.23	155 (54)	p=0.25	248 (87)	p=0.42
CRC- (n=392)	162 (41)	186 (48)	44 (11)		230 (59)		348 (89)	
Subject group (n=632)								
hMLH1 (n=353)	146 (41)	156 (44)	51 (15)	p=0.45	207 (59)	p=0.88	302 (85)	p=0.22
hMSH2 (n=279)	117 (42)	131 (47)	31 (11)		162 (58)		248 (89)	
Subject group (n=681)								
Female (n=409)	173 (42)	191 (47)	45 (11)	p=0.22	236 (58)	p=0.51	364 (89)	p=0.195
Male (n=272)	122 (45)	111 (41)	39 (14)		150 (55)		233 (86)	
Subject group (n=408)								
Gynaecological+ (n=80)	37 (46)	37 (46)	6 (8)	p=0.50	43 (54)	p=0.41	74 (92)	p=0.29
Gynaecological- (n=328)	135 (41)	155 (47)	38 (12)		193 (59)		290 (88)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (TT) compared to combination of heterozygous and homozygous variant genotypes (TC+CC)
[3]p-value: Homozygous variant genotype (CC) compared to combination heterozygous and wild type genotypes (TC+TT)

TABLE 3.6- ALLELE FREQUNCY DISTRIBUTION OF THE rs6983267 POLYMORPHISM IN THE AUSTRALIAN AND POLISH POPULATION COMBINED

SNP 6 Rs6983267	TT (%)	TG (%)	GG (%)	p-value[1]	TG+GG (%)	p-value[2]	TG+TT (%)	p-value[3]
Allele frequency								
Subject group (n=667)								
CRC+ (n=282)	60 (21)	151 (54)	71 (25)	p=0.60	222 (79)	p=0.52	211 (75)	p=0.59
CRC- (n=385)	90 (23)	191 (50)	104 (27)		295 (77)		281 (73)	
Subject group (n=621)								
hMLH1 (n=344)	77 (22)	185 (54)	82 (24)	p=0.27	267 (78)	p=0.59	262 (76)	p=0.11
hMSH2 (n=277)	57 (20)	138 (50)	82 (30)		220 (80)		195 (70)	
Subject group (n=670)								
Female (n=403)	99 (25)	195 (48)	109 (27)	p=0.11	304 (75)	p=0.097	294 (73)	p=0.50
Male (n=267)	51 (19)	150 (56)	66 (25)		216 (81)		201 (75)	
Subject group (n=402)								
Gynaecological+ (n=79)	15 (19)	36 (46)	28 (35)	p=0.14	64 (81)	p=0.19	51 (65)	p=0.063
Gynaecological- (n=323)	84 (26)	158 (49)	81 (25)		239 (74)		242 (75)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (TT) compared to combination of heterozygous and homozygous variant genotypes (TG+GG)
[3]p-value: Homozygous variant genotype (GG) compared to combination heterozygous and wild type genotypes (TG+TT)

TABLE 3.7- ALLELE FREQUNCY DISTRIBUTION OF THE rs7014346 POLYMORPHISM IN THE AUSTRALIAN AND POLISH POPULATION COMBINED

SNP 7 Rs7014346	GG (%)	GA (%)	AA (%)	p-value[1]	GA+AA (%)	p-value[2]	GA+GG (%)	p-value[3]
Allele frequency								
Subject group (n=671)								
CRC+ (n=282)	97 (34)	132 (47)	53 (19)	p=0.91	185 (66)	p=0.67	229 (81)	p=0.79
CRC- (n=389)	140 (36)	179 (46)	70 (18)		249 (64)		319 (82)	
Subject group (n=625)								
hMLH1 (n=347)	118 (34)	167 (48)	62 (18)	p=0.87	229 (66)	p=0.74	285 (82)	p=0.79
hMSH2 (n=278)	98 (35)	128 (46)	52 (19)		180 (65)		226 (81)	
Subject group (n=674)								
Female (n=405)	146 (36)	184 (45)	75 (19)	p=0.76	259 (64)	p=0.55	330 (81)	p=0.82
Male (n=269)	91 (34)	130 (48)	48 (18)		178 (66)		221 (82)	
Subject group (n=404)								
Gynaecological+ (n=79)	28 (35)	33 (42)	18 (23)	p=0.54	51 (65)	p=0.89	61 (77)	p=0.28
Gynaecological- (n=325)	118 (36)	150 (46)	57 (18)		207 (64)		268 (82)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (GG) compared to combination of heterozygous and homozygous variant genotypes (GA+AA)
[3]p-value: Homozygous variant genotype (AA) compared to combination heterozygous and wild type genotypes (GA+GG)

TABLE 3.8- ALLELE FREQUNCY DISTRIBUTION OF THE rs4779584 POLYMORPHISM IN THE AUSTRALIAN AND POLISH POPULATION COMBINED

SNP 8 Rs4779584	CC (%)	CT (%)	TT (%)	p-value[1]	CT+TT (%)	p-value[2]	CT+CC (%)	p-value[3]
Allele frequency								
Subject group (n=666)								
CRC+ (n=280)	173 (62)	91 (32)	16 (6)	p=0.68	107 (38)	p=0.48	264 (94)	p=0.76
CRC- (n=386)	228 (59)	138 (36)	20 (5)		158 (41)		366 (95)	
Subject group (n=620)								
hMLH1 (n=342)	209 (61)	118 (35)	15 (4)	p=0.18	133 (39)	p=0.21	327 (96)	p=0.093
hMSH2 (n=278)	156 (56)	101 (36)	21 (8)		122 (44)		257 (92)	
Subject group (n=670)								
Female (n=402)	232 (58)	152 (38)	18 (4)	p=0.038	170 (42)	p=0.11	384 (96)	p=0.15
Male (n=268)	171 (64)	78 (29)	19 (7)		97 (36)		249 (93)	
Subject group (n=400)								
Gynaecological+ (n=80)	44 (55)	32 (40)	4 (5)	p=0.85	36 (45)	p=0.58	76 (95)	p=0.766*
Gynaecological- (n=320)	187 (59)	119 (37)	14 (4)		133 (41)		306 (96)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (CC) compared to combination of heterozygous and homozygous variant genotypes (CT+TT)
[3]p-value: Homozygous variant genotype (TT) compared to combination heterozygous and wild type genotypes (CT+CC)
*p-value was calculated using Fisher's exact test

TABLE 3.9- ALLELE FREQUNCY DISTRIBUTION OF THE rs10795668 POLYMORPHISM IN THE AUSTRALIAN AND POLISH POPULATION COMBINED

SNP 9 Rs10795668	GG (%)	GA (%)	AA (%)	p-value[1]	GA+AA (%)	p-value[2]	GA+GG (%)	p-value[3]
Allele frequency								
Subject group (n=658)								
CRC+ (n=280)	128 (46)	121 (43)	31 (11)	p=0.61	152 (54)	p=0.38	249 (89)	p=0.84
CRC- (n=378)	160 (42)	178 (47)	40 (11)		218 (58)		338 (89)	
Subject group (n=612)								
hMLH1 (n=334)	142 (43)	152 (45)	40 (12)	p=0.20	192 (57)	p=0.25	294 (88)	p=0.10
hMSH2 (n=278)	131 (47)	125 (45)	22 (8)		147 (53)		256 (92)	
Subject group (n=661)								
Female (n=396)	174 (44)	173 (44)	49 (12)	p=0.23	223 (56)	p=0.91	347 (88)	p=0.10
Male (n=265)	117 (44)	126 (48)	22 (8)		148 (56)		243 (92)	
Subject group (n=395)								
Gynaecological+ (n=79)	33 (42)	37 (47)	9 (11)	p=0.83	46 (58)	p=0.70	70 (89)	p=0.85
Gynaecological- (n=316)	140 (44)	136 (43)	40 (13)		176 (56)		276 (87)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (GG) compared to combination of heterozygous and homozygous variant genotypes (GA+AA)
[3]p-value: Homozygous variant genotype (AA) compared to combination heterozygous and wild type genotypes (GA+GG)

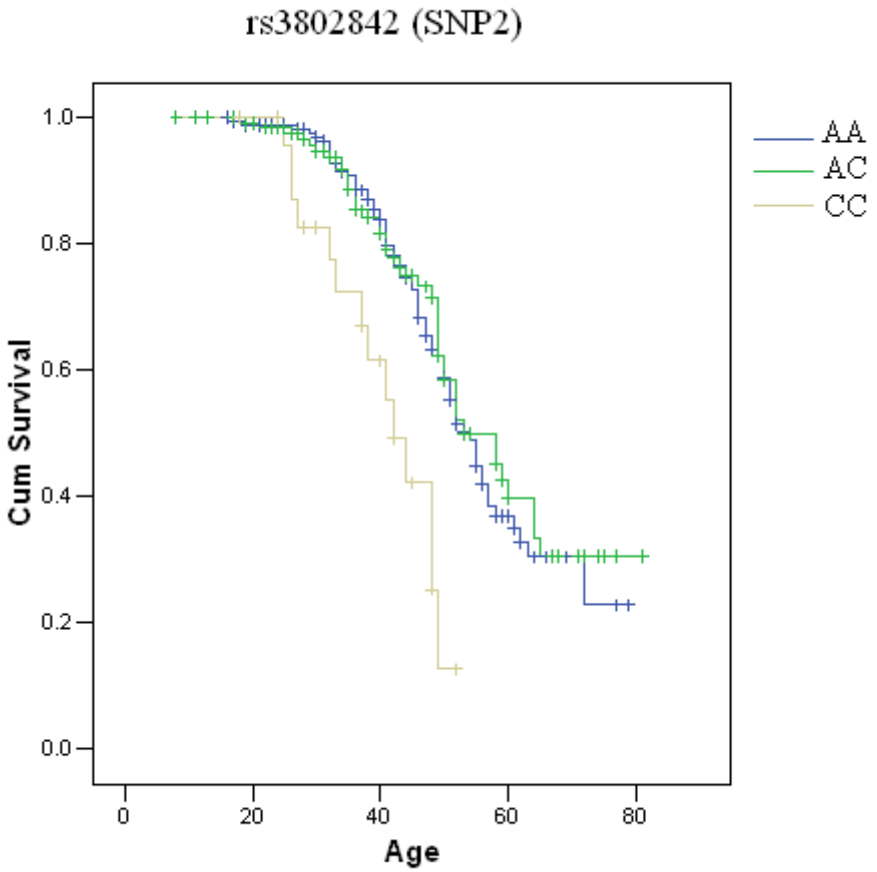


Figure 3.1: Kaplan-Meier estimated by rs3802842 (SNP2) genotype in hMLH1 mutation carriers-Australian+Polish combined. The graph shows the effect the three genotypes (AA vs AC vs CC) have on age of diagnosis of CRC in the Australian and Polish HNPCC patients. There is a statistical significant difference between the genotypes and the age of diagnosis of CRC (log-rank test: p=0.0002, Breslow test: p=0.0005 and Tarone-Ware test: p=0.0003)

Table 3.10: Median age of diagnosis of CRC (age at which 50% of the population is cancer free) in HNPCC participants for SNP rs3802842

Genotype	Subject group SNP rs3802842 [1]
Wild type (AA)	54 yrs (n=176)
Heterozygote (AC)	53 yrs (n=124)
Variant (CC)	42 yrs (n=25)

[1] The subject group includes 325 samples

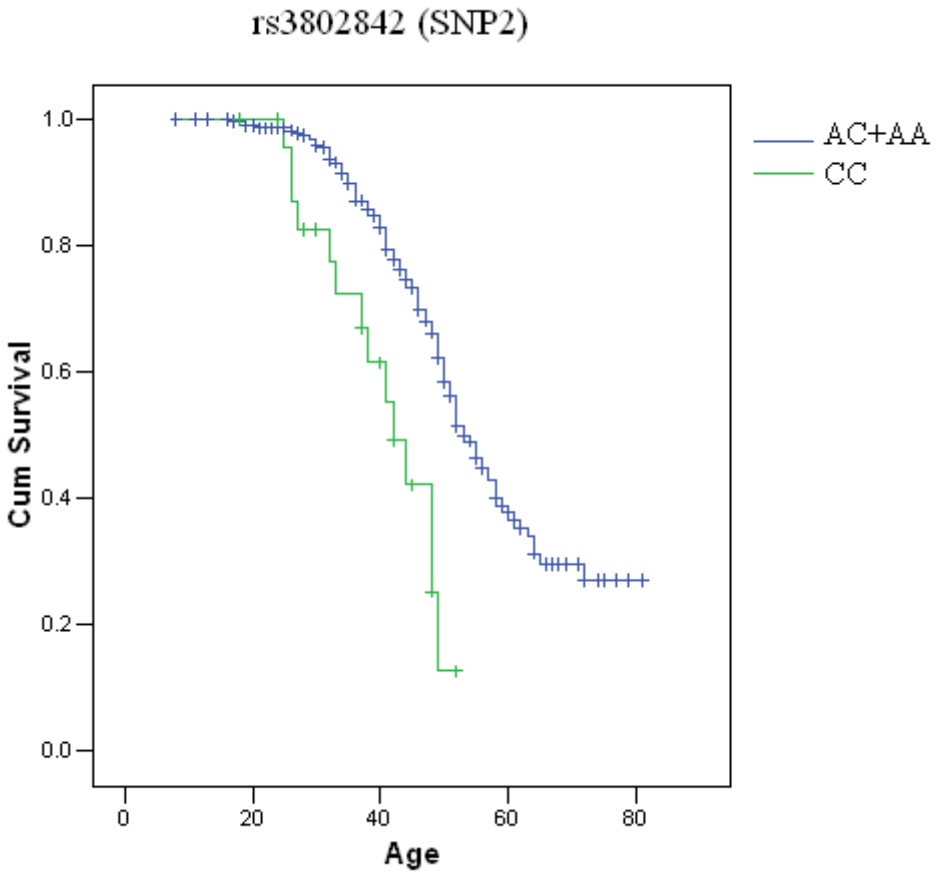


Figure 3.2: Kaplan-Meier estimated by rs3802842 (SNP2) genotype in hMLH1 mutation carriers-Australian+Polish combined. The graph shows the effect of homozygous variant genotype (CC) compared to combination of heterozygous and wild type genotypes (AC+AA) have on age of age of diagnosis of CRC in Australian and Polish HNPCC patients. There is a statistical significant difference between the variant genotype compared to combination of heterozygous and wild type and the age of diagnosis of CRC (log-rank test: p=0.0001, Breslow test: p=0.0001 and Tarone-Ware test: p=0.0001).

