Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

VARIATION IN DNA REPAIR GENES XRCC3, XRCC4, and XRCC5 AND SUSCEPTIBILITY TO MYELOMA

Patrick John Hayden

A thesis submitted for the degree of Doctor in Medicine

University of Dublin

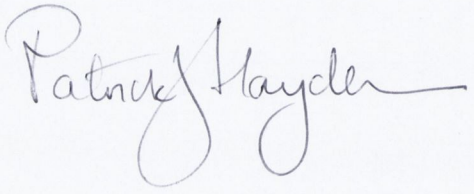
Trinity College

2010

DECLARATION

The work presented in this thesis, except where duly acknowledged, was carried out by me and has not been submitted as an exercise for a degree at this or any other university. I agree that the library may lend or copy the thesis on request.

Signed:



SUMMARY

Cytogenetic analysis in myeloma reveals marked chromosomal instability. Both widespread genomic alterations and evidence of aberrant class switch recombination, the physiological process that regulates maturation of the antibody response, implicate the DNA repair pathway in disease pathogenesis. We therefore assessed 27 SNPs in 3 genes (XRCC3, XRCC4 and XRCC5) central to DNA repair in patients with myeloma and controls from the EPILYMPH study and from an Irish hospital registry (n=306 cases, 263 controls). For the haplotype-tagging SNP (htSNP) rs963248 in XRCC4, Allele A was significantly more frequent in cases than in controls (86.4% vs. 80.8%; odds ratio (OR) 1.51; 95% confidence interval (CI) 1.10-2.08; P=0.0133), as was the AA genotype (74% vs. 65%) (P=0.026). Haplotype analysis was performed using Unphased for rs963248 in combination with additional SNPs in XRCC4. The strongest evidence of association came from the A-T haplotype from rs963248-rs2891980, (P=0.008). For XRCC5, the genotype GG from rs1051685 was detected in 10 cases from different national populations but in only 1 control (P=0.015). This SNP is located in the 3′ UTR of XRCC5. Overall, these data provide support for the hypothesis that common variation in the genes encoding DNA repair proteins contributes to susceptibility to myeloma.

ACKNOWLEDGEMENTS

I wish to acknowledge four key individuals – Dr Paul Browne, for his generous supervision and encouragement; Professor Mark Lawler, for his enthusiastic support and helpful discussions at all stages; Dr. Prerna Tewari, who worked closely with me in sample preparation, genotyping and data analysis, for her kindness, patience and scientific expertise; and Professor Anthony Staines, who had originally contacted the Haematology Department at St. James's Hospital to make us aware of the Epilymph samples and without whom this project would never have taken place.

Acknowledgments are also due to the post-graduate and post-doctoral scientists in the John Durkan Laboratories who made me welcome in Laboratory meetings and who were always available for advice. Dr. Derek Morris of the Neuropsychiatric Genetics Research Group in the Department of Psychiatry was of great assistance in the genotyping component of the study and was patient and generous with his time. I would also like to thank the following members of the Epilymph consortium – Alexandra Nieters, Nikolaus Becker, Silvia de Sanjose, Lenka Foretova, Mark Maynardie, Pier Luigi Cocco, Paolo Boffetta, Paul Brennan and Stephen Chanock.

DEDICATION

To my wife, Clare, and my parents, Aideen and Patrick Hayden

ABBREVIATIONS

ABI	Applied Biosystems Incorporated
AD	Allelic Discrimination
CEU	HapMap cohort of North-Western European Ancestry
CSR	Class Switch Recombination
DNA	Deoxyribonucleic Acid
DSB	Double Strand Break
EDTA	Ethylene-Diamine Tetra-acetic Acid
FPRP	False Positive Report Probability
HR	Homologous Recombination
GSTs	Glutathione S-transferases
GWAS	Genome Wide Association Studies
HtSNP	Haplotype-tagging Single Nucelotide Polymorphism
IARC	International Association for Research into Cancer
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency
MDA	Multiple Displacement Amplification
MGB	Minor Groove Binder
MMR	Mismatch Repair
NCBI	National Center for Biotechnology Information
NFQ	Non-Fluorescent Quencher
NHEJ	Non-homologous End-Joining
NTC	Non-Template Control
PCR	Polymerase Chain Reaction
PFS	Progression-Free Survival
QC	Quality Control
RNA	Ribonucleic Acid
SNP	Single Nucleotide Polymorphism
TNF	Tumour Necrosis Factor
UTR	Untranslated Region
WGA	Whole Genome Amplification
WHO	World Health Organisation

TABLES AND FIGURES **Page number**

CONTENTS

Chapter 1

Genetic Epidemiology and Myeloma

1.1 Inherited predisposition to cancer

The majority of cancers are likely to be polygenic in origin. Epidemiologically, consistent familial clustering is seen in many common subtypes such as cancers of the breast, bowel and lung. In general, risk is increased two- to four-fold in first-degree relatives of individuals with cancer (Goldgar, Easton *et al.* 1994). A landmark Scandinavian twin study investigating the contribution of hereditary factors to sporadic cancer found that, although genetic factors were found to make a minor contribution to overall susceptibility to most types of neoplasm, monozygotic twins of cases showed higher cancer rates than dizygotic twins or siblings, suggesting the importance of genetic susceptibility rather than environmental factors in accounting for this clustering (Lichtenstein, Holm *et al.* 2000). Much attention has focused on rare highly penetrant cancer susceptibility alleles, the best examples being BRCA1 and BRCA2 in breast cancer. Though these variants confer a high risk of disease in the individual, they are relatively rare in the population, accounting for less than 20% of the risk of breast cancer (Narod and Foulkes 2004; Foulkes 2008). It therefore appears that genetic susceptibility to most cases of cancer results from the combined effects of many common alleles (defined as allele frequency > 5%), the 'common variant, common disease' hypothesis. According to this framework, the genetic component of cancer ranges from a relatively small number of alleles of moderate risk to a larger number of alleles which each confer a minimal increase in susceptibility to cancer (Reich and Lander 2001). A meta-analysis of 301 genetic association studies supports this model (Lohmueller, Pearce *et al.* 2003).

1.2 Association studies

Where mutations in a single gene have a large effect on disease, the Mendelian inheritance pattern allows for gene mapping by family-based linkage studies. The use of this approach in the detection of moderate and low risk alleles is precluded by a number of factors including genetic heterogeneity, phenocopies and incomplete penetrance. In the setting of complex disease, a more appropriate design is the association study which allows for the comparison of the frequency of a genetic variant in individuals with and without a disease i.e. in cases and controls, respectively (Cardon and Bell 2001; Altshuler, Daly *et al.* 2008). The simple underlying assumption is that a statistically significant difference in genotype distribution supports an association between the locus and susceptibility to the disease.

1.3 Candidate gene approach

There are several methods used in the choice of candidate genes. Linkage peaks from family-based mapping studies are especially useful if a large number of multiple-case families are available. This is seldom the case in cancer studies, however, due to the age profile of those affected. Alternatively, the identification of genes at syntenic loci in animal models may prompt the investigation of their human counterparts. Genome-wide association studies (GWAS) have emerged as a powerful new approach to identifying susceptibility loci without any prior knowledge of function or position. They been performed in a number of common cancer types and have identified over 20 novel disease loci, often in pathways not suspected to play a role in oncogenesis (Easton and Eeles 2008). The risks conferred by the susceptibility alleles are low, generally 1.3-fold or less.

To date, the majority of genetic association studies have assessed a limited number of SNPs in 'candidate' genes. Candidate genes are usually chosen based on perceived biological plausibility. Studies involving cancer have concentrated on genes felt to have a key role in carcinogenesis, including those involved in cell-cycle control, apoptosis and DNA repair. In a recent meta-analysis of the contribution of candidate gene association studies to our understanding of genetic susceptibility to cancer, nearly one-third of gene-variant cancer associations were statistically significant, with variants in gene encoding for metabolizing enzymes among the most consistent and highly significant associations (Dong, Potter *et al.* 2008).

1.4 Single Nucleotide Polymorphisms

Early genetic association studies were limited by the small number of known polymorphisms. Two developments have facilitated the field of molecular epidemiological analysis. Firstly, the large-scale genotyping consortia have identified millions of single nucleotide polymorphisms (SNPs) across the human genome. Secondly, high-throughput genotyping technologies have become available to simultaneously assay these polymorphic sites.

It has been estimated that there are nearly 15 million SNPs common enough (minor allele frequency greater than 1%) to study (Botstein and Risch 2003). Most of these are concentrated around genes. Particular interest has been paid to non-synonymous coding SNPs or missense variants which lead to altered amino acid sequence within proteins. Regulatory SNPs (rSNPs) are also capable of influencing cancer risk by altering gene

expression levels. The use of these markers, however, was initially hindered by inconsistent validation across genotyping consortia and uncertainty regarding their distribution in different populations. Efforts have been made by several groups to address these concerns.

The International HapMap consortium represents a collaboration of several genotyping facilities (2003). In October 2005, the Phase 1 HapMap was released. This public database consisted of more than one million single nucleotide polymorphisms for which accurate and complete genotypes were obtained in 269 DNA samples from four populations (2005). The specific samples examined were (1) 90 individuals (30 parent-offspring trios) from the Yoruba in Ibadan, Nigeria (abbreviation YRI); (2) 90 individuals (30 trios) in Utah, USA, of North-Western European descent from the Centre d'Etude du Polymorphism Humain collection (abbreviation CEU); (3) 45 Han Chinese in Beijing, China (abbreviation CHB); (4) 44 Japanese in Tokyo, Japan (abbreviation JPT). In 2007, the Phase 2 HapMap was released, increasing the number of validated SNPs to over three million (Frazer, Ballinger *et al.* 2007).

For those working in cancer research, the Cancer Genome Anatomy Project at the National Cancer Institute (NCI) has created the SNP500cancer, a database of polymorphic variation in 500 candidate cancer susceptibility genes (Packer, Yeager *et al.* 2004). These genes have been resequenced in 102 Coriell samples drawn from 4 major ethnic groups in the United States. Data pertaining to genes and their variation are available on a public web-site (http://cgap.nih.gov). Both the HapMap and CGAP projects have allowed the genetic

epidemiology of complex inheritance to move from linkage-based studies into the association mapping of genes affecting disease (Morton 2008; Rhee and Lee 2009).

1.5 Haplotypes

Most studies to date have reported on the association between individual SNPs and the risk of a given disease. It has long been known, however, that SNPs within a gene are correlated to some degree. This is a reflection of linkage disequilibrium (LD), the phenomenon by which physically contiguous regions of a chromosome are likely to be preserved through meiosis. A haplotype is the combined set of alleles at a number of closely spaced sites on a single chromosome. SNPs which are in tight LD constitute a haplotypic block (Gabriel, Schaffner *et al.* 2002). Interest has therefore focused on identifying the allelic variants of genes in terms of haplotypes. Associations with disease outcome can then be based on haplotypes rather than individual SNPs (Fallin, Cohen *et al.* 2001; Halldorsson, Istrail *et al.* 2004). The use of haplotype block information increases the statistical power 15-50% compared with a SNP-based analysis (Thomas, Stram *et al.* 2003). Software programs have been developed to reliably generate haplotype structure from unphased genotypic data in unrelated individuals (Excoffier and Slatkin 1995; Schaid, Rowland *et al.* 2002). Stram has introduced a formal statistical measure (R_h^2) of the predictability of haplotypes based on genotypes (Stram, Haiman *et al.* 2003).

Computational algorithms are also available to select a minimal set of haplotype-tagging SNPs (ht-SNPs) to be used in genetic association studies. There are several possible statistical approaches. LDSelect uses a simple pairwise LD-based selection (Carlson, Eberle

et al. 2004). An alternative approach termed haplotype resolution infers haplotypes from genotypic data. TagSNPs (Stram 2004), HapBlock (Zhang, Qin *et al.* 2005) and HaploScore (Chapman, Cooper *et al.* 2003) are examples of this technique. A comparative assessment of the association information captured by the pairwise- and haplotype block-based SNP-tagging approaches has found them to be comparable (Nothnagel, Wollstein *et al.* 2007).

If an association is found between a haplotype and a disease, future analysis will concentrate on identifying the element within the haplotype responsible for altered gene expression or function. Studies to date suggest that a universal core set of ht-SNPs can generally be feasibly transferred between populations (Gu, Yu *et al.* 2008). Limitations are seen with rarer SNPs and in populations with recent African ancestry.

1.6 DNA repair

Severe inherited defects in DNA repair have long been known to confer an increased susceptibility to cancer. Interest in recent years has focused on the wide population variability in repair capacity phenotype, which appears to account for a several-fold variation in cancer risk. The recent completion of large-scale sequencing of population allelic variance in DNA repair genes has facilitated the investigation of their contribution to disease susceptibility. The effect of variant function in the DNA repair pathway is not limited to a few individuals with rare, high penetrance, loss-of-function alleles. Wide population variability in the capacity for repair of DNA damage has been observed. This was calculated to account for a several-fold variation in cancer risk (Berwick and Vineis

2000). Mohrenweiser and colleagues have also shown that a striking variation exists in the general DNA repair capacity phenotype such that a substantial proportion of the population have reduced repair function (60-75% of normal) and a several-fold increase in cancer risk (Mohrenweiser, Wilson *et al.* 2003). In mismatch repair specifically, a recent sequencing program identified between 6 and 8 variant alleles in 4 mismatch repair genes with an average allele frequency of 0.038 (Mohrenweiser, Xi *et al.* 2002). Meta-analyses have shown correlations between polymorphisms in repair genes and the risk of various cancers (Mitchell, Farrington *et al.* 2002; Hung, Hall *et al.* 2005; Wang, Chang *et al.* 2008). Defective DNA repair has also been implicated directly as an aetiological factor in haematological malignancy (Allan, Smith *et al.* 2004; Rudd, Sellick *et al.* 2006; Kruger, Kinzel *et al.* 2008). Vineis and colleagues recently conducted a meta-analysis of 241 putative associations between variants in DNA repair genes and cancer that had been tested in two or more independent studies (Vineis, Manuguerra *et al.* 2009). Although few strong associations were identified, they suggest gaps in evidence to be addressed in future studies.

1.7 Epidemiology of myeloma

Myeloma is a largely incurable cancer of differentiated B-cells or plasma cells which is characterised by anaemia, bone disease and renal impairment (Kuehl and Bergsagel 2005; Tosi, Gamberi *et al.* 2006; Kyle and Rajkumar 2009). It represents about 10% of haematological cancers in most registries (Morgan, Davies *et al.* 2002; Sirohi and Powles 2006). An update from the Swedish Cancer Registry suggested that disease incidence in that country has been constant for several decades (Altieri, Chen *et al.* 2006). In 2007, the

National Cancer Registry of Ireland reported 199 incident cases of myeloma, representing 0.7% of all cancers. The crude incidence rate was 4.2 per 100,000 in women and 4.97 per 100,000 in men (http://www.ncri.ie). A number of occupations and exposures have been related to myeloma. Apparent associations include farming, pesticide exposure, printing services and some specific chemical industries such as plastic and rubber.

1.8 Clinical course and treatment of myeloma

The median age of diagnosis is about 65 years and with conventional chemotherapy, a life expectancy of two to three years has been the norm. A proportion of younger patients undergo high-dose treatment with autologous stem cell rescue, a procedure that increases life expectancy from diagnosis to approximately five years (Kyle and Vincent Rajkumar 2006; Caers, Vande broek *et al.* 2008). The recent introduction of several novel therapeutic agents is likely to improve survival in the coming years. Overall outcomes nonetheless remain poor and cure a rare event. This is partially due to the incomplete understanding of the molecular mechanisms underlying disease causation and progression, and the consequent inability to develop rational biologically targeted therapy.

1.9 Biology of myeloma

Two oncogenic pathways are considered central to disease pathogenesis; hyperdiploid myeloma involves multiple trisomies of chromosomes 3, 5, 7, 9, 11, 15, 19 and 21, whereas in non-hyperdiploid myeloma, primary translocations involving the immunoglobulin heavy chain locus at 14q32 and a limited number of recurrent partner loci are present (Ho, Brown *et al.* 2001; Fonseca, Barlogie *et al.* 2004; Bergsagel and Kuehl 2005). These latter events

are felt to represent aberrant class switch recombination (CSR), a process that normally functions to alter immunoglobulin isotype with the maturation of the immune response (Fenton, Pratt *et al.* 2002; Liebisch and Dohner 2006). Such translocations result in the placing of a number of proto-oncogenes under the control of the strong immunoglobulin heavy chain enhancer. Dysregulated CSR is believed to be central to the pathogenesis of myeloma.

