Public Health Genomics: Exploiting SNP-NSAID interactions to prevent colorectal cancer

by

Peter Alexander Demarsh

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the M.Sc. degree in
Epidemiology and Community Medicine

Faculty of Medicine
University of Ottawa

Abstract

Colorectal cancer (CRC) is a common disease with a high mortality rate. Non-steroidal anti-inflammatory drugs (NSAIDs) such as aspirin and ibuprofen have shown considerable promise as preventive agents for CRC. However, due to concerns over the balance between benefits and harms NSAIDs are not recommended as a means of preventing CRC in average risk groups. Investigation into genetic modifiers of NSAIDs' chemopreventive action may help to identify those for whom such drugs are beneficial or ineffective. This thesis explored genetic mediation of the effectiveness of NSAID prophylaxis for CRC. A review of basic CRC biology and methods for investigating interactions, a systematic review of the literature to identify candidate interactions, and a secondary analysis of a GWA case-control study were performed. Candidate SNPs were screened for potential interactions, and several possible interactions were identified. The joint effects were similar for aspirin and ibuprofen, but not acetaminophen, implying true biological effects. Potential interactions were investigated further using a stepwise model building procedure. This resulted in a model containing three SNPs, aspirin and/or Ibuprofen use, their interactions, sex and age. This model was better able to discriminate cases and controls, demonstrated better calibration, and had greater information content (by AIC) than models without the interaction terms. Finally, recommendations for further research were given.

Acknowledgements

I would like to thank:

- CIHR, and the Interdisciplinary Team for Colorectal Cancer for access to the high quality ARCTIC dataset and for my studentship;
- Dr. Julian Little, for being a patient and wise advisor, for always being ready with encouragement or assistance, and for reluctantly shifting from carrot to stick as it became appropriate;
- Dr. Brenda Wilson, for encouragement, guidance, and reminding me that every sentence requires a verb;
- Dr. Pascal Roy of the Université Claude Bernard and Dr. Mia Hashibe of the International Agency for Research on Cancer for being excellent hosts in France, and helping untangle my confusion over the interpretation of interaction in logistic regression;
- Fay Draper, without whom my academic life would have been even more disorderly, for helping me with increasingly obscure red-tape issues;
- my fellow students (especially Roxanne, Badr, Maggie, Jen, Lara, Ryan, the Jessicae, and Michelle) for your continued friendship, and for graduating well before me so it didn't look nearly as impossible;
- Alex Yurkewiech, for being a most excellent parallel reviewer and all-around good chap;
- Steven Hawken, for being an excellent statistician and a pleasure to work with (despite his best efforts to the contrary);
- Diana-Lynn Fox, for being an entertaining roomie, and very gentle (yet persistent) prodder;
- all the other Epi faculty and admin staff for their assistance;
- my parents, friends and family for levity and support;

- and of course, an extra-special thank you to Dr. Lauren Reid for being my partner-in-all-things-adventurous, a frequent library date, and playing the tricky role of simultaneously being my compassionate confidante and an irresistible motivational force. My love always.

Contents

List of Tables

List of Figures

"Nor can it be supposed that the diversity of chemical structure and process stops at the boundary of the species, and that within that boundary, which has no real finality, rigid uniformity reigns. Such a conception is at variance with any evolutionary conception of the nature and origin of species. The existence of chemical individuality follows of necessity from that of chemical specificity, but we should expect the differences between individuals to be still more subtle and difficult of detection. Indications of their existence are seen, even in man, in the various tints of skin, hair, and eyes, and in the quantitative differences in those portions of the end-products of metabolism which are endogenous and are not affected by diet, such as recent researches have revealed in increasing numbers. Even those idiosyncrasies with regard to drugs and articles of food which are summed up in the proverbial saying that what is one man's meat is another man's poison presumably have a chemical basis."

-Archibald Garrod 'Inborn Errors of Metabolism'. The Croonian Lectures delivered before the Royal College of Physicians of London, in June, 1908 (1909), 2-3.

"We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest."

-J.D. Watson, and F.H.C. Crick In 'A Structure for Deoxyribose Nucleic Acid', Letter in Nature (25 Apr 1953), 171, 738.

Chapter 1

Introduction

Almost precisely fifty years after Watson and Crick's understated 1953 letter to Nature first appeared, two independent collaborations announced that they had successfully sequenced the bulk of the human genome [85] [52]. Now, with the source code of the species at our disposal, describing all interpersonal variation seemed within reach.

Unfortunately for the advancement of medicine (though perhaps fortunately for career prospects in Epidemiology), biology remains elusively complex. Garrod's "errors of metabolism" depend not merely on the genetic code, but on a complex web of interconnected, interacting genetic and environmental factors. Whether this sticky network can be untangled and the component causes teased out may determine whether genomics can be used to develop better, safer interventions, and ultimately improve human health. This thesis attempts to apply detection of biological interaction to the improvement of public health through targeted intervention.

Colorectal cancer (CRC) is a common disease, with a lifetime risk of approximately 5 to 6%. In Canada, 22,500 new cases and 9,100 deaths were projected for 2010 [14]. Only lung cancer, for which tobacco smoking is the major cause [21], is responsible for more cancer deaths. Tobacco control measures have proven challenging in some groups (such as teenage girls [21]), but largely these efforts have been successful in reducing lung cancer incidence [23]. In contrast, CRC is multifactorial, which makes primary prevention difficult. The risk of developing CRC is influenced by several environmental factors, genetics, and complex interactions between the two. Further, the behaviors which are thought to reduce risk (such as increasing physical activity and fiber intake, and decreasing obesity [55]) have proven difficult to change at the population level [45]. CRC control is thus heavily based on secondary prevention through organized population

and/or opportunistic screening; faecal occult blood testing (FOBT), sigmoidoscopy, and colonoscopy. Although screening is an efficacious means of detecting cancers and reducing CRC mortality when performed regularly on large numbers of at-risk individuals [34], uptake levels have not reached the desired levels [66]. Primary prevention of CRC remains a worthwhile goal.

Non-steroidal anti-inflammatory drugs (NSAIDs) such as Aspirin and Ibuprofen have shown considerable promise as preventive agents for CRC [68]. However, due to concerns over the balance between benefits and harms NSAIDs are not recommended as a means of preventing CRC in average risk groups [65]. This balance shifts when considering individuals at an increased risk of CRC due to familial adenomatous polyposis (FAP) [75]. There may be other genetically defined subgroups for whom NSAID might be similarly favourable. Investigation into genetic modifiers of NSAIDs' chemopreventive action may help to define such groups, identify those for whom such drugs are beneficial or ineffective, and make targeted interventions possible.

This thesis aims to explore the role of genetic variation in mediating the effectiveness of NSAID prophylaxis for CRC. A systematic review and meta-analysis were performed with the intent of identifying gene variants that altered the chemopreventive effect of NSAIDs. Next, exploratory analysis of a large case-control dataset was performed to detect potentially exploitable SNP-NSAID interactions. Finally, conclusions and recommendations for future research are given.

1.1 Why colorectal cancer?

With the publication of the draft sequences of the human genome [85] [52] and the subsequent and continuing refinement of molecular genetic tools, an analytic means of inquiry into the hereditary basis of many common diseases has become theoretically plausible. In the opinion of one researcher "The sequencing of the human genome offers the greatest opportunity for epidemiology since John Snow discovered the Broad Street Pump" [72], but there are several hurdles.

A significant effort is being directed to trawling these data for associations with human diseases. As of yet the major successes have been instances of rare genetic variants that independently confer a greatly increased risk of disease. The examples of the BRCA1/2 gene mutations and breast cancer are probably the best known. Such single gene, high risk variants have been identified mostly due to their high penetrance, meaning that there is a high likelihood that individuals with the relevant genotype will develop the

disorder. They are generally quite rare, which limits their usefulness as a means of targeting interventions at the population level.

There is hope [46] that genetic information could have broader public health applications, and this potential is a major part of the motivation for this project. By incorporating information from sets of genes, the products of which are involved in disease-causing pathways, it may be possible to predict disease or the response to interventions, provided targeted treatments for individuals, and improve health on a population scale. However, the relative risk (RR) of disease associated with most common alleles identified thus far is small (generally 1.2-1.5) [41]. If such information is to be useful, new approaches must be used which require an appropriate disease on which to be tested. CRC provides a test case for several reasons.

In addition to being common with relatively high mortality, the cancer's biology makes CRC an appropriate test case for genetic epidemiology's foray into common diseases. The major opportunity for improving health in this area is through elucidation of the composite factors which bring about multifactorial disease. CRC shows such a complex etiology, with large fraction of incidence thought to be driven by environmental exposures mediated by genetic vulnerability [27]. It is the opinion of some researchers [45] that the public health significance of genomic research on common complex diseases will be through better understanding of the interaction between exposures and genetic susceptibility, rather than in determining sole genetic causes. By this route risk behaviors could be modified or preventive measures targeted to those most likely to benefit.

Further, CRC allows multiple opportunities for effective intervention, as it is thought that there are approximately 10 years [86] between the development of a precancerous lesion (adenoma) and the advance to invasive neoplasm (carcinoma). During this period adenomas and early carcinomas can be detected and excised due to their physically accessible location. This makes the ability to identify individuals at an increased risk highly appealing, as there is the opportunity for successful intervention. As a corollary, tissue samples are readily available for a variety of tumor types and stages. This relative wealth of biological information has allowed significant exploration of the molecular and genetic events underlying CRCs. These events are thought to be fairly well understood, although there are some indications of unexpected heterogeneity in tumor mutations [76] [25]. Exploration of the potential for individualized prevention strategies, one of the primary motivations for this thesis, depends on thorough knowledge of the roots of the condition of interest.

The complex phenotype, opportunity for intervention, and abundance of biological

data make CRC an appropriate topic for this thesis. Epidemiological methods used to elucidate the role of genetics in common disease can be developed or assessed in CRC, and adapted for other conditions.

1.2 Why NSAIDs?

NSAIDs primarily inhibit the activity of the cyclooxygenase (COX) enzymes, which are involved in prostaglandin synthesis (for a full review, see [83]). Prostaglandins are multipotent signalling molecules involved in a wide range of physiological processes, including inflammation. Inflammation can be caused by infectious and non-infectious processes. Non-infectious inflammation may be caused by chronic tissue irritation or damage. Inflammation produces reactive oxygen species, which can damage DNA. Prostaglandins also elicits cellular proliferation and the growth of blood vessels, which is a key component of carcinogenesis. Chronic inflammation may be an independent risk factor for cancer.