Table 3.11: Median age of diagnosis of CRC (age at which 50% of the population is cancer free) in HNPCC participants for SNP rs3802842

Genotype	Subject group SNP rs3802842
Variant (CC)	42 yrs (n=299)
Heterozygote (AC) + wild type (AA)	53 yrs (n=25)

[1]The subject group includes 324 samples

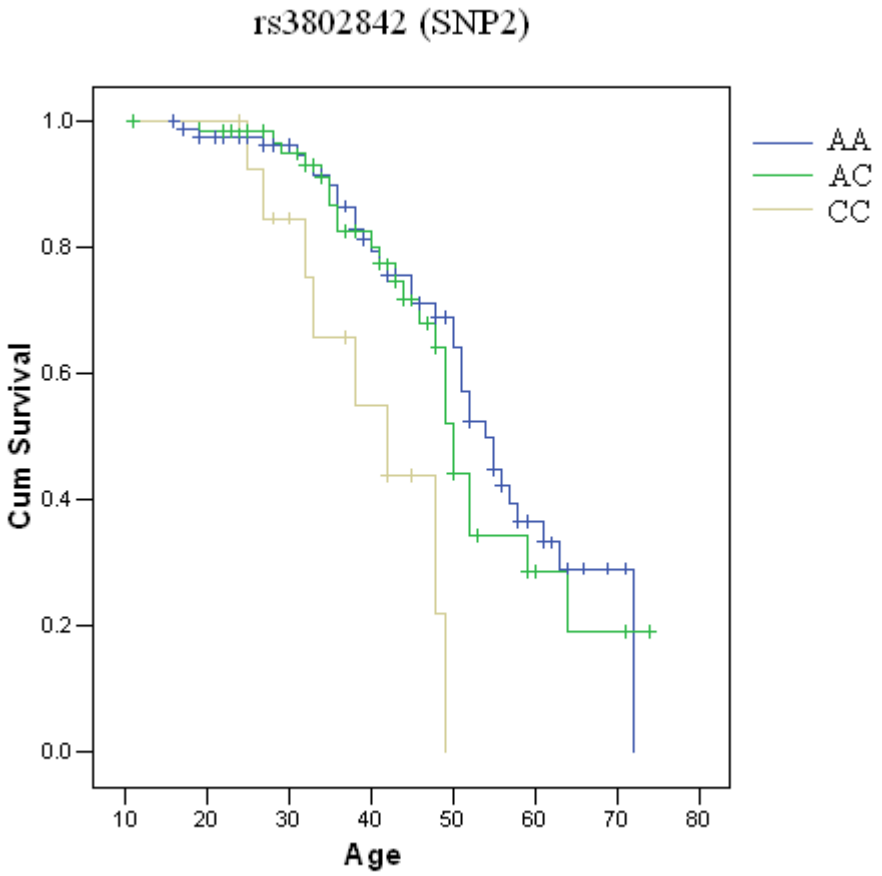


Figure 3.3: Kaplan-Meier estimated by rs3802842 (SNP2) genotype in hMLH1 mutation carriers-Australian only. The graph shows the effect the three genotypes (AA vs AC vs CC) have on age of age of diagnosis of CRC in Australian HNPCC patients. There is a statistical significant difference between the genotypes and the age of diagnosis of CRC (log-rank test: p=0.003, Breslow test: p=0.013 and Tarone-Ware test: p=0.006).

Table 3.12: Median age of diagnosis of CRC (age at which 50% of the population is cancer free) in HNPCC participants for SNP rs3802842

Genotype	Subject group SNP rs3802842 [1]
Wild type (AA)	42 yrs (n=14)
Heterozygote (AC)	50 yrs (n=64)
Variant (CC)	54 yrs (n=85)

[1] The subject group includes 163 samples

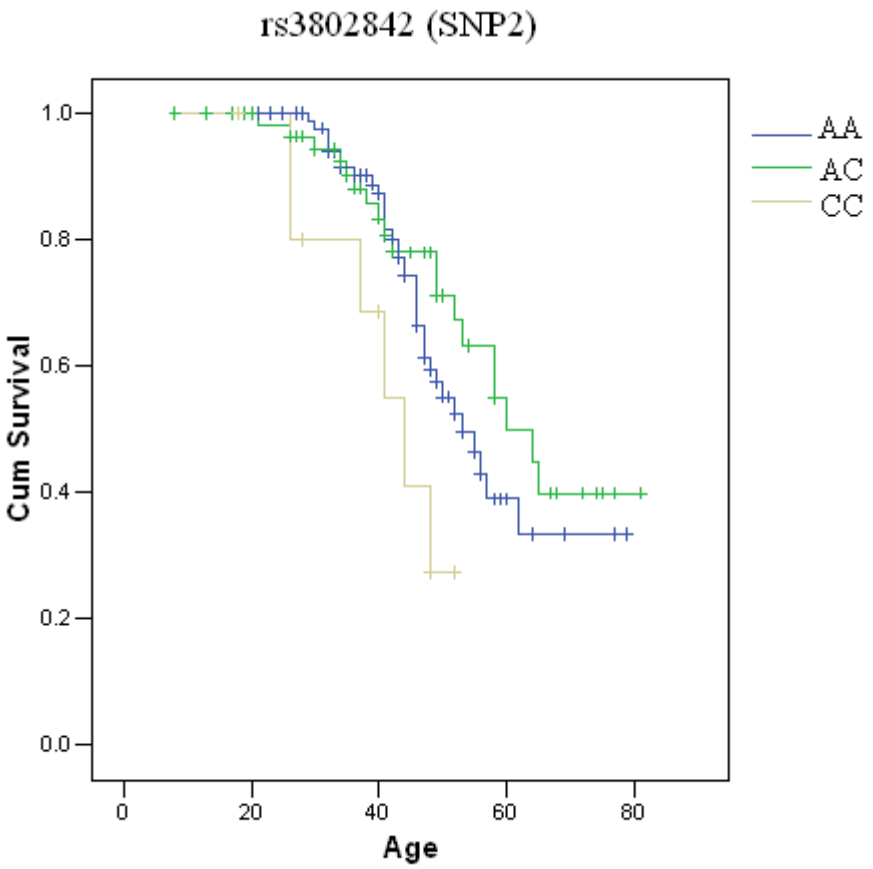


Figure 3.4: Kaplan-Meier estimated by rs3802842 (SNP2) genotype in hMLH1 mutation carriers-Polish only. The graph shows the effect the three genotypes (AA vs AC vs CC) have on age of age of diagnosis of CRC in Polish HNPCC patients. There is a statistical significant difference between the genotypes and the age of diagnosis of CRC (log-rank test: p=0.031, Breslow test: p=0.034 and the Tarone-Ware tests: p=0.030).

Table 3.13: Median age of diagnosis of CRC (age at which 50% of the population is cancer free) in HNPCC participants for SNP rs3802842

Genotype	Subject group SNP rs3802842 [1]
Wild type (AA)	53 yrs (n=91)
Heterozygote (AC)	60 yrs (n=60)
Variant (CC)	44 yrs (n=11)

4 DISCUSSION

The foundation for the great diversity observed in the human phenotype is based on genetic variation and accounts for the large variety of susceptibilities to common diseases [80]. Cancer is a result of the combination of germline susceptibility and somatic mutations, which over time accumulate resulting in an imbalance between cell proliferation and apoptosis. Combinations of polymorphisms in genes controlling cancer development are believed to be the cause of the differences seen in disease phenotype [81].

Since the identification of the genetic basis of HNPCC in early 1990's [1, 2] many studies have been undertaken to identify modifier genes that may explain at least some of the variation observed in disease expression in HNPCC patients [44, 48, 50, 52, 56, 58, 82]. Unlike other genetic predispositions to colorectal cancer, such as Familial adenomatous polyposis (FAP), HNPCC does not present with a premalignant phenotype, which as such makes it difficult to predict when or if an affected person will present with disease.

Observation of HNPCC patients who harbour germline mutation in different mismatch repair genes, as well as patients with the same mutation within the same gene do not always express the same disease patterns (age of diagnosis, severity of disease and appearance of extracolonic cancer) [43]. This indicates that other genes, modifier genes, and environmental factors are likely to influence disease expression. Identification of genes that influence disease expression can lead to routine use of molecular tests to diagnose this syndrome and establish interventions to prevent development of cancer [83].

The current study was undertaken to identify modifier genes that might alter the disease expression in patients with HNPCC who are mutation positive for genes (hMLH1, hMSH2 and hMSH6) already associated with the disease. The candidate SNPs were chosen on the basis of previously reported associations between the particular SNPs and development of CRC. In this study, the aim was to investigate if the nine SNPs elected have the ability to alter disease expression in HNPCC.

Both inherited and genetic factors are known to contribute to the development of HNPCC. Recently a genome wide association study has identified a colorectal cancer susceptibility locus on chromosome 11q23 [72]. SNP rs3802842 is located within a gene rich region of chromosome 11q23 and is located close to genes encoding POU transcription factors [84]. The POU family of transcription factor are divided into six classes and the rs3802842 is close to POU class 5 [84]. The regulation of cell function can occur via POU factors alone, in combination with other POU proteins or together with other transcription factors [84]. A change in POU protein levels have been found in several malignancies (for example melanoma) and a study performed by Tenesa et al (2008), showed that locus 11q23 is associated with CRC.

4.1 Age of diagnosis of CRC in hMLH1 mutation carriers

The results from the current study are highly suggestive of a protective effect against CRC development in HNPCC patients carrying a mutation in hMLH1 who harbour the wild type or heterozygous genotype (AA+AC) for SNP rs3802842 compared to those carrying the variant genotype (CC), see figure 3.1. Overall, the same significant result was observed for all three tests performed. Our study shows that hMLH1 mutation carriers harbouring the variant genotype (CC) develop CRC on average 11 years earlier than hMLH1 carriers with the heterozygous or wild type genotype (AC or AA, see table 3.10. This indicates that the rs3802842 polymorphism is likely to act as a modifier of disease expression in HNPCC patients harbouring the hMLH1 mutation. The same significant result was observed when the variant genotype was compared to combination of heterozygous and wild type genotypes, see figure 3.2. HNPCC is a dominantly inherited disease, meaning that an individual affected with the syndrome already has one defective allele of a gene and therefore only require a somatic mutation in the remaining allele to develop HNPCC related diseases. Individuals with the heterozygous genotype have one defective allele. So, when combining the wild type or variant genotype to the heterozygous genotype, we test to see if individuals harbouring that specific polymorphism act dominantly or recessive. In this case, when looking at figure 3.2 one can see that the age of diagnosis of CRC only differs between those individuals harbouring the variant genotype to those individuals harbouring either the wild type or heterozygous genotype. This means that

individuals harbouring SNP rs3802842 require both alleles to be defective to have an affect on the age of diagnosis of CRC.

The same trend was also observed when the Kaplan-Meier analysis was assessed on the Australian and Polish population separately (see figure 3.3 and 3.4) supporting the result found with the combined analysis. If SNP rs3802842 is a true modifier it is likely to influence the likelihood of disease in all populations studied. Figure 3.3 show that Australian hMLH1 mutation carriers with the variant genotype (CC) develop CRC at a median age of 42, while hMLH1 mutation carriers with heterozygous and wild type genotype develop CRC at a median age of 50 and 54 respectively. The major difference in age of diagnosis of CRC is between hMLH1 mutation carriers harbouring variant genotype compared to those individuals harbouring either the wild type or heterozygous genotype. Moreover, this effect was also observed in the Polish population, see figure 2.4. The median age of diagnosis of CRC for hMLH1 mutation carriers with the variant genotype was 44 years as compared to 60 and 53 years for the AA and AC genotype respectively. Therefore, when the populations are analysed separately, the same result is observed for the combined analysis. However, in the Polish population it appears that there could be a dosage effect as the heterozygous genotype (AC) is an intermediate of the other two genotypes (see figure 3.4). In the Australian population the curve for the wild type and heterozygous genotype seem to follow each other more closely. This result is highly significant as it was found in two separate populations and when combining the number of HNPCC mismatch repair positive individuals is one of the largest HNPCC populations in the world.