1.10 Class Switch Recombination

CSR is a DNA deletion-recombination event that requires formation and repair of DNA double-strand breaks (Gabrea, Leif Bergsagel *et al.* 2006; Stavnezer, Guikema *et al.* 2008). Once DNA damage sensor proteins such as p53, ATM, NBS1, BRCA1 and BRCA2 have sensed the double strand break, ligation of the cleaved ends makes use of common DNA damage repair machinery. In this regard, both the non-homologous end joining (NHEJ) and the homologous recombination (HR) pathways are utilised (Franco, Alt *et al.* 2006; Sonoda, Hochegger *et al.* 2006). The primary proteins involved in NHEJ are Ku70, Ku80 (XRCC5), DNA-Protein Kinase (catalytic subunit), XRCC4 and DNA Ligase 4. HR utilises the RAD50 family (RAD50, 51, 52, 54, etc) as well as BRCA1 and BRCA2. Variant expression and function of Ku protein has been reported in human myeloma cells (Tai, Teoh *et al.* 2000). BRCA1 and BRCA2, tumour suppressor genes which, when mutated, are associated with an increased risk of breast and ovarian cancer, function as enhancers of global genomic repair (GGR), a subset of the nucleotide excision repair pathway (Narod and Foulkes 2004). A BRCA2 mutation has been reported in the proband of a family with a high incidence of myeloma (Sobol, Vey *et al.* 2002).

Non-homologous end-joining (NHEJ) may result in the ligation of mismatched bases. A non-redundant pathway has specifically evolved to deal with this problem – DNA mismatch repair (MMR) (Li 2008). Deficiencies in this system predispose individuals to cancer (Peltomaki 2003). The obligatory mechanistic link between CSR and MMR is highlighted by the decreased efficiency of recombination seen in the presence of mutated MMR proteins (Schrader, Vardo *et al.* 2002; Jiricny 2006) .

1.11 First generation of genetic association studies

The first generation of genetic association studies was characterised by small sizes, a limited number of candidate SNPs and probable publication bias in favour of positive findings. A meta-analysis of 46 studies, which examined the effects of allelic variants of 18 different genes on breast cancer risk, reported that no definitive conclusions were possible (Dunning, Healey *et al.* 1999). The few positive findings were seldom replicated and the sample populations were small. Gonzalez and colleagues performed a similar meta-analysis on gastric cancer studies (Gonzalez, Sala *et al.* 2002). They analysed 31 articles based on 25 case-control studies carried out in Caucasian, Asian and African populations and concluded that the most important limitations precluding definite conclusions were the lack of appropriate control of potential sources of bias, the low number of cases analysed and the low number of studies offering concomitant analysis of genetic susceptibility and exposure to relevant cofactors.

Lohmueller and colleagues performed a meta-analysis of 25 inconsistent associations and 301 "replication" studies. i.e. those that attempted to replicate a prior positive result (Lohmueller, Pearce *et al.* 2003). A repeat positive finding was seen in 20% of cases, as opposed to an expected 5% under the null hypothesis. They concluded that these were genuine associations and not simply a reflection of publication bias or ethnic stratification.

In a review of association studies for finding cancer-susceptibility genetic variants, a candidate gene study in breast cancer was taken as a model (Pharoah, Dunning *et al.* 2004). This was a relatively large case-control study with 1715 cases and 2502 controls. Twenty-one haplotype-tagging SNPs were chosen after genotyping 50 SNPs spanning 109.4 kb of the BRCA2 gene. Nominally significant positive associations were observed for homozygous carriers of specific haplotypes in blocks 2 and 3 (Freedman, Penney *et al.* 2004). Though a well-designed approach, the extensive re-sequencing performed in the study population added to the complexity and cost. Dong and colleagues systematically examined the results of meta-analyses and pooled analyses for genetic polymorphisms and cancer risk published through March 2008 (Dong, Potter *et al.* 2008). They reported 98 statistically significant gene-variant cancer associations, of which 13 were considered noteworthy on the application of stricter statistical criteria.

1.12 Association studies in haematological malignancies

Researchers in leukaemia and lymphoma have studied variants in xenobiotic genes which might alter the metabolism of known carcinogens (Rothman, Wacholder *et al.* 2001; Thier, Bruning *et al.* 2003). Allelic variants in the cytochrome P450 pathway and glutathione-S-

transferases (GSTs) have been implicated in conferring an increased risk of paediatric acute lymphoblastic leukemia (Krajinovic, Labuda *et al.* 2002; Ye and Song 2005). It also appears that variants of DNA repair genes predispose to acute myeloid leukemia after previous chemotherapy (Allan, Smith *et al.* 2004; Jawad, Seedhouse *et al.* 2006; Seedhouse and Russell 2007).

1.13 Association studies in Myeloma

There have been two reports of families with clustering of cases of multiple myeloma (Lynch, Sanger *et al.* 2001; Sobol, Vey *et al.* 2002). In the apparent absence of environmental factors, these findings support the existence of an inherited element in at least some cases of the disease.

Several groups have investigated the association between genetic polymorphisms and susceptibility to myeloma. An initial report that haplotypes at the tumour necrosis factor locus on chromosome 6 may have been associated with myeloma was not borne out on further analysis (Davies, Rollinson *et al.* 2000; Morgan, Adamson *et al.* 2005). Parker and colleagues identified three polymorphisms of the IκBα gene which were more common in patients with myeloma than in controls (Parker, Ma *et al.* 2002). Subsequent work on IκBα revealed a risk haplotype [GCCTATCA] associated with myeloma (Spink, Gray *et al.* 2007). Preliminary support for an association between common variants in genes that mediate inflammation and multiple myeloma comes from a study of 78 common variants in 44 genes in Caucasian women in Connecticut (Brown, Lan *et al.* 2007). In this analysis of 126 cases and 545 population-based controls, tentative support for positive associations were noted for alleles of IL4R, a haplotype of TNF and a separate haplotype block covering

part of the LTA-TNF complex. Cozen and colleagues, in a study of 150 cases and both related (n=117) and population controls (n=131) found evidence that the -572 IL6 promoter SNP may be associated with risk of myeloma (Cozen, Gebregziabher *et al.* 2006).

On the basis of epidemiological data suggesting exposure to pesticides and solvents as risk factors for developing myeloma, xenobiotic gene polymorphisms were assessed in a case-control study of 90 Caucasians and 205 controls (Lincz, Kerridge *et al.* 2004). Patients with myeloma had a higher incidence of GST T1 null, PON1 BB and NAT2 slow acetylation genotypes. A similar study of 89 patients and 215 controls suggest that MDR1 gene SNPs may exert some effect in genetic predisposition (Jamroziak *et al*, 2004). Though lacking statistical power, both these results highlight the need for further epidemiological work on gene-environment interactions.

Only one research group has to date examined for any potential association between HLA class I or II alleles and myeloma. An analysis in South African Blacks consisting of 62 patients and 100 ethnically matched controls suggested a trend to association with HLA-18, and that a negative (protective) association existed for HLA Cw7 (Patel, Wadee *et al.* 2002).

There is very limited information on the effect of allelic variants of DNA repair genes. Two DNA Ligase IV polymorphisms were associated with a reduced risk of developing myeloma in a study of 270 patients and 220 controls (Roddam, Rollinson *et al.* 2002). As the 2 variant alleles were in strong linkage disequilibrium, the protective effect may be

linked to specific haplotypes. On the basis that genomic instability is a prominent feature in myeloma, an assessment of the proficiency of the DNA mismatch repair (MMR) pathway in clinical samples and cell lines was performed. Microsatellite analysis showed instability at one or more of nine loci examined in 15 of 92 patients (7.7% of MGUS/SMM, 20.7% of MM/plasma cell leukaemia) (Velangi, Matheson *et al.* 2004). The presence of increasing instability with disease progression, as here demonstrated, supports the need for further work on the role of allelic variance in the MMR pathway in the aetiology of myeloma.

Pharmacogenetic study of the interaction between therapeutic agents and patient outcome represents a new area of research activity. Glutathione-S-transferase (GSTP1) is a phase 2 drug metabolism enzyme involved in the metabolism and detoxification of a range of chemotherapeutic agents. The polymorphic variant - Ile105Val – has reduced enzymatic activity. In 222 patients in the UK myeloma VII trial, patients with the variant allele (105Val) had a significantly longer progression-free survival (PFS) and showed a trend to improved overall survival after treatment with melphalan (Dasgupta, Adamson *et al.* 2003). Thalidomide has anti-inflammatory and immunomodulatory functions and has been shown to inhibit tumour necrosis factor-α (TNF-α) production *in vitro*. In a study by Neben and colleagues, a SNP (-238A) in the TNF- α promoter was found to correspond both to higher pre-treatment blood TNF- α levels and a prolonged 12-month progression-free and overall survival when compared to those patients with the -238G allele (Neben, Mytilineos *et al.* 2002).

1.14 Hypothesis

Given both the widespread karyotypic abnormalities in this disease and the evidence that a process requiring DNA repair is central to the pathogenesis of myeloma, we selected 3 genes involved in DNA repair, XRCC3, XRCC4 and XRCC5, for analysis in a case control association study. XRCC3 (chromosome 14q32.3) was chosen based on its role in the HR repair pathway and on previous data suggesting an association between polymorphic variation and follicular lymphoma (Smedby, Lindgren *et al.* 2006). As both DNA Ligase IV polymorphisms and Ku protein have also been tentatively implicated in myeloma pathogenesis, XRCC4 (chromosome 5q13-q14) and XRCC5 (chromosome 2q35) were selected for assessment (Tai, Teoh *et al.* 2000; Roddam, Rollinson *et al.* 2002). A comprehensive haplotype-tagging approach was used to choose 31 SNPs across these genes for investigation. This allowed us to test a high proportion of common variation across these genes to determine if any particular SNP or haplotype confers susceptibility to myeloma.

Chapter 2

The Epilymph Study and The Irish Myeloma Database

2.1 The Epilymph study

Cases for this study were sourced from the Epilymph study and an Irish Myeloma study. The Epilymph study was carried out in six European countries (Germany, Italy, Spain, Ireland, France and the Czech Republic) from 1998 to 2003. A common core protocol and interview were used in all countries. There were 2302 incident lymphoma cases and 2417 controls.

Cases were defined as all consecutive patients who were given an initial diagnosis of lymphoid malignancy during the study period. The diagnosis of lymphoma was verified by histological testing, and 99 per cent of these tests were supplemented by immunohistochemistry tests and flow cytometry. Cases were categorised according to the World Health Organisation (WHO) Classification of Neoplastic Diseases of the Haematopoietic and Lymphoid Tissue and included all B-cell, T-cell and Natural Killer-cell neoplasms, as well as Hodgkin's lymphoma. Subjects with a diagnosis of uncertain malignant potential or monoclonal gammopathies of uncertain significance were excluded.

In Italy and Germany, controls were identified at the same time as the cases and were sampled from the general population on the basis of census lists. In the other study populations, controls were recruited from the same hospitals as the cases. In all cases, controls were matched to the cases by age (+/- 5 years), gender, and study centre. Potential hospital controls were excluded if the main reason for the hospitalisation at the time of recruitment was cancer, organ transplant, and/or systemic infection.

Informed consent was obtained from all subjects prior to enrolment, and the institutional review boards of participating centres approved the study. Cases and controls provided a blood sample for DNA extraction and serological testing and underwent a personal interview. Among cases, the participation rate was 87% (by study centre, refusal rates ranged from 7% to 18%). Among population controls, the participation rate was 75% (refusal rates by centre, 4-56%). For subjects with myeloma and matched controls, there were 142 samples from Germany, 10 from Italy, 122 from Spain, 38 from Ireland, 62 from France and 64 from the Czech Republic, giving a total of 438 samples. There were a total of 216 cases with myeloma and 222 controls in the Epilymph study.

2.2 The Irish Myeloma Study

The Irish Myeloma Study is a prospective observational single-centre study of the biology and clinical course of myeloma in Ireland. Ninety patients with myeloma have been enrolled to date. Age- and sex-matched controls for 41 of these patients were selected from the pool of controls for cases with diseases other than myeloma that had been enrolled in the Irish Epilymph study from 1998 to 2003. There was therefore no overlap of myeloma controls between the two studies. These selected controls were, as per Epilymph protocol, free from cancer, organ transplant, and/or systemic infection. This study was approved by the regional Ethics committee. Informed consent was obtained from each participant before interview and blood sampling. The diagnoses for the Irish cases were confirmed at multi-disciplinary histopathology meetings at St. James's Hospital in Dublin. The diagnoses for the control patients were confirmed through the Epilymph study. In total, therefore, the combined study consisted of 306 cases of myeloma and 263 controls.

Chapter 3

Selection of genes and SNPs

3.1 Selection of Genes and SNPs

The International HapMap Project has created a dense genome-wide haplotype map for populations of European (CEU – Utah residents with Northern and Western European Ancestry), West African (YRI – Yoruba in Ibadan, Nigeria) and East Asian (CHB - Han Chinese; JPT - Japanese) descent (2005; Frazer, Ballinger *et al.* 2007). It is an important source of validated SNPs with dense coverage of the genome and, importantly, allows for the selection of tag SNPs for use in association studies. Recent studies confirm that the HAPMAP project represents a valuable resource for choosing tag SNPs for European populations (De Bakker, Graham *et al.* 2006; Montpetit, Nelis *et al.* 2006; Gu, Yu *et al.* 2008).

SNPs from the chosen genes were identified from the International HapMap Project (http://www.hapmap.org), Public Release #20 in 2006. This release contains a remapping of the previous release (#19) on NCBI Build 35 coordinates. HapMap SNP genotyping data for the relevant genes was examined using Haploviewer Version 3.2 (http://www.broad.mit.edu/mpg/haploview) (Barrett, Fry *et al.* 2005).

Inter-marker linkage disequilibrium (LD) was measured using D' and r^2 using Haploview. Haplotype-tagging SNPs (htSNPs) were chosen using the *Tagger* (Paul de Bakker; http://www.broad.mit.edu/mpg/tagger) tag SNP selection algorithm available through Haploview (de Bakker, Yelensky *et al.* 2005).

Selection of tags is based on r^2 values between alleles of variable sites. This value is an important metric in linkage disequilibrium, being a correlation coefficient between alleles at 2 sites and a key determinant of power for association mapping. Setting the threshold to 1.0 will result in a non-redundant set of tag SNPs where all untyped SNPs will have a perfect proxy. Any value below 1.0 will result in a reduced number of SNPs while maximizing the ability to capture all identified haplotypes. *Tagger* employs both pairwise and effective haplotype predictors to capture alleles of interest. Pairwise tagging means that all tag SNPs will act as direct proxies to all other untyped SNPs because they are highly correlated with one another.

We used an r^2 threshold of 0.8 and the 2- and 3- marker haplotype 'aggressive tagging' option for tag SNP selection. *Tagger* improves tagging efficiency by aggressively searching for multi-marker predictors to capture all alleles of interest (SNPs and/or haplotypes).

The criteria for htSNP selection included Minor Allele Frequency (MAF) greater than 10%. Details of all variants genotyped at each gene in our reference panel are shown in Tables 3.1, 3.3 and 3.5. *Tagger* outputs for XRCC3, XRCC4 and XRCC5 are shown in Tables 3.2, 3.4 and 3.6, respectively.

Thirty-one ht SNPs were identified out of which 27 SNPs from 3 DSB repair genes XRCC3, XRCC4, XRCC5 were successfully genotyped.