In 1988, a large case-control study of colorectal cancer numerous identified potential associations with several medications [50]. An inverse association with Aspirin use and risk of colorectal cancer was identified, and persisted after controlling for known potential confounding factors. This initial human evidence of a potential role of Aspirin in prevention arose independent of the extensive previous research on animal models, which had also indicated a potential preventive application.

Many subsequent studies have confirmed this association [39]. A large body of evidence has indicated that Aspirin, Ibuprofen, and other non-steroidal anti-inflammatory drugs (NSAIDs) inhibit colorectal carcinogenesis, colorectal cancer (CRC) and colorectal adenoma (CRA). The evidence is diverse, coming from first from animal models (e.g. [49] [64] [62], and numerous others), subsequently from observational studies in humans [26] [30] [78] [80], before more recently from intervention trials of NSAIDs in patients with FAP [77] [35] [51] [28] and randomized controlled trials of Aspirin and selective COX-2 inhibitors in humans without FAP [9], [4] [6]. There is now extensive observational and experimental evidence that both colorectal adenomas (CRAs) [8] [5] and CRC [24] can be prevented through prolonged use of acetyl salicylic acid (ASA; most commonly Aspirin), Ibuprofen, and several other forms of non-steroidal anti-inflammatory drugs (NSAIDs). A recent systematic review [68] found that regular use of Aspirin reduced the incidence of CRAs in RCTs (RR 0.82, 95% CI 0.7 - 0.95), case-control studies (RR 0.87, 0.77 - 0.98), and cohort studies (RR 0.72, 0.61 - 0.85). The inverse relationship was observed

to become stronger with greater duration of NSAID use in both case-control and cohort studies, which is a strong indicator of causality.

Despite the likelihood of a protective effect, in 2007 the US Preventive Service Task Force advised against the adoption of NSAIDs as CRC preventive agents for asymptomatic individuals, citing "good evidence of at least moderate harms associated with Aspirin and NSAIDs" for the dose and duration of use associated with the putative protective effect [65]. Potential harms include serious gastrointestinal bleeding, as well as cardiac and liver toxicity. It was, however, explicitly noted that further efforts to quantify the net benefits and risks should be undertaken.

It is possible that genetically defined subgroups could derive additional benefit from NSAID prophylaxis. Inherited genetic factors may explain some inter-individual differences in NSAID metabolism and prostaglandin synthesis [83]. As an example, there is evidence that the preventive effect of NSAIDs on CRC is limited to those expressing the fast-metabolizing form of the drug-metabolizing enzyme CYP2C9 [10]. If genetic information could be used to shift NSAID use towards those likely to respond, it might be possible to provided targeted chemoprevention in such a way that provided maximal benefit and limited the potential for harms. Examining inherited differences in genes encoding enzymes involved in processing NSAIDs, a highly probable protective agent, seem to be a logical starting point for such inquiry. Thus, this project aims to assess the extent to which germline (inherited) genetic polymorphisms are associated with CRC risk in the presence of NSAID use, with the ultimate goal of determining appropriate targets for chemoprevention by genotype analysis.

Chapter 2

Background and Theory

2.1 Anatomy

The human colon is is a tube-shaped structure which extracts water and electrolytes from feces prior to excretion. It consists of the ascending colon, transverse colon, the descending colon, and the sigmoid colon. From the cecum to the splenic flexure (the junction between the transverse and descending colon) is also known as the right colon. The remainder is known as the left colon. The inner cavity of the colon (as with any tubular organ) is called the lumen. The colon is composed of four distinct layers: the mucosa epithelium, the submucosa, the muscalaris, and the adventitia (or serosa). Extending through these layers are simple tubular glands called crypts. The terminal segment in which feces accumulate is called the rectum. It is continuous with the sigmoid colon and extends 13 to 15 cm to the anus. A muscular sheet called the pelvic diaphragm runs perpendicular to the juncture of the rectum and anal canal and maintains a constriction between these two segments of the large intestine. A full description of colorectal anatomy can be found in any clinical anatomy text (e.g. [61]).

Stem cells at the base of the crypts renew the epithelium, which is composed of several different cell types (mostly mucus-secreting goblet cells, a type of glandular simple columnar epithelial cell). Differentiation occurs as cells move up the crypt towards the mucosa. Ultimately, neoplasia is believed to result from the accumulation of genetic mutations. Genes which encode growth-limiting proteins (tumor-suppressors) and genes giving rise to growth-promoting proteins (proto-oncogenes) maintain a dynamic equilibrium in normal cells of the colonic epithelium (colonocytes). Loss of a suppressor gene or the conversion of a proto-oncogene into an oncogene perturbs this balance, leading to the

loss of normal cellular behavior. Such genetic abnormalities can be inherited, acquired, or both, and result in markedly different behavior of cancer cells compared with normal cells; loss of senescence, loss of contact inhibition, autocrine growth, and independent growth.

2.2 Adenoma-Carcinoma Sequence

Most colorectal carcinomas are believed to develop from initially benign colorectal adenomas (CRAs), in a multistage process referred to as the adenoma-carcinoma sequence. In this model, somatic genetic alterations accumulate in colonic epithelial cells over time. Affected cells acquire a progressively perturbed biology. If these mutations lead to the silencing of a tumor suppressor gene or the activation of an oncogene, clones of the affected cell will have a relative growth advantage versus normal colonocytes. This allows for a process of clonal expansion, whereby clones of the mutant cell become locally dominant, forming an adenoma. Eventually, a cell in a small fraction of adenomas may undergo further mutations to gain invasive potential, and through clonal selection become a carcinoma. [19]

The molecular events underlying the adenoma-carcinoma model were first described in a seminal paper entitled "Genetic alterations during colorectal-tumor development" by Fearon and Vogelstein [22]. Presently, through investigations using tissue from excised human tumors and animal models, the genetic events common to most tumors are thought to have been identified [48], although there recently there has been indication that there may be additional heterogeneity [42].

The adenoma carcinoma sequence is described in depth in [19]. Although carcinogenesis is believed to be driven by the accumulation of mutations, rather than their order, typical tumors tend to follow a known progression; initiation of epithelial dysplasia, hyperproliferation of cells to form an adenoma, progression to severe dysplasia, and the acquisition of invasive potential as a malignant carcinoma. For the above progression to occur four critical cellular processes must be compromised; regulation of cell division by mitotic checkpoints, apoptosis (programmed cell death), intracellular and intercellular signaling, and anchorage to the basement membrane [15]. A neoplastic colonic epithelial cell is characterized by various aberrations in these processes due to direct changes in the nucleotide sequence (genetic), or changes in non-nucleotide factors controlling gene expression (epigenetic). The specific genetic changes in most CRCs are presented in Figure 2.1.

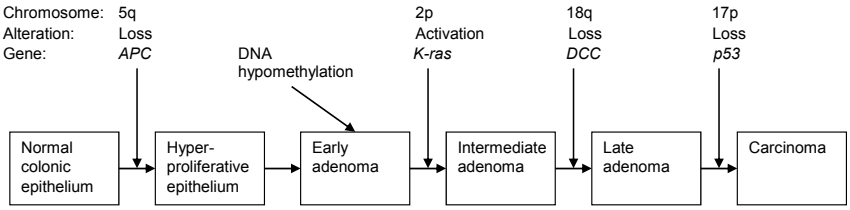

Figure 2.1: Typical genetic alterations in the adenoma-carcinoma sequence. Adapted from [22].

The molecular genetics of colorectal cancer are well understood (reviewed in [36]). Genomic instability enables the acquisition of these mutations, and progressively more disruption of cellular controls. In cells that eventually become cancerous, this instability affects genes involved in the maintenance of the non-neoplastic phenotype, such as those described above. Typically this instability takes one of two forms: chromosomal instability (CIS), or microsatellite instability (MSI). CIS is characterized by alterations in chromosome number (aneuploidy) and the rapid gain or loss of chromosomal regions. Up to 85% of CRCs show evidence of CIS. MSI involves instability at the level of the nucleotide bases, resulting from several types of mutation. In this tumorigenic pathway, chromosomes remain diploid. The inactivation of genes is caused by short repeating neucleotide sequences called microsatellites. In MSI tumors, microsatellites in the coding regions of several genes are poorly controlled (i.e. increase in length during DNA replication) and disrupt gene products, most apparently in mismatch repair genes.

2.3 Epidemiology of Colorectal cancer

The epidemiology of CRC has been recently been reviewed [55]. Age shows a strong, positive association with CRC. Sporadic CRC is extremely rare before 40 years of age, with 90 percent of cases occurring after age 50. Incidence rates increase rapidly between the ages of 40 and 50, and continue to do so with each successive decade. The disease tends to be more common in males than in females, a trend that is observed internationally for all age categories. A moderate increase in incidence has been observed over time in many populations, a trend that appears to be greater in or limited to males.

Worldwide more than a million new cases of cancer of the large bowel are diagnosed each year. It is a common cancer in both developed and developing countries, though the former accounts for approximately two thirds of the incidence. After controlling for age, a 25 fold difference in incidence rates between the highest and lowest incidence regions persists. The highest incidence rates, over 40 per 100 000 men, are seen in North America, Australia, and northern and western Europe. The lowest are seen in less developed countries, particularly in Africa and Asia, where the incidence rates in some regions are less than 10 per 100 000 men [63]. The distribution of incidence is approximately the same for women, though substantially lower (60-80 percent of the age-standardized rates, [55]).

The most important environmental risk factor appears to be diet. High vegetable consumption has the most supporting evidence of an association, while red meat consumption is the most supported risk factor. It is expected that incidence rates will rise in developing countries as the age of populations increase, and with the proliferation of the "western" diet high in red meat and animal fats and low in produce. Other risk factors include obesity and overweight, smoking, inflammatory bowel disease, and a previous history of adenomas. NSAIDs are a well-supported protective factor (reviewed in Chapter 1). Mortality also varies widely geographically, but taken cumulatively the mortality rate is approximately half of the incidence rate. Though some of these geographic differences appear to be attributable to differences in dietary and environmental exposures, they are imposed upon a background of genetically-influenced susceptibility.

2.4 What happens to NSAIDs in the body?

The metabolism of ASA (reviewed in [29]) is thought to be well understood. ASA is actually a prodrug (i.e. an inactive precursor, converted to an active form in vivo),

and is hydrolyzed to the pharmacologically active salicylic acid (SA) by esterases in the gastrointestinal tract and liver. The major pathways for the metabolism of salicylic acid are conjugation with glycine to form salicyluric acid and conjugation with glucuronic acid to form salicyl phenolic glucuronide.