4.2 Future directions

This is the first studiy to show that the rs3802842 polymorphism is specifically associated with the age of diagnosis of CRC in HNPCC patients with a hMLH1 mutation. The previous reports regarding this SNP have been on the development of CRC in the general population [72]. Little is known about the function of this SNP. However, the SNP is located near the POU5 gene which is involved in protein transcription and would play a role in disease development. The polymorphisms involved in this study need to be functionally evaluated to determine the role of the polymorphism and the effect it has on the protein function. By doing this we can

better understand the underlying causes of why this SNP appears to act as a modifier of disease.

There were differences found in the Polish and Australian population separately (See appendix 3 and 4), but since the differences were not found in both populations it appears as though the SNPs are not true modifiers of disease. They may, therefore be specific for a particular population. These findings obviously require further investigation in a larger independent cohort of MMR mutation positive HNPCC individuals.

4.3 Limitations of the study

Potential limitations of the study include population stratification. The most obvious cause of population stratification is migration where individuals from one population migrate into another population. However, this should not be affecting our results as we are searching for modifying polymorphisms affecting disease expression in HNPCC patients (defined group), but it cannot be ruled out. In this study, there have been several examples were we have found an association in one of the populations studied, but not in the other.

Environmental factors that are different in the two countries, could potentially affect the results as well as the genetic differences between two populations. Association studies must therefore be interpreted within the context of the genetic structure of the population being studied [85]. However, it has been shown that for most of the common disease-associated polymorphisms, ethnicity is likely to be a poor predictor of an individuals genotype [85].

This study on SNPs has provided results that indicate that a much larger population should be investigated. The association of SNP rs3802842 in hMLH1 mutation carriers and the age of diagnosis of CRC are interesting, and require further investigation searching for modifying polymorphisms influencing disease expression has proven to be a difficult task as controversial results seems to be the rule rather than the exception. Nevertheless, it is believed that the elucidation of modifiers in

HNPCC is important as they have the potential to improve predictive genetic counselling in HNPCC.

5 CONCLUSION

The results of the current study indicated that hMLH1 mutation carriers homozygote variant for the rs3802842 act as a modifier of disease development. However, more HNPCC populations need to be studied to confirm these results. For further analysis, a larger study would be required with an equal proportion of Australian and Polish samples. Additional studies on other HNPCC populations also need to be undertaken to confirm the association found in this study. Only by doing so, we might be able to understand the effect modifier genes have on cancer development in HNPCC patients.

6 REFERENCES

1. 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.
2. 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.
3. 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.
4. Peltomaki, P., *Deficient DNA mismatch repair: a common etiologic factor for colon cancer.* Hum Mol Genet, 2001. **10**(7): p. 735-40.
5. Lawes, D.A., S.B. SenGupta, and P.B. Boulos, *Pathogenesis and clinical management of hereditary non-polyposis colorectal cancer.* Br J Surg, 2002. **89**(11): p. 1357-69.
6. Watson, P. and H.T. Lynch, *Cancer risk in mismatch repair gene mutation carriers.* Fam Cancer, 2001. **1**(1): p. 57-60.
7. Lynch, J., *The genetics and natural history of hereditary colon cancer.* Semin Oncol Nurs, 1997. **13**(2): p. 91-8.
8. Lynch, H.T., T. Smyrk, and J. Lynch, *An update of HNPCC (Lynch syndrome).* Cancer Genet Cytogenet, 1997. **93**(1): p. 84-99.
9. Annie Yu, H.J., et al., *Hereditary nonpolyposis colorectal cancer: preventive management.* Cancer Treat Rev, 2003. **29**(6): p. 461-70.
10. Case, A.S., et al., *Clustering of Lynch syndrome malignancies with no evidence for a role of DNA mismatch repair.* Gynecol Oncol, 2008. **108**(2): p. 438-44.
11. Alarcon, F., et al., *Estimating cancer risk in HNPCC by the GRL method.* Eur J Hum Genet, 2007. **15**(8): p. 831-6.
12. Mitchell, R.J., et al., *Mismatch repair genes hMLH1 and hMSH2 and colorectal cancer: a HuGE review.* Am J Epidemiol, 2002. **156**(10): p. 885-902.
13. de Jong, M.M., et al., *Low-penetrance genes and their involvement in colorectal cancer susceptibility.* Cancer Epidemiol Biomarkers Prev, 2002. **11**(11): p. 1332-52.
14. Tamura, K., et al., *Mechanism of carcinogenesis in familial tumors.* Int J Clin Oncol, 2004. **9**(4): p. 232-45.
15. Mecklin, J.P. and H.J. Jarvinen, *Clinical features of colorectal carcinoma in cancer family syndrome.* Dis Colon Rectum, 1986. **29**(3): p. 160-4.
16. Mecklin, J.P., P. Sipponen, and H.J. Jarvinen, *Histopathology of colorectal carcinomas and adenomas in cancer family syndrome.* Dis Colon Rectum, 1986. **29**(12): p. 849-53.
17. Lanspa, S.J., et al., *Colorectal adenomas in the Lynch syndromes. Results of a colonoscopy screening program.* Gastroenterology, 1990. **98**(5 Pt 1): p. 1117-22.
18. 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.

19. Beck, N.E., et al., *Genetic testing is important in families with a history suggestive of hereditary non-polyposis colorectal cancer even if the Amsterdam criteria are not fulfilled.* Br J Surg, 1997. **84**(2): p. 233-7.
20. Julie, C., et al., *Identification in daily practice of patients with Lynch syndrome (hereditary nonpolyposis colorectal cancer): revised Bethesda guidelines-based approach versus molecular screening.* Am J Gastroenterol, 2008. **103**(11): p. 2825-35; quiz 2836.
21. Boland, C.R., et al., *A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer.* Cancer Res, 1998. **58**(22): p. 5248-57.
22. Lynch, H.T., et al., *Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: an updated review.* Gastroenterology, 1993. **104**(5): p. 1535-49.
23. Lynch, H.T. and A. de la Chapelle, *Hereditary colorectal cancer.* N Engl J Med, 2003. **348**(10): p. 919-32.
24. Peltomaki, P., *Role of DNA mismatch repair defects in the pathogenesis of human cancer.* J Clin Oncol, 2003. **21**(6): p. 1174-9.
25. Leach, F.S., et al., *Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer.* Cell, 1993. **75**(6): p. 1215-25.
26. Nicolaides, N.C., et al., *Mutations of two PMS homologues in hereditary nonpolyposis colon cancer.* Nature, 1994. **371**(6492): p. 75-80.
27. Panagopoulos, I., et al., *The POU5F1P1 pseudogene encodes a putative protein similar to POU5F1 isoform 1.* Oncol Rep, 2008. **20**(5): p. 1029-33.
28. Knudson, A.G., *Hereditary cancer: two hits revisited.* J Cancer Res Clin Oncol, 1996. **122**(3): p. 135-40.
29. Knudson, A.G., Jr., *Mutation and cancer: statistical study of retinoblastoma.* Proc Natl Acad Sci U S A, 1971. **68**(4): p. 820-3.
30. Liang, Y., et al., *DNA damage response pathways in tumor suppression and cancer treatment.* World J Surg, 2009. **33**(4): p. 661-6.
31. Ataian, Y. and J.E. Krebs, *Five repair pathways in one context: chromatin modification during DNA repair.* Biochem Cell Biol, 2006. **84**(4): p. 490-504.
32. Wood, J.L. and J. Chen, *DNA-damage checkpoints: location, location, location.* Trends Cell Biol, 2008. **18**(10): p. 451-5.
33. Lindahl, T., *Instability and decay of the primary structure of DNA.* Nature, 1993. **362**(6422): p. 709-15.
34. Wimmer, K. and J. Etzler, *Constitutional mismatch repair-deficiency syndrome: have we so far seen only the tip of an iceberg?* Hum Genet, 2008. **124**(2): p. 105-22.
35. Collins, F.S., L.D. Brooks, and A. Chakravarti, *A DNA polymorphism discovery resource for research on human genetic variation.* Genome Res, 1998. **8**(12): p. 1229-31.
36. Spurkland, A. and H.F. Harbo, *Genetiske analyser i genetisk epidemiologi.* Norsk Epidemiologi, 2002. **12**(2): p. 89-96.
37. Brookes, A.J., *The essence of SNPs.* Gene, 1999. **234**(2): p. 177-86.
38. Brown, T.A., ed. *Genetics A molecular approach*. Third edition ed. 1998, Stanley Thornes: Cheltenham. 184-189.
39. Ke, X., M.S. Taylor, and L.R. Cardon, *Singleton SNPs in the human genome and implications for genome-wide association studies.* Eur J Hum Genet, 2008. **16**(4): p. 506-15.

40. Riva, A. and I.S. Kohane, *SNPper: retrieval and analysis of human SNPs.* Bioinformatics, 2002. **18**(12): p. 1681-5.
41. Riva, A. and I.S. Kohane, *A SNP-centric database for the investigation of the human genome.* BMC Bioinformatics, 2004. **5**: p. 33.
42. Haider, N.B., et al., *Genetic modifiers of vision and hearing.* Hum Mol Genet, 2002. **11**(10): p. 1195-206.
43. 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.
44. 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.
45. Nebert, D.W., *Transcription factors and cancer: an overview.* Toxicology, 2002. **181-182**: p. 131-41.
46. Scott, R.J., *Modifier genes and HNPCC: variable phenotypic expression in HNPCC and the search for modifier genes.* Eur J Hum Genet, 2008. **16**(5): p. 531-2.
47. Zecevic, M., et al., *IGF1 gene polymorphism and risk for hereditary nonpolyposis colorectal cancer.* J Natl Cancer Inst, 2006. **98**(2): p. 139-43.
48. 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.
49. 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.
50. 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.
51. Levine, A.J., *p53, the cellular gatekeeper for growth and division.* Cell, 1997. **88**(3): p. 323-31.
52. Chen, J., et al., *Association between Aurora-A kinase polymorphisms and age of onset of hereditary nonpolyposis colorectal cancer in a Caucasian population.* Mol Carcinog, 2007. **46**(4): p. 249-56.
53. Sen, S., et al., *Amplification/overexpression of a mitotic kinase gene in human bladder cancer.* J Natl Cancer Inst, 2002. **94**(17): p. 1320-9.
54. Marumoto, T., et al., *Roles of aurora-A kinase in mitotic entry and G2 checkpoint in mammalian cells.* Genes Cells, 2002. **7**(11): p. 1173-82.
55. Aquilina, G., M. Crescenzi, and M. Bignami, *Mismatch repair, G(2)/M cell cycle arrest and lethality after DNA damage.* Carcinogenesis, 1999. **20**(12): p. 2317-26.
56. 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.
57. Donnellan, R. and R. Chetty, *Cyclin D1 and human neoplasia.* Mol Pathol, 1998. **51**(1): p. 1-7.
58. Bala, S. and P. Peltomaki, *CYCLIN D1 as a genetic modifier in hereditary nonpolyposis colorectal cancer.* Cancer Res, 2001. **61**(16): p. 6042-5.

59. Houlston, R., et al., *Explaining differences in the severity of familial adenomatous polyposis and the search for modifier genes.* Gut, 2001. **48**(1): p. 1-5.
60. Kemp, Z., et al., *An update on the genetics of colorectal cancer.* Hum Mol Genet, 2004. **13 Spec No 2**: p. R177-85.
61. Burghes, A.H., H.E. Vaessin, and A. de La Chapelle, *Genetics. The land between Mendelian and multifactorial inheritance.* Science, 2001. **293**(5538): p. 2213-4.
62. Tomlinson, I.P., et al., *A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3.* Nat Genet, 2008. **40**(5): p. 623-30.
63. Dong, Z. and J.T. Zhang, *Initiation factor eIF3 and regulation of mRNA translation, cell growth, and cancer.* Crit Rev Oncol Hematol, 2006. **59**(3): p. 169-80.
64. Zhang, L., et al., *An oncogenic role for the phosphorylated h-subunit of human translation initiation factor eIF3.* J Biol Chem, 2008. **283**(35): p. 24047-60.
65. Broderick, P., et al., *A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk.* Nat Genet, 2007. **39**(11): p. 1315-7.
66. Leng, A., et al., *Smad4/Smad7 balance: a role of tumorigenesis in gastric cancer.* Exp Mol Pathol, 2009.
67. Halder, S.K., et al., *Smad7 induces hepatic metastasis in colorectal cancer.* Br J Cancer, 2008. **99**(6): p. 957-65.
68. Hariharan, R., et al., *Mutational analysis of Smad7 in human cervical cancer.* Oncol Rep, 2009. **21**(4): p. 1001-4.
69. Namkoong, H., et al., *The bone morphogenetic protein antagonist gremlin 1 is overexpressed in human cancers and interacts with YWHAH protein.* BMC Cancer, 2006. **6**: p. 74.
70. Suzuki, M., et al., *DNA methylation-associated inactivation of TGFbeta-related genes DRM/Gremlin, RUNX3, and HPP1 in human cancers.* Br J Cancer, 2005. **93**(9): p. 1029-37.
71. Jaeger, E., et al., *Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk.* Nat Genet, 2008. **40**(1): p. 26-8.
72. Tenesa, A., et al., *Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21.* Nat Genet, 2008. **40**(5): p. 631-7.
73. Miller, S.A., D.D. Dykes, and H.F. Polesky, *A simple salting out procedure for extracting DNA from human nucleated cells.* Nucleic Acids Res, 1988. **16**(3): p. 1215.
74. Kwok, P.Y., *Methods for genotyping single nucleotide polymorphisms.* Annu Rev Genomics Hum Genet, 2001. **2**: p. 235-58.
75. Shi, M.M., *Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies.* Clin Chem, 2001. **47**(2): p. 164-72.
76. Higuchi, R., et al., *Kinetic PCR analysis: real-time monitoring of DNA amplification reactions.* Biotechnology (N Y), 1993. **11**(9): p. 1026-30.
77. Livak, K.J., et al., *Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization.* PCR Methods Appl, 1995. **4**(6): p. 357-62.