3.2 XRCC3 SNPs

	SNP ID	ABI ID	Location	HapMap-CEU First allele	HapMap-CEU Second allele
1 (A/G)	rs861528	Custom	intron 1	27.1	72.9
2 (A/G)	rs1799794	2983904	exon 2	81.6	18.4
3 (A/G)	rs 861530	8901544	intron 5	27.5	72.5
4 (G/T)	rs 861531	8901543	intron 6	59.2	40.8
5 (A/G)	rs1799796	2983922	intron 7	66.6	33.3(*)

Table 3.1 Selection of SNPs for XRCC3
*HapMap data not available, CEPH (Centre d'Etude du Polymorphisme Humain) data

	Alleles Captured
rs1799794	rs3212103, rs12432907
rs861530	rs861537, rs861530
rs861531	rs861539, rs861534, rs861531

Table 3.2 *Tagger* output for XRCC3

3.3 XRCC4 SNPs

	SNP ID	ABI ID	Location	HapMap-CEU First allele	HapMap-CEU Second allele
1 (C/T)	rs1478486	7601977	intron 1	57.6	42.4
2 (C/G)	rs1382376	7601970	intron 1	49.1	50.9
3 (A/G)	rs1011980	7601967	intron 1	72.5	27.5
4 (A/G)	rs1011981	3098183	intron 1	55	45
5 (C/T)	rs1478483	7601964	intron 2	86.7	13.3
6 (A/G)	rs963248	7440557	intron 6	85	15
7 (A/G)	rs1193693	7440553	intron 6	51.7	48.3
8 (A/G)	rs13178127	2963038	intron 7	89.2	10.8
9 (C/T)	rs2891980	16087720	intron 7	10.8	89.2

Table 3.3 Selection of SNPs for XRCC4

Test	Alleles Captured
rs1478486	rs4266384, rs1478487, rs1478486
rs1382376	rs7736289, rs7736296, rs7718284, rs7718472, rs11750799, rs1017794, rs1580310, rs10474081 , rs1478480, rs10041617, rs1382374, rs10076059, rs1478480, rs1564211, rs6897075, rs1120476, rs6452524, rs10043018, rs6452525, rs7721416, rs6887846, rs6452526, rs7732092, rs10055844, rs10077862, rs9293332, rs6894425, rs1382372, rs1382371, rs1382370, rs1382369, rs2126988, rs4591730, rs10474094, rs1382368, rs1382366, rs6452528, rs10080123, rs13436667, rs6452531, rs6890735, rs10040363, rs4290995, rs2974444, rs965674, rs2662241, rs2662242, rs2731861, rs2940544, rs2731865, rs2974448, rs2940540, rs2731846, rs12514607
rs1011980	rs13177759, rs13163743, rs1011980, rs13180356, rs17284253
rs1011981	rs1479569, rs1564379, rs10473864, rs10942325, rs17205699, rs10057303, rs17284218, rs2386235, rs2386237, rs6452527, rs10042854, rs2386241, rs2662238, rs4343818, rs2048215, rs13156510, rs10514249, rs11949301, rs10942325, rs2731850, rs6452532, rs35256
rs1478483	rs1478483
rs963248	rs35270,rs963248, rs177297, rs35271, rs35270, rs35268, rs301279, rs301281, rs301286, rs301289, rs445403, rs369619
rs1193693	rs1193693, rs1193695
rs13178127	rs13178127
rs2891980	rs10434637, rs7734849, rs7734849, rs7736592, rs10805813, rs1805377

Table 3.4 *Tagger* output for XRCC4

3.4 XRCC5 SNPs

	SNP ID	ABI ID	Location	HapMap CEU First allele	HapMap CEU Second allele
1 (A/C)	rs828704	8839934	intron 8	80	20
2 (C/T)	rs2303400	25473226	intron 12	37.5	62.5
3 (A/G)	rs207906	2496450	Exon 14	13.3	86.7
4 (A/T)	rs207908	2406448	intron 14	49.1	50.9
5 (A/G)	rs207916	2406437	intron 16	60.2	39.8
6 (C/T)	rs207922	2773429	intron 16	62.7	37.3
7 (C/T)	rs6753002	11522958	intron 16	22.5	77.5
8 (C/T)	rs207940	8840001	intron 16	51.7	48.3
9 (A/T)	rs3770500	27498761	intron 16	10.8	89.2
10 (A/G)	rs3770493	27479143	intron 16	11.7	88.3
11 (C/T)	rs1051677	8838367	3' UTR	14.5	85.5*
12 (A/G)	rs1051685	8838368	3' UTR	88.3	11.7
13 (C/T)	rs2440	3231046	3' UTR	74.2	25.8*

Table 3.5 Selection of SNPs for XRCC5

Test	Alleles Captured
rs828704	rs828704
rs2303400	rs2303400, rs10932647
rs207906	rs207905, rs207906
rs207908	rs207882, rs207908
rs207916	rs207912, rs207927, rs207917, rs207921, rs207926, rs207927, rs207929, rs207932, rs207938,rs3770496, rs9288518, rs10490363, rs2287559, rs1438162
rs207922	rs207911, rs207922, rs207923, rs207924, rs207925, rs207936, rs3770498, rs6704622, rs12470053, rs3821104, rs3770492, rs4674066, rs7573191
rs6753002	rs3821107, rs3770507, rs6753002, rs1015679
rs207940	rs207934, rs207939, rs207940, rs207943, rs207945, rs207946
rs3770500	rs16855534, rs16855552, rs3770500, rs2241320, rs2032765
rs3770493	rs3770497, rs6729441, rs3770493, rs16855563, rs7581055, rs2241321, rs12466253, rs7583902, rs7587831, rs6747119, rs1438161, rs3835, rs3834, rs12616505, rs12617423, rs1051685
rs1051677	rs1051677
rs1051685	rs1051685
rs2440	rs2440

Table 3.6 *Tagger* output for XRCC5

Chapter 4

Material and Methods

4.1 Sources of DNA samples

Three 96-well plates containing all the Epilymph samples except for those from Germany and the majority of the samples from Spain were sent to the John Durkan laboratory from the International Association for Research into Cancer (IARC) in Lyon, France. The samples from Germany and Spain were sent subsequently from the German Cancer Research Centre in Heidelberg, Germany. These plates contained 25 microlitres of DNA at a concentration of 500ng/µl.

Epilymph DNA samples were also received from 41 age- and sex-matched controls for patients with diseases other than myeloma in the Irish Epilymph study. These were then used as controls for patients in the separate Irish Myeloma Study.

Peripheral blood samples were drawn from the 90 Irish patients in the Irish Myeloma Study. DNA extraction was performed using the QIAamp DNA blood mini kit (Quiagen, Crawley, UK). The DNA was then quantitated using the NanoDrop® ND-1000 Spectrophotometer (Thermo Scientific, Delaware, USA).

As there was a limited amount of genomic DNA and in order to facilitate future studies, whole genome amplification (WGA) was performed on all samples using the REPLI-g Kit (Quiagen, Crawley, UK). Genotyping was performed using genomic DNA for the 90 Irish samples and WGA DNA for all other samples. The protocols for DNA extraction (1.1) and DNA quantitation (1.2) are in the Appendix.

4.2 Whole Genome Amplification

For all cases and controls in the EPILYMPH study, a peripheral blood sample was collected after written informed consent was obtained. The blood was processed using centrifugation and removal of buffy coat. In order to ensure sufficient material is available for future studies, whole genome amplification (WGA) was performed on all of the Epilymph samples using the REPLI-g Kit (Quiagen, Crawley, UK). DNA was therefore used for genotyping of the Epilymph cases and controls. This product has been validated by the manufacturers for use in genotyping studies.

For the Irish samples, genomic DNA was used for all the genotyping in this study. Subsequently, given the limited quantity obtained and the need to have enough material available for participation in future collaborative studies, we proceeded to Whole Genome Amplification of genomic DNA from the Irish patients. For this purpose, REPLI-g Kits (Quiagen, Crawley, UK) were used. The kits were shipped on dry ice and were immediately stored, upon receipt, at -70°C in a constant-temperature freezer.

The REPLI-g kit contains DNA polymerase, buffers, and reagents for whole genome amplification from small samples using Multiple Displacement Amplification (MDA) (Dean, Hosono *et al.* 2002). Typical DNA yields from a REPLI-g kit reaction are approximately 40µg per 50µl reaction. The average product length is typically greater than 10kb, with a range between 2kb and 100kb.

The REPLI-g kit provides highly uniform amplification across the entire genome with negligible sequence bias (Hosono, Faruqi *et al.* 2003). The method is based on MDA technology, which performs isothermal genome amplification using a DNA polymerase capable of replicating 100kb without dissociating from the genomic DNA template. The DNA polymerase has a 3'-5' exonuclease proof-reading activity to maintain high fidelity during replication and is used in the presence of exonuclease-resistant primers to achieve high yields of DNA product.

The flow chart for the protocol is shown in Figure 4.1. The protocol for Whole Genome Amplification (1.3) is in the Appendix.

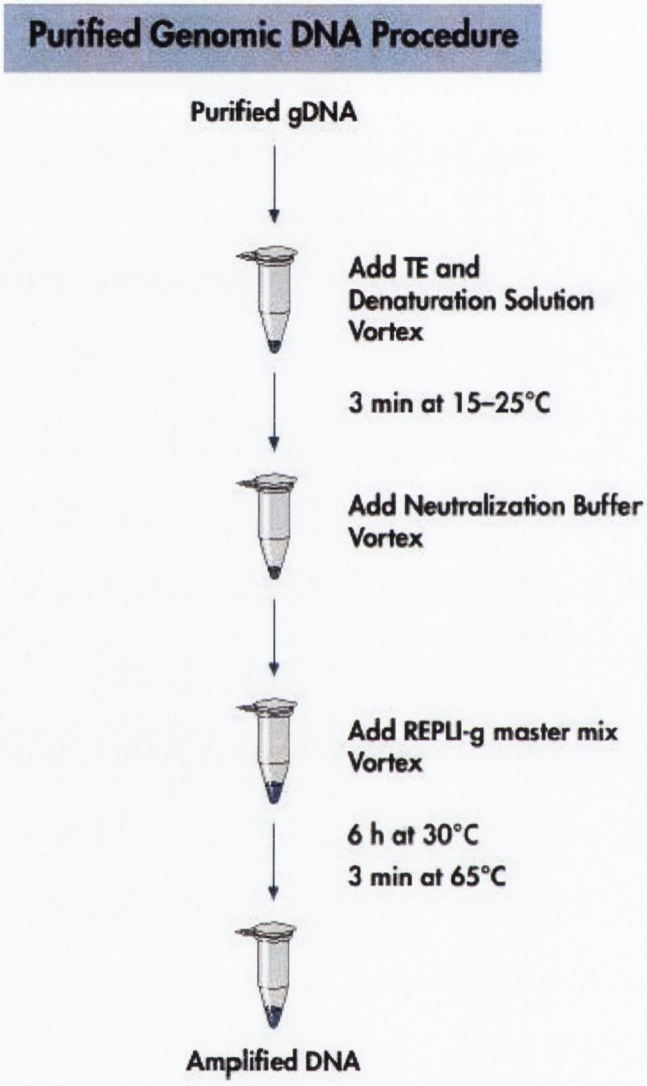


Figure 4.1 Flow Chart for Process of Whole Genome Amplification using the REPLI-g Kit (Quiagen, UK)

4.3 Statistical Analysis

The genotype frequencies were calculated for each SNP and comparisons between the observed and expected genotype frequencies in controls using the $\chi 2$ test allowed for the assessment of departures from Hardy-Weinberg equilibrium (HWE). All the genotype frequencies were in accordance with HWE.

Genotype frequencies in cases and controls were compared by χ^2 tests. The homozygous wild-type genotype was used as the reference group to calculate the odd's ratios (OR) and 95% confidence intervals (CI). The homozygous wild type genotype was also used as a referent for comparing with the heterozygote and homozygote variants (dominant model). The models were adjusted for age, sex and study centre. SNPs were also tested for association with the phenotype using a 2 x 2 (allele-wise) contingency table to calculate a χ^2 statistic.

We applied two methods - the False discovery rate (FDR) (Nyholt 2004) and the False Positive Report Probability (FPRP) (Li and Ji 2005) to address the issue of false positive SNP associations. The Benjamini-Hochberg method was used to control for FDR. FDR is defined as the expected ratio of erroneous rejections of the null hypothesis to the total number of rejected hypotheses. The FDR values were computed using the χ^2 P value for each genotype and allele frequency.

The FPRP method calculates the probability that a single SNP association is a false positive report for a range of prior probabilities, which were specified from 0.10 to 0.000001, we applied an FPRP cut-off value of 0.2 to identify which of the findings were noteworthy.

Haplotype association analysis was performed using UNPHASED (http://www.mrc-bsu.cam.ac.uk/personal/frank/software/unphased/) (Dudbridge 2003). This program calculates significance levels based on analysis of all haplotypes, described as an omnibus test. In addition, individual haplotypes can be tested for association. Given the non-independence of haplotype tests at a gene, UNPHASED also corrects for multiple testing using permutations.

4.4 Genotyping overview

Genotyping was performed using TaqMan®-based assays [Applied Biosystems (ABI), Warrington, UK] on a 7900HT ABI Sequence Detection System at the Durkan Leukaemia Laboratories in the Trinity Centre at St. James's Hospital in Dublin. The Procedure Flowchart for TaqMan SNP Genotyping Assays is shown in Figure 4.2.

Individual SNPs were genotyped using TaqMan ® Validated SNP Genotyping Assays, TaqMan® Pre-Designed SNP Genotyping Assays or Custom TaqMan® SNP Genotyping Assays.

Workflow The following diagram provides a simplified overview of the procedure for using TaqMan SNP Genotyping Assays.

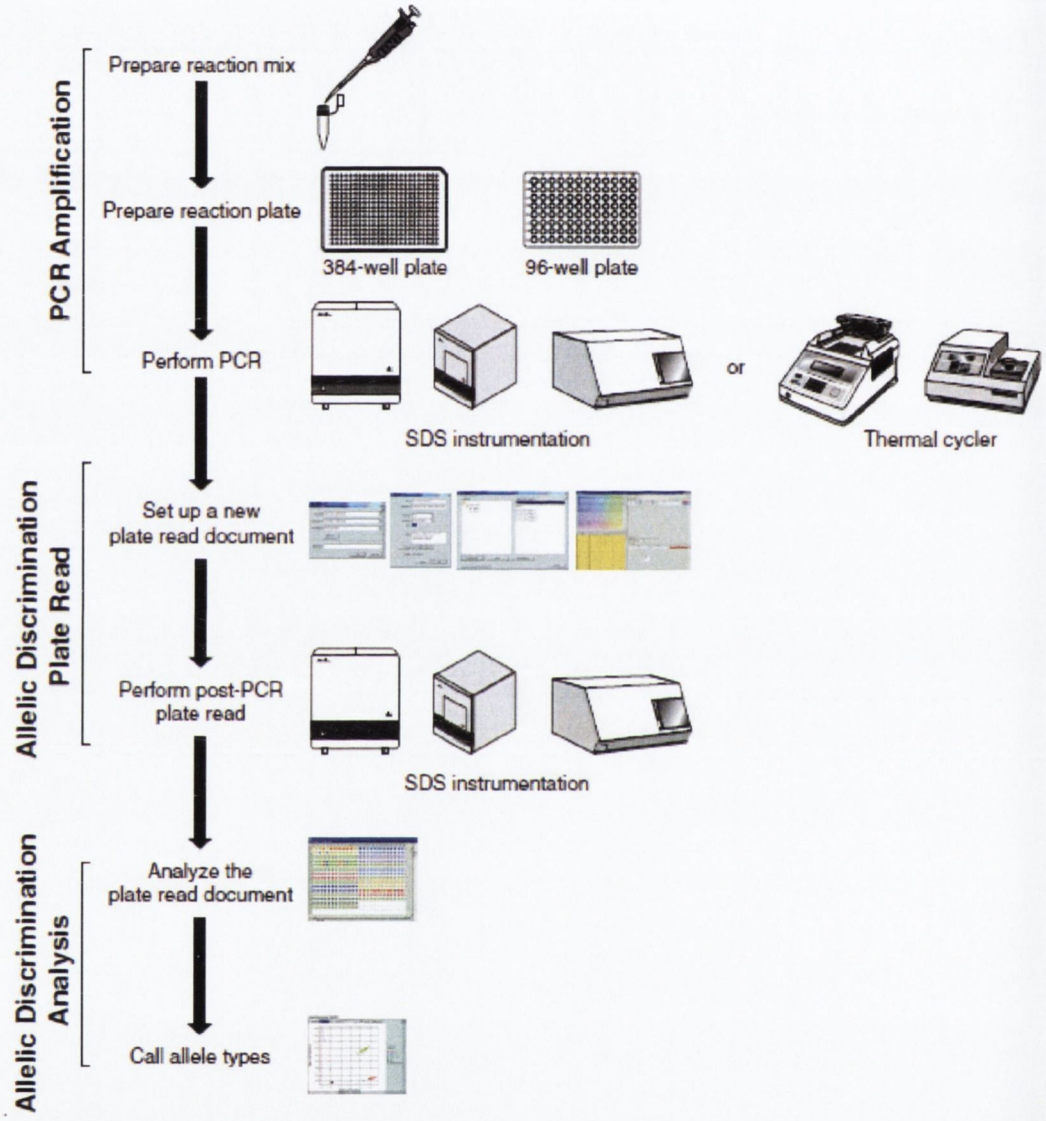


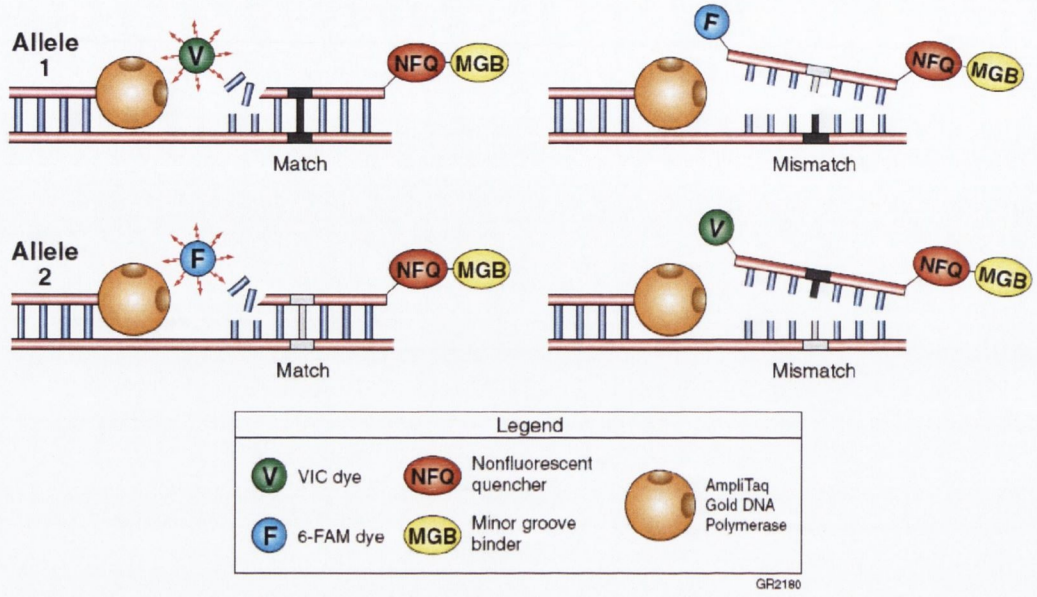
Figure 4.2 Procedure flow chart for TaqMan SNP Genotyping Assays (TaqMan® SNP Genotyping Assays Protocol - Applied Biosystems, Warrington, UK)

4.5 TaqMan SNP Genotyping Assays

Each SNP genotyping assay consists of two primers for amplifying the sequence of interest and two TaqMan MGB probes for detecting alleles (Figure 4.3). The probes contain a reported dye at the 5' end. VIC dye is linked to the 5' end of Allele 1 probe and FAM dye is linked to the 5' end of Allele 2 probe. MBG refers to Minor Groove Binder. This is a modification which increases the melting temperature (T_m) without increasing probe length, thereby allowing the design of shorter probes. This results in a greater difference in T_m values between matched and mismatched probes, which produces more accurate allelic discrimination. There is a nonfluorescent quencher (NFQ) at the 3' end of the probe which allows for a more accurate measurement of reporter dye fluorescence.