Although there are some differences in metabolism between the drug types, the same major pathways are involved in the metabolism of most NSAIDs. The main metabolic pathways leading to the inactivation and elimination of NSAIDs are oxidation by the cytochrome P450 enzymes (P450), glucuronide conjugation by the uridine-5'-diphosphate (UDP)-glucuronosyltransferases (UGTs) and, to a lesser extent, sulphate conjugation (sulphotransferases).

The primary mode of NSAID action as an analgesic and as a chemopreventive agent may be the inhibition of prostaglandin synthesis and COX enzymes [17]. ASA, Ibuprofen and most other NSAIDs are non-selective inhibitors of cyclooxygenase 1 and 2 (COX1/2) enzymes, binding to the active site of COX to prevent the combination of the enzyme with arachidonic acid. ASA is unique among NSAIDs in that it permanently renders the enzymes non-functional. ASA covalently modifies the COX proteins by the donation of an acetyl group. All other NSAIDs bind reversibly to the COX proteins. COX enzymes catalyze the synthesis of thromboxanes and prostaglandins. COX has two enzymatic activities, a cyclooxygenase activity that oxygenates arachidonic acid to a hydroperoxy endoperoside (prostaglandin G2), and a peroxidase that reduces prostaglandin G2 to prostaglandin H2. Increased levels of prostaglandins, particularly prostaglandin E2, have been detected in many CRCs. Prostaglandins influence angiogenesis, apoptosis, cell proliferation and migration [83], each of which are involved in tumorigenesis and/or metastasis. Acetaminophen is notable in that its structure is only weakly acidic; it is thus expected to have weak anti-inflammatory properties.

2.4.1 Application of Public Health Genomics

A polymorphism in the *ITG>3* gene encoding the glycoprotein IIIa subunit of the glycoprotein IIa/IIIa drug receptor may alter the anti-platelet activity of Aspirin [59]. If the finding is validated through replication, it could have clinical implications for the use of NSAIDs in the prevention of cardiovascular disease. However, the situation will likely be more convoluted for most attempts to use genetic information to inform prevention activities for other common diseases. Drug response is typically a multigenic trait [81], which may make it difficult to reach concrete conclusions about risk. Further, the field of

genetic epidemiology is plagued by non-reproduction of initially exciting results, perhaps the result of a combination of limited statistical power, selective reporting and publication biases [40]. The traditional approaches of epidemiology may be useful in distilling, assembling and translating genetic information into valid recommendations with public health relevance [54].

2.5 Statistical theory

2.5.1 Interaction

Interaction is a term that seems to lend itself to confused interpretation. The term means that the association between two variables is determined by the level of one or more other variables, but this can refer to a biological effect, or a statistical effect. A biological interaction exists in nature, while a statistical interaction may be dependent on the choice of model. A very clear explanation of the confusion is given in [1].

Interaction was the focus of a recent series of articles in the journal Epidemiology, and two of the articles were particularly relevant to this thesis. First, [44] gives a summary of this recent discussion around the evolving use of the term "interaction", and its implications for the practice of epidemiology. The consensus seems to be that while multiplicative interaction is more straightforward computationally, it is of dubious usefulness for applied public health action. It is recommended that effects of either exposure of interest and the joint effect of both versus those without either exposure be reported. The second [47] discussed the reporting of studies attempting to detect GxE interactions. It was found that interactions were usually reported as the p-value resulting from a statistical test of departure from a multiplicative model, while the reporting of strata-specific ORs would have greater potential for public-health application.

Hunter ([38]) reviewed the issue of detecting GxE interactions in human diseases in 2005. It was noted that a variety of models can describe such interactions, taking into account the variety of ways in which genetic effects can be altered by environmental exposures, the number of levels chosen for these exposures, and the genetic model chosen. Hunter notes that interactions are likely only to be accepted once replicated in at least two independent studies, which requires a high degree of coordination between studies. The issue of sample size is also explored. Hunter explores the rule of thumb that the sample size necessary to detect joint effects is four times the size required for each of the considered main effects. Of particular interest are methods of pooling information

across studies in order to increase power and prevent publication biases. As an additional means of addressing the issue of publication bias, and the possibility of chance associations arising from the testing of multiple hypotheses, Hunter recommends replicating results in independent samples. Additionally, obtaining high quality information on environmental exposures may be of critical importance for the assessment of GxE interactions. Particulary, while the use of departure from a multiplicative model is the most commonly used test of interactions, Hunter notes that the additive model is often said to be more relevant to assessing the public-health impact of an intervention. This is the approach followed in this thesis.

A second article by Thomas ([79]) investigated methods for investigating GxE interactions in GWAs. The author states that despite considerable enthusiasm new generation GWAs, there remains a great deal of "missing heritability", and that a portion of this is likely due to GxE interactions, or more complex pathways involving multiple genes and exposures. Hunter reviews the history of debate in the epidemiological and statistical literature over the use of the term "interaction", and makes an important distinction. In statistical modelling, interaction indicates that the joint effect of multiple factors be described in terms of a statistical model involving only main effects on some scale or requiring additional interaction terms. In contrast, in epidemiology synergy describes a joint effect that is greater than the sum of the component factors alone, while biological interaction describes an effect of one factor at the cellular or molecular level that depends on the presence or absence of the other.

Chapter 3

A systematic review of SNP-NSAID interactions and CRC

3.1 Introduction

Inflammation is an independent risk factor for CRC, as demonstrated by the increased risk for CRC of those with several inflammatory conditions, such as ulcerative colitis and Crohn's disease. Nonsteroidal anti-inflammatory drugs (NSAIDs) could decrease inflammation in the colonic epithelium, thereby decreasing the risk of neoplasia. NSAIDs have been shown to be effective chemopreventive agents for colorectal neoplasia. Polymorphisms in NSAID-metabolizing enzymes or other targets may alter the effectiveness of NSAIDs. However, NSAIDs are not recommended as chemopreventive agents due to the potential for harms. This has stimulated research attempting to identify genetic variations which might tip the balance of risks and benefits by tailoring chemoprevention recommendations to individuals. A systematic review of the literature was conducted in order to identify SNP-NSAID interactions for further analysis.

3.2 Methods

This section gives an overview of the methodology used to conduct the systematic review. First, methods of selecting studies for inclusion in the review are given, subdivided into various selection criteria. These include characteristics of the disease, participants in the study, the design of the study, and the means of measuring exposure. Next, the search methodology used to identify studies is given in a form that is reproducible and

transparent. Next, methods of selecting appropriate studies from the identified articles and methods of extracting data from these studies are given.

3.2.1 Selection Criteria

Disease

Cancers, adenomas, and polyps of the colon and rectum were considered in this review. Outcomes for inclusion were the incidence of CRC/CRA, or CRA recurrence.

Participants

This review targeted adults ($\geq$18 years of age) without a previous history of CRC. Secondary prevention studies of patients with previous CRC were excluded. No further restrictions were placed on participant age, sex, or ethnic background. Studies of participants Familial Adenomatous Polyposis (FAP) or Hereditary Non-Polyposis Colorectal Cancer (HNPCC) were not included, as individuals with these genetic conditions already constitute a high-risk group. Further profiling related to NSAID chemoprevention would not be useful. Secondary prevention studies of patients with a personal history of CRC were excluded for the same reason.

Study design

The review considered cohort, case-control and randomized trials for inclusion. Animal studies, cell line studies, review articles, abstracts and meeting proceedings were excluded. Genome-wide association (GWA) studies were excluded if only the exploratory scan is reported.

Exposure

Typically, HuGE reviews focus on one or several genes and a collection of variants. For the purposes of this review, studies were sought which had the potential to identify an interaction between or effect modification of NSAID prophylaxis and genotype. Studies must have included a statistical test for interaction or the reporting of stratum-specific estimates of the OR or RR to be included in this review. Any measure of NSAID use was accepted.

3.2.2 Identifying studies

Medline, EMBASE, BIOSIS, and the ISI Science Citations Index were searched for journal articles that investigated the genetic variation and the risk of CRC/CRA occurrence or recurrence, and included a measure of NSAID use. The reference lists of all included articles were screened to identify further potentially relevant articles.

The search strategy described here was used for the OVID Medline search, and adapted for other databases:

1. exp Colorectal Neoplasms
2. exp Colorectal Neoplasms
3. exp Adenomatous Polyps
4. (colorectal OR colon OR colonic OR rectal OR rectum OR rectosigmoid OR adenomat$) AND (cancer$ OR carcinoma$ OR adenocarcinoma$ OR malignan$ OR tumor$ OR tumour$ OR neoplasm$ OR polyp$)
5. OR 1-4
6. Exp Anti-Inflammatory Agents, Non-Steroidal
7. Chemoprevention OR (prevention or prevent or chemoprevent$ or chemoprophyl$)
8. 6 OR 7
9. exp Polymorphism, Genetic OR exp Polymorphism, Single Nucleotide
10. 9 AND 10, Limit to Human

The command *exp* means explode, and returns results for both the search term and the family of related concepts given by the Medical Subject Heading (MeSH) index. The operators *AND* and *OR* return items that are members of both or either set to which they refer, respectively. The symbol *$* is a wildcard. Final search was performed April 26, 2008, and was current to that date.

3.2.3 Data collection and analysis

Study selection

Articles from electronic databases were uploaded to RefWorks reference management software, and any duplicates were removed. Titles and abstracts were screened for relevance by two independent reviewers (the author, AD; and Alex Yurkewiech, AY), and any obviously irrelevant studies were excluded. Full text copies of all remaining reports were assessed independently by the two reviewers to determine the final set of included studies. Any disagreements which could not be resolved were to be taken to a third reviewer, but no such disputes arose.

Data collection

Data collection forms were developed based on an initial scan of the literature (see Appendix 3). These were used to extract data from individual studies. Two reviewers (AD and AY) extracted data from two randomly selected articles as a quality control check. There were no disagreements in this data collections. The remaining data were compiled in several Excel spreadsheets for review.

3.3 Results

231 articles were returned, of which 51 passed broadscreening and were retrieved for full-text review. 15 reported NSAID-SNP interactions, with colorectal neoplasia (cancer or adenoma) as the clinical outcome and passed other screening criteria, and were thus included in this review.

Characteristics of included articles are given in Table 3.1, below.