78. *Essentials of RT-PCR*. [cited 01.03.2009]; Available from: http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_039996.pdf.
79. *Real Time-PCR Vs. Traditional PCR*. [cited 01.03.2009]; Available from: http://www.appliedbiosystems.com/support/tutorials/pdf/rtpcr_vs_tradpcr.pdf.
80. Schmith, V.D., et al., *Pharmacogenetics and disease genetics of complex diseases.* Cell Mol Life Sci, 2003. **60**(8): p. 1636-46.
81. Wijnhoven, S.W. and H. van Steeg, *Transgenic and knockout mice for DNA repair functions in carcinogenesis and mutagenesis.* Toxicology, 2003. **193**(1-2): p. 171-87.
82. Heinimann, K., et al., *N-acetyltransferase 2 influences cancer prevalence in hMLH1/hMSH2 mutation carriers.* Cancer Res, 1999. **59**(13): p. 3038-40.
83. Souglakos, J., *Genetic alterations in sporadic and hereditary colorectal cancer: implementations for screening and follow-up.* Dig Dis, 2007. **25**(1): p. 9-19.
84. Cook, A.L. and R.A. Sturm, *POU domain transcription factors: BRN2 as a regulator of melanocytic growth and tumourigenesis.* Pigment Cell Melanoma Res, 2008. **21**(6): p. 611-26.
85. Lohmueller, K.E., et al., *Patterns of genetic variation in the hypertension candidate gene GRK4: ethnic variation and haplotype structure.* Ann Hum Genet, 2006. **70**(Pt 1): p. 27-41.

Appendix 1: Reagents, materials and equipment used in the study

Table 1: **Reagents.** Listed below are the concentration and supplier of the reagents used.

Reagent	Concentration	Supplier
TaqMan® PCR Universal Master Mix	2x	Applied Biosystems[1]
TaqMan® SNP genotyping Assays	40x	Assay-on-Demand[2]

[1] Applied Biosystems, Foster city, CA, USA
[2] Service offered by Applied Biosystems

Table 2: **Materials.** Listed below are the materials and its suppliers.

Materials	Supplier
96-well PCR Microplate	Axygen® Scientific[1]
96-well Masterblock® plate	Greiner Bio-One[2]

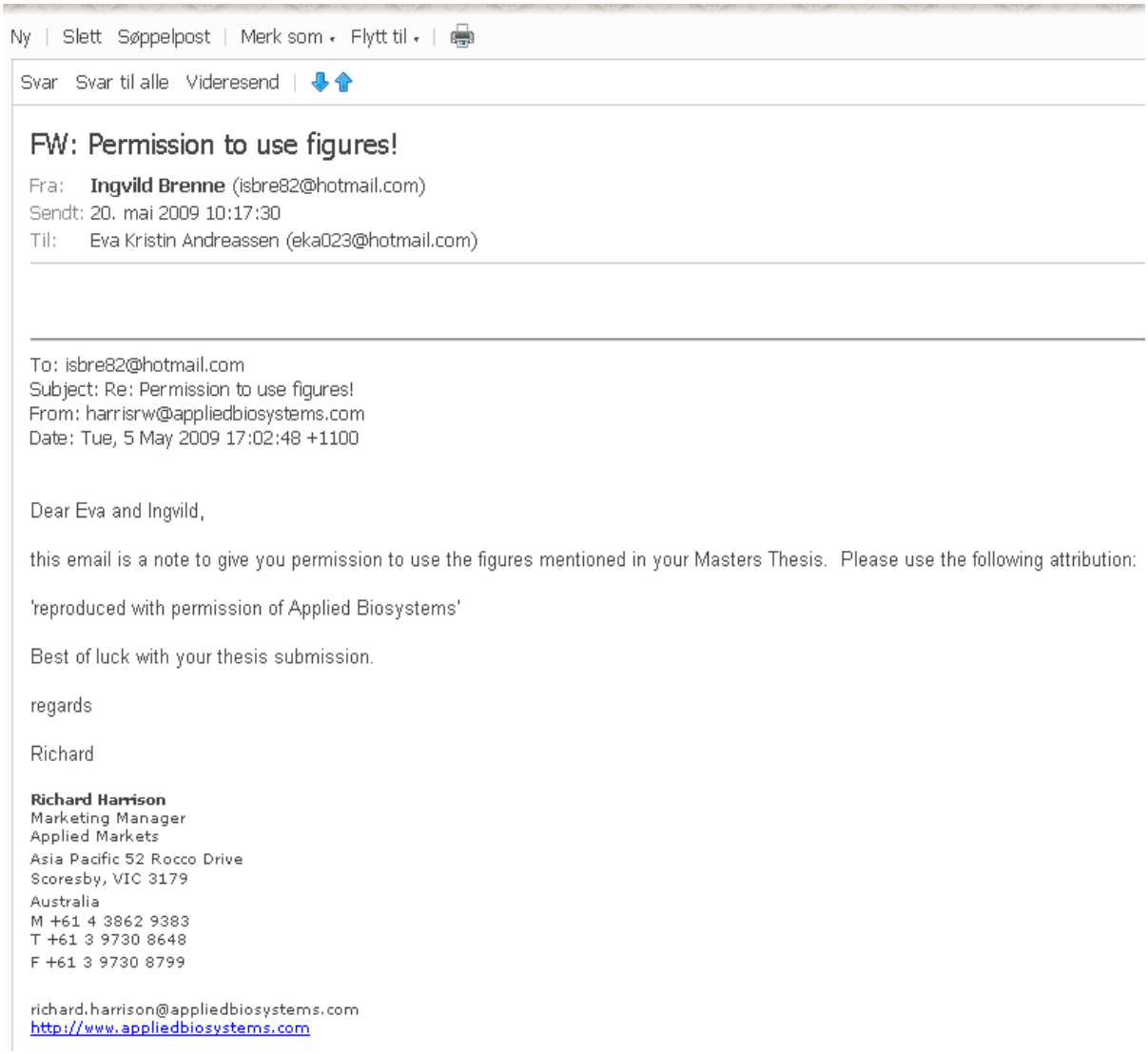
[1] Axygen® Scientific, Union city, CA, USA
[2] Greiner Bio-One, Frickenhausen, Germany

Table 3: **Equipment**. The equipment and its suppliers are listed below.

Equipment	Supplier
ABI PRISM® 7900HT SDS	Applied Biosystems[1]
ABI PRISM® Genetic analyser	Applied Biosystems[1]
Hybaid PCR Express Thermal Cycler	Hybaid[2]

[1] Applied Biosystems, Foster city, CA, USA
[2] Hybaid, Franklin, USA

Appendix 2: Permission to use figure from Applied Biosytems



Ny | Slett Søppelpost | Merk som | Flytt til |

Svar Svar til alle Videresend

FW: Permission to use figures!

Fra: **Ingvild Brenne** (isbre82@hotmail.com)
Sendt: 20. mai 2009 10:17:30
Til: Eva Kristin Andreassen (eka023@hotmail.com)

To: isbre82@hotmail.com
Subject: Re: Permission to use figures!
From: harrisrw@appliedbiosystems.com
Date: Tue, 5 May 2009 17:02:48 +1100

Dear Eva and Ingvild,

this email is a note to give you permission to use the figures mentioned in your Masters Thesis. Please use the following attribution:

'reproduced with permission of Applied Biosystems'

Best of luck with your thesis submission.

regards

Richard

Richard Harrison
Marketing Manager
Applied Markets
Asia Pacific 52 Rocco Drive
Scoresby, VIC 3179
Australia
M +61 4 3862 9383
T +61 3 9730 8648
F +61 3 9730 8799

richard.harrison@appliedbiosystems.com
http://www.appliedbiosystems.com

Appendix 3: Genotype distribution in the Australian HNPCC population

TABLE 1- ALLELE FREQUNCY DISTRIBUTION OF THE rs16892766 POLYMORPHISM IN THE AUSTRALIAN POPULATION

SNP 1 Rs16892766	AA (%)	AC (%)	CC (%)	p-value[1]
Allele frequency				
Subject group (n=365)				
CRC+ (n=164)	132 (80)	32 (20)	0	p=0.032
CRC- (n=201)	178 (89)	23 (11)	0	
				OR:1.88 CI:1.05-3.36
Subject group (n=344)				
hMLH1 (n=180)	154 (86)	26 (14)	0	p=0.96
hMSH2 (n=164)	140 (85)	24 (15)	0	
Subject group (n=368)				
Female (n=218)	190 (87)	28 (13)	0	p=0.17
Male (n=150)	123 (82)	27 (18)	0	
Subject group (n=217)				
Gynaecological+ (n=39)	35 (90)	4 (10)	0	p=0.793*
Gynaecological- (n=178)	154 (87)	24 (13)	0	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2] p-value: Homozygous wild type (AA) compared to combination of heterozygous and homozygous variant genotypes (AC+CC)
[3]p-value: Homozygous variant genotype (CC) compared to combination of wild type and heterozygous genotypes (AC+AA)
* p-value was calculated using Fisher's exact test.

TABLE 2- ALLELE FREQUNCY DISTRIBUTION OF THE rs3802842 POLYMORPHISM IN THE AUSTRALIAN POPULATION

SNP 2 Rs3802842	AA (%)	AC (%)	CC (%)	p-value[1]	AC+CC (%)	p-value[2]	AC+AA (%)	p-value[3]
Allele frequency								
Subject group (n=359)								
CRC+ (n=162)	86 (53)	59 (36)	17 (11)	p=0.82	76 (47)	p=0.97	145 (89)	p=0.55
CRC- (n=197)	105 (53)	75 (38)	17 (9)		92 (47)		180 (91)	
Subject group (n=338)								
hMLH1 (n=174)	90 (52)	69 (40)	15 (8)	p=0.66	84 (48)	p=0.75	159 (92)	p=0.37
hMSH2 (n=164)	82 (50)	63 (38)	19 (12)		82 (50)		145 (88)	
Subject group (n=362)								
Female (n=213)	112 (53)	80 (37)	21 (10)	p=0.91	101 (47)	p=0.74	192 (90)	p=0.72
Male (n=149)	81 (54)	55 (37)	13 (9)		68 (46)		136 (91)	
Subject group (n=212)								
Gynaecological+ (n=38)	21 (55)	17 (45)	0	p=0.07	17 (45)	p=0.74	38 (100)	p=0.017
Gynaecological- (n=174)	91 (52)	62 (36)	21 (12)		83 (48)		152 (88)	
							OR:0.093 CI:0.005-1.56	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (AA) compared to combination of heterozygous and homozygous variant genotypes (AC+CC)
[3]p-value: Homozygous variant genotype (CC) compared to combination of heterozygous and wild type genotypes (AC+AA)

TABLE 3- ALLELE FREQUNCY DISTRIBUTION OF THE rs10318 POLYMORPHISM IN THE AUSTRALIAN POPULATION

SNP 3 Rs10318	CC (%)	CT (%)	TT (%)	p-value[1]	CT+TT (%)	p-value[2]	CT+CC (%)	p-value[3]
Allele frequency								
Subject group (n=351)								
CRC+ (n=157)	113 (72)	42 (27)	2 (1)	p=0.09	44 (28)	p=0.32	155 (99)	p=0.044*
CRC- (n=194)	130 (67)	53 (27)	11 (6)		64 (33)		183 (94)	
							OR:0.215 CI:0.047-0.984	
Subject group (n=330)								
hMLH1 (n=169)	115 (68)	49 (29)	5 (3)	p=0.58	54 (32)	p=0.86	164 (97)	p=0.35
hMSH2 (n=161)	111 (69)	42 (26)	8 (5)		50 (31)		153 (95)	
Subject group (n=354)								
Female (n=209)	148 (71)	56 (27)	5 (2)	p=0.18	61 (29)	p=0.36	204 (98)	p=0.070
Male (n=145)	96 (66)	40 (28)	9 (6)		49 (34)		136 (94)	
Subject group (n=208)								
Gynaecological+ (n=37)	25 (68)	10 (27)	2 (5)	p=0.332*	12 (32)	p=0.60	35 (95)	p=0.19
Gynaecological- (n=171)	123 (72)	45 (26)	3 (2)		48 (28)		168 (98)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (CC) compared to combination of heterozygous and homozygous variant genotypes (CT+TT)
[3]p-value: Homozygous variant genotype (TT) compared to combination heterozygous and wild type genotypes (CT+CC)
* p-value was calculated using Fisher's exact test.