5' Nuclease assay

During PCR, each TaqMan PCR probe anneals specifically to a complementary sequence between the forward and reverse primer sites (Figure 4.4). When the probe is intact, the proximity of the reporter dye to the quencher dye results in suppression of the reporter fluorescence. The AmpliTaq Gold® DNA polymerase only cleaves probes that are hybridised to the target. Cleavage separates the reporter dye from the quencher dye, resulting in increased fluorescence by the reporter. This increase in fluorescence signal occurs only if the amplified target sequence is complementary to the probe. Thus, the fluorescence signal generated by PCR amplification indicates which alleles are present in the sample.



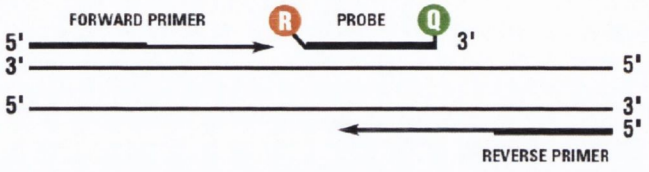
The table below shows the correlation between fluorescence signals and sequences present in the sample.

A substantial increase in...	Indicates...
VIC dye fluorescence only	homozygosity for Allele 1.
FAM dye fluorescence only	homozygosity for Allele 2.
both fluorescent signals	Allele 1-Allele 2 heterozygosity.

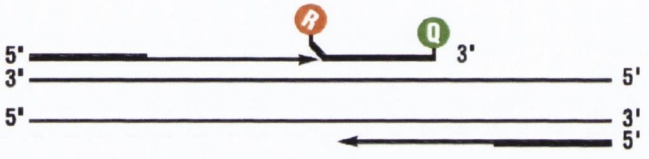
Figure 4.3 Structures of TaqMan MGB probes (TaqMan® SNP Genotyping Assays Protocol - Applied Biosystems, Warrington, UK)

TaqMan® Probe-Based Assay Chemistry

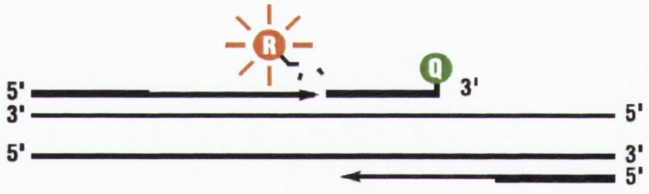
1. Polymerization: A fluorescent reporter (R) dye and a quencher (Q) are attached to the 5' and 3' ends of a TaqMan® probe, respectively.



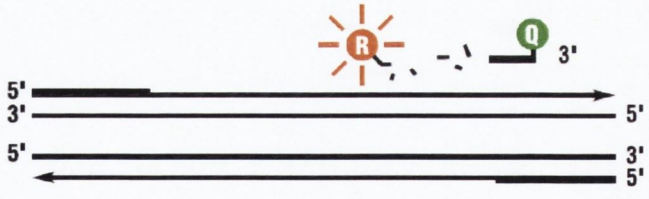
2. Strand displacement: When the probe is intact, the reporter dye emission is quenched.



3. Cleavage: During each extension cycle, the DNA polymerase cleaves the reporter dye from the probe.



4. Polymerization completed: Once separated from the quencher, the reporter dye emits its characteristic fluorescence.



www.appliedbiosystems.com

Figure 4.4 Chemistry of 5' Nuclease assay (TaqMan® SNP Genotyping Assays Protocol - Applied Biosystems, Warrington, UK)

The protocols for sample preparation and plating (1.4), the Realtime PCR run (1.5), the preparation of the AD file (1.6) and the Post-Read (1.7) are in the Appendix.

Figure 4.5 displays a typical Allelic Discrimination (AD) Plot showing clustering of allele calls based on the intensity of expression of the relevant labelling dye (FAM or VIC).

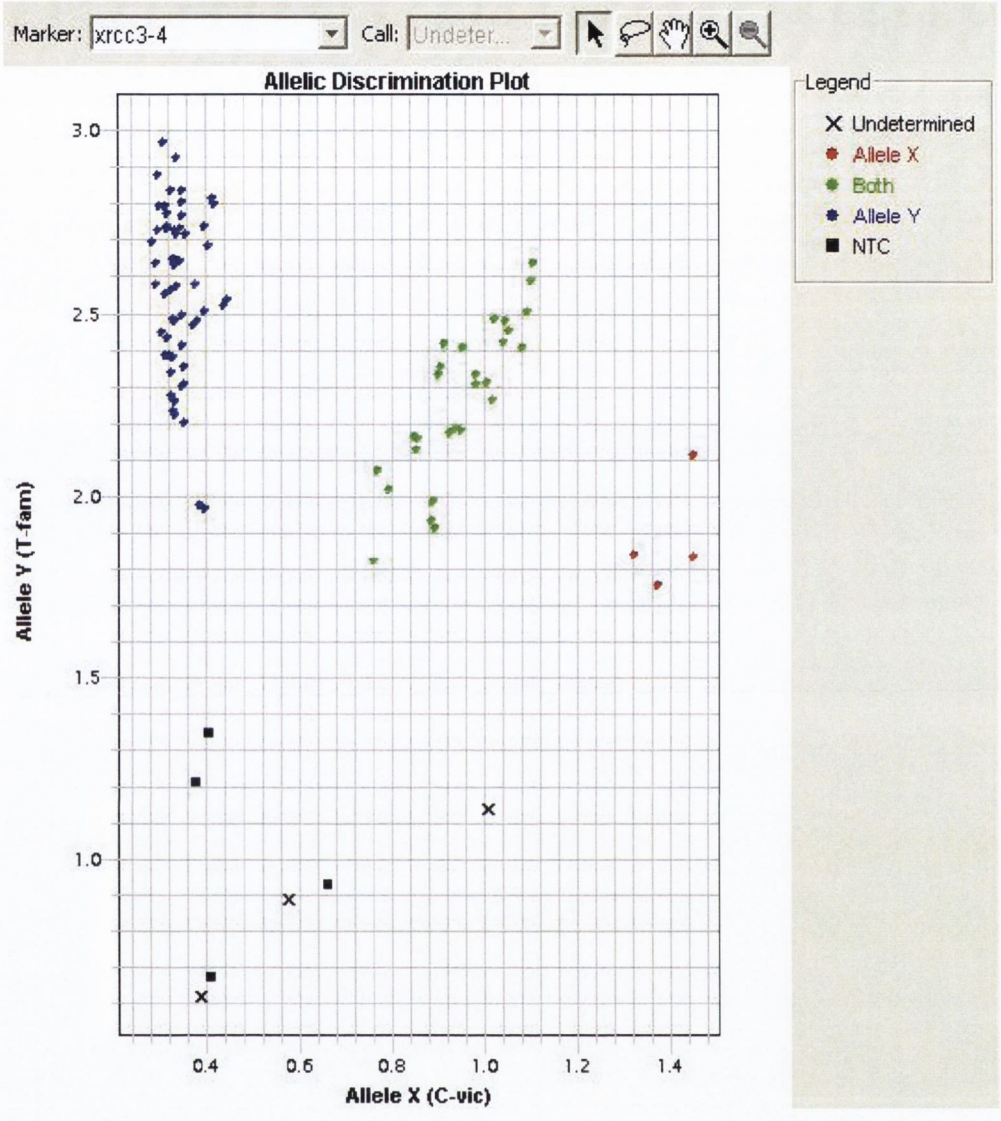


Figure 4.5 Allelic Discrimination Plot showing clustering of Allele calls based on the intensity of expression of the relevant labelling dye (FAM or VIC). This is data for XRCC3-4.

4.6 Genotyping and Quality Control

All assays were carried out in a 384-well format. For quality control (QC) purposes, duplicates of 10% of the samples were interspersed throughout the plates. The Allelic Discrimination Sequence Detection software (Applied Biosystems) was used to determine the genotypes, the genotyping was repeated if there was any discrepancy in the allele calls. For all the assays there was >99% concordance between duplicate samples. The drop-out rate for amplification or allele call was consistently < 3%.

Chapter 5

Genotyping results of SNPs in selected DNA repair genes

5.1 Result Overview

We have successfully genotyped 27 htSNPs in three DSB repair genes XRCC3, XRCC4 and XRCC5 in myeloma patients and controls from the Epilymph Study and from an Irish hospital registry (306 cases, 263 controls in total). 31 htSNPs were originally identified by *Tagger* but four assays failed design.

All SNPs were found to be in Hardy-Weinberg equilibrium ($p>0.01$).A random drop-out rate for amplification or allele call was observed. This drop-out rate was consistently less than 3% in all assays.

For quality control (QC) purposes, duplicates of 10% of the samples were interspersed throughout the plates. The concordance rates for QC samples were greater than 99% for all assays, indicating very low genotyping error.

5.2 Results for XRCC3

XRCC3 is a member of the RecA/Rad51-related protein family that participates in homologous recombination.

Details of the results of single marker association tests for each XRCC3 SNP chosen for analysis in our full case-control study population are presented in Tables 5.1 and 5.2.

There was no significant difference in allele or genotype frequencies between cases and controls for any of the 5 SNPs tested across XRCC3.

Gene	db SNP ID	Alleles	Cases n (%)	Controls n (%)	OR (95% CI)	*Allele P value*
XRCC3	rs861528	G	453(75)	393(76.8)	0.91(0.69-1.19)	0.5398
		A	151(25)	119(23.2)		
XRCC3	rs1799794	A	478(79.10	397(77.3)	1.12(0.84-1.5)	0.4867
		G	126(20.9)	117(22.7)		
XRCC3	rs861530	G	418(68.5)	346(66.8)	1.08(0.84 -1.40)	0.5788
		A	192(31.5)	172(33.2)		
XRCC3	rs861531	G	376(61.8)	336(65.1)	0.87(0.68-1.10)	0.2832
		T	232(38.2)	180(34.9)		
XRCC3	rs1799796	A	423(70)	351(67.5)	1.13(0.87-1.45)	0.3954
		G	181(30)	169(32.5)		

Table 5.1 Allele frequencies and overall associations with DNA repair gene XRCC3 alleles for multiple myeloma cases and controls. Significant associations are in **bold**.

Gene Polymorphism rs number	Genotype	Cases n (%)	Controls n (%)	OR	95% CI	*P value*	*P-value*
XRCC3 rs861528	GG	171 (56.6)	149 (58.2)	1.00	ref	-	0.615
	AG	111 (36.8)	95 (37.1)	0.98	0.69-.40	0.9916	
	AA	20 (6.2)	12 (4.7)	0.68	0.32-1.46	0.4266	
	AG/AA	131 (43.3)	107 (41.8)	0.94	0.67-1.31	0.7716	
XRCC3 rs1799794	AA	189 (62.6)	153 (59.5)	1.00	ref	-	0.742
	AG	100 (33.1)	91 (35.4)	1.12	0.79-1.60	0.5787	
	GG	13 (43.0)	13 (5.1)	1.23	0. 56-2.74	0.7524	
	AG/GG	113 (37.4)	104 (40.5)	1.14	0.80-1.60	0.5155	
XRCC3 rs861530	GG	144 (47.2)	119 (45.6)	1.00	ref	-	0.712
	AG	130 (42.6)	108 (41.7)	1.00	0.71-1.43	0.9765	
	AA	31 (10.2)	32 (12.3)	1.25	0.72-2.17	0.5142	
	AG/AA	161 (52.8)	140 (54.0)	1.05	0.76-1.47	0.8290	
XRCC3 rs 861531	GG	118 (38.8)	109 (42.2)	1.00	ref	-	0.496
	GT	140 (46.1)	118 (45.7)	0.91	0.64-1.30	0.6809	
	TT	46 (15.1)	31 (12.0)	0.73	0.43-1.23	0.2947	
	GT/TT	186 (61.2)	149 (42.8)	0.87	0.62-1.22	0.4592	
XRCC3 rs 1799796	AA	150 (49.7))	113 (43.5)	1.00	ref	-	0.216
	AG	123 (40.7)	125 (48.0)	1.35	0.95-1.91	0.1105	
	GG	29 (9.6)	22 (8.5)	1.00	0.55-1.84	0.9819	
	AG/GG	152 (50.3)	147 (56.5)	1.29	0.92-1.80	0.1658	

Table 5.2 Genotype frequencies and overall associations with XRCC3 genotypes for multiple myeloma cases and controls. Adjustment models included age, sex and study centre. Significant associations are in **bold**.

5.3 Results for XRCC4

The XRCC4 protein forms a complex with DNA ligase IV and DNA-dependent protein kinase in the repair of DNA double-strand breaks by non-homologous end joining.

Nine SNPs across XRCC4 were tested for evidence of association. The most significant result was for rs963248 where allele A was significantly more common in cases compared to controls (86.4% vs. 80.8%; odds ratio (OR) 1.51; 95% confidence interval (CI) 1.10-2.08; P=0.0133).

Genotype analysis also showed evidence of association (p=0.026) with the AA genotype in excess in cases compared to controls (74% vs. 65%).

Gene	db SNP ID	Alleles	Cases n (%)	Controls n (%)	OR (95% CI)	*Allele P value*
XRCC4	rs1478486	C	356(58.9)	305(59.1)	0.99(0.78-1.26)	0.9545
		T	248(41.8)	211(40.9)		
XRCC4	rs1382376	C	349(57.6)	285(55.2)	1.10(0.87-1.39)	0.4631
		G	257(42.4)	231(44.8)		
XRCC4	rs1011980	A	431(70.7)	377(72.5)	0.91(0.70-1.18)	0.5363
		G	179(29.3)	143(27.5)		
XRCC4	rs1011981	A	356(58.4)	306(58.8)	0.98(0.77-1.24)	0.9167
		G	254(41.6)	214(41.2)		
XRCC4	rs1478483	C	546(89.5)	453(87.1)	1.26(0.87-1.82)	0.2464
		T	64(10.5)	67(12.9)		
XRCC4	rs963248	A	529(86.4)	417(80.8)	1.51(1.10-2.08)	**0.0133**
		G	83(13.6)	99(19.2)		
XRCC4	rs1193693	A	318(52.6)	279(54.5)	0.93(0.73-1.18)	0.5789
		G	286(47.4)	233(45.5)		
XRCC4	rs1317812	A	570(94.4)	493(95.2)	0.85(0.50-1.45)	0.6408
		G	34(5.6)	25(4.8)		
XRCC4	rs2891980	T	553(91.6)	457(88.2)	1.45(0.98-2.14)	0.079
		C	51(8.4)	61(11.8)		

Table 5.3 Allele frequencies and overall associations with DNA repair gene XRCC4 alleles for multiple myeloma cases and controls. Significant associations are in **bold**.

