Genetic determinants of response to aspirin and warfarin and development of silicon nanowire based genotyping

Harsh Jayesh Sheth

Doctor of Philosophy (Ph.D.)

Institute of Genetic Medicine

Faculty of Medical Sciences

Newcastle University

October 2015

Abstract

Chronic diseases such as cardiovascular diseases and colorectal cancer are the leading cause of mortality worldwide. Commonly used drugs such as aspirin and warfarin are shown to effective at reducing the risk of chronic diseases but have a narrow therapeutic window and are associated with adverse drug reactions, particularly, hemorrhage. Identification of pharmacogenetic markers such as single nucleotide polymorphisms (SNPs) that could help deliver personalized dose could help improve the risk-benefit ratio. Furthermore, development of a rapid point of care genotyping device consisting of a pharmacogenetic SNP panel for aspirin and warfarin could help implement personalized medicine in the clinical setting.

Analysis of candidate SNPs in aspirin's pharmacokinetic and pharmacodynamic pathways was carried out to explain variation in aspirin's colorectal chemopreventive efficacy using two large population based case-control datasets. Associations and interactions were tested using logistic regression models and meta-analysis of the 2 datasets. A novel site-specific association for rs1799853 (OR=0.73, 95% CI=0.60-0.90, P=0.003) and rs1105879 (OR=1.16, 95% CI=1.02-1.32, P=0.03) with colon cancer risk was observed. Furthermore, stratification by aspirin use showed increased risk of colorectal cancer in aspirin users but not in non-users carrying variant allele of the SNPs rs4936367 and rs7112513 in *PAFAH1B2* gene and rs2070959 and rs1105879 in *UGT1A6* gene ($P_{interaction}<0.05$ for all). These results provide insight into aspirin's differential chemopreventive efficacy and the neoplastic transformation of cells in colon and rectum.

Utility of clinically validated pharmacogenetic dosing algorithms consisting of three warfarin dose associated SNPs from the European population needs to tested in the Gujarati Indians, an Indian sub-population. Dose prediction accuracy of the algorithms was compared between Gujarati Indian and European population. Mean squared difference of both pharmacogenetic algorithms was higher in Gujarati Indian compared to European population (Klein et al 2009, 216.3 v/s 160.7, P=0.05; Gage et al 2008, 170.6 v/s 143.2, P=0.07). Poor prediction accuracy could be explained by the presence of study subjects requiring dose for target INR range 2.5-3.5 and low frequency of the *VKORC1* rs9923231 variant, which is the most important genetic determinant of warfarin dosing in Europeans. Therefore, the SNP panel and dosing algorithms developed from European populations cannot be assumed to have utility in the Gujarati Indian population.

Finally, to help develop a rapid, point-of care, silicon nanowire (SiNW) based SNP genotyping device, a panel of isothermal melting probes were designed to genotype three warfarin dose associated SNPs. Testing of hybridization and washing conditions to have optimal hybridization kinetics between the probe and target DNA and high target sequence specificity was carried out using custom designed microarray platform. Accurate genotype calls for all 3 SNPs in 2 anonymised samples using empirically optimized hybridization and washing conditions was carried out successfully. Current work highlighted associations between probe characteristics and hybridization parameters, which would be useful in designing and testing probes on the SiNW platform.

Identification, validation and testing of clinical utility of population specific pharmacogenetic markers along with development and deployment of ultra-rapid point of care genotyping technologies could help deliver personalized risk-benefit ratio for aspirin and warfarin.

Dedication

To all those patients whose treatment and health could improve as a result of the work carried out and published within this thesis.

Acknowledgements

I express my deepest gratitude and thanks to my PhD supervisors- Prof. Sir John Burn, Dr. Mike Jackson, Dr. Mauro Santibanez-Koref and Prof. Ann Daly for their guidance, support, vision and wisdom throughout the duration of my PhD study. I have been privileged to work under their supervision, as I have not only learnt how to conduct science but have also learnt the true purpose and meaning of science itself.

Sincere thanks to the collaborators- Prof. D. Timothy Bishop and Prof. Cornelia Ulrich for the aspirin pharmacogenetics project; Dr. Jayesh Sheth, Dr. Frenny Sheth and Dr. Keyur Parikh for the warfarin pharmacogenetics project and; Mr. Jonathan O'Halloran, Dr. Sam Whitehouse and Dr. Stephen Osborne for the QuantuMDx's silicon nanowire project.

A round of thanks goes to Dr. John Tyson and Dr. Emma Northwood who have played a pivotal role in supervising during the planning of experiments, data analysis and result interpretation.

To all my fellow colleagues, comrades and friends (old and new) at Newcastle University, Leeds University and QuantuMDx Ltd., I express my gratitude for their generous support, enthusiasm and kinship during the tumultuous journey of this PhD degree. A special kudos to my dearest family who played a crucial role not only as collaborators but also as a patron of my dream to carry out "good" science which may one day lead to the fulfillment of my dream of saving and improving lives.

Sincere thanks to the Faculty of Medical Sciences, Newcastle University for bestowing me with Overseas Research Studentship and QuantuMDx Ltd. for providing funds to support my research work and livelihood. Special thanks for the work and support of the wonderful cleaning, IT and finance staff at the Institute of Genetic Medicine that has provided a comfortable framework and infrastructure in which I could carry out this work.

Last but not the least, a special applause goes to musicians Hans Zimmer and A.R. Rahman whose awe inspiring music has helped me concentrate and plough through the work; and to artists Bob Kane and Bill Finger who co-created arguably the best fictional character- Batman, from which I have derived great inspiration, courage, determination and joy.

"It is not who you are underneath, but what you do that defines you"

-The Dark Knight

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADR	Adverse drug reaction
AF	Atrial fibrillation
AFR	African population
ASA	Acetylsalicylic acid
BMI	Body mass index
CCO	Cancer Care Ontario
CDE	Common data element
CHB	Han Chinese living in Beijing population
CI	Confidence interval
CIMS	Care institute of medical sciences
CIN	Chromosomal instability
CRC	Colorectal cancer
CVD	Cardiovascular disease
DNA	Deoxyribonucleic acid
DVT	Deep vein thrombosis
EDTA	Ethylenediaminetetraacetic acid
EUR	European population
FAP	Familial adenomatous polyposis
FET	Field effect transistor
FFEQ	Food frequency and epidemiology questionnaire
FHCRC	Fred Hutchinson Cancer Research Centre
G x E	Gene X environment interaction
GI	Gastro-intestinal tract
GIH	Gujarati Indian living in Houston population
GWAS	Genome wide association study
HNPCC	Hereditary non-polyposis colorectal cancer
HR	Hazard ratio
HWE	Hardy Weinberg equilibrium
IHD	Ischemic heart disease
IHG	Institute of Human Genetics
INR	International normalized ratio
IRR	Incidence rate ratio

ISI	International sensitivity index
IWPC	International warfarin pharmacogenetics consortium
LD	Linkage disequilibrium
LVEF	Left ventricular ejection fraction
MAF	Minor allele frequency
MC	Mayo Clinic
MMR	Mismatch repair
MSD	Mean squared difference
NIH-CCFR	NIH- Colon cancer family registry
NSAID	Non-steroidal anti inflammatory drug
OR	Odds ratio
PCR	Polymerase chain reaction
PE	Pulmonary embolism
PT	Prothrombin time
RFLP	Restricted fragment length polymorphism
RR	Relative risk
SDS	Sodium dodecyl sulphate
SiNW	Silicon nanowire
SNP	Single nucleotide polymorphism
SSC	Saline sodium citrate
T_m	Melting temperature
UHI	University of Hawaii
UK-CCSG	UK- Colorectal cancer study group
UQM	University of Queensland and Melbourne
USC	University of Southern California

Chapter 1: Introduction

1.1 Need for personalized medicine

1.1.1 Burden of chronic diseases- cancer and cardiovascular disease

Cancer is one of the leading causes of death worldwide with the burden being expected to grow due to the growth and aging of the population, especially in the developing and under-developed countries that account for 82% of the world's population (Siegel et al., 2012). According to the GLOBOCAN and International Agency for Research on Cancer estimates, approximately 14.1 million new cancer cases and 8.2 million deaths due to cancer occurred in 2012 (Siegel et al., 2012). There is a growing trend for the shift in cancer burden to less developed countries due to an epidemiological transition to improved life expectancy, high-fat diets, smoking and sedentary lifestyle, which have been associated with an increase in cancer risk (Siegel et al., 2012, Gaziano, 2005, Yusuf et al., 2001). Amongst all cancer types, colorectal cancer is the leading cause of cancer related mortality in developed countries for both males and females, which lead to an estimated loss of productivity costs due to premature mortality of €6 billion in Europe alone in 2008 (Hanly et al., 2015, Siegel et al., 2012), thus posing a significant economic burden to the healthcare industry and society at large.

Similar to cancer, the burden of cardiovascular disease is on the rise worldwide and was the largest cause of deaths in 2002 (WHO, 2015, Yach et al., 2004). According to estimates for the year 2020, mortality from chronic diseases will triple in number worldwide and 71% and 75% of all deaths will occur due to ischemic heart disease (IHD) and stroke respectively (WHO, 2015, Yach et al., 2004). In developing countries, the burden of cardiovascular disease has surpassed that of the acute infectious diseases thus creating a polarized double burden of disease (WHO, 2015, Yach et al., 2004, Frenk et al., 1989). Due to the substantial increase in the percentage of people who are either overweight or obese in developing countries, the number of people with cardiovascular disease in India and China are higher than that of all the developed nations put together (Yach et al., 2004, Popkin, 2002). Together with cancer, cardiovascular diseases pose a major economic and healthcare burden, which require immediate prevention and treatment measures to improve the quality of life and reduce the global health burden.

1.1.2 The Human Genome Project

In 1990, the Human Genome Project was launched with the aim to map and sequence all 3 billion nucleotides in the human DNA. In 2001, when the first draft of the human genome was published (Venter et al., 2001, Lander et al., 2001), it heralded the start of implementation of genetics in mainstream medicine. Following it, Collins and McKusick in their paper described several applications of genetic knowledge in medicine such as individual diseases risk prediction, development of designer drugs and prediction of drug responsiveness which when utilized together would revolutionize disease diagnosis and treatment (Collins and McKusick, 2001). They argued that clues to the genetic risk of common diseases, pharmacogenetics and environmental risk factor assessment would be provided by the sequences that account for 0.1% of the variation between individuals.

Numerous types of sequence variation such as microsatellites, minisatellites and restriction fragment length polymorphisms (RFLPs) were identified through the polymerase chain reaction (PCR) technique (Collins et al., 1999). However, the focus soon shifted to non-repetitive sequence variants, which were most commonly found in the genome and were called single nucleotide polymorphisms (SNPs). A subsequent public-private partnership published a publicly available map of 1.42 million SNPs distributed throughout the genome and estimated that 60,000 SNPs would fall within exonic regions and 85% of the exons would be within 5kb of the nearest SNP (Sachidanandam et al., 2001). The authors suggested applications of SNPs in studying human population genetics, candidate gene analysis for disease association using genome wide association studies (GWAS) and studying human evolution.

Whilst GWAS studies have had limited success in identifying association between SNPs and several common diseases (Bodmer and Bonilla, 2008), they have been successful in pharmacogenetics where they have been used in identifying variants in drug metabolizing enzymes, drug target and disease susceptibility gene that are associated with adverse drug reactions (ADRs), drug response variation and output of a treatment in specific disease, respectively (McCarthy and Hilfiker, 2000, Krynetski and Evans, 1999, Drazen et al., 1999, Poirier et al., 1995). These examples vindicated the aim of applying genetic knowledge in disease diagnosis and treatment, which was put forward by the International Human Genome Project Consortium in 1990.

1.1.3 Pharmacogenetics, pharmacogenomics and personalized medicine

Pharmacogenetics is the study of the variability in drug response due to genetic differences whereas pharmacogenomics interrogates the entire genome to screen for the spectrum of genes involved in drug response. This aids in identifying the drug and selecting

optimal dose that is most likely to be effective and safe for an individual and hence frequently described as “personalized medicine”. The term pharmacogenetics and pharmacogenomics was coined by Friedrich Vogel in 1959 (Vogel, 1959) and Andrew Marshall in 1997 (Marshall, 1997) respectively (Pirmohamed, 2011).

Whilst the study of pharmacogenetics and pharmacogenomics required development of molecular genetic technologies, the first example of a pharmacogenetic trait called favism was mentioned by Pythagoras in 510 BC where certain Mediterranean populations developed red blood cell hemolysis after ingesting fava beans (Pirmohamed, 2011, Nebert et al., 2008, Cappellini and Fiorelli, 2008). The cause was later identified to be the deficiency of glucose-6-phosphate dehydrogenase (G6PD) that is the commonest human enzyme deficiency and affects 600 million people worldwide (Nebert et al., 2008, Cappellini and Fiorelli, 2008). The same deficiency caused hemolytic crisis in ~10% of African American soldiers and a small number of Caucasian soldiers during World War II after being administered with an anti-malarial drug primaquine or other chemically related drugs (Alving et al., 1956, Clayman et al., 1952).

The phenotype driven approach to understanding variation in drug response was used extensively between the 1950s and 1980s and it usually required administration of a probe drug and subsequent measurement of its metabolite (Pirmohamed, 2011, Meyer, 2004). The ratio of probe drug to its metabolite represented whether an individual had an absolute or partial deficiency of an enzyme. In the late 1970s, techniques like these helped define an individual’s capacity to hydroxylate debrisoquine, an anti-hypertensive drug, whose adverse reactions included significant cardiovascular effects such as hypotension (Mahgoub et al., 1977). Analysis of the urinary metabolite 4-hydroxy-debrisoquine revealed the source of differential metabolism to be due to the deficiency of a specific cytochrome P-450 enzyme in the liver microsome, which was later called CYP2D6 (Meier et al., 1983). Subsequent cloning and detailed sequence analysis identified two frequent alleles, *CYP2D6*4* and *CYP2D6*3*, which were associated with adverse drug reactions in poor metabolizers and thus the first PCR-based test to identify poor metabolisers of debrisoquine was developed (Gough et al., 1990). Furthermore, alleles containing 3 to 13 copies of the *CYP2D6* gene in individuals with ultra-rapid metabolism phenotype was later identified.

Since the completion of the Human Genome Project and increased utility of genome wide association studies (GWAS) in pharmacogenomics has led identification of several genes with drug response phenotype for several drugs such as: predisposition to stent thrombosis in clopidogrel users with *CYP2C19*2* allele, *HLA-B*5701* with abacavir hypersensitivity and flucloxacillin induced liver injury and *IL28B* with interferon-alpha

efficacy in treating hepatitis C infection etc. (Pirmohamed, 2011, Shastry, 2006, Meyer, 2004). Additionally, associations of polymorphisms with anticancer drugs were also identified, such as: polymorphism in thiopurine S-methyltransferase (*TPMT*) with mercaptopurine toxicity, polymorphism of UDP-glucuronosyltransferase (*UGT1A1*) with irinotecan toxicity and V600E mutation in the *BRAF* gene with vemurafenib efficacy etc. (Pirmohamed, 2011, Shastry, 2006, Meyer, 2004).

Recently, polymorphisms in the *CYP2C9* and *VKORC1* genes were shown to be associated with dose sensitivity of the drug warfarin, which is used to treat atrial fibrillation (Klein et al., 2009, Gage et al., 2008, Aithal et al., 1999). This suggests that pharmacogenetic testing can improve efficacy and prevent adverse drug reactions thus reducing cost of therapy to the healthcare system and improving the quality of life of patients. With the trend for rise in chronic diseases such as cancer and cardiovascular disease, implementation of pharmacogenetic testing for prophylactic and therapeutic agents could help reduce the burden of these diseases globally.

1.2 Genetics, aspirin and colorectal cancer

Colorectal cancer (CRC) is one of the most frequent cancers in the developed world with an incidence of 160,000 cases diagnosed in the US every year and more than 31,000 cases diagnosed in the UK alone in 2011 (CRUK, 2014, Markowitz and Bertagnolli, 2009). Additionally, both in the US and in the UK, it is the second leading cause of cancer related mortality (CRUK, 2014, Markowitz and Bertagnolli, 2009). The disease usually initiates as a benign polyp that develops in to advanced adenoma with dysplasia and finally progresses to an invasive cancer (Markowitz and Bertagnolli, 2009). Invasive cancers that are confined to the outer wall of the colon (stage I and II) are curable by surgical excision but if untreated, they can spread to lymph nodes (stage III) and then can metastasize to distant sites (stage IV) (Markowitz and Bertagnolli, 2009, Markowitz et al., 2002). Seventy three percent of stage III cancers are curable by surgery combined with adjuvant chemotherapy whereas stage IV cancers are usually incurable (Markowitz and Bertagnolli, 2009, Andre et al., 2004, Markowitz et al., 2002). Delineating underlying pathological and molecular changes that drive a normal epithelial cell to transform into a metastatic tumor cell could help determine individual susceptibility to CRC and the efficacy of antitumor agents.

1.2.1 Carcinogenesis of colorectal cancer

CRC occurs in both sporadic and familial (hereditary) cases, however, about 75-80% of tumors are of sporadic origin (Moran et al., 2010). Two major pathways are understood to be in involved in colorectal carcinogenesis- Chromosomal instability (CIN) or “suppressor”

pathway and the mutator or DNA mismatch repair pathway (Moran et al., 2010, Markowitz and Bertagnolli, 2009).

1.2.1.1 Chromosomal instability pathway (CIN)

CIN is the most common type of genomic instability observed in 80-85% of tumors; it is assumed to follow Fearon and Vogelstein model of carcinogenesis (Moran et al., 2010, Fearon and Vogelstein, 1990). The linear model proposed that specific genetic events were correlated with the evolving tissue morphology (Figure 1.1). In cancers caused due to CIN, rare inactivating mutations in genes involved in chromosome stability during replication are observed (Markowitz and Bertagnolli, 2009, Barber et al., 2008). This leads to physical loss of a wild-type copy of tumor suppressor genes such as *APC*, *P53* and *SMAD4* and gain of function of oncogenes such as *KRAS* (Moran et al., 2010, Markowitz and Bertagnolli, 2009). Furthermore, CIN tumors contain high frequency of allelic imbalance, most commonly on chromosome 5q, 8p, 17p and 18q (Moran et al., 2010, Markowitz and Bertagnolli, 2009)

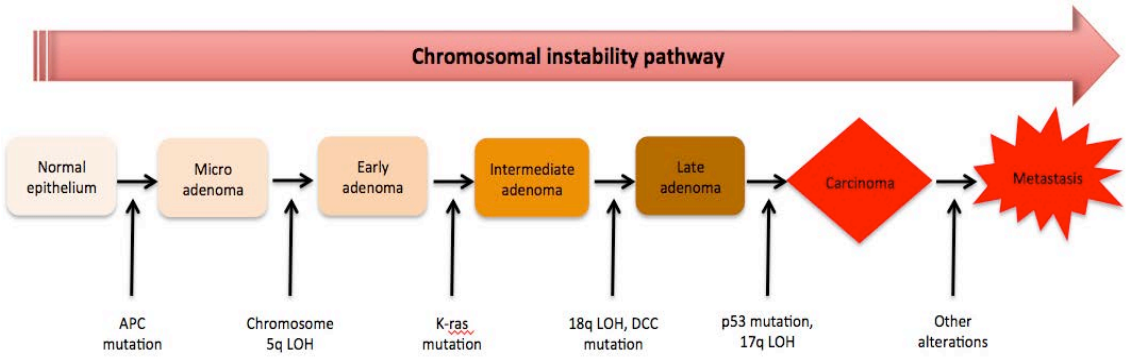


Figure 1.1 Colorectal carcinogenesis through chromosomal instability pathway.
Figure adapted with permission from Moran et al. 2010.

1.2.1.2 DNA mismatch defect pathway

DNA mismatch repair pathway defects are observed in approximately 15-20% of sporadic CRCs and the tumors are characterized by high mutation rates, 100- to 1000-fold more common in comparison to normal cells, mainly affecting microsatellite sequences (Moran et al., 2010, Pawlik et al., 2004). This is caused by the inactivation of mismatch repair (MMR) genes that are required for base-mismatch repair post DNA replication (Figure 1.2). In total, there are 7 MMR genes that encode functional proteins to carry out mismatch repair: *hMLH1*, *hMLH3*, *hMSH2*, *hMSH3*, *hMSH6*, *hPMS1* and *hPMS2* (Moran et al., 2010, Markowitz and Bertagnolli, 2009, Hoeijmakers, 2001). In sporadic tumors, epigenetic inactivation of *hMLH1* due to methylation and less frequently, mutation in *hMSH6* is observed (Imai and Yamamoto, 2008). In contrast, germ-line mutation in *hMLH1* and *hMSH2* genes leads to the hereditary form of CRC known as hereditary non-polyposis colorectal

cancer (HNPCC) or Lynch Syndrome that accounts for 2-4% of all CRC cases (Bronner et al., 1994, Fishel et al., 1993). In both sporadic and hereditary tumors with MMR deficiency, mutations in the mononucleotide or dinucleotide repeat sequences in the functional regions of the tumor suppressor genes such as *TGFβR2*, *BAX* and *IGF2R* etc. and epigenetic silencing of a number of normally functioning genes is observed (Moran et al., 2010, Markowitz and Bertagnolli, 2009).

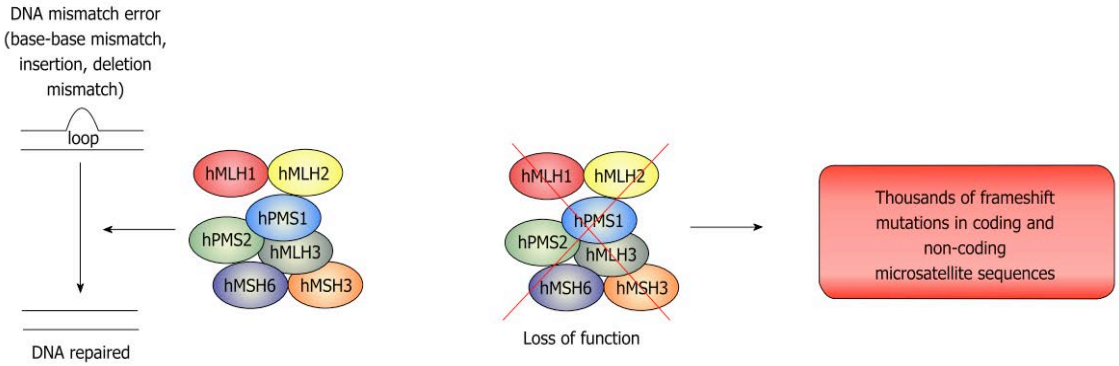


Figure 1.2 Colorectal carcinogenesis through defects in the DNA mismatch repair pathway.

Figure taken with permission from Moran et al., 2010.

1.2.2 Aspirin and colorectal cancer risk

Acetylsalicylic acid (ASA), Aspirin, is a non- steroidal anti- inflammatory drug (NSAID) which is used as an analgesic, anti-pyretic or as a prophylactic drug for cardiovascular diseases (CVD) (Fuster and Sweeny, 2011). Aspirin is made up of two components: a phenol ring consisting of 6- carbon benzene ring with a carboxyl group and an acetyl moiety (Fuster and Sweeny, 2011). It was first introduced into the market in 1899 and was registered under the name of "Aspirin" by Bayer (Fuster and Sweeny, 2011). Currently, an estimated 40,000 tons of aspirin are produced worldwide annually and approximately 10 to 20 billion tablets are consumed annually in the USA alone for CVD prophylaxis (Campbell et al., 2007), making it one of the most widely used drugs in the world.

1.2.2.1 Association with colorectal cancer risk

Despite being widely prescribed as an anti-platelet and anti-inflammatory drug, the first epidemiological evidence for an inverse association of its intake with colorectal cancer risk was reported from a population based case- control study in 1988 (Relative risk [RR] for males=0.58, 95% CI=0.38-0.88, P=0.02; RR for females=0.49, 95% CI=0.32-0.73, P<0.01) (Kune et al., 1988). Further evidence of an inverse association came from a meta-analysis of 11 case- control and 7 cohort studies that showed an inverse association between long term aspirin use and colorectal cancer (Meta-analysis of case-control studies RR=0.59, 95%

CI=0.54-0.64, P for heterogeneity= 0.008; Meta-analysis of cohort studies RR=0.85, 95% CI=0.78-0.92, P for heterogeneity= 0.006) (Cuzick et al., 2009, Bosetti et al., 2006). However, a significant heterogeneity was observed in the estimates between and within case-control and cohort studies as the studies were carried out in different populations, used different methods for case ascertainment, and used different types of controls.

To test the robustness of the inverse association, the first randomized 2 x 2 factorial double-blinded trials were launched, first in 1993 in Familial Adenomatous Polyposis patients (CAPP1) and then in 1998 in 1009 patients who were genetically predisposed to Lynch syndrome (CAPP2), receiving 600mg daily aspirin or placebo for a mean intervention period of 29 months (Burn et al., 2008). After a mean follow up of 55.7 months, per protocol analysis showed a reduced risk of primary cancer in the aspirin group compared to placebo (HR=0.41, 95% CI=0.19-0.86, P=0.02; IRR=0.37, 95% CI=0.18-0.78, P=0.008) (Burn et al., 2011b). The combined evidence from observational studies and randomized controlled trials provided compelling evidence that aspirin intake reduces CRC risk. Rothwell *et al.* returned to the randomized trials performed to assess the effects of aspirin on cardiovascular disease. Extended follow up of over 25,000 recruits displayed a highly significant reduction in colorectal and other cancers commencing around 5 years after the initial recruitment compared to placebo groups (Rothwell et al., 2012a). In 2013, Cook *et al.* published a long-term follow up of the only other aspirin randomized trial with cancer as an endpoint, the Women's Health Study. Alternate day 100mg aspirin resulted in an 18% reduction in gastrointestinal cancers with the effect commencing 10 years after randomization (Cook et al., 2013).

1.2.2.2 Association with colorectal adenoma risk

Since colorectal adenomas are the precursors to most CRCs, the chemopreventive effect of aspirin is likely to be observed in the adenomas as they form during the neoplastic transformation of normal to cancer cells. The largest randomized placebo controlled trial consisting of 206 patients who were genetically predisposed to a hereditary form of colorectal cancer, familial adenomatous polyposis (FAP), showed no significant reduction in the polyp count in the sigmoid colon and rectum of patients taking 600mg/day aspirin compared to placebo (RR=0.77, 95% CI=0.54-1.10) (Burn et al., 2011a). However, there was a reduction in the mean polyp size in patients randomized to aspirin (3.0mm versus 6.0mm; P=0.02) (Burn et al., 2011a). Despite the lack of clear evidence of the protective effect of aspirin on primary lesions in individual trials, a meta analysis of 4 randomized controlled trials (AFPPS (Baron et al., 2003), CALGB (Sandler et al., 2003), ukCAP (Logan et al., 2008) and APACC (Benamouzig et al., 2003)) that evaluated secondary prevention of sporadic colorectal

adenoma with aspirin showed a risk reduction for developing adenomas (pooled RR=0.83, 95% CI=0.72-0.96, P=0.012) and advanced lesions (pooled RR=0.72, 95% CI=0.57-0.90, P=0.005) (Cole et al., 2009). Thus in both sporadic and familial risk patients, there is convincing evidence of adenoma risk reduction with aspirin use.

1.2.2.3 Clinical utility of aspirin in prevention of CRCs

Despite the accumulation of evidence that supports the use of aspirin for prophylaxis and adjuvant therapy in reducing CRC risk, it is not being prescribed widely for chemoprevention due to its adverse side effect of gastrointestinal (GI) bleeding (Huang et al., 2011). A prospective study of 87,680 women over a 24 year follow up in the Nurses' Health Study showed an increased risk of GI bleeding in regular aspirin users (≥2 325mg tablets/ week) compared to non-regular users (RR=1.43, 95% CI=1.29-1.59) (Huang et al., 2011). Additionally, compared to non-users, the risk of bleeding increased with an increase in the number of tablets taken per week (P_{trend}<0.001) (Huang et al., 2011). However, after adjusting for dose, no difference in bleeding risk was observed for short term and long term users. Furthermore, in 2 randomized controlled trials that compared rates of GI bleeding in individuals taking low dose (81 mg) and high dose (325 mg) aspirin showed no significant difference between the rate of GI bleeding between the two groups (Baron et al., 2003, Taylor et al., 1999).

A recent review looking into the risks and benefit of prophylactic use of aspirin in the general population observed a delayed chemopreventive effect of aspirin by 3 years from the start of treatment (Cuzick et al., 2014). The review also mentioned an increase of 32-36% in hemorrhagic strokes, which is the most serious and potentially fatal side effect, in aspirin users from a baseline rate of 0.03% per annum whereas the risk of the GI bleeds increased by 30-70% from the baseline risk of 0.7/1000/ year in people taking aspirin (Cuzick et al., 2014). The authors estimate a relative reduction of ~9% and 7% in the number of men and women respectively with cancer, myocardial infarction or stroke event over a 15 year period if they have been taking aspirin for 10 years (Cuzick et al., 2014). Furthermore, they also calculated that 61-80% of the overall benefit would be accounted by the decrease in the cancer risk, especially, reduction in the CRC risk, which alone would account for 30-36% (Cuzick et al., 2014). Based on the evidence of risk and benefit of using aspirin, the authors concluded that prophylactic use of 75 to 325 mg/day aspirin for a minimum of 5 years would have a favorable risk-benefit ratio.

1.2.3 Pharmacogenetics of aspirin

Despite extensive evidence for aspirin's chemopreventive efficacy and accummulating evidence on the risk-benefit profile, data from several studies suggests inter-individual variation in the chemopreventive effect and the source of this variation has been attributed to the presence of somatic mutations (Nishihara et al., 2013, Liao et al., 2012) and germline variation in aspirin's pharmacokinetic and pharmacodynamic pathways (Wang et al., 2014, Reimers et al., 2014, Fink et al., 2014, Angstadt et al., 2014, Seufert et al., 2013, Nan et al., 2013, Pathi et al., 2012, Zell et al., 2009, Hubner et al., 2008, Chan et al., 2007, Hubner et al., 2006, Din et al., 2004, Stark et al., 2001). Aspirin's mode of action on cellular pathways within the colonic epithelial cells which in turn reduces the risk of CRC has yet to be elucidated. Furthermore, genetic variants that influence an individual's risk to adverse drug reaction associated with aspirin is currently under scrutiny (Agundez et al., 2009). Developing a panel consisting of SNPs associated with aspirin's efficacy along with SNPs associated with adverse drug reaction could help determine the optimal dose for an individual.

1.3 Genetics, warfarin and cardiovascular diseases

Cardiovascular diseases (CVDs) are estimated to account for 30% of all deaths worldwide, where 80% of the burden is attributed to developing countries (Gaziano, 2005). CVDs include stroke, atrial fibrillation (AF), sudden cardiac arrest, heart failure and coronary artery diseases etc. In the latest executive summary by the American Heart Association, CVDs accounted for ~1 in every 3 deaths in the USA alone in 2011 which averages to approximately 1 death every 40 seconds (Mozaffarian et al., 2015). In the USA, it is estimated that someone has a stroke every 40 seconds and death due to stroke occurs every 4 minutes (Mozaffarian et al., 2015). Furthermore, it is estimated that AF related stroke occurs every 15 seconds (AFA, 2015) thus making AF related stroke to be one of the biggest contributors to the burden from CVDs. AF can cause thrombosis, which is local coagulation or clotting of blood in a blood vessel, which when dislodged can travel to the capillaries in the brain and restrict or stem the flow of blood causing ischemic stroke.

1.3.1 Blood clotting and coagulation pathway

Blood clotting and coagulation is a process carried out in the damaged blood vessels to prevent loss of blood and the entire process is called as hemostasis (Silverthorn, 2009, Davie et al., 1991). Following the damage to the blood vessel, hemostasis leads to vasoconstriction to decrease blood flow and pressure within the vessel (Silverthorn, 2009). In addition to that, platelets rapidly create a mechanical block around the damaged and exposed vascular tissue through the process of clotting and finally, the exposed collagen and tissue factor of the blood

vessel initiates a reaction know as coagulation cascade that leads to the formation of fibrin mesh that stabilizes the platelet plug (Silverthorn, 2009).

1.3.1.1 Clotting through platelet activation

When a blood vessel is damaged, the exposed collagen and platelet-activating factor from endothelial cells activate platelets to adhere to the site of the injury (Camussi et al., 1983). Aggregated platelets form a clot and adhere to collagen with the help of integrins (Silverthorn, 2009, Davie et al., 1991). Binding of platelets leads to the release of signaling molecules such as serotonin, ADP and platelet-activating factor which feeds into a positive feedback loop to bind more platelets; arachidonic acid is converted to thromboxane A2 by the cyclooxygenase-1 enzyme, a target of aspirin, which leads to further vasoconstriction and; plasma proteins such as von Willebrand factor help in platelet adhesion with the vascular endothelial cells (Silverthorn, 2009, Davie et al., 1991, Ruggeri and Zimmerman, 1987, Girma et al., 1987). Platelet aggregation and platelet plug formation sets the stage for the coagulation cascade that ends with the formation of insoluble fibrin that creates a mesh around the platelet plug to provide strength and stability.

1.3.1.2 Coagulation cascade

The last step in hemostasis is coagulation, which ends in the formation of a gelatinous clot. Coagulation is divided in two pathways: intrinsic and extrinsic, that are activated by exposure to different cells of the blood vessels (Figure 1.3) (Silverthorn, 2009). The intrinsic pathway is initiated when a plasma protein called factor XII is activated when it comes in contact with the exposed vascular collagen (Silverthorn, 2009, Davie et al., 1991). Following that, each plasma protein involved in the intrinsic pathway is activated by minor proteolysis, Ca^{2+} ions or phospholipids (Davie et al., 1991, Macfarlane, 1964, Davie and Ratnoff, 1964). In contrast, the extrinsic pathway is initiated when the tissue factor called thromboplastin that is located in the tissue adventitia is exposed to the blood after vascular injury and activates factor VII (Silverthorn, 2009, Davie et al., 1991, Wilcox et al., 1989). Activated factor VII and factor IX in the presence of factor VIII from the extrinsic and intrinsic pathways respectively activate factor X in the presence of factor Ca^{2+} ions and phospholipids (Silverthorn, 2009, Davie et al., 1991). Active factor X converts prothrombin to thrombin in the presence of factor Va (Davie et al., 1991, Mann et al., 1982). Thrombin, a serine protease, converts fibrinogen into fibrin by limited proteolysis and at the same time activates factor XIII which helps in cross-linking fibrin to form an insoluble fibrin clot (Silverthorn, 2009, Davie et al., 1991). It is suggested that the intrinsic pathway is involved in the growth and

maintenance of fibrin formation whereas the extrinsic pathway is involved in the initiation of fibrin formation (Davie et al., 1991).

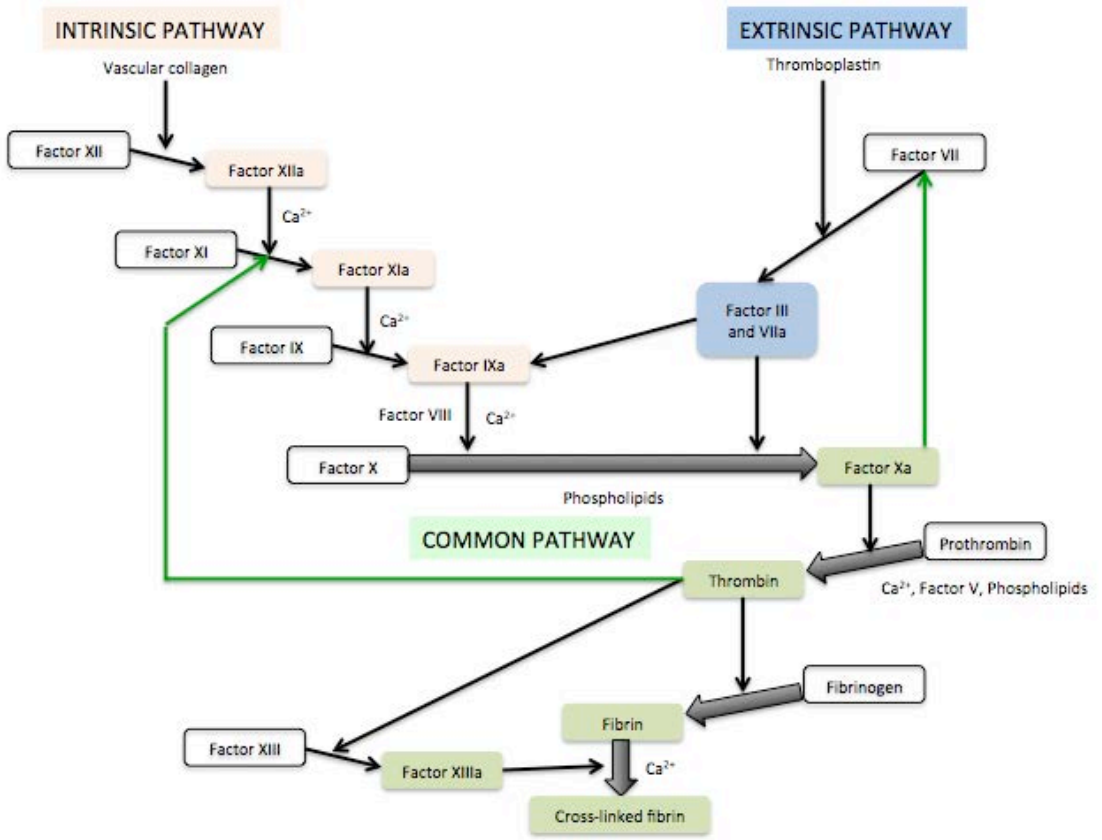


Figure 1.3 Intrinsic and extrinsic pathway in the coagulation cascade.
Green arrow indicates a positive feedback loop. Figure adapted with permission from Davie *et al.*, 1991.

1.3.2 Warfarin- from rat poison to oral anticoagulant

Whilst hemostasis helps in preventing blood loss, too little hemostasis can lead to excessive bleeding called a hemorrhage whereas too much can lead to a blood clot known as thrombus that adheres to the undamaged wall of a blood vessel. When the thrombus travels to the vessels in the brain, it can stop the blood flow and cause a stroke.

Two types of agents that affect platelet based clotting and coagulation cascades are known as anti-thrombotic and anti-coagulant agents respectively. Aspirin is one of the most widely used drugs in the world and is an anti-clotting agent; it affects platelet plug formation by inhibiting cyclooxygenase enzymes that activate platelet aggregation through generation of thromboxane A_2 (Silverthorn, 2009). In contrast, the coumarin based anti-coagulant drug warfarin affects the coagulation cascade by inhibiting regeneration of vitamin K, which is a

co-factor in the synthesis of factors II, VII, IX and X (McDonald et al., 2009b). Whilst aspirin and warfarin are used as prophylactic drugs for reducing the risk of stroke due to clots, warfarin is shown to be superior at reducing the risk of stroke in patients with heart failure in sinus rhythm (HR=0.52, 95% CI=0.33-0.82, P=0.005) and in elderly patients (75 years or over) with atrial fibrillation (RR=0.48, 95% CI=0.28-0.80, P=0.003) when compared to aspirin (Homma et al., 2012, Mant et al., 2007). This suggests that warfarin is superior at reducing the risk of stroke compared to aspirin.

Today, warfarin is commonly used to treat patients with CVDs but it was originally discovered when healthy cattle in the prairies of Canada and Northern parts of the USA began dying due to internal bleeding in the 1920s (Wardrop and Keeling, 2008, Francis, 2008, Pirmohamed, 2006). The substance responsible for bleeding- 3,3'-methylene-bis[4-hydroxycoumarin] was identified and extracted from mouldy sweet clover by Karl Paul Link at the University of Wisconsin (Stahnmann et al., 1941, Campbell and Link, 1941). In 1948, this substance was commercialized as a rodenticide under the brand name of Warfarin as the funding to extract and develop this substance was provided by the Wisconsin Alumni Research Foundation (Wardrop and Keeling, 2008, Francis, 2008). In 1954, it got approved for medical use and in 1955 it was given to President Dwight Eisenhower following an episode of myocardial infarction (Wardrop and Keeling, 2008, Francis, 2008).

1.3.3 Prescribing and monitoring warfarin dose

A substantial increase in the use of warfarin can be attributed to the evidence of its effectiveness in preventing strokes in patients with atrial fibrillation (Pirmohamed, 2006, Aguilar and Hart, 2005). Warfarin has a narrow therapeutic index and thus it is difficult to prescribe a dose and maintain patients within a defined anticoagulation range. In 1983, a standardized system of measuring sensitivity to warfarin was devised and was called as the International Normalized Ratio (INR) (Kirkwood, 1983). INR is the ratio of the Prothrombin time (PT) of the patient against a control sample raised to the power of the International Sensitivity Index (ISI), which is the measure of the sensitivity of the test. Baseline INR for an individual is given the value of 1.0. For a patient with atrial fibrillation, an ideal INR range is between 2.0 and 3.0 (Oden et al., 2006, Hylek et al., 1996). An INR of less than 2 increases risk of thrombotic events whereas an INR of more than 3 increases risk of hemorrhage (Jones et al., 2005). A meta-analysis of 33 studies calculated that the rate of major and fatal bleeding events occur at 7.2 and 1.3 per 100 patient years respectively in warfarin users and thus is placed at number 3 on the list of drugs implicated in hospital admission because of adverse drug reactions (Pirmohamed et al., 2004, Linkins et al., 2003). Furthermore, dose requirements between individuals can vary up to 20 fold, which helped explain results of the

analysis of 6545 patients with atrial fibrillation taking warfarin where ~50% of the time was spent outside the target INR range of 2.0-3.0 (Pirmohamed, 2006, Boulanger et al., 2006).

1.3.4 Pharmacogenetics of warfarin

To improve the time spent within the target INR range, several studies have suggested that the combined knowledge of three genetic polymorphisms, along with demographic and anthropometric variables such as age, gender and body mass index (BMI), could explain up to half of warfarin dose variability (Gage et al., 2008, Yang et al., 2009, Tatarunas et al., 2011). Presence of the variant allele of two SNPs in the *CYP2C9* gene (*2- rs1799853 and *3- rs1057910) have been shown to be associated with a 2-3 fold increased risk of an adverse event during treatment initiation (Aithal et al., 1999, Gage et al., 2008, Pavani et al., 2011) and -1639G>A SNP in the promoter of the *VKORC1* gene has been shown to reduce the quantity of active enzyme thus increasing sensitivity to warfarin (Wang et al., 2008, Yuan et al., 2005, D'Andrea et al., 2005).

In 2009, the International Warfarin Pharmacogenetics Consortium (IWPC) created a pharmacogenetic algorithm that predicted maintenance dose of warfarin based on the 3 genetic markers, age, gender, BMI, smoking and amiodarone and enzyme inducer use (Klein et al., 2009). The algorithm accounted for ~60% of the dose variation between individuals. A modified version of the algorithm was used in a randomized controlled trial European Pharmacogenetics of Anti Coagulant Therapy (EU-PACT) where the clinical utility of the pharmacogenetic guided dosing lead to the patients spending higher percentage of mean time in the therapeutic INR range (67.4%) compared to patients receiving standard dosing (60.3%) during the initiation of warfarin therapy (Adjusted difference, 7.0 percentage points; 95% CI, 3.3 to 10.6; $p<0.001$) (Pirmohamed et al., 2013). In contrast, the Clarification of Optimal Anticoagulation Through Genetics (COAG) trial that was conducted in the USA reached a different conclusion, which is discussed in chapter 5. Thus, the EU-PACT trial showed the benefit of applying pharmacogenetics-guided warfarin dosing in routine clinical practice. In addition, genotyping the three SNPs has recently been shown to help stratify patients based on their sensitivity to warfarin to be initiated on new oral anticoagulants such as Endoxaban, a factor Xa inhibitor, thus improving safety (Mega et al., 2015) and will likely reduce the burden of cost due to the introduction of new anticoagulants on the healthcare system. Therefore, introduction of pharmacogenetics-guided warfarin stratification and dosing is likely to improve safety and efficacy of the drug.

1.4 Silicon nanowire technology for rapid SNP genotyping

Pharmacogenetics and pharmacogenomics have become one of the most active areas of the personalized medicine paradigm with an increase in implementation of genetic tests before prescribing drugs over the past few years (Scott, 2011). Several trials that assessed the health resource utilization, potential cost savings for the healthcare system and improvement in the quality of life of the patient showed reduction in spending on medication (Chou et al., 2000), reduction in the length of hospital stay (Ruano et al., 2013), improvement in adherence to the prescribed medication (Fagerness et al., 2014) and reduction in overall pharmacy costs (Benitez et al., 2015). In the light of this evidence, several organizations have developed and curated pharmacogenetic gene and variant lists based on the relevant literature in an effort to summarize data on relevant markers; for example, PharmaADME "Core Gene List" (http://www.pharmaadme.org/) and Pharmacogenomics Knowledge Database (PharmGKB; http://www.pharmgkb.org/). Furthermore, websites that can predict drug dose based on the individual's genetic and anthropometric factors have also been developed and are now available for clinical use; for example, pharmacogenetics-guided warfarin dose can now be calculated using the free online tool available at http://www.warfarindosing.org/ which is based on the validated algorithm developed by Klein *et al.*, 2009.

1.4.1 Point of care genotyping device

Despite the widespread availability of several whole genome SNP genotyping platforms that are placed in central laboratories, one of the challenges of implementing pharmacogenetic biomarkers in routine clinical practice is the lack of a robust, portable and cheap genotyping platform that can carry out SNP genotyping at the point of care within a short turnaround time (Ong et al., 2012). Several pharmacoeconomic studies have shown that the influential factor for the cost-effectiveness of using pharmacogenetics in clinical practice is low genotyping cost and fast turnaround time (under 24 hours) (You, 2011, You et al., 2009, Eckman et al., 2009).

To aid in the implementation of pharmacogenetics in clinical practice, a Newcastle upon Tyne, UK based company called QuantuMDx Ltd. is currently developing a handheld, sample to result DNA diagnostic device that is designed to genotype SNPs in 20 minutes for as little as £20 per test (www.quantumdx.com) (Figure 1.4). The device consists of 4 components: Mechanical lysis leads to a DNA extraction cassette, which consists of densely packed sorbent filter with a unique property to bind with proteins and lipids and but not charged nucleotides including DNA (Figure 1.4 A); Third, a microfluidic based PCR cassette with two or three heating zones mimicking denaturation, annealing and amplification steps of PCR

(Figure 1.4 B) and; Fourth, a silicon nanowire (SiNW) based field effect transistor (FET) nanosensor for electrical detection and genotyping of oligonucleotides (Figure 1.4 C). Utilization of microfluidics and SiNW based genotyping platform would reduce the device footprint and increase sensitivity and selectivity for detection of label-free DNA in real time at the point of care.

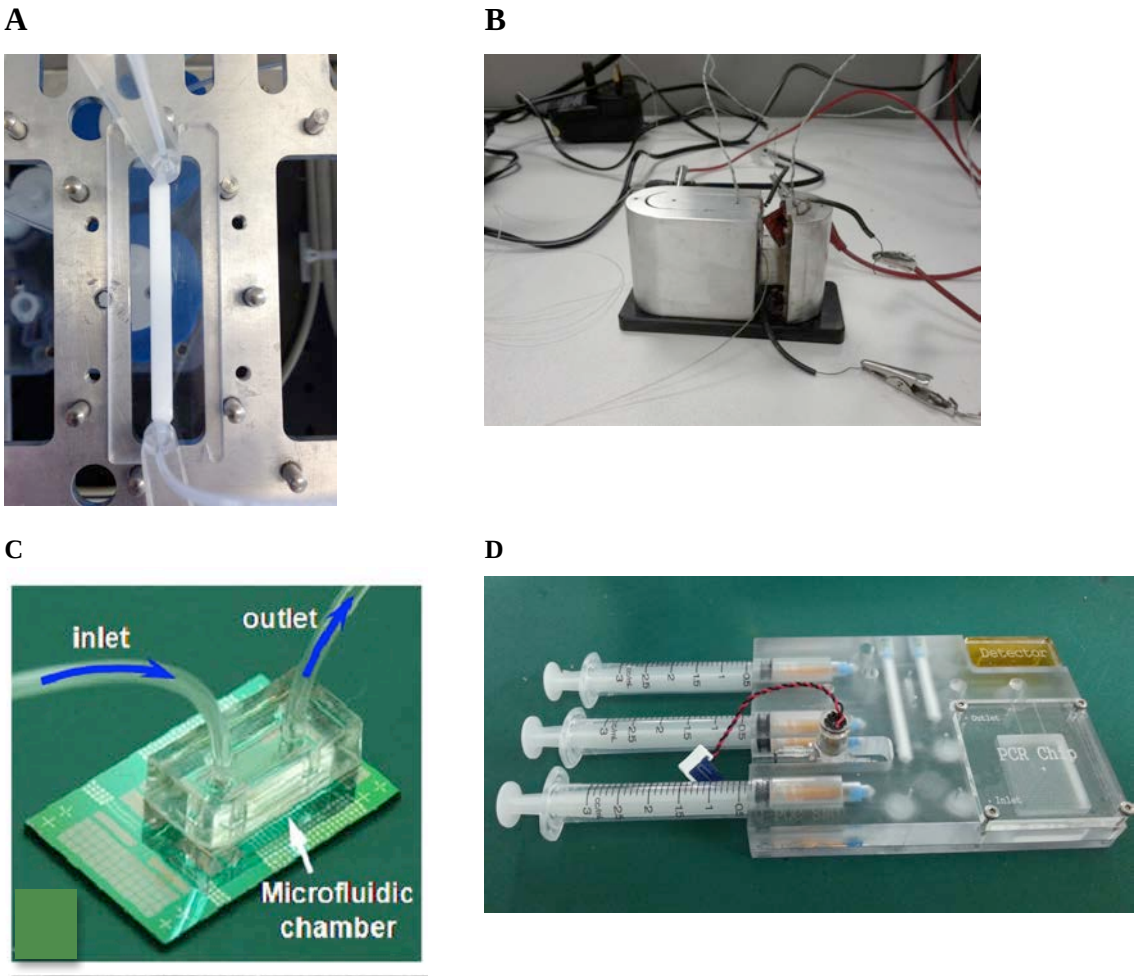


Figure 1.4 QuantuMDx's silicon nanowire based DNA genotyping platform.
(A) Microfluidics based DNA extraction cassette, (B) Microfluidics based thermal cycler with 95°C and 50°C heating zones in this early form, (C) Silicon nanowire based biosensor chip with microfluidic housing for delivering target DNA and (D) Prototype cassette housing DNA extraction, thermal cycler and silicon nanowire components. Images provided by QuantuMDx Ltd.

1.4.2 Chemistry of Silicon Nanowires

Silicon nanowires are 3-dimensional nanostructures that transport electrical charge and have semi-conducting properties (Cui et al., 2000). Advances in SiNWs have generated great interest as they offer high sensitivity due to their extremely large surface area to volume ratio and high sensitivity in electron conductance to the variations in the electric field or charge at the surface (Gao et al., 2011, Gao et al., 2007, Cui et al., 2000). SiNW devices are preferred over other semi conductor devices because they can be prepared using either "bottom up" or "top down" methods (Patolsky et al., 2006, Cui and Lieber, 2001). In the "bottom up" approach, nanoparticles self assemble into complex structures on an oxidized silicon substrate using gold nanoparticles and silane as a catalyst. This is followed by depositions of nanowires on a substrate, photolithographic wiring of source and drain electrodes and hydrofluoric acid etching of exposed SiNW (Patolsky et al., 2006). The device created using this technique suffers from issues such as nanowire uniformity and yield.

In contrast, in the "top down" approach, SiNWs are patterned and etched on silicon-on-insulator wafers using electron beam lithography followed by hydrofluoric acid etching (Gao et al., 2007). Whilst this approach is expensive in comparison to the "bottom up" approach, SiNWs created using this approach are uniform. However, SiNWs created using the "top down" approach could have small variations in the nanowire diameter, which could introduce parameter-coupling effects i.e. intra nanowire variation during conductance measurements (Li et al., 2011).

There are two types of SiNWs: p-type and n-type. The type of nanowire created is based on which element is used as a dopant during the SiNW fabrication process; diborane (B_2H_6) for p-type and phosphine (PH_3) for n-type (Patolsky et al., 2006). Doping provides surface charge screening by mobile charge carriers (Li et al., 2011). In 2011, Li *et al.* showed that SiNWs with low doping concentrations (10^{17} atoms/ cm^3) are 3.2x more sensitive to change in surface charge along with an improved sensor detection limit when compared with the SiNWs having high doping concentration (10^{19} atoms/ cm^3). Additionally, an inverse relationship between nanowire diameter and sensitivity to the surface charge was observed which suggested that thinner nanowires (20-80 nm) have higher sensitivity than thick nanowires (100-200 nm) (Li et al., 2011). Although, with decrease in diameter, there is an increase in susceptibility to background noise suggesting a trade-off between sensitivity and noise. Taking these parameters into consideration, SiNWs could be used to carry out ultra-rapid, label-free DNA detection.

1.4.3 DNA detection using silicon nanowires

Li *et al.*, 2004 showed detection of label-free DNA using both p- and n-type SiNWs (Li et al., 2004). Both p- and n-type SiNW surface were functionalized with a methoxy silane layer to which single stranded DNA oligonucleotide probes that were complementary to the target DNA sequence were covalently attached (Figure 1.5). Binding of target DNA to a complementary oligonucleotide probe produced signal that was >6 higher than the background noise. Furthermore, target DNA with a single base mismatch didn't produce signal above the background noise thus demonstrating the potential of the SiNW biosensor for detecting SNPs in the DNA. Several other studies have produced similar outcomes and showed that the DNA hybridization events can be observed *in situ* and in real time and can reliably detect target DNA at concentrations of 1fM with high specificity to detect SNPs (Gao et al., 2011, Gao et al., 2007). In addition to detecting label-free DNA, SiNWs have been shown to carry out multiplex detection of protein markers of cancer such as prostate specific antigen, carcinoembryonic antigen and mucin-1 from serum samples at femtomolar concentrations with high sensitivity (Zheng et al., 2005). This suggests that a SiNW based platform could be used for both SNP genotyping and disease specific protein marker detection.

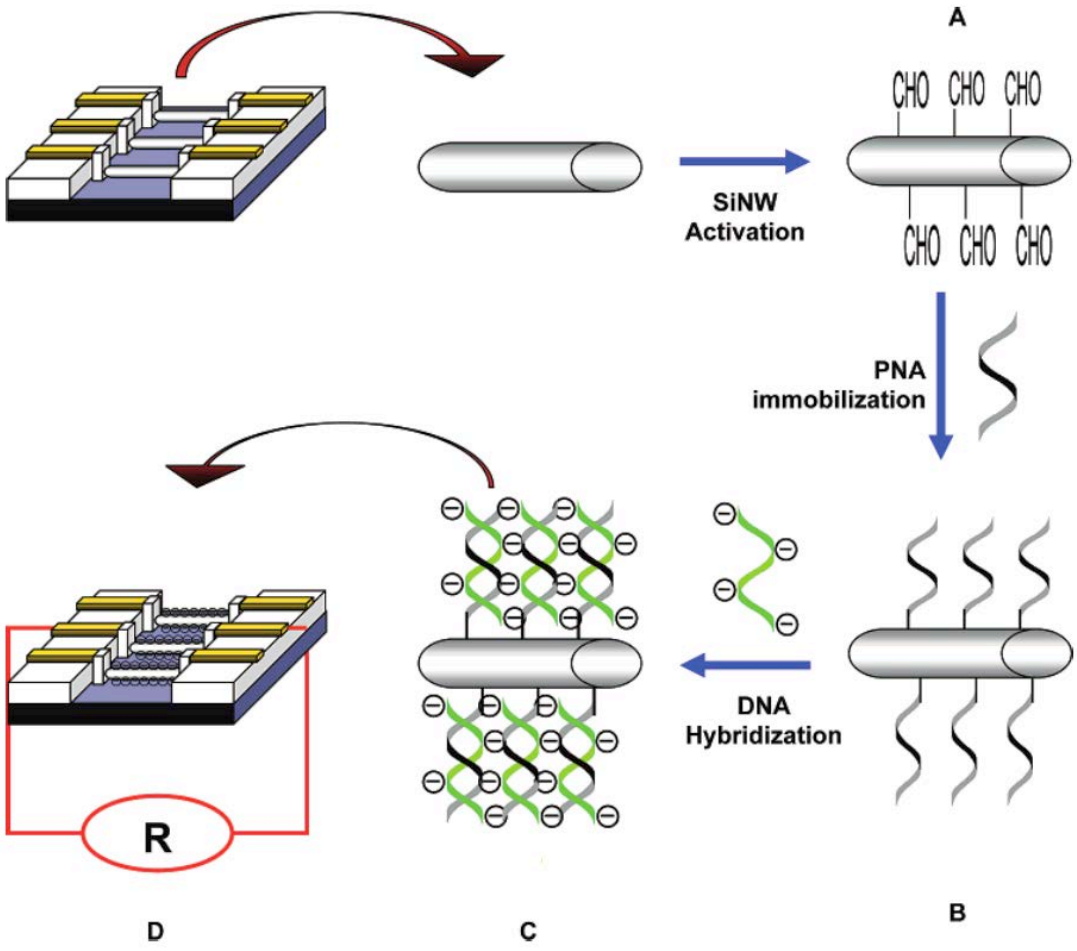


Figure 1.5 Schematic representation of target DNA sensing with complementary peptide nucleic acid (PNA) probe on silicon nanowire (SiNW). (A) Silicon nanowire surface functionalized with aldehyde moieties for covalently attaching complementary oligonucleotide probes, (B) monolayer of oligonucleotide probes assembled through silane chemistry, (C) hybridization of target DNA analyte with the probe and (D) measuring change in conductance through source and drain electrodes that are attached to silicon nanowires. Figure taken with permission from Gao et al., 2007.

1.5 Project aims and outline of results chapters

The results on the pharmacogenetics of aspirin and warfarin point to the potential clinical utility of genetic variants in improving efficacy and reduce the risk of adverse drug reactions. Moreover, development of the silicon nanowire platform based SNP genotyping platform could aid in the dissemination of clinically useful genetic markers into routine clinical practice. Therefore, the primary aim of this project was to develop a panel of SNPs that would have clinical utility in improving risk-benefit ratio of aspirin in relation to colorectal cancer prevention and warfarin in relation to cardiovascular diseases prophylaxis. Furthermore, the project aimed at developing DNA based oligonucleotide probes and optimizing hybridization conditions to genotype warfarin dose associated SNPs on custom designed microarray platform, which could subsequently be used to carry out ultra-rapid, label free SNP genotyping on the silicon nanowire platform.

In the first chapter (Chapter 3), I will use epidemiological and genome wide SNP data from two large population based case-control datasets to outline association between the known epidemiological risk factors such as BMI, smoking and alcohol consumption etc. with colorectal cancer risk and identify SNPs in aspirin's pharmacokinetic and pharmacodynamic pathways that are either associated with or modify the protective effect of aspirin use on colorectal cancer risk in individual datasets. I will also describe sample size estimates to identify significant associations by carrying out meta-analysis of the two datasets.

In the second chapter (Chapter 4), I will detail the results of association and interaction between the SNP variant allele & colorectal cancer risk and SNP variant allele, aspirin use & colorectal cancer risk respectively from the meta-analysis of the two large population based case-control datasets. I will also discuss drawbacks of current meta-analyses, infer biological implications and causality of the association and interaction results and provide an overview of the clinical utility of potential significant findings.

In the third chapter (Chapter 5), I will test for the clinical utility of published genotype-guided dosing algorithms by measuring their dose prediction accuracy in the Gujarati Indian population that lives on the western coast of India and compare them with the prediction accuracy observed in white European and South Indian populations. I will provide sample size estimate to identify a novel SNP that is specific to the Gujarati Indian population and can explain >5% of dose variability using genome wide association study.

In the last chapter (Chapter 6), I list a panel of custom designed probes which are designed to genotype 3 warfarin dose associated SNPs (rs1799853, rs1057910 and rs9923231), discern the interplay between hybridization conditions and target DNA specificity, delineate dynamics of 2-dimensional environments on hybridization kinetics and

based on the derived optimal experiment conditions, perform genotyping of anonymised DNA samples on a custom designed microarray platform. I will also discuss experiments that would be required to test sensitivity and specificity of the custom designed probe panel on the microfluidics based silicon nanowire genotyping platform.

Chapter 2: Materials and Methods

2.1 Aspirin Pharmacogenetics

2.1.1 Literature review and SNP selection

A comprehensive review of the literature was carried out using PubMed (http://www.ncbi.nlm.nih.gov/pubmed) and Google Scholar (http://scholar.google.co.uk) as search engines. A combination of keywords such as 'aspirin', 'NSAID', 'pharmacogenetics', 'polymorphisms', 'SNPs', 'gene variant' and 'colorectal cancer' was used to search for relevant literature in the databases. All studies presenting original data on SNP and colorectal adenoma or carcinoma risk association, or interaction between SNP and aspirin (or NSAID) use in relation to colorectal adenoma or carcinoma were retrieved and reviewed. The UK-CCSG study collaborators crosschecked candidate SNP selection from the literature. Results from ongoing studies and unpublished data were obtained from UK-CCSG and NIH-CCFR study collaborators and were also included in the analysis. Types of studies reviewed included case-control studies, cohort studies, prospective studies, meta-analysis and randomized controlled trials. Careful consideration during the review was given to the study design, study size, inclusion and exclusion criteria of the study subjects such as type of controls (matched, un-matched or sibling) and case ascertainment (population or family based), regular aspirin and NSAID use definition and type of statistical analysis employed. This helped in comparing results from the current study with published results. Throughout the duration of the study between March 2012 and May 2014, new SNPs were included in the current study based on the latest literature evidence.

2.1.1.1 Study population

2.1.1.2 UK-Colorectal Cancer Study Group

Study design, subject enrolment and data collection was carried out previously by the UK-CCSG study group (Turner et al., 2004, Barrett et al., 2003). Briefly, cases between the age of 45 and 80 years with histologically confirmed incident colorectal cancer and diagnosed in the period of 1997-2013, were identified at each of the three recruitment centers: Leeds (Leeds General Infirmary, St. James's Hospital), Dundee (Ninewells Hospital, Perth Royal Infirmary) and York (York District Hospital). In Dundee and York, where recruitment was carried out between 1997 and 2000, eligible patients were identified via ward diaries and patient notes, whereas, the pathology department in hospitals based in Leeds provided a

monthly update on patients diagnosed with colorectal cancer in previous 4 weeks between 1997 and 2013. Patients who had a primary cancer previously, history of coeliac disease, familial adenomatous polyposis, diverticular disease 2 years before current cancer diagnosis, non-adenocarcinoma colorectal cancer or ulcerative colitis diagnosed in previous 3 years were not recruited in the study. Cases were interviewed by a research nurse to collect epidemiological data and a blood sample. Informed written consent was obtained from cases prior to the interview.

Healthy population-based controls were identified through patient's GP practice list. An age and sex matched control with no history of previous cancer at the time of recruitment was identified for each case between 1997 and 2000 at all 3 study sites. Following 2000, friends or spouse of cases from Leeds with no history of cancer at the time of recruitment were collected for the study. Controls were contacted initially by post along with a standard letter from their GP, an information sheet and an addressed and stamped return envelope. Eligible controls were contacted by phone to arrange a time for an interview with a research nurse at home. Informed written consent, epidemiological data and blood sample was obtained during the interview. The research ethics committee at each participating center approved the study design and protocol (Supplementary Figure 1).

2.1.1.3 NIH-Colon Cancer Family Registry

Study design, subject enrolment and data collection was carried out previously by the NIH-CCFR study group (Newcomb et al., 2007). Briefly, NIH-CCFR employed two types of case ascertainment strategies: population based families from cancer registries and clinic based families from cancer family clinics. Population based cases and controls were enrolled at 6 study centers: Fred Hutchinson Cancer Research Centre (FHCRC), University of Hawaii (UHI), Cancer Care Ontario (CCO), Mayo Clinic (MC), University of Southern California Consortium (USC) and Universities of Queensland and Melbourne (UQM). All study centers used different recruitment strategies and sampling schemes (Table 2.1). For the current study, incident case probands identified through population based cancer registries recruited between 1997 and 2002 were included in the analysis. Depending on the study center, recruitment strategy differed based on age at diagnosis, race and family history of colorectal cancer, for example, only FHCRC invited all eligible cases on the study whereas in UQM, age limit of 18 to 59 years for age at diagnosis was applied in selecting case probands (Table 2.1).

Healthy population based and spouse controls were identified through medicare and driver's license files, telephone subscriber lists and electoral rolls and were randomly selected between 1997 and 2002. Only cases and controls of self reported non-Hispanic white ethnicity were included in the analysis. Informed consent was provided by all study participants and the

study design was approved by the Institutional Review Boards at each NIH-CCFR site. St. James's Hospital in Leeds approved the study design and joint analysis of UK-CCSG and NIH-CCFR data (Supplementary Figure 1).