TABLE 4- ALLELE FREQUNCY DISTRIBUTION OF THE rs4939827 POLYMORPHISM IN THE AUSTRALIAN POPULATION

SNP 4 Rs4939827	CC (%)	CT (%)	TT (%)	p-value[1]	CT+TT (%)	p-value[2]	CT+CC (%)	p-value[3]
Allele frequency								
Subject group (n=363)								
CRC+ (n=164)	57 (35)	81 (49)	26 (16)	p=0.40	107 (65)	p=0.18	138 (84)	p=0.57
CRC- (n=199)	56 (28)	107 (54)	36 (18)		143 (72)		163 (82)	
Subject group (n=342)								
hMLH1 (n=178)	53 (30)	97 (54)	28 (16)	p=0.65	125 (70)	p=0.92	150 (84)	p=0.36
hMSH2 (n=164)	48 (29)	84 (51)	32 (20)		116 (71)		132 (80)	
Subject group (n=366)								
Female (n=217)	57 (26)	119 (55)	41 (19)	p=0.031	160 (74)	p=0.010	176 (81)	p=0.17
Male (n=149)	58 (39)	71 (48)	20 (13)		91 (61)		129 (87)	
					OR:1.047 CI:0.443-0.477			
Subject group (n=216)								
Gynaecological+ (n=39)	14 (36)	17 (44)	8 (20)	p=0.25	25 (64)	p=0.12	31 (80)	p=0.92
Gynaecological- (n=177)	42 (24)	100 (56)	35 (20)		135 (76)		142 (80)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (CC) compared to combination of heterozygous and homozygous variant genotypes (CT+TT)
[3]p-value: Homozygous variant genotype (TT) compared to combination heterozygous and wild type genotypes (CT+CC)

TABLE 5- ALLELE FREQUNCY DISTRIBUTION OF THE rs4464148 POLYMORPHISM IN THE AUSTRALIAN POPULATION

SNP 5 Rs4464148	TT (%)	TC (%)	CC (%)	p-value[1]	TC+CC (%)	p-value[2]	TC+TT (%)	p-value[3]
Allele frequency								
Subject group (n=367)								
CRC+ (n=165)	83 (50)	70 (43)	12 (7)	p=0.41	82 (50)	p=0.20	153 (93)	p=0.47
CRC- (n=202)	88 (44)	95 (47)	19 (9)		114 (56)		183 (91)	
Subject group (n=346)								
hMLH1 (n=182)	80 (44)	81 (44)	21 (12)	p=0.39	102 (56)	p=0.92	161 (88)	p=0.18
hMSH2 (n=164)	73 (45)	79 (48)	12 (7)		91 (55)		152 (93)	
Subject group (n=370)								
Female (n=220)	95 (43)	106 (48)	19 (9)	p=0.19	125 (57)	p=0.095	201 (91)	p=0.82
Male (n=150)	78 (52)	58 (39)	14 (9)		72 (48)		136 (91)	
Subject group (n=219)								
Gynaecological+ (n=39)	20 (51)	17 (44)	2 (5)	p=0.45	19 (49)	p=0.25	37 (95)	p=0.747*
Gynaecological- (n=180)	74 (41)	90 (50)	16 (9)		106 (59)		164 (91)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (TT) compared to combination of heterozygous and homozygous variant genotypes (TC+CC)
[3]p-value: Homozygous variant genotype (CC) compared to combination heterozygous and wild type genotypes (TC+TT)
* p-value was calculated using Fisher's exact test

TABLE 6- ALLELE FREQUENCY DISTRIBUTION OF THE rs6983267 POLYMORPHISM IN THE AUSTRALIAN POPULATION

SNP 6 Rs6983267	TT (%)	TG (%)	GG (%)	p-value[1]	TG+GG (%)	p-value[2]	TG+TT (%)	p-value[3]
Allele frequency								
Subject group (n=356)								
CRC+ (n=161)	31 (19)	87 (54)	43 (27)	p=0.65	130 (81)	p=0.77	118 (73)	p=0.46
CRC- (n=195)	40 (21)	96 (49)	59 (30)		155 (79)		136 (70)	
Subject group (n=335)								
hMLH1 (n=173)	37 (21)	98 (57)	38 (22)	p=0.012	136 (79)	p=0.16	135 (78)	p=0.004
hMSH2 (n=162)	25 (16)	78 (48)	59 (36)		137 (84)		103 (64)	
							OR:0.491 CI:0.304-0.795	
Subject group (n=359)								
Female (n=214)	52 (24)	103 (48)	59 (28)	p=0.030	162 (76)	p=0.009	155 (72)	p=0.67
Male (n=145)	19 (13)	83 (57)	43 (30)		126 (87)		102 (70)	
					OR:1.131 CI:0.508-2.520			
Subject group (n=213)								
Gynaecological+ (n=38)	10 (26)	16 (42)	12 (32)	p=0.73	28 (74)	p=0.76	26 (68)	p=0.56
Gynaecological- (n=175)	42 (24)	86 (49)	47 (27)		133 (76)		128 (73)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (TT) compared to combination of heterozygous and homozygous variant genotypes (TG+GG)
[3]p-value: Homozygous variant genotype (GG) compared to combination heterozygous and wild type genotypes (TG+TT)

TABLE 7- ALLELE FREQUNCY DISTRIBUTION OF THE rs7014346 POLYMORPHISM IN THE AUSTRALIAN POPULATION

SNP 7 Rs7014346	GG (%)	GA (%)	AA (%)	p-value[1]	GA+AA (%)	p-value[2]	GA+GG (%)	p-value[3]
Allele frequency								
Subject group (n=360)								
CRC+ (n=161)	69 (43)	73 (45)	19 (12)	p=0.85	92 (57)	p=0.68	142 (88)	p=0.62
CRC- (n=199)	81 (41)	91 (46)	27 (13)		118 (59)		172 (87)	
Subject group (n=339)								
hMLH1 (n=176)	79 (45)	80 (45)	17 (10)	p=0.12	97 (55)	p=0.10	159 (90)	p=0.082
hMSH2 (n=163)	59 (36)	78 (48)	26 (16)		104 (64)		137 (84)	
Subject group (n=363)								
Female (n=216)	94 (44)	93 (43)	29 (13)	p=0.39	122 (56)	p=0.30	187 (87)	p=0.60
Male (n=147)	56 (38)	74 (50)	17 (12)		91 (62)		130 (88)	
Subject group (n=215)								
Gynaecological+ (n=38)	14 (37)	18 (47)	6 (16)	p=0.64	24 (63)	p=0.35	32 (84)	p=0.65
Gynaecological+(n=177)	80 (45)	74 (42)	23 (13)		97 (55)		154 (87)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (GG) compared to combination of heterozygous and homozygous variant genotypes (GA+AA)
[3]p-value: Homozygous variant genotype (AA) compared to combination heterozygous and wild type genotypes (GA+GG)

TABLE 8- ALLELE FREQUNCY DISTRIBUTION OF THE rs4779584 POLYMORPHISM IN THE AUSTRALIAN POPULATION

SNP 8 Rs4779584	CC (%)	CT (%)	TT (%)	p-value[1]	CT+TT (%)	p-value[2]	CT+CC (%)	p-value[3]
Allele frequency								
Subject group (n=355)								
CRC+ (n=159)	107 (67)	44 (28)	8 (5)	p=0.24	52 (33)	p=0.10	151 (95)	p=0.81
CRC- (n=196)	115 (59)	70 (36)	11 (5)		81 (41)		185 (95)	
Subject group (n=334)								
hMLH1 (n=171)	111 (65)	52 (30)	8 (5)	p=0.36	60 (35)	p=0.17	163 (95)	p=0.41
hMSH2 (n=163)	94 (58)	58 (35)	11 (7)		69 (42)		152 (93)	
Subject group (n=359)								
Female (n=213)	128 (60)	77 (36)	8 (4)	p=0.043	85 (40)	p=0.28	205 (96)	p=0.070
Male (n=146)	96 (66)	38 (26)	12 (8)		50 (34)		134 (92)	
Subject group (n=211)								
Gynaecological+ (n=39)	21 (54)	15 (38)	3 (8)	p=0.268*	18 (46)	p=0.37	36 (92)	p=0.167*
Gynaecological+(n=172)	106 (62)	61 (35)	5 (3)		66 (38)		167 (97)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (CC) compared to combination of heterozygous and homozygous variant genotypes (CT+TT)
[3]p-value: Homozygous variant genotype (TT) compared to combination heterozygous and wild type genotypes (CT+CC)
* p-value was calculated using Fisher's exact test

TABLE 9- ALLELE FREQUNCY DISTRIBUTION OF THE rs10795668 POLYMORPHISM IN THE AUSTRALIAN POPULATION

SNP 9 Rs10795668	GG (%)	GA (%)	AA (%)	p-value[1]	GA+AA (%)	p-value[2]	GA+GG (%)	p-value[3]
Allele frequency								
Subject group (n=348)								
CRC+ (n=159)	74 (47)	67 (42)	18 (11)	p=0.26	85 (53)	p=0.35	14 (89)	p=0.28
CRC- (n=189)	79 (42)	95 (50)	15 (8)		110 (58)		174 (92)	
Subject group (n=327)								
hMLH1 (n=164)	74 (45)	75 (46)	15 (9)	p=0.72	90 (55)	p=0.74	149 (91)	p=0.42
hMSH2 (n=163)	76 (46.5)	76 (46.5)	11 (7)		87 (53.5)		152 (93)	
Subject group (n=351)								
Female (n=208)	92 (44)	96 (46)	20 (10)	p=0.99	116 (56)	p=0.88	188 (90)	p=0.87
Male (n=143)	64 (45)	66 (46)	13 (9)		79 (55)		130 (91)	
Subject group (n=207)								
Gynaecological+ (n=38)	16 (42)	21 (55)	1 (3)	p=0.20	22 (58)	p=0.83	37 (97)	p=0.134*
Gynaecological+(n=169)	75 (44.5)	75 (44.5)	19 (11)		94 (55.5)		150 (89)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (GG) compared to combination of heterozygous and homozygous variant genotypes (GA+AA)
[3]p-value: Homozygous variant genotype (AA) compared to combination heterozygous and wild type genotypes (GA+GG)
* p-value was calculated using Fisher's exact test

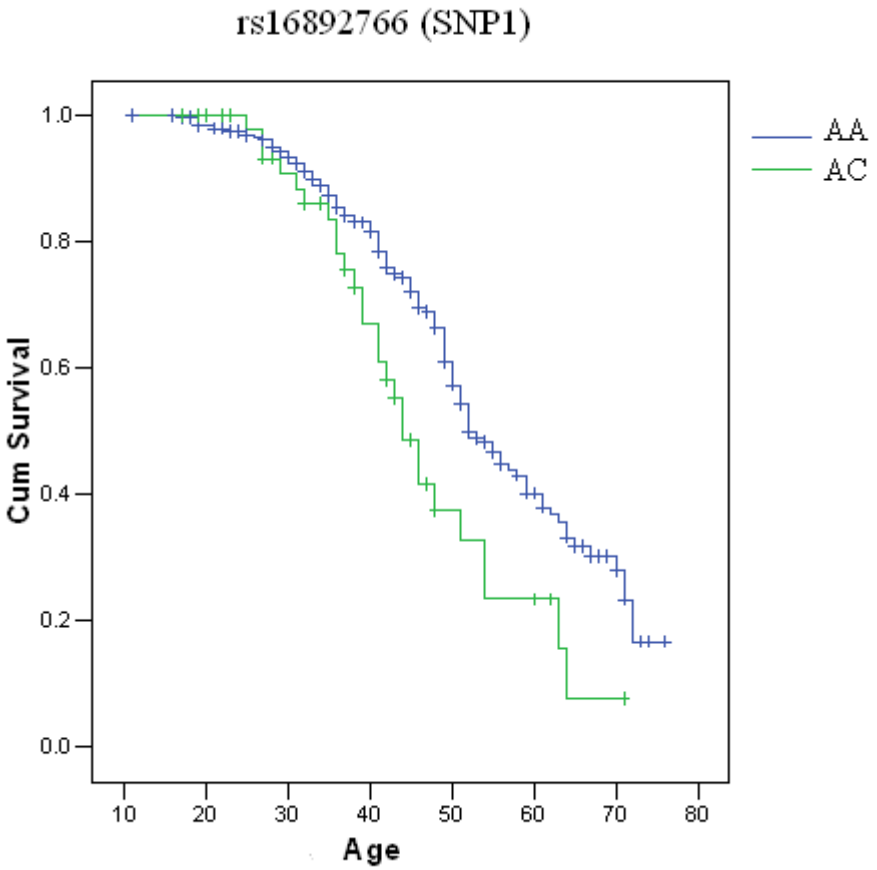


Figure 1: Kaplan-Meier estimated by rs16892766 (SNP1). The graph shows the effect of homozygous wild type genotypes (AA) compared to heterozygous genotype (AC) have on age of diagnosis of CRC in Australian HNPCC patients. There is a statistical significant difference between the genotypes and the age of diagnosis of CRC (log-rank test: p=0.004, Breslow test: p=0.014 and Tarone-Ware test: p=0.006).