Gene Polymorphism rs number	Genotype	Cases n (%)	Controls n (%)	OR	95% CI	*P value*	*P-value*
XRCC4 rs1478486	CC	104 (34.4)	94 (36.4)	1.00	ref	-	0.6798
	CT	148 (49.0)	117 (45.3)	0.88	0.60-1.27	0.5378	
	TT	50 (16.6)	47 (18.2)	1.04	0.64-1.69	0.9728	
	CT/TT	198 (65.6)	164 (63.6)	0.92	0.65-1.30	0.6862	
XRCC4 rs 1382376	GG	103 (33.9)	88 (34.2)	1.00	ref	-	0.315
	CG	143 (47.2)	109 (42.2)	0.89	0.61-1.30	0.6207	
	CC	57 (18.8)	61 (23.6)	1.25	0.79-1.98	0.3989	
	CG/CC	200 (66.0)	170 (65.9)	0.99	0.70-1.41	0.9771	
XRCC4 rs1011980	AA	151 (49.5)	136 (52.3)	1.00	ref	-	0.7855
	AG	129 (42.3)	105 (40.3)	0.90	0.64-1.28	0.6281	
	GG	25 (8.2)	19 (7.3)	0.84	0.44-1.60	0.7201	
	AG/GG	154 (50.5)	124 (47.7)	0.89	0.64-1.25	0.5626	
XRCC4 rs1011981	AA	107 (35.1)	100(38.5)	1.00	ref	-	0.3815
	AG	142(46.5)	106(40.8)	0.80	0.55-1.16	0.2742	
	GG	56(18.4)	54(20.8)	1.03	0.65-1.63	0.9884	
	AG/GG	198 (64.9)	160(61.5)	0.86	0.61-1.21	0.4572	
XRCC4 rs 1478483	CC	245(80.3)	200(76.9)	1.00	ref	-	0.3907
	CT	56(18.4)	53(20.4)	1.16	0.76-1.76	0.5592	
	TT	4(1.3)	7(2.7)	2.14	0.62-7.43	0.3557	
	CT/TT	60(19.7)	60(23.0)	1.23	0.81-1.83	0.3772	
XRCC4 rs 963248	AA	226(73.9)	167(64.7)	1.00	ref	-	**0.0256**
	AG	77(25.1)	83(32.2)	1.46	1.01-2.11	0.05	
	GG	3(1.0)	8(3.1)	3.61	0.94-13.8	0.0915	
	AG/GG	80(26.1)	91(35.3)	1.54	1.07-2.20	**0.024**	
XRCC4 rs13178127	AA	270(89.4)	234(90.3)	1.00	ref	-	0.8950
	AG	30(9.9)	25(9.7)	0.98	0.76-1.27	0.8906	
	GG	2(0.7)	0 (0)	0.23	0.01-4.83	0.5460	
	AG/GG	32(10.6)	25 (9.7)	0.90	0.52-1.57	0.8192	
XRCC4 rs11193693	AA	76(25.2)	82(31.7)	1.00	ref	-	0.0567
	AG	166(55.0)	115(44.4)	0.64	0.43-0.95	**0.0341**	
	GG	60(19.8)	59(22.8)	0.91	0.57-1.47	0.7943	
	AG/GG	226(74.8)	174(67.9)	0.71	0.49-1.03	0.0892	
XRCC4 rs 2891980	TT	251(83.1)	202(78)	1.00	ref	-	**0.0478**
	CT	51(16.9)	53(20.5)	1.29	0.84-1.98	0.2861	
	CC	0(0)	4(1.5)	11.17	0.59-208.98	**0.0406**	
	CT/CC	51(16.9)	57(22.0)	1.39	0.91-2.12	0.1539	

Table 5.4 Genotype frequencies and overall associations with XRCC4 genotypes for multiple myeloma cases and controls. Adjustment models included age, sex and study centre. Significant associations are in **bold**.

5.4 Results for XRCC5

XRCC5 encodes the 80-kilodalton subunit of the Ku heterodimer protein, the DNA-binding component of the DNA-dependent protein kinase.

No single marker at XRCC5 reached nominal significance levels. However, an interesting finding was noted for rs1051685 where the GG genotype was found in 10 cases from the different national populations but only in 1 control. This suggested a recessive model and was tested with Fisher's Exact Test and found to be significant (p=0.014). This SNP is located in the 3' UTR of XRCC5.

The models were adjusted for age, sex and study centre and no differences were seen.

Gene	db SNP ID	Alleles	Cases n (%)	Controls n (%)	OR (95% CI)	*Allele P value*
XRCC5	rs828704	A	480(78.4)	408(78.5)	0.99(0.75-1.33)	0.9902
		C	132(21.6)	112(21.5)		
XRCC5	rs2303400	T	347(56.9)	313(60.6)	0.85(0.67-1.08)	0.2223
		C	263(43.1)	203(39.3)		
XRCC5	rs207906	G	538(88.5)	456(87.7)	1.08(0.75-1.54)	0.7499
		A	70(11.5)	64(12.3)		
XRCC5	rs207908	A	319(52.3)	249(48.3)	1.17(0.93-1.49)	0.1967
		T	291(47.7)	267(51.7)		
XRCC5	rs207916	A	333(54.6)	304(59.4)	0.82(0.65-1.04)	0.1209
		G	277(45.4)	208(40.6)		
XRCC5	rs207922	C	398(65.2)	317(61.2)	1.19(0.94-1.52)	0.1657
		T	212(34.8)	201(38.8)		
XRCC5	rs6753002	T	487(79.8)	406(77.8)	1.13(0.85-1.50)	0.4396
		C	123(20.2)	116(22.2)		
XRCC5	rs207940	C	326(53.4)	257(49.8)	1.16(0.91-1.46)	0.2473
		T	284(46.6)	259(50.2)		
XRCC5	rs3770500	T	560(91.8)	467(91.2)	1.08(0.71-1.64)	0.8046
		A	50(8.2)	45(8.8)		
XRCC5	rs3770493	G	538(88.5)	462(89.2)	0.93(0.64-1.35)	0.7812
		A	70(11.5)	56(10.8)		
XRCC5	rs1051677	T	561(91.7)	475(91)	1.08(0.72-1.64)	0.7684
		C	51(8.3)	47(9)		
XRCC5	rs1051685	A	541(88.7)	464(88.9)	0.98(0.68-1.42)	0.9904
		G	69(11.3)	58(11.1)		
XRCC5	rs2440	C	352(57.5)	327(63.1)	0.79(0.62-1.00)	0.0631
		T	260(42.5)	191(36.9)		

Table 5.5 Allele frequencies and overall associations with DNA repair gene XRCC5 for multiple myeloma cases and controls. Significant associations are in **bold**.

Gene Polymorphism rs number	Genotype	Cases n (%)	Controls n (%)	OR	95% CI	*P value*	*P-value*
XRCC5 rs 828704	AA	192(62.7)	163(62.7)	1.00	ref	-	0.997813
	AC	96(31.4)	82(31.5)	1.00	0.70-1.44	0.9735	
	CC	18(5.9)	15(5.8)	0.98	0.48-2.01	0.9595	
	AC/CC	114(37.3)	97(37.3)	1.00	0.71-1.41	0.9897	
XRCC5 rs2303400	TT	97(31.8)	91(35.3)	1.00	ref	-	0.376677
	CT	153(50.2)	131(50.8)	0.91	0.63-1.32	0.6957	
	CC	55(18.0)	36(13.6)	0.70	0.42-1.16	0.2068	
	CT/CC	208(68.2)	167(64.7)	0.86	0.60-1.21	0.4356	
XRCC5 rs 207906	GG	243(79.4)	200(76.9)	1.00	ref	-	0.299593
	AG	54(17.6)	56(21.5)	1.26	0.83-1.91	0.3281	
	AA	9(2.9)	4(1.5)	0.54	0.16-1.78	0.4565	
	AG/AA	63(20.6)	60(23.0)	1.16	0.78-1.72	0.5398	
XRCC5 rs207908	TT	81(26.5)	70(27.1)	1.00	ref	-	0.847781
	AT	157(51.5)	127(49.2)	0.94	0.63-1.39	0.8213	
	AA	67(22.0)	61(23.6)	1.05	0.66-1.69	0.9234	
	AT/AA	224(73.4)	188(72.9)	0.97	0.67-1.41	0.9539	
XRCC5 rs 207916	AA	97(31.8)	88(34.4)	1.00	ref	-	0.11304
	AG	139(45.6)	128(50)	1.01	0.69-1.48	0.9379	
	GG	69(22.6)	40(15.6)	0.64	0.39-1.04	0.0903	
	AG/GG	208(68.2)	168(65.6)	0.89	0.63-1.27	0.5788	
XRCC5 rs 207922	CC	132(43.3)	95(37)	1.00	ref	-	0.281351
	CT	134(43.9)	127(49.4)	1.32	0.92-1.89	0.1569	
	TT	39(12.8)	37(14.4)	1.32	0.78-2.22	0.3647	
	CT/TT	173956.7)	164(63.8)	1.31	0.94-1.85	0.1320	
XRCC5 rs 6753002	TT	201(65.9)	159(61.9)	1.00	ref	-	0.316159
	CT	85(27.9)	88(33.7)	1.31	0.91-1.88	**0.0173**	
	CC	19(6.2)	14(5.4)	0.93	0.45-1.91	0.9922	
	CT/CC	104(34.1)	102(39.7)	1.21	0.88-1.75	0.2541	
XRCC5 rs 207940	CC	92(30.2)	59(22.9)	1.00	ref	-	0.11802
	CT	142(46.6)	139(53.9)	1.52	1.02-2.28	**0.0493**	
	TT	71(23.2)	60(23.2)	1.32	0.82-2.11	0.3076	
	CT/TT	213(69.8)	199(77.1)	1.46	0.99-2.13	0.0641	
XRCC5 rs3770500	TT	259(84.9)	215(84.0)	1.00	ref	-	0.940862
	AT	42(13.8)	37(14.4)	1.06	0.66-1.71	0.9029	
	AA	4(1.3)	4(1.6)	1.21	0.29-4.88	0.7937	
	AT/AA	46(15.1)	41(16.0)	1.07	0.68-1.69	0.8515	
XRCC5 rs 3770493	GG	244(80.3)	205(79.0)	1.00	ref	-	0.074405
	AG	50(16.4)	52(20.0)	1.24	0.80-1.90	0.3882	
	AA	10(3.3)	2(1.0)	0.24	0.05-1.09	0.0740	
	AG/AA	60(19.7)	54(20.8)	1.07	0.68-1.69	0.8242	
XRCC5 rs 1051677	TT	260(85.0)	218(83.5)	1.00	ref	-	0.868561
	CT	41(13.4)	39(15.0)	1.13	0.71-1.82	0.6885	
	CC	5(1.6)	4(1.5)	0.95	0.25-3.60	0.9447	
	CT/CC	46(15.0)	43(16.5)	1.12	0.71-1.75	0.7227	
XRCC5 rs 1051685	AA	246(81)	204(78.2)	1.00	ref	-	0.015143
	AG	49(16)	56(21.5)	0.73	0.47-1.11	0.1705	

	GG	10(3)	1(0. 4)	8.30	1.05-65.35	**0.0269**	
	AG/GG	59(19.4)	57(21.8)	0.86	0.57-1.29	0.5297	
XRCC5	CC	105(34.3)	101(39.0)	1.00	ref	-	0.098756
rs 2440	CT	142(46.4)	125(48.3)	0.92	0.63-1.32	0.7004	
	TT	59(19.3)	33(12.7)	0.58	0.35-0.96	**0.0443**	
	CT/TT	201(65.7)	158(61.0)	0.82	0.58-1.15	0.2871	

Table 5.6 Genotype frequencies and overall associations with XRCC5 for multiple myeloma cases and controls. Adjustment models included age, sex and study centre. Significant associations are in **bold**.

5.5 False Positive Reporting Probability value for detected association in XRCC4

The FPRP value for XRCC4 rs963248 (allelic frequency) was notable with a <20% chance of being a false positive.

Gene/SNP rs	*Odds ratio(95%CI)*	*Statistical power*	*Reported P value*	*Prior probability*				
				0.1	*0.01*	*0.001*	*0.0001*	*0.00001*
XRCC4 963248 A-G	1.50(1.10-2.08)	0.484	0.0133[a]	**0.192**	0.724	0.964	0.996	1.000
XRCC4 963248 A-G	1.54(1.07-2.20)	0.488	0.024[b]	0.307	0.829	0.980	0.998	1.000

[a] χ^2 *P* value for difference in allele frequencies between myeloma cases and controls

[b] χ^2 *P* value for difference in genotype frequencies

Table 5.7 False positive report probability values for an association between a variant in XRCC4 and multiple myeloma

Chapter 6

Analysis of haplotyping results in XRCC3, XRCC4 and XRCC5

6.1 Results overview

A haplotype is a set of statistically associated SNPs. The identification of a limited number of 'tagging' alleles in a given haplotypic block can unambiguously identify all other polymorphic sites in the region.

Haploview is a bioinformatics software tool which has been designed to perform analysis of raw genetic data and visualise patterns of linkage disequilibrium (LD). Haploview also provides functionality for performing association studies, choosing tagSNPs and estimating haplotype frequencies.

LD plots and tables showing haplotype block frequencies were calculated for the three genes under study using *Haploviewer*. The LD plots and overall haplotype block frequencies for XRCC3, XRCC4 and XRCC5 are shown below in Figures 6.1, 6.2 and 6.3, respectively. The Haploviewer outputs for comparisons of haplotype block frequencies between cases and controls for XRCC3, XRCC4 and XRCC5 are shown below in Tables 6.1, 6.2 and 6.3, respectively.

Haplotype association analysis was performed using UNPHASED (http://www.mrc-bsu.cam.ac.uk/personal/frank/software/unphased/) (Dudbridge 2003). This program calculates significance levels based on analysis of all haplotypes, described as an omnibus test. In addition, individual haplotypes can be tested for association. Given the non-independence of haplotype tests at a gene, UNPHASED also corrects for multiple testing using permutations.

As single marker analysis using UNPHASED suggested a trend towards association with rs963248, this SNP was combined in turn with each of the 8 other SNPs to assess for haplotypic association. The strongest evidence of association came from the A - T haplotype from rs963248-rs2891980 (80.9% v 74.5%; p = 0.008).

As shown in Table 6.2, *Haploviewer* also detected a haplotype block involving rs963248 and found that this similar block involving rs963248-rs1163963 was also significantly different between cases and controls (p=0.0124).

6.2 XRCC3 results

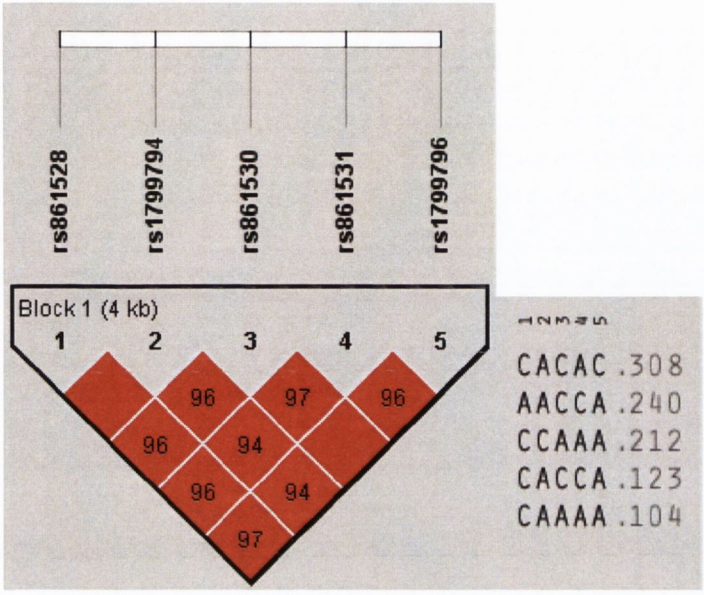
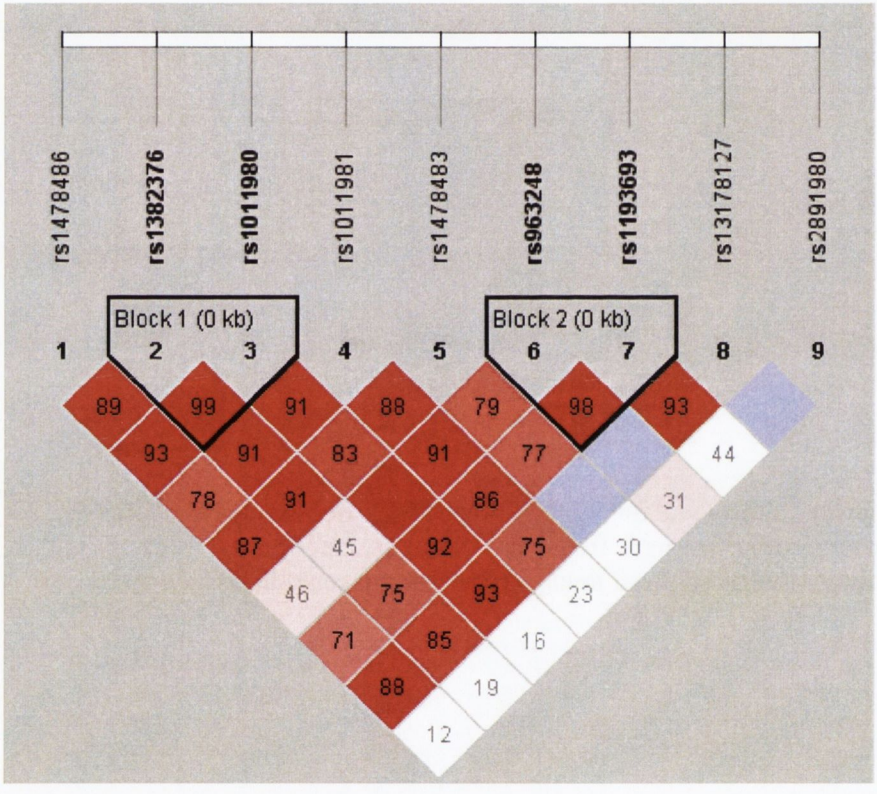