NIH-CCFR Study Site	Population-based recruitment criteria of cases
Fred Hutchinson Cancer Research Center	• Incident colorectal cancer diagnosed between January 1998 and June 2002. Age at diagnosis between 20 and 74 years. • No family history eligibility criteria. • All colorectal cancer affected first degree relatives were recruited.
University of Hawaii	• Cases diagnosed between 1997 and 2001 with adenocarcinoma of colon or rectum. • Only cases with more than 1 first degree relative with colorectal cancer were recruited. • No single case families were recruited.
Cancer Care Ontario	• Incident colorectal cancer cases diagnosed between July 1997 and June 2000 with age at diagnosis between 20 and 74 years. • All cases met Amsterdam-I criteria and were part of multiple case families. • All cases with FAP were excluded.
Mayo Clinic	• Cases diagnosed between 1997 and 2000. • Cases were either from multiple living-case families and were between the age of 18 and 74 years or were diagnosed under the age of 50 years with no family history criteria.
University of Southern California Consortium	• Cases diagnosed between 1997 and 1999 with the age range of 21 and 75 years. • Screened for 33% of whites >50 years, 66% of whites <50 years and 66% of all minorities between 21-75 years.
University of Queensland and	• Cases diagnosed between 1997 and 2001 with primary adenocarcinoma of colon or rectum.

Melbourne	• Cases diagnosed between the age of 18 and 59 years.

Table 2.1 Population-based case recruitment strategies employed by the NIH-CCFR study sites.

Table adopted from Newcomb *et al.* 2009.

2.1.2 Epidemiological questionnaire

2.1.2.1 UK-Colorectal Cancer Study Group

A research nurse carried out interviews of the study participants either in hospital or at home. Interviewees completed detailed diet and lifestyle questionnaire called the Food Frequency and Epidemiology Questionnaire (FFEQ) (Barrett et al., 2003), which was modeled on the questionnaire developed and validated by the European Prospective Investigation into Cancer and Nutrition (Kaaks et al., 1997). The questionnaire included information on suspected and established risk factors of colorectal cancer such as: medical history, medication use (including aspirin and NSAID use), reproductive history of female participants, physical activity, demographics, alcohol and tobacco use, race and ethnicity and detailed dietary data. Completed questionnaires were sent to University of Leeds for manual entry into a customized electronic database. Some of the questionnaires were checked for data entry accuracy in the database (Barrett et al., 2003). Regular use of aspirin or NSAIDs (such as ibuprofen, naproxen, diclofenac, indomethacin and pyroxicam) was defined as taking pain killer medication for 3 months or longer before the interview for controls and cancer diagnosis in cases. Other information such as number of pills per day, dose strength (mg) and duration of intake was also collected. All epidemiological data was collected previously by the UK-CCSG study group and was provided for the current study.

2.1.2.2 NIH-Colon Cancer Family Registry

Each study participant completed a standardized family history, personal exposure and baseline epidemiologic questionnaire either in person (USC), by telephone (FHCRC, USC, UQM) or by mail (UHI, CCO, MC) (Newcomb et al., 2007). The questionnaire included medical history and medication use, reproductive history of female participants, physical activity, demographics, alcohol and tobacco use, race and ethnicity and limited dietary data. Questionnaires were customized by the participating centers for local usage, in particular for different language conventions and brand names, and added some questions of local interests. Copy of the questionnaires from all centers can be downloaded from http://www.coloncfr.org/questionnaires. Regular use of aspirin or NSAIDs (such as naproxen, ketoprofen, diclofenac, ibuprofen, sulindac, pyroxicam and indomethacin) was defined as

regular use for at least twice per week for 1 month or more. Additionally, information on number of pills per day and duration of intake was also collected but dosing strength information was not collected. All epidemiological data was collected by the NIH-CCFR group previously and was made available for the current study.

2.1.3 SNP genotyping

2.1.3.1 UK-Colorectal Cancer Study Group

Venous blood sample from cases and controls was collected by the research nurse at the time of the interview, or shortly after, which was stored in an EDTA vacutainer tube at -20°C. Genomic DNA was extracted from leukocytes using Nucleon BACC2 Genomic DNA extraction kit at the study sites (Gen-Probe Life Sciences, Manchester, UK). An aliquot of genomic DNA sample was sent to either Tepnel Pharma Services Ltd. (Manchester, UK) or Wellcome Trust Clinical Research Facility (Edinburgh, UK) for genotyping >240,000 SNPs using the Illumina HumanExome BeadChip array V1.1 (Illumina, San Diego, USA). SNPs were automatically called using the Illumina GenomeStudio data analysis software (Illumina, San Diego, USA). Overall, genotyping call rate for the samples was 97.21%. A 100% match for the genotype call was observed for 64 sample replicates that were genotyped at both facilities. All the genotyping work was carried out by the UK-CCSG study group previously and the genotype data was made available for analysis.

SNPs with a minor allele frequency (MAF) of >3% that were absent on the Illumina HumanExome BeadChip array v1.1 were genotyped using TaqMan drug metabolizing genotyping assay for allelic discrimination (Applied Biosystems, Paisley, UK) at St. James's Hospital, Leeds (Table 2.2). To carry out the assay, genomic DNA samples were robotically replica-plated in a series of 96-well daughter plates. The end-point fluorescence was read using an ABI PRISM 7700 sequence detection system (Applied Biosystems, Paisley, UK) and analyzed using Sequence Detector Software v1.7a. For quality control, each 96 well plate included previously analyzed samples representative of each genotype where the genotype had been verified by sequencing along with multiple no-template control samples. In addition, 1% of the samples were selected at random for repeat analysis. Overall, the failure rate was <2%. Primer design, optimization and genotyping assay was carried out by the UK-CCSG study group and the genotype data was made available for analysis.

2.1.3.2 NIH-Colon Cancer Family Registry

Peripheral blood was collected using standardized procedures from cases and controls (Newcomb et al., 2007). Genomic DNA was extracted from leukocytes using the protocol of Lum and LeMarchand (Lum and Le Marchand, 1998) and quantified using the dsDNA

PicoGreen kit (Invitrogen, Paisley, UK). DNA samples were genotyped at USC on 3 separate platforms: Illumina 1M, Illumina 1M-Duo and Illumina HumanOmni1 arrays (Illumina, San Diego, USA) by the NIH-CCFR study group previously (Table 2.2). SNPs were automatically called using the Illumina GenomeStudio data analysis software (Illumina, San Diego, USA). Genome wide SNP data was provided by Prof. Graham Casey and was stored on a server at St. James's Hospital, Leeds. SNP rs20417 was previously imputed at USC and thus the imputed genotype data was provided by them for analysis. Before the analysis, SNPs that were absent on the arrays were manually imputed by using proxy SNPs (linkage disequilibrium R^2=1.0) from HapMap II CEU population (Table 2.2).

SNP ID	Genotyping platforms for UK-CCSG dataset		Genotyping platforms for NIH-CCFR dataset		
	Illumina Human Exome Array BeadChip v1.1	Taqman Allelic Discrimination Assay	Illumina Human 1M Array	Illumina Human 1M-Duo Array	Illumina Human Omni1 Array
rs1045642		✓	✓	✓	✓
rs1321311	✓		✓	✓	✓
rs1057910	✓		✓	✓	
rs1799853	✓			✓	✓
rs6983267	✓		✓	✓	✓
rs961253	✓		✓		
rs11694911		✓	✓	✓	✓
rs28362380		✓	✓	✓	Proxy SNP rs1405948
rs4936367	✓		✓	✓	Proxy SNP rs1351452
rs7112513	✓		✓	✓	Proxy SNP rs1351452
rs3842787	✓		✓	✓	✓
rs20417		✓		✓	✓
rs2070959	✓		✓	✓	✓
rs1105879	✓		✓	✓	✓
rs2619112		✓	✓	✓	✓
rs10958713		✓	✓	✓	✓
rs11986055		✓	✓	✓	✓
rs12910333		✓	✓	✓	✓

rs5995355		✓	Proxy SNP rs6000449	Proxy SNP rs6000449	Proxy SNP rs6000449
rs230490		✓	Proxy SNP rs1313925	Proxy SNP rs1313925	✓
rs5275			✓	✓	✓
rs4648310		✓	✓	✓	
rs5029748			✓	✓	
rs2745557		✓	✓		✓
rs6474387				✓	
rs16973225		✓			✓
rs2302615		✓			
rs2430420		✓			
rs5277		✓			✓
rs2965667		✓			
rs140461033	✓				
rs144410046	✓				
rs201103548	✓				
rs28382815	✓				
rs148026549	✓				
rs145407778	✓				
rs10852434	✓				
rs147942040	✓				
rs141625476	✓				
rs147070911	✓				
rs150408050	✓				
rs147694237	✓				
rs142710583	✓				
rs185651296	✓				
rs186808413	✓				
rs78428934	✓				

Table 2.2 List of SNPs genotyped across different platforms in UK-CCSG and NIH-CCFR datasets.

2.1.4 Statistical Analysis

Before analyzing UK-CCSG and NIH-CCFR datasets, common data elements (CDEs) were defined since each study employed unique recruitment and data collection protocols. Questionnaires and data dictionaries were examined to identify CDEs (eg. Age, height, weight, BMI, smoking, alcohol intake, exercise, aspirin/NSAID use) to produce common

definitions, standardized values and coding terms (Table 2.3). Resulting dataset was reviewed using logic checks to test for data distribution between and within two studies. Outlying sample data were removed from all downstream analyses or that variable was normalized. Data on individuals of non-Hispanic white ethnicity only was tested in all subsequent downstream analyses. Variables such as age and BMI etc. were coded as continuous variables whereas variable on aspirin/ NSAID or aspirin-only regular use were coded as dichotomous (0 and 1) variables. Non- regular users were used as the reference. The associations between known epidemiological risk factors (such as BMI, smoking, family history etc.), aspirin or NSAID use and colorectal cancer risk were assessed using conditional logistic regression models estimating odds ratios (OR) and 95% CI with two sided P-value adjusted for age, sex and study site.

Genotyped SNP data from UK-CCSG and NIH-CCFR were compared for the SNPs that were common between platforms in both studies (Table 2.2). Each genotyped SNP was coded as 0, 1 or 2 for the number of copies of variant allele and the imputed SNPs were coded based on the "expected" number of copies of variant allele. SNPs were excluded if they were tri-allelic, call rate (<98%), minor allele frequency (MAF) of <4% and inconsistency with Hardy Weinberg equilibrium (HWE) in controls ($p<0.001$, corrected for multiple testing). Comparison of SNP MAF between controls in two datasets was carried out using Fisher's exact test to check for genotyping consistency and population substructure. SNPs on the same chromosome were tested for linkage disequilibrium (R^2) in controls in both datasets. The associations between SNP genotype and colorectal cancer risk were tested using conditional logistic regression models estimating ORs and 95% CI with two sided P-value adjusted for age, sex and study site. Interaction of SNP genotype with only aspirin or NSAID use between regular users and non-users in relation to colorectal cancer risk was investigated. The two reference groups comprised of individuals who had homozygous wild type genotype and were either non-users or only aspirin (and/or NSAID) users. Gene-environment (G x E) interaction was tested using case- control logistic regression and cross product of the presence of variant allele and dichotomous regular use of NSAID or aspirin-only. The likelihood ratio test was used to determine the statistical significance of the interaction. Interaction P-values were adjusted for age, sex and study site. The significance threshold for type 1 error (P-value) in all tests was set at 0.05. For each SNP, the risk of colorectal cancer, colon cancer (combining proximal and distal colon) and rectal cancer was estimated using a dominant inheritance model. A co-dominant or recessive inheritance model was not used, as the sample size was relatively small in both datasets. All analyses were conducted in Stata v12 for Macintosh OS (Stata Corp., College Station, USA).

Common Data Element	UK-CCSG definition	NIH-CCFR definition	Standardized definition	Coding terms
Age	Cases: age at cancer diagnosis, Controls: age at interview	Cases: age at cancer diagnosis, Controls: age at interview	Cases: age at cancer diagnosis, Controls: age at interview	Continuous variable
Sex	Male coded as 0, female coded as 1	Male coded as 1, female coded as 2	Male coded as 0, female coded as 1	0 for male, 1 for female
BMI	BMI at 1 year before interview	BMI at 2 years before the interview	Keep respective study definitions	Continuous variable
BMI at 20 years of age	BMI at 20 years	BMI at 20 years	BMI at 20 years	Continuous variable
Regular Smoking+	Smoked 1 cigarette a day for 1 year	Smoked 1 cigarette a day for 3 months	Smoked 1 cigarette a day for 3 months	0 for non-smoker, 1 for smoker
Smoking before diagnosis+	Regular smoking 1 year before diagnosis	Regular smoking 1 year before diagnosis	Regular smoking 1 year before diagnosis	0 for non-smoker, 1 for smoker
Alcohol intake^	Alcohol intake of ≥1 drink a week at the age 40 years	Alcohol intake of ≥1 drink a week for at least 6 months at the age 30-40years	Alcohol intake of ≥1 drink a week at the age 30-40 years	0 for non-drinker, 1 for drinker
Alcohol intake unit^	Alcohol units based on intake quantity	Alcohol units based on intake quantity	Alcohol units based on intake quantity	Continuous variable
Exercise	Hours spent on activities in a week a year ago from interview	Hours spent on activities in a week in 30s and 40s	Keep respective study definitions	Continuous variable
Aspirin/ NSAIDs use	Regular aspirin/ NSAIDs use for 3 months or longer	At least twice a week for more than a month	Keep respective study definitions	0 for non user, 1 for user
Calorie intake	Based on average diet 1 year before the interview	Based on average diet 2 years before the interview	Keep respective study definitions	Continuous variable
Family risk	First degree and/or second degree relative with CRC	First degree and/or second degree relative with CRC	First degree and/or second degree relative with CRC	0 for no, 1 for yes

Table 2.3 Standardized definitions of common data elements between the UK-CCSG and NIH-CCFR datasets.

+ Smoking includes cigarettes, cigar and pipes

^Alcohol includes beer, cider, wine, sherry, other fortified wine, sake, champagne and spirits

2.1.5 Random effects meta-analysis

Log odds and standard error were estimated using unconditional logistic regression and a dominant inheritance model for the SNPs having significant threshold (P<0.08) for association with CRC (Table 3.4) or interaction with aspirin only use and CRC (Table 3.6) in either the UK-CCSG or NIH-CCFR datasets. Dominant inheritance model was used to increase the power of observing significant association. The associations were further tested using the co-dominant inheritance model to test for the robustness of associations observed using the dominant model. Estimates of log odds and standard error from both datasets, which were adjusted for age, sex and study site, were manually entered in a Microsoft Excel file which was used as the dataset to carry out random effects meta-analysis in Stata v12 for Macintosh OS (Stata Corp., College Station, USA). –metan- command was used to calculate the meta-analysis odds ratio, 95% confidence interval and two sided p-value. The –metan- command conducts meta-analysis of the data from more than one study by assessing the effect estimates (log odds) with corresponding standard errors and display the results graphically in a forest plot. To test for heterogeneity between the estimates from the two datasets, Cochran's Q-test was calculated. However, since Cochran's Q-test is not accurate for testing heterogeneity in meta-analysis studies with few datasets, Higgin's I-squared statistic (Higgins et al., 2003), which represents the percentage of variation between the estimates as a result of heterogeneity, was also calculated. The significance threshold for type 1 error (P-value) in all tests was set at 0.05.

2.1.6 Power calculation

To estimate the number of cases and controls that are required for an association test between SNP genotype and colorectal cancer risk and interaction test between SNP genotype, aspirin (and NSAID) use and colorectal cancer risk, power calculation was carried out using Quanto v1.2.4 (Quanto; hydra.usc.edu/gxe/). For association test, the design was set as "un-matched case-control" with 1 control per case ratio and the hypothesis as "Gene only". MAF was set as 0.10 and the inheritance pattern was assumed to be dominant. Baseline disease risk (P_0) was set at 0.05 and the R_G, which is the expected OR, was set between 1.20 and 2.0. The power to detect a SNP associated with colorectal cancer risk was set at 80% with the type 1 error rate set at 0.05. However, for the interaction test, the design was set as "un-matched case-control" with 1 control per case ratio and the hypothesis as "Gene-environment interaction". MAF was set as 0.10 and the inheritance pattern was assumed to be dominant.

Based on the number of aspirin users in the UK-CCSG and NIH-CCFR datasets, the binary environmental factor prevalence (P_E) was set at 0.25 with baseline disease risk (P_0) set at 0.05. R_G, R_E and R_{GE} which are the expected OR value for gene, environment and gene-environment interaction respectively was set at 1.20, 0.75 and 1.50. The power to detect an interaction between SNP and only-aspirin (or NSAID) use in relation to colorectal cancer risk was set at 80% with the type 1 error rate set at 0.05.

2.2 Warfarin Pharmacogenetics

2.2.1 Data collection and study population

SNP frequencies were analyzed from 501 individuals obtained from 2 sources. The first source consisted of genomic DNA samples from a cohort of 399 unrelated healthy individuals of self-reported Gujarati ethnicity, contributed by the Institute of Human Genetics (IHG), Ahmedabad, India. No clinical information was available for these samples and they were therefore only included in the investigation of SNP frequency distribution. The requirement for an informed consent was waived for them because the consent had been obtained previously by the IHG from the individuals for carrying out genetic studies.

The second source consisted of 102 patients of self-reported Gujarati ethnicity who were treated with warfarin. They were recruited between 1st November 2012 and 1st May 2013 at the Care Institute of Medical Sciences (CIMS) hospital, Ahmedabad, India. Clinical information including age, gender, height (in cm), weight (in kg), current smoking status, clinical indication for warfarin treatment, current warfarin dose (mg/day), initiation dose (mg/day), latest INR, concomitant medications (including any herbal medication) and adverse drug reactions were recorded by a research nurse at CIMS hospital. Records of routine INR values were also obtained for 88 patients. Therapeutic INR range was defined as being between 2 and 3 for patients with atrial fibrillation (AF), deep vein thrombosis (DVT), pulmonary embolism (PE) and left ventricular ejection fraction (LVEF); and 2.5 to 3.5 for patients with mechanical heart valves.

Stable or therapeutic warfarin dose was defined as the dose at which the 2 consecutive INR measurements, each being 1 week apart, were within the therapeutic range (Definition based on Klein et al. 2009). All 102 patients provided informed consent prior to enrolment in the study. During the interview, research nurse collected finger prick blood sample from 102 unrelated patients on Whatman FTA cards (Whatman plc, Kent, UK). The Institutional Ethics Committee at CIMS and IHG approved the study (Supplementary Figure 4 and Supplementary Figure 5).

2.2.2 Genotype analysis and quality control

Genomic DNA was isolated from Whatman FTA cards using ZyGEM *prepGEM* storage card blood kit (ZyGEM Corporation Ltd, Hamilton, New Zealand). Briefly, two 1.2mm^2 punches were made using a Harris micro punch. These were transferred to a 0.5ml safe-lock microfuge tube (Eppendorf, Hamburg, Germany) containing 100µl of MilliQ water and stored at room temperature for 15 minutes. Water was aspirated and the punches were re-suspended in 44µl of MilliQ water, 5µl of ZyGEM MAGENTA buffer and 1µl of *prep*GEM

enzyme. This solution was placed in a thermal cycler at 75°C for 15 minutes followed 95°C for 5 minutes. Tubes were centrifuged at 16,000g for 5 minutes and the supernatant containing genomic DNA was transferred to a labeled 1.5ml safe-lock tube (Eppendorf, Hamburg, Germany). Assessment of the quantity and quality the DNA extracted using ZyGEM *prepGEM* kit was not carried out since the extraction process didn't involve a cleanup step to remove cell debris and proteins. Since only spectrophotometric techniques were available to quantify DNA, which are prone to providing inaccurate results in the presence of protein contamination, DNA quantification was not carried out.

DNA samples were genotyped for the 4 SNPs in *CYP2C9* (8633C>T= rs1799853; 47639A>C= rs1057910; 47644C>G= rs28371686; 15625delA= rs9332131), 4 SNPs in *VKORC1* (3588G>A= rs9923231; 5332G>T= rs61742245; 8956G>A= rs7294; 5924C>T= rs17708472), 1 SNP in *CYP4F2* (23454G>A= rs2108622), 1 SNP in *GGCX* (16025G>C= rs11676382), 1 SNP in *CALU* (24879A>G= rs339097), 1 SNP in *CYP3A4* (25343G>A= rs2242480) and 1 SNP in *CYP2C19* (3583C>T= rs3814637). PCR and Sequenom iPLEX primers were designed for a multiplex reaction of 10 and 3 SNPs using Sequenom's Assay Designer 4.0 software (Sequenom, Hamburg, Germany) (Supplementary Table 5). Samples were amplified using Qiagen HotStar *Plus* DNA polymerase kit (Qiagen, Hilden, Germany) in 50ul reactions and the PCR reaction mixture was prepared according to Sequenom's recommendation (Sequenom, 2004). PCR amplicons generated from the DNA of 101 warfarin treated patients were purified using QIAquick PCR Purification kit (Qiagen, Hilden, Germany) to remove cell debris generated during the DNA isolation step, which may inhibit accurate genotype call to be made by Sequenom MassARRAY platform. DNA extraction and PCR amplification steps were carried out by the PhD candidate whereas genotyping by mass spectrometry with the use of Sequenom MassARRAY (Sequenom, Hamburg, Germany) was carried out at the High Throughput Genomics Group, Wellcome Trust Centre for Human Genetics, Oxford University.

Genotype calls from Sequenom MassARRAY were validated using restricted fragment length polymorphism (RFLP) PCR in a randomly selected 10% of the total number of samples for 5 SNPs (rs1799853, rs1057910, rs9923231, rs7294 and rs2108622). Missing genotype calls by Sequenom for the 3 SNPs (rs1799853, rs1057910 and rs9923231) in 53 of 102 patients treated with warfarin that reached therapeutic dose were re-genotyped using RFLP PCR for downstream analyses. Each SNP was genotyped with a specific PCR protocol and site-specific restriction enzyme (New England BioLabs, MA, USA) (Supplementary Table 6). Electrophoresis of 20µl of the digested PCR products was carried out according to the manufacturers' protocol on 2% E-gel pre-stained with SYBR Safe DNA Gel stain

(Invitrogen, New York, USA) and visualized with ultra violet light using i-Base (Invitrogen, New York, USA). All RFLP PCR based genotyping was carried out by the PhD candidate.

2.2.3 Statistical Analysis

Accordance with the Hardy-Weinberg equilibrium for polymorphisms was analyzed using the χ^2- test. Population allele frequencies between the 101 patients and 399 healthy individuals were compared using Fisher's exact test for concordance before merging the two datasets for further analyses. Linkage disequilibrium (LD) score between SNP pairs was carried out using PLINK v1.07 software (www.pngu.mgh.harvard.edu/~purcell/plink). Genotype frequency data for the European (EUR), Gujarati Indian from Houston (GIH), African (AFR) and Han Chinese (CHB) population was obtained from the 1000 Genomes database [Genome assembly: GRCh37] (browser.1000genomes.org/index.html). Genotype frequencies of the four populations were compared with the Gujarati population using Fisher's exact test. Fifty-three patients who had reached stable therapeutic warfarin dose were used to test the clinical utility of the published clinical and pharmacogenetic algorithms by Klein et al. 2009, Gage et al. 2008 and Pavani et al. 2012 (Pavani et al., 2012, 2009, Gage et al., 2008). Clinical and genotype data from the patients were used to predict therapeutic dose using these algorithms (Supplementary Information 1 (Klein et al., 2009 algorithm), Supplementary Information 2 (Gage et al., 2008 algorithm), Supplementary Information 3 (Pavani et al., 2012 algorithm)). Data was organized as per the requirements of the algorithm before carrying out the analysis. However, for the Pavani et al. 2012 algorithm, rs11676382 and rs7900194 SNPs in the *GGCX* and *CYP2C9* gene respectively were not included in the analysis as rs11676382 had <1% minor allele frequency (MAF) and rs7900194 had not been genotyped in this study. Missing genotype calls for rs7294 and rs17708472 SNPs were inputted as "0" in the algorithm. Predicted dose was compared with empirical therapeutic dose using ordinary linear regression to calculate coefficient of determination (R^2). R^2 value was used to compare the explained variance of warfarin dose by clinical and pharmacogenetic algorithms within a population.

Anonymised clinical and genetic data of European patients (part of the International Warfarin Pharmacogenetics Consortium (IWPC)) from which Klein et al. 2009 and Gage et al. 2008 pharmacogenetic algorithms were developed, was obtained from PharmGKB website (www.pharmgkb.org) whereas, South Indian patient data for the Pavani et al. 2012 algorithm was provided by the corresponding author. To compare the dose prediction accuracy of algorithms in the Gujarati Indian with European and South Indian population, mean squared difference (MSD) between the predicted and therapeutic dose was compared using Mann-

Whitney test. A P- value of ≤0.05 was considered statistically significant. All statistical analysis apart from LD score was carried out in Stata v12.0 (Stata Corp., Texas, USA).

2.2.4 Power calculation

To estimate the number of patients being treated with stable warfarin dose that are required for a genome wide association study (GWAS) analysis for future study, power calculation was carried out using Quanto v1.2.4 (Quanto; hydra.usc.edu/gxe/). The design was set as "Independent individuals" and the hypothesis as "Gene only". MAF was set as 0.05 and the inheritance pattern was assumed to be log additive. The R^2_G, which is the measure of coefficient of determination, was set between 0.05 and 0.30. The power to detect a SNP that explained a minimum of 5% variance of warfarin dose was set at 80% with the genome wide two-sided significance level of $5x10^{-8}$ (significance level based on Perera et al. 2013).

2.3 QuantuMDx Silicon Nanowire Technology

2.3.1 Probe design

Wild-type target sequences for probe design were obtained from the UCSC genome browser (GRCh37/ hg19; http://genome.ucsc.edu/) in FASTA format. Target sequence length was set between 30-40bp with the resulting sequence flanking either side of the SNP 15-20bp long. This would allow for designing probes with the SNP positioned at various locations within the probe. Isothermal melting probes were generated using OligoWiz2.2 online software (http://www.cbs.dtu.dk/services/OligoWiz2/) that was last visited in September 2011. After uploading the FATSA file, a series of parameters were set to calculate the scores of the probes (Wernersson et al., 2007). To calculate cross hybridization and low-complexity score, *H. sapiens* was chosen from the species database. Parameters like probe length, melting temperature (T_m) and cross hybridization were set according to the guidelines mentioned in (Gresham et al., 2010). The optimal length of the probe was set at 21bp for the 2 *CYP2C9* SNPs and 16bp for *VKORC1* SNP to generate probes with similar T_m as the GC content between the *CYP2C9* and *VKORC1* sequence varied. Lowest and highest length was set at 18 and 23bp for the *CYP2C9* SNPs and 12 and 18bp for the *VKORC1* SNP respectively.

Optimum T_m was set at 51°C with hybridization chemistry being DNA:DNA, cross hybridization minimum length of homology stretch was set at 0 and position was set at random prime. It is important to note that OligoWiz software calculates T_m using the Nearest Neighbor algorithm. The resulting probes were then further screened by setting the total score cut-off value to 0.4 and the minimum distance between oligos to 1. These parameters filtered out an array of probes, which were then exported in either FASTA or TAB format. From this array, probes having a T_m 51+/-1.5 °C were filtered out manually while the rest were discarded. These probes were then tested in OligoCalc V3.26 online software (http://www.basic.northwestern.edu/biotools/oligocalc.html) for potential hairpin formation, 3’ complementarity and potential self-annealing sites. Probes having potential for generating any of the 3 structures mentioned above were discarded and the rest of the probes were loaded on to ClustalX2.1 software (http://www.clustal.org/clustal2/) to observe the location of the SNP within the probes. Probes having SNP located between the center of the probe and 2bp away from the either end were filtered out whereas the rest were discarded. Filtered probes were then uploaded on the UCSC genome browser BLAT function (http://genome.ucsc.edu/cgi-bin/hgBlat?command=start) to confirm homology of the query sequence with the human genome. Lastly, mutant probes were manually generated by changing the SNP nucleotide of the selected probes and the final probe sets were given unique

identity numbers and names (Supplementary Table 9). Probe sequences with 5' C6-amino group modification were sent for synthesis to BioBasic (New York, USA) and 100nmoles of HPLC purified and lyophilized oligos were shipped to Newcastle.

2.3.2 PCR primer design

PCR primers were generated using a combination of 3 internet-based software packages. DNA sequence was retrieved from the UCSC genome browser (GRCh37/ hg19; http://genome.ucsc.edu/), which was uploaded on the Primer 3 software V0.4.0 (http://frodo.wi.mit.edu/), which provided with putative primer pairs. These were tested for absence of primer dimers and primer loops using OligoCalc V3.26 software (Supplementary Table 10). The length of the amplicon was kept between 140-250 bp. Primer sequences with 5' Cy3 label modification were sent to Metabion (Hilden, Germany) and 20nmoles of HPLC purified and lyophilized oligos were shipped to Newcastle. These were used to generate Cy3 labeled PCR amplicons.

2.3.3 Generating PCR amplicons

One hundred genomic DNA samples, which were previously genotyped for the three SNPs were provided by Prof. Ann Daly, Newcastle University and were then used as templates for PCR products. Sample GU45 which was genotyped as homozygous wild-type for all 3 warfarin SNPs was used to test the microarray platform. Since no homozygous mutant sequence for the three SNPs were available to test on the microarray platform, homozygous mutant sequence for all 3 SNPs were synthesized (Supplementary Table 11) and received lyophilized after being cloned in pEX-A2 vector (MWG Eurofins, Ebersberg, Germany). Cy3 labeled amplicons of *CYP2C9*2*, *CYP2C9*3*, *VKORC1* and *ACTB* were generated using Cy3 labeled primers. Qiagen HotStart *Plus* 1000U (Qiagen: Hilden, Germany) PCR kit and SensQuest LabCycler thermal cycler was used to generate PCR products. PCR reaction was carried out in 100µl solution consisting of 0.1 ng/µl of template DNA, 10µl of Qiagen 10X buffer, 200µM dNTP mix, 0.5µM of each primer and 1 U of HotStart *Plus Taq* polymerase. The PCR was performed with initial denaturation at 95°C for 5 minutes followed by 35 cycles of denaturation (95°C for 30 seconds), annealing (58°C for 30 seconds), extension (72°C for 30 seconds) and final extension (72°C for 2 minutes).

Electrophoresis was carried out on 2% E-gel (Invitrogen; New York, USA) and PCR products were viewed using i-Base (Invitrogen; New York, USA). Amplicons obtained after PCR were purified using QIAquick PCR purification kit (Qiagen: Hilden, Germany) to remove any residual primers, nucleotides, enzyme, salts and other impurities that may hinder the microarray experiment. A batch of PCR amplicons that were generated with primers not

labeled with Cy3 were sent to GATC Biotech (London, UK) for dual pass Sanger sequencing to ensure that the primers were amplifying the region of interest.

2.3.4 Custom microarray slide layout and printing probes

Slide layout design and probe spotting was carried out by ArrayJet Ltd. (Edinburgh, UK). The probe oligos were spotted on 20 epoxy silane layered glass slides. Sixty-six lyophilized 5' amine terminated oligos were provided to ArrayJet. They were re-suspended in water to a concentration of 100µM. Each oligo was printed in triplicate and was spotted in 5 different concentrations: 0.1, 0.5, 1.0, 5 and 20µM. Each slide was divided in to 8 miniarrays and each miniarray consisted of 66 oligos in 5 concentrations in triplicate and included ethylene glycol buffer spots as negative controls. Each drop dispensed on the slide was 100pL in volume containing the desired oligo. Two drops per spot were printed. Horizontal gap and vertical pitch between the spots was kept at 250µm. Printing was carried out at between 45-55% relative humidity and 20-22°C. Slides were incubated in the humidity chamber for 30 minutes after printing to favor DNA immobilization. Following that, slides were sent to Newcastle and were stored at 4°C.

2.3.5 Surface blocking of microarray slides

Slides were dried at 90°C for 90 minutes in the N-Biotek NB-205 shaking incubator (N-Biotek, Gyeonggi-do, Korea). Blocking was carried out with blocking solution containing 50mM ethanolamine (Sigma Aldrich, Dorset, England) in 0.1% sodium dodecyl sulfate (SDS) (Sigma Aldrich, Dorset, England) and 0.1M Tris at pH 9.0 (Sigma Aldrich, Dorset, England) at 50°C for 15 minutes. Slides were rinsed with MilliQ water for 1 minute and were dried by centrifugation.

2.3.6 Target DNA hybridization and slide scanning

Eight miniarrays on a slide were simultaneously divided and sealed using an 8 well ProPlate gasket (Grace Bio-Labs, Oregon, USA). Cy3 labeled DNA amplicon stocks were diluted to 10uM with MilliQ water. 200ul of hybridization solution containing PCR amplicon, 3x saline sodium citrate (SSC) buffer (Sigma Aldrich, Dorset, England) and 0.1% SDS was added to each miniarray. Hybridization solution containing amplicons was heated to 95°C for 3 minutes prior to adding it to the miniarray. Microarrays were sealed and kept in a moist shaking hybridization chamber at a specific temperature and 50 rpm for 16 hours. Slides were washed using a series of 300ml of SSC washing buffers with increase in stringency with every wash. Washing was carried out at 115 rpm in a shaking incubator with temperature set at 30°C. Slides were dried by centrifugation and were scanned using Axon Genepix4000B scanner (Molecular Devices, California USA) and Genepix Pro6.0 software. Scanner's PMT

gain was set to 625 and laser power was set at 100% in all experiments. Spot intensity data along with the probe spot information in each miniarray on the microarray slide was exported in a text file.

2.3.7 Statistical analysis

Spot intensity data from the text file was exported to a Microsoft excel sheet where data from each miniarray was separated manually. To calculate true intensity of each probe spot, median value of the intensity of the pixels representing the probe spot was subtracted by the median value of the intensity of the pixels representing the background around the probe spot. Since each probe was spotted in triplicate, an average value of the intensity for the probe was calculated. Background intensity for a miniarray was calculated using the intensity signals from *ACTB* control probes using the formula: average intensity values of probes (in a miniarray) + 3X standard deviation of intensity of probes. Probes whose mean intensity value was lower than that of the background intensity were replaced with background intensity value. Specificity, defined as the efficiency of a probe to identify and discriminate target for a given DNA sequence, was calculated as the ratio of the intensity of a probe when hybridized to complementary DNA compared to non-complementary DNA. All data sorting and statistical analysis was carried out in Microsoft excel.

Chapter 3. (Results 1): Pharmacogenetic influences on colorectal cancer chemoprevention using aspirin- Part 1

3.1 Introduction

3.1.1 Development of aspirin

Acetylsalicylic acid (ASA), also commonly known as aspirin, is a non- steroidal anti-inflammatory drug (NSAID) which is used as an analgesic, anti-pyretic or as a prophylactic drug for cardiovascular diseases (CVD) (Fuster and Sweeny, 2011). Aspirin is made up of two components: a phenol ring consisting of 6- carbon benzene ring with a carboxyl group and an acetyl moiety (Figure 3.1) (Fuster and Sweeny, 2011). In 1853, Charles Gerhardt created the first chemically synthesized compound of ASA but it was found to be impure and unstable ((Gerhardt, 1853) cited in (Fuster and Sweeny, 2011)). However in 1897, chemist Arthur Eichengrün assisted by Felix Hoffman at Friedrich Bayer & Company (now known as Bayer Pharma) obtained ASA in its purest form by acetylating the phenol group. This compound demonstrated desirable analgesic and anti-pyretic properties but had a low risk of gastric irritation. It was first introduced into the market in 1899 and was registered under the name of “aspirin” (Fuster and Sweeny, 2011). Currently, an estimated 40,000 tons of aspirin is produced worldwide annually and approximately 10 to 20 billion tablets is consumed annually in the USA alone for CVD prophylaxis (Campbell et al., 2007), making it one of the most widely used drugs in the world.

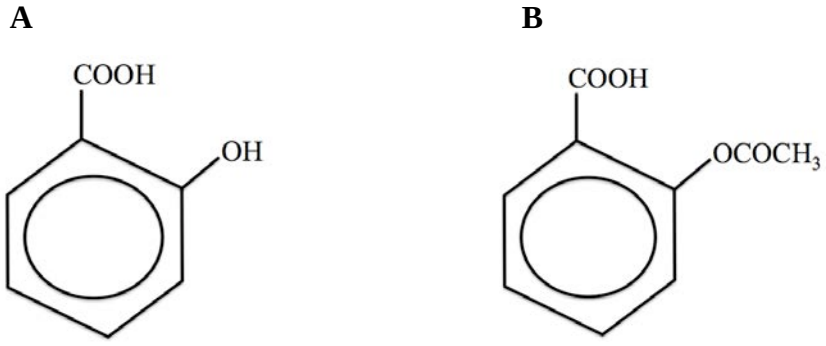


Figure 3.1 Two dimensional chemical structure of (A) Salicylic acid and (B) Acetylsalicylic acid.

Figure adopted from NCBI PubChem (NCBI, 2014).

3.1.2 Aspirin and colorectal neoplasia

Aspirin is widely prescribed for its anti-platelet and anti-inflammatory effects, however in 1988, the first epidemiological evidence for an inverse association of its intake with colorectal cancer risk was observed in a population based case- control study (Relative risk [RR] for males=0.58, 95% CI=0.38-0.88, p=0.02; RR for females=0.49, 95% CI=0.32-0.73, p<0.01) (Kune et al., 1988). Colorectal cancer is the third most common cancer in men and second most common in women with an estimate of 1.2 million new cases and 609,000 deaths worldwide in 2008 (Jemal et al., 2010). Thus, primary prevention of colorectal cancer is a priority. More recently, a meta-analysis carried out by Bosetti et al., 2006 showed an inverse association between long term aspirin use and colorectal cancer after pooling the risk estimates from 11 case- control studies (RR=0.59, 95% CI=0.54-0.64, p for heterogeneity= 0.008) and 7 cohort studies (RR=0.85, 95% CI=0.78-0.92, p for heterogeneity= 0.006) (Cuzick et al., 2009, Bosetti et al., 2006).

However, a significant heterogeneity was observed in the estimates between and within case- control and cohort studies as the studies were carried out in different populations, used different methods for case ascertainment, and used different types of controls. Two large scale randomized 2 x 2 factorial trials, the Physician's Health Study and the Women's Health Study, examined the effects of long term low dose aspirin use on the incidence of colorectal cancer in healthy men and women respectively. In the Physician's Health Study, 11037 and 11034 men were randomized to 325mg of alternate day aspirin and aspirin placebo respectively for mean treatment duration of 5 years. After an early termination of the trial follow up at 5 years due to a significant reduction in myocardial infarction incidence in the aspirin group, no significant association between aspirin use and colorectal cancer incidence was observed (RR=1.15, 95% CI=0.80-1.65) (Gann et al., 1993). Similarly in the Women's Health Study, 19934 and 19942 women were randomized to 100mg of alternate day aspirin and aspirin placebo respectively for mean treatment duration of 10 years. After a mean follow up duration of 10.1 years, no significant association between aspirin use and colorectal cancer incidence was observed (RR=0.97, 95% CI=0.77-1.24, p=0.83) (Cook et al., 2005).

In 2008, Burn et al. described the first randomized 2 x 2 factorial double-blinded trial in 1071 patients who were genetically predisposed to the hereditary form of colorectal cancer, Lynch syndrome, receiving 600mg daily aspirin or placebo for a mean intervention period of 29 months. At the end of the intervention period, no difference in the incidence of neoplasms (adenoma and cancer) between aspirin and placebo group was observed (RR=1.0, 95% CI=0.7-1.4, p=0.7) (Burn et al., 2008). However, after a mean follow up of 55.7 months, per protocol analysis showed a reduced risk of primary cancer in the aspirin group compared to

placebo (Hazard ratio [HR]=0.41, 95% CI=0.19-0.86, p=0.02; Incidence Rate Ratio [IRR]=0.37, 95% CI=0.18-0.78, p=0.008) (Burn et al., 2011b). In a follow up observational study of a randomized trial, the Women's Health Study, Cook *et al.* showed reduction in colorectal cancer by 20% (HR=0.80, 95% CI=0.67-0.76, P=0.021) in the group taking 100mg alternate day aspirin compared to the placebo group with the effect commencing 10 years after randomization (Cook et al., 2013). The combined evidence from observational studies and randomized controlled trials provided compelling evidence that aspirin intake reduces colorectal cancer risk.

Colorectal adenomas are the precursors to most colorectal cancers. The chemopreventive effect of aspirin is likely to be observed in the adenomas as they form during the neoplastic transformation of normal to cancer cells. The largest randomized placebo controlled trial consisting of 206 patients who were genetically predisposed to another form of hereditary colorectal cancer, familial adenomatous polyposis (FAP), showed no significant reduction in the polyp count in the sigmoid colon and rectum of patients taking 600mg/day aspirin compared to placebo (RR=0.77, 95% CI=0.54-1.10) (Burn et al., 2011a). However, there was a reduction in the mean polyp size in patients randomized to aspirin (3.0mm versus 6.0mm; p=0.02) (Burn et al., 2011a). Furthermore, similar results were observed in a trial involving 34 Japanese patients with FAP, which showed a trend for reduction in the mean diameter of polyps in patients randomized to 100mg/day aspirin for 6-10 months compared to the placebo group (Response rate= 2.33, 95% CI=0.72-7.55) (Ishikawa et al., 2013). Despite the lack of clear evidence of the protective effect of aspirin on primary lesions in individual trials, a meta analysis of 4 randomized controlled trials (AFPPS (Baron et al., 2003), CALGB (Sandler et al., 2003), ukCAP (Logan et al., 2008) and APACC (Benamouzig et al., 2003)) that evaluated secondary prevention of sporadic colorectal adenoma with aspirin showed a risk reduction for developing adenomas (pooled RR=0.83, 95% CI=0.72-0.96, p=0.012) and advanced lesions (pooled RR=0.72, 95% CI=0.57-0.90, p=0.005) (Cole et al., 2009). Thus in both sporadic and familial risk patients, there is convincing evidence of adenoma risk reduction with aspirin use.

In addition to the evidence from controlled trials, meta analysis of 5 observational studies, of which two focused on colorectal cancer, showed an inverse association between cancers with distant metastasis and regular aspirin use (pooled OR=0.69, 95% CI=0.57-0.83, p<0.0001) which were consistent with the observations from randomized trials (pooled OR=0.48, 95% CI=0.30-0.75, p=0.002) (Algra and Rothwell, 2012). However, the inverse association was not observed for cancers with regional spread (pooled OR for observational studies=0.98, 95%CI=0.88-1.09, p=0.71) (Algra and Rothwell, 2012). In a meta analysis of 5

large randomized trials from the UK of daily aspirin versus controls, patients with colorectal cancer without metastasis at the time of initial diagnosis had an overall significantly reduced risk of later metastasis in aspirin users (HR=0.26, 95% CI=0.11-0.57, p=0.0008) and further risk reduction of later metastasis in patients still on trial treatment at the time of diagnosis (HR=0.13, 95% CI=0.03-0.56, p=0.007) (Rothwell et al., 2012b). Furthermore, allocation to aspirin use halved the risk of death due to cancer in patients with adenocarcinoma without metastasis at initial diagnosis compared to controls (HR=0.50, 95% CI=0.34-0.74, p=0.0006) (Rothwell et al., 2012b). Thus there is a plethora of convincing evidence for aspirin to be prescribed for prophylaxis and adjuvant therapy in patients at high risk for colorectal cancer.

3.1.3 Variation in aspirin's chemopreventive efficacy

Despite the evidence of aspirin's chemopreventive potential, its efficacy varies between individuals. Chan et al. 2007, carried out immunohistochemistry staining to determine the expression of COX-2 in 636 incident colorectal cancers for whom baseline epidemiological data on aspirin use was available from two cohort studies [the Nurses' Health Study (NHS) and the Health Professional Follow-up Study (HPFS)]. Using Cox regression analysis for competing risks to compare colorectal cancer risk in regular aspirin users based on the expression of COX-2 in tumors, the authors showed a significant decrease in the risk of colorectal cancers that over-expressed COX-2 (RR=0.64, 95% CI=0.52-0.78) but no influence on cancers with weak or absent COX-2 expression (RR=0.96, 95% CI=0.73-1.26) (Chan et al., 2007). Prostaglandins synthesized by COX-2 are metabolized by hydroxyprostaglandin dehydrogenase 15-(nicotinamide adenine dinucleotide) (15-PGDH, HPGD) enzyme, thus HPGD functions as a metabolic antagonist of COX-2 (Yan et al., 2004). Using 270 colorectal cancer cases from the NHS and HPFS cohorts, Fink et al. showed that compared to non-use, regular aspirin use was associated with a low risk of colorectal cancer that was surrounded by colonic mucosa with high HPGD mRNA expression (HR=0.49, 95% CI=0.34-0.71, p=0.0002) but not in cancers surrounded with mucosa having low HPGD expression (HR=0.90, 95% CI=0.63-1.27, p=0.53) (Fink et al., 2014).

The phosphatidylinositol 3-kinase (PI3K) signaling pathway is involved in carcinogenesis of colorectal cancer, and activating mutations in the *PIK3CA* gene that occur in two "hotspots" (exon 9 and 20) are present in approximately 15-20% of colorectal cancers (Samuels et al., 2004). PIK3CA is a downstream mediator of COX-2 and mutations in PIK3CA leads to increased prostaglandin E2 synthesis and inhibition of apoptosis in colon cancer cells. Using *PIK3CA* mutation status and epidemiological data on 964 patients with colorectal cancer, from the NHS and HPFS study, Liao et al. showed an improved colorectal cancer specific survival among post-diagnosis regular aspirin users with mutated *PIK3CA*

colorectal cancer (HR=0.18, 95% CI=0.06-0.61, p<0.001) but not in patients with wild type *PIK3CA* cancer (HR=0.96, 95% CI=0.69-1.32, p=0.76) (Liao et al., 2012). Lastly, a recent study carried out by Reimers et al., found an improved overall survival benefit associated with post cancer diagnosis regular aspirin use (RR=0.53, 95% CI=0.38-0.74, p<0.001) in tumors expressing HLA class I antigen, whereas, no benefit was observed in tumors not expressing HLA class I antigen (HR=1.03, 95% CI=0.66-1.61, p=0.91) (Reimers et al., 2014).

Even though the majority of the literature describes modulation of aspirin's chemopreventive effect through COX-2 inhibition, other potential mechanisms such as inhibition of NFkB transcription factor (Seufert et al., 2013, Din et al., 2004, Stark et al., 2001), interaction with Wnt signaling pathway (Nan et al., 2013), polyamine metabolism (Zell et al., 2009, Hubner et al., 2008) and specificity protein (Sp) transcription factors (Pathi et al., 2012) may help explain the variation in efficacy. Furthermore, presence of a single nucleotide polymorphism (SNPs) in the UDP glucoronosyltransferase 1A6 (UGT1A6) and cytochrome 450 2C9 (CYP2C9) enzymes that metabolize aspirin have been implicated to modulate its chemopreventive efficacy and modify colorectal cancer risk in observational studies (Angstadt et al., 2014, Hubner et al., 2006, Wang et al., 2014) but lacked statistical power to correct for multiple tests due to small sample size. Genome wide association studies (GWAS) have identified novel SNPs that are associated with colorectal cancer risk (Peters et al., 2012, Dunlop et al., 2012, Tomlinson et al., 2008) however, they lack depth to test for a large number of SNPs within candidate gene loci or pathways. Therefore, re-assessing previously implicated SNPs along with the SNPs from candidate genes in larger datasets is imperative to develop a SNP panel, which can help explain variation in aspirin's chemopreventive efficacy.

3.1.4 Aims

The current study aimed to re-assess SNPs that are previously implicated to modulate aspirin's chemopreventive efficacy in small observational studies and identify novel SNPs in candidate genes that are associated with colorectal cancer risk or modulate aspirin's chemopreventive efficacy. A systematic review of the literature was carried out to assemble a panel of potentially informative SNPs from candidate gene loci and aspirin's pharmacokinetic and pharmacodynamic pathways. These SNPs were tested for association with colorectal cancer risk and interaction with aspirin use and cancer risk within two large population based case-control study datasets, UK-Colorectal Cancer Study Panel (UK-CCSG) and NIH-Colon Cancer Family Registry (NIH-CCFR), with the latter serving as a validation dataset. Identification of novel SNPs that are associated with cancer risk and modify aspirin's efficacy may help gain insight into the neoplastic transformation of epithelial cells in colon and rectum and relevant aspect of aspirin's mode of action respectively. Perception of the biological

implication of these SNPs could help in assembling a panel of genetic markers, which would serve as a prognostic test for patients at risk of colorectal cancer and being treated with aspirin.

3.2 Results

3.2.1 Baseline characteristics of cases and controls

Overall, baseline epidemiological and genotype data was available for 3186 and 2926 individuals from UK-CCSG and NIH-CCFR datasets respectively. UK-CCSG dataset consisted of 1910 cases and 1276 controls whereas the NIH-CCFR dataset consisted of 1941 cases and 986 controls (Table 3.1). As no controls were available from the USC, MC and UHI sites in the NIH-CCFR dataset, cases for these sites were removed from downstream analyses. Furthermore, 16 study subjects in the NIH-CCFR dataset that were not of non-Hispanic white ethnicity were removed from analysis. In comparison to the subjects in NIH-CCFR dataset, UK-CCSG dataset consisted of slightly higher number of males than females, both cases and controls that were older, lower number of subjects with first or second degree relative with colorectal cancer, higher number of current smokers, lower number of subjects that used only aspirin (or NSAID) regularly and subjects with slightly lower BMI but higher physical activity output and daily alcohol intake ($P<0.001$ for all) (Table 3.1). No difference was observed in the distribution of primary cancer across site (colon and rectum) between the two datasets ($P=0.15$).

Baseline characteristics	Colorectal Cancer Study Group				Colon Cancer Family Registry							P-value[b]
	Dundee	Leeds	York	**Total**	CCO	USC	UQM	MC	FHCRC	UHI	**Total[a]**	
Sex, n (%)												
Male	285 (54.7)	1343 (55.0)	136 (60.7)	**1764 (55.4)**	523 (51.3)	89 (43.2)	267 (50.9)	136 (44.7)	414 (48.4)	-	**1204 (50.2)**	<0.001
Female	236 (45.3)	1098 (45.0)	88 (39.3)	**1422 (44.6)**	497 (48.7)	117 (56.8)	258 (49.1)	168 (55.3)	442 (51.6)	-	**1197 (49.9)**	
Age at diagnosis in cases (in years), n (%)												
<40	0 (0)	20 (1.2)	0 (0)	**20 (1.1)**	23 (4.4)	16 (7.8)	62 (18.3)	39 (12.8)	40 (7.3)	-	**125 (8.8)**	<0.001
40-69	83 (57.2)	896 (54.2)	62 (55.4)	**1041 (54.5)**	463 (87.5)	169 (82.0)	277 (81.7)	253 (83.2)	420 (76.8)	-	**1160 (82.0)**	
>69	62 (42.8)	737 (44.6)	50 (44.6)	**849 (44.5)**	43 (8.1)	21 (10.2)	0 (0)	12 (4.0)	87 (15.9)	-	**130 (9.2)**	
Age at interview in controls (in years), n (%)												
<40	0 (0)	5 (0.6)	0 (0)	**5 (0.4)**	6 (1.2)	-	40 (21.5)	-	0 (0)	-	**46 (4.7)**	<0.001
40-69	218 (58.0)	415 (52.7)	64 (57.1)	**697 (54.6)**	357 (72.7)	-	146 (78.5)	-	209 (67.6)	-	**712 (72.2)**	
>69	158 (42.0)	368 (46.7)	48 (42.9)	**574 (45.0)**	128 (26.1)	-	0 (0)	-	100 (32.4)	-	**228 (23.1)**	
Family History of Cancer, n (%)												
Unaffected	370 (77.1)	1575 (74.3)	140 (76.9)	**2085 (74.9)**	645 (63.2)	100 (48.5)	309 (58.9)	145 (47.7)	641 (74.9)	-	**1595 (66.4)**	<0.001
Affected second degree relatives	29 (6.0)	173 (8.2)	10 (5.5)	**212 (7.6)**	163 (16.0)	9 (4.4)	117 (22.3)	67 (22.0)	62 (7.2)	-	**342 (14.2)**	
Affected first or first and second degree relatives	81 (16.9)	373 (17.6)	32 (17.6)	**486 (17.5)**	212 (20.8)	97 (47.1)	99 (18.9)	92 (30.3)	153 (17.9)	-	**464 (19.3)**	
Smoking, n (%)												
Never	211 (41.1)	977 (40.2)	80 (37.7)	**1268 (40.2)**	387 (38.0)	93 (45.2)	234 (44.7)	139 (46.3)	344 (40.2)	-	**965 (40.2)**	<0.001

Stopped	223 (43.4)	1043 (42.9)	101 (47.6)	**1367 (43.3)**	527 (51.8)	88 (42.7)	209 (39.9)	126 (42.0)	416 (48.6)	-	**1152 (48.0)**	
Current	80 (15.6)	410 (16.9)	31 (14.6)	**521 (16.5)**	104 (10.2)	25 (12.1)	81 (15.5)	35 (11.7)	96 (11.2)	-	**281 (11.7)**	
Any NSAID use, n (%)												
No	372 (72.5)	1733 (71.2)	167 (78.4)	**2272 (71.9)**	611 (60.6)	102 (49.5)	394 (75.3)	156 (53.1)	387 (45.4)	-	**1392 (58.4)**	<0.001
Yes	141 (27.5)	702 (28.8)	46 (21.6)	**889 (28.1)**	398 (39.4)	104 (50.5)	129 (24.7)	38 (46.9)	465 (54.6)	-	**992 (41.6)**	
Aspirin only use, n (%)												
No	372 (79.7)	1733 (79.0)	167 (86.5)	**2272 (79.6)**	611 (71.4)	102 (61.1)	394 (88.0)	156 (69.6)	387 (58.7)	-	**1392 (70.9)**	<0.001
Yes	95 (20.3)	460 (21.0)	26 (13.5)	**581 (20.4)**	245 (28.6)	65 (38.9)	54 (12.1)	68 (30.4)	272 (41.3)	-	**571 (29.1)**	
Primary Cancer Site, n (%)												
Colon	90 (63.8)	1154 (65.9)	62 (54.9)	**1306 (65.2)**	279 (65.8)	131 (69.3)	188 (58.4)	189 (63.6)	337 (62.6)	-	**804 (62.6)**	0.15
Rectum	51 (36.2)	596 (34.1)	51 (45.1)	**698 (34.8)**	145 (34.2)	58 (30.7)	136 (41.6)	108 (36.4)	201 (37.4)	-	**480 (37.4)**	
BMI at 20 years (kg/m^2), mean (SD)	22.3 (2.5)	22.3 (3.2)	22.5 (3.4)	**22.3 (3.1)**	22.4 (3.6)	22.1 (3.1)	22.8 (3.9)	23.1 (4.0)	22.6 (4.4)	-	**22.6 (4.0)**	0.002
Physical Activity (Hours/ week), mean (SD)	25.4 (17.1)	23.1 (13.4)	26.1 (19.7)	**23.7 (14.6)**	4.1 (7.1)	8.7 (9.4)	6.5 (7.2)	10.9 (15.7)	6.5 (8.8)	-	**5.4 (7.7)**	<0.0001
Alcohol (Units/ day), mean (SD)	1.7 (3.0)	3.2 (5.0)	2.2 (3.1)	**2.9 (4.6)**	0 (0)	1.6 (1.7)	1.8 (2.9)	2.0 (2.5)	2.2 (3.6)	-	**2.0 (3.3)**	<0.0001

Table 3.1 Baseline epidemiological characteristics data distribution within the UK-CCSG and NIH-CCFR datasets.

a, Total number of subjects calculated from only CCO, UQM and FHCRC study sites.

b, P-value calculated using Fisher's exact test for categorical variables and Student t-test for continuous variables.

3.2.1.1 Association between environmental risk factors and colorectal cancer

To test for association between environmental risk factors and colorectal cancer risk, continuous variables such as BMI, daily calorie intake, weekly exercise and alcohol intake per day were converted to dichotomous variables by placing a cut off point where 50% of the controls under the cut off value were coded at 0 and 50% of the controls above the cut off value were coded as 1. In the UK-CCSG dataset, higher BMI at the age of 20 (>21.8 kg/m^2), ever smoked, higher daily calorie intake (>2172.5 kcal), higher alcohol intake (>5.6 units/ day) and presence of first or second degree relative diagnosed with colorectal cancer were associated with an increased risk of colorectal cancer (all P<0.001; Table 3.2). Higher time spent in carrying out exercise every week (>22 hours/ week) showed a trend for a decreased risk of colorectal cancer but didn't reach statistical significance (P=0.076). The associations observed in the UK-CCSG dataset are consistent with the existing literature (Fedirko et al., 2011, Hannan et al., 2009, Larsson and Wolk, 2007, Slattery et al., 2003b, Giacosa et al., 1999, Slattery et al., 1997). In concordance with the existing literature, use of only aspirin and any NSAID (including aspirin) was inversely associated with colorectal cancer risk (Only aspirin OR=0.74, 95% CI=0.61-0.89, P=0.003; Any NSAID OR=0.65, 95% CI=0.55-0.76, P=1.79 x 10^{-7}). The mean daily dose of aspirin among regular aspirin only users was 107mg/day (Range= 25 mg/day to 4000 mg/day).

However, in the NIH-CCFR dataset, only BMI at the age of 20 (>21.5 kg/m^2), calorie intake in males (>2218 Kcal), presence of first or second degree relative diagnosed with colorectal cancer and higher weekly exercise (>3 hours/week) were associated with an increased risk of colorectal cancer (Table 3.2). The positive association between increase in exercise and colorectal cancer risk could be due to the socio-economic status of the study subjects (Doubeni et al., 2012) but the lack of data regarding socio-economic status precluded adjustment for it in the model. A significant association between family risk and cancer could be due to the case recruitment strategies employed by the CCO and UQM sites, which selected cases based on the family history and age at cancer diagnosis respectively (Site specific association: for CCO OR=13.2, 95% CI=9.46-18.53, P<0.001; for UQM OR=4.51, 95% CI=2.96-6.85, P<0.001) (Table 2.1). However, for FHCRC site where no family history based case recruitment strategy was employed, a high association between family history and CRC was observed (OR=7.54, 95% CI=4.68-12.14, P<0.001), which is because 35.5% of the cases had affected FDR and (or) SDR with CRC compared to controls that only had 6.8% of affected family members. The relatively high proportion of cases with affected family members at the FHCRC site is because all CRC affected FDR were recruited on the study which were substantially higher in cases than controls (Table 2.1). Smoking status and

alcohol consumption showed association trend with colorectal cancer risk but didn't reach statistical significance (P>0.05). Use of only aspirin and any NSAID (including aspirin) was associated with a reduced risk of colorectal cancer (Table 3.2). Compared to the un-adjusted model, the association strength between aspirin (or NSAID) use and cancer risk was poor in the adjusted model, where the association was adjusted for age, sex and study site (For any NSAID use OR=0.82, 95% CI=0.69-0.98, P=0.03; for aspirin only use OR=0.78, 95% CI=0.63-0.96, P=0.02). This could be because the mean age of regular aspirin only or NSAID users versus non-users in both cases (Mean age (SD) for aspirin only user v/s non-user=59.0 (10.5) v/s 51.7 (10.8), t-test P<0.0001; Mean age (SD) for any NSAID user v/s no-user=55.9 (10.9) v/s 51.7 (10.8), t-test P<0.0001) and controls (Mean age (SD) for aspirin only user v/s non-user=64.0 (8.2) v/s 57.4 (11.7), t-test P<0.0001; Mean age (SD) for any NSAID user v/s no-user=63.1 (8.9) v/s 57.4 (11.7), t-test P<0.0001) was significantly higher and thus adjusting for age in the regression model could be leading to adjusting for the case-control status. Furthermore, compared to the UK-CCSG dataset, the NIH-CCFR dataset didn't record the daily aspirin dose information and hence no further analysis regarding the relationship between the dose and CRC risk was carried out.

	UK-Colorectal Cancer Study Group (N=3186)				NIH-Colon Cancer Family Registry (N=2401)			
	Controls, n (%)	Cases, n (%)	Odds Ratio (95% CI)	P value	Controls, n (%)	Cases, n (%)	Odds Ratio (95% CI)	P value
BMI at 20 years [kg/m^2][a]								
Low	629 (49.8)	822 (44.8)	1.22 (1.06-1.41)	0.006	481 (49.5)	575 (41.3)	1.39 (1.18-1.64)	8.2 x 10^{-5}
High	633 (50.2)	1013 (55.2)			490 (50.5)	816 (58.7)		
Family history of cancer								
No	210 (47.5)	196 (29.6)	2.15 (1.68-2.76)	1.98 x 10^{-9}	876 (88.8)	719 (50.8)	7.71 (6.16-9.64)	3.3 x 10^{-150}
First or (and) second degree relative	232 (52.5)	466 (70.4)			110 (11.2)	696 (49.2)		
Cigarette smoking								
No	558 (43.8)	710 (37.6)	1.30 (1.12-1.50)	0.0005	408 (41.4)	557 (39.4)	1.08 (0.92-1.28)	0.34
Yes	716 (56.2)	1180 (62.4)			578 (58.6)	856 (60.6)		
Alcohol [units/ day][b]								
Low	644 (50.6)	741 (39.6)	1.57 (1.36-1.81)	1.01 x 10^{-9}	146 (48.2)	219 (42.3)	1.27 (0.95-1.69)	0.10
High	628 (49.4)	1131 (60.4)			157 (51.8)	299 (57.7)		
Physical activity [hours/ week][c]								

Low	625 (49.1)	976 (52.3)	0.88 (0.76-1.01)	0.079	395 (48.6)	493 (43.3)	1.24 (1.03-1.48)	0.02
High	647 (50.9)	889 (47.7)			418 (51.4)	645 (56.7)		
Regular aspirin only use[d]								
No	851 (76.9)	1421 (81.4)	0.76 (0.63-0.91)	0.004	518 (64.7)	874 (75.2)	0.60 (0.50-0.73)	4.80×10^{-7}
Yes	256 (23.1)	325 (18.6)			283 (35.3)	531 (24.8)		
Regular any NSAID use[d]								
No	851 (66.9)	1421 (75.2)	0.67 (0.57-0.78)	3.58×10^{-7}	518 (52.9)	874 (62.2)	0.68 (0.58-0.81)	6.04×10^{-6}
Yes	421 (33.1)	468 (24.8)			461 (47.1)	531 (37.8)		

Table 3.2 Association between baseline characteristics and colorectal cancer risk.

Continuous variables such as BMI, alcohol and physical activity were converted to dichotomous variables by placing a cut off point where 50% of the controls under the cut off value were coded at 0 and 50% of the controls above the cut off value were coded as 1.

a, Body Mass Index (BMI) cut off point in UK-CCSG and NIH-CCFR is 21.8 kg/m^2 and 21.5 kg/m^2 respectively

b, Alcohol intake cut off point in UK-CCSG and NIH-CCFR is 5.6 units/day and 0.89 units/day respectively

c, Physical activity cut off point in UK-CCSG and NIH-CCFR is 22 hours/week and 3 hours/week respectively

d, Regular aspirin or NSAID use is defined as regular intake for a period of 3 months or longer in the UK-CCSG dataset whereas it is defined as regular use of at least two pills per week for at least one month in the NIH-CCFR dataset

3.2.2 SNP frequency and linkage disequilibrium

Overall, 43 SNPs from 16 genes that are involved in aspirin's pharmacokinetic and pharmacodynamic pathways were selected for analysis in the current study (Table 3.3). Seventeen SNPs, which were absent on the Illumina Human Exome Array v1.1 platform in the UK-CCSG dataset, were genotyped using Taqman allelic discrimination assay (Table 2.2). However, 12 SNPs from the *CES2* gene and 4 from the *PAFAH1B2* gene were removed from all downstream analyses as the MAF was <4% (Supplementary Table 1).

A total of 27 SNPs were selected for analysis in the UK-CCSG dataset, out of which, SNP rs4648310 in *PTGS2*, rs4936367 in *PAFAH1B2* and rs11694911 in *ODC1* gene were observed to be inconsistent with the Hardy Weinberg equilibrium in controls (all $P<0.05$; Supplementary Table 1). On comparing the SNP genotype frequency in controls with the HapMap Phase I GBR population from the 1000 Genomes database using Fisher's exact test, 5 SNPs (rs1799853, rs4936367, rs7112513, rs28362380 and rs2430420) showed different genotype frequency in controls compared to the GBR population (all $P<0.05$; Supplementary Table 1). Overall, 6 linkage disequilibrium (LD) groups in controls were found on chromosomes 1, 2, 8, 10, 11 and 15 (Supplementary Figure 2). SNPs rs2070959 and rs1105879 in *UGT1A6* gene and rs7112513 and rs4936367 in *PAFAH1B2* gene had a high LD score (R^2) of 0.89 and 0.99 respectively.