Table 10: Median age of diagnosis of CRC (age at which 50% of the population is cancer free) in HNPCC participants for SNP rs16892766

Genotype	Subject group SNP rs16892766 [1]
Wild type (AA)	52 yrs (n=297)
Heterozygote (AC)	44 yrs (n=50)

[1] The subject group includes 347 samples

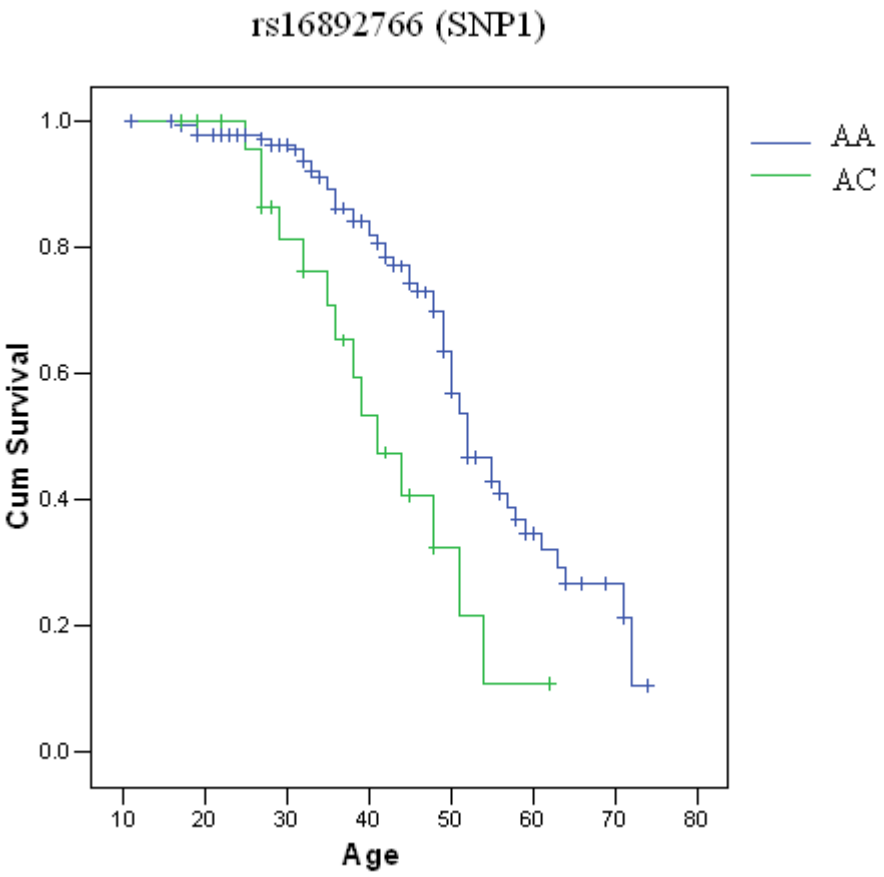


Figure 2: Kaplan-Meier estimated by rs16892766 (SNP1) genotype in hMLH1 mutation carriers. The graph shows the effect of homozygous wild type genotype (AA) compared to heterozygous genotype (AC) have on age of age of diagnosis of CRC in Australian HNPCC patients. There is a statistical significant difference between the genotypes and the age of diagnosis of CRC (log-rank test: p=0.001, Breslow test: p=0.001 and Tarone-Ware test: p=0.0008)

Table 11: Median age of diagnosis of CRC (age at which 50% of the population is cancer free) in HNPCC participants for SNP rs16892766

Genotype	Subject group SNP rs16892766 [1]
Wild type (AA)	52 yrs (n=144)
Heterozygot (AC)	41 yrs (n=25)

[1] The subject group includes 169 samples

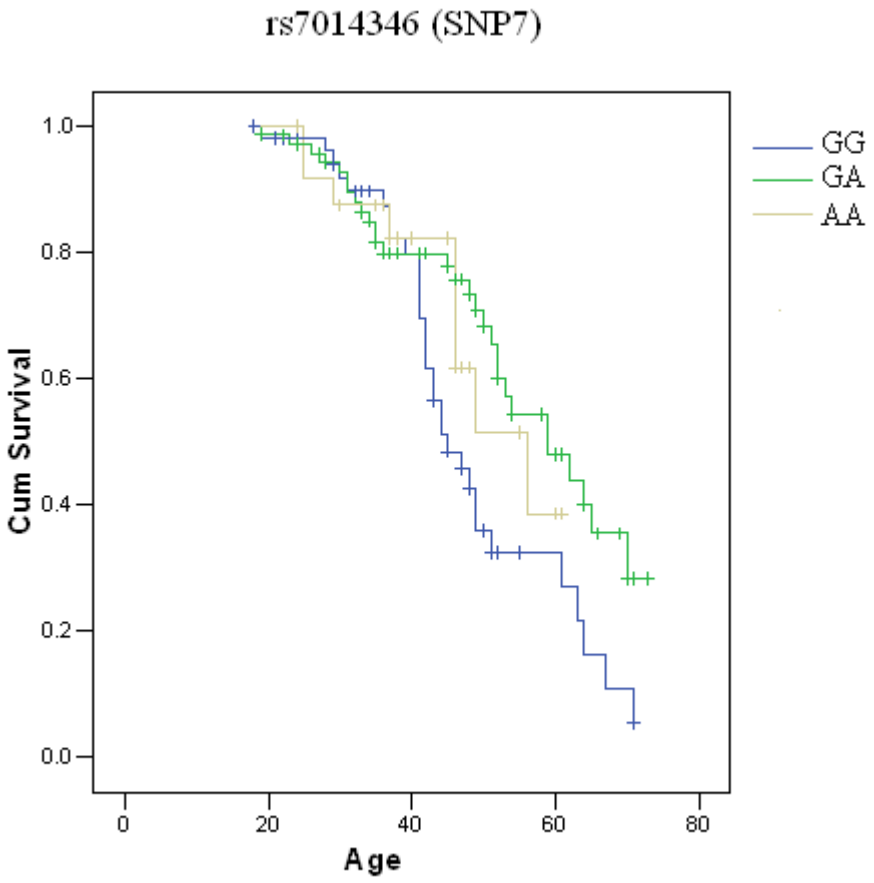


Figure 3: Kaplan-Meier estimated by rs7014346 (SNP7) genotype in hMSH2 mutation carriers. The graph shows the effect the three genotypes (GG vs GA vs AA) have on age of age of diagnosis of CRC in Australian HNPCC patients. There is a statistical significant difference between the genotypes for the log-rank test: p=0.031 and the age of diagnosis of CRC. The Breslow and the Tarone-Ware tests did not show statistical significant different results.

Table 12: Median age of diagnosis of CRC (age at which 50% of the population is cancer free) in HNPCC participants for SNP rs7014346

Genotype	Subject group SNP rs7014346 [1]
Wild type (GG)	45 yrs (n=54)
Heterozygote (GA)	59 yrs (n=72)
Variant (AA)	56 yrs (n=26)

[1] The subject group includes 152 samples

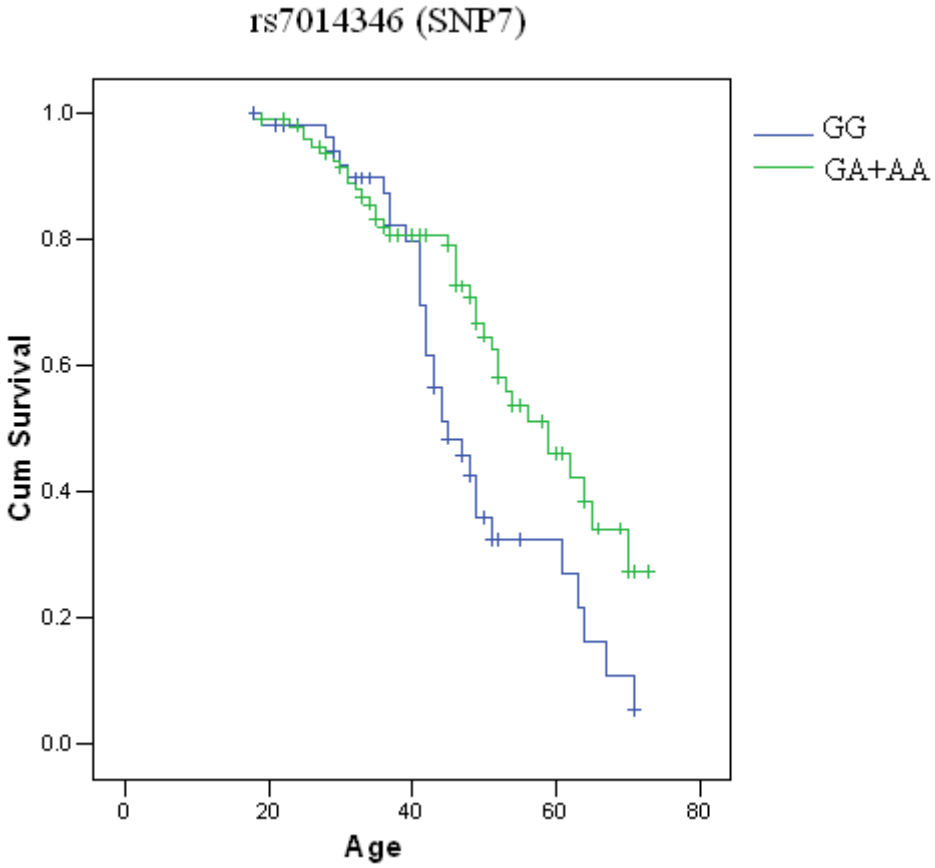


Figure 4: Kaplan-Meier estimated by rs7014346 (SNP7) genotype in hMSH2 mutation carriers. The graph shows the effect of homozygous wild type genotype (GG) compared to combination of heterozygous and variant genotypes (GA+AA) have on age of age of diagnosis of CRC in Australian HNPCC patients. There is a statistical significant difference between the wild type genotype compared to heterozygous and variant genotypes for the log-rank test: p=0.010 and the Tarone-Ware test: p=0.030 and the age of diagnosis of CRC. The Breslow test did not show a statistical significant different result.

Table 13: Median age of diagnosis of CRC (age at which 50% of the population is cancer free) in HNPCC participants for SNP rs7014346.

Genotype	Subject group SNP rs7014346 [1]
Wild type (GG)	45 yrs (n=54)
Heterozygote (GA) + variant (AA)	59 yrs (n=98)

[1] The subject group includes 152 samples

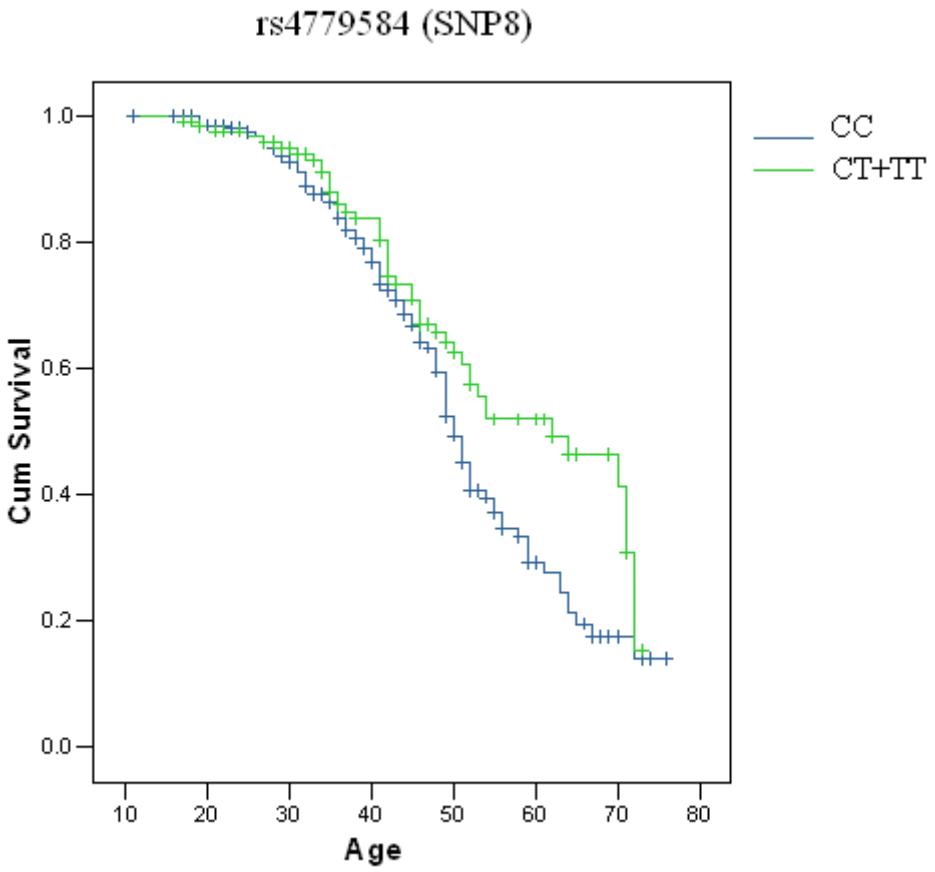


Figure 5: Kaplan-Meier estimated by rs4779584 (SNP8). The graph shows the effect of homozygous wild type genotype (CC) compared to combination of heterozygous and variant genotypes (CT+TT) have on age of age of diagnosis of CRC in Australian HNPCC patients. There is a statistical significant difference between the genotypes for the log-rank test: p=0.021 and the age of diagnosis of CRC. The Breslow and the Tarone-Ware tests did not show statistical significant different results.