Figure 6.1 Haploviewer LD plot and overall haplotype block frequencies for XRCC3
Of note, the increasing colour strength from blue through white through pink through red correlates with an increasing strength of linkage between adjacent alleles

Block Block 1	Haplotype	Freq.	Case, Control Ratios	Chi Square	P Value
	CACAC	0.308	180.3 : 427.7, 166.8 : 351.2	0.86	0.3536
	AACCA	0.24	149.7 : 458.3, 120.1 : 397.9	0.317	0.5733
	CCAAA	0.212	121.4 : 486.6, 117.8 : 400.2	1.277	0.2584
	CACCA	0.123	80.8 : 527.2, 57.8 : 460.2	1.174	0.2787
	CAAAA	0.104	67.2 : 540.8, 50.0 : 468.0	0.582	0.4455

Table 6.1 Haploviewer output for comparison of haplotype block frequencies between cases and controls for XRCC3 (Significant results are in **bold**)

6.3 XRCC4 results



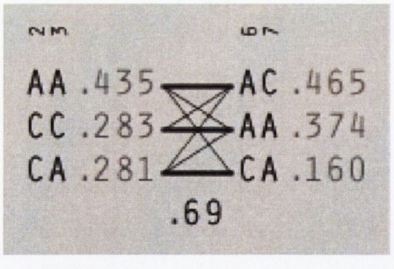
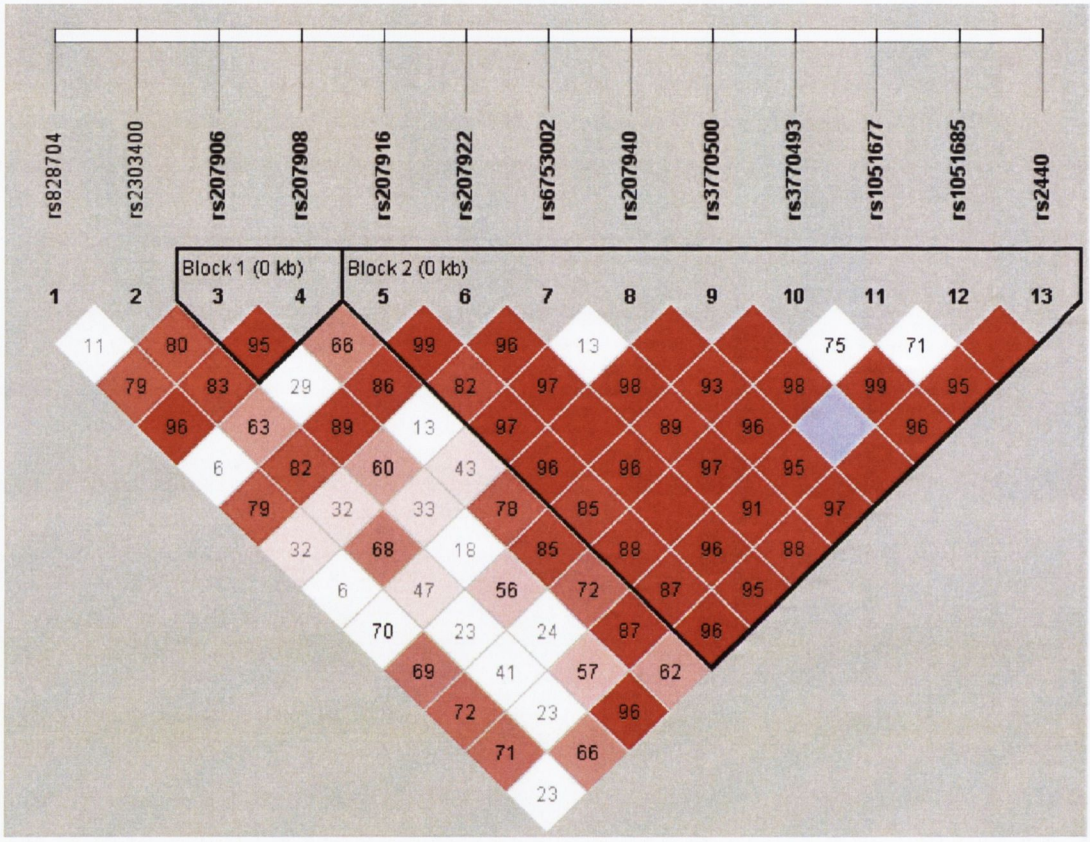


Figure 6.2 Haploviewer LD plot and overall haplotype block frequencies for XRCC4
Of note, the increasing colour strength from blue through white through pink through red correlates with an increasing strength of linkage between adjacent alleles

Block	Haplotype	Freq.	Case, Control Ratios	Chi Square	P Value
	Block 1				
	AA	0.435	258.2 : 351.8, 233.3 : 286.7	0.73	0.393
	CC	0.283	177.8 : 432.2, 141.8 : 378.2	0.484	0.4865
	CA	0.281	172.8 : 437.2, 144.7 : 375.3	0.034	0.8546
	Block 2				
	AC	0.465	290.1 : 321.9, 235.1 : 282.9	0.463	0.4964
	AA	0.374	238.9 : 373.1, 183.3 : 334.7	1.589	0.2075
	CA	0.16	82.7 : 529.3, 98.3 : 419.7	6.249	**0.0124**

Table 6.2 Haploviewer output for comparison of haplotype block frequencies between cases and controls for XRCC4 (Significant results are in **bold**)

6.4 XRCC5 results



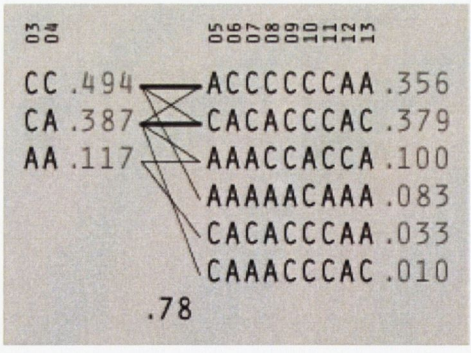


Figure 6.3 Haploviewer LD plot and overall haplotype block frequencies for XRCC5
Of note, the increasing colour strength from blue through white through pink through red correlates with an increasing strength of linkage between adjacent alleles

Block	Haplotype	Freq.	Case, Control Ratios	Chi Square	P Value
	Block 1				
	CC	0.494	291.6 : 320.4, 267.3 : 252.7	1.581	0.2086
	CA	0.387	249.7 : 362.3, 188.5 : 331.5	2.452	0.1174
	AA	0.117	70.2 : 541.8, 61.7 : 458.3	0.042	0.838
	Block 2				
	CACACCCAC	0.379	246.1 : 358.0, 177.7 : 329.5	3.803	0.0512
	ACCCCCCAA	0.356	203.9 : 400.1, 194.4 : 312.7	2.515	0.1127
	AAACCACCA	0.1	60.9 : 543.1, 50.9 : 456.2	0.001	0.9781
	AAAAACAAA	0.083	49.0 : 555.1, 44.0 : 463.1	0.114	0.7358
	CACACCCAA	0.033	18.7 : 585.3, 17.7 : 489.4	0.134	0.7138
	CAAACCCAC	0.01	4.8 : 599.3, 6.9 : 500.3	0.835	0.361

Table 6.3 Haploviewer output for comparison of haplotype block frequencies between cases and controls for XRCC5 (Significant results are in **bold**)

Chapter 7

Discussion

This is the first study to report an analysis of common variants in the three DNA DSB repair genes, XRCC3, XRCC4 and XRCC5, and the risk of developing myeloma. A comprehensive SNP-tagging approach was employed, incorporating the HAPMAP CEU reference panel data, to select SNPs across these genes that effectively capture the majority of common variants at these loci. We found that SNPs in XRCC4 and XRCC5 may alter the risk of developing myeloma. Although the risk estimates are modest, this nonetheless implicates the DNA repair pathway in disease susceptibility. Given the limited size of our study, these results need replication in larger independent samples to confirm and possibly elucidate their role in myelomagenesis.

An important strength of our study is in the use of htSNPs. The availability of comprehensive SNP frequency data through the HAPMAP consortium allows for more robust assessment of genomic regions of interest rather than simply genotyping SNPs of theoretical *a priori* significance. At XRCC3, the 5 SNPs genotyped in this study effectively tagged 11 SNPs across the gene. For XRCC4, the 9 SNPs analysed effectively tagged 116 SNPs and at XRCC5, the 13 tag SNPs selected effectively tagged 77 SNPs. Evidence is also accumulating that htSNPs selected using HAPMAP data efficiently tag for haplotypes in European populations (Montpetit, Nelis *et al.* 2006; Gu, Yu *et al.* 2008).

XRCC3 is a member of the RecA/Rad51-related protein family that participates in homologous recombination. A recent meta-analysis suggests that XRCC3 might be associated with cancer susceptibility, especially for cancer of breast, bladder, head and neck, and non-melanoma skin cancer (Han, Zhang *et al.* 2006). However, a huGE review of

XRCC3 variants in codon 241 found no definite associations between this commonly genotyped SNP and cancer (Manuguerra, Saletta *et al.* 2006). A Swedish study has suggested that rare homozygotes of three SNPs in the gene increase the risk of developing follicular lymphoma. The evidence for differences in XRCC3 haplotype distributions between follicular lymphoma cases and controls was, however, weak (Smedby, Lindgren *et al.* 2006). Our study found no difference in genotype or haplotype distributions between cases with myeloma and controls.

The XRCC4 protein forms a complex with DNA ligase IV and DNA-dependent protein kinase in the repair of DNA DSBs by non-homologous end joining. In our study, allele A of XRCC4 htSNP rs963248 was more frequent in cases than controls (86.4% vs. 80.8%; odds ratio (OR) 1.51; 95% confidence interval (CI) 1.10-2.08; P=0.0133). FastSNP predicts that this SNP is an intronic enhancer which may therefore be involved in affecting the stability of the XRCC4 mRNA transcript or altering its expression (Yuan, Chiou *et al.* 2006). Further analysis of this SNP by Transfac (http://www.gene.regulation.com/index.htm) suggests that the presence of the allele A instead of the G allele leads to loss of a GATA2 binding motif which may potentially have a pathological consequence. Additional work is required to characterise the functional aspects of this SNP and also to determine whether it is itself the high risk allele or in LD with a causal variant as on the basis of the HAPMAP dataset, rs963248 is in very high LD (r^2>0.8) with 10 other SNPs (rs177297, rs35271, rs35270, rs35268, rs301279, rs301281, rs301286, rs301289, rs445403, rs369619) which merit further investigation.

A number of groups have examined the potential role of XRCC4 in cancer susceptibility. The SNP rs2075685 is located approximately 0.5kb 5' of XRCC4 and was significantly associated (p=0.02) with the risk of breast cancer in a Taiwanese study (Fu, Yu *et al.* 2003). A recent evaluation of 4 XRCC4 htSNPs in breast cancer in Utah also found that 2 two-locus haplotypes were nominally associated with breast cancer risk (Allen-Brady, Cannon-Albright *et al.* 2006). Rs963248, which gave significant results in our study, was also one of their chosen htSNPs. In order to detect whether the Taiwanese and Utah findings were related, the Utah group assessed and found a high relative pair-wise LD (D'=0.81) between rs2075685 and rs963248. This raises the possibility that all three studies may be detecting the same haplotypic variant increasing the risk of cancer.

XRCC5 encodes the 80-kilodalton subunit of the Ku heterodimer protein, the DNA-binding component of the DNA-dependent protein kinase. A recent report suggests that a novel variable number of tandem repeats (VNTR) polymorphism containing Sp1 binding elements in the promoter of XRCC5 is a risk factor for human bladder cancer (Wang, Wang *et al.* 2008). No single marker at XRCC5 reached nominal significance levels in our study. It was found, however, that the GG genotype of Rs1051685 was found in 10 cases from the different national populations but only in 1 control. This suggested a recessive model and was tested with Fisher's Exact Test and found to be significant (p=0.014). Rs1051685 is located in the 3′ UTR of XRCC5 and may be of functional relevance since it is located in an exonic splice enhancer (ESE) sequence as determined by PupaSNP. SNPs in ESE sequences can result in exon skipping, lead to errors in alternative splicing patterns and affect mRNA stability and translation (Miller and Madras 2002). In addition, rs1051685

tags 15 other SNPs at this locus (r^2>0.8) (rs3770497, rs6729441, rs3770493, rs16855563, rs7581055, rs2241321, rs12466253, rs7583902, rs7587831, rs6747119, rs1438161, rs3835, rs3834, rs12616505, rs12617423, rs1051685) any of which can be in linkage disequilibrium with a causal variant.

Quality control is an essential component of current high-throughput technology. It has been proposed that an optimal genotyping approach should include 5-10% duplicate samples and a no-call rate less than 5% (Rebbeck, Ambrosone *et al.* 2004). Guidelines also emphasise the importance of including blind duplicates and blanks in every genotyping run. In this regard, the "360 rule" refers to the requirement for the inclusion of 24 controls in every 384-well plate. The potential for genotyping error can also be minimised by combining cases and controls during the genotyping process. In our study, 10% duplicate samples were randomly interspersed throughout the plates, there were blanks in each run and cases and controls were randomly combined. Our no-call rate was less than 2% and genotyping calls were made without knowledge of case-control status. Apart from introducing patient identification systems for biosample handling, estimation of Hardy-Weinberg proportions and tests of Hardy-Weinberg equilibrium will highlight deviations from expected results. In this regard, all 27 SNPs were found to be in Hardy-Weinberg equilibrium ($p>0.01$).

The emphasis on these safeguards stems from an appreciation that genotype misclassification can result in bias. Deitz and colleagues evaluated the effect of genotype misclassification on odds ratio estimates and sample size requirements for a study of NAT2

acetylation status, smoking, and bladder cancer risk (Deitz, Rothman *et al.* 2004). Though only a slightly smaller error rate was found using an 11-SNP assay rather than a 3-SNP assay in the assignment of NAT2 acetylation status, this resulted in a substantial decrease in sample size needed to detect a previously reported NAT2-smoking interaction for bladder cancer. This model illustrates how reducing genotype misclassification can result in substantial decreases in sample size requirements and possibly similar decreases in the cost of studies to evaluate interactions.

Controversy still exists as to how to analyse the genotypic or haplotypic data. It has been suggested that the most efficient approach remains the 'locus scoring' method (Chapman, Cooper *et al.* 2003). Tagging SNPs are analysed by a simple case-control method, removing the need for more complex haplotype-based approaches.

The inability to replicate positive study findings has been attributed to several potential causes. The three most common are false-positive reports, false-negative reports and actual population differences. False-positive reports or spurious associations due to chance (type 1 errors) are most likely. These result directly from the inappropriate use of statistical techniques. The application of a threshold for statistical significance of $p<0.05$ is now felt to be insufficiently stringent. The imposition of higher levels is needed to compensate for the multiple individual assays (multiple testing) performed in the typical association study. Another potential problem is false negatives (type 2 errors) due to small samples sizes resulting in underpowered studies. A recent review recognised this as the primary weakness of most first-generation studies (Colhoun, McKeigue *et al.* 2003). This has been a

reflection of the predominance of single-centre studies reported to date. An appreciation of the need for large sample sizes has led to the formation of consortia and the development of international networks of co-ordinated case-control studies.

A difference in the genetic backgrounds of participants represents another potential confounder. Though most likely to arise in international projects, epidemiologists have in general been alert to this possibility and have striven to ensure consistent ethnic homogeneity in study design. The majority of studies have been done in European populations. As this is the least genetically diverse group, stratification and bias are least likely to occur here (Wacholder, Rothman *et al.* 2000). In addition, Cardon and Palmer forward the argument that bias in association studies is more likely to be due to poor study design than population stratification (Cardon and Palmer 2003).