Number	Gene Name	SNP ID	Study Reference
1	*ALOX15*	rs2619112	(Kleinstein et al., 2013)
2	*CDKN1A*	rs1321311	(Dunlop et al., 2012)
3	*CES2*	rs140461033	(Kubo et al., 2005)
		rs44410046	
		rs201103548	
		rs28382815	
		rs148026549	
		rs145407778	
		rs10852434	
		rs14792040	
		rs141625476	
		rs147070911	
		rs150408050	
		rs147694237	
4	*CYP2C9*	rs1057910	(Barry et al., 2013)
		rs1799853	

5	*IkBkB*	rs10958713	(Seufert et al., 2013)
		rs11986055	
		rs5029748	
		rs6474387	
6	*IL16*	rs12910333	Unpublished data from NIH-CCFR group
		rs16973225	Unpublished data from NIH-CCFR group
7	*Intergenic*	rs6983267	(Nan et al., 2013)
		rs961253	Unpublished data from UK-CCSG group
8	*MDR1*	rs1045642	(Sharma et al., 2012)
9	*NCF4*	rs5995355	(Ryan et al., 2014)
10	*Near MGST1*	rs2965667	Unpublished data from NIH-CCFR group
11	*NFkB*	rs230490	(Seufert et al., 2013)
12	*ODC1*	rs11694911	(Barry et al., 2011)
		rs28362380	(Barry et al., 2011)
		rs2302615	(Barry et al., 2011, Hubner et al., 2008)
		rs2430420	(Barry et al., 2011)
13	*PAFAH1B2*	rs4936367	(Zhou et al., 2011)
		rs7112513	
		rs142710583	
		rs185651296	
		rs186808413	
		rs78428934	
14	*PTGS1*	rs3842787	(Makar et al., 2013)
15	*PTGS2*	rs20417	(Makar et al., 2013, Ulrich et al., 2005)
		rs5275	(Makar et al., 2013)
		rs4648310	(Barry et al., 2009)
		rs2745557	Unpublished data from NIH-CCFR group
		rs689469	(Kraus et al., 2013)
		rs5277	(Barry et al., 2009)
16	*UGT1A6*	rs2070959	(Scherer et al., 2014, Angstadt et al., 2014)
		rs1105879	

Table 3.3 List of SNPs from genes involved in aspirin's pharmacokinetic and pharmacodynamic pathways, which were included in the study.

In contrast to the UK-CCSG dataset, SNPs in the NIH-CCFR dataset were genotyped using three separate Illumina platforms (See section 2.1.3 SNP genotyping). Whilst the Illumina platform used in the UK-CCSG dataset covered exonic content only with >240,000 markers, the platforms used in the NIH-CCFR dataset covered the whole genome, including the exonic regions, using 1 million markers. Overall, only 31 SNPs were selected for analysis in the NIH-CCFR dataset as, some of the SNPs selected for the UK-CCSG dataset were not genotyped in the NIH-CCFR dataset (Supplementary Table 1). rs16973225 and rs5277 were removed from all downstream analyses as they were only genotyped in cases. rs689469 was also removed from all downstream analyses as the SNPs' MAF was <4%. Apart from rs11694911 (p=0.02), all SNPs were consistent with the Hardy Weinberg equilibrium in controls. Furthermore, with the exception of rs6683455, rs1799853 and rs961253, allele frequencies of all other SNPs were not found to be significantly different from that of the HapMap Phase I CEU population from the 1000 Genomes database (Supplementary Table 1). A total of 6 LD groups on controls were found on chromosomes 1, 2, 4, 8, 10 and 11 (Supplementary Figure 3). Similar to the UK-CCSG dataset, SNPs rs1105879 and rs2070959 in *UGT1A6* gene and rs7112513 and rs4936367 in *PAFAH1B2* showed a high LD score R^2>0.90 in the NIH-CCFR dataset.

3.2.3 Association between SNP genotype and colorectal cancer risk

Out of 28 SNPs analyzed in the UK-CCSG dataset, 1 showed a significant association with colorectal cancer risk (Supplementary Table 2). The variant T allele of a non-synonymous SNP rs1799853, which converts arginine to cysteine at 144th amino acid position in CYP2C9 enzyme, was associated with decrease in colorectal cancer risk (OR=0.82, 95% CI=0.69-0.98, P=0.026) in individuals with the variant allele (Table 3.4). Furthermore, presence of variant A, T and G alleles of SNPs in *CDKN1A* (rs1321311), *ODC1* (rs2302615) and *UGT1A6* (rs2070959) genes respectively showed trends for an association with colorectal cancer risk but didn't reach the type I error significance threshold of 0.05 (Table 3.4).

In contrast, none of the four SNPs, which were associated with colorectal cancer risk in the UK-CCSG dataset, showed significant association with cancer risk in the NIH-CCFR dataset (Table 3.4). However, presence of variant T allele of SNP rs6983267, which indirectly affects expression of target oncogenes including *MYC* (Nan et al., 2013) and was initially identified through GWA studies, showed a trend for decrease in colorectal cancer risk in the NIH-CCFR dataset (OR=0.83, 95% CI=0.70-1.00, P=0.057) but not in the UK-CCSG dataset using a dominant model (Table 3.4).

On carrying out association tests between SNPs and site-specific colorectal cancer risk (colon and rectum) in the UK-CCSG dataset, the variant T alleles of rs1799853 and rs2302615 were both associated with a site specific reduction of colon cancer risk only (rs1799853 OR=0.73, 95% CI=0.60-0.90, P=0.002; rs2302615 OR=0.78, 95% CI=0.65-0.93, P=0.012) (Table 3.5; Supplementary Table 3). However, the variant C allele of SNP rs1105879 in *UGT1A6* gene was associated with an increased risk of colon cancer only (OR=1.21, 95% CI=1.01-1.44, P=0.036) (Table 3.5). No site-specific associations were observed in the NIH-CCFR dataset (Supplementary Table 3).

			UK-Colorectal Cancer Study Group					NIH-Colon Cancer Family Registry				
Gene name	**SNP ID**	**Copies of rare allele**	**Controls, n (%)**	**Cases, n (%)**	**Odds Ratio**	**95% CI**	**P-value***	**Controls, n (%)**	**Cases, n (%)**	**Odds Ratio**	**95% CI**	**P-value***
CDKN1A	rs1321311	0	579 (59.3)	946 (56)				566 (57.8)	789 (55.9)			
		1 or 2	398 (40.7)	743 (44.0)	1.14	0.97, 1.34	0.07	414 (42.2)	622 (44.1)	1.08	0.91, 1.27	0.72
CYP2C9	rs1799853	0	709 (74.3)	1274 (78)				141 (78.8)	126 (79.3)			
		1 or 2	245 (25.7)	359 (22.0)	0.82	0.68, 0.98	**0.03**	38 (21.2)	33 (20.8)	0.97	0.58, 1.64	0.83
Intergenic	rs6983267	0	282 (28.9)	523 (31)				267 (27.1)	436 (30.9)			
		1 or 2	693 (71.1)	1167 (69.1)	0.91	0.76, 1.08	0.31	717 (72.9)	976 (69.1)	0.83	0.70, 1.00	0.06
ODC1	rs2302615	0	501 (51.6)	907 (56.1)				-	-	-	-	-
		1 or 2	470 (48.4)	710 (43.9)	0.83	0.71, 0.98	0.06	-	-	-	-	-
UGT1A6	rs2070959	0	497 (51)	812 (47.9)				433 (44.0)	636 (45.0)			
		1 or 2	477 (49.0)	882 (52.1)	1.13	0.97, 1.33	0.07	551 (56.0)	779 (55.1)	0.96	0.82, 1.13	0.90

Table 3.4 Association between SNP variant allele and colorectal cancer risk.

*P-value of association is adjusted for age, sex and study site within each dataset.

CI, Confidence Interval

n, Number of subjects

			Colon Cancer					Rectal Cancer				
Gene name	**SNP ID**	**Copies of rare allele**	**Controls, n (%)**	**Cases, n (%)**	**Odds Ratio**	**95% CI**	**P-value***	**Controls, n (%)**	**Cases, n (%)**	**Odds Ratio**	**95% CI**	**P-value***
CYP2C9	rs1799853	0	709 (74.3)	855 (79.8)				709 (74.3)	419 (74.7)			
		1 or 2	245 (25.7)	217 (20.2)	0.73	0.60, 0.90	**0.002**	245 (25.7)	142 (25.3)	0.98	0.77, 1.25	0.86
ODC1	rs2302615	0	501 (51.6)	613 (57.8)				501 (51.6)	294 (52.9)			
		1 or 2	470 (48.4)	448 (42.2)	0.78	0.65, 0.93	**0.012**	470 (48.4)	262 (47.1)	0.95	0.77, 1.17	0.86
UGT1A6	rs1105879	0	458 (47.1)	481 (43.1)				458 (47.1)	281 (48.1)			
		1 or 2	515 (52.9)	634 (56.9)	1.17	0.99, 1.39	**0.036**	515 (52.9)	303 (51.9)	0.96	0.78, 1.18	0.99

Table 3.5 Association between SNP variant allele and site-specific colorectal cancer risk in the UK-Colorectal Cancer Study Group dataset.

*P-value of association is adjusted for age, sex and study site.

CI, Confidence Interval

n, Number of subjects

3.2.4 Interaction between aspirin only use, SNP genotype and colorectal cancer risk

Aspirin use was defined as regular intake of aspirin for 3 months or longer in the UK-CCSG dataset whereas, it was defined as regular intake of aspirin twice a week for a month or longer in the NIH-CCFR dataset. Furthermore, aspirin only use was defined for subjects who regularly used aspirin but not other NSAIDs. Gene- environment (G x E) interaction was tested for SNP genotype and aspirin only use in relation to colorectal cancer risk.

Out of all the SNP variants investigated for interaction with aspirin only use and colorectal cancer risk, 2 SNP variants reached significance in the UK-CCSG dataset and the SNPs in high LD with them showed a similar trend for interaction but didn't reach significance threshold (Supplementary Table 4). The variant G allele of *PAFAH1B2* SNP rs4936367 showed a statistically significant interaction with aspirin only use ($P_{interaction}$=0.04), increasing the risk of colorectal cancer by 65% in aspirin users (OR=1.65, 95% CI=1.02-2.66) but not in non-users (OR=1.02, 95% CI=0.80-1.30) (Table 3.6). A SNP rs7112513, which is in high LD with rs4936367 (R^2=0.99) showed a similar association trend but didn't reach statistical significance ($P_{interaction}$=0.08; Table 3.6). Similarly, the variant G allele of *UGT1A6* gene SNP rs2070959 showed a statistically significant interaction with aspirin only use ($P_{interaction}$=0.05), increasing the risk of colorectal cancer by 48% in aspirin users (OR=1.48, 95% CI=1.03-2.11) but not in non-users (OR=1.02, 95% CI=0.84-1.24) (Table 3.6). A SNP rs1105879, which is in high LD with rs2070959 (R^2=0.89) showed a similar association trend but didn't reach statistical significance threshold for interaction ($P_{interaction}$=0.10; Table 3.6). In contrast, none of the 4 SNPs reached significance threshold for interaction in the NIH-CCFR dataset and no other SNP showed interaction trend with aspirin only use and colorectal cancer risk (Supplementary Table 4).

Since 3 SNPs showed site-specific association with colorectal cancer risk, they were tested for site-specific interaction between aspirin only use and both colon and rectal cancer risk. However, minor alleles of all 3 SNPs showed no significant evidence of interaction in either datasets (Table 3.7 and Table 3.8).

			UK-Colorectal Cancer Study Group					NIH-Colon Cancer Family Registry				
Gene name	**SNP ID**	**Copies of rare allele**	**Non-users**		**Aspirin users**		**P-value for interaction***	**Non-users**		**Aspirin users**		**P-value for interaction***
			OR (95% CI)	**P-value+**	**OR (95% CI)**	**P-value+**		**OR (95% CI)**	**P-value+**	**OR (95% CI)**	**P-value+**	
PAFAH1B2	rs4936367	0	n=1475 1		n=393 1			n=1114 1		n=456 1		
		1 or 2	n=372 1.02 (0.80, 1.30)	0.88	n=93 1.65 (1.02, 2.66)	**0.04**	**0.04**	n=278 0.92 (0.70, 1.20)	0.53	n=115 1.00 (0.66, 1.51)	1.00	0.39
	rs7112513	0	n=1515 1		n=404 1			n=1109 1		n=453 1		
		1 or 2	n=391 1.03 (0.81, 1.31)	0.80	n=95 1.55 (0.96, 2.48)	0.07	0.08	n=279 0.89 (0.68, 1.16)	0.38	n=115 1.00 (0.67, 1.51)	0.98	0.30
UGT1A6	rs1105879	0	n=863 1		n=237 1			n=598 1		n=234 1		
		1 or 2	n=1039 1.00 (0.83, 1.21)	1.00	n=263 1.38 (0.97, 1.98)	0.08	0.10	n=791 0.93 (0.75, 1.16)	0.53	n=337 1.19 (0.85, 1.66)	0.31	0.16
	rs2070959	0	n=927 1		n=253 1			n=638 1		n=246 1		
		1 or 2	n=973 1.02 (0.84, 1.24)	0.83	n=247 1.48 (1.03, 2.11)	**0.03**	**0.05**	n=752 0.87 (0.70, 1.09)	0.23	n=325 1.16 (0.83, 1.61)	0.39	0.12

Table 3.6 Association between SNP variant allele and colorectal cancer risk stratified by only aspirin use.

+P-value for association between SNP variant allele and colorectal cancer risk.

*P-value for interaction between SNP variant allele, aspirin use and colorectal cancer risk calculated using Likelihood ratio test. P-value is adjusted for age, sex and study site.

OR, Odds Ratio

CI, Confidence Interval

			UK-Colorectal Cancer Study Group					NIH-Colon Cancer Family Registry				
Gene name	**SNP ID**	**Copies of rare allele**	**Non-users**		**Aspirin users**		**P-value for interaction***	**Non-users**		**Aspirin users**		**P-value for interaction***
			OR (95% CI)	**P-value+**	**OR (95% CI)**	**P-value+**		**OR (95% CI)**	**P-value+**	**OR (95% CI)**	**P-value+**	
CYP2C9	rs1799853	0	n=1105 1		n=304 1			n=90 1		n=74 1		
		1 or 2	n=318 0.77 (0.60, 0.99)	**0.04**	n=93 0.58 (0.36, 0.93)	**0.02**	0.26	n=22 1.20 (0.46, 3.10)	0.71	n=17 0.98 (0.31, 3.13)	0.98	0.77
ODC1	rs2302615	0	n=753 1		n=219 1			-	-	-	-	
		1 or 2	n=616 0.78 (0.63, 0.97)	**0.02**	n=170 0.74 (0.50, 1.11)	0.15	0.66	-	-	-	-	-
UGT1A6	rs1105879	0	n=653 1		n=191 1			n=428 1		n=180 1		
		1 or 2	n=813 1.06 (0.86, 1.31)	0.57	n=214 1.50 (1.02, 2.23)	**0.04**	0.12	n=578 0.95 (0.74, 1.22)	0.71	n=270 1.42 (0.96, 2.12)	0.08	0.07

Table 3.7 Association between SNP variant allele and colon cancer risk stratified by aspirin only use.

+P-value for association between SNP variant allele and colon cancer risk.

*P-value for interaction between SNP variant allele, aspirin use and colon cancer risk calculated using Likelihood ratio test. P-value is adjusted for age, sex and study site.

OR, Odds Ratio

CI, Confidence Interval

n, Number of subjects

			UK-Colorectal Cancer Study Group					NIH-Colon Cancer Family Registry				
Gene name	**SNP ID**	**Copies of rare allele**	**Non-users**		**Aspirin users**		**P-value for interaction***	**Non-users**		**Aspirin users**		**P-value for interaction***
			OR (95% CI)	**P-value+**	**OR (95% CI)**	**P-value+**		**OR (95% CI)**	**P-value+**	**OR (95% CI)**	**P-value+**	
CYP2C9	rs1799853	0	n=782 1		n=212 1			n=80 1		n=67 1		
		1 or 2	n=261 1.01 (0.76, 1.35)	0.92	n=83 0.99 (0.57, 1.71)	0.96	0.96	n=22 1.72 (0.65, 4.56)	0.28	n=14 0.58 (0.12, 2.87)	0.50	0.09
ODC1	rs2302615	0	n=542 1		n=151 1			-	-	-	-	
		1 or 2	n=488 0.91 (0.71, 1.17)	0.47	n=141 1.18 (0.71, 1.95)	0.53	0.42	-	-	-	-	-
UGT1A6	rs1105879	0	n=500 1		n=152 1			n=339 1		n=168 1		
		1 or 2	n=575 0.89 (0.70, 1.14)	0.37	n=146 1.16 (0.71, 1.90)	0.54	0.31	n=470 1.01 (0.75, 1.35)	0.96	n=219 0.96 (0.61, 1.50)	0.84	0.90

Table 3.8 Association between SNP variant allele and rectal cancer risk stratified by aspirin only use.

+P-value for association between SNP variant allele and rectal cancer risk.

*P-value for interaction between SNP variant allele, aspirin use and rectal cancer risk calculated using Likelihood ratio test. P-value is adjusted for age, sex and study site.

OR, Odds Ratio

CI, Confidence Interval

n, Number of subjects

3.3 Discussion

Colorectal cancer is a heterogeneous disease and is the second leading cause of cancer mortality with the lifetime risk estimated between 5-6% (Peters et al., 2012). There is an extensive evidence of the chemopreventive effect of regular use of aspirin in relation to colorectal cancer; however, data from recent studies suggests inter-individual variation in the chemopreventive effect may be attributed to the presence of somatic mutations (Nishihara et al., 2013, Liao et al., 2012) and germline variation (Wang et al., 2014, Reimers et al., 2014, Fink et al., 2014, Angstadt et al., 2014, Seufert et al., 2013, Nan et al., 2013, Pathi et al., 2012, Zell et al., 2009, Hubner et al., 2008, Chan et al., 2007, Hubner et al., 2006, Din et al., 2004, Stark et al., 2001). Aspirin's mode of action on colonic epithelial cells for reducing colorectal cancer risk has yet to be elucidated. Since genetic variation in aspirin's pharmacokinetic and pharmacodynamic pathways have been shown to modulate its efficacy in prior studies with a small sample size, re-assessment of previously associated variants along with exploration and identification of novel variants using biostatistics and epidemiological studies in these pathways could help explain aspirin's mode of action and gain insight into the neoplastic transformation of colonic epithelial cells.

The current study employed two large population-based case-control datasets (UK-CCSG and NIH-CCFR) consisting of a combined total of 3851 colorectal cancer cases and 2262 controls of self reported non-Hispanic white ethnicity, where one dataset was used as a validation dataset for the other. Previously documented environmental risk factors such as BMI, smoking and calorie intake etc. were studied for their association with colorectal cancer risk amongst the two datasets. In total, 43 SNPs across 16 genes involved in aspirin's pharmacokinetic and pharmacodynamic pathways were investigated for association or modification of the protective effect of aspirin use on colorectal cancer risk. These included 18 novel SNPs, which had not been investigated previously.

3.3.1 Known epidemiological risk factors

Despite several key differences between the two datasets such as, significantly younger age at diagnosis of cases in the NIH-CCFR dataset and case ascertainment strategies, most of the key epidemiological risk factors showed similar association trends with colorectal cancer risk in both datasets (Table 3.2). All continuous variables such as BMI, daily calorie intake and physical activity etc. were converted to dichotomous variables with the cut-off value set at the point where 50% of the controls under the cut off value were used as reference. The aim for converting continuous variables to dichotomous variables was to assess the association between the variable and CRC risk at a set cut-off value. This method of assessment is in

contrast with the test of association between the variable and CRC risk that provides an estimate for the increase in risk per unit value of the variable, when the variable is continuous.

An increase in BMI >21 kg/m^2 at the age of 20 years was significantly associated with an increased risk of colorectal cancer in both datasets. A recent study conducted by de Mutsert et al. in Health Professional Follow-Up Study showed that compared to the BMI of 18.5-22.9 at the age of 21 years, risk of obesity related cancers such as colorectal, renal, pancreatic and oesophageal cancers increased with increase in BMI (For 23.0-24.9 Multivariate HR= 1.15; for 25.0-27.4 HR=1.24; for 27.5-29.9 HR=1.36 and; for ≥30 HR=1.90) (de Mutsert et al., 2014). Pathologically, humans with high BMI have an increased adipose tissue mass. This increase in adipose tissue triggers an increase in pro-inflammatory cytokines such as tumor necrosis factor- alpha (TNFa) and interleukin-6 (IL6) along with an increase in immune cell infiltration (Khandekar et al., 2011). TNFa activates several downstream signaling pathways, including nuclear factor-κB (NF-κB), which promotes carcinogenesis, angiogenesis and metastasis (Khandekar et al., 2011).

In both datasets, cigarette smoking was associated with an increased risk of colorectal cancer. This observation is in concordance with the results of Hannan et al., 2009 who showed an increased colorectal cancer incidence amongst current smokers (HR=1.27, 95% CI=1.06-1.52) and former smokers (HR=1.23, 95% CI=1.11-1.36) compared to lifelong non-smokers (Hannan et al., 2009). Furthermore, a nested case control study within the PLCO screening trial showed association between smoking and cotinine (r=0.88, p=3.45 x 10^{-94}), O-cresol sulfate (r=0.72, p=1.22 x 10^{-42}) and hydroxycotinine (r=0.67, p=5.41 x 10^{-32}) (Cross et al., 2014). Compared to subjects with undetectable levels, subjects with detectable levels of hydroxycotinine (OR=2.68, 95% CI=1.33-5.40) and O-cresol sulfate (OR=1.81, 95% CI=0.98-3.33) showed an elevated risk of colorectal cancer (Cross et al., 2014). Hydroxycotinine is a major metabolite of nicotine but its effect on colorectal carcinogenesis has yet to be elucidated.

Alcohol intake was associated with an increased risk of colorectal cancer in both datasets. This observation is consistent with the meta-analysis results of Fedriko et al., 2011 who showed an increased risk of colorectal cancer for moderate (RR=1.21, 95% CI=1.13-1.28) and heavy (≥4 drinks/day) (RR=1.52, 95% CI=1.27-1.81) drinkers compared to non-or occasional drinkers (Fedirko et al., 2011). A possible biological mechanism behind the increased risk could be alcohol induced hypermethylation of Alcohol dehydrogenase iron containing 1 (*ADHFE1*) gene, which leads to reduced gene expression and cell proliferation (Moon et al., 2014). Furthermore, hypermethylated *ADHFE1* gene has been reported to be associated with colorectal cancer differentiation (Moon et al., 2014).

Physical activity was associated with a decreased risk of colorectal cancer in the UK-CCSG dataset but conversely; it was associated with an increased risk in the NIH-CCFR dataset. Numerous epidemiological studies have shown consistent inverse association between physical activity and colon and rectal cancer risk (Slattery et al., 2003a), with several proposed biological mechanisms to explain association including increased gut motility, decreased insulin and obesity and enhancing immune system by influencing prostaglandin levels (Slattery, 2004). The association observed in the NIH-CCFR dataset is inconsistent with the literature, which could be due to the socio-economic status of the study subjects whereby subjects from the poor socio-economic background would have a limited access to early diagnosis and treatment, especially in the US where the health service is not publicly funded (Doubeni et al., 2012). However, the lack of data regarding the socio-economic status didn't allow adjusting for it in the regression model in current study.

Compared to the UK-CCSG dataset, an exceptionally high association was observed between family history and colorectal cancer risk in the NIH-CCFR dataset (Table 3.2). This is likely to be due to the case ascertainment strategies employed by the NIH-CCFR study sites such as CCO and UQM, which selected cases based on the family history and age at cancer diagnosis respectively (Table 2.1). In contrast, UK-CCSG study sites employed a case selection criterion, which only excluded cases that had Lynch syndrome.

Since the aspirin (and NSAID) dose information was available in the UK-CCSG dataset, a standardized variable was created that encompassed dose and duration information: 75mg aspirin tablet year. 75 mg aspirin tablet year was defined as the amount of 75 mg aspirin tablets taken each day for one year. This helped in investigating association between aspirin dose and duration combination with CRC risk (75mg aspirin tablet years <0.25 (Ref) OR=1.0; ≥0.25 and ≤4.2 OR=0.73 95% CI=0.53-1.0; >4.2 and ≤10.2 OR=0.68 95% CI=0.49-0.93; >10.2 OR=0.55 95% CI=0.40-0.76; $P_{trend}=1.13 \times 10^{-5}$). However, since the aspirin (and NSAID) dose information was not collected by the NIH-CCFR consortium, association of aspirin dose and duration with CRC risk was not investigated.

3.3.2 SNP variant allele and colorectal cancer risk

In the present study, the variant T allele of a non-synonymous SNP rs1799853, which converts arginine to cysteine at the 144th amino acid position in the CYP2C9 enzyme, was associated with a decrease in colorectal cancer risk (OR=0.82, 95% CI=0.69-0.98, P=0.026) in the UK-CCSG dataset (Table 3.4). This observation is consistent with the results from a recent meta-analysis consisting of 16 case-control studies where the presence of the variant T allele was associated with reduced colorectal cancer risk (Summary OR=0.92, 95% CI=0.86-0.98, P=0.012) (Wang et al., 2014). CYP2C9 is involved in the metabolism of several

xenobiotics (including aspirin), endogenous compounds and activation of pro-carcinogenic compounds in tobacco smoke (Panigrahy et al., 2010, Shou et al., 1996). Using (S)-warfarin as a substrate, the variant allele has been shown to produce a slow metabolizing enzyme variant which retain only ~5-30% of the wild type activity (Bigler et al., 2001, Takahashi et al., 1998, Rettie et al., 1994). This in turn would be expected to reduce risk of adenoma and cancer, as lower quantities of pro-carcinogens would be metabolized to carcinogenic metabolites. Barry et al. provided evidence to this hypothesis, where adenoma recurrence risk increased in smokers carrying the wild type genotype (Former smokers RR=1.26, 95% CI=0.99-1.58; Current smoker RR=1.60, 95% CI=1.19-2.15) but no change in risk for individuals with ≥1 variant allele ($P_{interaction}$=0.04) (Barry et al., 2013).

In the NIH-CCFR dataset, presence of the variant T allele of the intergenic SNP rs6983267 showed a borderline significant association for decrease in colorectal cancer risk (OR=0.83, 95% CI=0.70-1.00, P=0.057). The observation is in concordance with the results of Nan et al. where the presence of the variant allele was associated with reduced colorectal cancer risk (OR=0.83, 95% CI=0.74-0.94, P=0.002) (Nan et al., 2013) and with the results of several GWAS studies (Tenesa et al., 2008, Zanke et al., 2007, Tomlinson et al., 2007). The *MYC* oncogene is 355 kb downstream to rs6983267; *in vivo* and *in vitro* experiments have shown that the SNP impairs binding of transcription factor 7 like-2 (TCF7L2) to the *MYC* promoter, which reduces *MYC* expression and induces resistance to tumorogenesis (Sur et al., 2012, Pomerantz et al., 2009).

The current study showed novel site-specific association between the SNPs rs1799853, rs2302615 and rs1105879 and colon cancer. Whilst the direction of association is similar to that for colorectal cancer risk reported in the literature, these results require further validation as the association was based on few study subjects and observed in only the UK-CCSG dataset. A study carried out by Makar et al. suggested a site-specific association between rs20417 SNP in the *PTGS2* gene with approximately two fold increase in rectal cancer risk in the population based Diet, Activity and Lifestyle Study and nearly 5 fold increase in rectal cancer in risk in the NIH-CCFR study where sibling controls were used (Makar et al., 2013). Site-specific association between rs20417 and rectal cancer was not observed in the UK-CCSG dataset and was not replicated in the NIH-CCFR dataset as the current study used population controls in the NIH-CCFR dataset (Supplementary Table 3).

3.3.3 Modulation of aspirin's efficacy by SNPs

On carrying out interaction analysis between SNP, aspirin only use and colorectal cancer risk, a novel observation was made where the variant allele of rs2070959 in *UGT1A6* showed an increased risk for cancer in aspirin users but not in non-users. SNP rs1105879,

which is in high LD with rs2070959, showed the same direction of association with aspirin use in relation to cancer risk, but didn't reach the significance threshold of 0.05 for interaction (P=0.10). A study in the literature has shown that the presence of variant alleles of either SNPs reduce colon polyp risk in NSAID users (OR=0.53, 95% CI=0.33-0.86) but not in non-users (OR=1.04, 95% CI=0.76-1.43) (Bigler et al., 2001). Furthermore, presence of a variant allele of rs1105879 reduces the risk of rectal cancer in NSAID users (OR=0.66, 95% CI=0.37-1.15) but increases risk in non-users (OR=1.25, 95% CI=0.89-1.77). However, the interaction analysis was carried out in a small number of individuals and the study used unaffected siblings controls rather than population controls (Scherer et al., 2014). The UDP glucuronosyltransferase 1A6 (UGT1A6) enzyme is involved in metabolizing aspirin through glucuronidation of salicylic acid. Presence of variant alleles at rs2070959 and rs1105879 have been associated with ~30-50% reduced enzyme activity compared to the wild-type form (Ciotti et al., 1997). Therefore, the hypothesis was that the presence of variant allele of *UGT1A6* SNPs would reduce risk of colorectal cancer in aspirin users as longer duration would be required to metabolise aspirin. Interaction results in the current study are in contrast with the hypothesis and other studies thus requiring further validation.

Similar to the *UGT1A6* SNPs, presence of a variant allele of rs4936367 in *PAFAH1B2* was associated with increase in colorectal cancer risk amongst aspirin users but not in non-users. Type I PAF acetylhydrolase α_2 (PAFAH1B2) is an enzyme found in erythrocytes that hydrolyse acetyl group of aspirin (Zhou et al., 2011). *In silico* analysis using SIFT and PolyPhen for predicting functional consequence of the SNP suggests that the missense variation would not affect enzymatic activity, however, functional validation using *in vitro* and *in vivo* methods remains to be carried out. The novel interaction observed in the current study provides a new biological pathway to explore for understanding modulation of aspirin's chemopreventive efficacy.

3.3.4 Downstream analysis

In the current study, several known epidemiological and genetic associations with colorectal cancer risk were observed which were consistent with the existing literature, thus vindicating dataset's ability to identify known observations. Whilst the direction of association of all epidemiological risk factors, except physical activity, were consistent in both datasets, association of genetic variants with colorectal cancer was only observed in the UK-CCSG dataset. Inability to replicate these modest genetic associations in the NIH-CCFR dataset could be due to the case ascertainment bias as it affects genetic association estimates and hence, may lead to underestimation of the sample size required to detect genuine association with sufficient power (Zöllner and Pritchard, 2007).

Two novel gene-environment interactions observed in the UK-CCSG dataset were not replicated in the NIH-CCFR dataset thus prompting their validation in either a larger case-control dataset or by carrying out random-effects meta-analysis of the UK-CCSG and NIH-CCFR dataset that would provide higher power to observe interaction and would account for the variation in association effect sizes between the datasets. Power calculation for an unmatched case-control study was carried out where the baseline population risk for colorectal cancer was set at 5% (based on Peters et al. 2012) and 25% of the study subjects were estimated to be aspirin users (based on the percentage of aspirin users in UK-CCSG and NIH-CCFR datasets). From the calculations, it is estimated that a dataset containing 2923 cases and controls would have 80% power to observe an association between a SNP with an MAF of 10% and an increase in colorectal cancer by 20%, which would reach a significance threshold of 0.05. On the other hand, it is estimated that a dataset containing 3076 cases and controls would have 80% power to observe an interaction between a SNP with an MAF of 10% that increases colorectal cancer risk by 20% and 25% of study subjects being aspirin users, which would reach a significance threshold of 0.05. The combined UK-CCSG and NIH-CCFR dataset would consist of 3851 cases and 2262 controls, thus making meta-analysis an ideal methodology to re-assess associations and interactions that reached significance threshold of 0.05 in the two datasets.

3.4 Conclusion

The current chapter presented results from an exploratory analysis of candidate SNPs in aspirin's pharmacokinetic and pharmacodynamic pathways to be associated with colorectal cancer risk or modulate aspirin's chemopreventive efficacy. Two novel interactions between SNPs in *UGT1A6* and *PAFAH1B2* genes with aspirin use in relation to colorectal cancer risk were identified in one of the two datasets. Power analysis estimate suggests that conducting a meta-analysis by combining the two datasets for the associations and interactions with $P \leq 0.08$ identified in the current chapter would have power to be replicated with $P \leq 0.05$. Re-analysis of the results observed in the current chapter is carried out using the random effects meta-analysis of the UK-CCSG and NIH-CCFR datasets in the following chapter.

Chapter 4. (Results 2): Pharmacogenetic influences on colorectal cancer chemoprevention using aspirin- Part 2

4.1 Introduction

4.1.1 Aspirin and colorectal cancer prevention

With an estimated 1.2 million new cases and 609,000 deaths worldwide in 2008 (Jemal et al., 2010), colorectal cancer (CRC) is the third most common cancer in men and second most common in women. Development of a primary prevention strategy for CRC is imperitive. Aspirin is a non-steroidal anti-inflammatory drug (NSAID), which is commonly used as an analgesic, anti-pyretic or as a prophylactic drug to treat cardiovascular diseases (CVD) (Fuster and Sweeny, 2011). A surfeit of evidence from case- control studies, cohort studies, meta-analyses and randomized controlled trials show regular intake of aspirin reduces colorectal adenoma and cancer risk (Ishikawa et al., 2013, Burn et al., 2011b, Burn et al., 2011a, Cuzick et al., 2009, Logan et al., 2008, Bosetti et al., 2006, Cook et al., 2005, Kune et al., 1988). Furthermore, an inverse association between regular aspirin use and risk of distant metastasis and case-fatality has also been reported (Rothwell et al., 2012b, Algra and Rothwell, 2012). Thus there is convincing evidence for aspirin to be prescribed for prophylaxis and adjuvant therapy in patients with risk of CRC.

Despite the extensive evidence of aspirin's chemopreventive efficacy, data from several studies suggests inter-individual variation in the chemopreventive effect and the source of this variation has been attributed to the presence of somatic mutations (Nishihara et al., 2013, Liao et al., 2012) and germline variation in aspirin's pharmacokinetic and pharmacodynamic pathways (Wang et al., 2014, Reimers et al., 2014, Fink et al., 2014, Angstadt et al., 2014, Seufert et al., 2013, Nan et al., 2013, Pathi et al., 2012, Zell et al., 2009, Hubner et al., 2008, Chan et al., 2007, Hubner et al., 2006, Din et al., 2004, Stark et al., 2001). Aspirin's mode of action on cellular pathways within the colonic epithelial cells which in turn reduces the risk of CRC has yet to be elucidated. Delineating aspirin's metabolism pathway and the interaction of its metabolites with molecules involved in key cellular processes associated with tumuorogenesis could not only help validate variation in the chemopreventive efficacy, but also aid in understanding the neoplastic transformation of colonic epithelial cells.

4.1.2 Aspirin's pharmacokinetic pathway

Aspirin is made up of two components: a phenol ring consisting of 6- carbon benzene ring with a carboxyl group and an acetyl moiety (Fuster and Sweeny, 2011). Once aspirin is ingested orally, it crosses the mucosal lining of the stomach and small intestine with bioavailability of 40-50% (Floyd and Ferro, 2014, Pedersen and FitzGerald, 1984). It undergoes hydrolysis to salicylic acid (SA) by human carboxylesterase 2 (CES2) in liver and intestinal microsomes (Tang et al., 2006), and by PAFAH1B2 subunit of Type I platelet activating factor acetylhydrolase (PAFAH) in erythrocytes (Zhou et al., 2011) before entering the systemic circulation.

Approximately, 50% of the aspirin and SA absorbed in the stomach and small intestine is then conjugated during first-pass hepatic metabolism. Three different pathways, which are as follows, then clear SA: First, glucuronidation by UDP-glucuronosyltransferases (UGTs) to form salicyl acyl glucuronide and salicyl phenolic glucuronide (Chen et al., 2007, Kuehl et al., 2006); Second, glycine conjugation by glycine-N-acylase to form salicyluric acid which is further conjugated to produce salicyluric acid phenolic glucuronide (Chen et al., 2007, Campbell et al., 1988) and; Third, oxidation by cytochrome P450 (CYP) enzymes to generate a minor metabolite called gentisic acid (Dupont et al., 1999). High inter-individual variation in the clearance of SA could be explained by the variability in SA conjugation. A study carried out by Hutt et al. showed that the urinary recovery of conjugated metabolites as a percentage of administered dose (900 mg) varied greatly between 129 study participants (Hutt et al., 1986). Overall, plasma half-life of aspirin is ~15-20 minutes across a range of treatment doses.

Genetic polymorphisms in the enzymes UGT and CYP have been shown to encode protein variants with differential metabolic activities (Subramanian et al., 2012, Krishnaswamy et al., 2005). Furthermore, genetic variants (rs1105879 and rs2070959) in the *UGT1A6* gene have been shown to be associated with colorectal adenoma (RR=0.68, 95% CI= 0.52-0.89) and carcinoma risk (OR=3.87, 95% CI=1.04-14.45) (Scherer et al., 2014, Hubner et al., 2006). The same variants have been shown to interact with regular aspirin use in relation to colorectal adenoma where the carriers of the variant allele using aspirin regularly were at a reduced risk of colorectal adenoma (OR=0.66, 95% CI= 0.45-0.95) compared to individuals having wild-type genotype (OR=0.93, 95% CI= 0.60-1.44) ($P_{interaction}$=0.02) in population based case- control studies (Chan et al., 2005). Similarly, a meta-analysis of 16 case- control studies by Wang et al. showed a significant association between the genetic variant rs1799853 in the *CYP2C9* gene with colorectal adenoma (OR=1.39, 95% CI= 1.07-1.81, P=0.013) and cancer (OR=0.92, 95% CI=0.86-0.98, P=0.012)

(Wang et al., 2014). Thus, variation in the metabolism of aspirin could affect its efficacy in preventing colorectal neoplasia.

4.1.3 Aspirin's pharmacodynamic pathway

Once aspirin is absorbed, it enters the portal circulation where it irreversibly inhibits Cyclooxygenase (COX) or Prostaglandin H-synthase (PTGS) enzymes. The primary target of aspirin is the COX-1 enzyme, which it inhibits by irreversibly acetylating Ser529 residue, thereby preventing access of the substrate to the enzyme's catalytic site (Picot et al., 1994). COX-1 is constitutively expressed in platelets and gastric epithelial cells where it converts arachodonic acid to prostanoids such as Thromboxane A2, which is involved in the platelet activation cascade. Thus, inactivation of the COX-1 enzyme using aspirin provides cardio-protection by reducing platelet activation and subsequent aggregation. Experimental evidence suggests that cancer patients exhibit increased platelet activation, which in turn has been shown to support tumor metastasis by protecting metastatic cells from the immune system and aid in attaching cancer cells to the endothelial lining thereby initiating secondary lesions (Gay and Felding-Habermann, 2011). Therefore, observation of reduced risk of distant metastasis in regular aspirin users from observational studies and randomized controlled trials (Rothwell et al., 2012b, Algra and Rothwell, 2012) could be the result of the inactivation of the COX-1 enzyme in platelets.

Another COX enzyme- COX-2, which is induced in response to pro-inflammatory and mitogenic stimuli in monocyte and epithelial cells, is also inactivated by aspirin but has been shown to act in a dose dependent manner (Dovizio et al., 2013, Sharma et al., 2010). Requirement of higher dose and dosing frequency of aspirin could be attributed to the ability of nucleated monocytes and epithelial cells to resynthesize COX-2 enzyme. COX-2 has been shown to be overexpressed in colon cancer, and COX-2 expressing colorectal cancer tissues produce large amounts of prostaglandin E2 which can cause resistance to apoptosis and stimulate cell migration and angiogenesis (Alfonso et al., 2014, Dixon et al., 2013, Eberhart et al., 1994). Acetylation of COX-2 at Ser516 by aspirin (Dovizio et al., 2013) results in generation of lipoxins that has been shown to inhibit cancer cell proliferation and angiogenesis (Ferrandez et al., 2012, Claria and Serhan, 1995).

Further to the COX-dependent pathway, there is growing evidence of the chemopreventive effects of aspirin and salicylate through COX-independent pathways. To date, the only COX-independent target known to interact with aspirin is IκB kinase (IKK). *In vivo* and *in vitro* studies have shown that aspirin and salicylate inhibit IKK, which prevents activation of NF-κB thereby reducing inflammatory and angiogenic responses (McCarty and Block, 2006, Yin et al., 1998). However, a study carried out by Stark et al. 2001 showed

activation and nuclear translocation of NF-κB in colorectal cancer cell lines that was induced by aspirin which was followed by apoptosis. The study also showed that this effect was specific to the cells of colonic origin only, thus suggesting a tissue specific effect of aspirin on NF-κB signaling (Din et al., 2004, Stark et al., 2001). Furthermore, a causal link between the nuclear translocation of NF-κB and apoptosis preceded by the activation of an upstream regulator c-Src tyrosine kinase in CRC cells with aspirin has been demonstrated, suggesting c-Src as an upstream mediator of NF-κB signaling in CRC cells (Brady et al., 2011).

Other chemopreventive mechanisms that have been mentioned in the literature include nuclear caspase-dependent cleavage of Sp1, Sp3 and Sp4 specificity protein transcription factors induced by aspirin, which was associated with the downregulation of genes involved in cell survival, proliferation and angiogenesis (Pathi et al., 2012); decrease in the ATPase and selective inhibition of DNA cleavage activity of the topoisomerase IIα enzyme by the primary metabolite of aspirin- salicylic acid (Bau et al., 2014); decrease in cellular glucose consumption and inhibition of cell proliferation through inhibition of 6-phosphofructo-1-kinase activity by aspirin and salicylic acid (Spitz et al., 2009) and; activating polyamine catabolism by increasing the expression and activity of spermidine N-acetyltransferase in colonic mucosa by aspirin, thus reducing the risk of colorectal neoplasia (Martínez et al., 2003).

Genetic variation in both COX-dependent and independent pathways have been shown to be associated with CRC and modulate aspirin's chemopreventive efficacy. For example, presence of the single nucleotide polymorphism (SNP) rs20417 in *COX2* has shown to be associated with an increased risk of rectal cancer in two independent observational studies (Diet, Activity, Lifestyle Study OR=1.95, 95% CI=0.80-4.26, P=0.05; Colon Cancer Family Registry OR=4.88, 95% CI=1.54-15.45, P=0.01) (Makar et al., 2013). A study carried out by Seufert et al. 2013, which analyzed SNPs in genes involved in COX-independent pathway, showed association between 3 SNPs (rs9694958, rs10958713 and rs5029748) in the *IκBKβ* gene with lower risk of colorectal or colon cancer ($P<0.05$ for all). Furthermore, 2 SNPs (rs230490 and rs997476) in the *NF-κB* gene were associated with higher CRC risk among NSAID users compared to non-users ($P_{interaction}<0.05$ for both) (Seufert et al., 2013). Additionally, SNPs rs11694911 (RR=1.29, 95% CI=1.08-1.53, P=0.005) and rs2430420 (RR=1.20, 95% CI=1.03-1.40, P=0.022) in the *ODC1* gene have been shown to be associated with an increased risk for colorectal adenoma (Barry et al., 2011). Based on the literature evidence, there is clear evidence that the variation in aspirin's pharmacokinetic and pharmacodynamic pathway could influence CRC risk by modulating aspirin's chemopreventive efficacy.

4.1.4 Aims

In the previous chapter, 43 SNPs across 16 genes involved in aspirin's pharmacokinetic and pharmacodynamic pathways were tested for association and modification of the protective effect of aspirin use on colorectal cancer. Using two large population based case-control datasets, the UK-Colorectal Cancer Study Group (UK-CCSG) and the NIH-Colon Cancer Family Registry (NIH-CCFR), novel site-specific associations between SNPs in *CYP2C9, ODC1* and *UGT1A6* gene with colon cancer risk were observed ($P<0.05$ for all) (Table 3.5). Moreover, SNPs rs4936367 and rs2070959 in the *PAFAH1B2* and *UGT1A6* gene respectively showed novel significant interactions with aspirin only use in relation to CRC risk, where the variant allele for both SNPs was associated with an increase in CRC risk amongst aspirin users but not in non-users (Table 3.6). However, the aforementioned asssociations and interactions were not observed in the NIH-CCFR dataset.. The difference in findings between datasets could be due to the differences in the case acertainment strategies used within the two datasets which may affect the sample size required to detect genuine associations with sufficient power (Zöllner and Pritchard, 2007).

Based on the power calculation for an unmatched case-control study, where the baseline population risk for CRC is set at 5% and 25% of the study subjects are estimated to be aspirin only users, a dataset containing 2923 cases and controls would be required to test for association between the SNP and CRC and; 3076 cases and controls would be required to test for interaction between the SNP, aspirin only use and CRC with 80% power at a 0.05 significance level (see Downstream analysis). Based on the power calculation, carrying out a meta-analysis of the UK-CCSG and NIH-CCFR dataset would be an ideal approach as the combined dataset would consist of 3851 cases and 2262 controls. There are two popular models for meta-analysis: fixed effect and random effects that accounts for different assumptions. In the fixed effect model, it is assumed that a single true effect size exists for all the studies in the analysis and the source of variance between the studies is due to sampling error (Borenstein et al., 2010, Walker et al., 2008). In contrast, in the random effects model it is assumed that there is a distribution of true effect sizes and the aim is to deliniate the mean effect size (Borenstein et al., 2010, Walker et al., 2008). Since UK-CCSG and NIH-CCFR used different case ascertainment strategies (Table 2.1) which may be the source of within-study and between-study variance in the effect size, a random effects model was used to carry out meta-analysis.

The current analysis aimed to re-assess associations and interactions observed in the previous chapter that reached a significance threshold of 0.08 in either datasets using a random effects meta-analysis approach. The cut-off of $P\leq0.08$ was based on the power

calculation for an un-matched case-control study which estimated that 1949 cases and controls would be required to observe an association with 80% power between a SNP with a MAF of 15% and an increase in colorectal cancer risk by 20%, which would reach a significance threshold of 0.08. Furthermore, it was estimated that a dataset containing 2087 cases and controls would have 80% power to observe an interaction between a SNP with an MAF of 15% that increased colorectal cancer risk by 20% and 25% of the study subjects being aspirin users, which would reach a significance thresholf of 0.08. Since the UK-CCSG dataset consisted of 1910 cases and 1276 controls whereas the NIH-CCFR dataset consisted of 1415 cases and 986 controls, a P-value cut-off of 0.08 was used to select asssociations and interactions from the previous chapter for meta-analysis. Analysis of these novel associations and interactions could further our understanding of the neoplastic process in the colon and rectum and add evidence to the relevant aspects of aspirin's mode of action. Additionaly, novel SNPs identified from the analysis could be used to develope a panel of genetic markers, which would be utilised to predict the efficacy of aspirin's prophylactic treatment in patients at risk of CRC.

4.2 Results

4.2.1 Meta-analysis of association between SNP and CRC risk

Overall, 4 SNPs from 3 separate gene loci- *CDKN1A*, *CYP2C9* and *UGT1A6* within the UK-CCSG and 1 intergenic SNP at 8q24 locus within the NIH-CCFR datasets showed association with CRC risk with significance threshold of <0.08 (Supplementary Table 2). These were taken forward for meta-analysis (see Random effects meta-analysis) and the results are presented in Figure 4.1. The variant T alleles of SNPs rs1799853 and rs6983267, which lie in the *CYP2C9* gene and intergenic region respectively, were associated with a reduced risk of CRC (*CYP2C9* rs1799853 Meta-analysis P=0.03; Intergenic rs6983267 Meta-analysis P=0.04). For the other 3 SNPs, no significant association between the variant allele and CRC risk was observed (*CDKN1A* rs1321311 Meta-analysis P=0.12; *UGT1A6* rs2070959 Meta-analysis P=0.37; *UGT1A6* rs1105879 Meta-analysis P=0.20).

On carrying out a site-specific association test between SNPs and colon or rectal cancer, 3 SNPs from 3 separate gene loci- *CYP2C9, ODC1* and *UGT1A6* were found to be associated with colon cancer in the UK-CCSG dataset (Table 3.5 and Supplementary Table 3). Subsequent meta-analysis, showed that the variant T allele of rs1799853 in *CYP2C9* gene, and the variant C allele of rs1105879 in *UGT1A6* gene, were associated with colon cancer (Figure 4.2) (*CYP2C9* rs1799853 Meta-analysis P=0.003; *UGT1A6* rs1105879 Meta-analysis P=0.03). Furthermore, as rs2070959 was in high linkage disequilibrium (LD) with rs1105879 (R^2=0.90), meta-analysis showed a trend for an increase in the risk of colon cancer in the presence of the G allele but didn't reach statistical significance (Meta-analysis P=0.10). Meta-analysis for rs2302615 in *ODC1* gene was not carried out since the SNP was not genotyped in the NIH-CCFR dataset (Supplementary Table 1). No significant associations were observed in the meta-analysis between the SNPs and rectal cancer (Figure 4.3).

Since the test for association with cancer risk was carried out using a dominant inheritance model for the SNPs, the associations were re-assessed using a co-dominant inheritance model for the SNPs. Overall, 2 significant associations between the SNP and CRC in the dominant model were replicated successfully using the co-dominant model (*CYP2C9* rs1799853 Meta-analysis OR=0.85, 95% CI=0.72-1.00, P=0.05, I-squared=0%; Intergenic rs6983267 Meta-analysis OR=0.87, 95% CI=0.80-0.95, P=0.001, I-squared=0%). For the 2 significant associations between the SNPs and colon cancer using a dominant model, association for only rs1799853 in the *CYP2C9* gene was replicated successfully using the co-dominant model (*CYP2C9* rs1799853 Meta-analysis OR=0.78, 95% CI=0.65-0.93, P=0.007, I-squared=0%; *UGT1A6* rs1105879 Meta-analysis OR=1.09, 95% CI=0.99-1.20, P=0.10, I-

squared=0%). This suggests that the results from the dominant inheritance model largely mimic the results that would have been obtained using the co-dominant model.

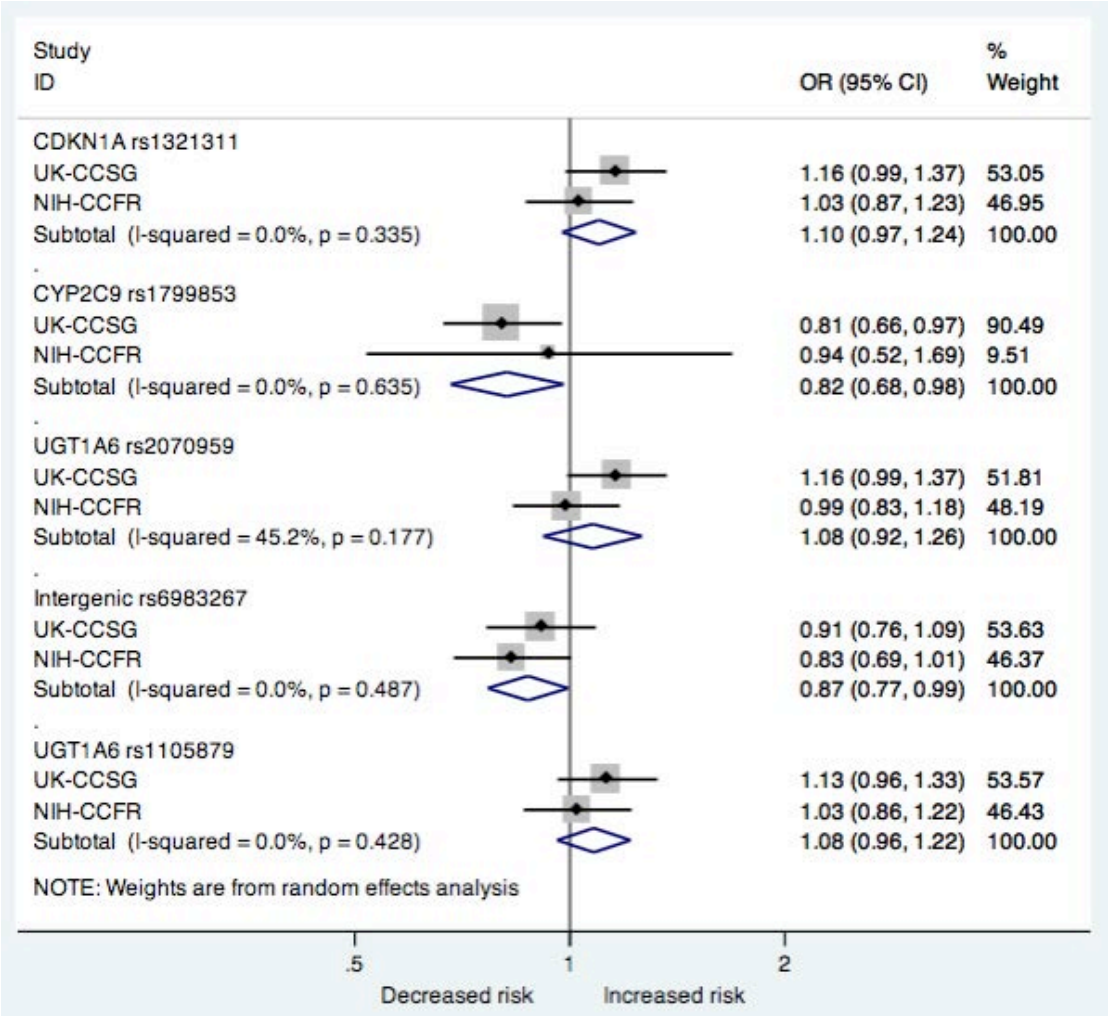


Figure 4.1 Meta analysis of association between the SNP variant allele and CRC risk.

Forest plot depicting meta-analysis odds ratio of association. I-square is the measure of the variation in odds ratio attributable to heterogeneity (Higgins et al., 2003) and p-value tests for heterogeneity between the UK-CCSG and NIH-CCFR dataset.

UK-CCSG, UK-Colorectal Cancer Study Group

NIH-CCFR, NIH-Colon Cancer Family Registry

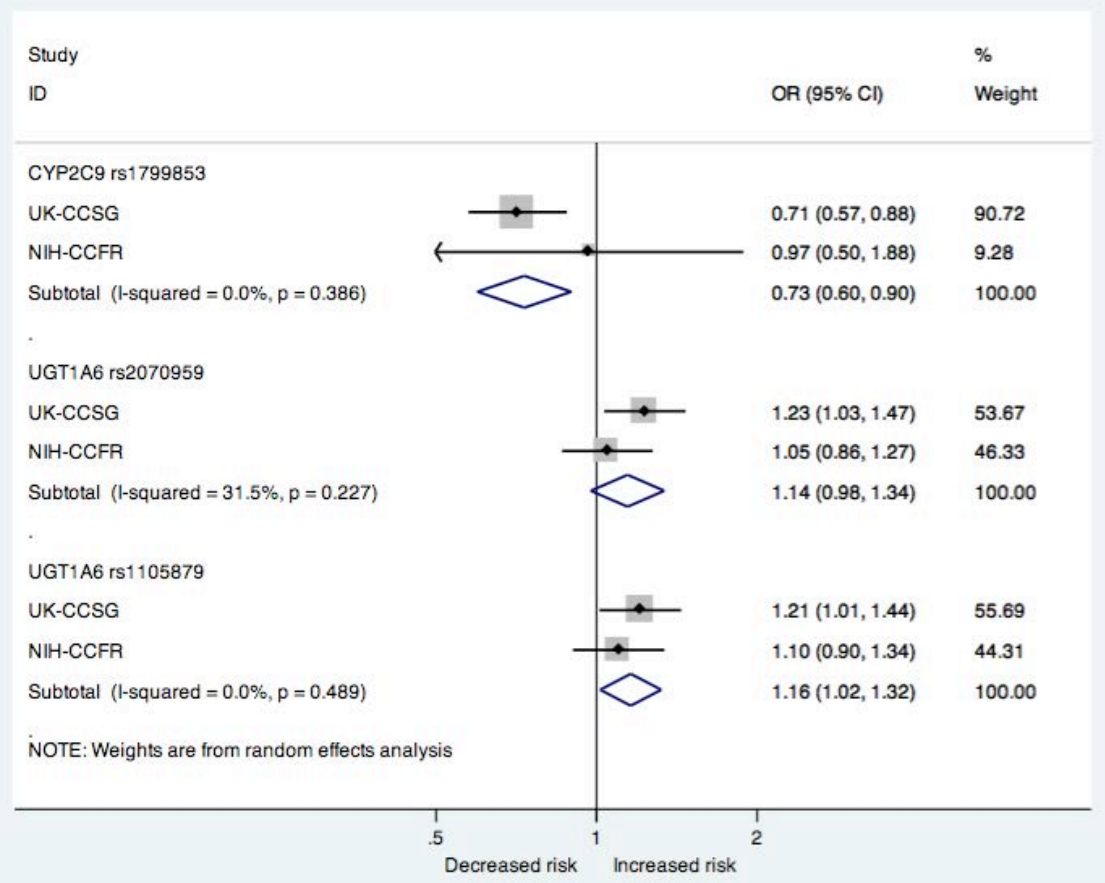


Figure 4.2 Meta- analysis of association between the SNP variant allele and colon cancer risk.

Forest plot depicting meta-analysis odds ratio of association. I-square is the measure of the variation in odds ratio attributable to heterogeneity (Higgins et al., 2003) and p-value tests for heterogeneity between the UK-CCSG and NIH-CCFR dataset.

UK-CCSG, UK-Colorectal Cancer Study Group

NIH-CCFR, NIH-Colon Cancer Family Registry

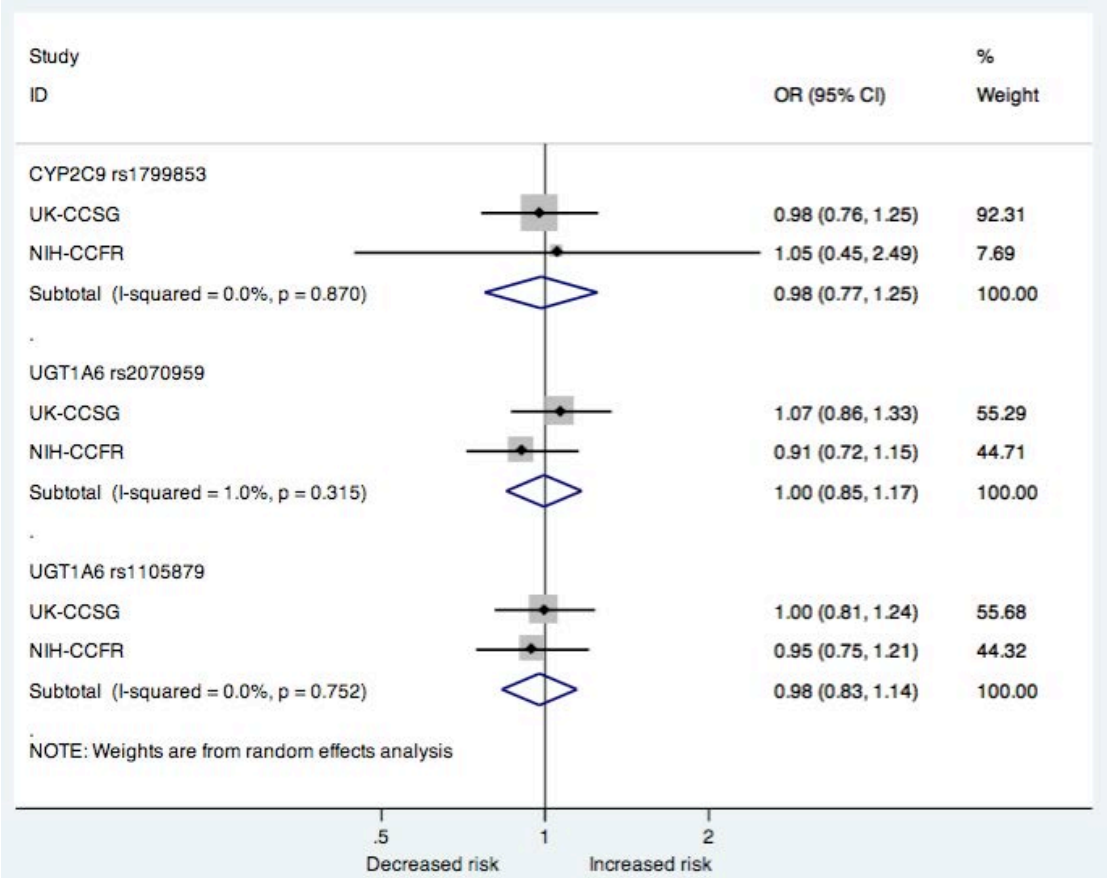


Figure 4.3 Meta-analysis of association between the SNP variant allele and rectal cancer risk.

Forest plot depicting meta-analysis odds ratio of association. I-square is the measure of the variation in odds ratio attributable to heterogeneity (Higgins et al., 2003) and p-value tests for heterogeneity between the UK-CCSG and NIH-CCFR dataset.

UK-CCSG, UK-Colorectal Cancer Study Group

NIH-CCFR, NIH-Colon Cancer Family Registry

4.2.2 Meta-analysis of interaction between aspirin only use, SNP genotype and colorectal cancer risk

Overall, 2 SNP variants in the UK-CCSG dataset, rs4936367 in *PAFAH1B2* and rs2070959 in *UGT1A6* showed a significant gene-environment interaction with aspirin only use in regards to CRC risk. The presence of the variant alleles was associated with an increased risk of CRC in aspirin users but not in non-users ($P_{interaction} \leq 0.05$) (Table 3.6). SNPs rs7112513 and rs1105879 that were in high linkage disequilibrium with rs4936367 and rs2070959 respectively, showed a similar trend for interaction but didn't reach significance threshold of <0.08. Since a significant interaction was only observed in the UK-CCSG dataset, all 4 SNPs were selected for the meta-analysis of interaction.

On carrying out meta-analysis of the gene X environment (GxE) interaction term, all 4 SNPs showed a significant interaction with aspirin only use and CRC risk (Figure 4.4 A). To explore the interaction, the association between the SNP variant allele and CRC risk was stratified by aspirin only users and non-users. For all 4 SNPs, presence of the variant allele was associated with an increase in risk of CRC in aspirin users (*PAFAH1B2* rs4936367 OR=1.32, 95% CI=0.79-2.22; *PAFAH1B2* rs7112513 OR=1.28, 95% CI=0.82-2.01; *UGT1A6* rs2070959 OR=1.38, 95% CI=1.06-1.79; *UGT1A6* rs1105879 OR=1.36, 95% CI=1.05-1.75) but not in non-users (*PAFAH1B2* rs4936367 OR=0.91, 95% CI=0.76-1.10; *PAFAH1B2* rs7112513 OR=0.90, 95% CI=0.72-1.14; *UGT1A6* rs2070959 OR=0.97, 95% CI=0.81-1.15; *UGT1A6* rs1105879 OR=0.99, 95% CI=0.85-1.15) (Figure 4.4 B-D).

The GxE interaction meta-analysis was re-assessed using the co-dominant inheritance model where only 2 SNPs in the *UGT1A6* gene showed significant interaction with aspirin only use and CRC risk (rs2070959 OR=1.35, 95% CI=1.08-1.70, $P_{interaction}$=0.008, I-squared=0%; rs1105879 OR=1.32, 95% CI=1.06-1.64, $P_{interaction}$=0.01, I-squared=0%) whereas, 2 SNPs in the *PAFAH1B2* gene showed a trend for interaction but didn't reach statistical significance threshold (rs4936367 OR=1.40, 95% CI=0.94-2.11, $P_{interaction}$=0.10, I-squared=25.7%; rs7112513 OR=1.36, 95% CI=0.96-1.91, $P_{interaction}$=0.08, I-squared=0%). When stratified by aspirin only users and non-users, all 4 SNPs suggested a trend for an increased risk of CRC in aspirin users (*PAFAH1B2* rs4936367 OR=1.32, 95% CI=0.74-2.34; *PAFAH1B2* rs7112513 OR=1.26, 95% CI=0.79-2.02; *UGT1A6* rs2070959 OR=1.27, 95% CI=1.03-1.55; *UGT1A6* rs1105879 OR=1.24, 95% CI=1.03-1.50) but not in non-users (*PAFAH1B2* rs4936367 OR=1.32, 95% CI=0.79-2.22; *PAFAH1B2* rs7112513 OR=1.28, 95% CI=0.82-2.01; *UGT1A6* rs2070959 OR=1.38, 95% CI=1.06-1.79; *UGT1A6* rs1105879 OR=1.36, 95% CI=1.05-1.75). This suggests that the results from dominant model largely agree with the results from co-dominant model.

Since rs1105879 and rs2070959 in the *UGT1A6* gene showed site-specific association with colon cancer in the meta-analysis, site-specific interaction between these SNPs, aspirin only use and colon or rectal cancer was carried out. Meta-analysis of the GxE interaction term showed a significant interaction with colon cancer (Figure 4.5 A) but not rectal cancer (Figure 4.6 A). When stratified by aspirin only users and non-users, the variant allele in both SNPs was associated with an increase in risk of colon cancer in aspirin users but not non-users (Figure 4.5 B and C). No significant association between the variant allele of both SNPs with rectal cancer in aspirin users and non-users was observed (Figure 4.6 B and C).