Table 14: Median age of diagnosis of CRC (age at which 50% of the population is cancer free) in HNPCC participants for SNP rs4779584

Genotype	Subject group SNP rs4779584 [1]
Wild type (CC)	50 yrs (n=210)
Heterozygote (CT) + Variant (TT)	62 yrs (n=127)

[1] The subject group includes 337 samples

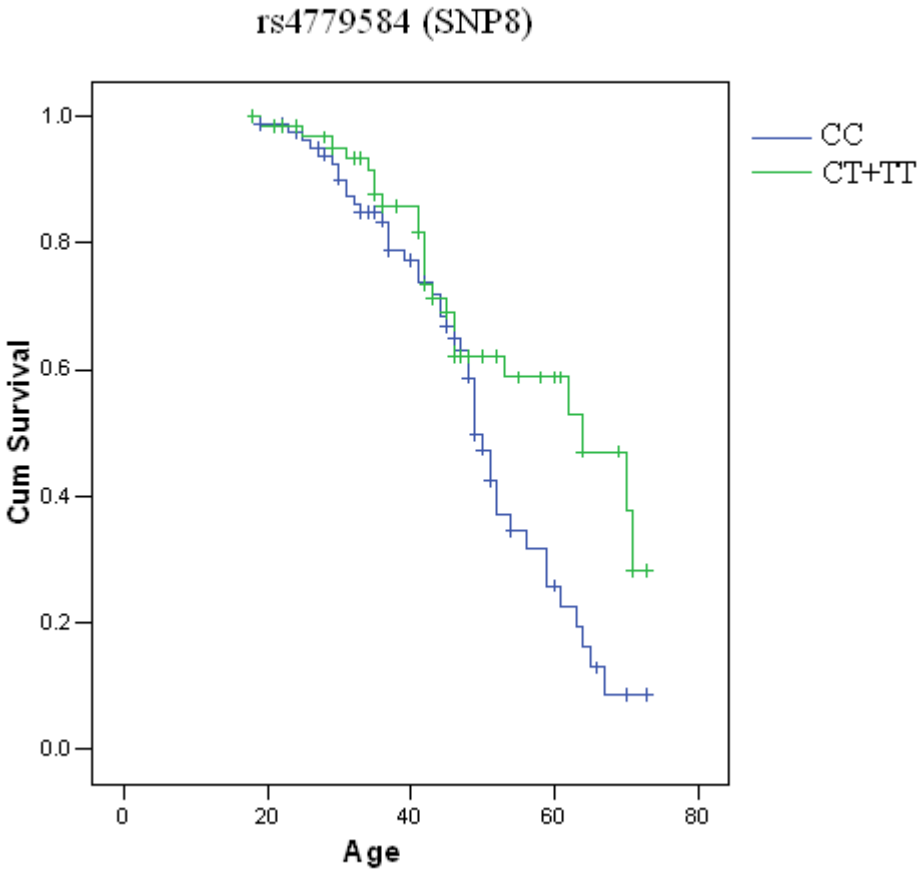


Figure 6: **Kaplan-Meier estimated by rs4779584 (SNP8) genotype in hMSH2 mutation carriers.** The graph shows the effect of homozygous wild type genotype (CC) compared to combination of heterozygous and variant genotypes (CT+TT) have on age of age of diagnosis of CRC in Australian HNPCC patients. There is a statistical significant difference between the wild type genotype compared to heterozygous and variant genotypes for the log-rank test: p=0.020 and the age of diagnosis of CRC. The Breslow and Tarone-Ware tests did not show a statistical significant different result.

Table 15: Median age of diagnosis of CRC (age at which 50% of the population is cancer free) in HNPCC participants for SNP rs4779584

Genotype	Subject group SNP rs4779584 [1]
Wild type (CC)	49 yrs (n=85)
Heterozygote (CT) + variant (TT)	64 yrs (n=67)

[1]The subject group includes 152 samples

Appendix 4: Genotype distribution in the Polish HNPCC population

TABLE 1- ALLELE FREQUNCY DISTRIBUTION OF THE rs16892766 POLYMORPHISM IN THE POLISH POPULATION

SNP 1 Rs16892766	AA (%)	AC (%)	CC (%)	p-value[1]	AC+CC (%)	p-value[2]	AC+AA (%)	p-value[3]
Allele frequency								
Subject group (n=311)								
CRC+ (n=121)	109 (90)	12 (10)	0	p=0.06	12 (10)	p=0.018	121 (100)	p=1.0*
CRC- (n=190)	152 (80)	37 (19.5)	1 (0.5)		38 (20)		189 (99.5)	
					OR:2.271 CI:1.134-4.547			
Subject group (n=286)								
hMLH1 (n=171)	145 (85)	26 (15)	0	p=0.46	26 (15)	p=0.92	171 (100)	p=0.402*
hMSH2 (n=115)	98 (85)	16 (14)	1 (1)		17 (15)		114 (99)	
Subject group (n=311)								
Female (n=189)	161 (85)	28 (15)	0	p=0.39	28 (15)	p=0.45	189 (100)	p=0.392*
Male (n=122)	100 (82)	21 (17)	1 (1)		22 (18)		121 (99)	
Subject group (n=189)								
Gynaecological+ (n=41)	32 (78)	9 (22)	0	p=0.15	9 (22)	p=0.15	41 (100)	
Gynaecological- (n=148)	129 (87)	19 (13)	0		19 (13)		148 (100)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (AA) compared to combination of heterozygous and homozygous variant genotypes (AC+CC)
[3]p-value: Homozygous variant genotype (CC) compared to combination of wild type and heterozygous genotypes (AC+AA)
* p-value was calculated using Fisher's exact test

TABLE 2- ALLELE FREQUNCY DISTRIBUTION OF THE rs3802842 POLYMORPHISM IN THE POLISH POPULATION

SNP 2 Rs3802842	AA (%)	AC (%)	CC (%)	p-value[1]	AC+CC (%)	p-value[2]	AC+AA (%)	p-value[3]
Allele frequency								
Subject group (n=311)								
CRC+ (n=121)	70 (58)	43 (35)	8 (7)	p=1.0	51 (42)	p=0.93	113 (93)	p=0.94
CRC- (n=190)	109 (57)	68 (36)	13 (7)		81 (43)		177 (93)	
Subject group (n=286)								
hMLH1 (n=171)	96 (56)	64 (37)	11 (7)	p=0.31	75 (44)	p=0.13	160 (93)	p=0.67
hMSH2 (n=115)	75 (65)	34 (30)	6 (5)		40 (35)		109 (95)	
Subject group (n=311)								
Female (n=189)	106 (56)	68 (36)	15 (8)	p=0.55	83 (44)	p=0.51	174 (92)	p=0.30
Male (n=122)	73 (60)	43 (35)	6 (5)		49 (40)		116 (95)	
Subject group (n=189)								
Gynaecological+ (n=41)	26 (64)	10 (24)	5 (12)	p=0.16	15 (36)	p=0.23	36 (88)	p=0.324*
Gynaecological- (n=148)	80 (54)	58 (39)	10 (7)		68 (46)		138 (93)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (AA) compared to combination of heterozygous and homozygous variant genotypes (AC+CC)
[3]p-value: Homozygous variant genotype (CC) compared to combination of heterozygous and wild type genotypes (AC+AA)
* p-value was calculated using Fisher's exact test.

TABLE 3- ALLELE FREQUNCY DISTRIBUTION OF THE rs10318 POLYMORPHISM IN THE POLISH POPULATION

SNP 3 Rs10318	CC (%)	CT (%)	TT (%)	p-value[1]	CT+TT (%)	p-value[2]	CT+CC (%)	p-value[3]
Allele frequency								
Subject group (n=311)								
CRC+ (n=121)	71 (59)	41 (34)	9 (7)	p=0.18	50 (41)	p=0.24	112 (93)	p=0.09
CRC- (n=190)	124 (65)	60 (32)	6 (3)		66 (35)		184 (97)	
Subject group (n=286)								
hMLH1 (n=171)	107 (63)	58 (34)	6 (3)	p=0.26	64 (37)	p=0.89	165 (97)	p=0.11
hMSH2 (n=115)	71 (62)	35 (30)	9 (8)		44 (38)		106 (92)	
Subject group (n=311)								
Female (n=189)	108 (57)	73 (39)	8 (4)	p=0.016	81 (43)	p=0.012	181 (96)	p=0.55
Male (n=122)	87 (71)	28 (23)	7 (6)		35 (29)		115 (94)	
						OR:0.536 CI:0.329-0.873		
Subject group (n=189)								
Gynaecological+ (n=41)	25 (61)	15 (37)	1 (2)	p=0.75	16 (39)	p=0.58	40 (98)	p=1.0*
Gynaecological- (n=148)	83 (56)	58 (39)	7 (5)		65 (44)		141 (95)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (CC) compared to combination of heterozygous and homozygous variant genotypes (CT+TT)
[3]p-value: Homozygous variant genotype (TT) compared to combination heterozygous and wild type genotypes (CT+CC)
* p-value was calculated using Fisher's exact test

TABLE 4- ALLELE FREQUNCY DISTRIBUTION OF THE rs4939827 POLYMORPHISM IN THE POLISH POPULATION

SNP 4 Rs4939827	CC (%)	CT (%)	TT (%)	p-value[1]	CT+TT (%)	p-value[2]	CT+CC (%)	p-value[3]
Allele frequency								
Subject group (n=311)								
CRC+ (n=121)	30 (25)	64 (53)	27 (22)	p=0.92	91 (75)	p=0.67	94 (78)	p=0.97
CRC- (n=190)	51 (27)	97 (51)	42 (22)		139 (73)		148 (78)	
Subject group (n=286)								
hMLH1 (n=171)	52 (31)	74 (43)	45 (26)	p=0.014	119 (69)	p=0.11	126 (74)	p=0.08
hMSH2 (n=115)	25 (22)	70 (61)	20 (17)		90 (78)		95 (83)	
Subject group (n=311)								
Female (n=189)	53 (28)	98 (52)	38 (20)	p=0.43	136 (72)	p=0.32	151 (80)	p=0.27
Male (n=122)	28 (23)	63 (52)	31 (25)		94 (77)		91 (75)	
Subject group (n=189)								
Gynaecological+ (n=41)	13 (32)	21 (51)	7 (17)	p=0.78	28 (68)	p=0.56	34 (83)	p=0.58
Gynaecological- (n=148)	40 (27)	77 (52)	31 (21)		108 (73)		117 (79)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (CC) compared to combination of heterozygous and homozygous variant genotypes (CT+TT)
[3]p-value: Homozygous variant genotype (TT) compared to combination heterozygous and wild type genotypes (CT+CC)

TABLE 5- ALLELE FREQUNCY DISTRIBUTION OF THE rs4464148 POLYMORPHISM IN THE POLISH POPULATION

SNP 5 Rs4464148	TT (%)	TC (%)	CC (%)	p-value[1]	TC+CC (%)	p-value[2]	TC+TT (%)	p-value[3]
Allele frequency								
Subject group (n=311)								
CRC+ (n=121)	48 (40)	47 (39)	26 (21)	p=0.11	73 (60)	p=0.90	95 (79)	p=0.053
CRC- (n=190)	74 (39)	91 (48)	25 (13)		116 (61)		165 (87)	
Subject group (n=286)								
hMLH1 (n=171)	66 (39)	75 (44)	30 (17)	p=0.96	105 (61)	p=0.95	141 (83)	p=0.82
hMSH2 (n=115)	44 (38)	52 (45)	19 (17)		71 (62)		96 (83)	
Subject group (n=311)								
Female (n=189)	78 (41)	85 (45)	26 (14)	p=0.27	111 (59)	p=0.36	163 (86)	p=0.12
Male (n=122)	44 (36)	53 (43)	25 (21)		78 (64)		97 (79)	
Subject group (n=189)								
Gynaecological+ (n=41)	17 (41)	20 (49)	4 (10)	p=0.68	24 (59)	p=0.98	37 (90)	p=0.608*
Gynaecological- (n=148)	61 (41)	65 (44)	22 (15)		87 (59)		126 (85)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (TT) compared to combination of heterozygous and homozygous variant genotypes (TC+CC)
[3]p-value: Homozygous variant genotype (CC) compared to combination heterozygous and wild type genotypes (TC+TT)
* p-value was calculated using Fisher's exact test