Table 3.1: Characteristics of included studies

Study	Genes	Study design	No. cases/controls	NSAID measure	Region	End point	Population	Covariates
Bigler 2001 [11]	CYP2C9 UGT1A6	Case-control	441/488	ASA/NSAID users:$\geq$ 1/day for $\geq$ 1 year	Minnesota	Ad. polyp	30-74 years old	Adjusted: age, sex, smoking, HRT
Lin 2002 [53]	COX2	Case-control	161/219	Not defined	California/ North Carolina	Ad. polyp; Cancer	African American; 30-74 years old; spoke English	Matched: age, sex, sigmoidoscopy date, center
Martinez 2003 [58]	ODC	Case-control	341/347	ASA use collected, not defined	Arizona	Ad. polyp	40-80 years old	Adjusted: age, sex, number of colonoscopies after baseline
Slattery, 2004 [74]	VDRIGF1 IGFBP3 IRS1 IRS2	Case-control	1346/1544	Not defined	USA	Cancer	White; 55-74 years old	Adjusted: adjusted: age; sex; physical activity; BMI; energy intake; dietary fiber; calcium; cigarettes
Macarthur 2004 [57]	IL1 IL10 TNFα TGFβ	Case-control	264/408	Regular ASA/ NSAID users: $\geq$ 1/day for $\geq$ 1 month	Scotland	Cancer	White; 50-81	Adjusted: age and sex.
Cox 2004 [20]	COX2	Case-control	292/274	Regular use for $\geq$ 6 consecutive months	Spain	Cancer	24-92 years old	Matched: age, sex
Ulrich 2004 [82]	COX1	Case-control	715/621	Regular ASA/NSAID use: $\geq$ 1/day	Minnesota	Ad., hp. polyp	White; English speaking; 30-74 years old	Adjusted: age, sex
Ali 2005 [3]	COX2	Case-control	726/729	Use of ASA, Ibuprofen, none or both (not clearly defined)	USA	Advanced Ad. polyp	White; 55-74 years old	Matched: age, sex. Adjusted: age, sex, smoking, NSAID
Chan 2005 [16]	UGT1A6	Case-control	530/532	ASA users: $\geq$ two/week	USA	Ad. polyp	Not reported	Adjusted: age, smoking hx, BMI, physical activity, family history of CRC, meat intake, alcohol intake, multivitamin use, folate intake, calcium intake
Ulrich 2005 [84]	COX2	Case-control	690/584	Regular ASA/ NSAID users: $\geq$ 1/day	Minnesota	Ad., hp. polyp	White; English—-speaking; 30-70 years old.	Adjusted: age, sex, BMI, calories, alcohol, fiber, hormone use, and smoking

Table 3.1 – Continued

Samowitz 2006 [70]	CYP2C9 UGT1A6 CYP2C9 UGT1A6	Case-control	2295/2903	ASA, Ibuprofen use collected, not defined	California, Minnesota, Utah	Cancer	predominantly White, 30-79 years old	Adjusted: age, sex, BMI, physical activity, energy intake, cigarette smoking, dietary calcium intake, dietary fiber intake
Sansbury 2006 [71]	COX2	Case-control	240/326	NSAID users: at least 3 /week for 3 months	North Carolina	Cancer	African American; 40-80 years old	Matched: age, sex, ethnicity. Adjusted: age, sex. Multivariate adjustment: smoking, activity, diet, supplement use, BMI, CRC
Siezen 2006 [73]	COX1 COX2	Case-control	724/682	NSAID users: more than 12 times/year	Netherlands	Ad. polyp	White; 18-75 years old	Adjusted: age, sex, smoking duration, endoscopy indication
Barry 2006 [7]	ODC	RCT	372 placebo, 377 ASA 81mg, 372 ASA 325mg	Randomized to ASA or Placebo	North America	Ad. polyp	Multi-ethnic study population; 21-80 years old	Adjusted: age, sex, race, ASA intervention, follow-up time.
Hubner 2006 [37]	CYP2C9	RCT	280placebo, 266 ASA	ASA: 300mg daily	UK	Ad. polyp recurrence	Mean age=57.4	Adjusted: age, sex, interval between entry and follow-up colonoscopy

Ten of the fifteen studies reported were of colorectal adenomatous polyps, while the remaining five examined colorectal cancer. All studies included age and sex as covariates, or where matched by these variables. Other potential confounders that were included in some studies were smoking, colonoscopy history, family history of CRC, diet (fiber, alcohol intake), and body mass index. Some studies, but not all, reported results of departure from Hardy-Weinburg equilibrium (HWE). Of those that did, no major departure from HWE were noted. The majority of studies consisted primarily of White participants, although 2 studies [71] [53] involved African-American populations. All but 3 were performed in the USA, with the exceptions all in Western Europe (Spain, the Netherlands, and UK). The studies were also of highly variable size. The definition of NSAID use was highly heterogenous, varying in dose, duration, and/or frequency. The definition of NSAID use was inconsistent. This has the potential to lead to differential exposure misclassification across studies, and makes comparability of results difficult. For this reason, and due to the low level of replication across studies, no quantitative synthesis of results was attempted.

3.3.1 *COX1*

Four *COX1* SNPs were investigated for interactions with NSAID use. These were; Arg8Trp (rs1236913), Leu15-Leu16del (rs3842787), Pro17Leu, and Leu237Met (rs5789). In one study [82], Pro17Leu (a single nucleotide polymorphism that results in an amino acid change in exon 2 of the *COX1* gene was associated with adenoma risk reduction only among wildtype NSAID users, while those with the variant allele did not show any benefit. No statistically significant interactions were reported for the rs1236913, rs3842787, or rs5789 polymorphisms.

3.3.2 *COX2*

Seventeen *COX2* SNPs were evaluated for interaction with NSAIDs. The most commonly evaluated polymorphisms were those occurring at G765C (rs20417) and Val511Ala (rs5273). The latter polymorphism is exclusively found in African-American populations. One study of 494 cases and 584 controls [53] reported large risk reductions in colorectal adenoma for rs20417 homozygous variant nonusers compared to wild-type nonusers (OR: 0.26, 95%CI: 0.07-0.89), whereas there was no decrease in risk among homozygous variant NSAID users (OR=0.82, 95% CI, 0.25-2.73). However, a smaller study of 337 adenoma

cases and 368 controls [73] found no evidence of interaction between this polymorphism and NSAID use.

The rs20417 polymorphism is relatively common (MAF=0.17). Two studies [71] [53] have reported on the *COX2* rs5273 polymorphism and NSAID. One examined this interaction in regards to colorectal cancer (240 cases and 326 controls) [71], while the other studied the interaction for distal adenoma [53]. The first [71] found decreased risk of cancer among wild-type NSAID users (OR=0.66, 95% CI 0.45-0.95) was even greater among those carrying at least one variant allele (OR=0.29, 95% CI, 0.08-1.06). The latter study [53] (161 cases, 219 controls) found significant reductions in risk among those who were either NSAID users or carried the variant allele (or both) compared to those with neither exposure, but did not formally evaluate interaction.

In a hospital-based case-control study conducted in Spain, subjects carrying at least one variant allele of the *COX2* gene rs689469 SNP showed an increased risk of colorectal cancer (OR=2.49, 95% CI, 1.17-5.32). [20] This small study (n=292/274 controls/cases) was the one identified in the literature that reported on this polymorphism with respect to colorectal neoplasia. However, statistical power with respect to interactions was limited, and the hospital-based nature of the study may introduce bias.

3.3.3 *ODC*

Two studies have examined the *ODC* 315G>A (rs2302615) polymorphism. One study [58] reported that the homozygous variant genotype was associated with a reduced risk for adenoma (OR=0.48, 95% CI, 0.24-0.99). The risk among homozygous variant NSAID users was greatly reduced compared to wild-type nonusers (OR=0.10, 95% CI, 0.02-0.66), whereas a risk reduction was not observed among homozygous variant nonusers vs wild-type nonusers (OR=0.68, 95% CI, 0.30-1.51). The second study [7] was a secondary analysis of data from a previous RCT, and did not detect an association between the *ODC* rs2302615 polymorphism and adenoma. However, they reported a statistically significant interaction of genotype and Aspirin use on adenoma risk. Aspirin users with one or more variant allele showed a reduction in risk for adenoma (RR=0.77, 95%CI, 0.63-0.95), as well as risk of advanced adenoma (RR=0.51, 95%CI 0.29-0.90) compared to members of the group receiving placebo who had at least one variant allele. There was no risk reduction associated with Aspirin use observed among those with two copies of the common allele. This seems to indicate that there is a cumulative protective effect for NSAID users who also have at least one variant *ODC* rs2302615 allele.

3.3.4 *UGT1A6*

Four studies examined the polymorphisms in *UGT1A6*, Thr181Ala+Arg184Ser (rs2070959) or Arg184Ser (rs1105879). In a study of 441 adenoma cases and 451 controls, the risk reduction for regular Aspirin users was seen only among those with at least one variant allele (OR=0.53, 95% CI 0.33-0.86). [11] Similarly, in a case-control study of 313 women with adenoma and 303 control women, the risk reduction associated with regular NSAID use was stronger among women with any variant UGT1A6 genotype compared to those with the wild-type alleles. [16] ORs were not reported, but did report the results of a test of departure from additivity (p-interaction=0.02). Two other studies [37][70] reported no interaction between *UGT1A6* genotype and NSAID use.

3.3.5 *CYP2C9*

Three studies [11][70] [37]investigated interactions between the "*2" and "*3" polymorphisms in *CYP2C9* and NSAID use related to colorectal neoplasia. The first [11] examined CRA risk reduction associated with Aspirin use, and found Aspirin to only be effective in those with the wild-type *CYP2C9* genotype (OR=0.50, 95% CI 0.32-0.78). There was no decrease in OR seen among non-Aspirin users. A second study [70] detected an interaction between the "*2" and "*3" genotypes and Ibuprofen use. Those with a variant allele showed a greater decrease in OR with regular Ibuprofen use than those with the common genotype. The third study [37] reported no interaction for *CYP2C9* genotypes and Aspirin in a RCT of adenoma recurrence. The inconsistency of findings indicates that further, coordinated efforts will be needed to ascertain the interaction of variant *CYP2C9* SNPs and NSAID use.

3.3.6 *Other*

Other studies suggest interactions between Aspirin and variants of the genes coding for interleukins IL10, IL1 as well as TNFα and TGFβ [57], the insulin receptor substrate 1 (IRS1) [74], and the vitamin D receptor (VDR) [74]. One study [74] reported an interaction between *IRS1* genotype and Aspirin/NSAIDs use and risk of colorectal cancer, with a reduced OR observed in those with the one or more variant allele, who did not use NSAIDs. They also observed an interaction between Aspirin/NSAIDs use and the *VDR* gene. Having a variant alleles was associated with a reduced risk of colorectal cancer among non-Aspirin/NSAID users. The other study [57] reported a statistically

significant interaction between an *IL10* polymorphism and Aspirin use. Subjects who carried the variant allele and who had used Aspirin had a decreased OR versus non-users with the common genotype.