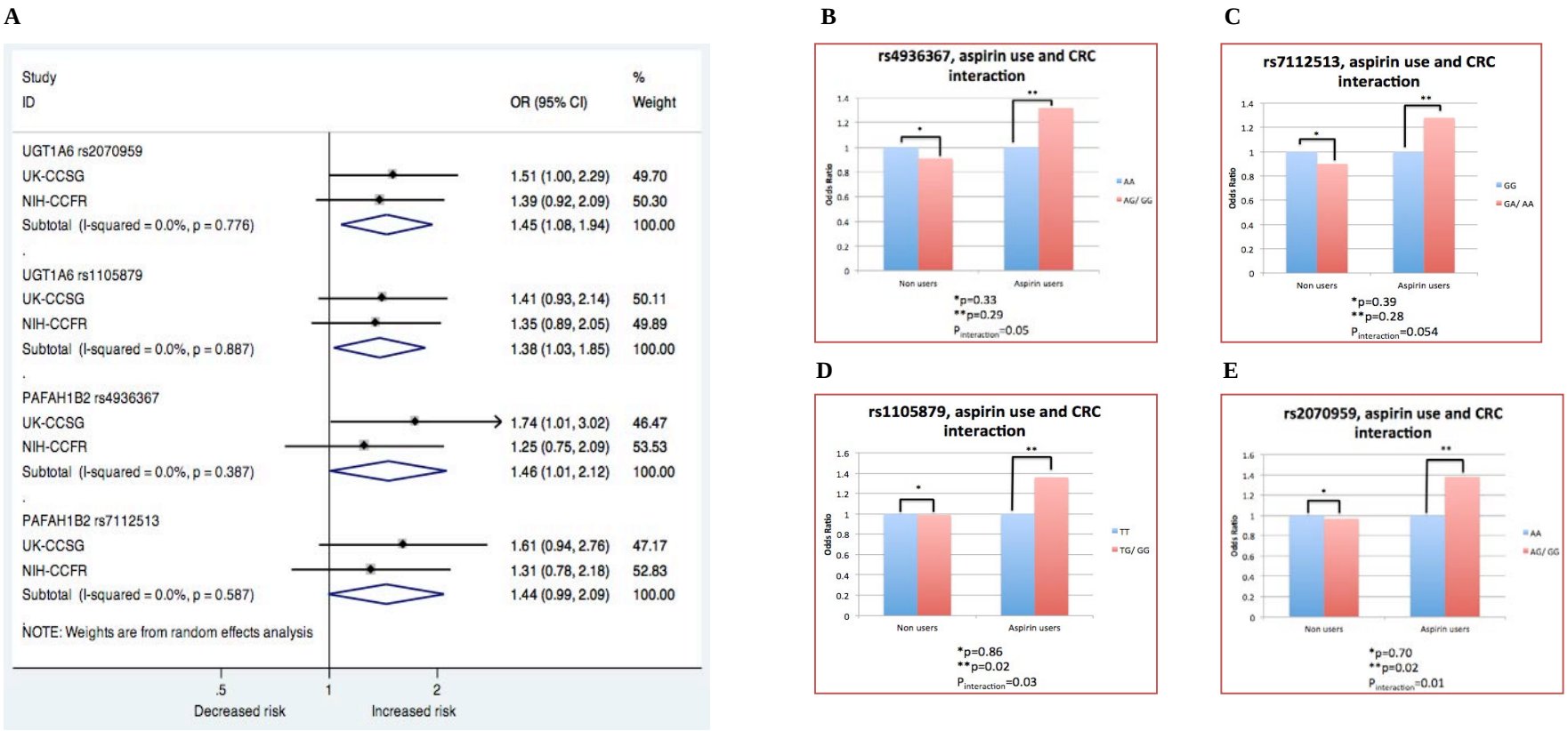


Figure 4.4 Meta-analysis of interaction between SNP variant allele, aspirin only use and colorectal cancer risk.

(A) Forest plot depicting meta-analysis odds ratio of GxE interaction term. I-squared is the measure of the variation in odds ratio attributable to heterogeneity (Higgins et al., 2003) and p-value tests for heterogeneity between the UK-CCSG and NIH-CCFR dataset.
(B) Association between *PAFAH1B2* SNP rs4936367 variant allele and colorectal cancer risk stratified by aspirin use.
(C) Association between *PAFAH1B2* SNP rs7112513 variant allele and colorectal cancer risk stratified by aspirin use.
(D) Association between *UGT1A6* SNP rs2070959 variant allele and colorectal cancer risk stratified by aspirin use.
(E) Association between *UGT1A6* SNP rs1105879 variant allele and colorectal cancer risk stratified by aspirin use.

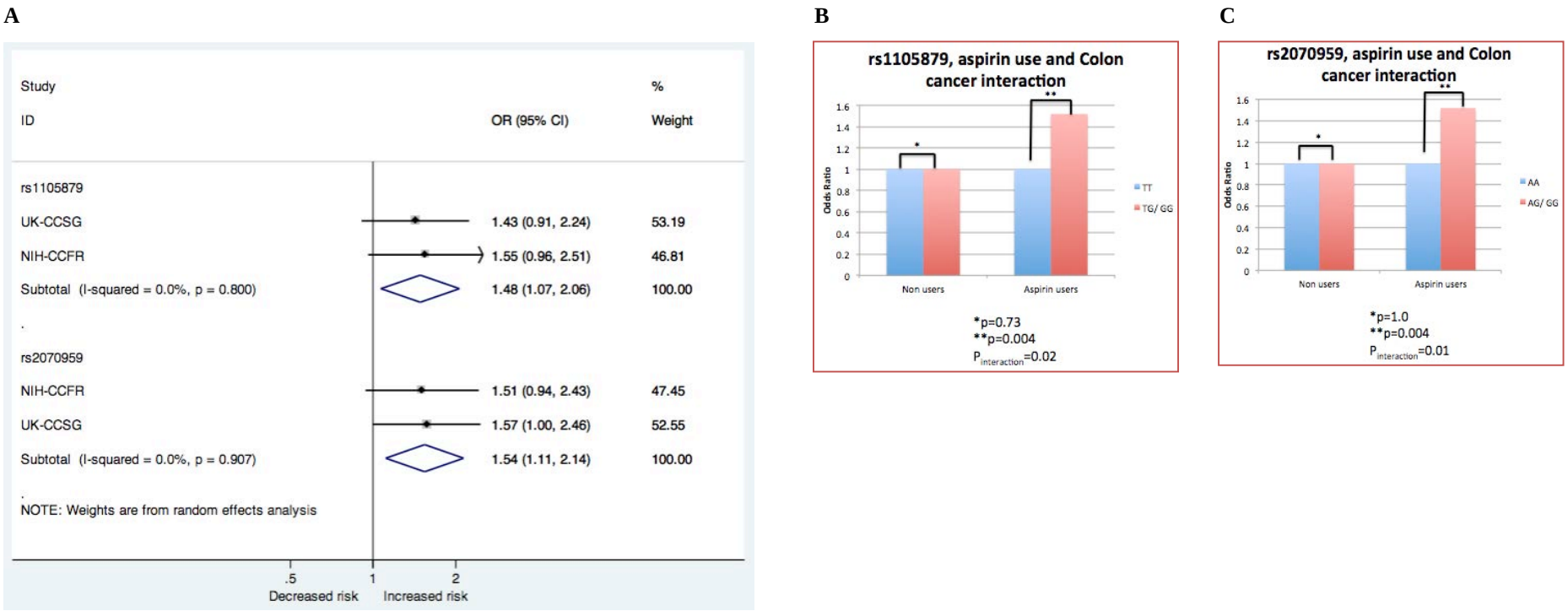
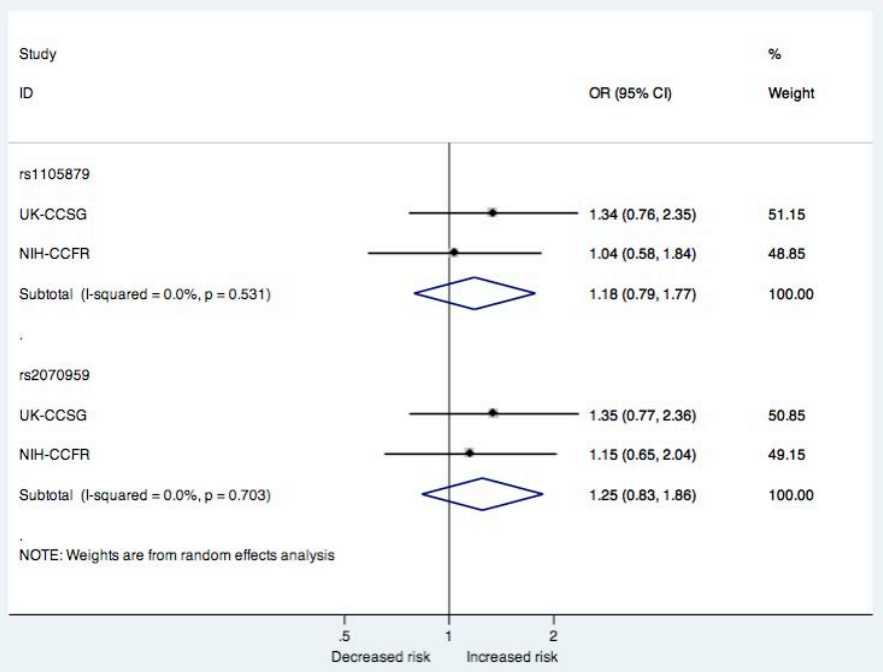


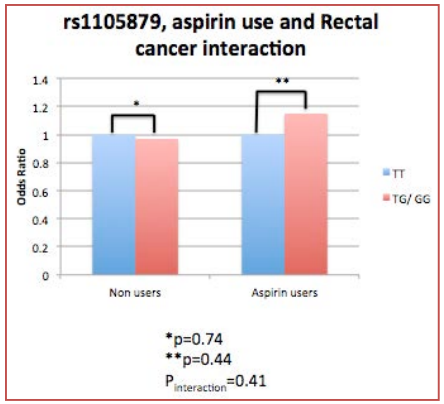
Figure 4.5 Meta-analysis of site-specific interaction between the SNP variant allele, aspirin only use and colon cancer risk.

(A) Forest plot depicting meta-analysis odds ratio of GxE interaction term. I-squared is the measure of the variation in odds ratio attributable to heterogeneity (Higgins et al., 2003) and p-value tests for heterogeneity between the UK-CCSG and NIH-CCFR dataset.
(B) Association between *UGT1A6* SNP rs1105879 variant allele and colon cancer risk stratified by aspirin use.
(C) Association between *UGT1A6* SNP rs2070959 variant allele and colon cancer risk stratified by aspirin use.

A



B



C

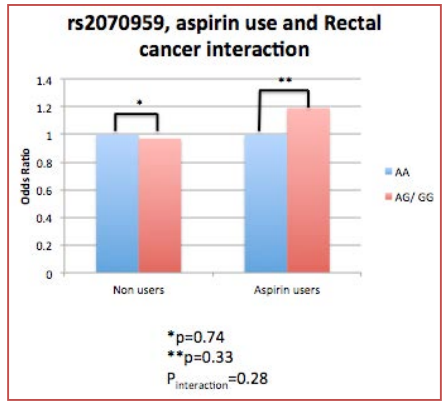


Figure 4.6 Meta-analysis of site-specific interaction between the SNP variant allele, aspirin only use and rectal cancer risk.

(A) Forest plot depicting meta-analysis odds ratio of GxE interaction term. I-squared is the measure of the variation in odds ratio attributable to heterogeneity (Higgins et al., 2003) and p-value tests for heterogeneity between the UK-CCSG and NIH-CCFR dataset.
(B) Association between *UGT1A6* SNP rs1105879 variant allele and rectal cancer risk stratified by aspirin use.
(C) Association between *UGT1A6* SNP rs2070959 variant allele and rectal cancer risk stratified by aspirin use.

4.2.3 Meta-analysis of interaction between aspirin only use, SNP haplotype and colorectal cancer risk

As the 2 SNPs in the *PAFAH1B2* and *UGT1A6* gene were in LD with each other, SNP haplotype interaction with aspirin only use and colorectal cancer risk was tested. However, since rs4936367 and rs7112513 in the *PAFAH1B2* gene showed a high degree of LD (R^2=0.99) in both datasets (Supplementary Figure 2 and Supplementary Figure 3), haplotype analysis was not carried out for these SNPs. In contrast, the *UGT1A6* SNPs have been shown to have two haplotype alleles, *UGT1A6*2* (rs1105879 and rs2070959) and *UGT1A6*4* (rs1105879 single mutation) (Hubner et al., 2006).

Overall, no association between the *UGT1A6* haplotypes and colorectal cancer risk in either of the datasets reached P<0.08 (UK-CCSG *UGT1A6*2* OR=1.16, 95% CI=0.98-1.37, P=0.08; UK-CCSG *UGT1A6*4* OR=1.03, 95% CI=0.70-1.50, P=0.89; NIH-CCFR *UGT1A6*2* OR=1.01, 95% CI=0.84-1.20, P=0.94; NIH-CCFR *UGT1A6*4* OR=1.16, 95% CI=0.74-1.83, P=0.52) and therefore, meta-analysis for association with cancer risk was not carried out. Furthermore, when the test for association of *UGT1A6* haplotypes with CRC risk stratified by aspirin only use, no interaction between either of the haplotypes with CRC and aspirin only use was observed in the NIH-CCFR dataset (For *UGT1A6*2*: Aspirin only use OR=1.25, 95% CI=0.88-1.79 versus Non-users OR=0.90, 95% CI=0.71-1.13, $P_{interaction}$=0.13; For *UGT1A6*4*: Aspirin only use OR=1.36, 95% CI=0.53-3.53 versus Non-users OR=1.20, 95% CI=0.68-2.13, $P_{interaction}$=0.88). Similarly, no significant association between the *UGT1A6* haplotypes and CRC risk was observed when stratified by aspirin only use in the UK-CCSG dataset (For *UGT1A6*2*: Aspirin only use OR=1.55, 95% CI=1.06-2.25 versus Non-users OR=1.05, 95% CI=0.86-1.28, $P_{interaction}$=0.07; For *UGT1A6*4*: Aspirin only use OR=1.18, 95% CI=0.53-2.66 versus Non-users OR=0.96, 95% CI=0.60-1.54, $P_{interaction}$=0.69). Since no significant interaction was observed in either of the datasets, meta-analysis for interaction between *UGT1A6* SNP haplotypes, aspirin only use and CRC risk was not carried out.

4.3 Discussion

There is an extensive evidence of the chemopreventive effect of regular use of aspirin in relation to colorectal cancer; however, data from recent studies suggests inter-individual variation in the chemopreventive effect that has been attributed to the presence of somatic mutations (Nishihara et al., 2013, Liao et al., 2012) and germline variation (Wang et al., 2014, Reimers et al., 2014, Fink et al., 2014, Angstadt et al., 2014, Seufert et al., 2013, Nan et al., 2013, Pathi et al., 2012, Zell et al., 2009, Hubner et al., 2008, Chan et al., 2007, Hubner et al., 2006, Din et al., 2004, Stark et al., 2001). Several COX-dependent and –independent pathways have been implicated in aspirin's mode of action on colonic epithelium for reducing CRC risk. Furthermore, genetic variants in these pathways have been suggested to modulate aspirin's chemopreventive efficacy in prior studies.

The association and interaction estimates reported in the literature however were based on datasets with small sample size or case ascertainment bias thus requiring re-assessment either in a new dataset or by carrying out meta-analysis using several datasets. The current study employed two large population-based case-control datasets (UK-CCSG and NIH-CCFR) consisting of a combined total of 3851 colorectal cancer cases and 2262 controls of self reported non-Hispanic white ethnicity. Based on the power calculations presented in the previous chapter (see Downstream analysis), the random effects meta-analysis approach using the two datasets had 80% power to observe associations and interactions for SNPs with a MAF of 10%. SNPs that reached significance threshold ($P \leq 0.08$) in either of the two datasets were investigated using the meta-analysis approach in the current chapter. In total, 5 SNPs were tested for association with CRC risk and 4 SNPs were tested for GxE interaction with aspirin only use and CRC risk. Furthermore, site-specific cancer (colon and rectum) risk was also tested for these SNPs.

4.3.1 Association with CRC risk

A total of 5 SNPs from 4 separate gene loci were tested for association with CRC risk using the meta-analysis. However, only 2 out of 5 SNPs showed significant association with CRC (Figure 4.1). The variant T allele of rs1799853 in the *CYP2C9* gene was associated with a decrease in risk of CRC. The current observation is in concordance with the study by Wang et al., 2014, where meta-analysis of 16 case-control studies showed a decreased risk of CRC in the presence of the variant T allele of rs1799853 (Summary OR= 0.92, 95% CI=0.86-0.98, P=0.012) (Wang et al., 2014). Cytochrome P450 (CYP2C9) is involved in metabolizing dietary carcinogens and several xenobiotic compounds. *In vitro* studies have shown that the enzyme carrying novel amino acids encoded by the SNP variants (rs1799853 or rs1057910)

retain only 5-30% of the activity of the wild-type enzyme (Haining et al., 1996). Since CYP2C9 is involved in the metabolism of pro-carcinogenic compounds such as benzo[*a*]pyrene (Shou et al., 1994), it can be hypothesized that the carriers of variant alleles have a reduced ability to metabolise pro-carcinogenic compounds and thus have a reduced risk of cancer. Further to the association with CRC risk, the variant T allele of rs1799853 also showed a novel site-specific association for reduction of colon cancer risk but not rectal cancer risk in the current study (Figure 4.2 and Figure 4.3). Several hypotheses such as differential expression and activity of the enzyme across the intestine leading to different rate of metabolism of pro-carcinogens (Läpple et al., 2003), spatial differences in the gut microbiome and distinct gene-specific methylation profile and somatic molecular characteristics of proximal and distal CRCs (Deng et al., 2008) could explain this observation. Assessment of the mRNA expression and protein expression of the CYP2C9 enzyme shows a 10 fold higher protein content in the liver than intestine ($P<0.001$) which suggests that majority of the pro-carcinogen metabolism takes place in the liver (Läpple et al., 2003). However, assessment of the difference in protein expression between colon and rectum in the Human Protein Atlas (http://www.proteinatlas.org) shows medium protein expression in the colon but no expression in the rectal tissue (Uhlen et al., 2015). This suggests the possibility of site-specific metabolism of pro-carcinogens in the colon and compliments the observation of site-specific reduction of colon cancer risk only in the presence of the variant allele made in the current study. Therefore, this association should be considered as a hypothesis generating observation and warrants further analysis in other datasets.

Variant T allele of the SNP rs6983267 was also observed to be associated with a decreased risk of CRC (Figure 4.1). This observation is in concordance with the three genome wide association studies (GWAS) that have consistently shown 15-18% reduced risk for CRC in association with the presence of the variant T allele of rs6983267 (Tenesa et al., 2008, Zanke et al., 2007, Tomlinson et al., 2007). The nearest gene locus to this SNP is the *MYC* oncogene, which has been implicated in tumorigenesis. Previous experiments have shown impaired binding of WNT pathway related transcription factor 7 like-2 (TCF7L2) protein with the *MYC* promoter in the presence of variant T allele, thus inhibiting *MYC* expression and inducing resistance to intestinal tumorigenesis (Sur et al., 2012, Tuupanen et al., 2009, Pomerantz et al., 2009). Additionally, MYC acts as a transcriptional activator of the downstream gene *ODC1* which is involved in polyamine synthesis (Zell et al., 2009). ODC activity and polyamine levels are observed to increase in colon cancer (Pegg et al., 1994), therefore, it could be hypothesized that the variant T allele of rs6983267 would lead to

reduced *MYC* expression which would lead to reduced ODC activity and polyamine levels thus reducing CRC risk.

Another novel association observed in the current study was the site-specific association of SNP rs1105879 in the *UGT1A6* gene with colon cancer but not rectal cancer (Figure 4.2 and Figure 4.3). SNP rs2070959 which is in high LD (R^2=0.90) with rs1105879 showed a similar trend for association but didn't reach statistical significance. In accordance with current results, a study conducted by Thompson et al., 2009 suggested a trend for an increased risk of colon cancer in the presence of *UGT1A6* SNPs but the association didn't reach statistical significance threshold possibly due to the small study size consisting of 422 cases and 481 population controls (Thompson et al., 2009). These functionally relevant SNPs have previously been associated with reduced risk of colorectal adenoma recurrence in an aspirin intervention trial (Hubner et al., 2006) but a 3 SNP genotype consisting of the 2 aforementioned SNPs is associated with an increased risk of CRC in the case- unaffected sibling control cohort of the NIH-CCFR dataset (Scherer et al., 2014). This novel site-specific association has not been reported in the literature and warrants further investigation and validation in other datasets to generate hypothesis that might explain the biological mechanism behind this association.

4.3.2 Association with CRC risk stratified by aspirin only use

A total of 4 SNPs from two gene loci, *PAFAH1B2* and *UGT1A6*, were tested for interaction with aspirin only use and CRC. All 4 SNPs reach significance threshold for the GxE interaction term (Figure 4.4 A) and the variant allele in all 4 SNPs was associated with an increased CRC risk in aspirin only users but not in non-users (Figure 4.4 B-E). This is the first study to report an interaction between SNPs in the *PAFAH1B2* gene, aspirin only use and CRC.

Type 1 platelet activating factor acetylhydrolase subunit 2 (PAFAH1B2) acetylates aspirin to salicylic acid in erythrocytes as a heterodimer with PAFAH1B3 (Zhou et al., 2011) and in plasma as a homomer (Zhou et al., 2013). Currently, no literature exists describing functional variants in the *PAFAH1B2* gene that modulates its acetylhydrolase activity. To predict the impact of SNPs on protein function, the Variant Effect Predictor (VEP) tool in 1000 Genomes database was used (http://browser.1000genomes.org/Homo_sapiens/UserData/UploadVariations?db=core). SNP rs7112513 is an intron variant whereas rs4936367 is a missense variant found in exon 7 and results in a valine to methionine change at amino acid position 151 in the protein. Since rs4936367 is a missense variant, it was hypothesized to affect protein function, however, the VEP tool predicts that the missense variation is tolerated by the protein (SIFT score= 0.23;

PolyPhen score=0.353) and suggests that it has very little or no impact on its function. Absence of functional data restricts hypothesis generation for the observed GxE interaction in the current study and the novelty of the observation warrants validation in other datasets.

Like the *PAFAH1B2* SNPs, two *UGT1A6* SNPs showed association for an increased risk of CRC in aspirin only users but not in non-users in the presence of the variant allele (Figure 4.4 E-F). However, a study conducted by Bigler et al., 2001 showed association for a decreased risk of colon adenoma in aspirin users (OR=0.53, 95% CI=0.33-0.86) but not in non-users (OR=1.04, 95% CI=0.76-1.43) in the presence of the variant allele (Bigler et al., 2001). Furthermore, a study by Chan et al., 2005 showed an interaction trend for colorectal adenoma similar to that observed in the Bigler et al., 2001 study ($P_{interaction}$=0.02) (Chan et al., 2005). This is the first study to report an interaction between the two *UGT1A6* SNPs, aspirin only use and CRC. The contrast in the association observed for cancer risk in the current study and adenoma risk in the literature could be attributed to a difference between the key genetic and epigenetic molecular differences between adenoma and carcinoma which may make carcinoma cells more sensitive to aspirin intervention. This hypothesis could be backed by the results from the CAPP trial where aspirin intervention didn't reduce polyp number in the sigmoid colon or rectum (RR=0.77, 95% CI=0.54-1.10) but reduced risk of CRC after 2 years of intervention (HR=0.41, 95% CI=0.19-0.86, P=0.02; IRR=0.37, 95% CI=0.18-0.78, P=0.008) (Burn et al., 2011b, Burn et al., 2008).

Additionally, the two *UGT1A6* SNPs in the current study showed significant interaction with aspirin only use and site specific colon cancer but not rectal cancer where the risk of colon cancer was increased in aspirin users and not non-users having the variant allele (Figure 4.5 and Figure 4.6). A similar trend for interaction was observed between the two SNPs, NSAID use and colon cancer in a study by Thompson et al., 2009 but the interaction didn't reach the significance threshold possibly due to the small study size consisting of 422 cases and 481 population controls. In contrast, a study conducted by Scherer *et al.,* 2014 showed that the SNP rs1105879 was associated with lower risk of rectal cancer in NSAID users but not non-users in the presence of the variant allele ($P_{interaction}$=0.02) (Scherer et al., 2014). However, this observation was based on a small study sample of 445 rectal cancer cases and 743 sibling controls and hasn't been replicated in any other datasets.

Salicylic acid formed after hydroxylation of aspirin in stomach and systemic circulation is metabolized through glucoronidation by UDP-glucuronosyltransferases (UGTs) to form salicyl acyl glucuronide and salicyl phenolic glucuronide (Chen et al., 2007, Kuehl et al., 2006). *In vitro* biochemical assays showed that the enzyme variant containing rs2070959 and rs1105879 had 41-74% metabolic activity of the wild-type form at certain pH levels (Ciotti et

al., 1997), whereas, a study conducted on liver microsomes showed that the enzyme variant containing rs2070959 and rs1105879 had higher metabolic activity compared to the wild-type form (Nagar et al., 2004). Although, when metabolism of salicylic acid in urine following aspirin dosing to young volunteers was tested, excretion of aspirin and its metabolites after 2-4 hour period was found to be higher in individuals homozygous for rs2070959 and rs1105879 compared to the wild-type individuals suggesting that the variant form of enzyme may confer more rapid glucoronidation compared to wild-type form (Chen et al., 2007). The Chen et al. 2007, study didn't account for the possible change in the pharmacokinetic parameters of the other 2 enzymes involved in salicylic acid metabolism but does support the interaction result of the current study where the carriers of the variant allele had an increased risk of CRC compared to the wild-type individuals.

4.3.3 Clinical utility of the SNPs

Whilst the novel associations and interactions in the current study provide plausible hypotheses of the mechanistic processes of colorectal neoplasia and aspirin chemoprevention, it also provides an opportunity to explore the utility of genotyping SNPs before prescribing aspirin as a prophylactic or an adjuvant drug for preventing CRC. One of the ways to test for their clinical utility is by carrying out burden tests of association, which models the effect of "mutation load" by accumulating minor alleles of several SNPs within a functional unit- in this case the aspirin pharmacokinetic and pharmacodynamic pathway. In this test, each variant is weighted and is assumed that the direction of the association for all variants with the phenotype is same (Moutsianas and Morris, 2014). Since some variants in the current study showed positive association whereas certain variants showed inverse association with CRC, burden tests of association cannot be conducted. However, generalized burden tests, which don't assume the same direction of association for all variants could be used to test the clinical utility of the SNPs identified in the current study (Moutsianas and Morris, 2014). Nonetheless, carrying out generalized burden tests is beyond the scope of the current study as the novel associations and interactions have yet to be validated in other datasets and functional characterization of variant alleles of some SNPs needs to be undertaken.

4.3.4 Study limitations

Whilst the current study highlights plausible biological mechanisms to explore in the future, it also has several drawbacks. First, no multiple test correction procedure to control the type 1 error rate was used due to the relatively small study size of 3851 cases and 2262 controls and modest effect size for associations and interactions. Furthermore, since a flexible approach to study design and analysis was employed which encompassed multiple hypotheses

and multiple tests for each hypothesis, appropriate multiple test adjustment for a global conclusion could be difficult to perform (Bender and Lange, 2001). Hence any "significant" or "novel" observations in the current study should be regarded as exploratory results that warrants further testing in other datasets. Second, even though the log odds and standard error for associations in the NIH-CCFR dataset were adjusted for study site, the case ascertainment strategies employed at the NIH-CCFR sites were hugely different and thus, adjusting association for study site may not encompass the variance in the effect size estimates observed at different sites. However, carrying out meta-analysis where NIH-CCFR study sites were stratified into individual datasets would have reduced power to observe significant association and increased between study heterogeneity. Third, a dominant inheritance model was used for carrying out association and interaction tests in the meta-analysis to increase statistical power. However, upon carrying the tests using co-dominant model, majority of the results had similar estimates to that observed in the dominant model and reached significance threshold of 0.05 that was set for the current study thus suggesting that the associations are robust as they are observed irrespective of the inheritance model used in the test.

4.4 Conclusion

The current chapter presented exploratory meta-analysis results of candidate SNPs in aspirin's pharmacokinetic and pharmacodynamic pathways that were associated with colorectal cancer risk or modulated aspirin's chemopreventive efficacy using two population based case-control datasets- UK-CCSG and NIH-CCFR. SNP rs1799853 in the *CYP2C9* gene and rs6983267 near *MYC* gene were associated with CRC and rs1105879 in *UGT1A6* gene showed a novel site-specific association with colon cancer risk only. Furthermore, novel interactions between SNPs in *UGT1A6* and *PAFAH1B2* genes with aspirin only use in relation to colorectal cancer risk were observed. Lastly, two SNPs in the *UGT1A6* gene showed novel site-specific interaction with aspirin only use and colon cancer risk. Together, these results provide hypothesis-generating observations that involve new biological mechanisms that could be investigated further to help explain aspirin's differential efficacy and provide insight into the neoplastic transformation of cells in colon and rectum. All novel associations and interactions identified in the current study warrant further investigation in other case-control datasets.

Chapter 5. (Results 3): Testing clinical utility of published warfarin dosing algorithms in the Gujarati population from India

5.1 Introduction

5.1.1 Warfarin pharmacogenetic pathway

Warfarin is an oral anticoagulant, which is widely prescribed to manage thromboembolic diseases such as atrial fibrillation, pulmonary embolism and deep vein thrombosis. It is administered as a racemic mixture of 2 optically active enantiomers, S- and R- warfarin, with the S- isomer being 3 to 5 times more potent at anti-coagulating than the R-isomer (Tatarunas et al., 2011). S-warfarin is primarily metabolized by the cytochrome P450 (CYP2C9) enzyme, which catalyzes its conversion to inactive 6-hydroxy and 7-hydroxy metabolites, whereas R-warfarin is metabolized to 10-hydroxywarfarin by CYP1A2 and CYP3A4 (Kaminsky and Zhang, 1997). Two non-synonymous single nucleotide polymorphisms (SNPs) in the *CYP2C9* gene, *CYP2C9**2 (rs1799853) and *CYP2C9**3 (rs1057910), reduce enzymatic activity by 12% and 5% respectively compared to the wild type genotype (Tatarunas et al., 2011). The presence of the variant allele of either of the two SNPs is associated with an increased risk for an adverse event by 2-3 fold during treatment initiation (Aithal et al., 1999, Gage et al., 2008, Pavani et al., 2011). The *VKORC1* gene codes for vitamin K epoxide reductase complex subunit 1 (VKORC1), an enzyme that activates clotting factors by regeneration of vitamin K_1 from vitamin K_1 2,3-epoxide and is the target for warfarin (Choonara et al., 1988, Bell, 1978). A SNP within the *VKORC1* gene promoter, -1639G>A (rs9923231) leads to creation of an E-box binding site which may repress transcription resulting in decreased mRNA levels and therefore resulting in lower levels of active enzyme (Wang et al., 2008, Yuan et al., 2005, D'Andrea et al., 2005).

Several studies have suggested that the combined knowledge of these 3 SNPs, along with demographic and anthropometric variables such as age, gender and body mass index (BMI), could explain up to half of warfarin dose variability (Gage et al., 2008, Yang et al., 2009, Tatarunas et al., 2011). Addition of other genetic polymorphisms in genes involved in warfarin pharmacokinetics and pharmacodynamics (Reider et al., 2007, McDonald et al., 2009a, Voora et al., 2010, Lane et al., 2011, Pavani et al., 2011, Universiy, 2012), concomitant medication use (McDonald et al., 2012, Whitley et al., 2007), indication for warfarin medication (Vink et al., 2003), ethnicity (Whitley et al., 2007), smoking

(Nathisuwan et al., 2011) and vitamin K intake (Lubetsky et al., 1999) could further improve dosing accuracy and explain dose variability.

5.1.2 Benefit of genotype based warfarin dosing

Response to warfarin therapy is monitored using International Normalised Ratio (INR), which is the ratio of patient's prothrombin time to that of a reference population. Several anticoagulation trials report the optimal target INR level to be between 2 and 3 (Oden et al., 2006, Hylek et al., 1996). To improve the time spent within target INR, in 2007, the Food and Drug Administration (FDA) added pharmacogenetic information to the warfarin drug label, but didn't provide any dosing regime to make use of the genetic information. Subsequently, the label was updated in 2010 to include a genotype-stratified dosing table (Lee and Klein, 2013, Klein et al., 2009). A retrospective study carried out by Finkelman et al. concluded that the therapeutic dose predicted using pharmacogenetic algorithms were more accurate compared to the dose calculated using genetic tables or empirical dosing alone (Finkelman et al., 2011). In 2010, the Medco-Mayo Warfarin Effectiveness Study showed 43% lower risk of hospitalization due to bleeding or thromboembolism (HR: 0.57, 95% CI: 0.39 to 0.83, P=0.003) for patients who were given genotyped guided dose during treatment initiation (Epstein et al., 2010).

A recent randomized clinical trial, CoumaGen-II, demonstrated that in intention to treat analysis, patients whose dosage was calculated using a pharmacogenetic algorithm had lower percentage of out of range INR (31% versus 42% at 1 month; 30% versus 42% at 3 months, P<0.001 for both) and spent a higher percentage of time within the therapeutic range (69% versus 58% at 1 month; 71% versus 59% at 3 months, P<0.001 for both) compared to people on a standard dosing regime (Anderson et al., 2012). Two further randomized clinical trials, European Pharmacogenetics of Anti Coagulant Therapy (EU-PACT) (Pirmohamed et al., 2013) and Clarification of Optimal Anticoagulation through Genetics (COAG) (Kimmel et al., 2013), tested the clinical utility and effectiveness of pharmacogenetic guided dosing regimes. In the EU-PACT trial, patients receiving a pharmacogenetic-guided warfarin dose spent a higher percentage of mean time in the therapeutic INR range (67.4%) compared to controls (60.3%) during the initiation of warfarin therapy (Adjusted difference, 7.0 percentage points; 95% CI, 3.3 to 10.6; P<0.001). However, in the COAG trial, no difference was observed in the mean percentage of time in the therapeutic range between the genotype guided group (45.2%) and control group (45.4%) (Adjusted difference, -0.2; 95% CI, -3.4 to 3.1; P=0.91) but a significant interaction between ethnicity and dosing strategy was observed (P=0.003).

Two potential factors could explain conflicting results from the two aforementioned trials. First, the EU-PACT trial used a loading dose algorithm (Avery et al., 2011) whereas the COAG trial used a maintenance dose algorithm (Gage et al., 2008). The loading dose algorithm incorporates pharmacokinetic parameters of S-warfarin based on the *CYP2C9* genotype. A simulation carried out by Avery et al., 2011 illustrated improved plasma S-warfarin and anticoagulation response time profile compared to the maintenance dose algorithm. Second, 1.3% and 27% of the patients on the EU-PACT and COAG trial respectively were Africans. Both trials genotyped patients for *VKORC1, CYP2C9**2 and *CYP2C9**3 SNPs which have been shown to explain a significant proportion of warfarin dose variability in white Europeans but not in non-European populations, suggesting absence of population specific SNPs in the algorithm. In a recent genome wide association study (GWAS) in African American individuals, a novel SNP upstream of *CYP2C18* (rs12777823) was found to be significantly associated with warfarin dose requirement ($P= 4.5 \times 10^{-12}$) and improved dosing accuracy of the Klein et al., 2009 dosing algorithm by 21% (Perera et al., 2014, Perera et al., 2013b). Thus, the EU-PACT trial showed the benefit of genotype guided dosing approach in a genetically homogeneous population whereas the COAG trial failed to show benefit of the approach in a genetically diverse population. This suggests that despite the strong evidence for the clinical utility of pharmacogenetic based dosing strategy, ethnic variations at the genetic and demographic level should be taken into account. Alternatively, another dosing algorithm incorporating additional genotype data such as that for rs12777823 could be used.

5.1.3 Clinical utility of warfarin pharmacogenetics in India

It is estimated that 30% of all deaths worldwide are attributed to cardiovascular diseases (CVDs) out of which ~80% of the burden is from developing countries (Gaziano, 2005). Epidemiological transition to improved life expectancy, high-fat diets, smoking and sedentary lifestyle have been implicated to be causal factors for mortality due to atherosclerotic CVDs, especially at ages below 50 years in urban India (Gaziano, 2005, Yusuf et al., 2001). It is estimated that India suffers the highest loss of potentially productive life years due to deaths from cardiovascular disease (9.2 million years lost in 2000) compared to all other countries (Reddy et al., 2005). With low cost and high efficacy, warfarin is anticipated to remain the drug of choice for preventing cardiovascular related diseases and deaths. Most of the existing literature on warfarin pharmacogenetics is related to the white European population. However, there is an increasing interest in developing genotype-guided warfarin dose algorithm for the Indian population.

Results on the clinical utility of pharmacogenetic algorithms from the studies carried out in the white European population imply that they cannot be translated to the Indian population due to demographic, dietary and genetic based differences between the two populations. A study carried out by the Indian Genome Variation Consortium showed that the heterogeneous Indian population could be divided into 5 clusters based on their genotype frequencies. It suggested that the effect of population stratification on association studies carried out in individuals from one cluster, regardless of ethnicity, may be small (Consortium, 2008). To date, allele frequencies of warfarin dose associated SNPs have been established and a population specific pharmacogenetic algorithm has been developed for the South Indian population (Adithan et al., 2003, Pavani et al., 2011, Pavani et al., 2012). However, the clinical utility of existing pharmacogenetic algorithms for the Gujarati Indians, a sub-population of 60 million which belongs to a different population cluster compared to South Indian population, has not been tested.

5.1.4 Aims

The current study aims to: (A) analyze allele frequencies of a panel of 13 warfarin dose associated SNPs in the Gujarati population and compare them with allele frequencies observed in European, African and Han Chinese populations and (B) test the clinical utility of published genotype- guided dosing algorithms by measuring their dose prediction accuracy in the Gujarati population and comparing them with the prediction accuracy observed in white European and South Indian populations.

5.2 Results

5.2.1 Clinical characteristics of patients treated with warfarin

To further investigate the clinical utility of the published algorithms, 102 patients of Gujarati ethnicity were recruited on the study and had clinical information recorded (see Data collection and study population). Nearly two thirds of the study participants were male and 8 out of 102 patients were current smokers. The mean age, BMI and body surface area (BSA) was 51.3 years, 23.91 kg/m^2 and 1.68 m^2 respectively (Table 5.1). Current warfarin dose was not available for 2 patients but the mean weekly warfarin dose for the rest (100 patients) was 24.4 mg/week.

Parameter	Number of subjects	Mean	Standard deviation	Minimum	Maximum
Age (years)*	102	51.3	13.7	15	76
BMI (kg/m^2)	102	23.91	4.08	16	36.46
BSA (m^2)	102	1.68	0.20	1.15	2.19
Warfarin (mg/ week)	100	24.4	10.4	7	52.5
Number of concomitant medications	102	5.5	2.6	1	14

Table 5.1 Clinical characteristics of 102 patients treated with warfarin

Demographics, warfarin intake and concomitant medication use summary of 102 patients recruited on the study.

* Age is calculated as the difference between the interview date and date of birth.

The mean number of concomitant medications (including herbal medications) was 5.5 with the maximum number of concomitant drugs in one patient being 14 (Table 5.1). Overall, patients co-administered with amiodarone and azole antifungal drugs had lower mean warfarin dose requirement whereas, patients co-administered with statins had a higher mean warfarin dose requirement compared to other drug users thus suggesting an association between concomitant drug use and warfarin dose (Figure 5.1). Furthermore, upon stratifying current warfarin dose into 3 groups to define patient's sensitivity to warfarin: sensitive (≤21 mg/week), intermediate (>21 - 49≤ mg/week) and resistant (>49 mg/week) (based on Klein et al. 2009), 69% of the amiodarone drug users and 100% of the azole antifungal drug users were warfarin sensitive, whereas ~64% of the statin users were warfarin intermediate suggesting an association between concomitant drug use and warfarin dose (Table 5.2).

Warfarin dose stratification	Limits	Number of subjects	Concomitant drugs			
			Amiodarone	Azole	Statin	Others*
Sensitive	21≤ mg/week	51	18	5	11	17
Intermediate	>21 - 49≤ mg/week	47	8	-	21	18
Resistant	>49 mg/week	2	-	-	1	1

Table 5.2 Current warfarin dose (mg/week) and concomitant medication use

Dose stratification based on sensitivity to warfarin and use of concomitant drugs for 100 patients whose current warfarin dose information was available.

Others include aspirin, clopidogrel, multivitamins, antibiotics, non-steroidal anti-inflammatory drugs, folic acid supplement, beta-blockers etc.

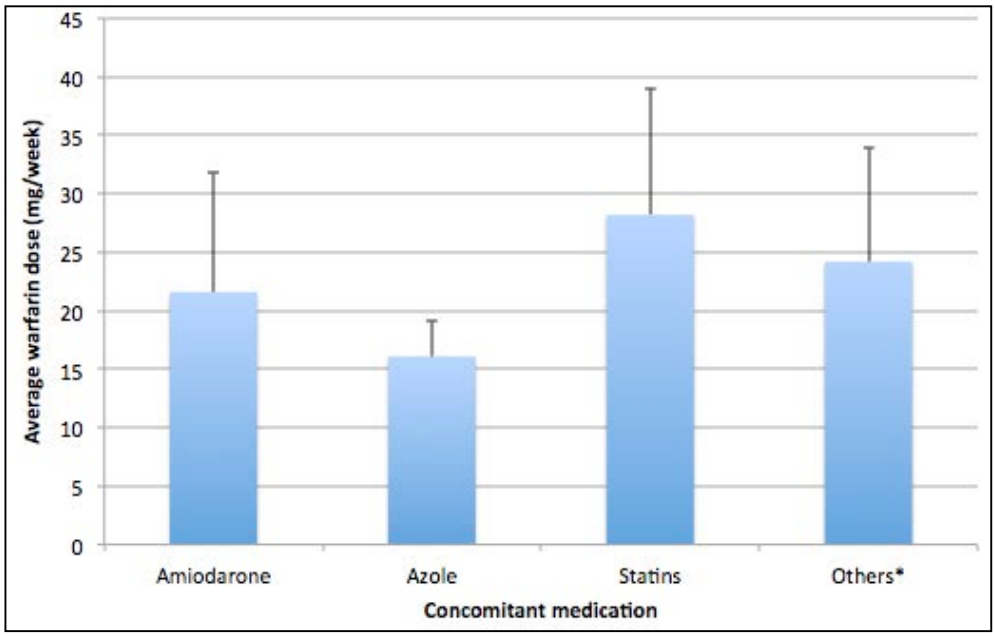


Figure 5.1 Average warfarin dose (mg/week) based on concomitant medication use

Error bars indicate standard deviation.

*Others include aspirin, clopidogrel, multivitamins, antibiotics, non-steroidal anti-inflammatory drugs, folic acid supplement, beta-blockers etc.

Fifty percent of the patients were being treated for valvular heart disease whereas only ~20% of the patients were indicated treatment for atrial fibrillation (Supplementary Table 7). The high percentage of patients treated for valvular heart disease could be either due to high prevalence of rheumatic fever in the Indian population (Seckeler and Hoke, 2011) or due to recruitment bias as the recruitment hospital specializes in heart valve replacement. Two patients were re-hospitalized with adverse reaction during their warfarin treatment.

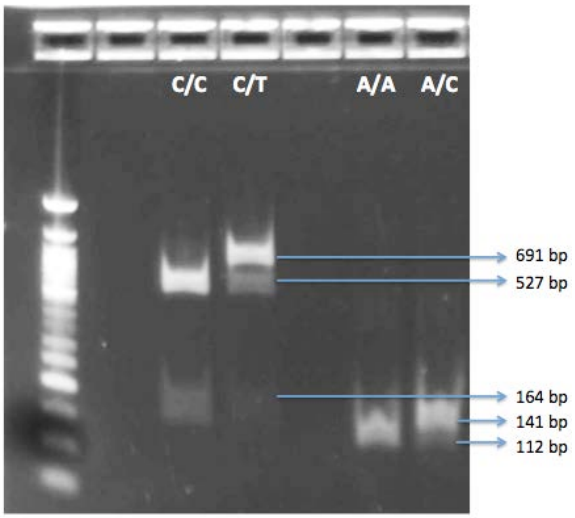
5.2.2 Genetic makeup of the Gujarati population

Out of 102 patients that were treated with warfarin, genetic analysis was not carried out on the 102nd patient as the sample was received after the recruitment deadline; thus removed from all subsequent downstream genetic analysis. A total of 13 SNPs were genotyped in 500 subjects recruited on the study (see Genotype analysis and quality control). These SNPs were selected for the current study based on the literature evidence of their effect on daily warfarin dose or their utility in predicting daily warfarin dose. Out of 13 SNPs analyzed in the study, 2 SNPs (rs2242480 in *CYP3A4* gene and rs3814637 in *CYP2C19* gene) had <80% call rate on the Sequenom and were therefore not included in the downstream analyses (Supplementary Table 8). Validity of the genotype calls made by Sequenom platform was carried out using RFLP PCR on 10% of randomly selected samples for 5 SNPs (rs1799853, rs1057910, rs9923231, rs7294 and rs2108622) (Figure 5.2). 100% concordance was observed between the genotype calls from Sequenom MassARRAY platform and RFLP PCR techniques. Four samples where the SNP variant rs9923231 was called as homozygous for an extremely rare third allele (-1639 G>T), bi-directional Sanger sequencing (GATC Biotech, UK) was carried out for genotype confirmation. From the sequencing chromatograph, it was observed that three of the samples were homozygous wildtype and one sample was homozygous mutant for the second allele (-1639 G>A) (Figure 5.3).

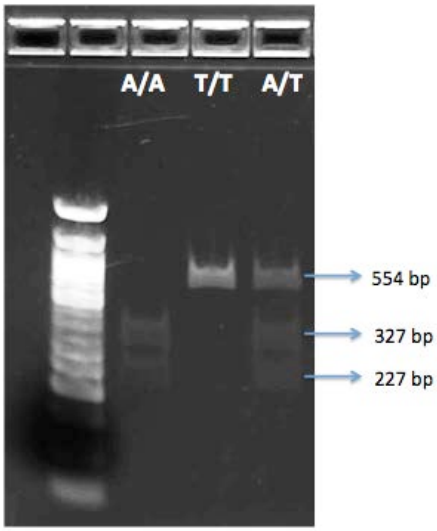
Using Fisher's exact test, no evidence was observed for differences in any of the SNP frequencies between the cohort of 101 patients and 399 healthy individuals with exception of the SNP rs2108622 (P=0.012), and therefore, SNP data from both cohorts were merged (Table 5.3). All SNPs were in concordance with the Hardy- Weinberg equilibrium and no SNP pairs were observed to be in linkage disequilibrium (LD). On comparing the MAF observed in the population enrolled on the current study with the Gujarati Indian in Houston (GIH) population in 1000 Genomes database, the allele frequency for all the SNPs were similar between the 2 populations thus vindicating the genotype calls made in the current study. Overall 5 out of 11 SNPs (rs28371686, rs9332131, rs61742245, rs11676382 and rs339097) had <1% MAF in the Gujarati population (Table 5.3). The highest MAF was observed for rs7294 (variant A allele=68%) and rs2108622 (variant A allele= 43%) in the

Gujarati population as compared to the 3 reference populations- EUR, AFR and CHB. All 6 SNPs with MAF >1% had significantly different frequencies compared to all 3 populations (Table 5.3).

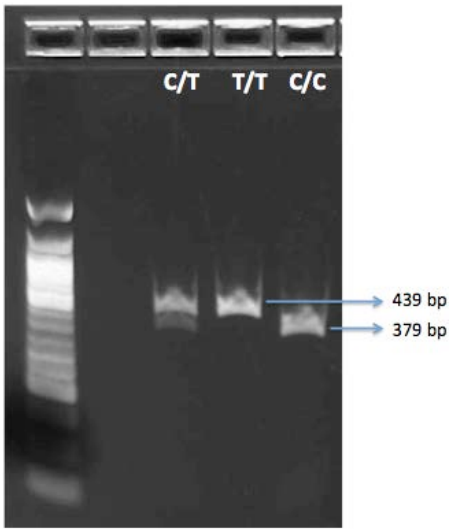
A. *CYP2C9*- rs1799853 and rs1057910



B. *VKORC1*- rs9923231



C. *CYP4F2*- rs2108622



D. *VKORC1*- rs7294

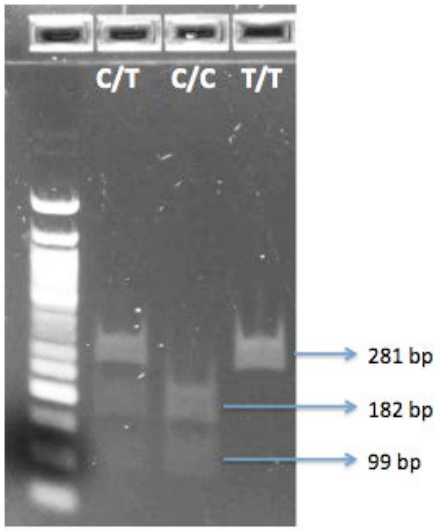


Figure 5.2 Restriction fragment length polymorphism (RFLP) PCR to genotype SNPs. Genotyping rs1799853 and rs1057910 in the *CYP2C9* gene. Lanes consists of homozygous WT and heterozygous MT sample for rs1799853 and rs1057910 SNPs; (B) Genotyping rs9923231 in the *VKORC1* gene. Lanes consists of homozygous WT, homozygous MT and heterozygous MT sample; (C) Genotyping rs2108622 in the *CYP4F2* gene. Lanes consists of heterozygous MT, homozygous MT and homozygous WT sample; (D) Genotyping rs7294 in the *VKORC1* gene. Lanes consists of heterozygous MT, homozygous WT and homozygous MT sample respectively. Lane 1 in all figures represents 2- log 10kb DNA ladder.

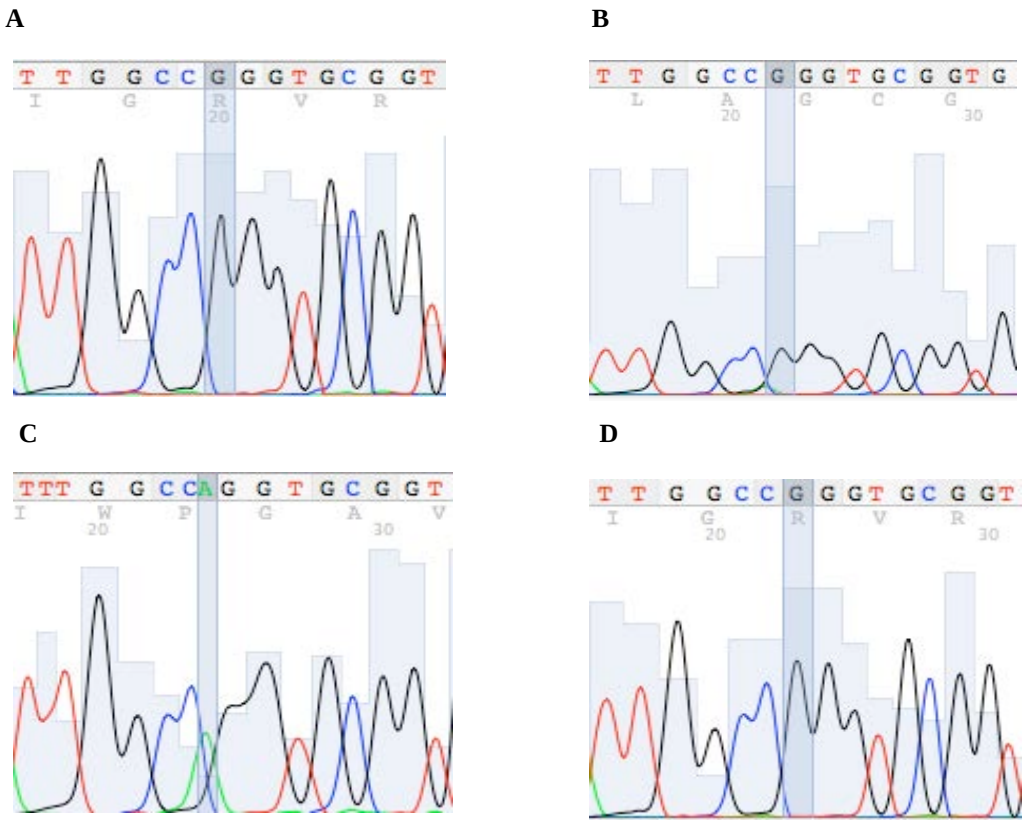


Figure 5.3 rs9923231 SNP genotyping using Sanger sequencing for 4 samples.

Highlighted peak and letter on the chromatograph represents rs9923231 SNP position. Samples re-genotyped are (A) IUKWP63, (B) IUKWP49, (C) IUKWP8 and (D) IUKWP71.

Gene Name	HGVS Name	Polymorphism (N=500)	Cohort comparison (P-value)	Population specific Minor Allele Frequency*				
				Study population	Gujarati Indian (GIH)	European (EUR)	African (AFR)	Han Chinese (CHB)
CYP2C9	8633C>T	rs1799853 (n=451)	0.62	0.06	0.05 (P=0.27)	0.12 (P<0.001)	0.02 (P<0.001)	0.00 (P<0.001)
	47639A>C	rs1057910 (n=455)	0.20	0.08	0.13 (P=0.06)	0.06 (P=0.049)	0.006 (P<0.001)	0.04 (P=0.013)
	47644C>G	rs28371686 (n=446)	1.0	0.00	0.00	0.00	0.02 (P<0.001)	0.00
	15625delA	rs9332131+ (n=400)	0.11	0.001				
VKORC1	3588G>A	rs9923231 (n=472)	0.16	0.21	0.18 (P=0.65)	0.40 (P<0.001)	0.07 (P<0.001)	0.95[a] (P<0.001)
	5332G>T	rs61742245+ (n=420)	1.00	0.00				
	8956G>A	rs7294 (n=425)	0.65	0.68	0.67 (P=0.17)	0.35 (P<0.001)	0.48 (P<0.001)	0.05 (P<0.001)
	5924C>T	rs17708472 (n=434)	0.12	0.15	0.16 (P=0.86)	0.23 (P<0.001)	0.05 (P<0.001)	0.00 (P<0.001)
CYP4F2	23454G>A	rs2108622 (n=465)	0.012	0.43	0.44 (P=0.51)	0.27 (P<0.001)	0.09 (P<0.001)	0.22 (P<0.001)
GGCX	16025G>C	rs11676382 (n=446)	1.0	0.008	0.01 (P=0.68)	0.09 (P<0.001)	0.006 (P=1.0)	0.00 (P=0.36)
CALU	24879A>G	rs339097 (n=386)	1.0	0.005	0.00 (P=0.58)	0.00 (P=0.12)	0.17 (P<0.001)	0.01 (P=0.35)

Table 5.3 Summary of minor allele frequency of 11 SNPs in 5 populations

Comparison of SNP allele frequency between cohort of 101 patients and 399 healthy individuals was carried out using Fisher's exact test. Comparison of allele frequency of 11 SNPs in the Gujarati population with four other populations was carried out using the Fisher's exact test.

* Minor allele frequency data obtained from the 1000 Genomes database for Gujarati Indian in Houston (GIH), European (EUR), African (AFR) and Han Chinese in Beijing (CHB)

+Frequency data unavailable in the 1000 Genomes database for Gujarati Indian in Houston, European, African and Han Chinese in Beijing

a, Only 97 samples were genotyped in the 1000 Genomes database from the Han Chinese population for rs9923231 SNP

N, Total number of study subjects

n, Number of samples called for the genotype in the current study population

5.2.3 Dose prediction accuracy of algorithms within the population

The list of variables incorporated in the two clinical and three pharmacogenetic algorithms to predict therapeutic warfarin dose is shown in Table 5.4. For the Pavani et al. 2012 algorithm, rs11676382 and rs7900194 SNPs in the *GGCX* and *CYP2C9* gene respectively were not included in the following analysis as rs11676382 had <1% minor allele frequency (MAF) and rs7900194 had not been genotyped in this study. The coefficient of determination (R^2) was calculated using the sum of squared errors (SSE) and sum of squared total (SST) from linear regression between the predicted and therapeutic dose using the equation R^2= 1-(SSE/SST). The R^2 value was used to compare the correlation between the predicted of the published algorithms and therapeutic dose within the population since the SST will be same for the population. To compare the predicted dose against therapeutic dose, 53 patients who were receiving therapeutic dose in the current study population were selected for analysis. Furthermore, anonymised clinical and genetic data of 1100 European patients (part of the IWPC) that had reached a stable therapeutic dose and from which Klein et al. 2009 and Gage et al. 2008 pharmacogenetic algorithms were developed, was obtained from PharmGKB website (www.pharmgkb.org) while, data from 121 South Indian patients from which the Pavani et al. 2012 algorithm was developed were provided by the corresponding author for analysis.

Compared to the clinical algorithms, doses predicted with pharmacogenetic algorithms published by Klein et al. 2009 and Gage et al. 2008 had higher correlation with the therapeutic dose in both Gujarati and European populations (Table 5.5). However, dose predicted using the pharmacogenetic algorithm published by Pavani et al. 2012 had a notably lower correlation with the therapeutic dose compared to the other two clinical and the pharmacogenetic algorithms in the Gujarati population (Table 5.5). This algorithm was not tested in the European population since some of the key variables used to predict dose were not available (Table 5.4). It is important to note that the R^2 value can't be used to compare prediction accuracy of the algorithms between populations, as the SST value is different for different populations (Study population=6002.05; IWPC population=311162.57; South Indian=12952.98). This helps explain the relatively lower R^2 values for the algorithms in the Gujarati population compared to the European population, however, the general trend for a higher R^2 value of pharmacogenetic algorithms compared to clinical algorithms remain consistent in both populations.

Variables incorporated	Clinical algorithms		Pharmacogenetic algorithms		
	Gage et al. 2008	Klein et al. 2009	Gage et al. 2008	Klein et al. 2009	Pavani et al. 2012[+]
Age	✓	✓	✓	✓	✓
Height*	✓	✓	✓	✓	✓
Weight*	✓	✓	✓	✓	✓
Race	✓	✓	✓	✓	
Concomitant medication~	✓	✓	✓	✓	
Target INR	✓		✓		
Smoking status	✓		✓		
Indication for treatment	✓		✓		
***CYP2C9* rs1799853**			✓	✓	✓
***CYP2C9* rs1057910**			✓	✓	✓
***VKORC1* rs9923231**			✓	✓	✓
***CYP4F2* rs2108622**					✓
***VKORC1* rs7294**					✓
***VKORC1* rs17708472**					✓
***GGCX* rs11676383**					✓
***CYP2C9* rs75838422**					✓

Table 5.4 Variables incorporated in clinical and pharmacogenetic algorithms

*Height and weight are used to calculate BMI (kg/m^2) for Klein et al. 2009 and Pavani et al. 2012 and body surface area (BSA) (m^2) for Gage et al. 2008 algorithm.

+Gender variable is only incorporated by Pavani et al. 2012 algorithm. The dose predicted is further adjusted for thyroid status.

~Information on amiodarone use only is incorporated in Gage et al. 2008 algorithm whereas, information on amiodarone and enzyme inducers (eg. Carbamazepine, phenytoin, rifampin or rifampicin) is incorporated in Klein et al. 2009 algorithm.

Dosing algorithms		Study population		
		Gujarati Indian (N=53)	IWPC European (N=1100)*	South Indian (N=121)
Clinical only	Klein et al. 2009	4.67%	21.36%	-
	Gage et al. 2008	2.47%	23.12%	-
Pharmaco-genetic	Klein et al. 2009	10.06%	43.19%	-
	Gage et al. 2008	8.20%	49.74%	-
	Pavani et al. 2012	0.04%	-	44.18%

Table 5.5 Coefficient of determination (R^2) for predicting therapeutic dose using clinical and pharmacogenetic algorithms

*881 out of 1100 subjects were used to test Gage et al. 2008 algorithm as subjects with missing variables were dropped from this analysis.

N, Number of subjects

5.2.4 Dose prediction accuracy of algorithms between populations

Mean squared difference (MSD) between the predicted and therapeutic dose was used to compare the dose prediction accuracy of algorithms between populations since the SST was different between populations. Using the Mann-Whitney test for comparing MSD between populations, prediction accuracy of one of the two clinical algorithms was better in Gujarati Indians compared to Europeans (P=0.04 for Klein et al. 2009 algorithm; P=0.22 for Gage et al. 2008 algorithm) (Table 5.6). Prediction accuracy of all three pharmacogenetic algorithms was, however, poorer in the Gujarati Indians, one significantly so (P=0.05 for Klein et al. 2009 algorithm; P=0.002 for Pavani et al. 2012 algorithm; P=0.07 for Gage et al. 2008 algorithm) (Table 5.6). Since half of study subjects in the Gujarati population had mechanical heart valve replacement due to which they had a higher target INR requirement than other subjects, a *post hoc* sub-group analysis of was carried out where MSD for the subjects with the target INR range of 2-3 were compared with the European population. For clinical algorithms, no difference in the MSD between the two populations was observed (Klein et al. 2009 144.08 v/s 222.45, P=0.48; Gage et al. 2008 133.75 v/s 219.50, P=0.33). Similarly for the pharmacogenetic algorithms, no difference in the MSD between the two populations was observed (Klein et al. 2009 193.98 v/s 160.69, P=0.29; Gage et al. 2008 130.78 v/s 143.22, P=0.59). This suggests that the poorer prediction accuracy of the pharmacogenetic algorithms is likely to be associated with genetic variables.

To test that the univariate effect size of the 3 SNP variables (rs1799853, rs1057910 and rs9923231) used in the pharmacogenetic algorithms is similar between current study population and estimate of the effect size from the European population (obtained from Gage et al., 2008), a Wald test was carried out (Gage et al., 2008). No evidence to reject the null hypothesis that the Gujarati population encompassed the effect size reported for the European population was observed (P>0.05 for all 3 SNPs) (Table 5.7). Despite the R^2 of the clinical algorithm being lower than pharmacogenetic algorithm, the MSD of the pharmacogenetic algorithm is higher than the clinical algorithm in the Gujarati population. This conflict in results was due to a few subjects where the squared difference between therapeutic dose and dose predicted with pharmacogenetic algorithm was substantially higher (Figure 5.4).

Overall, the allele frequency result suggests a distinct genetic makeup of the Gujarati population as compared to 3 other populations. Simultaneously, from the dose prediction accuracy results, it can be hypothesized that the SNP panel used in the pharmacogenetic algorithms that are developed from European population do not include SNPs that are specific to genetically distinct populations.

Dosing algorithms		Study population		
		Gujarati Indian (N=53)	IWPC European (N=1100)*	South Indian (N=121)
Clinical only	Klein et al. 2009	119.97	222.45 P=0.04	-
	Gage et al. 2008	129.16	219.50 P=0.22	-
Pharmaco-genetic	Klein et al. 2009	216.34	160.69 P=0.05	-
	Gage et al. 2008	170.64	143.22 P=0.07	-
	Pavani et al. 2012	256.13	-	59.75 P=0.002

Table 5.6 Mean squared difference (MSD) between the predicted and therapeutic dose

P-value for the difference in MSD between the Gujarati Indian and IWPC European or South Indian population was tested using Mann-Whitney test.

*881 out of 1100 subjects were used to test Gage et al. 2008 algorithm as subjects with missing variables were dropped from this analysis.

N, Number of subjects

SNP id	Gujarati Indian		IWPC European		Wald Test P- value
	R^2	Coefficient (95% CI)	R^2	Coefficient (95% CI)	
rs1799853	0.4%	-0.28 (-1.42 to 0.85)	5%	-0.19 (-0.22 to -0.15)	0.88
rs1057910	0.08%	-0.09 (-1.04 to 0.86)	6%	-0.33 (-0.37 to -0.29)	0.62
rs9923231	7.63%	-0.74 (-1.46 to -0.02)	25%	-0.28 (-0.30 to -0.25)	0.21

Table 5.7 Comparison of SNP effect size in a univariate linear regression model

Effect size obtained from univariate linear regression model between the therapeutic dose and the SNP in Gujarati Indian population and effect size estimates from the IWPC European population were compared using the Wald test.

A **B**

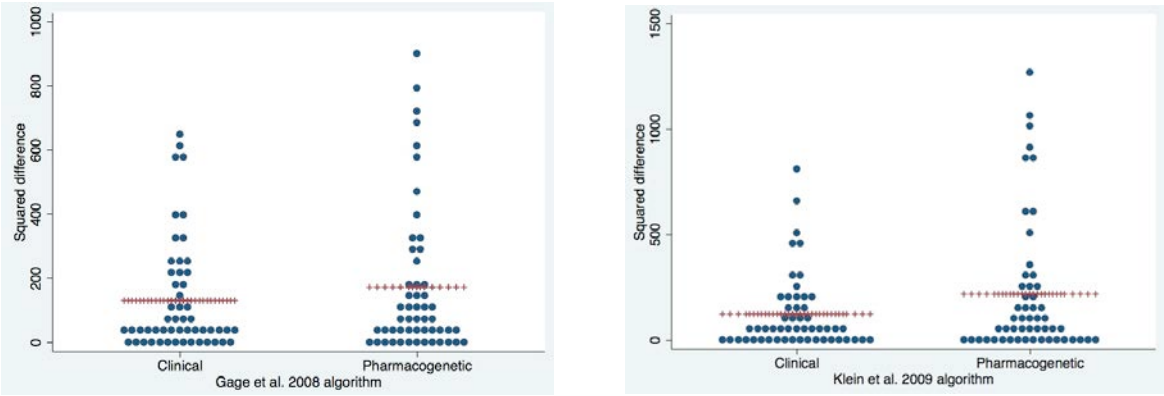


Figure 5.4 Dotplot comparing squared difference between predicted and therapeutic dose using (A) Gage et al. 2008 and (B) Klein et al. 2009 clinical and pharmacogenetic algorithms in the Gujarati population.

Blue dot indicates squared difference for each subject and horizontal red line indicates mean of squared difference.

5.3 Discussion

The current chapter details tests for the clinical utility of published clinical and pharmacogenetic algorithms and compares the allele frequencies of warfarin dose associated SNPs between the Gujarati Indian population and Europeans, Africans and Han Chinese populations. When comparing the dose prediction accuracy of the published algorithms within population, the results show that in both Gujarati Indian and white European populations, the dose prediction accuracy of the pharmacogenetic algorithms was higher than the clinical algorithms, which is in concordance with the results published by Klein et al. 2009 for the white European population (Table 5.5). However, the prediction accuracy of one of the two clinical algorithms was slightly better in Gujarati Indians compared to white Europeans but the prediction accuracy of all three pharmacogenetic algorithms was poorer in the Gujarati Indians compared to white Europeans and South Indians (Table 5.6). In a *post hoc* sub-group analysis where only subjects requiring target INR of 2-3 were included, no significant difference in the prediction accuracy of both clinical and pharmacogenetic algorithms between the Gujarati Indians and Europeans was observed.