The issue of how to interpret the results of association studies remains unresolved. Wacholder and colleagues at the NCI have suggested that there are 3 potential categories into which a result can be placed; it may be 'noteworthy', an important association can be found to be absent, or finally, the evidence can be deemed insufficient to make a decision either way. In order to prevent the over-interpretation of statistically significant findings that are not likely to signify a true association, they have forwarded the concept of false positive report probability (FPRP) (Wacholder, Chanock *et al.* 2004). FPRP is based on the understanding that the probability of no true association between a genetic variant and disease given a statistically significant finding depends not only on the observed P value but also on both the prior probability that the association between the genetic variant and

the disease is real and the statistical power of the test. If the prior probability is high, the FPRP is low and the association is more likely to be correct. Though an attractive attempt to integrate what is known about the underlying pathogenesis of disease into the interpretation of results, some feel that the choice of prior probability remains relatively arbitrary given that most genes identified as noteworthy by family-based linkage studies had not been considered previously as candidate genes.

There are some potential limitations to our study. Hospital-based as opposed to population controls were used in most participating Epilymph centres. Although this is an important issue in the analysis of, for example, inflammatory gene polymorphisms, there is no *a priori* reason to assume that any given DNA repair gene alleles would be over-represented in hospital controls from which patients with cancer or systemic infection have been excluded. Nonetheless, the use of hospital controls could represent a potential confounding variable in genetic epidemiological studies. Equally, the choice of controls from the Epilymph study for the cases enrolled on the prospective Irish study should be noted. Though this is not methodologically ideal, they are derived from the same background population and are acceptable for an exploratory study aimed at producing a restricted set of hypotheses that can then be tested in studies employing more robust study designs.

In summary, we have genotyped 27 htSNPs in 3 DNA repair genes, XRCC3, XRCC4 and XRCC5. A number of SNPs show evidence of association with myeloma and are promising candidate SNPs for replication in larger studies. Even though the selected htSNPs efficiently tag variation within the available HAPMAP dataset, we cannot comment on the

potential importance of as-yet unidentified variants at these genes or SNPs located in distant regulatory regions. The investigation of rare variants at these genes was beyond the scope of this study. The significant results for XRCC4 and XRCC5 have not been adjusted for multiple hypothesis testing. Since SNPs within the same gene may be in linkage disequilibrium, the standard methods for multiple testing, such as Bonferroni correction, may be too conservative. We therefore assessed the robustness of our significant findings by the FDR and FPRP methods. The FPRP value for XRCC4 rs963248 (allelic frequency) was notable with a <20% chance of being a false positive (Table 6.3). Adjustment for FDR values did not yield significant associations. The other significant association for SNP rs 1015685 was also not noteworthy. However, it should be noted that it is unlikely that any study of our sample size can generate positive results using these more stringent criteria under the 'common disease – common variant' hypothesis. The observation that the SNP rs 963248 has also been reported in two independent breast cancer studies suggests that it might be a true disease-causing variant or possibly tagging for the true causal variant.

To conclude, we report a significant association between two regulatory SNPs in DNA double strand break repair genes XRCC4, XRCC5 and myeloma. Our findings should be considered in the context of both the strengths and limitations of the study and should be viewed as exploratory. The genetic epidemiology of rarer malignancies such as myeloma requires that promising SNPs such as those reported in the study are subjected to further analysis by large international disease-specific consortia. Multiple investigations should be pooled in order to assess the robustness of positive findings. Further evaluation of the functional relevance of identified variants and identification of the contribution of other

genes involved in the DNA double strand break repair pathways as well as potential interactions with other risk factors may eventually lead to a better understanding of myeloma pathogenesis.

Presentations/Publications

Variation in DNA repair genes XRCC3, XRCC4, and XRCC5 and risk of myeloma
Hayden PJ, Tewari P, Morris D, Staines A, Crowley D, Nieters A, Becker N, de Sanjose S, Foretova L, Maynardie M, Cocco PL, Boefetta P, Brennan P, Browne PV, Lawler M.
48th American Society for Hematology Annual Meeting, Dec. 2006, Orlando, USA
Blood 2006; Abstract 3416

Variation in DNA repair genes XRCC3, XRCC4, XRCC5 and susceptibility to myeloma
Patrick J. Hayden, Prerna Tewari, Derek W. Morris, Anthony Staines, Dominique Crowley, Alexandra Nieters, Nikolaus Becker, Silvia de Sanjose, Lenka Foretova, Marc Maynadié, Pier Luigi Cocco, Paolo Boffetta, Paul Brennan, Beth Browne, Steve J Chanock, Paul V Browne, Mark Lawler.
Human Molecular Genetics 2007 Dec 15;16(24):3117-27.
PMID: 17901044

Investigation of genetic variation in DNA repair genes and disease susceptibility to multiple myeloma by application of a candidate pathway strategy
Tewari P, Hayden PJ, Fortune A, Catherwood M, Epilymph Investigators, Chanock SJ, Staines A, Browne PV, Lawler M.
Annual Meeting of the Haematology Association of Ireland (HAI), October 2008, Armagh, Northern Ireland

References

(2003). "The International HapMap Project." Nature **426**(6968): 789-796.

(2005). "A haplotype map of the human genome." Nature **437**(7063): 1299-1320.

Allan, J. M., A. G. Smith, *et al.* (2004). "Genetic variation in XPD predicts treatment outcome and risk of acute myeloid leukemia following chemotherapy." Blood **104**(13): 3872-3877.

Allen-Brady, K., L. A. Cannon-Albright, *et al.* (2006). "A role for XRCC4 in age at diagnosis and breast cancer risk." Cancer Epidemiol Biomarkers Prev **15**(7): 1306-1310.

Altieri, A., B. Chen, *et al.* (2006). "Familial risks and temporal incidence trends of multiple myeloma." Eur J Cancer **42**(11): 1661-1670.

Altshuler, D., M. J. Daly, *et al.* (2008). "Genetic mapping in human disease." Science **322**(5903): 881-888.

Barrett, J. C., B. Fry, *et al.* (2005). "Haploview: analysis and visualization of LD and haplotype maps." Bioinformatics **21**(2): 263-265.

Bergsagel, P. L. and W. M. Kuehl (2005). "Molecular pathogenesis and a consequent classification of multiple myeloma." J Clin Oncol **23**(26): 6333-6338.

Berwick, M. and P. Vineis (2000). "Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review." J Natl Cancer Inst **92**(11): 874-897.

Botstein, D. and N. Risch (2003). "Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease." Nat Genet **33 Suppl**: 228-237.

Brown, E. E., Q. Lan, *et al.* (2007). "Common variants in genes that mediate immunity and risk of multiple myeloma." Int J Cancer **120**(12): 2715-2722.

Caers, J., I. Vande broek, *et al.* (2008). "Multiple myeloma--an update on diagnosis and treatment." Eur J Haematol **81**(5): 329-343.

Cardon, L. R. and J. I. Bell (2001). "Association study designs for complex diseases." Nat Rev Genet **2**(2): 91-99.

Cardon, L. R. and L. J. Palmer (2003). "Population stratification and spurious allelic association." Lancet **361**(9357): 598-604.

Carlson, C. S., M. A. Eberle, *et al.* (2004). "Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium." Am J Hum Genet **74**(1): 106-120.

Chapman, J. M., J. D. Cooper, *et al.* (2003). "Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power." Hum Hered **56**(1-3): 18-31.

Colhoun, H. M., P. M. McKeigue, *et al.* (2003). "Problems of reporting genetic associations with complex outcomes." Lancet **361**(9360): 865-872.

Cozen, W., M. Gebregziabher, *et al.* (2006). "Interleukin-6-related genotypes, body mass index, and risk of multiple myeloma and plasmacytoma." Cancer Epidemiol Biomarkers Prev **15**(11): 2285-2291.

Dasgupta, R. K., P. J. Adamson, *et al.* (2003). "Polymorphic variation in GSTP1 modulates outcome following therapy for multiple myeloma." Blood **102**(7): 2345-2350.

Davies, F. E., S. J. Rollinson, *et al.* (2000). "High-producer haplotypes of tumor necrosis factor alpha and lymphotoxin alpha are associated with an increased risk of myeloma and have an improved progression-free survival after treatment." J Clin Oncol **18**(15): 2843-2851.

De Bakker, P. I., R. R. Graham, *et al.* (2006). "Transferability of tag SNPs to capture common genetic variation in DNA repair genes across multiple populations." Pac Symp Biocomput: 478-486.

de Bakker, P. I., R. Yelensky, *et al.* (2005). "Efficiency and power in genetic association studies." Nat Genet **37**(11): 1217-1223.

Dean, F. B., S. Hosono, *et al.* (2002). "Comprehensive human genome amplification using multiple displacement amplification." Proc Natl Acad Sci U S A **99**(8): 5261-5266.

Deitz, A. C., N. Rothman, *et al.* (2004). "Impact of misclassification in genotype-exposure interaction studies: example of N-acetyltransferase 2 (NAT2), smoking, and bladder cancer." Cancer Epidemiol Biomarkers Prev **13**(9): 1543-1546.

Dong, L. M., J. D. Potter, *et al.* (2008). "Genetic susceptibility to cancer: the role of polymorphisms in candidate genes." JAMA **299**(20): 2423-2436.

Dudbridge, F. (2003). "Pedigree disequilibrium tests for multilocus haplotypes." Genet Epidemiol **25**(2): 115-121.

Dunning, A. M., C. S. Healey, *et al.* (1999). "A systematic review of genetic polymorphisms and breast cancer risk." Cancer Epidemiol Biomarkers Prev **8**(10): 843-854.

Easton, D. F. and R. A. Eeles (2008). "Genome-wide association studies in cancer." Hum Mol Genet **17**(R2): R109-115.

Excoffier, L. and M. Slatkin (1995). "Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population." Mol Biol Evol **12**(5): 921-927.

Fallin, D., A. Cohen, *et al.* (2001). "Genetic analysis of case/control data using estimated haplotype frequencies: application to APOE locus variation and Alzheimer's disease." Genome Res **11**(1): 143-151.

Fenton, J. A., G. Pratt, *et al.* (2002). "Isotype class switching and the pathogenesis of multiple myeloma." Hematol Oncol **20**(2): 75-85.

Fonseca, R., B. Barlogie, *et al.* (2004). "Genetics and cytogenetics of multiple myeloma: a workshop report." Cancer Res **64**(4): 1546-1558.

Foulkes, W. D. (2008). "Inherited susceptibility to common cancers." N Engl J Med **359**(20): 2143-2153.

Franco, S., F. W. Alt, *et al.* (2006). "Pathways that suppress programmed DNA breaks from progressing to chromosomal breaks and translocations." DNA Repair (Amst) **5**(9-10): 1030-1041.

Frazer, K. A., D. G. Ballinger, *et al.* (2007). "A second generation human haplotype map of over 3.1 million SNPs." Nature **449**(7164): 851-861.

Freedman, M. L., K. L. Penney, *et al.* (2004). "Common variation in BRCA2 and breast cancer risk: a haplotype-based analysis in the Multiethnic Cohort." Hum Mol Genet **13**(20): 2431-2441.

Fu, Y. P., J. C. Yu, *et al.* (2003). "Breast cancer risk associated with genotypic polymorphism of the nonhomologous end-joining genes: a multigenic study on cancer susceptibility." Cancer Res **63**(10): 2440-2446.

Gabrea, A., P. Leif Bergsagel, *et al.* (2006). "Distinguishing primary and secondary translocations in multiple myeloma." DNA Repair (Amst) **5**(9-10): 1225-1233.

Gabriel, S. B., S. F. Schaffner, *et al.* (2002). "The structure of haplotype blocks in the human genome." Science **296**(5576): 2225-2229.

Goldgar, D. E., D. F. Easton, *et al.* (1994). "Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands." J Natl Cancer Inst **86**(21): 1600-1608.

Gonzalez, C. A., N. Sala, *et al.* (2002). "Genetic susceptibility and gastric cancer risk." Int J Cancer **100**(3): 249-260.

Gu, C. C., K. Yu, *et al.* (2008). "On transferability of genome-wide tagSNPs." Genet Epidemiol **32**(2): 89-97.

Halldorsson, B. V., S. Istrail, *et al.* (2004). "Optimal selection of SNP markers for disease association studies." Hum Hered **58**(3-4): 190-202.

Han, S., H. T. Zhang, *et al.* (2006). "DNA repair gene XRCC3 polymorphisms and cancer risk: a meta-analysis of 48 case-control studies." Eur J Hum Genet **14**(10): 1136-1144.

Ho, P. J., R. D. Brown, *et al.* (2001). "Illegitimate switch recombinations are present in approximately half of primary myeloma tumors, but do not relate to known prognostic indicators or survival." Blood **97**(2): 490-495.

Hosono, S., A. F. Faruqi, *et al.* (2003). "Unbiased whole-genome amplification directly from clinical samples." Genome Res **13**(5): 954-964.

Hung, R. J., J. Hall, *et al.* (2005). "Genetic polymorphisms in the base excision repair pathway and cancer risk: a HuGE review." Am J Epidemiol **162**(10): 925-942.

Jawad, M., C. H. Seedhouse, *et al.* (2006). "Polymorphisms in human homeobox HLX1 and DNA repair RAD51 genes increase the risk of therapy-related acute myeloid leukemia." Blood **108**(12): 3916-3918.

Jiricny, J. (2006). "The multifaceted mismatch-repair system." Nat Rev Mol Cell Biol **7**(5): 335-346.

Krajinovic, M., D. Labuda, *et al.* (2002). "Glutathione S-transferase P1 genetic polymorphisms and susceptibility to childhood acute lymphoblastic leukaemia." Pharmacogenetics **12**(8): 655-658.

Kruger, S., M. Kinzel, *et al.* (2008). "Homozygous PMS2 germline mutations in two families with early-onset haematological malignancy, brain tumours, HNPCC-associated tumours, and signs of neurofibromatosis type 1." Eur J Hum Genet **16**(1): 62-72.

Kuehl, W. M. and P. L. Bergsagel (2005). "Early genetic events provide the basis for a clinical classification of multiple myeloma." Hematology Am Soc Hematol Educ Program: 346-352.

Kyle, R. A. and S. V. Rajkumar (2009). "Criteria for diagnosis, staging, risk stratification and response assessment of multiple myeloma." Leukemia **23**(1): 3-9.

Kyle, R. A. and S. Vincent Rajkumar (2006). "Treatment of multiple myeloma: an emphasis on new developments." Ann Med **38**(2): 111-115.

Li, G. M. (2008). "Mechanisms and functions of DNA mismatch repair." Cell Res **18**(1): 85-98.

Li, J. and L. Ji (2005). "Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix." Heredity **95**(3): 221-227.

Lichtenstein, P., N. V. Holm, *et al.* (2000). "Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland." N Engl J Med **343**(2): 78-85.

Liebisch, P. and H. Dohner (2006). "Cytogenetics and molecular cytogenetics in multiple myeloma." Eur J Cancer **42**(11): 1520-1529.

Lincz, L. F., I. Kerridge, *et al.* (2004). "Xenobiotic gene polymorphisms and susceptibility to multiple myeloma." Haematologica **89**(5): 628-629.

Lohmueller, K. E., C. L. Pearce, *et al.* (2003). "Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease." Nat Genet **33**(2): 177-182.

Lynch, H. T., W. G. Sanger, *et al.* (2001). "Familial multiple myeloma: a family study and review of the literature." J Natl Cancer Inst **93**(19): 1479-1483.

Manuguerra, M., F. Saletta, *et al.* (2006). "XRCC3 and XPD/ERCC2 single nucleotide polymorphisms and the risk of cancer: a HuGE review." Am J Epidemiol **164**(4): 297-302.

Miller, G. M. and B. K. Madras (2002). "Polymorphisms in the 3'-untranslated region of human and monkey dopamine transporter genes affect reporter gene expression." Mol Psychiatry **7**(1): 44-55.

Mitchell, R. J., S. M. Farrington, *et al.* (2002). "Mismatch repair genes hMLH1 and hMSH2 and colorectal cancer: a HuGE review." Am J Epidemiol **156**(10): 885-902.

Mohrenweiser, H. W., D. M. Wilson, 3rd, *et al.* (2003). "Challenges and complexities in estimating both the functional impact and the disease risk associated with the extensive genetic variation in human DNA repair genes." Mutat Res **526**(1-2): 93-125.

Mohrenweiser, H. W., T. Xi, *et al.* (2002). "Identification of 127 amino acid substitution variants in screening 37 DNA repair genes in humans." Cancer Epidemiol Biomarkers Prev **11**(10 Pt 1): 1054-1064.

Montpetit, A., M. Nelis, *et al.* (2006). "An evaluation of the performance of tag SNPs derived from HapMap in a Caucasian population." PLoS Genet **2**(3): e27.

Morgan, G. J., P. J. Adamson, *et al.* (2005). "Haplotypes in the tumour necrosis factor region and myeloma." Br J Haematol **129**(3): 358-365.

Morgan, G. J., F. E. Davies, *et al.* (2002). "Myeloma aetiology and epidemiology." Biomed Pharmacother **56**(5): 223-234.

Morton, N. E. (2008). "Into the post-HapMap era." Adv Genet **60**: 727-742.

Narod, S. A. and W. D. Foulkes (2004). "BRCA1 and BRCA2: 1994 and beyond." Nat Rev Cancer **4**(9): 665-676.