3.4 Discussion

This review identified multiple studies investigating potential interactions between polymorphisms in *COX1*, *COX2*, *ODC*, *UGT1A6* and *CYP2C9*, and NSAID use and the risk of colorectal adenoma or cancer. Single studies investigated interleukins *IL10*, *IL1* as well as *TNF*α and *TGF*β, the insulin receptor substrate 1 (*IRS1*), and the vitamin D receptor (*VDR*). The bulk of the identified research has focused on *COX2* polymorphisms. However, there were no statistically significant *COX2*-NSAID interactions identified, possibly attributable to inadequate power to detect true interactions. Interactions were reported for the *COX1* polymorphism Pro17Leu, *ODC* 315G>A, *UGT1A6* Thr181Ala/Arg184Ser and *CYP2C9* *2 and *3. The interactions for *ODC*, *UGT1A6* and *CYP2C9* have been replicated in more than one study. In summary, most studies were of modest sample size and there was substantial inter-study heterogeneity in the NSAID measure used. HWE was not consistently addressed in the studies reviewed. There is no strong evidence of any gene-NSAID interaction.

Understanding interaction of NSAID and genetics for colorectal cancer may allow optimization of the risk-benefit balance associated with NSAIDs, and improve the effectiveness of preventive use. However, the identification of true gene-NSAID interactions will require larger samples, and coordination across studies. Studies investigating potential interactions between NSAID use and genetic polymorphisms in inflammation have been of limited power due to inadequate sample size; studies with fewer than 400 cases and 400 controls are most likely underpowered for detecting most gene-NSAID interactions. [72] Achieving the sample sizes necessary to detect modest GxE interactions may be prohibitive in single studies, and may require harmonization of many studies. Particularly, the harmonization of exposure data across studies is vital to determining the role of interactions in human disease. [73]

Chapter 4

Assessment of SNP-NSAID interaction

4.1 Introduction

In this section, an assessesment of the joint effect of single nucleotide polymorphism (SNP) genetic variants and non-steroidal anti-inflammatory drug (NSAID) use and the risk of colorectal cancer (CRC) were evaluated in a large case-control study. We analysed data from the Assessment of Risk for Colorectal Tumors in Canada (ARCTIC) Study, a population-based genome-wide association (GWA) case-control study of CRC, [87]. Through a meta-analysis and systematic review [13], 8 polymorphisms in genes affecting xenobiotic metabolism (*CYP1A1, CYP2E1*); 4 affecting inflammation and immune response (*IL6*, *IL8*, *PPARg*, *TNF*α); and 2 polymorphisms identified from ARCTIC GWA and replicated in Scotland and Seattle (the 8q24 locus) were selected.

As a component of the original ARCTIC study, a χ^2 test was used to check for departures from Hardy-Weinberg equilibrium (HWE) [87]. No serious departures from HWE were detected. To address missing values for SNP covariates in multivariate modeling, missing genotypes were imputed using the haplotype clustering method implemented in the open-source software BEAGLE (See [33] for a full review of methods). The resulting data with imputations were used for all analysis. All analyses were carried out in SAS version 9.1 (SAS Institute, Cary, NC).

4.2 Materials and Methods

4.2.1 Study

The Assessment of Risk of Colorectal Tumors In Canada (ARCTIC) is a study involving 1,257 CRC cases and 1,336 matched community controls from the Ontario Familial Colorectal Cancer Registry. The population-based Ontario Cancer Registry is used to identify living, incident colorectal cancer cases (probands) aged 20-74 who were diagnosed between July 1, 1997, and June 30, 2000. Cases with known germline mutations that confer large familial risk of CRC (*APC*, *MSH2*, *MLH1*, *MSH6* or biallelic *MUTYH*) were excluded. Genotyping was performed using several large genotyping arrays involving markers for over 600,000 single nucleotide polymorphisms (SNPs). Full detail on genotyping methods is given elsewhere [87], as is the full study design [18].

The study also collected questionnaire-based epidemiological data on environmental factors and lifestyle, including Aspirin, Ibuprophen, acetominophen, or "Other NSAID" use. The questions pertaining to NSAID use were of the form "Did you take Aspirin/Acetaminophen/Ibuprofen two or more times per week during the past month?". Further data on to the frequency and duration of use were sought, but these were not included in analyses due to sparse coverage in the study population (i.e. no measurement for $\geq 40\%$ of the study population reporting NSAID use). There was insufficient data to determine whether the pattern of missingness was informative.

Polymorphisms were chosen based on their availability in ARCTIC, and their putative involvement with NSAID metabolism or inflammation, and on the basis of a previously published systematic review [13]. Additionally, the 8q24 risk allele was included because its association with CRC was newly identified in the original ARCTIC study, and its interaction with NSAIDs does not appear to have been previously assessed in the literature. Allele frequencies were calculated and can be found in 4.2. Variants which had been identified by the systematic review presented in this thesis (3) as potentially modifiers of NSAID chemoprevention of CRC were not available in ARCTIC.

4.2.2 Statistical Methods

All data analysis was performed with SAS 9.2 (SAS Institute). Univariate analyses of all candidate SNPs were conducted, and multivariate logistic regression analyses and model validation procedures were used to test for interactions. All multivariate models included sex, and the age at which the survey was completed.

In these analyses, both the genetic and environmental factors have been presented as dichotomous measures. Initially, analyses were performed using three levels for the genetic factor (homozygous wildtype, heterozygous, and homozygous variant), but this lead to small cell sizes with uninterpretable results. It was thus decided to collapse the genetic factor into two categories. The possibility that this choice of a dichotomous genetic model may have influenced the results was explored by removing homozygous variant cases and controls and re-running the univariate analyses. This did not substantially change the observed effects.

Several indexes of additive interaction (departure from additivity) were calculated. First, odds ratios (OR) and 95% confidence intervals (95% CI) were estimated for each exposure category (NSAID user/common genotype, NSAID non-user/variant genotype, and NSAID user/variant genotype) versus NSAID non-user/common genotype subjects. The relative excess risk due to interaction (RERI), and attributable proportion (AP; the proportion of disease among those with both "exposures" that is attributable to their interaction) [56] were calculated. Ninety-five % CIs were estimated for each of these measures using the SAS program by [56]. The null values of RERI and AP are 0.

Missing data for at least one polymorphism was common, so missing data were imputed. Missing genotypes were imputed using the haplotype clustering method implemented in the open-source software BEAGLE [12]. This imputation was performed by the Public Health Genomic Applied Research Ottawa group's statistician (Steven Hawken; SH).

Next, a multivariate predictive model was developed using stepwise logistic regression to construct models that concentrated NSAID responders. Receiver operator characteristic (ROC) curves which plot sensitivity versus the compliment of specificity were used to assess the performance of multivariate logistic regression models. Discrimination of the model was tested using areas under the ROC curves (AUC). ROC plots sensitivity versus the compliment of specificity for all possible thresholds. AUC gives the probability that, given a vector of covariates (i.e., the sex, age, NSAID status, and genotypes of an individual subject), the model correctly assigns cases and controls. An AUC of 0.5 indicates discrimination no better than chance, while an AUC of 1.0 indicates perfect discrimination. This is equivalent to the probability that a predictive model (or more generally, any diagnostic test) correctly assigns a higher probability of being a case to actual cases, and vice-versa for controls [32]. AUC was measured using the c-statistic (concordance index) [43]. The calibration of the final model was also tested. Calibration quantifies how closely the predicted probabilities of case status given a vector of covariate

values associated with the participant (e.g. age, sex, NSAID status, genotype) match their actual status (case vs. control). This was assessed using the Hosmer-Lemeshow (HL) goodness-of-fit test.

Model validaty was estimated using the Akaike information criterion (AIC), which is an entropy-based measure which penalizes model complexity. AIC is commonly used to compare logistic regression models, and is defined by the equation:

$$AIC = -2logL(M) + 2 * K,$$

where logL(M) is the logarithm of the maximum likelihood for the fitted model, N is the sample size, and K is the number of covariates including the intercept. AIC is intended to serve as measure of cross validation - models with the lowest AIC score are expected to perform best when tested in new data. Models with smaller AIC scores are expected to have better external validity ([69], page 426). This assumption is used here to rank model validity, but models developed using this method will require further validation using an external dataset.

4.3 Results

Demographic and other selected characteristics of cases and controls are given in Table 4.1. The mean age of participants was 62.9 (SD=8.4) years for cases and 63.6 (SD=8.6) years for controls. A higher proportion of cases than controls reported using NSAIDs (54% vs. 48%). Cases were more likely to be missing data for all variables.

Table 4.1: Demographic characteristics of participants

		Cases (%)		Controls (%)	
Total		1133		1125	
Sex					
	Female	598	(53)	491	(44)
	Male	407	(36)	632	(56)
	Missing	128	(11)	2	(<1)
Age					
	Under 40	10	(1)	8	(1)
	40 to 49	54	(5)	60	(5)
	50 to 59	261	(23)	258	(23)
	60 to 69	437	(39)	489	(43)
	70 or over	243	(21)	302	(27)
	Missing	128	(11)	8	(1)
NSAID					
	Never User	375	(33)	576	(51)
	Ever User	608	(54)	534	(48)
	Missing	151	(13)	14	(1)
Aspirin					
	Never User	701	(26)	664	(40)
	Ever User	294	(62)	448	(59)
	Missing	138	(12)	13	(1)
Ibuprofen					
	Never User	857	(76)	927	(82)
	Ever User	120	(11)	180	(16)
	Missing	156	(13)	18	(2)
Acetaminophen					
	Never User	161	(14)	934	(83)
	Ever User	831	(73)	177	(16)
	Missing	141	(13)	14	(1)

Allele frequencies and minor allele frequencies (MAF) were calculated for each of the SNPs considered in these analyses. These can be found in Table 4.2.