In the main analysis, the poorer performance of the pharmacogenetic algorithms in Gujarati Indians could in majority be explained by the presence of the study subjects that require higher warfarin dose due to their target INR range being 2.5-3.5. Since the published clinical and pharmacogenetic algorithms have been designed to predict dose for the target INR range of 2-3, the prediction dose by the algorithms would be underestimated for the

subjects requiring target INR of 2.5-3.5. Furthermore, the poor performance of the pharmacogenetic algorithms in the main analysis could in part be explained by the MAFs of rs1799853 and rs9923231 that were almost half compared to Europeans but were higher than African and Han Chinese populations. This suggests that the 3 SNPs assayed by the pharmacogenetic algorithms have reduced ability to explain dose variance in the Gujarati population. Additionally, allele frequency of the SNPs in the *VKORC1* (rs7294) and *CYP4F2* (rs2108622) had significantly higher MAF in Gujarati Indians compared to Europeans, Africans and Han Chinese populations. This suggests that population specific SNPs have not been accommodated by the European population specific algorithms. Therefore, the SNP panel and dosing algorithms developed from white European populations cannot be assumed to have utility in genetically distinct populations.

5.3.1 Clinical variables and warfarin dose

Overall, 102 patients were recruited on the study who were on warfarin treatment at the time of recruitment out of which only 53 were prescribed therapeutic dose of warfarin. Approximately 50% of the patients prescribed warfarin had mechanical heart valve replacement to treat valvular heart disease. Heart valve disease, which is a downstream consequence of acute rheumatic fever (ARF), is suggested to occur due to an auto-immune response by T cells and macrophages that recognize laminin present in the basement membrane of the heart valves (Carapetis and McDonald, 2005). The high percentage of patients treated for valvular heart disease could be either due to high prevalence of acute rheumatic fever (ARF) in the Indian population (Seckeler and Hoke, 2011) or due to recruitment bias as the recruitment hospital specializes in heart valve replacement.

Sixty nine percent of the amiodarone drug users and 100% of the azole anti-fungal drug users had a warfarin dose requirement of <21 mg/week in the study. Amiodarone is prescribed to patients with cardiac arrhythmias and has been shown to inhibit CYP2C9 enzyme that catalyzes oxidation of S- and R-warfarin thus, reducing warfarin clearance and potentiating the anticoagulation effect of warfarin (Heimark et al., 1992). However, use of amiodarone is included in the IWPC algorithm so should not have affected the findings in the current study. Similarly, both *in vitro* and *in vivo* studies have showed that azole anti-fungal drugs such as fluconazole inhibits CYP2C9 and CYP3A4 enzyme activity by up to 70% and 45% respectively thereby potentiating warfarin's anticoagulation effect (Kunze et al., 1996, Black et al., 1996). Lastly, whilst the average weekly warfarin dose of statin users was higher in comparison with other drug users, 97% of the statin users required <7mg/day of warfarin dose. Statin co-administration has been shown to potentiate warfarin anticoagulation so

patients taking statins require lower warfarin doses (Bellosta et al., 2004) due to competitive inhibition of cytochromes P450 such as CYP3A4 (Corsini et al., 1999).

Whilst these results are in concordance with the existing literature, they should be inferred with caution due to two reasons: First, 49 of 102 patients hadn't reached maintenance dose and therefore the current dose comparison is subjected to change and: second, the comparison was carried out with the other drug users as the reference group rather than no concomitant drug users as the later group was absent in the current study. Since other drug user group consisted of a number of anti-platelets and herbal medications, which can either potentiate or inhibit the effect of warfarin, the group isn't ideal for carrying out dose comparisons. However, the contribution of concomitant medication information to warfarin dose was observed in the Gujarati Indian population when the information was removed from the Klein et al. 2009 dosing algorithm. The coefficient of determination (R^2) reduced from 10.06% to 4.13% after removing co-medication data from the dosing algorithm. This illustrates the importance of incorporating co-medication data in the dosing algorithm for dosing accuracy.

5.3.2 SNP frequency difference between populations

The relative contribution of demographic features such as age, sex, BMI, disease indication and concomitant medication to warfarin dose variability has been shown to be independent of ethnicity (Dang et al., 2005). Comparing MSD of clinical algorithms between the Gujarati Indian and white European population showed no evidence of differential performance of the algorithms between populations (Table 5.6), a finding consistent with the observations of Dang et al. 2005. This suggests that the performance of pharmacogenetic algorithms, which consist of clinical and genetic variables, is non-consistent between populations in part due to genetic differences.

The contribution of genetic factors to warfarin dose variability depends on the distribution of genetic variants in a particular population. This is the first study to report the MAF of SNPs associated with warfarin dose in the Gujarati population and compare them against the frequencies found in Europeans, Africans and Han Chinese. Additionally, the MAF of SNPs reported in the study are in concordance with the MAF reported in the 1000 Genomes database for the Gujarati Indian in Houston (GIH) population thus vindicating the accuracy of the genotype calls reported in the current study and also supporting the use of the genotype data from the database for downstream studies. SNPs included in the study have been shown to affect warfarin's pharmacokinetic and pharmacodynamic pathways. The highest MAF of 68% and 43% was observed for rs7294 and rs2108622 respectively and all 6 SNPs with MAF of >1% had significantly different frequencies compared to 3 other

populations (Table 5.3). This shows differences in *VKORC1* haplotype frequencies between the Gujarati Indian and the other 3 populations.

A study by Lee et al., 2006 showed that the variability in warfarin dose requirement between different ethnicities could be explained by the difference in *VKORC1* haplotype frequencies thus suggesting that the *VKORC1* haplotypes play a vital role in predicting therapeutic warfarin dose accurately (Lee et al., 2006b). In the current study, no evidence for difference in the effect sizes of the 3 SNPs used in pharmacogenetic algorithms between the Gujarati Indian and white European population was observed, however, MSD of the pharmacogenetic algorithms was higher in Gujarati Indians compared to white Europeans suggesting absence of population specific SNPs in the algorithm that would explain higher percentage of dose variance.

Despite having a high R^2 and low MSD value in the South Indian population, the pharmacogenetic algorithm developed by Pavani et al. 2012 had the lowest R^2 value when compared with other algorithms in the Gujarati Indian population and the highest MSD of all algorithms across different populations. The algorithm incorporated more than twice the number of SNPs than the algorithms of Gage et al. 2008 and Klein et al. 2009 but lacked critical clinical variables such as race and concomitant medication. This suggests the algorithm's predicted dose data could be over-fitted to the observed dose data and thus may explain its poor performance in the Gujarati Indian population. However, the algorithm's poor performance due to the genetic differences between population clusters in the Indian population can't be ruled out (Consortium, 2008).

5.3.3 Other variables affecting warfarin dose

Smoking status was only incorporated in the Gage et al. 2008 algorithm and none of the algorithms incorporated vitamin K intake data. In a meta-analysis carried out by Nathisuvan et al., 2011, smoking was associated with a 12.13% (91% CI, 6.99-17.27, P<0.001) increase in warfarin dose compared to non-smokers (Nathisuwan et al., 2011). Similarly, another study showed that patients initiating on warfarin treatment and consuming high amount of vitamin K (>250 ug/day) had lower day 5 INR and needed more warfarin to achieve INR>2.0 compared to normal vitamin K consumers (<250 ug/day) (32.0±9.2 mg/week versus 25.4±6.4 mg/week, P=0.009) (Lubetsky et al., 1999). However, calculating daily vitamin K intake is very difficult and the intake varies by geographic location due to difference in diet. However, since the vitamin K intake influences INR response, incorporating INR response in the pharmacogenetic dosing algorithm after starting warfarin treatment could be beneficial. A dose revision algorithm developed by Lenzini et al. 2010 incorporated INR response on day 4 of warfarin treatment and target INR to revise maintenance dose prediction (Lenzini et al.,

2010). This algorithm together with the initiation dose algorithm developed by Avery et al. 2011 was used in the EU-PACT trial which reported the clinical benefit of using pharmacogenetic based algorithm over standard dosing in anticoagulation control (Pirmohamed et al., 2013, Avery et al., 2011). Based on the evidence, addition of smoking status and INR response to the existing algorithms may potentially improve dosing accuracy.

5.3.4 Study limitation

Whilst the current study is the first to test the clinical utility of know clinical and pharmacogenetic algorithms in the Gujarati Indian population and report genotype frequencies of known warfarin dose associated SNPs, there are several key limitations: First, sample size used in the study to carry out tests for the clinical utility is small. Due to the small sample size, there is a lack of statistical power to observe significant difference in the prediction accuracy of pharmacogenetic algorithms between populations and thus derive a definitive conclusion. Second, half of the patients recruited on the current study had a higher target INR range of 2.5-3.5 since they were being treated for mechanical heart valve replacement. As the published clinical and pharmacogenetic algorithms are derived from the population having target INR of 2-3, they are likely to underestimate the maintenance dose for the subjects having target INR of 2.5-3.5. Results from *post hoc* sub-group analysis indicates that the poor dose prediction of the pharmacogenetic algorithm in the Gujarati population in the main analysis is likely to be due to the presence of subjects having target INR requirement of 2.5-3.5 and thus warrants caution in interpreting results. However, association of poor performance of the algorithms with population specific genetic makeup cannot be excluded since the study lacks statistical power to test the association. With a larger sample size, it is plausible to test the interaction between the SNP, ethnicity and maintenance dose as shown by (Limdi et al., 2015).

5.3.5 Identification of population specific novel SNPs

Results from the current study suggest the need for a downstream study to identify SNPs that can explain dose variability within the Gujarati Indian population. A recent GWAS study carried out in African Americans found a novel rs12777823 genotype that explained 5% of the dose variability whereas the rs1799853 and rs1057910 genotype which explains 5% dose variability in the people with European ancestry explained only 1-2% variability in the African American population (Perera et al., 2013a). SNP rs12777823 has an MAF of 33% in the GIH population in 1000 Genomes database thus it could be hypothesized that genotyping this SNP could help explain higher proportion of warfarin dose variability in the Gujarati Indian population. Since this SNP was not genotyped in the current study, the aforementioned

hypothesis can't be tested but a similar approach to identify relevant genotypes could be conducted in the Gujarati Indian population.

Based on the power calculation where the type 1 error threshold 5×10^{-8} was set, a GWAS study in the Gujarati Indian population would require a minimum of 772 patients on stable warfarin dose to identify a SNP having MAF of 0.05 that explains 5% of warfarin dose variability with 80% power. A similar GWAS approach is needed across genetically distinct clusters of population in India to identify population specific SNPs that can explain more than 5% of warfarin dose variability and a list of SNPs that represent as a genetic signature for an individual which could help in population cluster identification (Consortium, 2008). Development of a pan-India warfarin pharmacogenetic algorithm that incorporates population specific SNPs along with a rapid genotyping platform could help in introducing warfarin pharmacogenetics across the population.

5.4 Conclusion

Results in the current chapter suggest limited utility of the published pharmacogenetic algorithms for the Gujarati Indian population. The poor dose prediction accuracy of the published pharmacogenetic algorithms in the Gujarati Indian population in the main analysis could be explained by the presence of subjects having target INR of 2.5-3.5 and low frequency of the *VKORC1* rs9923231 variant which is the most important genetic determinant of warfarin dosing in Europeans. Furthermore, the absence of variables such as smoking status and INR in the algorithms may help explain their poor prediction accuracy. Therefore, the SNP panel and dosing algorithms developed from European populations cannot be assumed to have utility in current population. The results suggests need for a population specific GWAS study to identify novel genetic markers which can help explain dose variance and development of pharmacogenetic algorithms that can incorporate variation in warfarin dose based on ethnicity. Developing improved algorithms that allow prediction of dose independent of ethnicity by incorporating additional genetic markers is also important. Population specific pharmacogenetic algorithms are needed urgently to allow effective deployment of increasingly cheap and reliable DNA diagnostics.

Chapter 6. (Results 4): Optimisation of hybridization parameters for multiplex SNP genotyping on custom designed microarray platform

6.1 Introduction

6.1.1 Health economics of pharmacogenetic testing

Emergence of *omics* technologies has led to an outpouring of putative biomarkers that could help enhance the effectiveness and safety profile of many commonly prescribed drugs (Wu, 2011). Testing for these biomarkers in routine clinical practice could have a profound benefit to the health of the patients and the economics of healthcare. Selection of a biomarker for clinical utilization is primarily based on its ability to stratify patients to a specific drug by reducing the risk of adverse drug reaction (ADR) and predicting accurate dose (Wu, 2011). One of the best example for pharmacogenetic guided dosing is warfarin where, 2 SNPs (rs1799853 and rs1057910) in the *CYP2C9* gene and 1 SNP (rs9923231) in the *VKORC1* gene along with anthropometric variables such as age, sex, BMI and concomitant medication use explain ~60% of warfarin dose variance (Klein et al., 2009).

Economic analysis conducted by the Brookings Joint Centre for Regulatory Studies, based on assumption of genotyping cost of $350 and an inflated 15 and 50% reduction in bleeding events and stroke respectively, suggested that pharmacogenetic warfarin dosing would lead to savings of up to $2 billion in the US alone per year (McWilliams et al., 2010). In contrast, using a genotyping test costing $400 with 3 days turn around time was concluded not to be cost-effective. However, a sub-analysis indicated a benefit if the genotyping results were delivered in under 24 hours (at less than $200 per test) or if the test is conducted in patients at high risk for hemorrhage (Eckman et al., 2009). A pharmacoeconomic study by Prof. Joyce Hoi-sze You showed that the influential factors for cost-effectiveness were low genotyping cost and fast turn around time, improvement in anticoagulation control, and genotyping patients at high risk of bleeding (You, 2011). Reduction in the cost of genotyping to $47 per test (You et al., 2009) and a subsequent improvement in the time spent in therapeutic range (TTR) to >77% (You et al., 2012) would make genotype guided warfarin dosing cost effective.

6.1.2 Current warfarin SNP genotyping platforms

A study by King *et al.*, 2008 tested 3 molecular methods developed by instrument manufacturers to calculate accuracy and turnaround time to genotype 3 warfarin SNPs. The

INFINITI analyzer, developed by AutoGenomics, is a fully automated microarray platform that uses allele specific primer extension technology and fluorescent labeled PCR product binding to capture probes on a BioFilmChip (Vairavan, 2004). Despite a relatively low number of manual steps in the assay and 100% genotyping accuracy for all 3 SNPs, it takes ~8 hours to genotype and the instrument has a large footprint (King et al., 2008).

The second method, called the Invader assay, uses primary hybridization and cleavage of allele specific primer extension reaction followed by fluorescence resonance energy transfer based secondary signal amplification and detection using the GENios FL fluorescence plate reader (TECAN, Zurich, Switzerland) (Lyamichev et al., 1999). It consists of a relatively low number of manual steps and records ~100% genotyping accuracy for all 3 SNPs but it takes ~3 hours to genotype and requires a large quantity of DNA sample (250ng) (King et al., 2008). The last method, called the Tag-It mutation detection assay, utilizes multiplex PCR, allele specific primer extension and bead hybridization (Strom et al., 2005). It has the largest number of manual steps out of the 3 methods tested, and has an ~8 hour turnaround time but requires least amount of DNA sample (15ng) and provides ~100% genotyping accuracy for all 3 SNPs (King et al., 2008).

The 3 aforementioned methods are designed to be high throughput by genotyping >20 samples per run, restricting their utility for research purpose only since the genotype results in clinics are required on the day patients are initiated on warfarin therapy. To genotype SNPs from blood sample to result in 2 hours, BioAnalytical Innovations at LGC (Teddington, UK) developed HyBeacon probes where the internal nucleotides are tagged with fluorophore moieties and the 3' end consists of a blocker to prevent PCR extension of probes (French et al., 2001). Hybridization of probes to complementary DNA leads to a measurable elevation in probe fluorescence emission and thus can be used to carry out allele discrimination and genotyping using a real time PCR assay. This technology was used by Pirmohamed et al. in their randomized trial (EU-PACT) to genotype 3 warfarin SNPs (rs1799853, rs1057910 and rs9923231) from sample to result in approximately 2 hours after randomization of patients (Pirmohamed et al., 2013). Despite the short turn around time and high genotyping accuracy, the assay requires the reagents to be frozen at -20°C and has a short shelf life thus restricting its utility, particularly in developing countries in some urban hospitals where power cuts are frequent.

6.1.3 QuantuMDx's silicon nanowire platform

QuantuMDx is a Newcastle based biotechnology company partnered with Newcastle University, which is currently developing a handheld, sample to result DNA diagnostic device (www.quantumdx.com). The device consists of 4 components: Mechanical lysis leads to a

DNA extraction cassette, which consists of densely packed sorbent filter with a unique property to bind with proteins and lipids and but not charged nucleotides including DNA; Third, a microfluidic based PCR cassette with two or three heating zones mimicking denaturation, annealing and amplification steps of PCR and; Fourth, a silicon nanowire (SiNW) based field effect transistor (FET) nanosensor for electrical detection and genotyping of oligonucleotides.

Fabrication and electrical characterization of SiNWs have been described previously (Patolsky et al., 2006, Cui and Lieber, 2001). The first functional application of boron-doped SiNWs was shown in 2001 where SiNWs were demonstrated to exhibit pH dependent change in electrical conductance when amine or oxide functionalized surface of the nanowires were protonated and deprotonated (Cui et al., 2001). To demonstrate detection of charged biological species, detection of streptavidin binding to biotin modified SiNWs and reversible binding of antibody to antigen coated SiNW was also carried out (Cui et al., 2001). Due to their small footprint and highly sensitive and selective detection nature, SiNWs offer the capability to carry out label-free and real time detection of charged biological species such as DNA to carry out diagnostics at the point of care.

Li et al. showed detection of label-free DNA using both boron and phosphorous doped SiNWs (Li et al., 2004). They were able to detect complementary single stranded target DNA in sample solution using covalently attached single stranded DNA probes to methoxy silane functionalized SiNWs with a signal to noise ratio of >6. Moreover, a single base mismatch in the target DNA didn't produce signal above the background noise thus demonstrating the potential of the SiNW biosensor for detecting SNPs in DNA. Several other studies produced similar outcomes and showed that the DNA hybridization event can be observed *in situ* and in real time and can reliably detect target DNA at concentrations of 1fM with high specificity to detect SNPs (Gao et al., 2011, Gao et al., 2007).

The results demonstrating SiNW's ability to detect SNPs *in situ* were obtained using custom designed, short length (<50 bp) single stranded target DNA. However, in QuantuMDx's device, the target DNA with a length between 100-150bp would be generated using the microfluidic PCR cassette. Length of target DNA (Gibriel, 2014), along with several other parameters such as probe length (Chou et al., 2004), surface probe density (Peterson et al., 2001), hybridization temperature (Gibriel, 2014), hybridization chamber dimensions (Gibriel, 2014) and washing stringency (Gibriel, 2014) play a critical role in designing genotyping assays which have the capability to carry out multiplex SNP genotyping with high sensitivity and selectivity. As described in the next section, these parameters can be optimized using custom designed DNA microarrays.

6.1.4 Custom designed DNA microarrays

DNA microarray technology has evolved rapidly in the past decade and has enabled whole genome- and transcriptome-based experiments to be conducted in a single run. This technology has been used for several purposes such as microbial detection, SNP genotyping, comparative genomic hybridization (CGH), chromatin immune-precipitation (ChIP) on chip analysis and miRNA detection (Poulsen et al., 2008). Microarrays consist of several probes, which are complementary to a specific sequence in the target nucleic acid, covalently attached to either glass or silicon surface.

Hybridization of complementary nucleic acids in a 3 dimensional medium (eg. solution) has been well studied using the Nearest Neighbor model (Poulsen et al., 2008) but microarray provides 2 dimensional surfaces during hybridization, similar to that on a SiNW platform. Constraints induced by attaching one end of the probe to a solid surface include inability of the probe to diffuse in the hybridization solution, thus reducing the hybridization rate and steric hindrance to the approaching target nucleic acid from the surface (Poulsen et al., 2008). Steric hindrance could be due to the physical constraint of the 2 dimensional surfaces or electrostatic repulsion from the neighboring DNA probes or target nucleic acids. Furthermore, accuracy, sensitivity and specificity of a probe may be affected by several factors such as probe length, probe GC content and target nucleic acid concentration (Koltai and Weingarten-Baror, 2008). Sensitivity is also affected by probe concentration and is dependent upon the availability of free probes to bind with target nucleic acid (Koltai and Weingarten-Baror, 2008). Therefore, characterizing parameters associated with probes and optimization of several hybridization conditions such as hybridization temperature and duration and washing stringency on the microarray platform would enable replication and optimize target DNA hybridization and SNP genotyping on the SiNW platform.

6.1.5 Aims

The current study aimed to identify optimal hybridization and washing conditions for carrying out simultaneous SNP genotyping, understand the interplay between hybridization conditions and specificity of probes towards target DNA sequence, understand dynamics of the microarray based 2 dimensional environments and based on the derived optimal experiment conditions, perform genotyping of anonymised DNA samples. Isothermal melting DNA probes were designed *in silico* to carry out genotyping of 3 warfarin dose associated SNPs: rs1799853 (CYP2C9*2), rs1057910 (CYP2C9*3) and rs9923231 (VKORC1 -1639G>A). Specificity, sensitivity and kinetics of hybridization between probe and target DNA for probe concentration, hybridization temperature, washing stringency and hybridization duration was determined. Optimal condition was identified for each

hybridization condition parameter and was used to carry out genotyping of 2 anonymised samples for 3 SNPs to test the genotyping accuracy of the assay. Understanding the interaction of these parameters with hybridization kinetics, specificity and sensitivity would enable probes to be designed and selected for QuantuMDx's SiNW platform.

6.2 Results

To identify optimal genotyping conditions for all 3 SNPs, several hybridization temperatures and washing stringency strategies were tested followed by testing of 4 hybridization durations to determine the duration at which hybridization between the probe and its complementary target DNA reaches equilibrium thereby providing high target DNA specificity. Each parameter was tested and optimized by genotyping each SNP in a separate miniarray, which allowed for interrogating the association between probes' characteristics and hybridization parameter. Since the target DNA specificity value of 1 or <1 for a probe depicts inability of the probe to genotype the SNP in target DNA accurately, condition of the hybridization parameter at which probes for all 3 SNPs had target DNA specificity >1 was selected.

6.2.1 Assay reproducibility and sensitivity

Initially, PCR step for generating target DNA amplicons of 3 SNPs was optimized. Since the melting temperature of primers for all 3 SNPs and the *ACTB* control sequence was similar, with the lowest melting temperature of the primer being 61°C (Supplementary Table 10), 58°C was selected as the annealing temperature for the PCR reaction for all 3 SNPs and the *ACTB* control sequence. All 4 PCR reactions generated amplicons of the expected size with no non-specific bands or primer dimers (Figure 6.1).

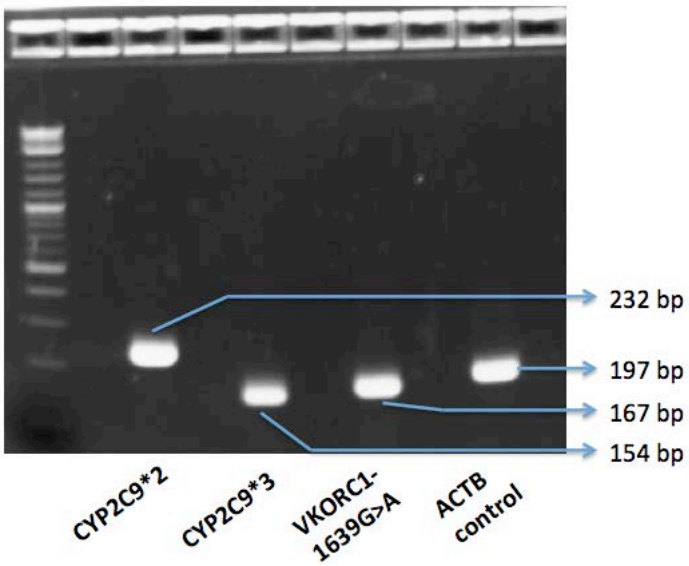


Figure 6.1 PCR to generate amplicons for microarray experiments

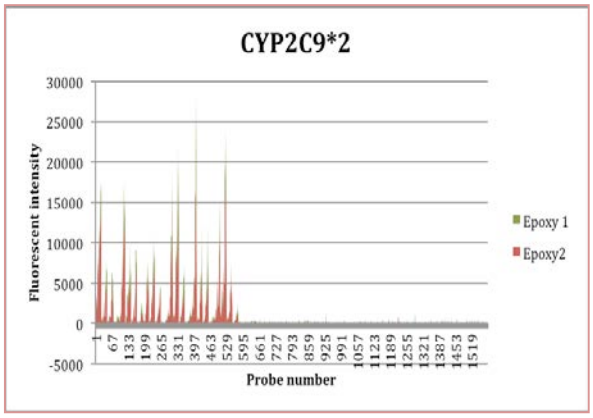
Since only 20 microarray slides were available for testing due to funding constraints, intra- and inter- experimental reproducibility were tested to ensure that the subsequent optimization assays could be carried out without the need of having duplicates for each assay. To test for intra-experimental reproducibility, assays were analysed simultaneously on two separate microarray slides at hybridization temperature of 42°C and subjected to the same washing conditions. Overall, there was a high degree of selectivity for the microarray probes in detecting the target DNA amplicon, with no cross hybridization observed (Figure 6.2). Paired sample t-test for the mean spot intensity for all probes showed a high correlation between the two microarray slides (R^2=0.88, P<0.0001). Furthermore, to test for inter-experimental reproducibility, the same assays on two separate microarray slides were also carried out on consecutive days at 42°C and with the same washing conditions. Similar to the intra-experiment analysis, the assays showed a high degree of reproducibility (R^2=0.84, P<0.0001).

6.2.2 Probe concentration and orientation

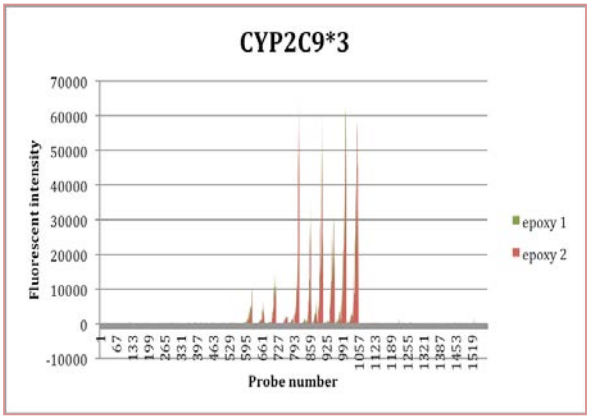
To investigate the relationship between fluorescent intensity and probe concentration, 5 different probe concentrations were tested viz: 0.1, 0.5, 1.0, 5 and 20µM. Figure 6.3 is a typical graph observed for most probes. Between 0.1 to 5µM concentration, an exponential increase in fluorescent intensity with increase in probe concentration is observed but between 5 and 20µM probe concentration the fluorescent intensity plateaus. This suggests that above 20µM, further increase in probe concentration will not make significant difference in the probe fluorescent intensity.

Overall, there were 6 probe sets for *CYP2C9*2*, 5 for *CYP2C9*3* and 4 for *VKORC1*-16939G>A SNP. Each probe set consisted of 1 wild type and 1 mutant probes in sense and anti-sense strand orientation, thus making 4 probes in total for each probe set. Strikingly, probes with high specificity towards target DNA within a probe set of all 3 SNPs showed a strand orientation bias. For example, *CYP2C9*2* probes showed sense strand orientation bias whereas *CYP2C9*3* and *VKORC1*-16939G>A probes showed anti-sense strand orientation bias. However, this observation couldn't be explained in the current study.

A



B



C

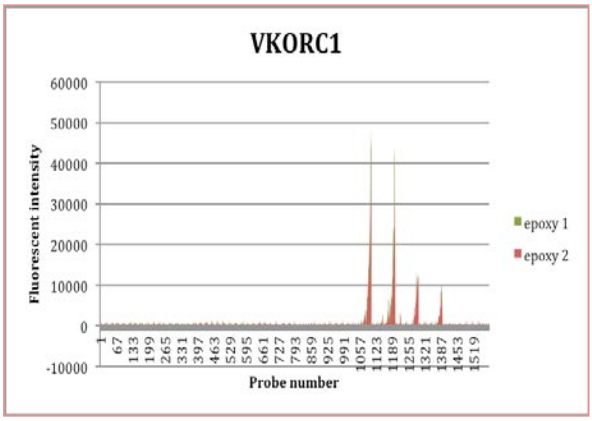


Figure 6.2 Microarray intra-experimental reproducibility and probe sensitivity to target DNA.

Mean fluorescent intensity of all probes within a miniarray that is hybridized with a specific target DNA is represented on a single graph. Each graph consists of data obtained from the 2 microarray slides called epoxy 1 and epoxy 2. (A) Probe 1 to 576 are designed to bind to *CYP2C9*2* amplicon only. (B) Probe 577 to 1056 are designed to bind to *CYP2C9*3* amplicon only. (C) Probe 1057 to 1512 are designed to bind to *VKORC1*-1639G>A amplicon only. Probes show high sensitivity towards their complementary DNA and no non-specific hybridization is observed.

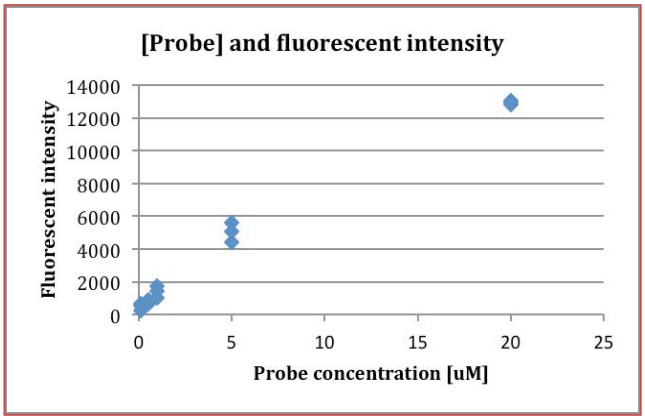


Figure 6.3 Relationship between probe surface concentration and fluorescent intensity post hybridization

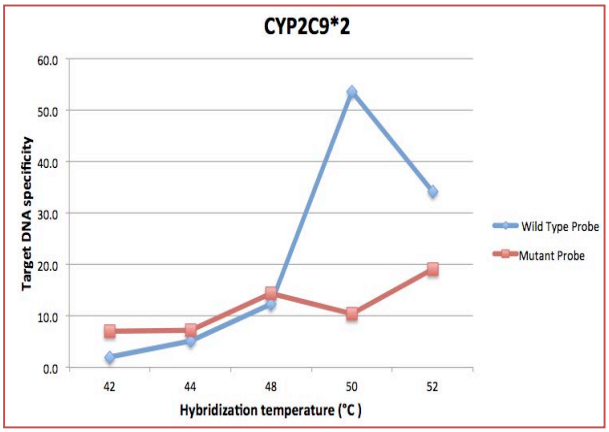
Probe fluorescence at all 5 concentrations viz: 0.1, 0.5, 1, 5 and 20µM. Each probe is spotted in triplicate at all 5 concentrations.

6.2.3 Optimising hybridization temperature

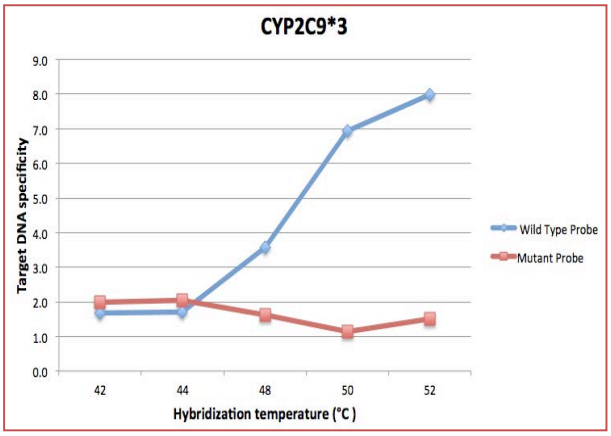
In the current study, 5 hybridization temperatures viz: 42, 44, 48, 50 and 52°C were tested to identify optimal hybridization temperature at which high specificity is achieved without losing fluorescent signal intensity. Selection of the hybridization temperatures was based on the study by (Gresham et al., 2010).

Overall, an increase in fluorescent signal intensity was observed between 42 and 44°C hybridization temperature but a rapid fall in the intensity was observed at hybridization temperatures above 44°C for all 3 SNP probes. For *CYP2C9*2* probes, specificity increased between 42 and 50°C hybridization temperature for the wild-type probes, whereas the mutant probes also showed a trend for an increase in specificity with increase in hybridization temperature (Figure 6.4 A). Similar to the *CYP2C9*2* wild-type probes, *CYP2C9*3* wild-type probes showed an increase in specificity with increase in temperature. In contrast however, *CYP2C9*3* mutant probes showed a trend towards an inverse relationship between specificity and hybridization temperature (Figure 6.4 B). Lastly, for the *VKORC1*-1639G>A probes, highest specificity was observed at 44°C and 42°C for wild-type and mutant probes respectively (Figure 6.4 C). For all 3 SNPs, wild-type probes had a higher specificity compared to mutant probes suggesting a higher affinity of the wild-type probe to its complementary target DNA compared to the mutant probes. Based on the relationship between hybridization temperature and specificity of probes for all 3 SNPs, 48°C was selected as the optimal hybridization temperature since above 48°C, *CYP2C9*3* probes have very low specificity where below 48°C, *CYP2C9*2* probes have a low specificity.

A



B



C

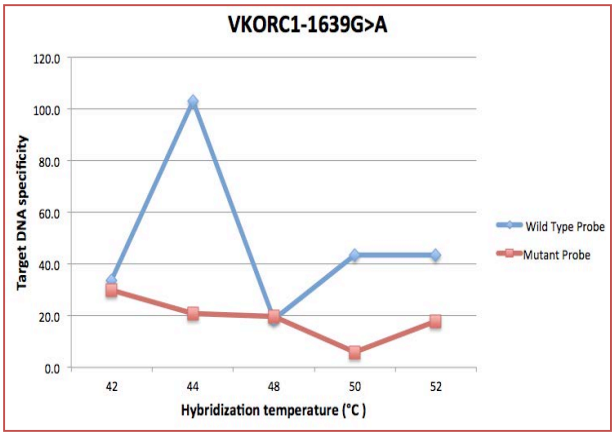


Figure 6.4 Effect of hybridization temperature on probe's specificity towards target DNA

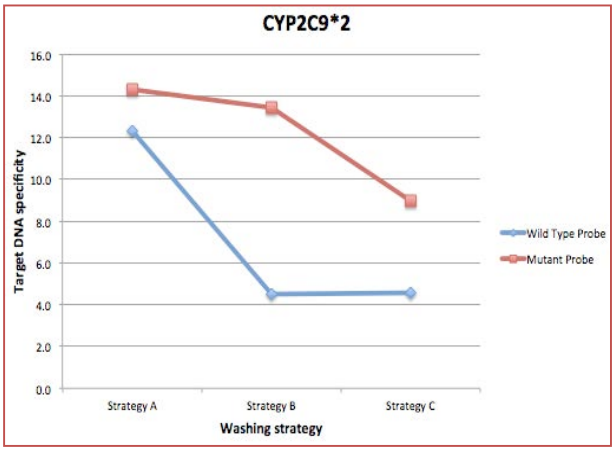
Graphs represent average specificity of 20µM wild-type (in blue) and mutant probes (in pink) to wild-type and mutant target DNA amplicon respectively across 5 hybridization temperatures.

6.2.4 Optimising post-hybridization washing

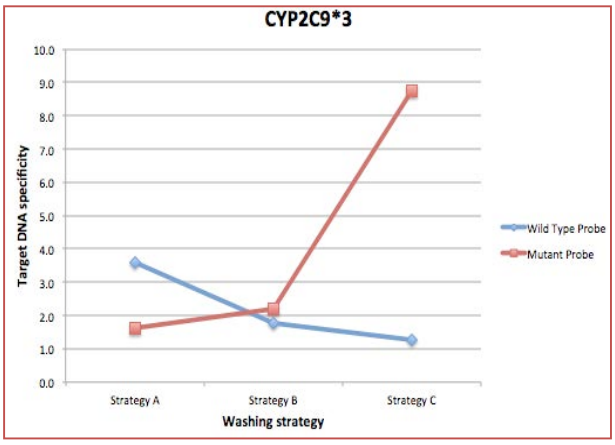
Following optimization of hybridization temperature at 48°C, the effect of post-hybridization washing on probe specificity was tested using three washing strategies: (A) first 2 washes with 2x SSC buffers followed by 0.2x SSC buffer; (B) first wash with 2x SSC, second wash with 1x SSC and last wash with 0.2x SSC buffer and; (C) first wash with 2x SSC, 2nd wash with 1x SSC and last wash with 0.1x SSC buffer. The strategies were designed to test if the increasing in wash buffer stringency would increase probe specificity.

For the *CYP2C9*2* wild-type and mutant probes, target DNA specificity was highest when the least stringent washing strategy was used (Figure 6.5 A). Similar to the *CYP2C9*2* wild-type probes, *CYP2C9*3* wild-type probes had the highest specificity when the least stringent washing strategy was used however in contrast, the mutant probes had the highest specificity when the most stringent washing strategy was used (Figure 6.5 B). This contrast in the effect of washing stringency on probe's specificity could partially be explained by the fact that the mutant probe has G instead of T at SNP position, thus increasing GC content of the probe. Lastly, for the *VKORC1*-1639G>A probes, both wild-type and mutant probes had the highest target DNA specificity when the most stringent washing condition was applied (Figure 6.5 C). Although specificity of *VKORC1*-1639G>A probes at the least stringent washing strategy was higher than the *CYP2C9*2* and *CYP2C9*3* probes at their optimal washing strategy. Thus washing strategy A that includes first 2 washes with 2x SSC buffers followed by 0.2x SSC buffer was concluded as the optimal washing strategy, although, the effect of an increase in washing temperature was not tested in the current study.

A



B



C

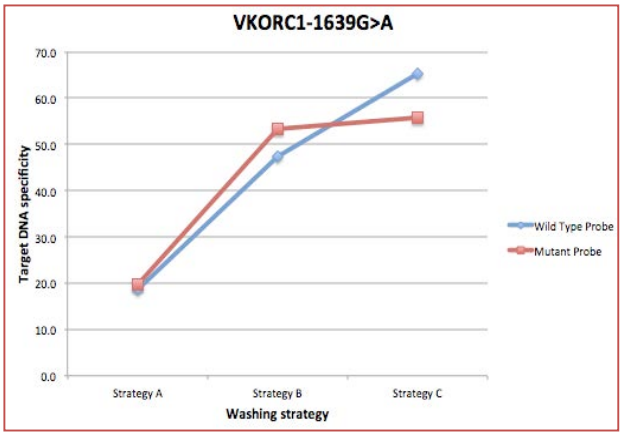


Figure 6.5 Effect of washing buffer stringency on probe specificity towards target DNA

Graphs represent average specificity of 20µM wild-type (in blue) and mutant probes (in pink) to wild-type and mutant target DNA amplicon respectively across 3 washing strategies: (A) Two washes with 2x SSC buffer followed by 0.2x SSC buffer, (B) Washing with 2x SSC, 1x SSC and 0.2x SSC buffers and (C) Washing with 2x SSC, 1x SSC and 0.1x SSC buffers.

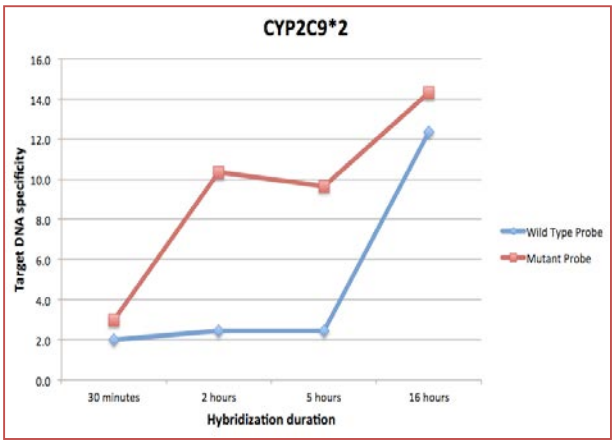
6.2.5 Optimising hybridization duration when genotyping a single SNP

A total of 4 hybridization durations were tested in the current study: 30 minutes, 2 hours, 5 hours and 16 hours. The aim of the test was to identify optimal hybridization duration at which the hybridization reaction reached equilibrium, which is the point at which the rate of association and dissociation between target DNA and probe is equal thus obtaining high specificity. The experiment was carried out using optimal hybridization temperature of 48°C and washing strategy where the first 2 washes were with 2x SSC buffers followed by 0.2x SSC buffer, and the results are shown in Figure 6.6 and Figure 6.7.

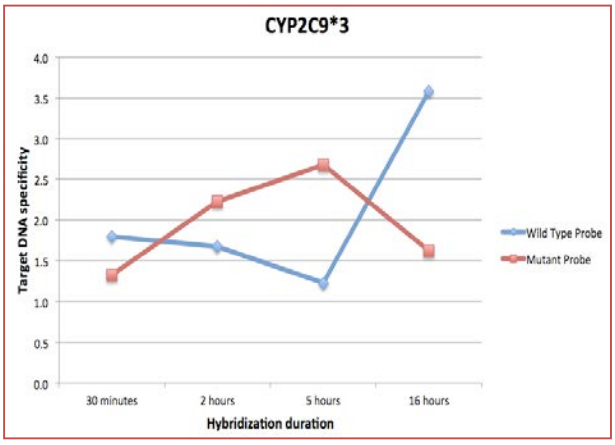
When the probe surface concentration was 20µM, *CYP2C9*2* wild-type and mutant probes had the highest specificity at 16 hours, *CYP2C9*3* wild-type and mutant probes had highest specificity at 16 and 5 hours respectively and *VKORC1*-1639G>A wild-type and mutant probes had highest specificity at 2 hours (Figure 6.6). It is to be noted that even after 30 minutes of hybridization duration, *CYP2C9*2* and *VKORC1*-1639G>A probes showed high specificity whereas, *CYP2C9*3* probes took longer duration to exhibit the level of specificity that could be utilized for genotyping. Since the hybridization kinetics is also determined by probe surface concentration, specificity of probes with surface concentration of 5 µM was also determined. Wild-type and mutant probes for all 3 SNPs having 5µM surface concentration showed similar trend and level of specificity in relation to hybridization duration compared to probes with 20µM surface concentration (Figure 6.7). This suggests that the hybridization kinetics of probes with surface concentration of 5µM is similar to that of 20µM and is in concordance with the fluorescent intensity results shown in Figure 6.3.

To rule out the specificity results at different hybridization durations being the result of probe spotting error, post-hybridization spot morphology was observed during the scanning stage. Figure 6.8 represents a typical spot morphology for the probes spotted at 20µM concentration. At 30 minutes and 2 hours, uneven fluorescent intensity is observed across the spot whereas at 5 and 16 hours, a uniform fluorescent intensity is observed across the spot. The observation provides evidence for no spotting error and suggests that there are various stages of hybridization kinetics that are dependent upon the duration of hybridization.

A



B



C

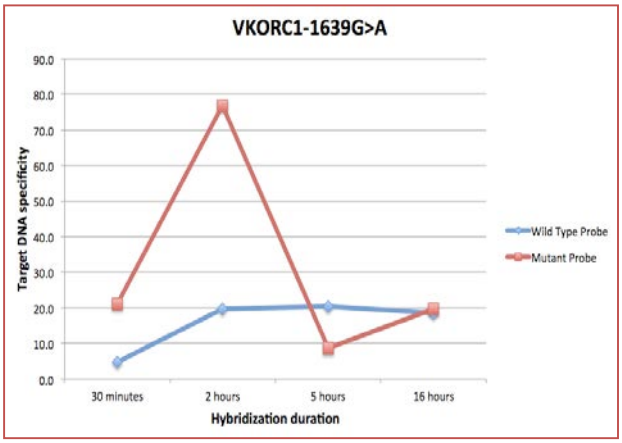
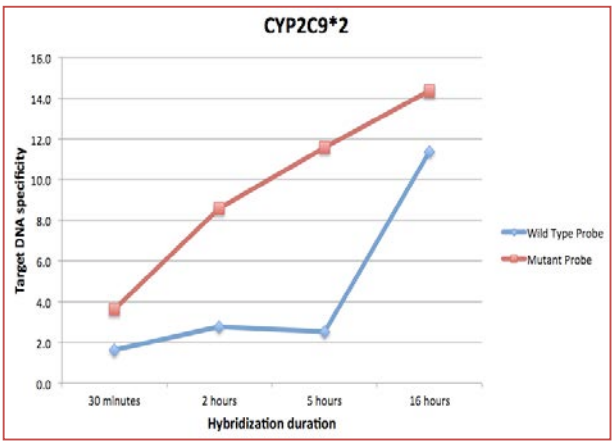


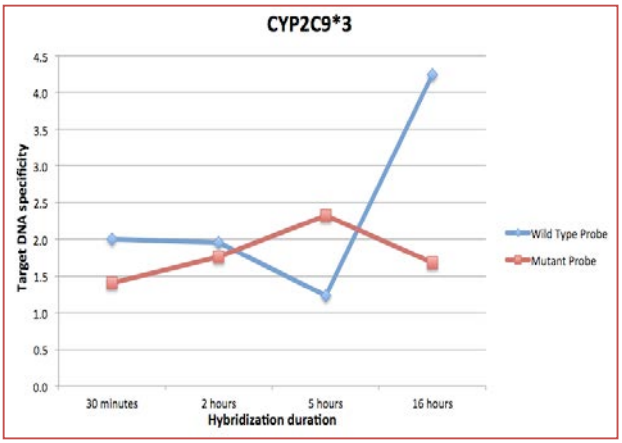
Figure 6.6 Effect of hybridization duration on probe specificity towards target DNA for 20µM surface probe concentration

Graphs represent average specificity of 20µM wild-type (in blue) and mutant probes (in pink) to wild-type and mutant target DNA amplicon respectively across 4 hybridization durations: 30 minutes, 2 hours, 5 hours and 16 hours.

A



B



C

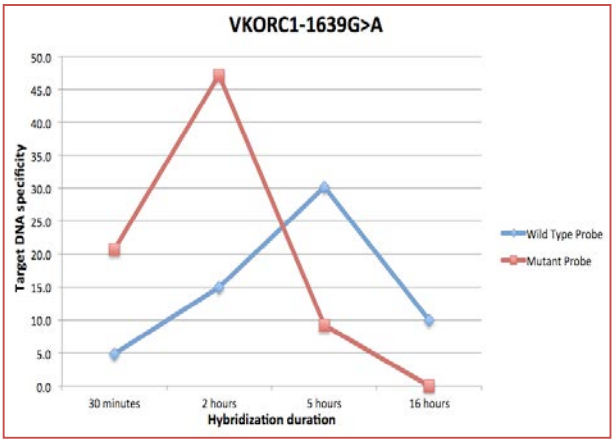


Figure 6.7 Effect of hybridization duration on probe specificity towards target DNA for 5µM surface probe concentration

Graphs represent average specificity of 5µM wild-type (in blue) and mutant probes (in pink) to wild-type and mutant target DNA amplicon respectively across 4 hybridization durations: 30 minutes, 2 hours, 5 hours and 16 hours.

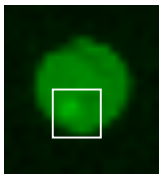
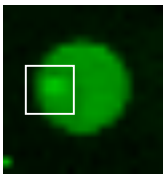
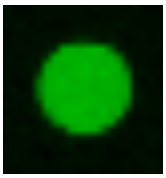
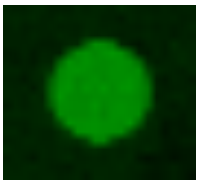
Hybridization Duration	30 minutes	2 hours	5 hours	16 hours
Har_017 (20uM) spot				

Figure 6.8 Effect of hybridization duration on probe spot morphology

Image of 20µM spot of Har_017_051011_CYP2C9_R144C_SENSE_WT probe hybridized with *CYP2C9*2* wild-type target DNA for different hybridization duration. White square for 30 minutes and 2 hours spot indicates region of uneven hybridization.

Thus, probe specificity data coupled with spot morphology information suggests that reliable specificity and fluorescent intensity could be achieved when the hybridization duration is set between 2 and 5 hours for the current set of probes. However, optimization of hybridization duration was carried out using a single target DNA per miniarray to genotype a single SNP at a time. The optimal hybridization duration may vary when multiple target DNAs are added simultaneously in a miniarray to genotype multiple SNPs.

6.2.6 Testing optimal hyrbidization conditions for genotyping multiple SNPs

Since genotyping of all 3 SNPs on the SiNW will be carried out simultaneously in a multiplex format, testing of optimal hybridization conditions for genotyping multiple SNPs on the microarray platform was carried out. Furthermore, since the concentration of PCR amplicons could vary between samples on the SiNW platform, 8 different amplicon concentration combinations were tested on the microarray platform to test for the robustness of probe specificity (Table 6.1). Two microarray slides were used in this experiment where on one slide, homozygous wild-type target DNA and on another slide homozygous mutant target DNA was tested. All optimal hybridization conditions except hybridization duration of 16 hours were used in the experiment to allow hybridization of multiplex amplicons to their respective probes.

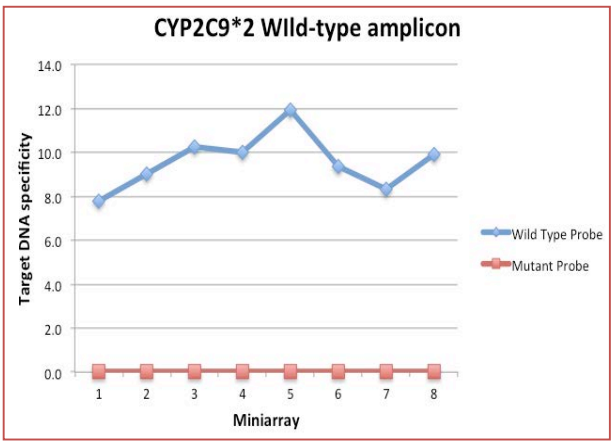
Overall, probes for all 3 SNPs were able to detect homozygous wild-type and mutant target DNA at various concentrations of target DNA amplicon in multiplex format (Figure 6.9). Maximum variation of specificity was observed for the *VKORC1*-1639G>A probes in relation to the target DNA amplicon concentration but the same probes showed highest degree of specificity compared to the probes of *CYP2C9*2* and *CYP2C9*3*.

To test the specificity of probes for a heterozygous target DNA sequence, an experiment similar to the one carried out for homozygous target DNA sequence was carried out in the same hybridization and washing conditions with a range of target DNA amplicon concentrations in multiplex format as shown in Table 6.1. For *CYP2C9*2* and *CYP2C9*3*, both wild-type and mutant probes had specificity of ~1 thus suggesting similar quantity of hybridization of wild-type and mutant probes to their complementary single stranded target DNA (Figure 6.10 A and B). However for *VKORC1*-1639G>A, wild-type probes had a substantially higher specificity than mutant probes (Figure 6.10 C) which could either be due to the target DNA being prepared by mixing equimolar quantity of homozygous wild-type and mutant target DNA which could have led to a pipetting error or possibility of hybridization bias towards wild-type probes as they have a higher GC content compared to mutant probes. In total, the results indicate the ability of the probes of all 3 SNPs to identify and genotype multiplex target DNA sequence in optimal hybridization conditions.

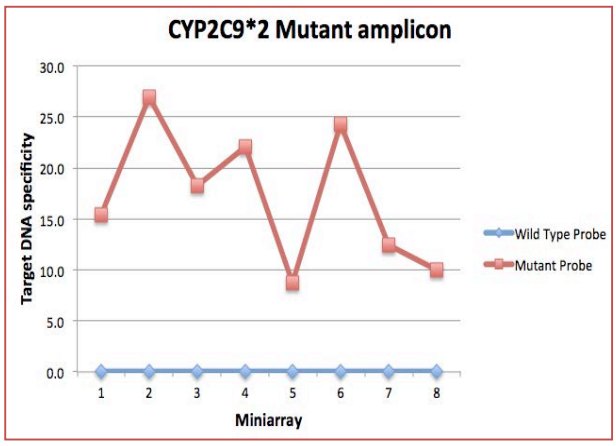
Target DNA	**Miniarrays: Concentration of amplicons (x 10nM)**							
Amplicon	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
CYP2C9*2	1	1	0.5	0.5	1	1.5	1.5	1
CYP2C9*3	1	0.5	1	0.5	1.5	1	1.5	1
VKORC1-1639G>A	1	0.5	0.5	1	1.5	1.5	1	2

Table 6.1 Concentration (nM) of target DNA amplicons of *CYP2C9*2, CYP2C9*3* and *VKORC1* -1639G>A SNP added in 8 miniarrays on a microarray slide

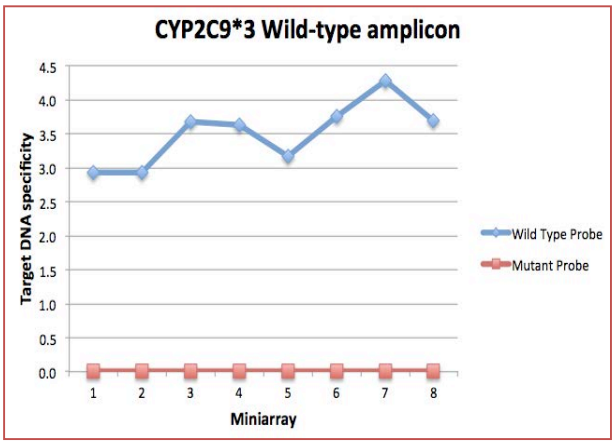
A



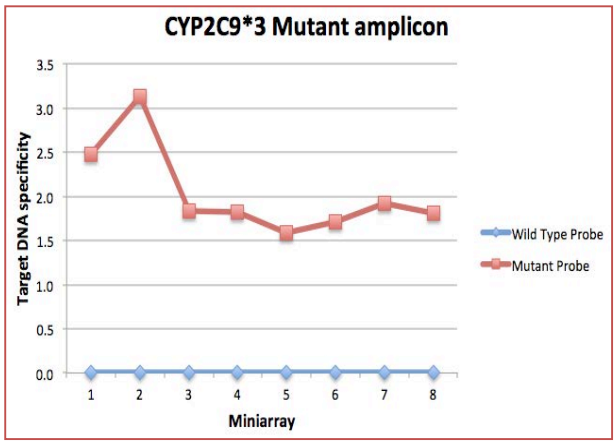
B



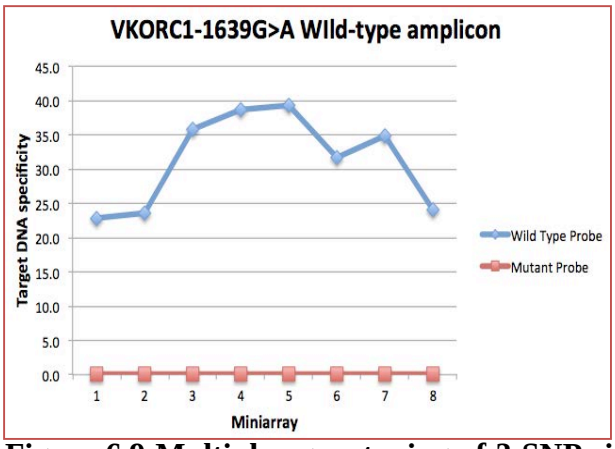
C



D



E



F

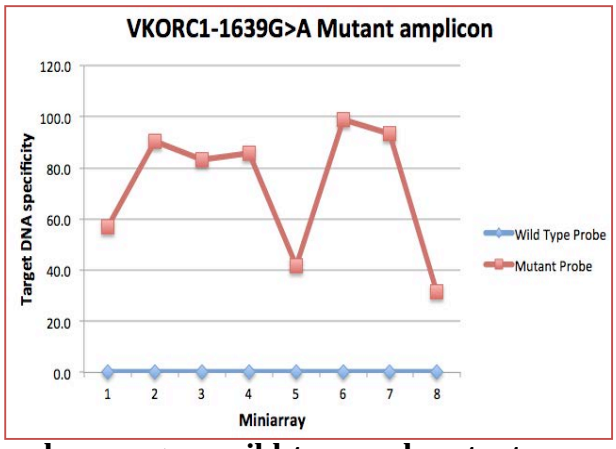
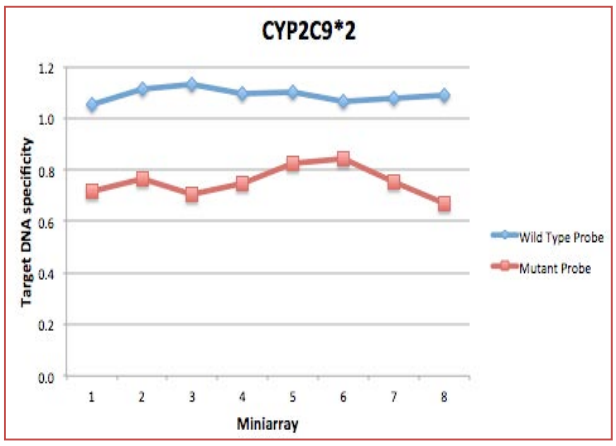


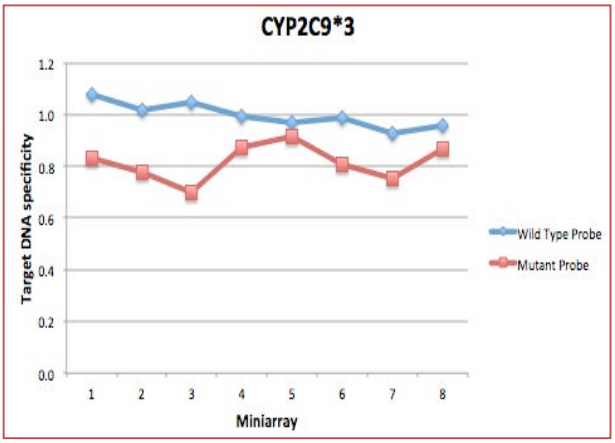
Figure 6.9 Multiplex genotyping of 3 SNPs in a homozygous wild-type and mutant DNA sample

Graphs represent average specificity of wild-type (in blue) and mutant (in pink) probes. Graphs A and B represent genotyping of wild-type and mutant sample for *CYP2C9*2* SNP respectively, C and D represents genotyping of wild-type and mutant sample for *CYP2C9*3* SNP respectively and E and F represent genotyping of wild-type and mutant sample for *VKORC1*-1639G>A SNP respectively.

A



B



C

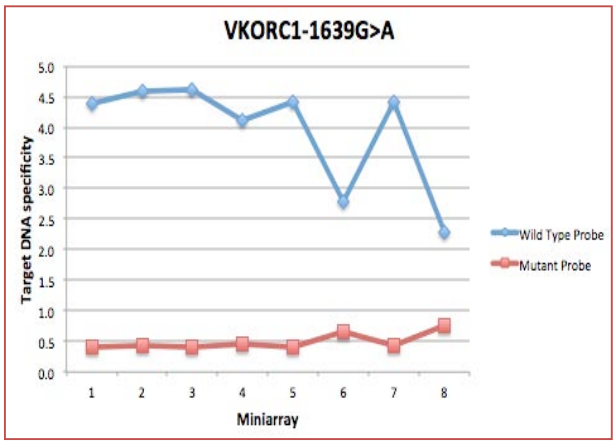


Figure 6.10 Multiplex genotyping of 3 SNPs in a heterozygous sample

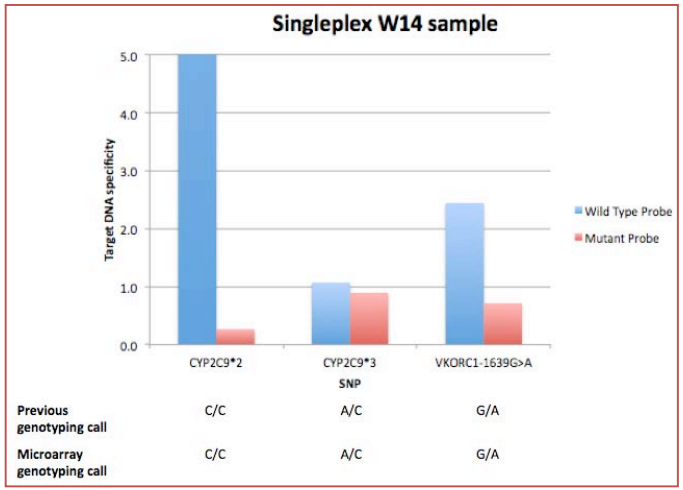
Graphs represent average specificity of wild-type (in blue) and mutant (in pink) probes across 8 miniarrays.

6.2.7 Genotyping of DNA samples

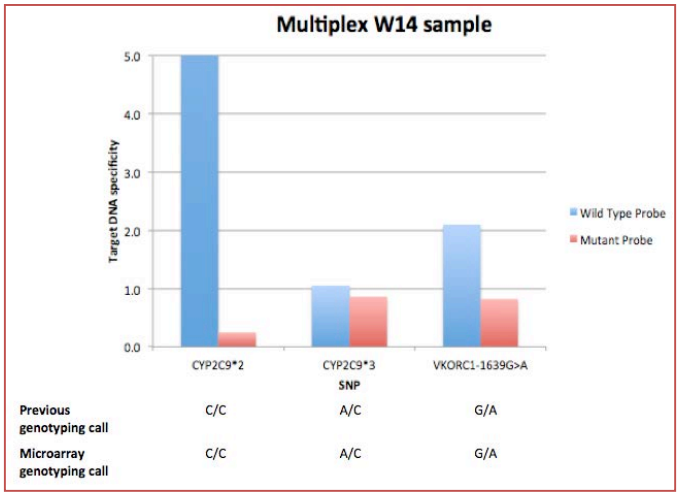
Lastly, to test the genotyping accuracy of the probes using the microarray platform and optimal hybridization and washing conditions, 2 previously genotyped DNA samples- W14 and FRW62 were randomly chosen to be genotyped for all 3 SNPs. Probe specificity was tested using 10nM singleplex and 10nM equimolar multiplex target DNA amplicons for both DNA samples.

Using the specificity as an indicator for the genotype call, sample W14 was homozygous wild-type for *CYP2C9*2* and heterozygous mutant for *CYP2C9*3* and *VKORC1*-1639G>A SNP (Figure 6.11 A and B). For sample FRW62, the specificity indicated the sample to be heterozygous mutant for *CYP2C9*2* and *CYP2C9*3* but homozygous mutant for *VKORC1*-1639G>A SNP (Figure 6.11 C and D). For both samples, target DNA specificity data for both wild-type and mutant probes for all 3 SNPs corroborated well for singleplex and multiplex target DNA format. Since these samples were previously genotyped for all 3 SNPs in another study, the genotype call from the microarray platform was compared against the genotype data for the samples from the previous study. Genotype calls for the sample W14 corroborated perfectly with genotype call from the previous study, but for sample FRW62 the genotype call for *VKORC1*-1639G>A didn't match with the genotype call made in the previous study. Subsequent bi-directional Sanger sequencing confirmed that the genotype call for *VKORC1*-1639G>A SNP in FRW62 using this microarray platform was correct (Figure 6.12). This suggests either an error in the genotype call made in the previous study or sample mix-up. Irrespective of the underlying reason for this discrepancy, probes designed for genotyping 3 warfarin dose associated SNPs along with optimized hybridization and washing conditions were able to genotype samples accurately.

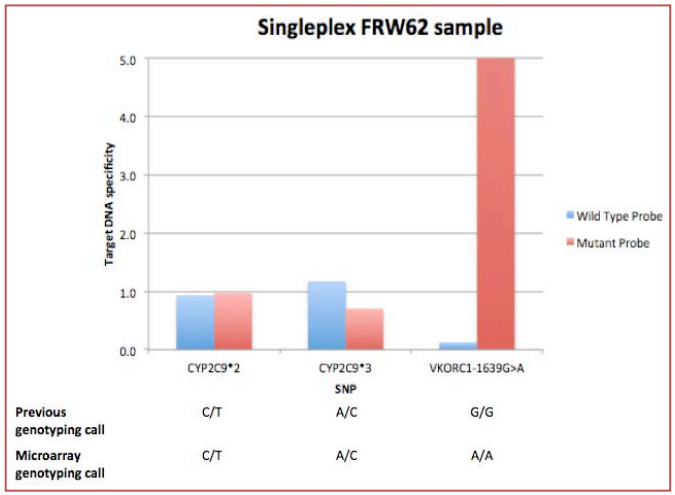
A



B



C



D

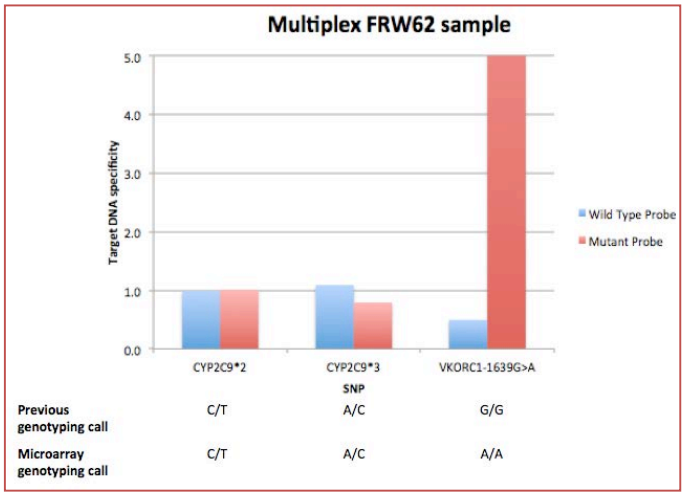


Figure 6.11 Genotyping of 3 SNPs in W14 and FRW62 DNA samples

Graphs represent average specificity of wild-type (in blue) and mutant probes (in pink) of *CYP2C9*2*, *CYP2C9*3* and *VKORC1*-1639G>A SNPs in singleplex (A and C) and multiplex (B and D) target DNA amplicons.