Neben, K., J. Mytilineos, *et al.* (2002). "Polymorphisms of the tumor necrosis factor-alpha gene promoter predict for outcome after thalidomide therapy in relapsed and refractory multiple myeloma." Blood **100**(6): 2263-2265.

Nothnagel, M., A. Wollstein, *et al.* (2007). "Comparative assessment of the association information captured by SNP tagging." Hum Hered **64**(1): 27-34.

Nyholt, D. R. (2004). "A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other." Am J Hum Genet **74**(4): 765-769.

Packer, B. R., M. Yeager, *et al.* (2004). "SNP500Cancer: a public resource for sequence validation and assay development for genetic variation in candidate genes." Nucleic Acids Res **32**(Database issue): D528-532.

Parker, K. M., M. H. Ma, *et al.* (2002). "Identification of polymorphisms of the IkappaBalpha gene associated with an increased risk of multiple myeloma." Cancer Genet Cytogenet **137**(1): 43-48.

Patel, M., A. A. Wadee, *et al.* (2002). "HLA class I and class II antigens associated with multiple myeloma in southern Africa." Clin Lab Haematol **24**(4): 215-219.

Peltomaki, P. (2003). "Role of DNA mismatch repair defects in the pathogenesis of human cancer." J Clin Oncol **21**(6): 1174-1179.

Pharoah, P. D., A. M. Dunning, *et al.* (2004). "Association studies for finding cancer-susceptibility genetic variants." Nat Rev Cancer **4**(11): 850-860.

Rebbeck, T. R., C. B. Ambrosone, *et al.* (2004). "SNPs, haplotypes, and cancer: applications in molecular epidemiology." Cancer Epidemiol Biomarkers Prev **13**(5): 681-687.

Reich, D. E. and E. S. Lander (2001). "On the allelic spectrum of human disease." Trends Genet **17**(9): 502-510.

Rhee, H. and J. S. Lee (2009). "MedRefSNP: a database of medically investigated SNPs." Hum Mutat **30**(3): E460-466.

Roddam, P. L., S. Rollinson, *et al.* (2002). "Genetic variants of NHEJ DNA ligase IV can affect the risk of developing multiple myeloma, a tumour characterised by aberrant class switch recombination." J Med Genet **39**(12): 900-905.

Rothman, N., S. Wacholder, *et al.* (2001). "The use of common genetic polymorphisms to enhance the epidemiologic study of environmental carcinogens." Biochim Biophys Acta **1471**(2): C1-10.

Rudd, M. F., G. S. Sellick, *et al.* (2006). "Variants in the ATM-BRCA2-CHEK2 axis predispose to chronic lymphocytic leukemia." Blood **108**(2): 638-644.

Schaid, D. J., C. M. Rowland, *et al.* (2002). "Score tests for association between traits and haplotypes when linkage phase is ambiguous." Am J Hum Genet **70**(2): 425-434.

Schrader, C. E., J. Vardo, *et al.* (2002). "Role for mismatch repair proteins Msh2, Mlh1, and Pms2 in immunoglobulin class switching shown by sequence analysis of recombination junctions." J Exp Med **195**(3): 367-373.

Seedhouse, C. and N. Russell (2007). "Advances in the understanding of susceptibility to treatment-related acute myeloid leukaemia." Br J Haematol **137**(6): 513-529.

Sirohi, B. and R. Powles (2006). "Epidemiology and outcomes research for MGUS, myeloma and amyloidosis." Eur J Cancer **42**(11): 1671-1683.

Smedby, K. E., C. M. Lindgren, *et al.* (2006). "Variation in DNA repair genes ERCC2, XRCC1, and XRCC3 and risk of follicular lymphoma." Cancer Epidemiol Biomarkers Prev **15**(2): 258-265.

Sobol, H., N. Vey, *et al.* (2002). "Re: familial multiple myeloma: a family study and review of the literature." J Natl Cancer Inst **94**(6): 461-462; author reply 463.

Sonoda, E., H. Hochegger, *et al.* (2006). "Differential usage of non-homologous end-joining and homologous recombination in double strand break repair." DNA Repair (Amst) **5**(9-10): 1021-1029.

Spink, C. F., L. C. Gray, *et al.* (2007). "Haplotypic structure across the I kappa B alpha gene (NFKBIA) and association with multiple myeloma." Cancer Lett **246**(1-2): 92-99.

Stavnezer, J., J. E. Guikema, *et al.* (2008). "Mechanism and regulation of class switch recombination." Annu Rev Immunol **26**: 261-292.

Stram, D. O. (2004). "Tag SNP selection for association studies." Genet Epidemiol **27**(4): 365-374.

Stram, D. O., C. A. Haiman, *et al.* (2003). "Choosing haplotype-tagging SNPS based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study." Hum Hered **55**(1): 27-36.

Tai, Y. T., G. Teoh, *et al.* (2000). "Ku86 variant expression and function in multiple myeloma cells is associated with increased sensitivity to DNA damage." J Immunol **165**(11): 6347-6355.

Thier, R., T. Bruning, *et al.* (2003). "Markers of genetic susceptibility in human environmental hygiene and toxicology: the role of selected CYP, NAT and GST genes." Int J Hyg Environ Health **206**(3): 149-171.

Thomas, D. C., D. O. Stram, *et al.* (2003). "Bayesian spatial modeling of haplotype associations." Hum Hered **56**(1-3): 32-40.

Tosi, P., B. Gamberi, *et al.* (2006). "Biology and treatment of multiple myeloma." Biol Blood Marrow Transplant **12**(1 Suppl 1): 81-86.

Velangi, M. R., E. C. Matheson, *et al.* (2004). "DNA mismatch repair pathway defects in the pathogenesis and evolution of myeloma." Carcinogenesis **25**(10): 1795-1803.

Vineis, P., M. Manuguerra, *et al.* (2009). "A field synopsis on low-penetrance variants in DNA repair genes and cancer susceptibility." J Natl Cancer Inst **101**(1): 24-36.

Wacholder, S., S. Chanock, *et al.* (2004). "Assessing the probability that a positive report is false: an approach for molecular epidemiology studies." J Natl Cancer Inst **96**(6): 434-442.

Wacholder, S., N. Rothman, *et al.* (2000). "Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias." J Natl Cancer Inst **92**(14): 1151-1158.

Wang, F., D. Chang, *et al.* (2008). "DNA repair gene XPD polymorphisms and cancer risk: a meta-analysis based on 56 case-control studies." Cancer Epidemiol Biomarkers Prev **17**(3): 507-517.

Wang, S., M. Wang, *et al.* (2008). "A novel variable number of tandem repeats (VNTR) polymorphism containing Sp1 binding elements in the promoter of XRCC5 is a risk factor for human bladder cancer." Mutat Res **638**(1-2): 26-36.

Ye, Z. and H. Song (2005). "Glutathione s-transferase polymorphisms (GSTM1, GSTP1 and GSTT1) and the risk of acute leukaemia: a systematic review and meta-analysis." Eur J Cancer **41**(7): 980-989.

Yuan, H. Y., J. J. Chiou, *et al.* (2006). "FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization." Nucleic Acids Res **34**(Web Server issue): W635-641.

Zhang, K., Z. Qin, *et al.* (2005). "HapBlock: haplotype block partitioning and tag SNP selection software using a set of dynamic programming algorithms." Bioinformatics **21**(1): 131-134.

Appendix

1.1 Protocol for DNA extraction

Collect a 9ml EDTA blood sample

Preparation

1 Amp DNA blood mini kit (cat. No. 51104).

2 Set heating block to 56°C.

3 Equilibrate Buffer AE to room temperature.

4 If Buffer AL has a precipitate, dissolve by incubating at 56 °C.

Procedure

1 Equilibrate sample to room temperature.

2 For plasma isolation and preparation of buffy coat,
Spin sample at 2500g for 10 minutes at room temperature.

3 3 layers are now visible:
Upper clear layer – plasma
Middle white layer – buffy coat containing concentrated leucocytes
Lower red layer – concentrated erythrocytes

4 Place 20µl of Quiagen protease into a 1.5ml eppendorf.

5 Add 4µl of RNAase A (100µg/ml) into the same 1.5ml eppendorf.

6 Aspirate upper plasma layer into a 2ml starstedt tube and label clearly.
Store plasma at -80 °C.

7 Carefully remove buffy coat layer with a Pasteur pipette. Add to the protease. If the sample volume is less than 200µl, bring volume up to 200µl with PBS. If volume is greater than 200µl but less than 400µl, increase volumes of protease and Buffer AL accordingly.

8 Add 200µl of Buffer AL. Mix by pulse vortexing for 15 seconds.
Note: mix thoroughly. Solution should be homogeneous.

9 Incubate tube at 56 °C for 10 minutes. Spin down to remove drops from lid.

10 Add 200µl of ethanol (96-100%) to the sample. Pulse vortex for 15 seconds. Spin down to remove drops from the lid. If sample volume is greater than 200µl, increase ethanol volume proportionally.

11 Carefully apply sample to QIAamp spin column in a 2 ml collection tube, without wetting the rim.

12 Centrifuge at full speed (20000g) for 1 minute.
Note: if lysate has not completely passed through column after centrifugation, centrifuge again until spin column is completely empty.

13 Transfer spin column to a fresh 2ml collection tube. Add 500µl of buffer AW1.

14 Centrifuge at 6000g for 1 minute.

15 Transfer spin column to a fresh collection tube. Add 500µl of buffer AW2.

16 Centrifuge at full speed (20000g) for 3 minutes.

17 Transfer spin column to a fresh collection tube and spin again at 20000g for 1 minute.

18 Transfer spin column to a fresh, labelled 1.5ml eppendorf tube. Add 200µl of Buffer AE. Incubate at room temperature for 5 minutes.

19 Centrifuge at 6000g for 1 minute.

20 Repeat steps 17 and 18, and combine samples.

21 Quantitate DNA. The typical yield should be 6µg. The A260/A280 ratio should be between 1.7 and 1.9.

1.2 Protocol for DNA quantification

DNA quantitation using the Nanodrop facility

1 Open Nanodrop User on desktop.

2 Open default user account.

3 Open default nucleic acid measurement (DNA or RNA).

4 Pre-set Gilson pipette to 1.1µl.

5 Perform a water-blank with 1µl H_20.

6 Leave the same tip on the Gilson pipette.

7 Remove the water and discard.

8 Clean the reader with tissue paper only between samples.

9 Start report: sample type DNA50, name trial 1

10 Place 1µl, and then click measure.

11 Record readings for A260/A280 ratio and quantity in ng/µl for all samples.

1.3 Protocol for Whole Genome Amplification

The protocol is optimised for WGA from greater than 10ng genomic DNA template.
The template DNA should be stored at a concentration of greater than 4ng/µl.
REPLI-g DNA polymerase should be stored on ice.
Buffer D1 and Buffer N1 should not be stored longer than a week.
All Buffers and reagents should be vortexed before use to ensure thorough mixing.

1 Prepare 500µl Solution A by adding 40µl 5M KOH and 10µl 0.5M EDTA (pH 8) into 450µl deionised water.
2 Solution A can be stored for 1 week at room temperature. The container must be tightly sealed to avoid neutralisation with CO_2.
3 Set a water bath or heating block to 30°C.

1 Prepare sufficient Buffer D1 (denaturation buffer) and Buffer N1 (neutralization buffer) for the total number of whole genome amplifcation reactions. Buffer D1 is made up of 5µl Solution A and 35µl nuclease-free water. Buffer N1 is made up of 8µl Solution B (Stop solution) and 72µl of nuclease-free water.

2 Place 2.5µl template DNA (>4ng/µl) into a microcentrifuge tube.

3 Add 2.5µl buffer D1 to the DNA. Mix by vortexing and centrifuge briefly.

4 Incubate the samples at room temperature for 3 minutes.

5 Add 5µl Buffer N1 to the samples. Mix by vortexing and centrifuge briefly.

6 Thaw REPLI-g DNA polymerase on ice. Thaw all other components at room temperature, vortex and centrifuge briefly.

7 Prepare a master mix on ice with the following components: nuclease-free water 27µl, 12.5µl REPLI-g buffer 4X and 0.5µl REPLI-g DNA polymerase. Mix and centrifuge briefly.

8 Add 40µl of master mix to 10µl of denatured DNA.

9 Incubate at 30°C for 6-16 hours.

10 Inactivate REPLI-g DNA polymerase by heating the sample for 3 minutes at 65°C.

11 Store amplified DNA at 4°C for short-term storage or -20°C for long-term storage.

1.4 Protocol for sample preparation and plating

Preparation of samples and assay mixes, and plating of 384-well plate

1 Prepare 5 2ml Starstedt tubes prior to loading the robot.

2 The first 4 2ml tubes are for each of the 4 individual SNP assays.

3 The final 2ml tube is for water for the Non-Template Controls (NTC).

For 96 reactions, prepare sufficient material for 105 reactions for each SNP.

Each reaction is made up of:

DNA (20ng/microl. Approx.) 3 µl x 105 (In Plate)

Sterile Water	0.25 µl	x 105	26.25 µl
TaqMan Mastermix	2.5 µl	x 105	262.5 µl
20XSNP Assay	0.25 µl	x 105	26.25 µl
Total	6 µl		315 µl

1 Add water, then TaqMan Mastermix and finally the SNP assay to total mix.

2 Vortex and spin the Starstedt tubes down in the centrifuge to ensure there are no bubbles prior to placement in the robot.

3 Bring 1 plate and 5 Starstedt tubes down to Robot.

4 Check Sterile Water supply for robot is above upper black line.

5 Check Vircon tray is full (should be changed daily).

6 Remove parafilm and cling film from plate.

7 Place uncovered plate in Position P2.

8 Place 384-well plate in Position P3.

9 Place the 4 SNPs, SNP1 to SNP4, in positions 13, 14, 15 and 16 respectively.

10 Ensure caps are removed from Starstedt tubes.

11 Place the 5th Starstedt tube with water for the NTC in position 12.

12 Initialise the robot.

13 Open File.

14 Open Myeloma folder (name). Open script.

15 Double-click on green triangle key.

16 Double-click on each of the confirmation prompts (important).
The robot run takes 75 minutes.

1.5 Realtime PCR run

1 When the robot run is complete, remove the 384-well plate and cover the plate with the optical adhesive cover.

2 Remove the 96-deepwell plate and cover with cling film. Then place the adherent top firmly onto the deepwell plate and cover it all with parafilm.

3 Double-click on the SDS 2.1 icon on the desktop.

4 Double-click on file, the New.

5 Double-click on New.

6 Choose Absolute Quantitation from the drop-down menu.

7 Choose 384-well option from the drop-down menu.

8 Double-click on OK.

9 Double-click on Add Detector.

10 Choose FAM-NFQ and copy to plate document.

11 Choose VIC-NFQ and copy to plate document.

12 Double-click on Done.

13 Highlight entire plate by double-clicking on the top left-hand corner of the plate.

14 Place an X beside both dyes (FAM-NFQ and VIC-NFQ).

15 Proceed to Instrument tab.

16 Change Sample Volume to 6 microlites.

17 Disable 9600 emulation.

18 Save As: Myeloma group, with Realtime suffix.
(copy previous file name and modify)

19 Double-click on Connect (software to 7900HT hardware).

20 Double-click on Open/Close.

21 Place 384-well plate in 7900HT, ensuring that A1 is at top left hand corner.

22 Double-click on Open/Close again to close.

23 Double-click on Start.

24 Wait 2 minutes to ensure run is proceeding correctly.

25 The Realtime PCR run takes 90 minutes.

1.6 Preparation of AD file

N.B. This preparation for the AD run can be done straight after commencing the Realtime PCR run so as to minimise the time required for the final Post Read.

1 Open File, then New, choose AD from dropdown menu, double-click on OK.

2 Import Sample sheet.

3 Open File, then Import, then Myeloma Group, then Sample Sheet.

4 Add SNP Markers.

5 Highlight Markers.

6 Copy to Plate document.

7 If there are New SNP assays, then write name and identify markers (Mondays)

8 Identify where in plate each marker will be as follows:

9 Highlight first 6 columns for the first of the 4 SNPs and click the relevant marker with X.

10 Identify the 16 NTC by holding down the Ctrl key and double-clicking on the relevant row.

11 On right hand-side, change unknown to NTC four times.

12 Open File. Save As – change Realtime to AD (Allelic Discrimination).

1.7 Post read

1 Double-click on Window.

2 Open AD document.

3 Open Instrument.

4 Double-click on Connect.

5 Double-click on Post Read.

6 This takes 30 seconds.

7 Double-Click on Analysis and Settings button (this is the button with the small green triangle and a blue line).

8 Choose All Markers.

9 Double-click on Auto Caller enabled.

10 Double-click on OK.

11 Double-click on Analyse button (Large green triangle only).

12 Double-click on results tab.

13 Double-click on File, then Export, then place results suffix to previous Realtime file.

14 Open in Excel.