Table 4.2: Allele frequencies for selected SNPs

Pathway		Cases				Controls			
		MM	Mm	mm	MAF	MM	Mm	mm	MAF
Xenobiotic metabolism									
cyp1a1 T461N	rs1799814	1037	94	9	0.049	1046	78	6	0.040
cyp1a1 m2 (a462g BsaI)	rs1048943	1021	106	7	0.053	1028	92	6	0.046
cyp1a2 -164C¿A rsq=0.622	rs1378942	472	433	138	0.340	423	471	156	0.373
cyp1a2 -164C¿A rsq=1.0	rs2472300	572	398	89	0.272	547	426	96	0.289
cyp2E1 RsaI	rs2031920	941	53	2	0.029	925	73	1	0.038
cyp2E1 PstI rsq=1.0	rs2070674	980	55	1	0.028	974	82	1	0.040
MDR1 -129T¿C rsq=1.0	rs2188525	1131	98	1	0.041	1116	93	0	0.038
mdr1 exon 1b 2677G¿T/A rsq=1.0	rs4148738	316	522	220	0.455	302	540	226	0.464
Inflammation/Cytochine signalling									
IL6 -174G¿C	rs1800795	349	485	141	0.393	386	517	184	0.407
TNF-alpha -308G¿A	rs3099844	772	211	25	0.129	858	252	13	0.124
PPARG P10A	rs1801282	919	303	14	0.134	978	227	22	0.1108
IRS2 snp1	rs2289046	450	432	114	0.331	488	499	127	0.338
Other									
Arctic 8q24	rs10505477	354	553	248	0.454	287	581	300	0.506
ARCTIC 8q24 rsq=0.569	rs10956369	322	516	211	0.4479	375	518	165	0.401

4.3.1 Preliminary assessment of interaction

The ORs for exposure specific and joint gene-NSAID effects, as well as the two indexes of additive interaction, are given below. "Any NSAID" was defined as reported use of Aspirin, Ibuprofen, or both. Odds ratios were calculated comparing NSAID non-users with the common genotype to NSAID non-users with the variant genotype, NSAID users with the common genotype, and NSAID users with the variant genotype. Results for metabolic genes are given in Table 4.3, results for Inflammatory genes are given in Table 4.4, and 8q24 SNPs are given in Table 4.11. These SNP-NSAID interactions were considered in individual logistic models.

Any NSAID

Table 4.3: Stratified and joint gene-Non Steroidal Anti-Inflammatory Drug Odds Ratios, Relative Excess Risk due to Interactions, and Attributable Proportions with 95% Confidence Intervals for Single Nucleotide Polymorphisms in Metabolic genes

NSAID status, genotype	OR	(95% CI)
CYP1A1		
rs1799814		
Non-user, Common	1.00	Reference
Non-user, Variant	1.28	(0.85-1.93)
User, Common	0.69	(0.58-0.83)
User, Variant	0.91	(0.55-1.48)
	Interaction effects	
RERI	-0.07	(-1.02 to 0.89)
AP	-0.08	(-1.04 to 0.89)
rs1048943		
Non-user, Common	1.00	Reference
Non-user, Variant	1.15	(0.74-1.8)
User, Common	0.69	(0.57-0.82)
User, Variant	0.85	(0.54-1.35)
	Interaction effects	
RERI	0.01	(-0.90 to 0.92)
AP	0.01	(-0.97 to 1.00)
CYP1A2		
rs1378942		
Non-user, Common	1.00	Reference
Non-user, Variant	0.93	(0.74-1.17)
User, Common	0.65	(0.50-0.86)
User, Variant	0.67	(0.53-0.86)
	Interaction effects	
RERI	0.09	(-0.24 to 0.42)
AP	0.13	(0.36 to 0.62)
rs2472300		
Non-user, Common	1.00	Reference
Non-user, Variant	1.04	(0.83-1.30)
User, Common	0.67	(0.52-0.86)
User, Variant	0.74	(0.58-0.95)
	Interaction effects	
RERI	0.03	(-0.33 to 0.40)
AP	0.05	(-0.44 to 0.53)
CYP2E1		
rs2031920		
Non-user, Common	1.00	reference
Non-user, Variant	0.94	(0.58-1.52)
User, Common	0.71	(0.59-0.85)
User, Variant	0.46	(0.26-0.84)
	Interaction effects	
RERI	-0.18	(-1.08 to 0.71)
AP	-0.39	(-2.13 to 1.34)
rs2070674		
Non-user, Common	1.00	Reference
Non-user, Variant	0.82	(0.51-1.33)
User, Common	0.71	(0.59-0.85)
User, Variant	0.42	(0.23-0.76)
	Interaction effects	
RERI	-0.11	(-0.91 to 0.69)
AP	-0.25	(-1.96 to 1.45)
MDR1		
rs2188525		
Non-user, Common	1.00	Reference
Non-user, Variant	1.02	(0.68-1.52)
User, Common	0.69	(0.57-0.83)
User, Variant	0.75	(0.46-1.22)
	Interaction effects	
RERI	0.04	(-0.74 to 0.82)

Table 4.3 – Continued

AP	0.06	(-0.89 to 1.00)
rs4148738		
Non-user, Common	1.00	Reference
Non-user, Variant	0.92	(0.72-1.19)
User, Common	0.68	(0.49-0.94)
User, Variant	0.65	(0.49-0.84)
	Interaction effects	
RERI	0.04	(-0.29 to 0.37)
AP	0.07	(-0.47 to 0.61)

The OR comparing NSAID users with two copies of the common *CYP1A1* rs1799814 allele to non-users also having the common genotype was 0.69 (95% CI 0.58-0.83), while this reduction in risk was not observed in users with the variant genotype (OR=0.91, 95%CI 0.55-1.48). A similar result was observed for the *CYP1A1* rs1048943 polymorphism, where NSAID users with the common genotype showed a reduced OR (0.69, 95% CI 0.57-0.82) which was not present among carriers of the variant genotype (OR=0.85, 95%CI 0.54-1.35).

For *CYP2E1*, the OR was reduced further in NSAID users carrying a variant allele than those with two copies of the common allele. This was observed for both the rs2031920 and the rs2070674 polymorphism. For rs2031920, the difference was 25% (0.46, 95% CI 0.26-0.84, versus 0.71, 95% CI 0.59-0.85). For rs2070674, the difference was 29% (0.42, 95% CI 0.23-0.76, versus 0.71, 95% CI 0.59-0.85).

Table 4.4: Stratified and joint gene-NSAID ORs, RERIs, and APs with 95% CIs for SNPs in Inflammatory genes

NSAID status, genotype	OR	(95% CI)
IL6		
rs1800795		
Non-user, Common	1.00	Reference
Non-user, Variant	1.07	(0.85-1.36)
User, Common	0.81	(0.60-1.09)
User, Variant	0.69	(0.53-0.89)
	Interaction effects	
RERI	-0.19	(-0.58 to 0.20)
AP	-0.28	(-0.84 to 0.29)
TNFα		
rs3099844		
Non-user, Common	1.00	Reference
Non-user, Variant	1.06	(0.81-1.39)
User, Common	0.75	(0.61-0.92)
User, Variant	0.56	(0.40-0.76)
	Interaction effects	
RERI	-0.25	(-0.77 to 0.27)
AP	-0.46	(-1.34 to 0.42)
PPARg		
rs1801282		
Non-user, Common	1.00	Reference
Non-user, Variant	1.28	(0.98-1.67)
User, Common	0.69	(0.56-0.85)
User, Variant	0.91	(0.67-1.24)
	Interaction effects	
RERI	-0.06	(-0.64 to 0.52)
AP	-0.06	(-0.65 to 0.53)
IRS2		
rs2289046		
Non-user, Common	1.00	Reference
Non-user, Variant	1.07	(0.85-1.35)
User, Common	0.77	(0.59-1.01)
User, Variant	0.69	(0.53-0.89)
	Interaction effects	
RERI	-0.15	(-0.53 to 0.22)
AP	-0.23	(-0.77 to 0.32)

It was observed that some carriers of variant alleles in inflammatory genes may have derived greater benefit from NSAID chemoprevention. The OR comparing NSAID users with two copies of the common *IL6* rs1800795 allele to non-users also having the common genotype was 0.81 (95% CI 0.60-1.09), while that of NSAID users with the variant genotype was reduced further (OR=0.69, 95%CI 0.53-0.89). A similar effect was observed for the *TNF*α polymorphism rs3099844. The OR for NSAID users with the common genotype was 0.75 (95% CI 0.61-0.92), while that of users with the variant genotype was reduced further still (OR=0.56, 95%CI 0.40-0.76).

A different relationship was observed with *PPARg* SNP rs1801282. A statistically significant decrease in OR was observed only for NSAID users with the common genotype (OR=0.69, 95%CI 0.56-0.85), while users with the variant genotype showed no such decrease (OR=0.91, 95%CI 0.67-1.24).

Table 4.5: Stratified and joint gene-NSAID ORs, RERIs, and APs with 95% CIs for SNPs at the 8q24 locus

NSAID status, genotype	OR	(95% CI)
ARCTIC 8q24		
rs10505477		
Non-user, Common	1.00	Reference
Non-user, Variant	0.73	(0.56-0.94)
User, Common	0.66	(0.47-0.93)
User, Variant	0.51	(0.39-0.67)
	Interaction effects	
RERI	0.12	(-0.18 to 0.43)
AP	0.24	(-0.40 to 0.88)
rs10956369		
Non-user, Common	1.00	Reference
Non-user, Variant	1.06	(0.83-1.35)
User, Common	0.56	(0.41-0.76)
User, Variant	0.82	(0.63-1.06)
	Interaction effects	
RERI	0.21	(-0.11 to 0.53)
AP	0.25	(-0.16 to 0.66)

Two polymorphisms at the 8q24 locus were investigated for potential interactions with NSAID use. Having one or more copy of the minor allele for the rs10505477 SNP was associated with a decreased OR for NSAID non-users (OR=0.73, 95%CI 0.56-0.94), while the effect was greater still for users of NSAIDs with the variant genotype (OR=0.51, 95%CI 0.39-0.67). No statistically significant departure from additivity was observed for this allele (RERI=0.12, 95%CI -0.18 to 0.43; AP=0.24, 95%CI -0.40 to 0.88).

The relationship observed for the rs10956369 polymorphism was essentially the converse. Having one or more copy of the minor allele did not decrease the OR for non-NSAID users (OR=1.06, 95%CI 0.83-1.35), and users of NSAIDs with the variant genotype showed less of a decrease in OR than did users with the common genotype (OR=0.82,

95%CI 0.63-1.06 vs. OR=0.56, 95%CI 0.41-0.76). No statistically significant departure from additivity was observed for this allele (RERI=0.12, 95%CI -0.18 to 0.43; AP=0.24, 95%CI -0.40 to 0.88). Next, NSAID use was examined by NSAID type (Aspirin or Ibuprofen).

Aspirin

The ORs for exposure specific and the joint gene-Aspirin effects, as well as the two measures of additive interaction, are given below. Results for metabolic genes are given in Table 4.6, results for inflammatory genes are given in Table 4.7, and 8q24 SNPs are given in Table 4.8.