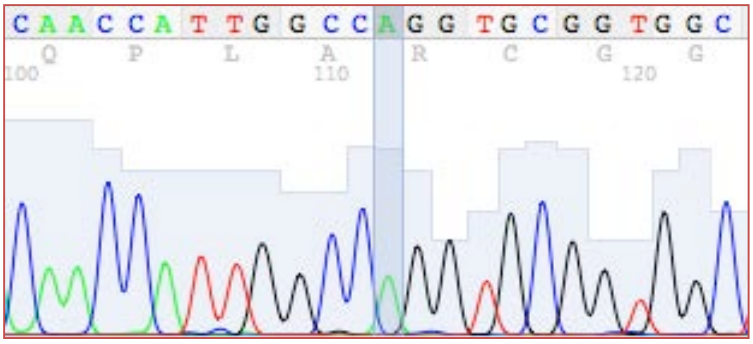


Figure 6.12 Sanger sequencing chromatograph for *VKORC1*-1639G>A SNP amplicon from FRW62 DNA sample

Highlighted peak represent *VKORC1*-1639G>A SNP position in the chromatograph

6.3 Discussion

Identification of putative biomarkers for commonly prescribed drugs has led to an increase in demand for point of care companion diagnostics that can genotype the biomarkers rapidly and cheaply thus improving safety and effectiveness of the drug. Implementation of these biomarkers in routine clinical practice has been suggested to reduce cost to the healthcare industry and improve quality of life for patients (McWilliams et al., 2010). Warfarin, which is one of the most commonly prescribed drugs, is an oral anticoagulant that is used to manage thromboembolic diseases. Two SNPs in the *CYP2C9* gene (*CYP2C9*2* and *CYP2C9*3*) and a SNP in the *VKORC1* gene (*VKORC1*-1639G>A) have been previously shown to affect warfarin dose (Pavani et al., 2011, McWilliams et al., 2010, Wang et al., 2008, Gage et al., 2008, Yuan et al., 2005, D'Andrea et al., 2005, Aithal et al., 1999). Furthermore, a randomized clinical trial (EU-PACT trial) showed that patients receiving warfarin dose based on their genotype for the 3 SNPs spent a higher percentage of mean time in the therapeutic INR range (67.4%) compared to controls (60.3%) during the initiation of warfarin therapy (Adjusted difference, 7.0 percentage points; 95% CI, 3.3 to 10.6; $p<0.001$) (Pirmohamed et al., 2013). However, there is no point of care test available to genotype these SNPs in a turn around time of 20 minutes at low cost thus inhibiting the implementation of genotype-guided warfarin dosing in clinical practice.

QuantuMDx's SiNW based platform is aiming to genotype SNPs in less than 20 minutes for around £20. One of the genotyping strategies that are under development uses differential hybridization kinetics between the probe attached to the silicon surface and free floating complementary or non-complementary target DNA. However, to design probes that have a high specificity towards complementary target DNA, the effect of various hybridization conditions on hybridization kinetics of the probe has to be studied. The aim of

the current section of work was to design a set of probes for the 3 warfarin dose-associated SNPs and characterize the effect of hybridization conditions on their target DNA specificity by carrying out the test on a custom-designed microarray slide. Characterization of the probe's specificity will help elucidate optimal hybridization conditions where probes would have highest specificity to complementary target DNA. Furthermore, these would help guide the testing of probe specificity on the SiNW platform.

6.3.1 Hybridization temperature and probe specificity

The current study tested 5 hybridization temperatures in order to identify the optimal temperature at which probes for all 3 SNPs have optimal signal intensity and target DNA specificity. Probes for *CYP2C9*2* and *CYP2C9*3* SNPs showed higher specificity at hybridization temperatures above 48°C whereas probes for the *VKORC1*-1639G>A SNP had a higher specificity below 48°C (Figure 6.4). An association between probe length and hybridization temperature was observed since the probe length for *CYP2C9*2*, *CYP2C9*3* and *VKORC1*-1639G>A SNP probes was 20, 22 and 16 respectively. This association has been reported in previous studies where longer probes had higher signal intensity compared to shorter probes resulting in higher specificity (Letowski et al., 2004, Chou et al., 2004). The association has previously been explained using the thermodynamics of duplex formation where the free energy of duplex formation (ΔG°) is estimated using the Nearest Neighbor model that assumes that the stability of a given nucleotide pair is dependent upon the identity and orientation of the neighboring nucleotide pair (Gresham et al., 2010). As enthalpy (ΔH°), which is the measure of total energy of a thermodynamic system and entropy (ΔS°) which is the measure of energy available for useful work within a thermodynamic system has been empirically determined for all 10 nearest neighbors nucleotide pairs, ΔG° of a duplex can be determined using the equation $\Delta G° = \Delta H° - T\Delta S°$, where T is the melting temperature of the duplex (Gresham et al., 2010). Since mismatch destabilizes duplex formation, T for the mismatch duplex is lower than that of the perfectly matched duplex. Therefore, the $\Delta G°_{PM}$ of a perfectly matched duplex is significantly lower than the $\Delta G°_{MM}$ of a duplex with a mis-match at a given hybridization temperature (ΔS°). Furthermore, it also explains that with the increase in probe length, ΔG° between the probe and non-complementary target DNA decreases thus enabling increase in hybridization temperature to increase ΔG° for mis-match duplex and thus increasing target DNA specificity. However, it also suggests that shorter probes have higher specificity which explains higher target DNA specificity of the *VKORC1*-1639G>A probes compared to that of *CYP2C9*2* and *CYP2C9*3* (Figure 6.4). Poulsen et al. observed a general decrease in specificity with increase in probe length, which corroborates with the aforementioned mathematical hypothesis (Poulsen et al., 2008).

6.3.2 Washing stringency and probe specificity

Three different washing stringency strategies were tested in the current study to identify optimal washing strategy that would provide highest specificity for the probes from all 3 SNPs. Probes for *CYP2C9*2* and *CYP2C9*3* SNPs showed higher specificity when the least stringent washing strategy A was used whereas, probes for the *VKORC1*-1639G>A SNP had the highest specificity when most stringent washing strategy was applied (Figure 6.5). Since probes for the *VKORC1*-1639G>A SNP had the highest GC content compared to the probes of *CYP2C9*2* and *CYP2C9*3* SNPs, an association between probe's GC content and washing stringency was observed whereby, probes with higher GC content had higher specificity with increasingly stringent washing strategy. Since GC rich probes form more hydrogen bonds compared to the AT rich probes when hybridizing with the target DNA molecule, a more stringent washing strategy would be required to disassociate non-complementary target DNA from the probe to improve probe's specificity. Poulsen et al. reported a similar observation where AT rich probes had lower signal intensity compared to GC rich probes when wash buffer stringency was increased which resulted in lower specificity (Poulsen et al., 2008). In the current study, the least stringent washing strategy provided optimal specificity for the probes of all 3 SNPs.

6.3.3 Hybridization duration and kinetics

Since hybridization duration plays a critical role in affecting probe's specificity, four different hybridization durations were tested to identify the duration at which probes from all 3 SNPs showed high specificity. Compared to the probes of *CYP2C9*2* and *CYP2C9*3* SNPs, probes of *VKORC1*-1639G>A SNP reached high specificity at shorter hybridization duration (Figure 6.6 and Figure 6.7). This suggests that probes with short length and high GC content require shorter hybridization duration to achieve high specificity compared to the probes with long length and AT rich sequence.

The rate of hybridization and disassociation of perfectly matched (PM) and mis-matched (MM) target DNA dictates hybridization kinetics and thus hybridization duration. Hybridization kinetics can be mathematically represented by the formula $R + L \rightleftarrows C$ where R, L and C denotes number of free probes available for hybridization, number of free target DNA molecules (in Molar) and number of bound probe-target DNA complexes respectively (Dai et al., 2002). Assume K_f and K_r to be association ($Molar^{-1}\,time^{-1}$) and dissociation rates ($time^{-1}$) respectively. Now assuming that the K_f is same for PM and MM target DNA but the dissociation rate of PM is less than MM target DNA ($K_{rPM} << K_{rMM}$), mathematical representation of temporal based hybridization kinetics would be $\tau_{L0} = 1/ K_f (L_0 + K_D)$ where, L_0 is the concentration of probes at time=0 of the experiment, τ is the time at which the

hybridization reaction has reached equilibrium and $K_D=K_r/K_f$ (Dai et al., 2002). Since $K_{rPM}<<K_{rMM}$, $K_{D\text{-}PM}<<K_{D\text{-}MM}$ and thus $\tau_{PM}>>\tau_{MM}$. This represents the temporal aspect of hybridization kinetics and explains the requirement for a longer hybridization duration to achieve high signal intensity and target DNA specificity when genotyping many SNPs.

The mathematical representation of temporal based hybridization kinetics also explains the post-hybridization spot morphology observed at hybridization duration of 30 minutes and 2 hours for all 3 SNPs (Figure 6.8). At short hybridization durations, K_f is same for PM and MM duplex with the probe but since $K_{rPM}<<K_{rMM}$, the number of free probes available for hybridization are high. Stopping the hybridization reaction before it has reached equilibrium coupled with high probe surface density that would be creating electrostatic hindrance to forming new probe-target DNA duplex may be leading to formation of uneven spot morphologies at short hybridization durations. However, in the current study, uneven spot morphology was not detected at 5 and 16 hours of hybridization durations suggesting the hybridization reaction reaching equilibrium. In the current study, 16 hours of hybridization duration was found to be optimal for carrying out multiplex SNP genotyping.

6.3.4 Optimisation on SiNW platform

Since the SiNW platform will carry out multiplex genotyping, the current microarray platform was used to test probe specificity when multiplex target DNA was added in the miniarray. Probes of all 3 SNPs showed high specificity when multiplex homozygous and heterozygous target DNA was added at different concentrations, thus suggesting low non-specific hybridization between probes and non-complementary DNAs (Figure 6.9 and Figure 6.10). Furthermore, the probes were able to genotype 2 anonymised samples for all 3 SNPs accurately in singleplex and multiplex format (Figure 6.11). Therefore, the current set of probes for all 3 SNPs can carry out genotyping on the microarray platform accurately.

The current study didn't test probes on the SiNW platform. In comparison to the microarray platform, microfluidic based SiNW platform offers several advantages such as: smaller hybridization chamber volume, which could help accelerate hybridization kinetics by increasing the probability of collision between the probe and target DNA as the diffusion distance is reduced (Peytavi et al., 2005, McQuain et al., 2004, Axelrod and Wang, 1994); recirculating hybridization buffer containing target DNAs within microfluidic environment, which has been shown to accelerate hybridization kinetics and reduce hybridization duration (Lee et al., 2006a) and; lower probe surface concentration that could improve target DNA capture rate due to lower electrostatic repulsion from neighboring probe-target DNA duplexes (Peterson et al., 2001).

Owing to the high sensitivity of the SiNW platform, there are several additional issues that would require to be optimized before carrying out multiplex SNP genotyping. In contrast to the microarray platform where hybridization temperature is provided externally to accelerate hybridization reaction and improve specificity, SiNWs can't be used at higher working temperatures as it has a negative effect on the use of SiNWs in electronic circuits (Hashim and Sidek, 2012). However, QuantuMDx has recently been able to heat the SiNW to 50°C without compromising its detection capabilities (Data not shown), thereby allowing the use of high temperatures to carry out SNP genotyping. Additionally, hybridization and washing buffers containing high concentrations of ions can't be used on the SiNW platform as electrolytes in solution have been shown to reduce SiNW sensitivity due to their "screening effect" on the nanowires which leads to reduction in the current passing across the nanowires (Nozaki et al., 2014). Therefore, it is imperative to spot the current set of probes of all 3 SNPs on to the SiNW platform and carry out further optimization of the hybridization conditions in order to carry out rapid sample to result SNP genotyping.

6.4 Conclusion

Results in the current chapter demonstrates designing, optimization, testing and validation of 64 DNA oligonucleotide probes that can carry out simultaneous genotyping of *CYP2C9*2, CYP2C9*3* and *VKORC1*-1639G>A SNPs on a custom designed microarray platform. Several experimental conditions such as hybridization temperature, hybridization duration and washing buffer stringency were optimized. Association of probe length with hybridization temperature, probe's GC content with washing buffer stringency and probe surface concentration and target DNA multiplexing with hybridization duration was observed. Accurate genotype calls for all 3 SNPs in 2 anonymised samples using empirically optimized hybridization and washing conditions was carried out successfully on the microarray platform. Whilst further optimization and validation of probe specificity on microfluidic based SiNW platform would be required due to differences in the hybridization kinetics of probe-target DNA duplex between the two platforms, the current work highlighted associations between probe characteristics and hybridization parameters which would be useful in designing probes for genotyping other SNPs in the future.

Chapter 7: General discussion and future work

Pharmacogenetics and pharmacogenomics can aid in the identification and selection of the appropriate drug and its optimal dose that will improve the risk-benefit ratio for the individual; a demonstration of “personalised medicine”. Study and application of personalised medicine for commonly prescribed drugs such as aspirin and warfarin has seen a rapid increase and the outcomes are likely to improve the delivery of these drugs and reduce the burden of chronic diseases like colorectal cancer and cardiovascular diseases.

The current thesis described three separate projects that had the objective of identifying new clinically useful genetic variants that could help explain variation in the chemopreventive effect of aspirin in relation to colorectal cancer; test the clinical utility of published and clinically validated genotype-guided warfarin dosing algorithms in the Gujarati Indian population and; develop a panel of oligonucleotide probes and delineate optimal hybridization and washing conditions to carry out hybridization based detection of warfarin dose associated SNPs on a custom designed microarray platform that could then be transferred to the silicon nanowire based platform.

In the first and second results chapters (Chapter 3 and 4), out of 43 candidate SNPs from 16 genes that are involved in aspirin’s pharmacokinetic and pharmacodynamic pathways which were selected for the analysis, 3 SNPs were observed to be associated with colorectal cancer risk with 2 of them having site-specific association with colon cancer risk in the meta-analysis. Furthermore, 4 SNPs in 2 genes showed interaction with aspirin use and colorectal cancer risk whereby the variant allele was associated with an increase in cancer risk amongst aspirin users but not in non-users. This exploratory analysis identified 2 SNPs that showed site-specific association with cancer risk and 4 SNPs that showed interaction with aspirin use and cancer risk.

Whilst the associations and interactions in the meta-analysis reached the type 1 error threshold of 0.05, they were not corrected for multiple tests due to number of hypotheses tested and lack of statistical power. Hence, the current results should be interpreted as hypothesis generating observations that require testing in further confirmatory studies and no specific conclusions can be drawn. This was largely due to the relatively small effect size of the associations and interactions that were observed in the current study. However, the size and the direction of association of *CYP2C9* and intergenic SNPs, rs1799853 and rs6983267 respectively, with colorectal cancer risk that was observed in the current study was similar to

the effect size observed in the previous studies (Wang et al., 2014, Tenesa et al., 2008, Zanke et al., 2007, Tomlinson et al., 2007) thus suggesting that the novel associations and interactions noted in the current study could be *bona fide* observations rather than a false discovery. To support this, these SNPs would have to be tested in a new case-control cohort known as Genetic Epidemiology of Colorectal Cancer Consortium (GECCO) that has combined genetic and epidemiological data from 5 case-control and 5 cohort studies thus making a total of 8634 cases and 8553 controls (Nan et al., 2015).

Using this dataset, the GECCO consortium recently showed an interaction between the SNP rs2965667 in *MGST1* gene with aspirin or NSAID use and colorectal cancer risk (Genome wide $P_{interaction} = 4.6 \times 10^{-9}$) whereby the wild-type TT genotype was associated with lower risk of cancer amongst aspirin or NSAID users (OR=0.66, 95% CI=0.61-0.70) compared to the individuals with the rare TA or AA genotype which was associated with an increased risk of cancer (OR=1.89, 95% CI=1.27-2.81) (Nan et al., 2015). Additionally, the consortium also showed an interaction between the SNP rs16973225 near the *IL16* gene with aspirin or NSAID use and colorectal cancer risk (Genome wide $P_{interaction} = 8.2 \times 10^{-9}$) when they carried out case-only interaction analysis. It is to be noted that both SNPs were tested in the current study for association and interaction but no significant results were observed which highlights the issue of the effect size of associations and interactions for these SNPs. Furthermore, previously mentioned SNPs from literature were not identified in the latest GWAS study which suggests that they didn't have power to validate those SNPs low effect sizes.

Like previous studies, the current study was unsuccessful in identifying SNPs with a large effect size that could be utilized in clinical practice. Furthermore, the current study only tested for the effect of SNPs on aspirin's efficacy but didn't test for the association with aspirin's adverse reaction. Moreover, it was beyond the scope of the current study to carry out a generalized burden test to assess the clinical utility of the novel SNPs identified in the study since these SNPs were not validated in other datasets and functional characterization of the variant allele of some SNPs is still pending. Whilst these results highlight lack of evidence for the clinical utility of the SNPs in predicting aspirin dose for prophylaxis and adjuvant therapy for colorectal cancer, it also highlights novel pathways that may be explored in the future to explain the variation in aspirin's chemopreventive efficacy and colorectal tumorigenesis.

In the third result chapter (Chapter 5), the clinical utility of published genotype guided warfarin dosing algorithms that were developed using white European and South Indian population data were tested in the Gujarati Indian population. Unlike aspirin, warfarin dose is significantly associated with the variant alleles in the *CYP2C9* gene (*CYP2C9*2* and

*CYP2C9*3*) and the *VKORC1* gene (-1639G>A) (Jorgensen et al., 2012). Furthermore, development of a pharmacogenetic dose prediction algorithm based on the three SNPs (Klein et al., 2009) and successful demonstration of its clinical utility by improving time spent in the therapeutic range has also been shown in the white European population (Pirmohamed et al., 2013). When the algorithm was tested for its clinical utility in the Gujarati Indian population in the current study, the dose prediction accuracy of the genotype guided dosing algorithm was found to be inferior in the Gujarati Indians compared to white Europeans. This is most likely to be due to the presence of the study subjects requiring warfarin dose for the target INR range 2.5-3.5. Since the pharmacogenetic algorithms were designed to predict dose for the target INR range 2-3, it is highly likely that the algorithms were underestimating dose for the subjects with higher target INR range and thus leading to poor prediction accuracy. However, due to the small study size and lack of power, testing for the influence of genetics on algorithm's performance in the Gujarati Indian population couldn't be carried out. Thus the hypothesis for the influence of genetics on algorithm's performance could not rejected. Hence, one of the explanations for the poor dose prediction accuracy was suggested to be the low frequency of the *VKORC1* rs9923231 variant in Gujarati Indians, which is the most important genetic determinant of warfarin dosing in the white Europeans (Klein et al., 2009, Gage et al., 2008).

Whilst this is the first study to test the clinical utility of published pharmacogenetic algorithms and calculate allele frequencies of various warfarin dose associated SNPs in the Gujarati Indian population, recent study has reported allele frequencies of SNPs in the North Indian population (Giri et al., 2014) and a second group has developed pharmacogenetic based dosing algorithm for the South Indian population (Pavani et al., 2012). When the pharmacogenetic algorithm for the South Indian was tested for its utility in the Gujarati Indian population, the dose prediction accuracy was observed to be significantly lower in the Gujarati Indians. This could either be due to the inaccuracy in the development of the dosing algorithm whereby the predicted dose was over-fitted with the therapeutic dose or it could be due to genetic differences between the sub-populations in India that have been previously reported (Consortium, 2008).

The current study didn't involve identification of the population specific novel SNP that could explain variance in dose of warfarin by 5% or more and further developing the IWPC algorithm to better predict warfarin dose for the Gujarati Indians. This was due to the very low number of patients in the current study that reached therapeutic dose (N=53) whereas, based on the power calculation estimates, a minimum of 772 patients on therapeutic dose of warfarin would be required to carry out aforementioned analysis. Despite the shortfalls of the

current study, it did provide an insight into the clinical utility of published algorithms in the Indian population and re-affirmed the need to carry out GWAS studies in ethnically diverse populations to help implement warfarin pharmacogenetics at a large scale.

Whilst warfarin has been the first line anti-coagulant for more than 6 decades, newer oral anti-coagulants (NOACs) such as dabigatran, rivaroxaban, apixaban and edoxaban that directly inhibit thrombin and factor Xa in the coagulation cascade have been introduced into clinics recently (Pirmohamed et al., 2015). Whilst NOACs offer the advantage of standardized dosing without the need of regular monitoring and have fewer drug interactions than warfarin, they have several disadvantages such as lack of specific antidotes in case of excess bleeding, lack of reliable marker to assess the degree of anticoagulation, possibility of poor adherence and much higher costs than warfarin (Pirmohamed et al., 2015). Many clinical trials (such as RELY trial) promoted the advantage of switching all patients from warfarin to NOACs, a recent double-blind trial called ENGAGE AF-TIMI 48 showed that the use of edoxaban is useful in patients who carry one or more of the warfarin dose associated SNPs whereas in the rest of the patients who were wild-type for all 3 SNPs, no benefit was derived by switching to endoxaban (Mega et al., 2015). The latest study highlighted the benefit of using a 2-stage approach whereby the patients are first stratified to either NOACs or warfarin based on their sensitivity to warfarin and then deriving a genotype guided dose for the patients who are normal responders to warfarin. This approach may become a cost-effective solution for the healthcare bodies however, clinical data to support this approach is lacking currently.

In the final chapter (Chapter 6), development of oligonucleotide probes for genotyping warfarin dose associated SNPs and optimization of hybridization and washing conditions to carry out genotyping on a custom designed microarray platform was described. Since 3 SNPs have been previously shown to be significantly associated with warfarin dose and benefit of genotyping them in the clinics been proven, developing probes for genotyping these SNPs on an ultra-rapid point of care genotyping device would help in implementing warfarin pharmacogenetics in routine clinical practice. The current work identified optimal hybridization and washing conditions for genotyping 3 SNPs using 64 probes on a microarray platform with the aim of replicating the conditions on the silicon nanowire platform once the probes have been spotted on them.

Whilst the microarray platform successfully genotyped 3 SNPs in 2 anonymised samples using optimal conditions, more samples couldn’t be tested as only a certain number of microarray slides were available for testing due to monetary constraints. Furthermore, optimization of conditions using a microfluidic channel and chamber on the microarray platform was not carried out in the current study but several studies have previously reported

an improvement in the hybridization kinetics and target DNA specificity after using a microfluidic channel on a microarray platform (Henry and O'Sullivan, 2012). Moreover, the probe panel and conditions were not tested on the silicon nanowire platform due to the unavailability of a reliable supply of high quality nanowire chips and lack of test station facility consisting of probe station with semiconductor property analyzer where electrical detection using nanowires could be carried out.

Despite the relative success and methodological ease of carrying out SNP genotyping using differential hybridization between the probe and the target, several other methods could also be used for SNP genotyping on the silicon nanowire platform. One such method that is currently under development carries out SNP detection using base extension technique that is similar to the one employed on the Sequenom platform whereby a single base that is complementary to the SNP position on the target DNA is added on the 3'-end of the probe. However, to carry out SNP detection on the silicon nanowire platform, the base used for extending the probe would be conjugated with a heavily charged molecule that will disrupt the current passing through the nanowires thus aiding in SNP detection. This method along with the differential hybridization based SNP genotyping is currently under development at QuantuMDx Ltd. that will use the silicon nanowire platform in their ultra rapid point of care genotyping device.

In conclusion, the current work has aided in the identification of SNPs that are associated with CRC risk and modulate aspirin's chemopreventive efficacy in a exploratory meta-analysis, shown poor dose prediction accuracy of published genotype guided warfarin dosing algorithms in the Gujarati Indian population and developed a probe set for carrying out warfarin dose associated SNPs on custom designed microarray platform where optimal hybridization and washing conditions were identified for rapid genotyping. The work presented here provides a unique overview of the entire pharmacogenetic process, from the methods employed in the identification of clinically useful genetic markers to testing clinical utility of genotype guided dosing algorithms in predicting personalised dose to developing a technology that would help deploy pharmacogenetic knowledge into clinical practice.

Chapter 8: Appendix

NHS
Health Research Authority

NRES Committee Yorkshire & The Humber - Leeds East

Jarrow REC Centre
Room 002
Jarrow Business Centre
Rolling Mill Road
Jarrow
Tyne and Wear
NE32 3DT

Tel: 0191 4283548

04 April 2014

Dr Emma L Northwood
Project Manager
Leeds Cancer Research UK Centre
Section of Epidemiology and Biostatistics
Leeds Inst Mol Medicine
St James's Hospital
Beckett Street
Leeds
LS9 7TF

Dear Dr Northwood

Study title:	**Genetic and Environmental risk factors for Colorectal Cancer**
REC reference:	**LE02/267**
Amendment number:	**Amendment 10, 04/03/14**
Amendment date:	**19 March 2014**

The above amendment was reviewed by the Sub-Committee in correspondence.

Ethical opinion

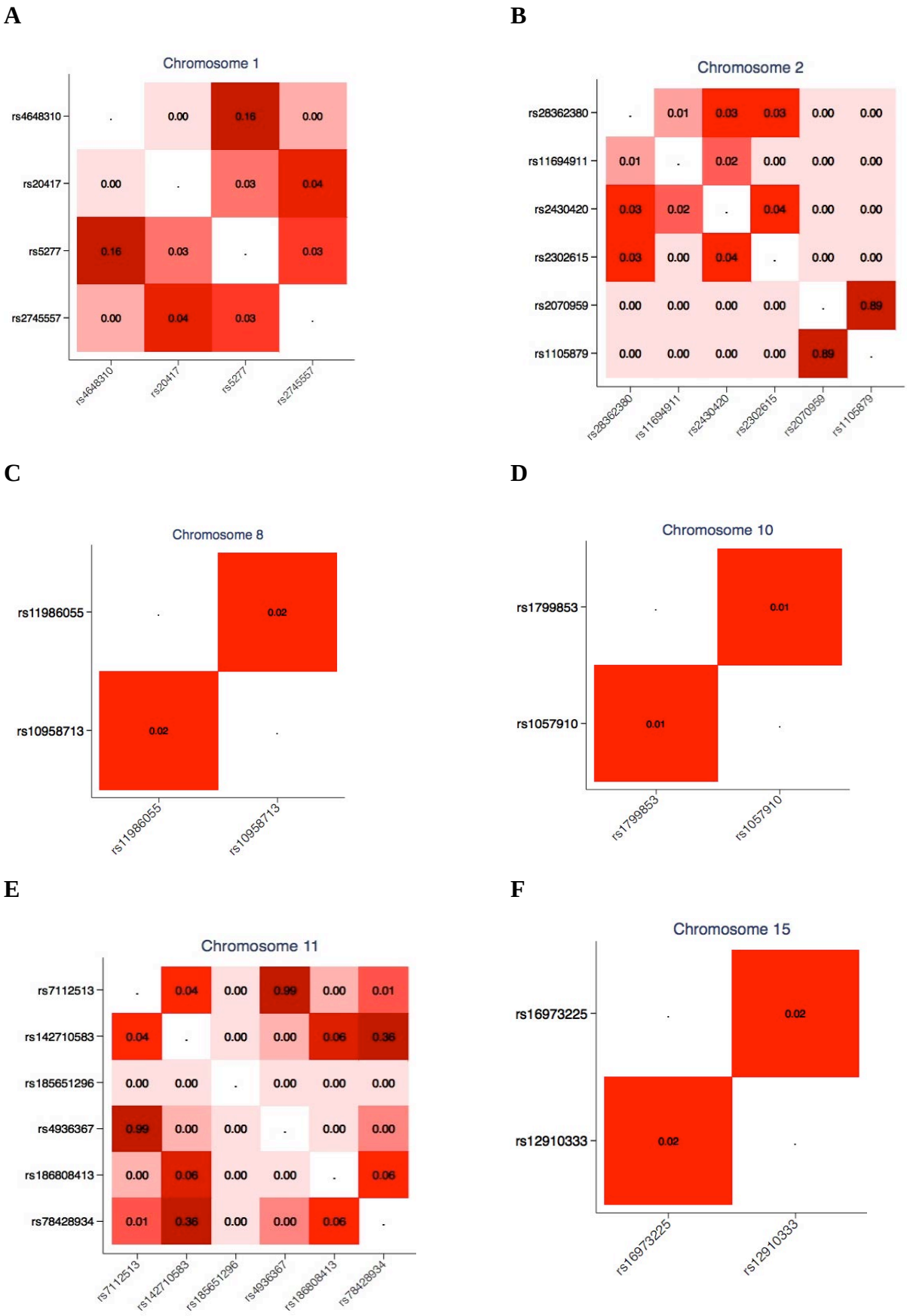
The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents

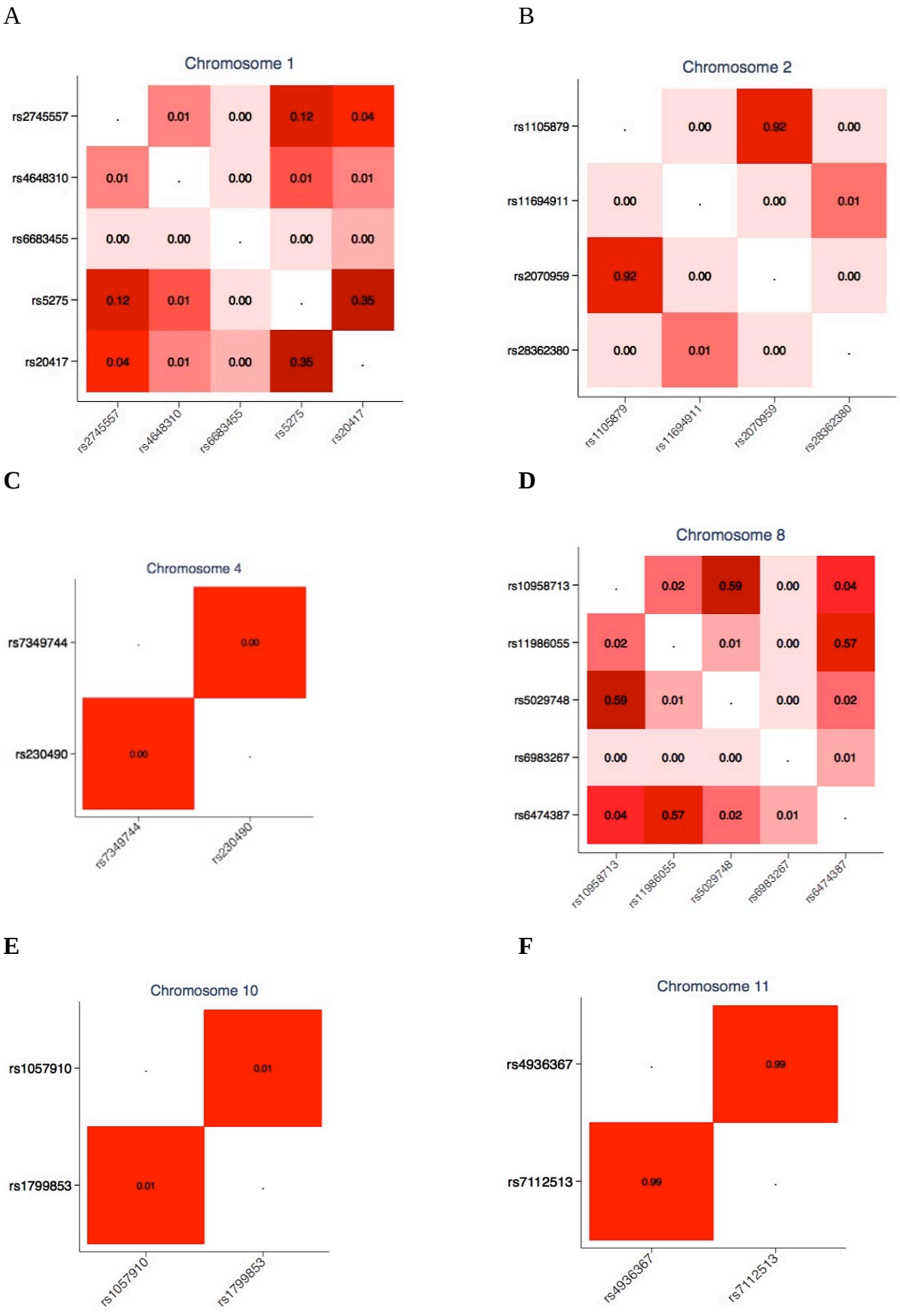
The documents reviewed and approved at the meeting were:

Document	Version	Date
Investigator CV	H.J. Sheth	
Academic Visitor Pro Forma		11 February 2014
Notice of Substantial Amendment (non-CTIMPs)	Amendment 10, 04/03/14	19 March 2014
Protocol	6.0	04 March 2014
Letter of Support from University of Leeds	T. Bishop	

Supplementary Figure 1 St. James' Hospital ethics committee approval letter for carrying out analysis in UK-CCSG and NIH-CCFR dataset.



Supplementary Figure 2 Linkage disequilibrium heat maps for SNPs in the UK-Colorectal Cancer Study Group dataset.



Supplementary Figure 3 Linkage disequilibrium heat maps for SNPs in the NIH-Colon Cancer Family Registry dataset.

SNP ID	UK-Colorectal Cancer Study Group			NIH-Colon Cancer Family Registry			MAF comparison between datasets+
	Observed MAF	Hardy-Weinberg equilibrium (P-value)	MAF comparison (P-value)*	Observed MAF	Hardy-Weinberg equilibrium (P-value)	MAF comparison (P-value)*	
rs1045642	0.46	0.85	0.51	0.48	0.48	0.24	0.31
rs1321311	0.23	0.65	0.41	0.24	0.34	0.26	0.64
rs1057910	0.07	0.13	0.66	0.08	0.25	0.80	0.77
rs1799853	0.14	0.89	**0.04**	0.11	0.23	**0.02**	0.12
rs6983267	0.47	0.40	0.36	0.48	0.95	0.12	0.67
rs961253	0.37	0.13	0.24	0.35	0.26	**0.02**	0.08
rs11694911	0.12	**0.04**	0.79	0.12	**0.02**	0.25	0.91
rs28362380	0.09	0.69	**0.04**	0.09	0.30	0.67	0.89
rs4936367	0.10	**0.03**	**0.002**	0.11	0.87	0.65	0.07
rs7112513	0.10	0.60	**0.01**	0.11	0.87	0.66	0.59
rs3842787	0.09	0.66	0.76	0.07	0.65	0.59	0.19
rs20417	0.14	0.52	0.74	0.18	0.59	0.72	0.36
rs2070959	0.29	**0.05**	0.56	0.33	0.22	0.91	**0.005**
rs1105879	0.32	0.27	0.29	0.35	0.58	0.96	0.06
rs2619112	0.46	1.00	0.80	0.45	**0.05**	0.58	0.30
rs10958713	0.37	0.06	0.57	0.36	0.73	0.98	0.41
rs11986055	0.04	0.09	1.00	0.04	1.0	0.69	0.46
rs12910333	0.28	0.25	0.50	0.30	0.76	0.66	0.35
rs5995355	0.06	0.16	0.77	0.06	0.02	0.85	0.83
rs230490	0.44	0.73	0.66	0.41	0.43	0.53	0.24
rs5275	-	-	-	0.36	0.88	0.55	-
rs4648310	0.04	**0.0003**	0.90	0.04	0.63	0.65	**0.03**
rs5029748	-	-	-	0.25	0.45	0.27	-
rs2745557	0.18	0.91	0.53	0.17	0.10	0.14	0.34
rs6474387	-	-	-	0.06	1.0	0.49	-
rs16973225[A]	0.06	1.0	1.0	-	-	-	-
rs2302615	0.28	0.52	0.94	-	-	-	-
rs2430420	0.34	0.22	**0.009**	-	-	-	-
rs5277[A]	0.14	0.90	0.06	-	-	-	-
rs2965667	0.04	0.22	1.0	-	-	-	-
rs140461033	**0.01**	**<0.0001**	1.0	-	-	-	-
rs144410046	**0.004**	**<0.0001**	1.0	-	-	-	-
rs201103548	**0.005**	**<0.0001**	-	-	-	-	-
rs28382815	**0.002**	**<0.0001**	**0.03**	-	-	-	-
rs148026549	**0.0005**	0.99	1.0	-	-	-	-
rs145407778	**0.002**	0.96	**0.006**	-	-	-	-
rs10852434	**0.00**	1.0	-	-	-	-	-
rs147942040	**0.005**	**<0.0001**	1.0	-	-	-	-
rs141625476	**0.004**	**<0.0001**	**0.001**	-	-	-	-
rs147070911	**0.005**	**<0.0001**	**0.001**	-	-	-	-

rs150408050	**0.004**	**<0.0001**	1.0	-	-	-	-
rs147694237	**0.005**	**<0.0001**	**0.001**	-	-	-	-
rs142710583	**0.004**	**<0.0001**	**0.001**	-	-	-	-
rs185651296	**0.003**	**<0.0001**	0.30	-	-	-	-
rs186808413	**0.015**	0.64	1.0	-	-	-	-
rs78428934	**0.002**	0.96	1.0	-	-	-	-

Supplementary Table 1 Comparison of observed minor allele frequency of SNPs between UK-Colorectal Cancer Study Group and NIH-Colon Cancer Study Registry.

*Observed minor allele frequency (MAF) in controls was compared to the MAF reported for Phase I GBR and Phase I CEU population from 1000 Genomes database in UK-CCSG and NIH-CCFR datasets respectively using Fisher's exact test.

+Observed MAF in controls of the two datasets were compared using Fisher's exact test.

[A]rs16973225 and rs5277 was only genotyped in cases in the NIH-CCFR dataset.

			UK-Colorectal Cancer Study Group					NIH-Colon Cancer Family Registry				
Gene name	**SNP ID**	**Copies of rare allele**	**Controls, n (%)**	**Cases, n (%)**	**Odds Ratio**	**95% CI**	**P-value***	**Controls, n (%)**	**Cases, n (%)**	**Odds Ratio**	**95% CI**	**P-value***
MDR1	rs1045642	0	297 (29.2)	476 (28.4)				257 (26.1)	383 (27.1)			
		1 or 2	722 (70.9)	1202 (71.6)	1.04	0.87, 1.23	0.87	728 (73.9)	1032 (72.9)	0.95	0.79, 1.14	0.93
CDKN1A	rs1321311	0	579 (59.3)	946 (56)				566 (57.8)	789 (55.9)			
		1 or 2	398 (40.7)	743 (44.0)	1.14	0.97, 1.34	0.07	414 (42.2)	622 (44.1)	1.08	0.91, 1.27	0.72
CYP2C9	rs1057910	0	848 (86.9)	1487 (87.8)				841 (85.8)	992 (86.5)			
		1 or 2	128 (13.1)	207 (12.2)	0.92	0.73, 1.17	0.49	139 (14.2)	155 (13.5)	0.95	0.74, 1.21	0.80
	rs1799853	0	709 (74.3)	1274 (78)				141 (78.8)	126 (79.3)			
		1 or 2	245 (25.7)	359 (22.0)	0.82	0.68, 0.98	**0.03**	38 (21.2)	33 (20.8)	0.97	0.58, 1.64	0.83
Intergenic	rs6983267	0	282 (28.9)	523 (31)				267 (27.1)	436 (30.9)			
		1 or 2	693 (71.1)	1167 (69.1)	0.91	0.76, 1.08	0.31	717 (72.9)	976 (69.1)	0.83	0.70, 1.00	0.06
	rs961253	0	401 (41)	654 (38.5)				406 (41.3)	573 (40.5)			
		1 or 2	576 (59.0)	1044 (61.5)	1.11	0.95, 1.30	0.13	578 (58.7)	842 (59.5)	1.03	0.87, 1.22	0.99
ODC1	rs28362380	0	841 (83.4)	1365 (82.6)				822 (83.4)	1149 (81.3)			
		1 or 2	168 (16.7)	287 (17.4)	1.05	0.85, 1.30	0.93	164 (16.6)	2665 (18.7)	1.16	0.93, 1.43	0.21
	rs11694911	0	788 (77.6)	1356 (80.9)				772 (78.3)	1135 (80.2)			
		1 or 2	228 (22.4)	321 (19.1)	0.82	0.68, 0.99	0.11	214 (21.7)	280 (19.8)	0.89	0.73, 1.09	0.31
	rs2430420	0	452 (44.4)	748 (44.5)				-	-	-	-	-
		1 or 2	565 (55.6)	932 (55.5)	1	0.85, 1.17	0.78	-	-	-	-	-

	rs2302615	0	501 (51.6)	907 (56.1)				-	-	-	-	-	
		1 or 2	470 (48.4)	710 (43.9)	0.83	0.71, 0.98	0.06	-	-	-	-	-	
PAFAH1B2	rs4936367	0	779 (80.8)	1294 (79.3)				786 (79.7)	1125 (79.5)				
		1 or 2	185 (19.2)	337 (20.7)	1.1	0.90, 1.34	0.45	200 (20.3)	290 (20.5)	1.01	0.83, 1.24	0.64	
	rs7112513	0	789 (80.6)	1340 (79.1)				779 (79.4)	1122 (79.5)				
		1 or 2	190 (19.4)	355 (20.9)	1.1	0.90, 1.34	0.40	202 (20.6)	289 (20.5)	0.99	0.81, 1.22	0.51	
PTGS1	rs3842787	0	629 (82.8)	974 (79.9)				843 (85.9)	1242 (88.0)				
		1 or 2	131 (17.2)	245 (20.1)	1.21	0.96, 1.53	0.37	139 (14.2)	169 (12.0)	0.83	0.65, 1.05	0.37	
PTGS2	rs4648310	0	951 (93.1)	1568 (93.6)				917 (93.1)	1085 (94.1)				
		1 or 2	70 (6.9)	108 (6.4)	0.94	0.69, 1.28	0.91	68 (6.9)	68 (5.9)	0.85	0.6, 1.20	0.28	
	rs20417	0	750 (73.2)	1179 (70.6)				672 (68.2)	976 (69.0)				
		1 or 2	275 (26.8)	492 (29.4)	1.14	0.96, 1.35	0.17	313 (31.8)	438 (31.0)	0.96	0.81, 1.15	0.80	
	rs2745557	0	612 (67.4)	1150 (70.6)				558 (69.7)	838 (66.9)				
		1 or 2	296 (32.6)	480 (29.5)	0.86	0.72, 1.03	0.14	243 (30.3)	414 (33.1)	1.13	0.94, 1.37	0.11	
	rs5277	0	759 (73.7)	1193 (71.2)				-	-	-	-	-	
		1 or 2	271 (26.3)	482 (28.8)	1.13	0.95, 1.35	0.10	-	-	-	-	-	
	rs5275	0	-	-	-	-	-	385 (41.3)	618 (45.3)				
		1 or 2	-	-	-	-	-	547 (58.7)	746 (54.7)	0.85	0.72, 1.01	0.17	
UGT1A6	rs1105879	0	458 (47.1)	762 (44.9)				412 (41.8)	594 (42.0)				
		1 or 2	515 (52.9)	937 (55.1)	1.09	0.93, 1.28	0.14	573 (58.2)	819 (58.0)	0.99	0.84, 1.17	0.77	
	rs2070959	0	497 (51)	812 (47.9)				433 (44.0)	636 (45.0)				
		1 or 2	477 (49.0)	882 (52.1)	1.13	0.97, 1.33	0.07	551 (56.0)	779 (55.1)	0.96	0.82, 1.13	0.90	

IL16	rs16973225	0	773 (87.6)	1421 (89.0)				-	-	-	-	-
		1 or 2	109 (12.4)	176 (11.0)	0.88	0.68, 1.13	0.23	-	-	-	-	-
	rs12910333	0	455 (51.0)	853 (53.7)				482 (48.9)	737 (52.1)			
		1 or 2	438 (49.1)	735 (46.3)	0.9	0.76, 1.05	0.32	504 (51.1)	677 (47.9)	0.88	0.75, 1.03	0.27
IKBKB	rs11986055	0	851 (93.3)	1487 (92.2)				914 (92.9)	1305 (92.3)			
		1 or 2	61 (6.7)	126 (7.8)	1.18	0.86, 1.62	0.28	70 (7.1)	109 (7.7)	1.09	0.80, 1.49	1.00
	rs10958713	0	373 (41.0)	679 (41.9)				406 (41.2)	609 (43.0)			
		1 or 2	537 (59.0)	940 (58.1)	0.96	0.82, 1.13	0.80	579 (58.8)	806 (57.0)	0.93	0.79, 1.09	0.58
	rs5029748	0	-	-	-	-	-	557 (56.6)	643 (55.8)			
		1 or 2	-	-	-	-	-	428 (43.5)	510 (44.2)	1.03	0.87, 1.23	0.60
	rs6474387	0	-	-	-	-	-	161 (87.0)	146 (90.1)			
		1 or 2	-	-	-	-	-	24 (13.0)	16 (9.9)	0.74	0.38, 1.44	0.24
NCF4	rs5995355	0	791 (88.0)	1409 (88.0)				874 (88.6)	1224 (86.6)			
		1 or 2	108 (12.0)	193 (12.1)	1	0.78, 1.29	0.56	112 (11.4)	190 (13.4)	1.21	0.94, 1.55	0.18
ALOX15	rs2619112	0	263 (29.4)	434 (27.6)				288 (29.2)	390 (27.6)			
		1 or 2	632 (70.6)	1137 (72.4)	1.09	0.91, 1.31	0.50	698 (70.8)	1024 (72.4)	1.08	0.90, 1.30	0.55
NFKB	rs230490	0	289 (32.2)	498 (31.2)				335 (34.0)	478 (33.8)			
		1 or 2	609 (67.8)	1098 (68.8)	1.05	0.88, 1.25	0.77	651 (66.0)	937 (66.2)	1.01	0.85, 1.20	0.73
MGST1	rs2965667	0	668 (93.2)	1361 (92.8)				-	-	-	-	-
		1 or 2	49 (6.8)	105 (7.2)	1.05	0.74, 1.49	0.86	-	-	-	-	-
IL23R	rs6683455	0	-	-	-	-	-	755 (77.3)	867 (75.5)			
		1 or 2	-	-	-	-	-	222 (22.7)	282 (24.5)	1.11	0.90, 1.35	0.53

PGDH	rs7349744	0	-	-	-	-	-	101 (54.6)	205 (48.2)			
		1 or 2	-	-	-	-	-	84 (45.4)	220 (51.8)	1.29	0.91, 1.82	0.46
FLAP	rs17239025	0	-	-	-	-	-	162 (87.6)	153 (93.9)			
		1 or 2	-	-	-	-	-	23 (12.4)	10 (6.1)	0.46	0.21, 1.00	0.14

Supplementary Table 2 Association between SNP variant allele and risk of colorectal cancer.

*P-value is adjusted for age, sex and study site.

CI, confidence interval

n, number of subjects

			UK-Colorectal Cancer Study Group					NIH-Colon Cancer Family Registry				
Gene name	**SNP ID**	**Copies of rare allele**	**Colon, n (%)**	**Rectum, n (%)**	**Odds Ratio**	**95% CI**	**P-value***	**Colon, n (%)**	**Rectum, n (%)**	**Odds Ratio**	**95% CI**	**P-value***
MDR1	rs1045642	0	314 (28.5)	162 (28.1)				218 (27.1)	125 (26.0)			
		1 or 2	788 (71.5)	414 (71.9)	1.02	0.81, 1.27	0.66	586 (72.9)	355 (74.0)	1.06	0.82, 1.37	0.53
CDKN1A	rs1321311	0	623 (56.3)	323 (55.5)				462 (57.7)	254 (53.0)			
		1 or 2	484 (43.7)	259 (44.5)	1.03	0.84, 1.26	0.70	339 (42.3)	225 (47.0)	1.21	0.96, 1.52	0.18
CYP2C9	rs1057910	0	990 (88.9)	497 (85.7)				553 (86.3)	345 (86.3)			
		1 or 2	124 (11.1)	83 (14.3)	1.33	0.99, 1.80	0.12	88 (13.7)	55 (13.8)	1	0.70, 1.44	0.86
	rs1799853	0	855 (79.8)	419 (74.7)				73 (79.4)	50 (79.4)			
		1 or 2	217 (20.2)	142 (25.3)	1.34	1.05, 1.70	**0.008**	19 (20.7)	13 (20.6)	1	0.45, 2.21	0.88
Intergenic	rs6983267	0	346 (31.3)	177 (30.4)				249 (31.0)	150 (31.4)			
		1 or 2	761 (68.7)	406 (69.6)	1.04	0.84, 1.30	0.58	554 (69.0)	328 (68.6)	0.98	0.77, 1.25	0.77
	rs961253	0	432 (38.8)	222 (38.0)				332 (41.3)	182 (37.9)			
		1 or 2	681 (61.2)	363 (62.1)	1.04	0.84, 1.27	0.71	472 (58.7)	298 (62.1)	1.15	0.91, 1.45	0.29
ODC1	rs28362380	0	908 (83.7)	457 (80.6)				656 (81.7)	390 (81.3)			
		1 or 2	177 (16.3)	110 (19.4)	1.23	0.95, 1.61	0.06	147 (18.3)	90 (18.8)	1.03	0.77, 1.38	0.83
	rs11694911	0	894 (81.4)	462 (79.9)				632 (78.6)	392 (81.7)			
		1 or 2	205 (18.7)	116 (20.1)	1.09	0.85, 1.41	0.57	172 (21.4)	88 (18.3)	0.82	0.62, 1.10	0.18
	rs2430420	0	481 (43.5)	267 (46.4)				-	-	-	-	-

		1 or 2	624 (56.5)	308 (53.6)	0.89	0.73, 1.09	0.16	-	-	-	-	-
	rs2302615	0	613 (57.8)	294 (52.9)				-	-	-	-	-
		1 or 2	448 (42.2)	262 (47.1)	1.22	0.99, 1.50	0.06	-	-	-	-	-
PAFAH1B2	rs4936367	0	858 (80.3)	436 (77.6)				643 (80.0)	377 (78.5)			
		1 or 2	211 (19.7)	126 (22.4)	1.18	0.92, 1.51	0.37	161 (20.0)	103 (21.5)	1.09	0.83, 1.44	0.75
	rs7112513	0	890 (80.1)	450 (77.1)				640 (79.9)	377 (78.7)			
		1 or 2	221 (19.9)	134 (23.0)	1.2	0.94, 1.53	0.27	161 (20.1)	102 (21.3)	1.08	0.81, 1.42	0.81
PTGS1	rs3842787	0	642 (80.6)	332 (78.7)				704 (88.0)	423 (88.1)			
		1 or 2	155 (19.5)	90 (21.3)	1.12	0.84, 1.50	0.47	96 (12.0)	57 (11.9)	0.99	0.70, 1.40	0.91
PTGS2	rs4648310	0	1032 (93.9)	536 (92.9)				605 (93.9)	380 (94.5)			
		1 or 2	67 (6.1)	41 (7.1)	1.18	0.79, 1.76	0.59	39 (6.1)	22 (5.5)	0.9	0.52, 1.54	0.78
	rs20417	0	768 (69.9)	411 (71.9)				553 (68.8)	338 (70.4)			
		1 or 2	331 (30.1)	161 (28.2)	0.91	0.73, 1.14	0.76	251 (31.2)	142 (29.6)	0.93	0.72, 1.18	0.41
	rs2745557	0	748 (70.0)	402 (71.7)				485 (68.4)	273 (65.6)			
		1 or 2	321 (30.0)	159 (28.3)	0.92	0.74, 1.15	0.45	224 (31.6)	143 (34.4)	1.13	0.88, 1.47	0.29
	rs5277	0	791 (71.9)	402 (69.9)				-	-	-	-	-
		1 or 2	309 (28.1)	173 (30.1)	1.1	0.88, 1.37	0.40	-	-	-	-	-
	rs5275	0	-	-	-	-	-	361 (46.3)	202 (44.0)			
		1 or 2	-	-	-	-	-	418 (53.7)	257 (56.0)	1.1	0.87, 1.39	0.43
UGT1A6	rs1105879	0	491 (43.1)	281 (48.1)				322 (40.2)	208 (43.3)			
		1 or 2	634 (56.8)	303 (51.9)	0.82	0.67, 1.00	0.06	480 (59.9)	272 (56.7)	0.88	0.70, 1.10	0.29

	rs2070959	0	517 (46.5)	295 (50.7)				348 (43.3)	221 (46.0)				
		1 or 2	595(53.5)	287 (49.3)	0.85	0.69, 1.03	0.11	456 (56.7)	259 (54.0)	0.89	0.71, 1.12	0.39
IL16	rs16973225	0	940 (89.7)	481 (87.6)				-	-	-	-	-
		1 or 2	108 (10.3)	68 (12.4)	1.23	0.89, 1.70	0.14	-	-	-	-	-
	rs12910333	0	561 (53.9)	292 (53.3)				432 (53.8)	239 (49.8)			
		1 or 2	479 (46.1)	256 (46.7)	1.03	0.83, 1.26	0.85	371 (46.2)	241 (50.2)	1.17	0.34, 1.47	0.14
IKBKB	rs11986055	0	987 (92.9)	500 (90.9)				737 (91.7)	445 (92.9)			
		1 or 2	76 (7.2)	50 (9.1)	1.3	0.89, 1.89	0.21	67 (8.3)	34 (7.1)	0.84	0.55, 1.29	0.46
	rs10958713	0	456 (43.0)	223 (40.0)				364 (45.3)	195 (40.6)			
		1 or 2	605 (57.0)	335 (60.0)	1.13	0.92, 1.39	0.12	440 (54.7)	285 (59.4)	1.21	0.96, 1.52	0.07
	rs5029748	0	-	-	-	-	-	370 (57.5)	216 (53.7)			
		1 or 2	-	-	-	-	-	274 (45.6)	186 (46.3)	1.16	0.90, 1.49	0.14
	rs6474387	0	-	-	-	-	-	83 (87.4)	59 (93.7)			
		1 or 2	-	-	-	-	-	12 (12.6)	4 (6.4)	0.47	0.14, 1.53	0.21
NCF4	rs5995355	0	928 (88.1)	481 (87.6)				695 (86.6)	417 (86.9)			
		1 or 2	125 (11.9)	68 (12.4)	1.05	0.77, 1.44	0.67	108 (13.5)	63 (13.1)	0.97	0.70, 1.36	0.80
ALOX15	rs2619112	0	279 (27.0)	155 (28.9)				221 (27.5)	133 (27.7)			
		1 or 2	755 (73.0)	382 (71.1)	0.91	0.72, 1.15	0.38	582 (72.5)	347 (72.3)	0.99	0.77, 1.28	0.78
NFKB	rs230490	0	319 (30.5)	179 (32.6)				264 (32.8)	167 (34.8)			
		1 or 2	727 (69.5)	371 (67.5)	0.91	0.73, 1.14	0.36	540 (67.2)	313 (65.2)	0.92	0.72, 1.16	0.42
MGST1	rs2965667	0	885 (93.2)	476 (92.3)				-	-	-	-	-
		1 or 2	65 (6.8)	40 (7.8)	1.14	0.76, 1.72	0.48	-	-	-	-	-

IL23R	rs6683455	0	-	-	-	-	-	487 (75.7)	289 (72.4)			
		1 or 2	-	-	-	-	-	156 (24.3)	110 (27.6)	1.19	0.89, 1.58	0.24
PGDH	rs7349744	0	-	-	-	-	-	126 (49.4)	62 (43.7)			
		1 or 2	-	-	-	-	-	129 (50.6)	80 (56.3)	1.26	0.83, 1.90	0.22
FLAP	rs17239025	0	-	-	-	-	-	90 (94.7)	59 (92.2)			
		1 or 2	-	-	-	-	-	5 (5.3)	5 (7.8)	1.53	0.42, 5.50	0.32

Supplementary Table 3 Association between SNP variant allele and site-specific colorectal cancer risk.

*P-value is adjusted for age, sex and study site.

CI, confidence interval

n, number of subjects

			UK-Colorectal Cancer Study Group					NIH-Colon Cancer Family Registry				
Gene name	**SNP ID**	**Copies of rare allele**	**Non-users**		**Aspirin users**		**P-value for interaction***	**Non-users**		**Aspirin users**		**P-value for interaction***
			OR (95% CI)	**P-value+**	**OR (95% CI)**	**P-value+**		**OR (95% CI)**	**P-value+**	**OR (95% CI)**	**P-value+**	
MDR1	rs1045642	0	n=537 1		n=139 1			n=392 1		n=138 1		
		1 or 2	n=1307 1.01 (0.82, 1.24)	0.94	n=348 1.22 (0.82, 1.80)	0.33	0.49	n=999 0.97 (0.76, 1.24)	0.83	n=433 0.98 (0.67, 1.44)	0.94	1.00
CDKN1A	rs1321311	0	n=1078 1		n=287 1			n=779 1		n=320 1		
		1 or 2	n=820 1.09 (0.90, 1.33)	0.36	n=212 0.96 (0.67, 1.37)	0.81	0.45	n=608 1.10 (0.88, 1.37)	0.41	n=248 0.94 (0.67, 1.30)	0.70	0.64
CYP2C9	rs1057910	0	n=1677 1		n=434 1			n=1082 1		n=425 1		
		1 or 2	n=226 1.00 (0.75, 1.35)	0.99	n=64 0.82 (0.48, 1.38)	0.45	0.54	n=162 0.99 (0.71, 1.39)	0.97	n=72 1.13 (0.68, 1.87)	0.63	0.85
	rs1799853	0	n=1417 1		n=371 1			n=114 1		n=90 1		
		1 or 2	n=423 0.85 (0.8, 1.07)	0.16	n=119 0.70 (0.46, 1.06)	0.09	0.36	n=32 1.46 (0.66, 3.23)	0.35	n=19 0.80 (0.28, 2.22)	0.67	0.25
Intergenic	rs6983267	0	n=566 1		n=164 1			n=427 1		n=169 1		
		1 or 2	n=1333 0.93 (0.75, 1.14)	0.48	n=334 0.77 (0.52, 1.13)	0.18	0.41	n=960 0.80 (0.63, 1.02)	0.07	n=402 0.82 (0.57, 1.18)	0.29	0.70
	rs961253	0	n=750 1		n=195 1			n=568 1		n=223 1		
		1 or 2	n=1156 1.07 (0.88, 1.30)	0.49	n=304 1.40 (0.97, 2.02)	0.07	0.17	n=822 0.99 (0.79, 1.23)	0.92	n=348 1.17 (0.84, 1.65)	0.35	0.29

ODC1	rs28362380	0	n=1499 1		n=399 1			n=1126 1		n=471 1			
		1 or 2	n=325 0.99 (0.77, 1.27)	0.92	n=77 1.30 (0.79, 2.14)	0.30	0.36	n=266 1.15 (0.87, 1.52)	0.33	n=100 1.19 (0.77, 1.83)	0.43	0.48
	rs11694911	0	n=1473 1		n=384 1			n=1095 1		n=457 1		
		1 or 2	n=372 0.79 (0.62, 0.99)	**0.04**	n=99 0.65 (0.41, 1.01)	**0.05**	0.58	n=297 0.83 (0.63, 1.08)	0.16	n=114 1.07 (0.71, 1.61)	0.75	0.25
	rs2430420	0	n=819 1		n=206 1			-	-	-	-	-
		1 or 2	n=1026 0.97 (0.80, 1.18)	0.79	n=280 0.92 (0.64, 1.33)	0.67	0.95	-	-	-	-	-
	rs2302615	0	n=967 1		n=260 1			-	-	-	-	-
		1 or 2	n=798 0.83 (0.69, 1.00)	**0.05**	n=213 0.86 (0.60, 1.24)	0.42	0.94	-	-	-	-	-
PAFAH1B2	rs4936367	0	n=1475 1		n=393 1			n=1114 1		n=456 1		
		1 or 2	n=372 1.02 (0.80, 1.30)	0.88	n=93 1.65 (1.02, 2.66)	**0.04**	**0.04**	n=278 0.92 (0.70, 1.20)	0.53	n=115 1.00 (0.66, 1.51)	1.00	0.39
	rs7112513	0	n=1515 1		n=404 1			n=1109 1		n=453 1		
		1 or 2	n=391 1.03 (0.81, 1.31)	0.80	n=95 1.55 (0.96, 2.48)	0.07	0.08	n=279 0.89 (0.68, 1.16)	0.38	n=115 1.00 (0.67, 1.51)	0.98	0.30
PTGS1	rs3842787	0	n=1147 1		n=294 1			n=1220 1		n=478 1		
		1 or 2	n=261 1.18 (0.89, 1.57)	0.26	n=75 1.45 (0.86, 2.45)	0.17	0.39	n=165 1.01 (0.72, 1.42)	0.95	n=92 0.71 (0.46, 1.12)	0.14	0.27

PTGS2	rs4648310	0	n=1722 1		n=458 1			n=1176 1		n=470 1		
		1 or 2	n=127 0.89 (0.62, 1.29)	0.55	n=30 0.89 (0.43, 1.87)	0.77	0.76	n=75 1.06 (0.66, 1.71)	0.80	n=31 0.51 (0.23, 1.13)	0.10	0.12
	rs20417	0	n=1339 1		n=342 1			n=951 1		n=399 1		
		1 or 2	n=507 1.08 (0.88, 1.34)	0.46	n=146 1.34 (0.90, 1.99)	0.15	0.43	n=439 0.98 (0.78, 1.23)	0.84	n=172 0.80 (0.56, 1.14)	0.22	0.26
	rs2745557	0	n=1245 1		n=345 1			n=835 1		n=320 1		
		1 or 2	n=557 0.88 (0.71, 1.09)	0.23	n=132 0.84 (0.56, 1.26)	0.39	0.67	n=404 1.27 (0.99, 1.63)	0.06	n=140 1.01 (0.68, 1.51)	0.94	0.38
	rs5277	0	n=1330 1		n=347 1			-	-	-	-	-
		1 or 2	n=522 1.20 (0.97, 1.49)	0.09	n=140 0.92 (0.62, 1.37)	0.70	0.20	-	-	-	-	-
	rs5275	0	-	-	-	-	-	n=580 1		n=232 1		
		1 or 2	-	-	-	-	-	n=748 0.80 (0.64, 1.01)	0.06	n=313 0.88 (0.63, 1.24)	0.46	0.97
UGT1A6	rs1105879	0	n=863 1		n=237 1			n=598 1		n=234 1		
		1 or 2	n=1039 1.00 (0.83, 1.21)	1.00	n=263 1.38 (0.97, 1.98)	0.08	0.10	n=791 0.93 (0.75, 1.16)	0.53	n=337 1.19 (0.85, 1.66)	0.31	0.16
	rs2070959	0	n=927 1		n=253 1			n=638 1		n=246 1		
		1 or 2	n=973 1.02 (0.84, 1.24)	0.83	n=247 1.48 (1.03, 2.11)	**0.03**	**0.05**	n=752 0.87 (0.70, 1.09)	0.23	n=325 1.16 (0.83, 1.61)	0.39	0.12

IL16	rs16973225	0	n=1063 1		n=399 1			-	-	-	-	-	
		1 or 2	n=186 0.93 (0.68, 1.28)	0.93	n=60 0.76 (0.44, 1.31)	0.33	0.69	-	-	-	-	-	
	rs12910333	0	n=925 1		n=248 1			n=697 1		n=291 1			
		1 or 2	n=847 0.85 (0.70, 1.04)	0.12	n=208 1.18 (0.81, 1.73)	0.38	0.28	n=694 0.91 (0.73, 1.13)	0.40	n=280 0.84 (0.60, 1.17)	0.30	0.50	
IKBKB	rs11986055	0	n=1659 1		n=447 1			n=1280 1		n=534 1			
		1 or 2	n=133 1.35 (0.91, 1.99)	0.14	n=28 0.99 (0.45, 2.17)	0.99	0.21	n=110 1.00 (0.66, 1.49)	0.99	n=37 1.17 (0.60, 2.28)	0.65	0.69	
	rs10958713	0	n=752 1		n=209 1			n=591 1		n=238 1			
		1 or 2	n=1048 0.99 (0.82, 1.21)	0.96	n=262 0.96 (0.66, 1.40)	0.84	0.79	n=801 0.94 (0.75, 1.17)	0.58	n=332 0.75 (0.54, 1.05)	0.10	0.50	
	rs5029748	0	-	-	-	-	-	n=704 1		n=267 1			
		1 or 2	-	-	-	-	-	n=547 0.93 (0.74, 1.17)	0.52	n=233 1.08 (0.76, 1.54)	0.66	0.49	
	rs6474387	0	-	-	-	-	-	n=137 1		n=93 1			
		1 or 2	-	-	-	-	-	n=15 1.48 (0.50, 4.38)	0.48	n=18 0.49 (0.16, 1.48)	0.21	0.14	
NCF4	rs5995355	0	n=1563 1		n=407 1			n=1208 1		n=512 1			
		1 or 2	n=215 0.94 (0.70, 1.28)	0.71	n=61 1.82 (1.01, 3.29)	**0.05**	0.09	n=183 1.63 (1.15, 2.30)	**0.005**	n=59 0.88 (0.51, 1.50)	0.63	0.13	

ALOX15	rs2619112	0	n=505 1		n=129 1			n=390 1		n=159 1			
		1 or 2	n=1252 1.00 (0.80, 1.25)	0.99	n=330 1.41 (0.93, 2.12)	0.10	0.20	n=1002 1.16 (0.91, 1.48)	0.22	n=412 0.91 (0.63, 1.31)	0.60	0.14
NFKB	rs230490	0	n=547 1		n=143 1			n=485 1		n=178 1		
		1 or 2	n=1226 1.14 (0.92, 1.41)	0.23	n=324 0.79 (0.52, 1.18)	0.25	0.17	n=907 0.95 (0.76, 1.20)	0.66	n=393 1.13 (0.79, 1.61)	0.50	0.16
MGST1	rs2965667	0	n=1441 1		n=367 1			-	-	-	-	-
		1 or 2	n=115 1.19 (0.77, 1.82)	0.43	n=28 0.72 (0.33, 1.57)	0.41	0.25	-	-	-	-	-
IL23R	rs6683455	0	-	-	-	-	-	n=940 1		n=382 1		
		1 or 2	-	-	-	-	-	n=301 1.00 (0.77, 1.31)	0.98	n=117 1.04 (0.69, 1.58)	0.85	0.86
PGDH	rs7349744	0	-	-	-	-	-	n=154 1		n=86 1		
		1 or 2	-	-	-	-	-	n=140 1.36 (0.80, 2.32)	0.26	n=95 1.64 (0.89, 3.02)	0.11	0.40
FLAP	rs17239025	0	-	-	-	-	-	n=144 1		n=97 1		
		1 or 2	-	-	-	-	-	n=9 0.45 (0.11, 1.86)	0.27	n=14 0.34 (0.09, 1.31)	0.12	0.77

Supplementary Table 4 Association between SNP variant allele and colorectal cancer risk stratified by only aspirin use.