TABLE 6- ALLELE FREQUNCY DISTRIBUTION OF THE rs6983267 POLYMORPHISM IN THE POLISH POPULATION

SNP 6 Rs6983267	TT (%)	TG (%)	GG (%)	p-value[1]	TG+GG (%)	p-value[2]	TG+TT (%)	p-value[3]
Allele frequency								
Subject group (n=311)								
CRC+ (n=121)	29 (24)	64 (53)	28 (23)	p=0.87	92 (76)	p=0.64	93 (77)	p=0.91
CRC- (n=190)	50 (26)	95 (50)	45 (24)		140 (74)		145 (76)	
Subject group (n=286)								
hMLH1 (n=171)	40 (23)	87 (51)	44 (26)	p=0.47	131 (77)	p=0.40	127 (74)	p=0.26
hMSH2 (n=115)	32 (28)	60 (52)	23 (20)		83 (72)		92 (80)	
Subject group (n=311)								
Female (n=189)	47 (25)	92 (49)	50 (26)	p=0.30	142 (75)	p=0.79	139 (74)	p=0.12
Male (n=122)	32 (26)	67 (55)	23 (19)		90 (74)		99 (81)	
Subject group (n=189)								
Gynaecological+ (n=41)	5 (12)	20 (49)	16 (39)	p=0.039	36 (88)	p=0.034	25 (61)	p=0.039
Gynaecological- (n=148)	42 (28)	72 (49)	34 (23)		106 (72)		114 (77)	
					OR:0.35 CI:0.128-0954		OR:2.146 CI:1.029-4.477	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (TT) compared to combination of heterozygous and homozygous variant genotypes (TG+GG)
[3]p-value: Homozygous variant genotype (GG) compared to combination heterozygous and wild type genotypes (TG+TT)

TABLE 7- ALLELE FREQUNCY DISTRIBUTION OF THE rs7014346 POLYMORPHISM IN THE POLISH POPULATION

SNP 7 Rs7014346	GG (%)	GA (%)	AA (%)	p-value[1]	GA+AA (%)	p-value[2]	GA+GG (%)	p-value[3]
Allele frequency								
Subject group (n=311)								
CRC+ (n=121)	28 (23)	59 (49)	34 (28)	p=0.27	93 (77)	p=0.13	87 (72)	p=0.28
CRC- (n=190)	59 (31)	88 (46)	43 (23)		131 (69)		147 (77)	
Subject group (n=286)								
hMLH1 (n=171)	39 (23)	87 (51)	45 (26)	p=0.12	132 (77)	p=0.039	126 (74)	p=0.48
hMSH2 (n=115)	39 (34)	50 (43)	26 (23)		76 (66)		89 (77)	
					OR:0.576 CI:0.340-0.974			
Subject group (n=311)								
Female (n=189)	52 (28)	91 (48)	46 (24)	p=0.93	137 (72)	p=0.82	143 (76)	p=0.83
Male (n=122)	35 (29)	56 (46)	31 (25)		87 (71)		91 (75)	
Subject group (n=189)								
Gynaecological+ (n=41)	14 (34)	15 (37)	12 (29)	p=0.25	27 (66)	p=0.28	29 (71)	p=0.41
Gynaecological- (n=148)	38 (26)	76 (51)	34 (23)		110 (74)		114 (77)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (GG) compared to combination of heterozygous and homozygous variant genotypes (GA+AA)
[3]p-value: Homozygous variant genotype (AA) compared to combination heterozygous and wild type genotypes (GA+GG)

TABLE 8- ALLELE FREQUNCY DISTRIBUTION OF THE rs4779584 POLYMORPHISM IN THE POLISH POPULATION

SNP 8 Rs4779584	CC (%)	CT (%)	TT (%)	p-value[1]	CT+TT (%)	p-value[2]	CT+CC (%)	p-value[3]
Allele frequency								
Subject group (n=311)								
CRC+ (n=121)	66 (54)	47 (39)	8 (7)	p=0.62	55 (46)	p=0.39	113 (93)	p=0.48
CRC- (n=190)	113 (59)	68 (36)	9 (5)		77 (41)		181 (95)	
Subject group (n=286)								
hMLH1 (n=171)	98 (57)	66 (39)	7 (4)	p=0.27	73 (43)	p=0.57	164 (96)	p=0.11
hMSH2 (n=115)	62 (54)	43 (37)	10 (9)		53 (46)		105 (91)	
Subject group (n=311)								
Female (n=189)	104 (55)	75 (40)	10 (5)	p=0.47	85 (45)	p=0.26	179 (95)	p=0.87
Male (n=122)	75 (61)	40 (33)	7 (6)		47 (39)		115 (94)	
Subject group (n=189)								
Gynaecological+ (n=41)	23 (56)	17 (42)	1 (2)	p=0.65	18 (44)	p=0.88	40 (98)	p=0.693*
Gynaecological- (n=148)	81 (55)	58 (39)	9 (6)		67 (45)		139 (94)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2]p-value: Homozygous wild type (CC) compared to combination of heterozygous and homozygous variant genotypes (CT+TT)
[3]p-value: Homozygous variant genotype (TT) compared to combination heterozygous and wild type genotypes (CT+CC)
* p-value was calculated using Fisher's exact test

TABLE 9- ALLELE FREQUNCY DISTRIBUTION OF THE rs10795668 POLYMORPHISM IN THE POLISH POPULATION

SNP 9 Rs10795668	GG (%)	GA (%)	AA (%)	p-value[1]	GA+AA (%)	p-value[2]	GA+GG (%)	p-value[3]
Allele frequency								
Subject group (n=311)								
CRC+ (n=121)	54 (44.5)	54 (44.5)	13 (11)	p=0.81	67 (55.5)	p=0.73	108 (89)	p=0.53
CRC- (n=190)	81 (43)	83 (44)	25 (13)		109 (57)		165 (87)	
Subject group (n=286)								
hMLH1 (n=171)	68 (40)	77 (45)	25 (15)	p=0.28	103 (60)	p=0.18	146 (85)	p=0.21
hMSH2 (n=115)	55 (48)	49 (43)	11 (9)		60 (52)		104 (91)	
Subject group (n=311)								
Female (n=189)	82 (44)	77 (41)	29 (15)	p=0.080	107 (56)	p=0.99	160 (85)	p=0.036
Male (n=122)	53 (44)	60 (49)	9 (7)		69 (56)		113 (93)	
							OR:2.276 CI:1.037-4.994	
Subject group (n=188)								
Gynaecological+ (n=41)	17 (41)	16 (39)	8 (20)	p=0.72	24 (59)	p=0.78	33 (80)	p=0.40
Gynaecological- (n=147)	65 (44)	61 (42)	21 (14)		83 (56)		127 (86)	

CRC+ = Colorectal cancer patients, CRC- = Unaffected with colorectal cancer
Gynaecological+ = Endometrial and ovarian cancer patients, Gynaecological- = Unaffected with Endometrial and Ovarian cancer
[1]p-value: Comparison of the three genotype frequencies using Pearson's Chi-square
[2] p-value: Homozygous wild type (GG) compared to combination of heterozygous and homozygous variant genotypes (GA+AA)
[3]p-value: Homozygous variant genotype (AA) compared to combination heterozygous and wild type genotypes (GA+GG)

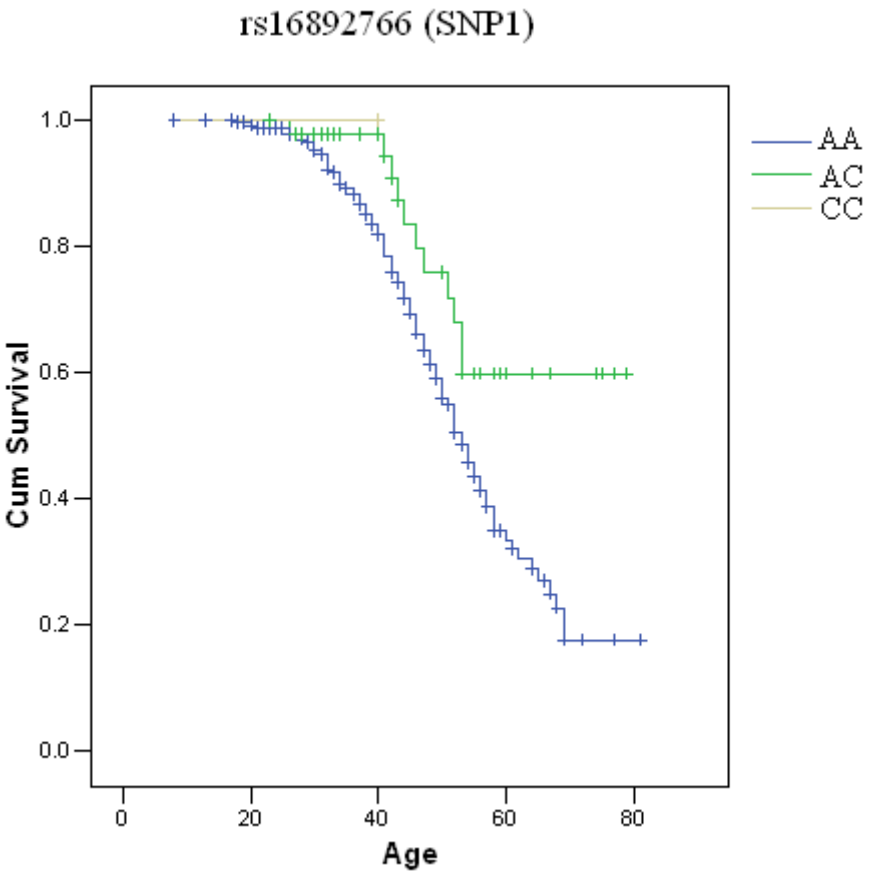


Figure 1: Kaplan-Meier estimated by rs16892766 (SNP1). The graph shows the effect the three genotypes (AA vs AC vs CC) have on age of age of diagnosis of CRC in Polish HNPCC patients. There is a statistical significant difference between the genotypes for the log-rank test: p=0.031 and the age of diagnosis of CRC. The Breslow and the Tarone-Ware tests did not show statistical significant different results.

Table 10: Median age of diagnosis of CRC (age at which 50% of the population is cancer free) in HNPCC participants for SNP rs16892766

Genotype	Subject group SNP rs16892766 [1]
Wild type (AA)	Do not reach 50% (n=252)
Heterozygote (AC)	53 yrs (n=45)
Variant (CC)	Do not reach 50% (n=1)

[1] The subject group includes 298 samples

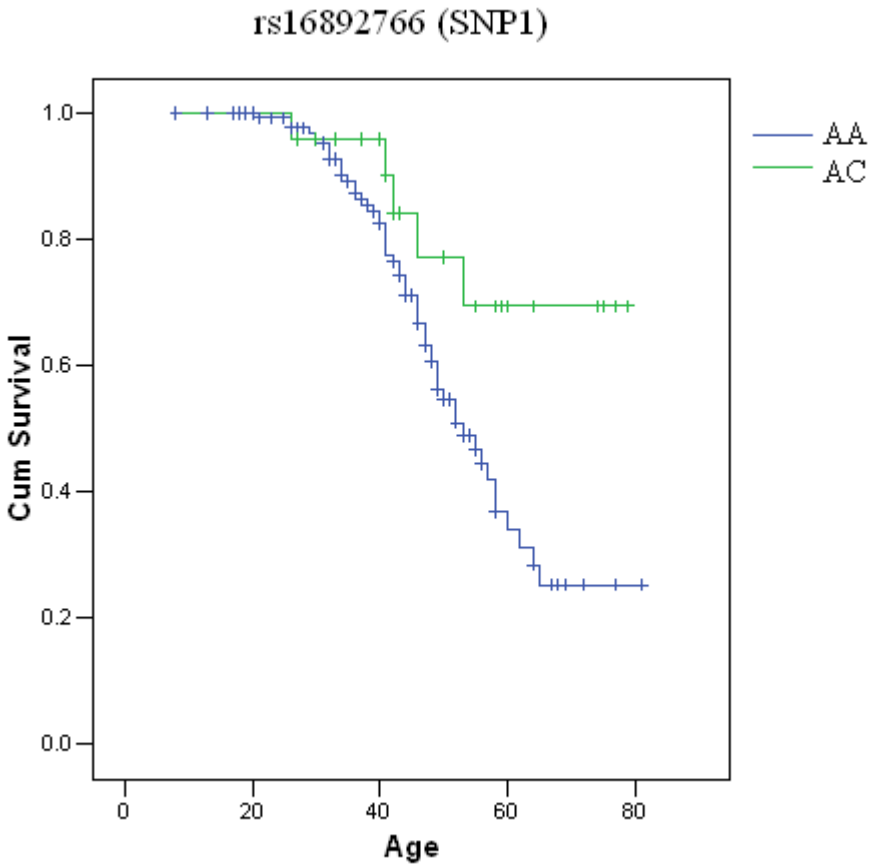


Figure 2: **Kaplan-Meier estimated by rs16892766 (SNP1) genotype in hMLH1 mutation carriers.** The graph shows the effect the three genotypes (AA vs AC vs CC) have on age of age of diagnosis of CRC in Polish HNPCC patients. There is a statistical significant difference between the genotypes for the log-rank test: p=0.023 and the age of diagnosis of CRC. The Breslow and the Tarone-Ware tests did not show statistical significant different results.

Table 11: Median age of diagnosis of CRC (age at which 50% of the population is cancer free) in HNPCC participants for SNP rs16892766

Genotype	Subject group SNP rs16892766 [1]
Homozygote wild type	Do not reach 50% (n=252)
Heterozygote + Variant	53 yrs (n=46)

[1]The subject group includes 298 samples