Table 4.6: Stratified and joint gene-Aspirin ORs, RERIs, and APs with 95% CIs for SNPs in Metabolic genes

Aspirin status, genotype	OR	(95% CI)
CYP1A1		
rs1799814		
Non-user, Common	1.00	Reference
Non-user, Variant	1.26	(0.86-1.86)
User, Common	0.67	(0.55-0.82)
User, Variant	0.93	(0.55-1.58)
	Interaction effects	
RERI	0.00	(-0.95 to 0.95)
AP	0.00	(-0.91 to 0.91)
rs1048943		
Non-user, Common	1.00	Reference
Non-user, Variant	1.20	(0.79-1.84)
User, Common	0.67	(0.55-0.82)
User, Variant	0.81	(0.50-1.33)
	Interaction effects	
RERI	-0.06	(-0.99 to 0.87)
AP	-0.07	(-1.11 to 0.96)
CYP1A2		
rs1378942		
Non-user, Common	1.00	Reference
Non-user, Variant	0.96	(0.77-1.19)
User, Common	0.65	(0.49-0.87)
User, Variant	0.66	(0.51-0.86)
	Interaction effects	
RERI	0.05	(-0.29 to 0.39)
AP	0.08	(-0.42 to 0.58)
rs2472300		
Non-user, Common	1.00	Reference
Non-user, Variant	1.05	(0.84-1.30)
User, Common	0.65	(0.49-0.84)
User, Variant	0.74	(0.57-0.95)
RERI	Interaction effects	
AP	0.05	(-0.32 to 0.41)
	0.06	(-0.41 to 0.54)
CYP2E1		
rs2031920		
Non-user, Common	1.00	Reference
Non-user, Variant	0.86	(0.54-1.37)
User, Common	0.68	(0.56-0.83)
User, Variant	0.52	(0.28-0.96)
	Interaction effects	
RERI	-0.03	(-0.85 to 0.80)
AP	-0.05	(-1.44 to 1.34)
rs2070674	1.00	Reference
Non-user, Common	0.76	(0.48-1.21)
Non-user, Variant	0.68	(0.56-0.83)
User, Common	0.47	(0.25-0.85)
User, Variant		
	Interaction effects	
RERI	0.02	(-0.73 to 0.76)
AP	0.04	(-1.36 to 1.44)
MDR1		
rs2188525		
Non-user, Common	1.00	Reference
Non-user, Variant	1.06	(0.72-1.56)
User, Common	0.68	(0.56-0.82)
User, Variant	0.72	(0.43-1.21)
	Interaction effects	
RERI	-0.02	(-0.83 to 0.80)
AP	-0.02	(-1.03 to 0.98)

Table 4.6 – Continued

rs4148738		
Non-user, Common	1.00	Reference
Non-user, Variant	0.93	(0.74-1.18)
User, Common	0.66	(0.47-0.92)
User, Variant	0.64	(0.49-0.84)
	Interaction effects	
RERI	0.05	(-0.29 to 0.38)
AP	0.07	(-0.47 to 0.61)

The OR comparing Aspirin users users with two copies of the common *CYP1A1* rs1799814 allele to non-users also having the common genotype was 0.67 (95% CI 0.55-0.82), while this reduction in risk was not observed in users with the variant genotype (OR=0.93, 95%CI 0.55-1.58). A similar result was observed for the *CYP1A1* rs1048943 polymorphism, where Aspirin users with the common genotype showed a reduced OR (0.67, 95% CI 0.55-0.82) which was not observed in carriers of the variant genotype (OR=0.81, 95%CI 0.50-1.33).

Table 4.7: Stratified and joint gene-Aspirin ORs, RERIs, and APs with 95% CIs for SNPs in Inflammatory genes

Aspirin status, genotype	OR	(95% CI)
IL6		
rs1800795		
Non-user, Common	1.00	Reference
Non-user, Variant	1.03	(0.82-1.29)
User, Common	0.73	(0.53-1.01)
User, Variant	0.67	(0.51-0.87)
	Interaction effects	
RERI	-0.09	(-0.46 to 0.28)
AP	-0.14	(-0.69 to 0.41)
TNFα		
rs3099844		
Non-user, Common	1.00	Reference
Non-user, Variant	1.04	(0.80-1.34)
User, Common	0.73	(0.59-0.91)
User, Variant	0.53	(0.38-0.75)
	Interaction effects	
RERI	-0.24	(-0.75 to 0.28)
AP	-0.45	(-1.33 to 0.44)
PPARg		
rs1801282		
Non-user, Common	1.00	Reference
Non-user, Variant	1.26	(0.98-1.63)
User, Common	0.66	(0.53-0.82)
User, Variant	0.91	(0.66-1.27)
	Interaction effects	
RERI	-0.01	(-0.58 to 0.56)
AP	-0.01	(-0.58 to 0.56)
IRS2		
rs2289046		
Non-user, Common	1.00	Reference
Non-user, Variant	1.12	(0.90-1.39)
User, Common	0.82	(0.62-1.08)
User, Variant	0.66	(0.50-0.85)
	Interaction effects	
RERI	-0.28	(-0.69 to 0.12)
AP	-0.43	(-1.04 to 0.18)

It was observed that some carriers of variant alleles may have benefitted more from Aspirin chemoprevention. The OR comparing Aspirin users with the common genotype *TNFα* rs3099844 to non-users with the common genotype was 0.73 (95% CI 0.59-0.91), while users with the variant genotype showed a greater reduction in risk (OR=0.53, 95%CI 0.38-0.75). For this SNP, the RERI was -0.24 (95% CI -0.75 to 0.28), and the AP was -0.45 (95% CI -1.33 to 0.44). Neither is statistically significant, and thus chance variation may be the cause of the effect, but the point estimates indicate that there is a 24% reduction in risk due to interaction, and that 45% of the total reduction in risk is attributable to the interaction of genotype and Aspirin use.

A similar relationship was observed for the *IRS2* SNP rs2289046. The OR comparing Aspirin users with the common genotype to non-users with the common genotype was 0.82 (95% CI 0.62-1.08), while users with the variant genotype showed a greater reduction in risk (OR=0.66, 95%CI 0.50-0.85). or this SNP, the RERI was -0.28 (95% CI -0.69 to 0.12), and the AP was -0.43 (95% CI -1.04 to 0.18). Neither is statistically significant, and thus chance variation may be the cause of the effect, but the point estimates indicate that there is a 28% reduction in risk due to interaction, and that 45% of the total reduction in risk is attributable to the interaction of genotype and Aspirin use.

The converse was observed for the *PPARg* SNP rs1801282. While Aspirin users with the common genotype showed a reduced OR (0.66, 95% CI 0.53-0.82), those with the variant genotype did not show a benefit (OR=0.91, 95% CI 0.66-1.27).

Table 4.8: Stratified and joint gene-Aspirin ORs, RERIs, and APs with 95% CIs for SNPs at the 8q24 locus

Aspirin status, genotype	OR	(95% CI)
8q24		
rs10505477		
Non-user, Common	1.00	Reference
Non-user, Variant	0.74	(0.58-0.94)
User, Common	0.66	(0.46-0.94)
User, Variant	0.51	(0.38-0.66)
	Interaction effects	
RERI	0.10	(-0.21 to 0.42)
AP	0.21	(-0.45 to 0.86)
rs10956369		
Non-user, Common	1.00	Reference
Non-user, Variant	1.25	(0.99-1.57)
User, Common	0.69	(0.50-0.96)
User, Variant	0.83	(0.64-1.08)
	Interaction effects	
RERI	-0.11	(-0.51 to 0.29)
AP	-0.13	(-0.61 to 0.34)

Two polymorphisms at the 8q24 locus were investigated for potential interactions with Aspirin use. Having one or more copy of the minor allele for the rs10505477 SNP was associated with a decreased OR for Aspirin non-users (OR=0.74, 95%CI 0.58-0.94), while the effect was greater still for users of Aspirin with the variant genotype (OR=0.51, 95%CI 0.38-0.66). No statistically significant departure from additivity was observed for this allele (RERI=0.10, 95%CI -0.21 to 0.42; AP=0.21, 95%CI -0.45 to 0.86).

The relationship observed for the rs10956369 polymorphism was essentially the converse. Having one or more copy of the minor allele did not decrease the OR for Aspirin non-users (OR=1.25, 95%CI 0.99-1.57), and users of Aspirin with the variant genotype showed less of a decrease in OR than did users with the common genotype (OR=0.83,

95%CI 0.64-1.08 vs. OR=0.69, 95%CI 0.50-0.96). No statistically significant departure from additivity was observed for this allele (RERI= -0.11, 95%CI -0.51 to 0.29; AP= -0.13, 95%CI -0.61 to 0.34).

Ibuprofen

Exposure specific and the joint gene-Ibuprofen ORs, as well as the two measures of additive interaction, are given below. Results for Metabolic genes are given in Table 4.3, results for inflammatory genes are given in Table 4.4, and 8q24 SNPs are given in Table 4.11.

Table 4.9: Stratified and joint gene-Ibuprofen ORs, RERIs, and APs with 95% CIs for SNPs in metabolic genes

Ibuprofen status, genotype	OR	(95% CI)
CYP1A1		
rs1799814		
Non-user, Common	1.00	Reference
Non-user, Variant	1.38	(0.98-1.93)
User, Common	0.67	(0.52-0.88)
User, Variant	0.75	(0.30-1.88)
	Interaction effects	
RERI	-0.31	(-1.74 to 1.12)
AP	-0.41	(-1.99 to 1.16)
rs1048943		
Non-user, Common	1.00	Reference
Non-user, Variant	1.10	(0.77-1.57)
User, Common	0.65	(0.50-0.84)
User, Variant	0.89	(0.39-2.06)
	Interaction effects	
RERI	0.15	(-0.94 to 1.23)
AP	0.16	(-0.79 to 1.12)
CYP1A2		
rs1378942		
Non-user, Common	1.00	Reference
Non-user, Variant	0.93	(0.77-1.13)
User, Common	0.58	(0.38-0.87)
User, Variant	0.67	(0.48-0.93)
	Interaction effects	
RERI	0.16	(-0.25 to 0.56)
AP	0.23	(-0.29 to 0.76)
rs2472300		
Non-user, Common	1.00	Reference
Non-user, Variant	1.02	(0.85-1.24)
User, Common	0.64	(0.45-0.92)
User, Variant	0.69	(0.49-0.99)
	Interaction effects	
RERI	0.03	(-0.43 to 0.49)
AP	0.04	(-0.53 to 0.61)
CYP2E1		
rs2031920		
Non-user, Common	1.00	Reference
Non-user, Variant	0.86	(0.58-1.27)
User, Common	0.68	(0.52-0.88)
User, Variant	0.33	(0.10-1.04)
	Interaction effects	
RERI	-0.21	(-1.33 to 0.92)
AP	-0.63	(-3.15 to 1.90)
rs2070674		
Non-user, Common	1.00	Reference
Non-user, Variant	0.74	(0.50-1.10)
User, Common	0.67	(0.52-0.87)
User, Variant	0.33	(0.10-1.03)
	Interaction effects	
RERI	-0.09	(-1.07 to 0.9)
AP	-0.03	(-2.40 to 1.87)
MDR1		
rs2188525		
Non-user, Common	1.00	Reference
Non-user, Variant	0.96	(0.68-1.34)
User, Common	0.64	(0.49-0.84)
User, Variant	0.84	(0.37-1.94)
	Interaction effects	
RERI	0.24	(-0.70 to 1.18)
AP	0.29	(-0.56 to 1.14)

Table 4.9 – Continued

rs4148738		
Non-user, Common	1.00	Reference
Non-user, Variant	0.97	(0.78-1.19)
User, Common	0.71	(0.46-1.10)
User, Variant	0.61	(0.44-0.86)
	Interaction effects	
RERI	-0.06	(-0.52 to 0.39)
AP	-0.10	(-0.80 to 0.59)

The OR comparing Ibuprofen users with two copies of the common *CYP1A1* rs1799814 allele to non-users also having the common genotype was 0.67 (95% CI 0.52-0.88), while this reduction in risk was not observed in users with the variant genotype (OR=0.75, 95%CI 0.30-1.88). A similar result was observed for the *CYP1A1* rs1048943 polymorphism, where Ibuprofen users with the common genotype showed a reduced OR (0.65, 95% CI 0.50-0.84) which was not present among carriers of the variant genotype (OR=0.89, 95%CI 0.39-2.06).