+P-value for association between SNP variant allele and colorectal cancer risk.

*P-value for interaction between SNP variant allele, aspirin use and colorectal cancer risk calculated using Likelihood ratio test. P-value is adjusted for age, sex and study site.

OR, Odds Ratio

CI, Confidence Interval

Ethics Committee of

CIMS®

Care Institute of Medical Sciences

CIMS Hospital Pvt. Ltd.

IRB Regd. No. : IROG 0006900

(Recognised by US Department of Health and Human Services)

24th Sep 2012

To,
Dr.Keyur Parikh
Principal Investigator
Care Institute of Medical Science
Opp. Panchamrut Bunglow,
Nr. Shukan Mall,
Off Science City Road,
Sola, Ahmedabad 380060

Reference: Indo-UK Warfarin Pharmacogenetics Project Investigating the influence of genetics on the safety of Warfarin in Gujarat, India.

Subject: Ethics Committee approval for the conduct of the referenced study.

Dear Dr. Parikh,

The Ethics Committee of Care Institute of Medical Sciences reviewed and discussed of the above mentioned study at the meeting held on 15th Sep 2012 at 3:00 p.m. at Care Institute of Medical Sciences.

The Ethics committee has reviewed and discussed the following study documents for the above referenced study

- Clinical Trial Protocol
- Informed Consent Form: English
- Informed Consent Form: Gujarati
- Translation Certificate From English to Gujarati
- Case Report Form
- Indo-UK Warfarin Pharmacogenetics Project collaboration agreement
- CV and MRC of Principal Investigator

The following members of the Ethics Committee were present at the meeting.

Name	Qualification	Designation/ Role in EC	Affiliations as to the Institution Yes/No
Dr Adarshjit Singh	M.D,D.M,Diplomate NB,D.Pharm	Pharmacologist	No

Page **1** of **2**

CIMS Hospital : Nr. Shukan Mall, Off Science City Road, Sola, Ahmedabad-380060, Gujarat, INDIA.
Ph. : +91-79-2771 2771-75 (5 lines) Fax: +91-79-2771 2770 Mobile : +91-98250 66664, 98250 66668
www.cims.me

Supplementary Figure 4 Ethics approval from Care Institute of Medical Sciences hospital for the warfarin study.

Foundation For Research In Genetics & Endocrinology (FRIGE : E-13237) | ISO 9001:2008

INSTITUTE OF HUMAN GENETICS

GENETICS CENTRE Reg. No. : 952

FRIGE HOUSE, Jodhpur Gam Road, Satellite, Ahmedabad - 380 015. Gujarat. INDIA.

Project approval

Ethical committee & Scientific & Research committee meeting comprising following members, has been scheduled on 25th August, 2012 at FRIGE House, Institute of Human Genetics, Satellite for the project entitled **"Investigating the influence of genetics on the safety of Warfarin in Gujarat, India"** will as a part of Indo-UK collaboration project. This study involves blood collection of patients suffering from cardio-vascular diseases and on Warfarin.

Following members have approved the project without any suggestions except informed consent to the family.

Members:

Chairman	:	Dr. Bipin Shah
Member Secretary	:	Dr. Jayesh Sheth
Haemato-Oncologist	:	Dr. Ashwin Patel
Pediatrician	:	Dr. Nidhish Nanavaty
Scientist	:	Dr. Harish Padh
Gynacologist	:	Dr. Bindu Shah
Gynacologist	:	Dr. Atul munshi
Diabetologist	:	Dr. Navneet Shah
Scientist	:	Dr. Sunil Trivedi
Pediatrician	:	Dr. Raju C. Shah
Pharmacologist	:	Dr. R. K. Goyal
Social Worker	:	Mrs. Manisha Panchal
Lawyer	:	Ms. Tammi Vin
Neurophysician	:	Dr. Sudhir Shah

Dr. Bipin Shah
Chairman
Ethical Committee, FRIGE

Dr. Jayesh Sheth
Member Secretary
Ethical Committee, FRIGE

Prof. V. C. Shah
Executive Trustee
Dr. Jayesh J. Sheth
Hon. Director

'Recognized as Research Organization (SIRO)'
By Govt. of India
Ministry of Science & Technology
(14/409/2005-Tu-V)
&

Tel. : 079 - 2692 14 14, 6512 84 44
Fax : 079 - 2692 14 15
E-mail : jshethad1@gmail.com
fshethad1@googlemail.com

Supplementary Figure 5 Ethics approval from Institute of Human Genetics for the warfarin study

Number	SNP Number	PCR Primer	iPLEX Primer
1	rs61742245	F-ACGTTGGATGTTAGTGCTCTCGCTCTACGC R-ACGTTGGATGAAGACGCGCGAACAGCTGAT	CGCGCGGTAATCCCGGT
2	rs11676382	F-ACGTTGGATGGCCGCAGGTAAGTTCACAAC R-ACGTTGGATGTCTAGAGTTACTCTCCCCAG	AGGGGAAAGTTACCAAG
3	rs1799853	F-ACGTTGGATGCAGTGATATGGAGTAGGGTC R-ACGTTGGATGCTGCGGAATTTTGGGATGG	AAGAGGAGCATTGAGGAC
4	rs17708472	F-ACGTTGGATGGCCCGGCCCTTAAGTAATTC R-ACGTTGGATGCCCAGTCTCTGATGCAAAAC	ACCGAGTGAACCGTTATAC
5	rs339097	F-ACGTTGGATGTCTGTCTTTCCCCTTTAGCC R-ACGTTGGATGCCTTGGATTCTGAATCTGGC	CTGAATCTGGCCAATACTTA
6	rs28371686	F-ACGTTGGATGACATGCCCTACACAGATGCT R-ACGTTGGATGTGTCACAGGTCACTGCATGG	CGCGGTCCAGAGATACATTGA
7	rs2242480	F-ACGTTGGATGTGCTAAGGTTTCACCTCCTC R-ACGTTGGATGGCAGGAGGAAATTGATGCAG	ACCCAATAAGGTGAGTGGATG
8	rs9332131	F-ACGTTGGATGACATGAACAACCCTCAGGAC R-ACGTTGGATGCAAGCAGTCACATAACTAAGC	AGCTTTTGTTTACATTTTACCT
9	rs7294	F-ACGTTGGATGAAAAAAGAGCGAGCGTGTGG R-ACGTTGGATGTTCTAGATTACCCCCTCCTC	TTACCCCCTCCTCCTGCCATACCC
10	rs3814637	F-ACGTTGGATGCGACAATACTTACACAAAGCC R-ACGTTGGATGAGAGAACTGGAAATAACCTC	CTCATTAGGAAATTTAGAACAAATA
11	rs2108622	F-ACGTTGGATGCATCAGTGTTTTCGGAACCC R-ACGTTGGATGGGACAAAAACAGAGAGAGGG	CTCAGGGTCCGGCCACA
12	rs1057910	F-ACGTTGGATGTGTCACAGGTCACTGCATGG R-ACGTTGGATGCTACACAGATGCTGTGGTGC	GCACGAGGTCCAGAGATAC
13	rs9923231	F-ACGTTGGATGTCTGGGAAGTCAAGCAAGAG R-ACGTTGGATGGCTAGGATTATAGGCGTGAG	ATAGGCGTGAGCCACCGCACC

Supplementary Table 5 PCR and Sequenom iPLEX primer sequences for genotyping 13 SNPs on Sequenom MassARRAY platform.

F- Forward primer

R-Reverse primer

SNP Number	Primers	Thermal Cycler Programme				Restriction Enzyme	Restriction site	Digested Product		Reference
		Steps	Temperature (°C)	Time (sec)	Cycles			Allele	Product Size	
rs1799853	F-TACAAATACAATGAAAATATCATG R- CTAACAACCAGACTCATAATG	Denature	95	30	35	*Ava*II	Destroyed	WT	527 + 164	(Gaikwad et al., 2013, Pavani et al., 2011, Adithan et al., 2003)
		Anneal	57	30				Variant	691	
		Extend	72	60						
rs1057910	F-AATAATAATATGCACGAGGTCCAGAGATGC R- GATACTATGAATTTGGGACTTC	Denature	95	30	35	*Nsi*I	Destroyed	WT	112 + 29	(Gaikwad et al., 2013, Pavani et al., 2011, Adithan et al., 2003)
		Anneal	57	30				Variant	141	
		Extend	72	60						
rs9923231	F- CAAGTTCCAGGGATTCATGC R- CAAGACGCTAGACCCAATG	Denature	95	30	30	*Msp*I	Destroyed	WT	327+227	(Gaikwad et al., 2013)
		Anneal	60	30				Variant	554	
		Extend	72	60						
rs2108622	F- CGGAACTTGGACCATCTACA R- CTACTCTCCCACAGGCATTA	Denature	95	30	35	*Pvu*II	Destroyed	WT	379+60	(Pavani et al., 2012)
		Anneal	60	30				Variant	439	
		Extend	72	60						
rs7294	F- TTTGCTTTGGCATGTGAGCCTTGC R- ACAGTCCATGGCAGACACATGGTT	Denature	95	30	30	*Aci*I	Created	WT	182+99	(Pavani et al., 2011)
		Anneal	64	30				Variant	281	
		Extend	72	60						

Supplementary Table 6 Restriction Fragment Length Polymorhism (RFLP) PCR protocol for genotyping 5 SNPs.

F- Forward primer

R-Reverse primer

Clinical Indication						Total
Atrial Fibrillation	Deep Vein Thrombosis	Left Ventricular Ejection Fraction	Pulmonary Embolism	Valvular Heart Disease	Other*	
21	6	9	5	50	11	102

Supplementary Table 7 Clinical indications for warfarin treatment in 102 patients.

*Other clinical indications include arterial thrombosis, coronary artery bypass graft, pacemaker implantation, left branchial paresis and diabetes mellitus.

Assay	QC Assessment	GD Pass Rate*	Project Assessment	Clustering+
rs1057910	Pass	93.1	Pass	2
rs11676382	Pass	99.3	Pass	2
rs17708472	Pass	96.7	Pass	1
rs1799853	Pass	93.3	Pass	1
rs2108622	Pass	96.5	Pass	1
rs2242480	Pass	76.2	Fail	2
rs28371686	Pass	99.3	Pass	2
rs339097	Pass	86.0	Pass	2
rs3814637	Pass	68.8	Fail	1
rs61742245	Pass	93.5	Pass	2
rs7294	Pass	94.7	Pass	1
rs9332131	Pass	89.1	Pass	2
rs9923231	Pass	95.6	Pass	1
Average GD Pass Rate		90.93		
Number of failed assays			2	

Supplementary Table 8 Assay performance on Sequenom MassARRAY platform.

*The pass rate of each assay is calculated with failed DNAs removed (conservative + moderate + aggressive + user calls)/ (total calls – minus bad spectra). An assay is considered a 'pass' if the rate if 80% or over whereas a failed assay fails to reach 80% call rate.

SNP	Sequence Name	Sequence 5'-3'
CYP2C9*2	Har001_051011_CYP2C9_R144C_SENSE_WT	GCA TTG AGG ACC GTG TTC A
	Har002_051011_CYP2C9_R144C_ANTISENSE_WT	TGA ACA CGG TCC TCA ATG C
	Har003_051011_CYP2C9_R144C_SENSE_M	GCA TTG AGG ACT GTG TTC A
	Har004_051011_CYP2C9_R144C_ANTISENSE_M	TGA ACA CAG TCC TCA ATG C
	Har005_051011_CYP2C9_R144C_SENSE_WT	GAG CAT TGA GGA CCG TGT TC
	Har006_051011_CYP2C9_R144C_ANTISENSE_WT	GAA CAC GGT CCT CAA TGC TC
	Har007_051011_CYP2C9_R144C_SENSE_M	GAG CAT TGA GGA CTG TGT TC
	Har008_051011_CYP2C9_R144C_ANTISENSE_M	GAA CAC AGT CCT CAA TGC TC
	Har009_051011_CYP2C9_R144C_SENSE_WT	AGG ACC GTG TTC AAG AGG AA
	Har010_051011_CYP2C9_R144C_ANTISENSE_WT	TTC CTC TTG AAC ACG GTC CT
	Har011_051011_CYP2C9_R144C_SENSE_M	AGG ACT GTG TTC AAG AGG AA
	Har012_051011_CYP2C9_R144C_ANTISENSE_M	TTC CTC TTG AAC ACA GTC CT
	Har013_051011_CYP2C9_R144C_SENSE_WT	GAG GAC CGT GTT CAA GAG GA
	Har014_051011_CYP2C9_R144C_ANTISENSE_WT	TCC TCT TGA ACA CGG TCC TC
	Har015_051011_CYP2C9_R144C_SENSE_M	GAG GAC TGT GTT CAA GAG GA
	Har016_051011_CYP2C9_R144C_ANTISENSE_M	TCC TCT TGA ACA CAG TCC TC
	Har017_051011_CYP2C9_R144C_SENSE_WT	AGC ATT GAG GAC CGT GTT C
	Har018_051011_CYP2C9_R144C_ANTISENSE_WT	GAA CAC GGT CCT CAA TGC T
	Har019_051011_CYP2C9_R144C_SENSE_M	AGC ATT GAG GAC TGT GTT C
	Har020_051011_CYP2C9_R144C_ANTISENSE_M	GAA CAC AGT CCT CAA TGC T
	Har021_051011_CYP2C9_R144C_SENSE_WT	TGA GGA CCG TGT TCA AGA GG
	Har022_051011_CYP2C9_R144C_ANTISENSE_WT	CCT CTT GAA CAC GGT CCT CA
	Har023_051011_CYP2C9_R144C_SENSE_M	TGA GGA CTG TGT TCA AGA GG
	Har024_051011_CYP2C9_R144C_ANTISENSE_M	CCT CTT GAA CAC AGT CCT CA
CYP2C9*3	Har025_051011_CYP2C9_I359L_SENSE_WT	ACG AGG TCC AGA GAT ACA TTG AC
	Har026_051011_CYP2C9_I359L_ANTISENSE_WT	GTC AAT GTA TCT CTG GAC CTC GT
	Har027_051011_CYP2C9_I359L_SENSE_M	ACG AGG TCC AGA GAT ACC TTG AC
	Har028_051011_CYP2C9_I359L_ANTISENSE_M	GTC AAG GTA TCT CTG GAC CTC GT
	Har029_051011_CYP2C9_I359L_SENSE_WT	CGA GGT CCA GAG ATA CAT TGA C
	Har030_051011_CYP2C9_I359L_ANTISENSE_WT	GTC AAT GTA TCT CTG GAC CTC G
	Har031_051011_CYP2C9_I359L_SENSE_M	CGA GGT CCA GAG ATA CCT TGA C
	Har032_051011_CYP2C9_I359L_ANTISENSE_M	GTC AAG GTA TCT CTG GAC CTC G
	Har033_051011_CYP2C9_I359L_SENSE_WT	GAG ATA CAT TGA CCT TCT CCC C
	Har034_051011_CYP2C9_I359L_ANTISENSE_WT	GGG GAG AAG GTC AAT GTA TCT C
	Har035_051011_CYP2C9_I359L_SENSE_M	GAG ATA CCT TGA CCT TCT CCC C
	Har036_051011_CYP2C9_I359L_ANTISENSE_M	GGG GAG AAG GTC AAG GTA TCT C
	Har037_051011_CYP2C9_I359L_SENSE_WT	GAT ACA TTG ACC TTC TCC CCA
	Har038_051011_CYP2C9_I359L_ANTISENSE_WT	TGG GGA GAA GGT CAA TGT ATC
	Har039_051011_CYP2C9_I359L_SENSE_M	GAT ACC TTG ACC TTC TCC CCA
	Har040_051011_CYP2C9_I359L_ANTISENSE_M	TGG GGA GAA GGT CAA GGT ATC
	Har041_051011_CYP2C9_I359L_SENSE_WT	AGA TAC ATT GAC CTT CTC CCC A
	Har042_051011_CYP2C9_I359L_ANTISENSE_WT	TGG GGA GAA GGT CAA TGT ATC T
	Har043_051011_CYP2C9_I359L_SENSE_M	AGA TAC CTT GAC CTT CTC CCC A

	Har044_051011_CYP2C9_I359L_ANTISENSE_M	TGG GGA GAA GGT CAA GGT ATC T
***VKORC1*-1639G>A**	Har045_051011_VKORC1_G1639A_SENSE_WT	CAC CCG GCC AAT GGT T
	Har046_051011_VKORC1_G1639A_ANTISENSE_WT	AAC CAT TGG CCG GGT G
	Har047_051011_VKORC1_G1639A_SENSE_M	CAC CTG GCC AAT GGT T
	Har048_051011_VKORC1_G1639A_ANTISENSE_M	AAC CAT TGG CCA GGT G
	Har049_051011_VKORC1_G1639A_SENSE_WT	ACC CGG CCA ATG GTT G
	Har050_051011_VKORC1_G1639A_ANTISENSE_WT	CAA CCA TTG GCC GGG T
	Har051_051011_VKORC1_G1639A_SENSE_M	ACC TGG CCA ATG GTT G
	Har052_051011_VKORC1_G1639A_ANTISENSE_M	CAA CCA TTG GCC AGG T
	Har053_051011_VKORC1_G1639A_SENSE_WT	CGC ACC CGG CCA AT
	Har054_051011_VKORC1_G1639A_ANTISENSE_WT	ATT GGC CGG GTG CG
	Har055_051011_VKORC1_G1639A_SENSE_M	CGC ACC TGG CCA AT
	Har056_051011_VKORC1_G1639A_ANTISENSE_M	ATT GGC CAG GTG CG
	Har057_051011_VKORC1_G1639A_SENSE_WT	ACC GCA CCC GGC C
	Har058_051011_VKORC1_G1639A_ANTISENSE_WT	GGC CGG GTG CGG T
	Har059_051011_VKORC1_G1639A_SENSE_M	ACC GCA CCT GGC C
	Har060_051011_VKORC1_G1639A_ANTISENSE_M	GGC CAG GTG CGG T
	Har061_051011_VKORC1_G1639A_ANTISENSE_M	IGG CCA GGT GCG GT
	Har062_051011_VKORC1_G1639A_ANTISENSE_M	IIG GCC AGG TGC GGT
	Har063_051011_VKORC1_G1639A_ANTISENSE_M	III GGC CAG GTG CGG T
	Har076_251011_ACTB_COTROL_SENSE_PROBE	GGT CCC GGC CAG CC
	Har079_271011_ACTB_CONTROL_SENSE_PROBE	CTC GTA GAT GGG CAC AGT GT

Supplementary Table 9 Probe set designed to genotype 3 warfarin dose SNPs.

I, Inosine

Har076 and Har079 probes used a positive controls.

WT= wild type, M= mutant

SNP/ Gene	Sequence Name	Sequence 5'-3'	Melting temperature (°C)	Amplicon size
CYP2C9*2	Har064_051011_CYP2C9_R144C_FP	CCT CCT AGT TTC GTT TCT CTT CCT GT	64	232 bp
	Har065_051011_CYP2C9_R144C_RP	CAT ATC ACT GAC CTT ACT GGA CTA CTA TCT TCT CTA C	65	
CYP2C9*3	Har066_051011_CYP2C9_I359L_FP	TGC ATG CAA GAC AGG AGC C	64	154 bp
	Har067_051011_CYP2C9_I359L_RP	GGA GAA ACA AAC TTA CCT TGG GAA	62	
VKORC1-1639G>A	Har068_051011_VKORC1_G1639A_FP	AGC CAG CAG GAG AGG GAA ATA	62	167 bp
	Har069_051011_VKORC1_G1639A_RP	GCC TCC CAA AAT GCT AGG ATT	62	
ACTB	Har077_251011_ACTB_CONTROL_FP	GTG GTG GTG AAG CTG TAG CC	61	197 bp
	Har078_251011_ACTB_CONTROL_RP	GCT GTG CTA TCC CTG TAC GC	61	

Supplementary Table 10 PCR primer sequence for generating PCR products for the 3 warfarin SNPs and *ACTB* control sequence.

FP= Forward primer, RP= Reverse primer.

bp, base pair

SNP	DNA sequence (5’ – 3’ sense strand)
CYP2C9*2	GCTCCTCGGGCAGAGCTTGGCCCATCCACATGGCTGCCCAGTGTCAGCTTCCTCTTTCTTGCCTGGGATCTCCCTCCTAGTTTCGTTTCTCTTCCT GTTAGGAATTGTTTTCAGCAATGGAAAGAAATGGAAGGAGATCCGGCGTTTCTCCCTCATGACGCTGCGGAATTTTGGGATGGGGAAGAGGAG CATTGAGGAC[C/T]GTGTTCAAGAGGAAGCCCGCTGCCTTGTGGAGGAGTTGAGAAAAACCAAGGGTGGGTGACCCTACTCCATATCACTGACC TTACTGGACTACTATCTTCTCTACTGACATTCTTGGAAACATTTCAGGGGTGGCCATATCTTTCATTATGAGTCCTGGTTGTTAGCTCATGTGAAG CGGGGGTTTGAAGCTGAGAGCCAA
CYP2C9*3	CCCCTGAATTGCTACAACAAATGTGCCATTTTTCTCCTTTTCCATCAGTTTTTACTTGTGTCTTATCAGCTAAAGTCCAGGAAGAGATTGAACGT GTGATTGGCAGAAACCGGAGCCCCTGCATGCAAGACAGGAGCCACATGCCCTACACAGATGCTGTGGTGCACGAGGTCCAGAGATAC[A/C]TT GACCTTCTCCCCACCAGCCTGCCCCATGCAGTGACCTGTGACATTAAATTCAGAAACTATCTCATTCCCAAGGTAAGTTTGTTTCTCCTACACTG CAACTCCATGTTTTCGAAGTCCCCAAATTCATAGTATCATTTTTAAACCTCTACCATCACCGGGTGAGAGAAGTGCATAACTCATATGTA
VKORC1-1639G>A	TGGACTACAGGTGCCTGCCACCATGTCTGGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAGGCTTGTCTTAAACTCCT GACCTCAAGTGATCCACCCACCTCGGCCTCCCAAAATGCTAGGATTATAGGCGTGAGCCACCGCACC[C/T]GGCCAATGGTTGTTTTTCAGGTCT TCTCTTGCTTGACTTCCCAGAGGGATCCCTTACTGTTGCACCTACCCTTCTGGGAACTCTCTTCCTCTGGCGTCTGTGATATTTCCCTCTCCTGCT GGCTCCTCCCTCTCCAGATGCTGTTTCTCACATCTACTCTCTTCTAGAGAGTGTGGTAGACAGAATAATGGTCACCAAAGATGTCCC

Supplementary Table 11 DNA sequence encompassing the SNP sent for synthesis

Letters in red color show SNP position within the DNA sequence

[Ancestral nucleotide/Variant nucleotide]

8.1 Supplementary Information 1 (Klein et al., 2009 algorithm)

Predicted clinical dose (mg/week)= [4.0376-(0.2546*Age in decades)+(0.0118*Height in cm)+(0.0134*Weight in Kg)-(0.6752*Asian race)+(0.4060*Black or African American)+(0.0443*Missing or Mixed race)+(1.2799*Enzyme inducer status)-(0.5695*Amiodarone status)]^2

Predicted pharmacogenetic dose (mg/week)= [5.6044-(0.2614*Age in decades)+(0.0087*Height in cm)+(0.0128*Weight in Kg)-(0.8677*VKORC1 A/G)-(1.6974*VKORC1 A/A)-(0.4854*VKORC1 genotype unknown)-(0.5211*CYP2C9 *1/**2)-(0.9357*CYP2C9 *1/**3)-(1.0616*CYP2C9 *2/**2)-(1.9206*CYP2C9 *2/**3)-(2.3312*CYP2C9 *3/**3)-(0.2188*CYP2C9 genotype unknown)-(0.1092*Asian race)-(0.2760*Black or African American)-(0.1032*Missing or Mixed race)+(1.1816*Enzyme inducer status)-(0.5503*Amiodarone status)]^2

Legend for use of algorithms:

• Age in decades = 1 for 10-19, 2 for 20-29, etc…

• VKORC1 G/A = 1 if heterozygous for rs9923231, otherwise zero

• VKORC1 A/A = 1 if homozygous for A at rs9923231, otherwise zero

• VKORC1 genotype unknown = 1 if rs9923231 genotype missing or unknown, otherwise zero

• CYP2C9 *1/*2 = 1 if CYP2C9 genotype is *1/*2, otherwise zero

• CYP2C9 *1/*3 = 1 if CYP2C9 genotype is *1/*3, otherwise zero

• CYP2C9 *2/*2 = 1 if homozygous for CYP2C9 *2 allele, otherwise zero

• CYP2C9 *2/*3 = 1 if CYP2C9 genotype is *2/*3, otherwise zero

• CYP2C9 *3/*3 = 1 if homozygous for CYP2C9 *3 allele, otherwise zero

• CYP2C9 genotype unknown = 1 if CYP2C9 genotype unknown, otherwise zero • Asian Race = 1 if self-reported race is Asian, otherwise zero

• Black/African American = 1 if self-reported race is Black or African American, otherwise zero

• Missing or Mixed race = 1 if self-reported race is unspecified or mixed, otherwise zero

• Enzyme inducer status = 1 if patient taking carbamazepine, phenytoin, rifampin, or rifampicin, otherwise zero

• Amiodarone status = 1 if patient taking amiodarone, otherwise zero

8.2 Supplementary Information 2 (Gage et al., 2008 algorithm)

Predicted clinical dose (mg/week)= [exp(0.613+(0.425*BSA)-(0.0075*Age)+(0.156*African American race)+(0.216*target INR)-(0.257*Amiodarone status)+(0.108*Smoking status)+(0.0784*DVT/PE))] *7

Predicted pharmacogenetic dose (mg/week)= [exp(0.9751−(0.3238*VKORC1 genotype)+(0.4317*BSA)−(0.4008*CYP2C9*3)−(0.00745*Age)−(0.2066*CYP2C9*2)+(0.2029*target INR)−(0.2538*Amiodarone status)+(0.0922*Smoking status)−(0.0901*African American race)+(0.0664*DVT/PE))]*7

Legend for use of algorithms:

- Age = 1,2,3…..99 etc.
- BSA = Body surface area in m^2
- VKORC1 genotype = 0 if homozygous wild-type, 1 if heterozygous and 2 if homozygous mutant
- CYP2C9*2 = 0 if homozygous wild-type, 1 if heterozygous and 2 if homozygous mutant
- CYP2C9*3 = 0 if homozygous wild-type, 1 if heterozygous and 2 if homozygous mutant
- Target INR = 2.5 for patients with AF, DVT, PE or LVEF; 3 for patients with mechanical heart valve replacement
- Amiodarone status = 1 if patient is taking Amiodarone, otherwise zero
- Smoking status = 1 if patient is smoking, otherwise zero
- African American race = 1 if self reported race is African American, otherwise zero
- DVT/ PE = 1 if patient is treated for DVT or PE, otherwise zero

8.3 Supplementary Information 3 (Pavani et al., 2012 algorithm)

Predicted pharmacogenetic dose for males (mg/week)= -(0.1013885349*Age)+(1.449999606*BMI)+(8.054730665*CYP2C9*2)+(1.726919455*CYP2C9*3)-(4.437335987*VKORC1*3)-(2.771903482*VKORC1*4)+ (1.511628517*VKORC1 -1639)+(1.570215716*CYP4F2 V433M)+ (2.409742997*GGCX)+(11.05198035*CYP2C9*8)+7.970140851

Predicted pharmacogenetic dose for females (mg/week)= -(0.05440552061*Age)-(0.2938201651*BMI)-(1.576151039*CYP2C9*2) -(5.950436495*CYP2C9*3)+(2.983528309*VKORC1*3)+(8.699010214*VKORC1*4)-(11.00733747*VKORC1 -1639)-(2.282918521*CYP4F2 V433M)-(4.097105716*GGCX)-(2.96671589*CYP2C9*8)+44.53497515

Legend for use of algorithms:

- Age = 1,2,3…..99 etc.
- BMI = Body Mass Index in Kg/m^2
- CYP2C9*2 = 0 if homozygous wild-type, 1 if heterozygous and 2 if homozygous mutant
- CYP2C9*3 = 0 if homozygous wild-type, 1 if heterozygous and 2 if homozygous mutant
- VKORC1*3 genotype = 0 if homozygous wild-type, 1 if heterozygous and 2 if homozygous mutant
- VKORC1*4 genotype = 0 if homozygous wild-type, 1 if heterozygous and 2 if homozygous mutant
- VKORC1 -1639 genotype = 0 if homozygous wild-type, 1 if heterozygous and 2 if homozygous mutant
- CYP4F2 V433M genotype = 0 if homozygous wild-type, 1 if heterozygous and 2 if homozygous mutant
- GGCX genotype = 0 if homozygous wild-type, 1 if heterozygous and 2 if homozygous mutant
- CYP2C9*8 genotype = 0 if homozygous wild-type, 1 if heterozygous and 2 if homozygous mutant

Bibliography

2009. Estimation of the Warfarin Dose with Clinical and Pharmacogenetic Data. *New England Journal of Medicine,* 360**,** 753-764.

ADITHAN, C., GERARD, N., VASU, S., BALAKRISHNAN, R., SHASHINDRAN, C. H. & KRISHNAMOORTHY, R. 2003. Allele and genotype frequency of CYP2C9 in Tamil Nadu population. *Eur J Clin Pharmacol (2003),* 59**,** 707–709.

AFA. 2015. *Atrial Fibrillation Association* [Online]. Available: http://www.atrialfibrillation.org.uk.

AGUILAR, M. I. & HART, R. 2005. Oral anticoagulants for preventing stroke in patients with non-valvular atrial fibrillation and no previous history of stroke or transient ischemic attacks. *Cochrane Database Syst Rev***,** CD001927.

AGUNDEZ, J. A. G., MARTINEZ, C., PEREZ-SALA, D., CARBALLO, M., TORRES, M. J. & GARCIA-MARTIN, E. 2009. Pharmacogenomics in Aspirin Intolerance. *Current Drug Metabolism,* 10**,** 998-1008.

AITHAL, G., DAY, C., KESTEVEN, P. & DALY, A. 1999. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. *Lancet,* 353**,** 717-719.

ALFONSO, L., AI, G., SPITALE, R. C. & BHAT, G. J. 2014. Molecular targets of aspirin and cancer prevention. *Br J Cancer,* 111**,** 61-7.

ALGRA, A. M. & ROTHWELL, P. M. 2012. Effects of regular aspirin on long-term cancer incidence and metastasis: a systematic comparison of evidence from observational studies versus randomised trials. *Lancet Oncol,* 13**,** 518-27.

ALVING, A. S., CARSON, P. E., FLANAGAN, C. L. & ICKES, C. E. 1956. Enzymatic deficiency in primaquine-sensitive erythrocytes. *Science,* 124**,** 484-5.

ANDERSON, J. L., HORNE, B. D., STEVENS, S. M., WOLLER, S. C., SAMUELSON, K. M., MANSFIELD, J. W., ROBINSON, M., BARTON, S., BRUNISHOLZ, K., MOWER, C. P., HUNTINGHOUSE, J. A., ROLLO, J. S., SILER, D., BAIR, T. L., KNIGHT, S., MUHLESTEIN, J. B. & CARLQUIST, J. F. 2012. A randomized and clinical effectiveness trial comparing two pharmacogenetic algorithms and standard care for individualizing warfarin dosing (CoumaGen-II). *Circulation,* 125**,** 1997-2005.

ANDRE, T., BONI, C., MOUNEDJI-BOUDIAF, L., NAVARRO, M., TABERNERO, J., HICKISH, T., TOPHAM, C., ZANINELLI, M., CLINGAN, P., BRIDGEWATER, J., TABAH-FISCH, I. & DE GRAMONT, A. 2004. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *N Engl J Med,* 350**,** 2343-51.

ANGSTADT, A. Y., HARTMAN, T. J., LESKO, S. M., MUSCAT, J. E., ZHU, J., GALLAGHER, C. J. & LAZARUS, P. 2014. The effect of UGT1A and UGT2B polymorphisms on colorectal cancer risk: Haplotype associations and gene–environment interactions. *Genes, Chromosomes and Cancer,* 53**,** 454-466.

AVERY, P. J., JORGENSEN, A., HAMBERG, A. K., WADELIUS, M., PIRMOHAMED, M. & KAMALI, F. 2011. A proposal for an individualized pharmacogenetics-based warfarin initiation dose regimen for patients commencing anticoagulation therapy. *Clin Pharmacol Ther,* 90**,** 701-6.

AXELROD, D. & WANG, M. D. 1994. Reduction-of-dimensionality kinetics at reaction-limited cell surface receptors. *Biophys J,* 66**,** 588-600.

BARBER, T. D., MCMANUS, K., YUEN, K. W., REIS, M., PARMIGIANI, G., SHEN, D., BARRETT, I., NOUHI, Y., SPENCER, F., MARKOWITZ, S., VELCULESCU, V. E., KINZLER, K. W., VOGELSTEIN, B., LENGAUER, C. & HIETER, P. 2008. Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. *Proc Natl Acad Sci U S A,* 105**,** 3443-8.

BARON, J. A., COLE, B. F., SANDLER, R. S., HAILE, R. W., AHNEN, D., BRESALIER, R., MCKEOWN-EYSSEN, G., SUMMERS, R. W., ROTHSTEIN, R., BURKE, C. A., SNOVER, D. C., CHURCH, T. R., ALLEN, J. I., BEACH, M., BECK, G. J., BOND, J. H., BYERS, T., GREENBERG, E. R., MANDEL, J. S., MARCON, N., MOTT, L. A., PEARSON, L., SAIBIL, F. & VAN STOLK, R. U. 2003. A Randomized Trial of Aspirin to Prevent Colorectal Adenomas. *New England Journal of Medicine,* 348**,** 891-899.

BARRETT, J. H., SMITH, G., WAXMAN, R., GOODERHAM, N., LIGHTFOOT, T., GARNER, R. C., AUGUSTSSON, K., WOLF, C. R., BISHOP, D. T. & FORMAN, D. 2003. Investigation of interaction between N-acetyltransferase 2 and heterocyclic amines as potential risk factors for colorectal cancer. *Carcinogenesis,* 24**,** 275-82.

BARRY, E. L., MOTT, L. A., SANDLER, R. S., AHNEN, D. J. & BARON, J. A. 2011. Variants downstream of the ornithine decarboxylase gene influence risk of colorectal adenoma and aspirin chemoprevention. *Cancer Prev Res (Phila),* 4**,** 2072-82.

BARRY, E. L., POOLE, E. M., BARON, J. A., MAKAR, K. W., MOTT, L. A., SANDLER, R. S., AHNEN, D. J., BRESALIER, R. S., MCKEOWN-EYSSEN, G. E. & ULRICH, C. M. 2013. CYP2C9 variants increase risk of colorectal adenoma recurrence and modify associations with smoking but not aspirin treatment. *Cancer Causes Control,* 24**,** 47-54.

BARRY, E. L., SANSBURY, L. B., GRAU, M. V., ALI, I. U., TSANG, S., MUNROE, D. J., AHNEN, D. J., SANDLER, R. S., SAIBIL, F., GUI, J., BRESALIER, R. S., MCKEOWN-EYSSEN, G. E., BURKE, C. & BARON, J. A. 2009. Cyclooxygenase-2 polymorphisms, aspirin treatment, and risk for colorectal adenoma recurrence--data from a randomized clinical trial. *Cancer Epidemiol Biomarkers Prev,* 18**,** 2726-33.

BAU, J. T., KANG, Z., AUSTIN, C. A. & KURZ, E. U. 2014. Salicylate, a catalytic inhibitor of topoisomerase II, inhibits DNA cleavage and is selective for the alpha isoform. *Mol Pharmacol,* 85**,** 198-207.

BELL, R. G. 1978. Metabolism of vitamin K and prothrombin synthesis: anticoagulants and the vitamin K--epoxide cycle. *Fed Proc,* 37**,** 2599-604.

BELLOSTA, S., PAOLETTI, R. & CORSINI, A. 2004. Safety of statins: focus on clinical pharmacokinetics and drug interactions. *Circulation,* 109**,** III50-7.

BENAMOUZIG, R., DEYRA, J., MARTIN, A., GIRARD, B., JULLIAN, E., PIEDNOIR, B., COUTURIER, D., COSTE, T., LITTLE, J. & CHAUSSADE, S. 2003. Daily soluble aspirin and prevention of colorectal adenoma recurrence: one-year results of the APACC trial. *Gastroenterology,* 125**,** 328-36.

BENDER, R. & LANGE, S. 2001. Adjusting for multiple testing—when and how? *Journal of Clinical Epidemiology,* 54**,** 343-349.

BENITEZ, J., JABLONSKI, M. R., ALLEN, J. D. & WINNER, J. G. 2015. The clinical validity and utility of combinatorial pharmacogenomics: Enhancing patient outcomes. *Applied & Translational Genomics.*

BIGLER, J., WHITTON, J., LAMPE, J. W., FOSDICK, L., BOSTICK, R. M. & POTTER, J. D. 2001. CYP2C9 and UGT1A6 genotypes modulate the protective effect of aspirin on colon adenoma risk. *Cancer Res,* 61**,** 3566-9.

BLACK, D. J., KUNZE, K. L., WIENKERS, L. C., GIDAL, B. E., SEATON, T. L., MCDONNELL, N. D., EVANS, J. S., BAUWENS, J. E. & TRAGER, W. F. 1996. Warfarin-fluconazole. II. A metabolically based drug interaction: in vivo studies. *Drug Metab Dispos,* 24**,** 422-8.

BODMER, W. & BONILLA, C. 2008. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet,* 40**,** 695-701.

BORENSTEIN, M., HEDGES, L. V., HIGGINS, J. P. T. & ROTHSTEIN, H. R. 2010. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods,* 1**,** 97-111.

BOSETTI, C., GALLUS, S. & LA VECCHIA, C. 2006. Aspirin and cancer risk: an updated quantitative review to 2005. *Cancer Causes Control,* 17**,** 871-88.

BOULANGER, L., KIM, J., FRIEDMAN, M., HAUCH, O., FOSTER, T. & MENZIN, J. 2006. Patterns of use of antithrombotic therapy and quality of anticoagulation among patients with non-valvular atrial fibrillation in clinical practice. *Int J Clin Pract,* 60**,** 258-64.

BRADY, R. R., LOVERIDGE, C. J., DUNLOP, M. G. & STARK, L. A. 2011. c-Src dependency of NSAID-induced effects on NF-kappaB-mediated apoptosis in colorectal cancer cells. *Carcinogenesis,* 32**,** 1069-77.

BRONNER, C. E., BAKER, S. M., MORRISON, P. T., WARREN, G., SMITH, L. G., LESCOE, M. K., KANE, M., EARABINO, C., LIPFORD, J., LINDBLOM, A. & ET AL. 1994. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature,* 368**,** 258-61.

BURN, J., BISHOP, D. T., CHAPMAN, P. D., ELLIOTT, F., BERTARIO, L., DUNLOP, M. G., ECCLES, D., ELLIS, A., EVANS, D. G., FODDE, R., MAHER, E. R., MOSLEIN, G., VASEN, H. F., COAKER, J., PHILLIPS, R. K., BULOW, S. & MATHERS, J. C. 2011a. A randomized placebo-controlled prevention trial of aspirin and/or resistant starch in young people with familial adenomatous polyposis. *Cancer Prev Res (Phila),* 4**,** 655-65.

BURN, J., BISHOP, D. T., MECKLIN, J. P., MACRAE, F., MOSLEIN, G., OLSCHWANG, S., BISGAARD, M. L., RAMESAR, R., ECCLES, D., MAHER, E. R., BERTARIO, L., JARVINEN, H. J., LINDBLOM, A., EVANS, D. G., LUBINSKI, J., MORRISON, P. J., HO, J. W., VASEN, H. F., SIDE, L., THOMAS, H. J., SCOTT, R. J., DUNLOP, M., BARKER, G., ELLIOTT, F., JASS, J. R., FODDE, R., LYNCH, H. T. & MATHERS, J. C. 2008. Effect of aspirin or resistant starch on colorectal neoplasia in the Lynch syndrome. *N Engl J Med,* 359**,** 2567-78.

BURN, J., GERDES, A. M., MACRAE, F., MECKLIN, J. P., MOESLEIN, G., OLSCHWANG, S., ECCLES, D., EVANS, D. G., MAHER, E. R., BERTARIO, L., BISGAARD, M. L., DUNLOP, M. G., HO, J. W., HODGSON, S. V., LINDBLOM, A., LUBINSKI, J., MORRISON, P. J., MURDAY, V., RAMESAR, R., SIDE, L., SCOTT, R. J., THOMAS, H. J., VASEN, H. F., BARKER, G., CRAWFORD, G., ELLIOTT, F., MOVAHEDI, M., PYLVANAINEN, K., WIJNEN, J. T., FODDE, R., LYNCH, H. T., MATHERS, J. C. & BISHOP, D. T. 2011b. Long-term effect of aspirin on cancer risk in carriers of hereditary colorectal cancer: an analysis from the CAPP2 randomised controlled trial. *Lancet,* 378**,** 2081-7.

CAMPBELL, C. L., SMYTH, S., MONTALESCOT, G. & STEINHUBL, S. R. 2007. Aspirin dose for the prevention of cardiovascular disease: a systematic review. *JAMA,* 297**,** 2018-24.

CAMPBELL, H. A. & LINK, K. P. 1941. Studies on the hemorrhagic sweet clover disease. IV. The isolation and crystallization of the hemorrhagic agent. *Journal of Biological Chemistry,* 138**,** 21-33.

CAMPBELL, L., WILSON, H. K., SAMUEL, A. M. & GOMPERTZ, D. 1988. Interactions of m-xylene and aspirin metabolism in man. *Br J Ind Med,* 45**,** 127-32.

CAMUSSI, G., AGLIETTA, M., MALAVASI, F., TETTA, C., PIACIBELLO, W., SANAVIO, F. & BUSSOLINO, F. 1983. The release of platelet-activating factor from human endothelial cells in culture. *The Journal of Immunology,* 131**,** 2397-403.

CAPPELLINI, M. D. & FIORELLI, G. 2008. Glucose-6-phosphate dehydrogenase deficiency. *The Lancet,* 371**,** 64-74.

CARAPETIS, J. & MCDONALD, M. 2005. Acute rheumatic fever. *Lancet,* 366**,** 155-168.

CHAN, A. T., OGINO, S. & FUCHS, C. S. 2007. Aspirin and the Risk of Colorectal Cancer in Relation to the Expression of COX-2. *New England Journal of Medicine,* 356**,** 2131-2142.

CHAN, A. T., TRANAH, G. J., GIOVANNUCCI, E. L., HUNTER, D. J. & FUCHS, C. S. 2005. Genetic variants in the UGT1A6 enzyme, aspirin use, and the risk of colorectal adenoma. *J Natl Cancer Inst,* 97**,** 457-60.

CHEN, Y., KUEHL, G. E., BIGLER, J., RIMORIN, C. F., SCHWARZ, Y., SHEN, D. D. & LAMPE, J. W. 2007. UGT1A6 polymorphism and salicylic acid glucuronidation following aspirin. *Pharmacogenet Genomics,* 17**,** 571-9.

CHOONARA, I. A., MALIA, R. G., HAYNES, B. P., HAY, C. R., CHOLERTON, S., BRECKENRIDGE, A. M., PRESTON, F. E. & PARK, B. K. 1988. The relationship between inhibition of vitamin K1 2,3-epoxide reductase and reduction of clotting factor activity with warfarin. *Br J Clin Pharmacol,* 25**,** 1-7.

CHOU, C.-C., CHEN, C.-H., LEE, T.-T. & PECK, K. 2004. Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression. *Nucleic Acids Research,* 32**,** e99.

CHOU, W. H., YAN, F. X., DE LEON, J., BARNHILL, J., ROGERS, T., CRONIN, M., PHO, M., XIAO, V., RYDER, T. B., LIU, W. W., TEILING, C. & WEDLUND, P. J. 2000. Extension of a pilot study: impact from the cytochrome P450 2D6 polymorphism on outcome and costs associated with severe mental illness. *J Clin Psychopharmacol,* 20**,** 246-51.

CIOTTI, M., MARRONE, A., POTTER, C. & OWENS, I. S. 1997. Genetic polymorphism in the human UGT1A6 (planar phenol) UDP-glucuronosyltransferase: pharmacological implications. *Pharmacogenetics,* 7**,** 485-95.

CLARIA, J. & SERHAN, C. N. 1995. Aspirin triggers previously undescribed bioactive eicosanoids by human endothelial cell-leukocyte interactions. *Proc Natl Acad Sci U S A,* 92**,** 9475-9.

CLAYMAN, C. B., ARNOLD, J., HOCKWALD, R. S., YOUNT, E. H., JR., EDGCOMB, J. H. & ALVING, A. S. 1952. Toxicity of primaquine in Caucasians. *J Am Med Assoc,* 149**,** 1563-8.

COLE, B. F., LOGAN, R. F., HALABI, S., BENAMOUZIG, R., SANDLER, R. S., GRAINGE, M. J., CHAUSSADE, S. & BARON, J. A. 2009. Aspirin for the chemoprevention of colorectal adenomas: meta-analysis of the randomized trials. *J Natl Cancer Inst,* 101**,** 256-66.

COLLINS, A., LONJOU, C. & MORTON, N. E. 1999. Genetic epidemiology of single-nucleotide polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America,* 96**,** 15173-15177.

COLLINS, F. S. & MCKUSICK, V. A. 2001. Implications of the Human Genome Project for medical science. *JAMA,* 285**,** 540-4.

CONSORTIUM, I. G. V. 2008. Genetic landscape of the people of India: a canvas for disease gene exploration. *J Genet,* 87**,** 3-20.

COOK, N. R., LEE, I. M., GAZIANO, J. M., GORDON, D., RIDKER, P. M., MANSON, J. E., HENNEKENS, C. H. & BURING, J. E. 2005. Low-dose aspirin in the primary prevention of cancer: the Women's Health Study: a randomized controlled trial. *JAMA,* 294**,** 47-55.

COOK, N. R., LEE, I. M., ZHANG, S. M., MOORTHY, M. V. & BURING, J. E. 2013. Alternate-day, low-dose aspirin and cancer risk: long-term observational follow-up of a randomized trial. *Ann Intern Med,* 159**,** 77-85.

CORSINI, A., BELLOSTA, S., BAETTA, R., FUMAGALLI, R., PAOLETTI, R. & BERNINI, F. 1999. New insights into the pharmacodynamic and pharmacokinetic properties of statins. *Pharmacol Ther,* 84**,** 413-28.

CROSS, A. J., BOCA, S., FREEDMAN, N. D., CAPORASO, N. E., HUANG, W.-Y., SINHA, R., SAMPSON, J. N. & MOORE, S. C. 2014. Metabolites of tobacco smoking and colorectal cancer risk. *Carcinogenesis*.

CRUK 2014. Cancer incidence and mortality in the UK for the 10 most common cancers. Cancer Research UK.

CUI, Y., DUAN, X., HU, J. & LIEBER, C. M. 2000. Doping and Electrical Transport in Silicon Nanowires. *The Journal of Physical Chemistry B,* 104**,** 5213-5216.

CUI, Y. & LIEBER, C. M. 2001. Functional nanoscale electronic devices assembled using silicon nanowire building blocks. *Science,* 291**,** 851-3.

CUI, Y., WEI, Q., PARK, H. & LIEBER, C. M. 2001. Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species. *Science,* 293**,** 1289-92.

CUZICK, J., OTTO, F., BARON, J. A., BROWN, P. H., BURN, J., GREENWALD, P., JANKOWSKI, J., LA VECCHIA, C., MEYSKENS, F., SENN, H. J. & THUN, M. 2009. Aspirin and non-steroidal anti-inflammatory drugs for cancer prevention: an international consensus statement. *The Lancet Oncology,* 10**,** 501-507.

CUZICK, J., THORAT, M. A., BOSETTI, C., BROWN, P. H., BURN, J., COOK, N. R., FORD, L. G., JACOBS, E. J., JANKOWSKI, J. A., LA VECCHIA, C., LAW, M., MEYSKENS, F., ROTHWELL, P. M., SENN, H. J. & UMAR, A. 2014. Estimates of benefits and harms of prophylactic use of aspirin in the general population. *Annals of Oncology,* 26**,** 47-57.

D'ANDREA, G., D'AMBROSIO, PERNA, R. D., CHETTA, P., TACROCE, M. S.-., BRANCACCIO, R., GRANDONE, E. & MARGAGLIONE, M. 2005. A polymorphism in the VKORC1 gene is associated with an interindividual variability in the dose-anticoagulant effect of warfarin. *Blood,* 105**,** 645–649.

DAI, H., MEYER, M., STEPANIANTS, S., ZIMAN, M. & STOUGHTON, R. 2002. Use of hybridization kinetics for differentiating specific from non-specific binding to oligonucleotide microarrays. *Nucleic Acids Research,* 30**,** e86-e86.

DANG, M. T., HAMBLETON, J. & KAYSER, S. R. 2005. The influence of ethnicity on warfarin dosage requirement. *Ann Pharmacother,* 39**,** 1008-12.

DAVIE, E. W., FUJIKAWA, K. & KISIEL, W. 1991. The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry,* 30**,** 10363-10370.

DAVIE, E. W. & RATNOFF, O. D. 1964. WATERFALL SEQUENCE FOR INTRINSIC BLOOD CLOTTING. *Science,* 145**,** 1310-2.

DE MUTSERT, R., SUN, Q., WILLETT, W. C., HU, F. B. & VAN DAM, R. M. 2014. Overweight in early adulthood, adult weight change, and risk of type 2 diabetes, cardiovascular diseases, and certain cancers in men: a cohort study. *Am J Epidemiol,* 179**,** 1353-65.

DENG, G., KAKAR, S., TANAKA, H., MATSUZAKI, K., MIURA, S., SLEISENGER, M. H. & KIM, Y. S. 2008. Proximal and distal colorectal cancers show distinct gene-specific methylation profiles and clinical and molecular characteristics. *Eur J Cancer,* 44**,** 1290-301.

DIN, F. V., DUNLOP, M. G. & STARK, L. A. 2004. Evidence for colorectal cancer cell specificity of aspirin effects on NF kappa B signalling and apoptosis. *Br J Cancer,* 91**,** 381-8.

DIXON, D. A., BLANCO, F. F., BRUNO, A. & PATRIGNANI, P. 2013. Mechanistic aspects of COX-2 expression in colorectal neoplasia. *Recent Results Cancer Res,* 191**,** 7-37.

DOUBENI, C. A., LAIYEMO, A. O., MAJOR, J. M., SCHOOTMAN, M., LIAN, M., PARK, Y., GRAUBARD, B. I., HOLLENBECK, A. R. & SINHA, R. 2012. Socioeconomic status and the risk of colorectal cancer: an analysis of more than a half million adults in the National Institutes of Health-AARP Diet and Health Study. *Cancer,* 118**,** 3636-44.

DOVIZIO, M., BRUNO, A., TACCONELLI, S. & PATRIGNANI, P. 2013. Mode of Action of Aspirin as a Chemopreventive Agent. *In:* CHAN, A. T. & DETERING, E. (eds.) *Prospects for Chemoprevention of Colorectal Neoplasia.* Springer Berlin Heidelberg.

DRAZEN, J. M., YANDAVA, C. N., DUBE, L., SZCZERBACK, N., HIPPENSTEEL, R., PILLARI, A., ISRAEL, E., SCHORK, N., SILVERMAN, E. S., KATZ, D. A. & DRAJESK, J. 1999. Pharmacogenetic association between ALOX5 promoter genotype and the response to anti-asthma treatment. *Nat Genet,* 22**,** 168-70.

DUNLOP, M. G., DOBBINS, S. E., FARRINGTON, S. M., JONES, A. M., PALLES, C., WHIFFIN, N., TENESA, A., SPAIN, S., BRODERICK, P., OOI, L. Y., DOMINGO, E., SMILLIE, C., HENRION, M., FRAMPTON, M., MARTIN, L., GRIMES, G., GORMAN, M., SEMPLE, C., MA, Y. P., BARCLAY, E., PRENDERGAST, J., CAZIER, J. B., OLVER, B., PENEGAR, S., LUBBE, S., CHANDER, I., CARVAJAL-CARMONA, L. G., BALLEREAU, S., LLOYD, A., VIJAYAKRISHNAN, J., ZGAGA, L., RUDAN, I., THEODORATOU, E., STARR, J. M., DEARY, I., KIRAC, I., KOVACEVIC, D., AALTONEN, L. A., RENKONEN-SINISALO, L., MECKLIN, J. P., MATSUDA, K., NAKAMURA, Y., OKADA, Y., GALLINGER, S., DUGGAN, D. J., CONTI, D., NEWCOMB, P., HOPPER, J., JENKINS, M. A., SCHUMACHER, F., CASEY, G., EASTON, D., SHAH, M., PHAROAH, P., LINDBLOM, A., LIU, T., SMITH, C. G., WEST, H., CHEADLE, J. P., MIDGLEY, R., KERR, D. J., CAMPBELL, H., TOMLINSON, I. P. & HOULSTON, R. S. 2012. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat Genet,* 44**,** 770-6.

DUPONT, I., BERTHOU, F., BODENEZ, P., BARDOU, L., GUIRRIEC, C., STEPHAN, N., DREANO, Y. & LUCAS, D. 1999. Involvement of cytochromes P-450 2E1 and 3A4 in the 5-hydroxylation of salicylate in humans. *Drug Metab Dispos,* 27**,** 322-6.

EBERHART, C. E., COFFEY, R. J., RADHIKA, A., GIARDIELLO, F. M., FERRENBACH, S. & DUBOIS, R. N. 1994. Up-regulation of cyclooxygenase 2 gene expression in human colorectal adenomas and adenocarcinomas. *Gastroenterology,* 107**,** 1183-8.

ECKMAN, M. H., ROSAND, J., GREENBERG, S. M. & GAGE, B. F. 2009. Cost-effectiveness of using pharmacogenetic information in warfarin dosing for patients with nonvalvular atrial fibrillation. *Ann Intern Med,* 150**,** 73-83.

EPSTEIN, R. S., MOYER, T. P., AUBERT, R. E., O'KANE, D. J., XIA, F., VERBRUGGE, R. R., GAGE, B. F. & TEAGARDEN, R. 2010. Warfarin Genotyping Reduces Hospitalization Rates. *Journal of the American College of Cardiology,* 55**,** 2804-2812.

FAGERNESS, J., FONSECA, E., HESS, G. P., SCOTT, R., GARDNER, K. R., KOFFLER, M., FAVA, M., PERLIS, R., BRENNAN, F. X. & LOMBARD, J. 2014. Pharmacogenetic-guided psychiatric intervention associated with increased adherence and cost savings. *Am J Manag Care,* 20**,** e146-56.

FEARON, E. R. & VOGELSTEIN, B. 1990. A genetic model for colorectal tumorigenesis. *Cell,* 61**,** 759-67.

FEDIRKO, V., TRAMACERE, I., BAGNARDI, V., ROTA, M., SCOTTI, L., ISLAMI, F., NEGRI, E., STRAIF, K., ROMIEU, I., LA VECCHIA, C., BOFFETTA, P. & JENAB, M. 2011. Alcohol drinking and colorectal cancer risk: an overall and dose-response meta-analysis of published studies. *Ann Oncol,* 22**,** 1958-72.

FERRANDEZ, A., PIAZUELO, E. & CASTELLS, A. 2012. Aspirin and the prevention of colorectal cancer. *Best Pract Res Clin Gastroenterol,* 26**,** 185-95.

FINK, S. P., YAMAUCHI, M., NISHIHARA, R., JUNG, S., KUCHIBA, A., WU, K., CHO, E., GIOVANNUCCI, E., FUCHS, C. S., OGINO, S., MARKOWITZ, S. D. & CHAN, A. T. 2014. Aspirin and the Risk of Colorectal Cancer in Relation to the Expression of 15-Hydroxyprostaglandin Dehydrogenase (HPGD). *Science Translational Medicine,* 6**,** 233re2.

FINKELMAN, B. S., GAGE, B. F., JOHNSON, J. A., BRENSINGER, C. M. & KIMMEL, S. E. 2011. Genetic Warfarin Dosing: Tables Versus Algorithms. *Journal of the American College of Cardiology,* 57**,** 612-618.

FISHEL, R., LESCOE, M. K., RAO, M. R., COPELAND, N. G., JENKINS, N. A., GARBER, J., KANE, M. & KOLODNER, R. 1993. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell,* 75**,** 1027-38.

FLOYD, C. N. & FERRO, A. 2014. Mechanisms of aspirin resistance. *Pharmacology & Therapeutics,* 141**,** 69-78.

FRANCIS, C. W. 2008. Warfarin: an historical perspective. *Hematology Am Soc Hematol Educ Program***,** 251.

FRENCH, D. J., ARCHARD, C. L., BROWN, T. & MCDOWELL, D. G. 2001. HyBeacon probes: a new tool for DNA sequence detection and allele discrimination. *Mol Cell Probes,* 15**,** 363-74.

FRENK, J., BOBADILLA, J., SEPUULVEDA, J. & CERVANTES, M. L. 1989. Health transition in middle-income countries: new challenges for health care. *Health Policy and Planning,* 4**,** 29-39.

FUSTER, V. & SWEENY, J. M. 2011. Aspirin: a historical and contemporary therapeutic overview. *Circulation,* 123**,** 768-78.

GAGE, B., EBY, C., JOHNSON, J., DEYCH, E., RIEDER, M., RIDKER, P., MILLIGAN, P., GRICE, G., LENZINI, P., RETTIE, A., AQUILANTE, C., GROSSO, L., MARSH, S., LANGAEE, T., FARNETT, L., VOORA, D., VEENSTRA, D., GLYNN, R., BARRETT, A. & MCLEOD, H. 2008. Use of Pharmacogenetic and Clinical Factors to Predict the Therapeutic Dose of Warfarin. *Nature* 84**,** 10-15.

GAIKWAD, T., GHOSH, K., KULKARNI, B., KULKARNI, V., ROSS, C. & SHETTY, S. 2013. Influence of CYP2C9 and VKORC1 gene polymorphisms on warfarin dosage, over anticoagulation and other adverse outcomes in Indian population. *Eur J Pharmacol,* 710**,** 80-4.

GANN, P. H., MANSON, J. E., GLYNN, R. J., BURING, J. E. & HENNEKENS, C. H. 1993. Low-Dose Aspirin and Incidence of Colorectal Tumors in a Randomized Trial. *Journal of the National Cancer Institute,* 85**,** 1220-1224.

GAO, A., LU, N., DAI, P., LI, T., PEI, H., GAO, X., GONG, Y., WANG, Y. & FAN, C. 2011. Silicon-nanowire-based CMOS-compatible field-effect transistor nanosensors for ultrasensitive electrical detection of nucleic acids. *Nano Lett,* 11**,** 3974-8.

GAO, Z., AGARWAL, A., TRIGG, A. D., SINGH, N., FANG, C., TUNG, C. H., FAN, Y., BUDDHARAJU, K. D. & KONG, J. 2007. Silicon nanowire arrays for label-free detection of DNA. *Anal Chem,* 79**,** 3291-7.

GAY, L. J. & FELDING-HABERMANN, B. 2011. Contribution of platelets to tumour metastasis. *Nat Rev Cancer,* 11**,** 123-34.

GAZIANO, T. A. 2005. Cardiovascular Disease in the Developing World and Its Cost-Effective Management. *Circulation,* 112**,** 3547-3553.

GERHARDT, C. H. 1853. Untersuchungen über die wasserfreien organischen Säuren. *Justus Liebigs Annalen der Chemie,* 87**,** 149-179.

GIACOSA, A., FRANCESCHI, S., LA VECCHIA, C., FAVERO, A. & ANDREATTA, R. 1999. Energy intake, overweight, physical exercise and colorectal cancer risk. *Eur J Cancer Prev,* 8 Suppl 1**,** S53-60.

GIBRIEL, A. 2014. Effect of Target Length on Specificity and Sensitivity of Oligonucleotide Microarrays: A Comparison between Dendrimer and Modified PCR based Labelling Methods. *Open Biochem J,* 8**,** 11-20.

GIRI, A. K., KHAN, N. M., BASU, A., TANDON, N., SCARIA, V. & BHARADWAJ, D. 2014. Pharmacogenetic landscape of clopidogrel in north Indians suggest distinct interpopulation differences in allele frequencies. *Pharmacogenomics,* 15**,** 643-53.

GIRMA, J.-P., MEYER, D., VERWEIJ, C. L., PANNEKOEK, H. & SIXMA, J. J. 1987. Structure-function relationship of human von Willebrand factor. *Blood,* 70**,** 605-611.

GOUGH, A. C., MILES, J. S., SPURR, N. K., MOSS, J. E., GAEDIGK, A., EICHELBAUM, M. & WOLF, C. R. 1990. Identification of the primary gene defect at the cytochrome P450 CYP2D locus. *Nature,* 347**,** 773-6.

GRESHAM, D., CURRY, B., WARD, A., GORDON, D. B., BRIZUELA, L., KRUGLYAK, L. & BOTSTEIN, D. 2010. Optimized detection of sequence variation in heterozygous genomes using DNA microarrays with isothermal-melting probes. *Proc Natl Acad Sci U S A,* 107**,** 1482-7.

HAINING, R. L., HUNTER, A. P., VERONESE, M. E., TRAGER, W. F. & RETTIE, A. E. 1996. Allelic variants of human cytochrome P450 2C9: baculovirus-mediated expression, purification, structural characterization, substrate stereoselectivity, and prochiral selectivity of the wild-type and I359L mutant forms. *Arch Biochem Biophys,* 333**,** 447-58.

HANLY, P., SOERJOMATARAM, I. & SHARP, L. 2015. Measuring the societal burden of cancer: the cost of lost productivity due to premature cancer-related mortality in Europe. *Int J Cancer,* 136**,** E136-45.

HANNAN, L. M., JACOBS, E. J. & THUN, M. J. 2009. The association between cigarette smoking and risk of colorectal cancer in a large prospective cohort from the United States. *Cancer Epidemiol Biomarkers Prev,* 18**,** 3362-7.

HASHIM, Y. & SIDEK, O. 2012. Effect of temperature on the characteristics of silicon nanowire transistor. *J Nanosci Nanotechnol,* 12**,** 7849-52.

HEIMARK, L. D., WIENKERS, L., KUNZE, K., GIBALDI, M., EDDY, A. C., TRAGER, W. F., O'REILLY, R. A. & GOULART, D. A. 1992. The mechanism of the interaction between amiodarone and warfarin in humans. *Clin Pharmacol Ther,* 51**,** 398-407.

HENRY, O. Y. F. & O'SULLIVAN, C. K. 2012. Rapid DNA hybridization in microfluidics. *TrAC Trends in Analytical Chemistry,* 33**,** 9-22.

HIGGINS, J. P., THOMPSON, S. G., DEEKS, J. J. & ALTMAN, D. G. 2003. Measuring inconsistency in meta-analyses. *BMJ,* 327**,** 557-60.

HOEIJMAKERS, J. H. 2001. Genome maintenance mechanisms for preventing cancer. *Nature,* 411**,** 366-74.

HOMMA, S., THOMPSON, J. L. P., PULLICINO, P. M., LEVIN, B., FREUDENBERGER, R. S., TEERLINK, J. R., AMMON, S. E., GRAHAM, S., SACCO, R. L., MANN, D. L., MOHR, J. P., MASSIE, B. M., LABOVITZ, A. J., ANKER, S. D., LOK, D. J., PONIKOWSKI, P., ESTOL, C. J., LIP, G. Y. H., DI TULLIO, M. R., SANFORD, A. R., MEJIA, V., GABRIEL, A. P., DEL VALLE, M. L. & BUCHSBAUM, R. 2012. Warfarin and Aspirin in Patients with Heart Failure and Sinus Rhythm. *New England Journal of Medicine,* 366**,** 1859-1869.

HUANG, E. S., STRATE, L. L., HO, W. W., LEE, S. S. & CHAN, A. T. 2011. Long-term use of aspirin and the risk of gastrointestinal bleeding. *Am J Med,* 124**,** 426-33.

HUBNER, R. A., MUIR, K. R., LIU, J. F., LOGAN, R. F., GRAINGE, M., ARMITAGE, N., SHEPHERD, V., POPAT, S. & HOULSTON, R. S. 2006. Genetic variants of UGT1A6 influence risk of colorectal adenoma recurrence. *Clin Cancer Res,* 12**,** 6585-9.

HUBNER, R. A., MUIR, K. R., LIU, J. F., LOGAN, R. F., GRAINGE, M. J. & HOULSTON, R. S. 2008. Ornithine decarboxylase G316A genotype is prognostic for colorectal adenoma recurrence and predicts efficacy of aspirin chemoprevention. *Clin Cancer Res,* 14**,** 2303-9.

HUTT, A. J., CALDWELL, J. & SMITH, R. L. 1986. The metabolism of aspirin in man: a population study. *Xenobiotica,* 16**,** 239-49.

HYLEK, E. M., SKATES, S. J., SHEEHAN, M. A. & SINGER, D. E. 1996. An analysis of the lowest effective intensity of prophylactic anticoagulation for patients with nonrheumatic atrial fibrillation. *N Engl J Med,* 335**,** 540-6.

IMAI, K. & YAMAMOTO, H. 2008. Carcinogenesis and microsatellite instability: the interrelationship between genetics and epigenetics. *Carcinogenesis,* 29**,** 673-80.

ISHIKAWA, H., WAKABAYASHI, K., SUZUKI, S., MUTOH, M., HIRATA, K., NAKAMURA, T., TAKEYAMA, I., KAWANO, A., GONDO, N., ABE, T., TOKUDOME, S., GOTO, C., MATSUURA, N. & SAKAI, T. 2013. Preventive effects of low-dose aspirin on colorectal adenoma growth in patients with familial adenomatous polyposis: double-blind, randomized clinical trial. *Cancer Medicine,* 2**,** 50-56.

JEMAL, A., CENTER, M. M., DESANTIS, C. & WARD, E. M. 2010. Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiol Biomarkers Prev,* 19**,** 1893-907.

JONES, M., MCEWAN, P., MORGAN, C. L., PETERS, J. R., GOODFELLOW, J. & CURRIE, C. J. 2005. Evaluation of the pattern of treatment, level of anticoagulation control, and outcome of treatment with warfarin in patients with non-valvar atrial fibrillation: a record linkage study in a large British population. *Heart,* 91**,** 472-7.

JORGENSEN, A. L., FITZGERALD, R. J., OYEE, J., PIRMOHAMED, M. & WILLIAMSON, P. R. 2012. Influence of CYP2C9 and VKORC1 on patient response to warfarin: a systematic review and meta-analysis. *PLoS One,* 7**,** e44064.

KAAKS, R., SLIMANI, N. & RIBOLI, E. 1997. Pilot phase studies on the accuracy of dietary intake measurements in the EPIC project: overall evaluation of results. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol,* 26 Suppl 1**,** S26-36.

KAMINSKY, L. S. & ZHANG, Z.-Y. 1997. Human P450 Metabollism of Warfarin. *Pharmacology and therapeutics,* 73**,** 67-74.

KHANDEKAR, M. J., COHEN, P. & SPIEGELMAN, B. M. 2011. Molecular mechanisms of cancer development in obesity. *Nat Rev Cancer,* 11**,** 886-895.

KIMMEL, S. E., FRENCH, B., KASNER, S. E., JOHNSON, J. A., ANDERSON, J. L., GAGE, B. F., ROSENBERG, Y. D., EBY, C. S., MADIGAN, R. A., MCBANE, R. B., ABDEL-RAHMAN, S. Z., STEVENS, S. M., YALE, S., MOHLER, E. R., FANG, M. C., SHAH, V., HORENSTEIN, R. B., LIMDI, N. A., MULDOWNEY, J. A. S., GUJRAL, J., DELAFONTAINE, P., DESNICK, R. J., ORTEL, T. L., BILLETT, H. H., PENDLETON, R. C., GELLER, N. L., HALPERIN, J. L., GOLDHABER, S. Z., CALDWELL, M. D., CALIFF, R. M. & ELLENBERG, J. H. 2013. A Pharmacogenetic versus a Clinical Algorithm for Warfarin Dosing. *New England Journal of Medicine,* 369**,** 2283-2293.

KING, C. R., PORCHE-SORBET, R. M., GAGE, B. F., RIDKER, P. M., RENAUD, Y., PHILLIPS, M. S. & EBY, C. 2008. Performance of commercial platforms for rapid genotyping of polymorphisms affecting warfarin dose. *Am J Clin Pathol,* 129**,** 876-83.

KIRKWOOD, T. B. 1983. Calibration of reference thromboplastins and standardisation of the prothrombin time ratio. *Thromb Haemost,* 49**,** 238-44.

KLEIN, T. E., ALTMAN, R. B., ERIKSSON, N., GAGE, B. F., KIMMEL, S. E., LEE, M. T., LIMDI, N. A., PAGE, D., RODEN, D. M., WAGNER, M. J., CALDWELL, M. D. & JOHNSON, J. A. 2009. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N Engl J Med,* 360**,** 753-64.

KLEINSTEIN, S. E., HEATH, L., MAKAR, K. W., POOLE, E. M., SEUFERT, B. L., SLATTERY, M. L., XIAO, L., DUGGAN, D. J., HSU, L., CURTIN, K., KOEPL, L., MUEHLING, J., TAVERNA, D., CAAN, B. J., CARLSON, C. S., POTTER, J. D. & ULRICH, C. M. 2013. Genetic variation in the lipoxygenase pathway and risk of colorectal neoplasia. *Genes Chromosomes Cancer,* 52**,** 437-49.

KOLTAI, H. & WEINGARTEN-BARON, C. 2008. Specificity of DNA microarray hybridization: characterization, effectors and approaches for data correction. *Nucleic Acids Res,* 36**,** 2395-405.

KRAUS, S., HUMMLER, S., TORIOLA, A. T., POOLE, E. M., SCHERER, D., KOTZMANN, J., MAKAR, K. W., KAZANOV, D., GALAZAN, L., NAUMOV, I., COGHILL, A. E., DUGGAN, D., GIGIC, B., ARBER, N. & ULRICH, C. M. 2013. Impact of genetic polymorphisms on adenoma recurrence and toxicity in a COX2 inhibitor (celecoxib) trial: results from a pilot study. *Pharmacogenet Genomics,* 23**,** 428-37.

KRISHNASWAMY, S., HAO, Q., AL-ROHAIMI, A., HESSE, L. M., VON MOLTKE, L. L., GREENBLATT, D. J. & COURT, M. H. 2005. UDP glucuronosyltransferase (UGT) 1A6 pharmacogenetics: II. Functional impact of the three most common nonsynonymous UGT1A6 polymorphisms (S7A, T181A, and R184S). *J Pharmacol Exp Ther,* 313**,** 1340-6.

KRYNETSKI, E. Y. & EVANS, W. E. 1999. Pharmacogenetics as a molecular basis for individualized drug therapy: the thiopurine S-methyltransferase paradigm. *Pharm Res,* 16**,** 342-9.

KUBO, T., KIM, S. R., SAI, K., SAITO, Y., NAKAJIMA, T., MATSUMOTO, K., SAITO, H., SHIRAO, K., YAMAMOTO, N., MINAMI, H., OHTSU, A., YOSHIDA, T., SAIJO, N., OHNO, Y., OZAWA, S. & SAWADA, J. 2005. Functional characterization of three naturally occurring single nucleotide polymorphisms in the CES2 gene encoding carboxylesterase 2 (HCE-2). *Drug Metab Dispos,* 33**,** 1482-7.

KUEHL, G. E., BIGLER, J., POTTER, J. D. & LAMPE, J. W. 2006. Glucuronidation of the aspirin metabolite salicylic acid by expressed UDP-glucuronosyltransferases and human liver microsomes. *Drug Metab Dispos,* 34**,** 199-202.

KUNE, G. A., KUNE, S. & WATSON, L. F. 1988. Colorectal cancer risk, chronic illnesses, operations, and medications: case control results from the Melbourne Colorectal Cancer Study. *Cancer Res,* 48**,** 4399-404.

KUNZE, K. L., WIENKERS, L. C., THUMMEL, K. E. & TRAGER, W. F. 1996. Warfarin-fluconazole. I. Inhibition of the human cytochrome P450-dependent metabolism of warfarin by fluconazole: in vitro studies. *Drug Metab Dispos,* 24**,** 414-21.

LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A., HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM, J., MESIROV, J. P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R., SHERIDAN, A., SOUGNEZ, C., STANGE-THOMANN, N., STOJANOVIC, N., SUBRAMANIAN, A., WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J., CLEE, C., CARTER, N., COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I., DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S., HUNT, A., JONES, M., LLOYD, C., MCMURRAY, A., MATTHEWS, L., MERCER, S., MILNE, S., MULLIKIN, J. C., MUNGALL, A., PLUMB, R., ROSS, M., SHOWNKEEN, R., SIMS, S., WATERSTON, R. H., WILSON, R. K., HILLIER, L. W., MCPHERSON, J. D., MARRA, M. A., MARDIS, E. R., FULTON, L. A., CHINWALLA, A. T., PEPIN, K. H., GISH, W. R., CHISSOE, S. L., WENDL, M. C., DELEHAUNTY, K. D., MINER, T. L., DELEHAUNTY, A., KRAMER, J. B., COOK, L. L., FULTON, R. S., JOHNSON, D. L., MINX, P. J., CLIFTON, S. W., HAWKINS, T., BRANSCOMB, E., PREDKI, P., RICHARDSON, P., WENNING, S., SLEZAK, T., DOGGETT, N., CHENG, J. F., OLSEN, A., LUCAS, S., ELKIN, C., UBERBACHER, E., FRAZIER, M., et al. 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

LANE, S., AL-ZUBIEDI, S., JORGENSEN, A. L., AARONS, L. & PIRMOHAMED, M. 2011. The population pharmacokinetics of R- and S-warfarin: effect of genetic and clinical factors. *British Journal of Clinical Pharmacology,* 73**,** 66-76.

LÄPPLE, F., VON RICHTER, O., FROMM, M. F., RICHTER, T., THON, K. P., WISSER, H., GRIESE, E.-U., EICHELBAUM, M. & KIVISTÖ, K. T. 2003. Differential expression and function of CYP2C isoforms in human intestine and liver. *Pharmacogenetics and Genomics,* 13**,** 565-575.

LARSSON, S. C. & WOLK, A. 2007. Obesity and colon and rectal cancer risk: a meta-analysis of prospective studies. *Am J Clin Nutr,* 86**,** 556-65.

LEE, H. H., SMOOT, J., MCMURRAY, Z., STAHL, D. A. & YAGER, P. 2006a. Recirculating flow accelerates DNA microarray hybridization in a microfluidic device. *Lab Chip,* 6**,** 1163-70.

LEE, M. T. M. & KLEIN, T. E. 2013. Pharmacogenetics of warfarin: challenges and opportunities. *Journal of Human Genetics,* 58**,** 334-338.

LEE, S. C., NG, S. S., OLDENBURG, J., CHONG, P. Y., ROST, S., GUO, J. Y., YAP, H. L., RANKIN, S. C., KHOR, H. B., YEO, T. C., NG, K. S., SOONG, R. & GOH, B. C. 2006b. Interethnic variability of warfarin maintenance requirement is explained by VKORC1 genotype in an Asian population. *Clin Pharmacol Ther,* 79**,** 197-205.

LENZINI, P., WADELIUS, M., KIMMEL, S., ANDERSON, J. L., JORGENSEN, A. L., PIRMOHAMED, M., CALDWELL, M. D., LIMDI, N., BURMESTER, J. K., DOWD, M. B., ANGCHAISUKSIRI, P., BASS, A. R., CHEN, J., ERIKSSON, N., RANE, A., LINDH, J. D., CARLQUIST, J. F., HORNE, B. D., GRICE, G., MILLIGAN, P. E., EBY, C., SHIN, J., KIM, H., KURNIK, D., STEIN, C. M., MCMILLIN, G., PENDLETON, R. C., BERG, R. L., DELOUKAS, P. & GAGE, B. F. 2010. Integration of genetic, clinical, and INR data to refine warfarin dosing. *Clin Pharmacol Ther,* 87**,** 572-8.

LETOWSKI, J., BROUSSEAU, R. & MASSON, L. 2004. Designing better probes: effect of probe size, mismatch position and number on hybridization in DNA oligonucleotide microarrays. *J Microbiol Methods,* 57**,** 269-78.

LI, J., ZHANG, Y., TO, S., YOU, L. & SUN, Y. 2011. Effect of Nanowire Number, Diameter, and Doping Density on Nano-FET Biosensor Sensitivity. *ACS Nano,* 5**,** 6661-6668.

LI, Z., CHEN, Y., LI, X., KAMINS, T. I., NAUKA, K. & WILLIAMS, R. S. 2004. Sequence-Specific Label-Free DNA Sensors Based on Silicon Nanowires. *Nano Letters,* 4**,** 245-247.

LIAO, X., LOCHHEAD, P., NISHIHARA, R., MORIKAWA, T., KUCHIBA, A., YAMAUCHI, M., IMAMURA, Y., QIAN, Z. R., BABA, Y., SHIMA, K., SUN, R., NOSHO, K., MEYERHARDT, J. A., GIOVANNUCCI, E., FUCHS, C. S., CHAN, A. T. & OGINO, S. 2012. Aspirin Use, Tumor PIK3CA Mutation, and Colorectal-Cancer Survival. *New England Journal of Medicine,* 367**,** 1596-1606.

LIMDI, N. A., BROWN, T. M., YAN, Q., THIGPEN, J. L., SHENDRE, A., LIU, N., HILL, C. E., ARNETT, D. K. & BEASLEY, T. M. 2015. Race influences warfarin dose changes associated with genetic factors. *Blood*.

LINKINS, L. A., CHOI, P. T. & DOUKETIS, J. D. 2003. Clinical impact of bleeding in patients taking oral anticoagulant therapy for venous thromboembolism: a meta-analysis. *Ann Intern Med,* 139**,** 893-900.

LOGAN, R. F. A., GRAINGE, M. J., SHEPHERD, V. C., ARMITAGE, N. C. & MUIR, K. R. 2008. Aspirin and Folic Acid for the Prevention of Recurrent Colorectal Adenomas. *Gastroenterology,* 134**,** 29-38.

LUBETSKY, A., DEKEL-STERN, E., CHETRIT, A., LUBIN, F. & HALKIN, H. 1999. Vitamin K intake and sensitivity to warfarin in patients consuming regular diets. *Thromb Haemost,* 81**,** 396-9.

LUM, A. & LE MARCHAND, L. 1998. A simple mouthwash method for obtaining genomic DNA in molecular epidemiological studies. *Cancer Epidemiol Biomarkers Prev,* 7**,** 719-24.

LYAMICHEV, V., MAST, A. L., HALL, J. G., PRUDENT, J. R., KAISER, M. W., TAKOVA, T., KWIATKOWSKI, R. W., SANDER, T. J., DE ARRUDA, M., ARCO, D. A., NERI, B. P. & BROW, M. A. 1999. Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat Biotechnol,* 17**,** 292-6.

MACFARLANE, R. G. 1964. An Enzyme Cascade in the Blood Clotting Mechanism, and its Function as a Biochemical Amplifier. *Nature,* 202**,** 498-499.

MAHGOUB, A., IDLE, J. R., DRING, L. G., LANCASTER, R. & SMITH, R. L. 1977. Polymorphic hydroxylation of Debrisoquine in man. *Lancet,* 2**,** 584-6.

MAKAR, K. W., POOLE, E. M., RESLER, A. J., SEUFERT, B., CURTIN, K., KLEINSTEIN, S. E., DUGGAN, D., KULMACZ, R. J., HSU, L., WHITTON, J., CARLSON, C. S., RIMORIN, C. F., CAAN, B. J., BARON, J. A., POTTER, J. D., SLATTERY, M. L. & ULRICH, C. M. 2013. COX-1 (PTGS1) and COX-2 (PTGS2) polymorphisms, NSAID interactions, and risk of colon and rectal cancers in two independent populations. *Cancer Causes Control,* 24**,** 2059-75.

MANN, K. G., NESHEIM, M. E., TRACY, P. B., HIBBARD, L. S. & BLOOM, J. W. 1982. Assembly of the Prothrombinase Complex. *Biophysical Journal,* 37**,** 106-107.

MANT, J., HOBBS, F. D., FLETCHER, K., ROALFE, A., FITZMAURICE, D., LIP, G. Y. & MURRAY, E. 2007. Warfarin versus aspirin for stroke prevention in an elderly community population with atrial fibrillation (the Birmingham Atrial Fibrillation Treatment of the Aged Study, BAFTA): a randomised controlled trial. *Lancet,* 370**,** 493-503.

MARKOWITZ, S. D. & BERTAGNOLLI, M. M. 2009. Molecular Basis of Colorectal Cancer. *New England Journal of Medicine,* 361**,** 2449-2460.

MARKOWITZ, S. D., DAWSON, D. M., WILLIS, J. & WILLSON, J. K. 2002. Focus on colon cancer. *Cancer Cell,* 1**,** 233-6.

MARSHALL, A. 1997. Genset-Abbott deal heralds pharmacogenomics era. *Nat Biotechnol,* 15**,** 829-30.

MARTÍNEZ, M. E., O'BRIEN, T. G., FULTZ, K. E., BABBAR, N., YERUSHALMI, H., QU, N., GUO, Y., BOORMAN, D., EINSPAHR, J., ALBERTS, D. S. & GERNER, E. W. 2003. Pronounced reduction in adenoma recurrence associated with aspirin use and a polymorphism in the ornithine decarboxylase gene. *Proceedings of the National Academy of Sciences,* 100**,** 7859-7864.

MCCARTHY, J. J. & HILFIKER, R. 2000. The use of single-nucleotide polymorphism maps in pharmacogenomics. *Nat Biotech,* 18**,** 505-508.

MCCARTY, M. F. & BLOCK, K. I. 2006. Preadministration of high-dose salicylates, suppressors of NF-kappaB activation, may increase the chemosensitivity of many cancers: an example of proapoptotic signal modulation therapy. *Integr Cancer Ther,* 5**,** 252-68.

MCDONALD, M. G., AU, N. T., WITTKOWSKY, A. K. & RETTIE, A. E. 2012. Warfarin-amiodarone drug-drug interactions: determination of [I](u)/K(I,u) for amiodarone and its plasma metabolites. *Clin Pharmacol Ther,* 91**,** 709-17.

MCDONALD, M. G., RIEDER, M. J., NAKANO, M., HSIA, C. K. & RETTIE, A. E. 2009a. CYP4F2 Is a Vitamin K1 Oxidase: An Explanation for Altered Warfarin Dose in Carriers of the V433M Variant. *Molecular Pharmacology,* 75**,** 1337-1346.

MCDONALD, M. G., RIEDER, M. J., NAKANO, M., HSIA, C. K. & RETTIE, A. E. 2009b. CYP4F2 is a vitamin K1 oxidase: An explanation for altered warfarin dose in carriers of the V433M variant. *Mol Pharmacol,* 75**,** 1337-46.

MCQUAIN, M. K., SEALE, K., PEEK, J., FISHER, T. S., LEVY, S., STREMLER, M. A. & HASELTON, F. R. 2004. Chaotic mixer improves microarray hybridization. *Anal Biochem,* 325**,** 215-26.

MCWILLIAMS, A., LUTTER, R. & NARDINELLI, C. 2010. Health care savings from personalized medicine using genetic testing: The case of warfarin. AEI-Brookings Joint Center for Regulatory Studies.

MEGA, J. L., WALKER, J. R., RUFF, C. T., VANDELL, A. G., NORDIO, F., DEENADAYALU, N., MURPHY, S. A., LEE, J., MERCURI, M. F., GIUGLIANO, R. P., ANTMAN, E. M., BRAUNWALD, E. & SABATINE, M. S. 2015. Genetics and the clinical response to warfarin and edoxaban: findings from the randomised, double-blind ENGAGE AF-TIMI 48 trial. *The Lancet,* 385**,** 2280-2287.

MEIER, P. J., MUELLER, H. K., DICK, B. & MEYER, U. A. 1983. Hepatic monooxygenase activities in subjects with a genetic defect in drug oxidation. *Gastroenterology,* 85**,** 682-92.

MEYER, U. A. 2004. Pharmacogenetics - five decades of therapeutic lessons from genetic diversity. *Nat Rev Genet,* 5**,** 669-676.

MOON, J. W., LEE, S. K., LEE, Y. W., LEE, J. O., KIM, N., LEE, H. J., SEO, J. S., KIM, J., KIM, H. S. & PARK, S. H. 2014. Alcohol induces cell proliferation via hypermethylation of ADHFE1 in colorectal cancer cells. *BMC Cancer,* 14**,** 377.

MORAN, A., ORTEGA, P., DE JUAN, C., FERNANDEZ-MARCELO, T., FRIAS, C., SANCHEZ-PERNAUTE, A., TORRES, A. J., DIAZ-RUBIO, E., INIESTA, P. & BENITO, M. 2010. Differential colorectal carcinogenesis: Molecular basis and clinical relevance. *World J Gastrointest Oncol,* 2**,** 151-8.

MOUTSIANAS, L. & MORRIS, A. P. 2014. Methodology for the analysis of rare genetic variation in genome-wide association and re-sequencing studies of complex human traits. *Briefings in Functional Genomics,* 13**,** 362-370.

MOZAFFARIAN, D., BENJAMIN, E. J., GO, A. S., ARNETT, D. K., BLAHA, M. J., CUSHMAN, M., DE FERRANTI, S., DESPRÉS, J.-P., FULLERTON, H. J., HOWARD, V. J., HUFFMAN, M. D., JUDD, S. E., KISSELA, B. M., LACKLAND, D. T., LICHTMAN, J. H., LISABETH, L. D., LIU, S., MACKEY, R. H., MATCHAR, D. B., MCGUIRE, D. K., MOHLER, E. R., MOY, C. S., MUNTNER, P., MUSSOLINO, M. E., NASIR, K., NEUMAR, R. W., NICHOL, G., PALANIAPPAN, L., PANDEY, D. K., REEVES, M. J., RODRIGUEZ, C. J., SORLIE, P. D., STEIN, J., TOWFIGHI, A., TURAN, T. N., VIRANI, S. S., WILLEY, J. Z., WOO, D., YEH, R. W. & TURNER, M. B. 2015. Executive Summary: Heart Disease and Stroke Statistics—2015 Update: A Report From the American Heart Association. *Circulation,* 131**,** 434-441.

NAGAR, S., ZALATORIS, J. J. & BLANCHARD, R. L. 2004. Human UGT1A6 pharmacogenetics: identification of a novel SNP, characterization of allele frequencies and functional analysis of recombinant allozymes in human liver tissue and in cultured cells. *Pharmacogenetics,* 14**,** 487-99.

NAN, H., HUTTER, C. M., LIN, Y. & ET AL. 2015. Association of aspirin and nsaid use with risk of colorectal cancer according to genetic variants. *JAMA,* 313**,** 1133-1142.

NAN, H., MORIKAWA, T., SUURINIEMI, M., IMAMURA, Y., WERNER, L., KUCHIBA, A., YAMAUCHI, M., HUNTER, D. J., KRAFT, P., GIOVANNUCCI, E. L., FUCHS, C. S., OGINO, S., FREEDMAN, M. L. & CHAN, A. T. 2013. Aspirin use, 8q24 single nucleotide polymorphism rs6983267, and colorectal cancer according to CTNNB1 alterations. *J Natl Cancer Inst,* 105**,** 1852-61.

NATHISUWAN, S., DILOKTHORNSAKUL, P., CHAIYAKUNAPRUK, N., MORARAI, T., YODTING, T. & PIRIYACHANANUSORN, N. 2011. Assessing evidence of interaction between smoking and warfarin: a systematic review and meta-analysis. *Chest,* 139**,** 1130-9.

NCBI. 2014. *PubChem* [Online]. Available: http://pubchem.ncbi.nlm.nih.gov/compound/aspirin - section=Top.
NEBERT, D. W., ZHANG, G. & VESELL, E. S. 2008. From Human Genetics and Genomics to Pharmacogenetics and Pharmacogenomics: Past Lessons, Future Directions. *Drug metabolism reviews,* 40**,** 187-224.
NEWCOMB, P. A., BARON, J., COTTERCHIO, M., GALLINGER, S., GROVE, J., HAILE, R., HALL, D., HOPPER, J. L., JASS, J., LE MARCHAND, L., LIMBURG, P., LINDOR, N., POTTER, J. D., TEMPLETON, A. S., THIBODEAU, S. & SEMINARA, D. 2007. Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. *Cancer Epidemiol Biomarkers Prev,* 16**,** 2331-43.
NISHIHARA, R., LOCHHEAD, P., KUCHIBA, A., JUNG, S., YAMAUCHI, M., LIAO, X., IMAMURA, Y., QIAN, Z. R., MORIKAWA, T., WANG, M., SPIEGELMAN, D., CHO, E., GIOVANNUCCI, E., FUCHS, C. S., CHAN, A. T. & OGINO, S. 2013. Aspirin use and risk of colorectal cancer according to BRAF mutation status. *JAMA,* 309**,** 2563-71.
NOZAKI, D., KUNSTMANN, J., ZÖRGIEBEL, F., PREGL, S., BARABAN, L., WEBER, W., MIKOLAJICK, T. & CUNIBERTI, G. 2014. Ionic effects on the transport characteristics of nanowire-based FETs in a liquid environment. *Nano Research,* 7**,** 380-389.
ODEN, A., FAHLEN, M. & HART, R. G. 2006. Optimal INR for prevention of stroke and death in atrial fibrillation: a critical appraisal. *Thromb Res,* 117**,** 493-9.
ONG, F. S., DEIGNAN, J. L., KUO, J. Z., BERNSTEIN, K. E., ROTTER, J. I., GRODY, W. W. & DAS, K. 2012. Clinical utility of pharmacogenetic biomarkers in cardiovascular therapeutics: a challenge for clinical implementation. *Pharmacogenomics,* 13**,** 465-75.
PANIGRAHY, D., KAIPAINEN, A., GREENE, E. R. & HUANG, S. 2010. Cytochrome P450-derived eicosanoids: the neglected pathway in cancer. *Cancer Metastasis Rev,* 29**,** 723-35.
PATHI, S., JUTOORU, I., CHADALAPAKA, G., NAIR, V., LEE, S. O. & SAFE, S. 2012. Aspirin inhibits colon cancer cell and tumor growth and downregulates specificity protein (Sp) transcription factors. *PLoS One,* 7**,** e48208.
PATOLSKY, F., ZHENG, G. & LIEBER, C. M. 2006. Fabrication of silicon nanowire devices for ultrasensitive, label-free, real-time detection of biological and chemical species. *Nat Protoc,* 1**,** 1711-24.
PAVANI, A., NAUSHAD, S., MISHRA, R., MALEMPATI, A., PINJALA, R., KUMAR, T. & KUTALA, V. 2012. Retrospective evidence for clinical validity of expanded genetic model in warfarin dose optimization in a South Indian population. *Pharmacogenomics,* 13**,** 869-878.
PAVANI, A., RUPASREE, Y., NAUSHAD, S., KUMAR, T., MALEMPATI, A., MISHRA, R., PINJALA, R. & KUTALA, V. 2011. Optimization of warfarin dose by population-specific pharmacogenomic algorithm. *The Pharmacogenomics Journal***,** 1-6.
PAWLIK, T. M., RAUT, C. P. & RODRIGUEZ-BIGAS, M. A. 2004. Colorectal carcinogenesis: MSI-H versus MSI-L. *Dis Markers,* 20**,** 199-206.
PEDERSEN, A. K. & FITZGERALD, G. A. 1984. Dose-related kinetics of aspirin. Presystemic acetylation of platelet cyclooxygenase. *N Engl J Med,* 311**,** 1206-11.
PEGG, A. E., SHANTZ, L. M. & COLEMAN, C. S. 1994. Ornithine decarboxylase: structure, function and translational regulation. *Biochem Soc Trans,* 22**,** 846-52.
PERERA, M. A., CAVALLARI, L. H. & JOHNSON, J. A. 2014. Warfarin pharmacogenetics: an illustration of the importance of studies in minority populations. *Clin Pharmacol Ther,* 95**,** 242-4.
PERERA, M. A., CAVALLARI, L. H., LIMDI, N. A., GAMAZON, E. R., KONKASHBAEV, A., DANESHJOU, R., PLUZHNIKOV, A., CRAWFORD, D. C.,

WANG, J., LIU, N., TATONETTI, N., BOURGEOIS, S., TAKAHASHI, H., BRADFORD, Y., BURKLEY, B. M., DESNICK, R. J., HALPERIN, J. L., KHALIFA, S. I., LANGAEE, T. Y., LUBITZ, S. A., NUTESCU, E. A., OETJENS, M., SHAHIN, M. H., PATEL, S. R., SAGREIYA, H., TECTOR, M., WECK, K. E., RIEDER, M. J., SCOTT, S. A., WU, A. H., BURMESTER, J. K., WADELIUS, M., DELOUKAS, P., WAGNER, M. J., MUSHIRODA, T., KUBO, M., RODEN, D. M., COX, N. J., ALTMAN, R. B., KLEIN, T. E., NAKAMURA, Y. & JOHNSON, J. A. 2013a. Genetic variants associated with warfarin dose in African-American individuals: a genome-wide association study. *Lancet,* 382**,** 790-6.

PERERA, M. A., CAVALLARI, L. H., LIMDI, N. A., GAMAZON, E. R., KONKASHBAEV, A., DANESHJOU, R., PLUZHNIKOV, A., CRAWFORD, D. C., WANG, J., LIU, N., TATONETTI, N., BOURGEOIS, S., TAKAHASHI, H., BRADFORD, Y., BURKLEY, B. M., DESNICK, R. J., HALPERIN, J. L., KHALIFA, S. I., LANGAEE, T. Y., LUBITZ, S. A., NUTESCU, E. A., OETJENS, M., SHAHIN, M. H., PATEL, S. R., SAGREIYA, H., TECTOR, M., WECK, K. E., RIEDER, M. J., SCOTT, S. A., WU, A. H. B., BURMESTER, J. K., WADELIUS, M., DELOUKAS, P., WAGNER, M. J., MUSHIRODA, T., KUBO, M., RODEN, D. M., COX, N. J., ALTMAN, R. B., KLEIN, T. E., NAKAMURA, Y. & JOHNSON, J. A. 2013b. Genetic variants associated with warfarin dose in African-American individuals: a genome-wide association study. *The Lancet,* 382**,** 790-796.

PETERS, U., HUTTER, C. M., HSU, L., SCHUMACHER, F. R., CONTI, D. V., CARLSON, C. S., EDLUND, C. K., HAILE, R. W., GALLINGER, S., ZANKE, B. W., LEMIRE, M., RANGREJ, J., VIJAYARAGHAVAN, R., CHAN, A. T., HAZRA, A., HUNTER, D. J., MA, J., FUCHS, C. S., GIOVANNUCCI, E. L., KRAFT, P., LIU, Y., CHEN, L., JIAO, S., MAKAR, K. W., TAVERNA, D., GRUBER, S. B., RENNERT, G., MORENO, V., ULRICH, C. M., WOODS, M. O., GREEN, R. C., PARFREY, P. S., PRENTICE, R. L., KOOPERBERG, C., JACKSON, R. D., LACROIX, A. Z., CAAN, B. J., HAYES, R. B., BERNDT, S. I., CHANOCK, S. J., SCHOEN, R. E., CHANG-CLAUDE, J., HOFFMEISTER, M., BRENNER, H., FRANK, B., BEZIEAU, S., KURY, S., SLATTERY, M. L., HOPPER, J. L., JENKINS, M. A., LE MARCHAND, L., LINDOR, N. M., NEWCOMB, P. A., SEMINARA, D., HUDSON, T. J., DUGGAN, D. J., POTTER, J. D. & CASEY, G. 2012. Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Hum Genet,* 131**,** 217-34.

PETERSON, A. W., HEATON, R. J. & GEORGIADIS, R. M. 2001. The effect of surface probe density on DNA hybridization. *Nucleic Acids Research,* 29**,** 5163-5168.

PEYTAVI, R., RAYMOND, F. R., GAGNE, D., PICARD, F. J., JIA, G., ZOVAL, J., MADOU, M., BOISSINOT, K., BOISSINOT, M., BISSONNETTE, L., OUELLETTE, M. & BERGERON, M. G. 2005. Microfluidic device for rapid (<15 min) automated microarray hybridization. *Clin Chem,* 51**,** 1836-44.

PICOT, D., LOLL, P. J. & GARAVITO, R. M. 1994. The X-ray crystal structure of the membrane protein prostaglandin H2 synthase-1. *Nature,* 367**,** 243-9.

PIRMOHAMED, M. 2006. Warfarin: almost 60 years old and still causing problems. *British Journal of Clinical Pharmacology,* 62**,** 509-511.

PIRMOHAMED, M. 2011. Pharmacogenetics: past, present and future. *Drug Discov Today,* 16**,** 852-61.

PIRMOHAMED, M., BURNSIDE, G., ERIKSSON, N., JORGENSEN, A. L., TOH, C. H., NICHOLSON, T., KESTEVEN, P., CHRISTERSSON, C., WAHLSTRÖM, B., STAFBERG, C., ZHANG, J. E., LEATHART, J. B., KOHNKE, H., MAITLAND-VAN DER ZEE, A. H., WILLIAMSON, P. R., DALY, A. K., AVERY, P., KAMALI, F. & WADELIUS, M. 2013. A Randomized Trial of Genotype-Guided Dosing of Warfarin. *New England Journal of Medicine,* 369**,** 2294-2303.

PIRMOHAMED, M., JAMES, S., MEAKIN, S., GREEN, C., SCOTT, A. K., WALLEY, T. J., FARRAR, K., PARK, B. K. & BRECKENRIDGE, A. M. 2004. *Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients*.

PIRMOHAMED, M., KAMALI, F., DALY, A. K. & WADELIUS, M. 2015. Oral anticoagulation: a critique of recent advances and controversies. *Trends in Pharmacological Sciences,* 36**,** 153-163.

POIRIER, J., DELISLE, M. C., QUIRION, R., AUBERT, I., FARLOW, M., LAHIRI, D., HUI, S., BERTRAND, P., NALBANTOGLU, J. & GILFIX, B. M. 1995. Apolipoprotein E4 allele as a predictor of cholinergic deficits and treatment outcome in Alzheimer disease. *Proceedings of the National Academy of Sciences of the United States of America,* 92**,** 12260-12264.

POMERANTZ, M. M., AHMADIYEH, N., JIA, L., HERMAN, P., VERZI, M. P., DODDAPANENI, H., BECKWITH, C. A., CHAN, J. A., HILLS, A., DAVIS, M., YAO, K., KEHOE, S. M., LENZ, H. J., HAIMAN, C. A., YAN, C., HENDERSON, B. E., FRENKEL, B., BARRETINA, J., BASS, A., TABERNERO, J., BASELGA, J., REGAN, M. M., MANAK, J. R., SHIVDASANI, R., COETZEE, G. A. & FREEDMAN, M. L. 2009. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet,* 41**,** 882-4.

POPKIN, B. M. 2002. The shift in stages of the nutrition transition in the developing world differs from past experiences! *Public Health Nutr,* 5**,** 205-14.

POULSEN, L., SOE, M. J., SNAKENBORG, D., MOLLER, L. B. & DUFVA, M. 2008. Multi-stringency wash of partially hybridized 60-mer probes reveals that the stringency along the probe decreases with distance from the microarray surface. *Nucleic Acids Res,* 36**,** e132.

REDDY, K. S., SHAH, B., VARGHESE, C. & RAMADOSS, A. 2005. Responding to the threat of chronic diseases in India. *Lancet,* 366**,** 1746-1751.

REIDER, M., REINER, A. & RETTIE, A. 2007. gamma-Glutamyl carboxylase (GGCX) tagSNPs have limited utility for predicting warfarin maintenance dose. *Journal of Thrombosis and Haemostasis,* 5.

REIMERS, M. S., BASTIAANNET, E., LANGLEY, R. E., VAN EIJK, R., VAN VLIERBERGHE, R. L., LEMMENS, V. E., VAN HERK-SUKEL, M. P., VAN WEZEL, T., FODDE, R., KUPPEN, P. J., MORREAU, H., VAN DE VELDE, C. J. & LIEFERS, G. J. 2014. Expression of HLA class I antigen, aspirin use, and survival after a diagnosis of colon cancer. *JAMA Intern Med,* 174**,** 732-9.

RETTIE, A. E., WIENKERS, L. C., GONZALEZ, F. J., TRAGER, W. F. & KORZEKWA, K. R. 1994. Impaired (S)-warfarin metabolism catalysed by the R144C allelic variant of CYP2C9. *Pharmacogenetics,* 4**,** 39-42.

ROTHWELL, P. M., PRICE, J. F., FOWKES, F. G. R., ZANCHETTI, A., RONCAGLIONI, M. C., TOGNONI, G., LEE, R., BELCH, J. F. F., WILSON, M., MEHTA, Z. & MEADE, T. W. 2012a. Short-term effects of daily aspirin on cancer incidence, mortality, and non-vascular death: analysis of the time course of risks and benefits in 51 randomised controlled trials. *The Lancet,* 379**,** 1602-1612.

ROTHWELL, P. M., WILSON, M., PRICE, J. F., BELCH, J. F., MEADE, T. W. & MEHTA, Z. 2012b. Effect of daily aspirin on risk of cancer metastasis: a study of incident cancers during randomised controlled trials. *Lancet,* 379**,** 1591-601.

RUANO, G., SZAREK, B. L., VILLAGRA, D., GOROWSKI, K., KOCHERLA, M., SEIP, R. L., GOETHE, J. W. & SCHWARTZ, H. I. 2013. Length of psychiatric hospitalization is correlated with CYP2D6 functional status in inpatients with major depressive disorder. *Biomark Med,* 7**,** 429-39.

RUGGERI, Z. M. & ZIMMERMAN, T. S. 1987. von Willebrand factor and von Willebrand disease. *Blood,* 70**,** 895-904.

RYAN, B. M., ZANETTI, K. A., ROBLES, A. I., SCHETTER, A. J., GOODMAN, J., HAYES, R. B., HUANG, W. Y., GUNTER, M. J., YEAGER, M., BURDETTE, L., BERNDT, S. I. & HARRIS, C. C. 2014. Germline variation in NCF4, an innate immunity gene, is associated with an increased risk of colorectal cancer. *Int J Cancer,* 134**,** 1399-407.

SACHIDANANDAM, R., WEISSMAN, D., SCHMIDT, S. C., KAKOL, J. M., STEIN, L. D., MARTH, G., SHERRY, S., MULLIKIN, J. C., MORTIMORE, B. J., WILLEY, D. L., HUNT, S. E., COLE, C. G., COGGILL, P. C., RICE, C. M., NING, Z., ROGERS, J., BENTLEY, D. R., KWOK, P. Y., MARDIS, E. R., YEH, R. T., SCHULTZ, B., COOK, L., DAVENPORT, R., DANTE, M., FULTON, L., HILLIER, L., WATERSTON, R. H., MCPHERSON, J. D., GILMAN, B., SCHAFFNER, S., VAN ETTEN, W. J., REICH, D., HIGGINS, J., DALY, M. J., BLUMENSTIEL, B., BALDWIN, J., STANGE-THOMANN, N., ZODY, M. C., LINTON, L., LANDER, E. S. & ALTSHULER, D. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature,* 409**,** 928-33.

SAMUELS, Y., WANG, Z., BARDELLI, A., SILLIMAN, N., PTAK, J., SZABO, S., YAN, H., GAZDAR, A., POWELL, S. M., RIGGINS, G. J., WILLSON, J. K., MARKOWITZ, S., KINZLER, K. W., VOGELSTEIN, B. & VELCULESCU, V. E. 2004. High frequency of mutations of the PIK3CA gene in human cancers. *Science,* 304**,** 554.

SANDLER, R. S., HALABI, S., BARON, J. A., BUDINGER, S., PASKETT, E., KERESZTES, R., PETRELLI, N., PIPAS, J. M., KARP, D. D., LOPRINZI, C. L., STEINBACH, G. & SCHILSKY, R. 2003. A Randomized Trial of Aspirin to Prevent Colorectal Adenomas in Patients with Previous Colorectal Cancer. *New England Journal of Medicine,* 348**,** 883-890.

SCHERER, D., KOEPL, L. M., POOLE, E. M., BALAVARCA, Y., XIAO, L., BARON, J. A., HSU, L., COGHILL, A. E., CAMPBELL, P. T., KLEINSTEIN, S. E., FIGUEIREDO, J. C., LAMPE, J. W., BUCK, K., POTTER, J. D., KULMACZ, R. J., JENKINS, M. A., HOPPER, J. L., WIN, A. K., NEWCOMB, P. A., ULRICH, C. M. & MAKAR, K. W. 2014. Genetic variation in UGT genes modify the associations of NSAIDs with risk of colorectal cancer: colon cancer family registry. *Genes Chromosomes Cancer,* 53**,** 568-78.

SCOTT, S. A. 2011. Personalizing medicine with clinical pharmacogenetics. *Genet Med,* 13**,** 987-95.

SECKELER, M. & HOKE, T. 2011. The worldwide epidemiology of acute rheumatic fever and rheumatic heart disease. *Clinical Epidemiology,* 3**,** 67-84.

SEQUENOM 2004. Multiplexing the Homogeneous MassEXTEND Assay. http://www.sequenom.com. Sequenom.

SEUFERT, B. L., POOLE, E. M., WHITTON, J., XIAO, L., MAKAR, K. W., CAMPBELL, P. T., KULMACZ, R. J., BARON, J. A., NEWCOMB, P. A., SLATTERY, M. L., POTTER, J. D. & ULRICH, C. M. 2013. IkappaBKbeta and NFkappaB1, NSAID use and risk of colorectal cancer in the Colon Cancer Family Registry. *Carcinogenesis,* 34**,** 79-85.

SHARMA, N. P., DONG, L., YUAN, C., NOON, K. R. & SMITH, W. L. 2010. Asymmetric acetylation of the cyclooxygenase-2 homodimer by aspirin and its effects on the oxygenation of arachidonic, eicosapentaenoic, and docosahexaenoic acids. *Mol Pharmacol,* 77**,** 979-86.

SHARMA, V., KAUL, S., AL-HAZZANI, A., PRABHA, T. S., RAO, P. P., DADHEECH, S., JYOTHY, A. & MUNSHI, A. 2012. Association of C3435T multi drug resistance gene-1 polymorphism with aspirin resistance in ischemic stroke and its subtypes. *J Neurol Sci,* 315**,** 72-6.

SHASTRY, B. S. 2006. Pharmacogenetics and the concept of individualized medicine. *Pharmacogenomics J,* 6**,** 16-21.

SHOU, M., KORZEKWA, K. R., CRESPI, C. L., GONZALEZ, F. J. & GELBOIN, H. V. 1994. The role of 12 cDNA-expressed human, rodent, and rabbit cytochromes P450 in the metabolism of benzo[a]pyrene and benzo[a]pyrene trans-7,8-dihydrodiol. *Mol Carcinog,* 10**,** 159-68.

SHOU, M., KRAUSZ, K. W., GONZALEZ, F. J. & GELBOIN, H. V. 1996. Metabolic activation of the potent carcinogen dibenzo[a,h]anthracene by cDNA-expressed human cytochromes P450. *Arch Biochem Biophys,* 328**,** 201-7.

SIEGEL, R., NAISHADHAM, D. & JEMAL, A. 2012. Cancer statistics, 2012. *CA: A Cancer Journal for Clinicians,* 62**,** 10-29.

SILVERTHORN, D. U. 2009. *Human Physiology: An Integrated Approach,* San Francisco, USA, Pearson Education Inc.

SLATTERY, M. 2004. Physical Activity and Colorectal Cancer. *Sports Medicine,* 34**,** 239-252.

SLATTERY, M. L., CAAN, B. J., POTTER, J. D., BERRY, T. D., COATES, A., DUNCAN, D. & EDWARDS, S. L. 1997. Dietary energy sources and colon cancer risk. *Am J Epidemiol,* 145**,** 199-210.

SLATTERY, M. L., EDWARDS, S., CURTIN, K., MA, K., EDWARDS, R., HOLUBKOV, R. & SCHAFFER, D. 2003a. Physical Activity and Colorectal Cancer. *American Journal of Epidemiology,* 158**,** 214-224.

SLATTERY, M. L., LEVIN, T. R., MA, K., GOLDGAR, D., HOLUBKOV, R. & EDWARDS, S. 2003b. Family history and colorectal cancer: predictors of risk. *Cancer Causes Control,* 14**,** 879-87.

SPITZ, G. A., FURTADO, C. M., SOLA-PENNA, M. & ZANCAN, P. 2009. Acetylsalicylic acid and salicylic acid decrease tumor cell viability and glucose metabolism modulating 6-phosphofructo-1-kinase structure and activity. *Biochem Pharmacol,* 77**,** 46-53.

STAHNMANN, M. A., CAMPBELL, H. A. & LINK, K. P. 1941. Studies on the hemorrhagic sweet clover disease. V. Identification and synthesis of the hemmorhagic agent. *Journal of Biological Chemistry,* 138**,** 513-527.

STARK, L. A., DIN, F. V., ZWACKA, R. M. & DUNLOP, M. G. 2001. Aspirin-induced activation of the NF-kappaB signaling pathway: a novel mechanism for aspirin-mediated apoptosis in colon cancer cells. *FASEB J,* 15**,** 1273-5.

STROM, C. M., JANECZKO, R. A., ANDERSON, B., REDMAN, J., QUAN, F., BULLER, A., MCGINNISS, M. J. & SUN, W. M. 2005. Technical validation of a multiplex platform to detect thirty mutations in eight genetic diseases prevalent in individuals of Ashkenazi Jewish descent. *Genet Med,* 7**,** 633-9.

SUBRAMANIAN, M., AGRAWAL, V., SANDEE, D., TAM, H. K., MILLER, W. L. & TRACY, T. S. 2012. Effect of P450 oxidoreductase variants on the metabolism of model substrates mediated by CYP2C9.1, CYP2C9.2, and CYP2C9.3. *Pharmacogenet Genomics,* 22**,** 590-7.

SUR, I. K., HALLIKAS, O., VAHARAUTIO, A., YAN, J., TURUNEN, M., ENGE, M., TAIPALE, M., KARHU, A., AALTONEN, L. A. & TAIPALE, J. 2012. Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science,* 338**,** 1360-3.

TAKAHASHI, H., KASHIMA, T., NOMOTO, S., IWADE, K., TAINAKA, H., SHIMIZU, T., NOMIZO, Y., MURAMOTO, N., KIMURA, S. & ECHIZEN, H. 1998. Comparisons between in-vitro and in-vivo metabolism of (S)-warfarin: catalytic activities of cDNA-expressed CYP2C9, its Leu359 variant and their mixture versus unbound clearance in patients with the corresponding CYP2C9 genotypes. *Pharmacogenetics,* 8**,** 365-73.

TANG, M., MUKUNDAN, M., YANG, J., CHARPENTIER, N., LECLUYSE, E. L., BLACK, C., YANG, D., SHI, D. & YAN, B. 2006. Antiplatelet agents aspirin and clopidogrel are hydrolyzed by distinct carboxylesterases, and clopidogrel is transesterificated in the presence of ethyl alcohol. *J Pharmacol Exp Ther,* 319**,** 1467-76.

TATARUNAS, V., LESAUSKAITĖ, V., JAKUŠKA, P., VEIKUTIENĖ, A. & BENETIS, R. 2011. The Influence of CYP2C9 and VKORC1 Gene Polymorphisms on Optimal Warfarin Doses After Heart Valve Replacement. *Medicina,* 47**,** 25-30.

TAYLOR, D. W., BARNETT, H. J., HAYNES, R. B., FERGUSON, G. G., SACKETT, D. L., THORPE, K. E., SIMARD, D., SILVER, F. L., HACHINSKI, V., CLAGETT, G. P., BARNES, R. & SPENCE, J. D. 1999. Low-dose and high-dose acetylsalicylic acid for patients undergoing carotid endarterectomy: a randomised controlled trial. ASA and Carotid Endarterectomy (ACE) Trial Collaborators. *Lancet,* 353**,** 2179-84.

TENESA, A., FARRINGTON, S. M., PRENDERGAST, J. G., PORTEOUS, M. E., WALKER, M., HAQ, N., BARNETSON, R. A., THEODORATOU, E., CETNARSKYJ, R., CARTWRIGHT, N., SEMPLE, C., CLARK, A. J., REID, F. J., SMITH, L. A., KAVOUSSANAKIS, K., KOESSLER, T., PHAROAH, P. D., BUCH, S., SCHAFMAYER, C., TEPEL, J., SCHREIBER, S., VOLZKE, H., SCHMIDT, C. O., HAMPE, J., CHANG-CLAUDE, J., HOFFMEISTER, M., BRENNER, H., WILKENING, S., CANZIAN, F., CAPELLA, G., MORENO, V., DEARY, I. J., STARR, J. M., TOMLINSON, I. P., KEMP, Z., HOWARTH, K., CARVAJAL-CARMONA, L., WEBB, E., BRODERICK, P., VIJAYAKRISHNAN, J., HOULSTON, R. S., RENNERT, G., BALLINGER, D., ROZEK, L., GRUBER, S. B., MATSUDA, K., KIDOKORO, T., NAKAMURA, Y., ZANKE, B. W., GREENWOOD, C. M., RANGREJ, J., KUSTRA, R., MONTPETIT, A., HUDSON, T. J., GALLINGER, S., CAMPBELL, H. & DUNLOP, M. G. 2008. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet,* 40**,** 631-7.

THOMPSON, C. L., PLUMMER, S. J., MERKULOVA, A., CHENG, I., TUCKER, T. C., CASEY, G. & LI, L. 2009. No association between cyclooxygenase-2 and uridine diphosphate glucuronosyltransferase 1A6 genetic polymorphisms and colon cancer risk. *World J Gastroenterol,* 15**,** 2240-4.

TOMLINSON, I., WEBB, E., CARVAJAL-CARMONA, L., BRODERICK, P., KEMP, Z., SPAIN, S., PENEGAR, S., CHANDLER, I., GORMAN, M., WOOD, W., BARCLAY, E., LUBBE, S., MARTIN, L., SELLICK, G., JAEGER, E., HUBNER, R., WILD, R., ROWAN, A., FIELDING, S., HOWARTH, K., SILVER, A., ATKIN, W., MUIR, K., LOGAN, R., KERR, D., JOHNSTONE, E., SIEBER, O., GRAY, R., THOMAS, H., PETO, J., CAZIER, J. B. & HOULSTON, R. 2007. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet,* 39**,** 984-8.

TOMLINSON, I. P., WEBB, E., CARVAJAL-CARMONA, L., BRODERICK, P., HOWARTH, K., PITTMAN, A. M., SPAIN, S., LUBBE, S., WALTHER, A., SULLIVAN, K., JAEGER, E., FIELDING, S., ROWAN, A., VIJAYAKRISHNAN, J., DOMINGO, E., CHANDLER, I., KEMP, Z., QURESHI, M., FARRINGTON, S. M., TENESA, A., PRENDERGAST, J. G., BARNETSON, R. A., PENEGAR, S., BARCLAY, E., WOOD, W., MARTIN, L., GORMAN, M., THOMAS, H., PETO, J., BISHOP, D. T., GRAY, R., MAHER, E. R., LUCASSEN, A., KERR, D., EVANS, D. G., SCHAFMAYER, C., BUCH, S., VOLZKE, H., HAMPE, J., SCHREIBER, S., JOHN, U., KOESSLER, T., PHAROAH, P., VAN WEZEL, T., MORREAU, H., WIJNEN, J. T., HOPPER, J. L., SOUTHEY, M. C., GILES, G. G., SEVERI, G., CASTELLVI-BEL, S., RUIZ-PONTE, C., CARRACEDO, A., CASTELLS, A., FORSTI, A., HEMMINKI, K., VODICKA, P., NACCARATI, A., LIPTON, L., HO,

J. W., CHENG, K. K., SHAM, P. C., LUK, J., AGUNDEZ, J. A., LADERO, J. M., DE LA HOYA, M., CALDES, T., NIITTYMAKI, I., TUUPANEN, S., KARHU, A., AALTONEN, L., CAZIER, J. B., CAMPBELL, H., DUNLOP, M. G. & HOULSTON, R. S. 2008. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet,* 40**,** 623-30.

TURNER, F., SMITH, G., SACHSE, C., LIGHTFOOT, T., GARNER, R. C., WOLF, C. R., FORMAN, D., BISHOP, D. T. & BARRETT, J. H. 2004. Vegetable, fruit and meat consumption and potential risk modifying genes in relation to colorectal cancer. *Int J Cancer,* 112**,** 259-64.

TUUPANEN, S., TURUNEN, M., LEHTONEN, R., HALLIKAS, O., VANHARANTA, S., KIVIOJA, T., BJORKLUND, M., WEI, G., YAN, J., NIITTYMAKI, I., MECKLIN, J. P., JARVINEN, H., RISTIMAKI, A., DI-BERNARDO, M., EAST, P., CARVAJAL-CARMONA, L., HOULSTON, R. S., TOMLINSON, I., PALIN, K., UKKONEN, E., KARHU, A., TAIPALE, J. & AALTONEN, L. A. 2009. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet,* 41**,** 885-90.

UHLEN, M., FAGERBERG, L., HALLSTROM, B. M., LINDSKOG, C., OKSVOLD, P., MARDINOGLU, A., SIVERTSSON, A., KAMPF, C., SJOSTEDT, E., ASPLUND, A., OLSSON, I., EDLUND, K., LUNDBERG, E., NAVANI, S., SZIGYARTO, C. A., ODEBERG, J., DJUREINOVIC, D., TAKANEN, J. O., HOBER, S., ALM, T., EDQVIST, P. H., BERLING, H., TEGEL, H., MULDER, J., ROCKBERG, J., NILSSON, P., SCHWENK, J. M., HAMSTEN, M., VON FEILITZEN, K., FORSBERG, M., PERSSON, L., JOHANSSON, F., ZWAHLEN, M., VON HEIJNE, G., NIELSEN, J. & PONTEN, F. 2015. Proteomics. Tissue-based map of the human proteome. *Science,* 347**,** 1260419.

ULRICH, C. M., WHITTON, J., YU, J. H., SIBERT, J., SPARKS, R., POTTER, J. D. & BIGLER, J. 2005. PTGS2 (COX-2) -765G > C promoter variant reduces risk of colorectal adenoma among nonusers of nonsteroidal anti-inflammatory drugs. *Cancer Epidemiol Biomarkers Prev,* 14**,** 616-9.

UNIVERSIY, W. 2012. http://www.warfarindosing.org [Online].

VAIRAVAN, R. 2004. AutoGenomics, Inc. *Pharmacogenomics,* 5**,** 585-588.

VENTER, J. C., ADAMS, M. D., MYERS, E. W., LI, P. W., MURAL, R. J., SUTTON, G. G., SMITH, H. O., YANDELL, M., EVANS, C. A., HOLT, R. A., GOCAYNE, J. D., AMANATIDES, P., BALLEW, R. M., HUSON, D. H., WORTMAN, J. R., ZHANG, Q., KODIRA, C. D., ZHENG, X. H., CHEN, L., SKUPSKI, M., SUBRAMANIAN, G., THOMAS, P. D., ZHANG, J., GABOR MIKLOS, G. L., NELSON, C., BRODER, S., CLARK, A. G., NADEAU, J., MCKUSICK, V. A., ZINDER, N., LEVINE, A. J., ROBERTS, R. J., SIMON, M., SLAYMAN, C., HUNKAPILLER, M., BOLANOS, R., DELCHER, A., DEW, I., FASULO, D., FLANIGAN, M., FLOREA, L., HALPERN, A., HANNENHALLI, S., KRAVITZ, S., LEVY, S., MOBARRY, C., REINERT, K., REMINGTON, K., ABU-THREIDEH, J., BEASLEY, E., BIDDICK, K., BONAZZI, V., BRANDON, R., CARGILL, M., CHANDRAMOULISWARAN, I., CHARLAB, R., CHATURVEDI, K., DENG, Z., DI FRANCESCO, V., DUNN, P., EILBECK, K., EVANGELISTA, C., GABRIELIAN, A. E., GAN, W., GE, W., GONG, F., GU, Z., GUAN, P., HEIMAN, T. J., HIGGINS, M. E., JI, R. R., KE, Z., KETCHUM, K. A., LAI, Z., LEI, Y., LI, Z., LI, J., LIANG, Y., LIN, X., LU, F., MERKULOV, G. V., MILSHINA, N., MOORE, H. M., NAIK, A. K., NARAYAN, V. A., NEELAM, B., NUSSKERN, D., RUSCH, D. B., SALZBERG, S., SHAO, W., SHUE, B., SUN, J., WANG, Z., WANG, A., WANG, X., WANG, J., WEI, M., WIDES, R., XIAO, C., YAN, C., et al. 2001. The sequence of the human genome. *Science,* 291**,** 1304-51.

VINK, R., KRAAIJENHAGEN, R. A., HUTTEN, B. A., VAN DEN BRINK, R. B., DE MOL, B. A., BULLER, H. R. & LEVI, M. 2003. The optimal intensity of vitamin K antagonists in patients with mechanical heart valves: a meta-analysis. *J Am Coll Cardiol,* 42**,** 2042-8.

VOGEL, F. 1959. Moderne Probleme der Humangenetik. *In:* HEILMEYER, L., SCHOEN, R. & DE RUDDER, B. (eds.) *Ergebnisse der Inneren Medizin und Kinderheilkunde.* Springer Berlin Heidelberg.

VOORA, D., KOBOLDT, D., KING, C., LENZINI, P., EBY, C., PORCHE-SORBET, R., DEYCH, E., CRANKSHAW, M., MILLIGAN, P., MCLEOD, H., PATEL, S., CAVALLARI, L., RIDKER, P., GRICE, G., MILLER, R. & GAGE, B. 2010. A Polymorphism in the VKORC1 Regulator Calumenin Predicts Higher Warfarin Dose Requirements in African Americans. *Clinical Pharmacology and Therapeutics,* 87**,** 445-451.

WALKER, E., HERNANDEZ, A. V. & KATTAN, M. W. 2008. Meta-analysis: Its strengths and limitations. *Cleve Clin J Med,* 75**,** 431-9.

WANG, D., CHEN, H., MOMARY, K. M., CAVALLARI, L. H., JOHNSON, J. A. & SADEE, W. 2008. Regulatory polymorphism in vitamin K epoxide reductase complex subunit 1 (VKORC1) affects gene expression and warfarin dose requirement. *Blood,* 112**,** 1013-21.

WANG, H., REN, L., HE, Y., WEI, Y., CHEN, Z., YANG, W., FU, Y., XU, X., FU, W., HU, G. & LOU, W. 2014. Association between cytochrome P450 2C9 gene polymorphisms and colorectal cancer susceptibility: evidence from 16 case-control studies. *Tumour Biol,* 35**,** 4317-22.

WARDROP, D. & KEELING, D. 2008. The story of the discovery of heparin and warfarin. *Br J Haematol,* 141**,** 757-63.

WERNERSSON, R., JUNCKER, A. S. & NIELSEN, H. B. 2007. Probe selection for DNA microarrays using OligoWiz. *Nat Protoc,* 2**,** 2677-91.

WHITLEY, H. P., FERMO, J. D., CHUMNEY, E. C. G. & BRZEZINSKI, W. A. 2007. Effect of patient-specific factors on weekly warfarin dose. *Therapeutics and Clinical Risk Management,* 3**,** 499-504.

WHO. 2015. *The global burden of chronic disease* [Online]. Available: http://www.who.int/nutrition/topics/2_background/en/ 2015].

WILCOX, J. N., SMITH, K. M., SCHWARTZ, S. M. & GORDON, D. 1989. Localization of tissue factor in the normal vessel wall and in the atherosclerotic plaque. *Proceedings of the National Academy of Sciences of the United States of America,* 86**,** 2839-2843.

WU, A. B. 2011. Economics of Pharmacogenomic Testing in Clinical Practice. *In:* WU, A. H. B. & YEO, K.-T. J. (eds.) *Pharmacogenomic Testing in Current Clinical Practice.* Humana Press.

YACH, D., HAWKES, C., GOULD, C. L. & HOFMAN, K. J. 2004. The global burden of chronic diseases: overcoming impediments to prevention and control. *JAMA,* 291**,** 2616-22.

YAN, M., RERKO, R. M., PLATZER, P., DAWSON, D., WILLIS, J., TONG, M., LAWRENCE, E., LUTTERBAUGH, J., LU, S., WILLSON, J. K., LUO, G., HENSOLD, J., TAI, H. H., WILSON, K. & MARKOWITZ, S. D. 2004. 15-Hydroxyprostaglandin dehydrogenase, a COX-2 oncogene antagonist, is a TGF-beta-induced suppressor of human gastrointestinal cancers. *Proc Natl Acad Sci U S A,* 101**,** 17468-73.

YANG, L., GE, W., YU, F. & ZHU, H. 2009. Impact of VKORC1 gene polymorphism on interindividual and interethnic warfarin dosage requirement— A systematic review and meta analysis. *Thrombosis Research,* 125**,** 159-166.

YIN, M. J., YAMAMOTO, Y. & GAYNOR, R. B. 1998. The anti-inflammatory agents aspirin and salicylate inhibit the activity of I(kappa)B kinase-beta. *Nature,* 396**,** 77-80.

YOU, J. H. 2011. Pharmacoeconomic evaluation of warfarin pharmacogenomics. *Expert Opin Pharmacother,* 12**,** 435-41.

YOU, J. H., TSUI, K. K., WONG, R. S. & CHENG, G. 2009. Potential clinical and economic outcomes of CYP2C9 and VKORC1 genotype-guided dosing in patients starting warfarin therapy. *Clin Pharmacol Ther,* 86**,** 540-7.

YOU, J. H., TSUI, K. K., WONG, R. S. & CHENG, G. 2012. Cost-effectiveness of dabigatran versus genotype-guided management of warfarin therapy for stroke prevention in patients with atrial fibrillation. *PLoS One,* 7**,** e39640.

YUAN, H. Y., CHEN, J. J., LEE, M. T., WUNG, J. C., CHEN, Y. F. & CHARNG, M. J. 2005. A novel functional VKORC1 promoter polymorphism is associated with inter-individual and inter-ethnic differences in warfarin sensitivity. *Human Molecular Genetics,* 14**,** 1745-1751.

YUSUF, S., REDDY, S., ÔUNPUU, S. & ANAND, S. 2001. Global Burden of Cardiovascular Diseases: Part I: General Considerations, the Epidemiologic Transition, Risk Factors, and Impact of Urbanization. *Circulation,* 104**,** 2746-2753.

ZANKE, B. W., GREENWOOD, C. M., RANGREJ, J., KUSTRA, R., TENESA, A., FARRINGTON, S. M., PRENDERGAST, J., OLSCHWANG, S., CHIANG, T., CROWDY, E., FERRETTI, V., LAFLAMME, P., SUNDARARAJAN, S., ROUMY, S., OLIVIER, J. F., ROBIDOUX, F., SLADEK, R., MONTPETIT, A., CAMPBELL, P., BEZIEAU, S., O'SHEA, A. M., ZOGOPOULOS, G., COTTERCHIO, M., NEWCOMB, P., MCLAUGHLIN, J., YOUNGHUSBAND, B., GREEN, R., GREEN, J., PORTEOUS, M. E., CAMPBELL, H., BLANCHE, H., SAHBATOU, M., TUBACHER, E., BONAITI-PELLIE, C., BUECHER, B., RIBOLI, E., KURY, S., CHANOCK, S. J., POTTER, J., THOMAS, G., GALLINGER, S., HUDSON, T. J. & DUNLOP, M. G. 2007. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet,* 39**,** 989-94.

ZELL, J. A., ZIOGAS, A., IGNATENKO, N., HONDA, J., QU, N., BOBBS, A. S., NEUHAUSEN, S. L., GERNER, E. W. & ANTON-CULVER, H. 2009. Associations of a polymorphism in the ornithine decarboxylase gene with colorectal cancer survival. *Clin Cancer Res,* 15**,** 6208-16.

ZHENG, G., PATOLSKY, F., CUI, Y., WANG, W. U. & LIEBER, C. M. 2005. Multiplexed electrical detection of cancer markers with nanowire sensor arrays. *Nat Biotech,* 23**,** 1294-1301.

ZHOU, G., MARATHE, G. K., HARTIALA, J., HAZEN, S. L., ALLAYEE, H., TANG, W. H. & MCINTYRE, T. M. 2013. Aspirin hydrolysis in plasma is a variable function of butyrylcholinesterase and platelet-activating factor acetylhydrolase 1b2 (PAFAH1b2). *J Biol Chem,* 288**,** 11940-8.

ZHOU, G., MARATHE, G. K., WILLARD, B. & MCINTYRE, T. M. 2011. Intracellular erythrocyte platelet-activating factor acetylhydrolase I inactivates aspirin in blood. *J Biol Chem,* 286**,** 34820-9.

ZÖLLNER, S. & PRITCHARD, J. K. 2007. Overcoming the Winner's Curse: Estimating Penetrance Parameters from Case-Control Data. *American Journal of Human Genetics,* 80**,** 605-615.