For *CYP2E1*, the OR was reduced further in Ibuprofen users carrying a variant allele than those with two copies of the common allele. This was observed for both the rs2031920 (OR=0.33, 95% CI 0.10-1.04; versus 0.68, 95% CI 0.52-0.88) and the rs2070674 polymorphism (OR=0.33, 95% CI 0.10-1.03; versus 0.67, 95% CI 0.52-0.87).

Table 4.10: Stratified and joint gene-Ibuprofen ORs, RERIs, and APs with 95% CIs for SNPs in Inflammatory genes

Ibuprofen status, genotype	OR	(95% CI)
IL6		
rs1800795		
Non-user, Common	1.00	Reference
Non-user, Variant	1.03	(0.85-1.26)
User, Common	0.82	(0.54-1.25)
User, Variant	0.61	(0.43-0.85)
	Interaction effects	
RERI	-0.24	(-0.75 to 0.26)
AP	-0.40	(-1.19 to 0.39)
TNFα		
rs3099844		
Non-user, Common	1.00	Reference
Non-user, Variant	0.94	(0.75-1.17)
User, Common	0.68	(0.51-0.90)
User, Variant	0.56	(0.33-0.94)
	Interaction effects	
RERI	-0.06	(-0.65 to 0.53)
AP	-0.11	(-0.96 to 0.75)
PPARg		
rs1801282		
Non-user, Common	1.00	Reference
Non-user, Variant	1.27	(1.02-1.59)
User, Common	0.67	(0.50-0.89)
User, Variant	0.81	(0.48-1.36)
	Interaction effects	
RERI	-0.13	(-0.90 to 0.63)
AP	-0.16	(-0.94 to 0.62)
IRS2		
rs2289046		
Non-user, Common	1.00	Reference
Non-user, Variant	0.97	(0.80-1.17)
User, Common	0.60	(0.41-0.87)
User, Variant	0.69	(0.49-0.98)
	Interaction effects	
RERI	0.13	(-0.29 to 0.55)
AP	0.19	(-0.33 to 0.70)

It was observed that some carriers of variant alleles in inflammatory genes may have derived greater benefit from Ibuprofen chemoprevention. The OR comparing Ibuprofen users with the common genotype for the *IL6* rs1800795 SNP to non-users with the common genotype was 0.82 (95% CI 0.54-1.25), while that comparing Ibuprofen users with the variant genotype to non-users with the common genotype was 0.61 (95% CI 0.43-0.85). For this SNP, the RERI was -0.24 (95% CI -0.75 to 0.26), and the AP was -0.40 (95% CI -1.19 to 0.39). Neither is statistically significant, and thus chance variation may be the cause of the effect, but the point estimates indicate a 24% reduction in risk due to the interaction of Ibuprofen and genotype, and that 45% of the total reduction in risk is attributable to the interaction of genotype and Ibuprofen use.

The converse was observed for the *PPARg* SNP rs1801282. While Aspirin users with the common genotype showed a reduced OR (0.67, 95% CI 0.50-0.89), those with the variant genotype did not show a benefit (OR=0.81, 95% CI 0.48-1.36).

Table 4.11: Stratified and joint gene-NSAID ORs, RERIs, and APs with 95% CIs for SNPs at the 8q24 locus

Ibuprofen status, genotype	OR	(95% CI)
rs10505477		
Non-user, Common	1.00	Reference
Non-user, Variant	0.75	(0.61-0.93)
User, Common	0.72	(0.44-1.16)
User, Variant	0.48	(0.35-0.67)
	Interaction effects	
RERI	0.01	(-0.41 to 0.44)
AP	0.03	(-0.84 to 0.89)
rs10956369		
Non-user, Common	1.00	Reference
Non-user, Variant	1.16	(0.95-1.42)
User, Common	0.52	(0.33-0.80)
User, Variant	0.88	(0.63-1.22)
	Interaction effects	
RERI	0.20	(-0.24 to 0.65)
AP	0.23	(-0.20 to 0.66)

Two polymorphisms at the 8q24 locus were investigated for potential interactions with Ibuprofen use. Having one or more copy of the minor allele for the rs10505477 SNP was associated with a decreased OR for Ibuprofen non-users (OR=0.75, 95%CI 0.61-0.93), while the effect was greater still for users of Ibuprofen users with the variant genotype (OR=0.48, 95%CI 0.35-0.67).

The relationship observed for the rs10956369 polymorphism was essentially the converse. Having one or more copy of the minor allele did not decrease the OR for Ibuprofen non-users (OR=1.16, 95%CI 0.95-1.42), and users of Ibuprofen with the variant genotype did not benefit from Ibuprofen use (OR=0.88, 95%CI 0.63-1.22).

Acetaminophen

Acetaminophen has very weak anti-inflammatory properties [60], and would thus not be expected to demonstrate the same interaction effects or preventive properties as Aspirin or Ibuprofen. As a control for the possibility that observed interactions were the result of chance, interactions between acetaminophen use and the SNPs listed above were tested using the same procedure as previously described. None of the differential responses by genotype, which were observed for both Aspirin and Ibuprofen, were observed for acetaminophen. All subgroup ORs included the null value, save for the 8q24 SNP rs10956369. Acetaminophen non-users with the common genotype had a reduced OR (0.78, 95% CI 0.67-0.97), as did users with the variant genotype (OR=0.71, 95% CI 0.52-0.97), compared to users with the common genotype. This result is not biologically coherent, and is unlikely to represent more than the result of chance.

4.3.2 Multivariate analysis

Finally, based on the results from the above analysis, stepwise logistic regression was used to construct models that concentrated NSAID responders. SNP-environment interactions that achieved nominal significance ($P<0.10$) in individual models were included as candidates in multivariate modeling. Once included in a model, the threshold for removal was $P>0.20$. Model hierarchy was respected in the modeling process, meaning in order for an interaction to be included in the model, both of the main effects were forced into the model. Seven SNPs that demonstrated a differential response by genotype were considered. These were *CYP1A1* rs1799814 and rs1048943; *CYP2E1* rs2031920 and rs2070674; *MDR1* rs2188525; *IL6*rs1800795 and rs3099844; *PPARg* rs1801282 and rs2289046; and 8q24 rs10956369.

These SNPs and other NSAID covariates were recoded so that the protective or predisposing class was the comparator. A simple model containing only age and sex was compared to more complex models containing information on genotype and NSAID use. Plots of the ROC curves with AUC are given in Figure 4.1, Figure 4.2, and Figure 4.3, below. Numeric results (AIC, AUC, and the HL test) for the three models are given in Table 4.12, below.

The final model from this procedure included variants: rs1799814, rs1801282, and rs10956369, and the interaction of each SNP with NSAID use. AUC as measured by the c-statistic (concordance index) was 0.57. Models additionally including age and sex gave an AUC of 62%. In contrast, a model with only age and sex gave an AUC of 59%, and

Table 4.12: Numeric results for AIC, AUC, and the HL test in three models.

	Age and Sex Score		Age, Sex, and NSAID use Score		Full Model Score	
AIC	2885.696		2849.054		2843.947	
AUC	0.59		0.60		0.62	
HL (p)	27.4063	(0.0006)	10.7557	(0.2159)	13.6796	(0.0905)

a model with age, sex, and NSAID use gave an AUC of 60%.

The full model, containing SNPs, NSAID use, their interaction, sex, and age was better calibrated than the model without the interaction terms (χ^2=13.6795, p=0.09 versus χ^2=10.7557, p=0.22). The full model is anticipated to have better external validity, with an AIC score of 2843.947. The model without interaction terms has an AIC score of 2849.054.

4.4 Discussion

This study investigated the potential GxE interaction of multiple genetic factors and NSAID use in a large case-control study. Based on the RERI and AP criteria, no statistically significant interaction between use of any NSAID and the investigated genes was observed. All results for the two measures of additive interaction used in these analyses included the null value. However, several apparent differences in the response to NSAID use were observed. Substituting Aspirin and Ibuprofen as the environmental exposure did not perceptibly alter the point estimates and confidence intervals, while including Acetaminophen instead of Aspirin or Ibuprofen resulted in large differences for all SNPs.

One result of potential interest was observed in two polymorphisms at the 8q24 locus, where rs10956369 showed a statistically significant decrease in the odds of CRC among NSAID users with the common genotype (OR=0.65, 95% CI 0.41-0.76), while NSAID users with the variant genotype did not show a decrease in risk (OR=0.82, 95% OR 0.63-1.06). The two measures of additive interaction used in this study indicated a potentially interesting relationship between NSAID use and the variant genotype, with a RERI of 0.21 (95% CI -0.11 to 0.53), and an AP of 0.25 (95% CI -0.16 to 0.66). Although these

results are not statistically significant, they appear to warrant further investigation.

Another result that warrants further investigation is that of several genes encoding proteins involved in the inflammatory response. The OR comparing NSAID users with two copies of the common *IL6* rs1800795 allele to non-users also having the common genotype was 0.81 (95% CI 0.60-1.09), while that of NSAID users with the variant genotype was reduced further (OR=0.69, 95%CI 0.53-0.89). A similar effect was observed for the *TNF*α polymorphism rs3099844. The OR for NSAID users with the common genotype was 0.75 (95% CI 0.61-0.92), while that of users with the variant genotype was reduced further still (OR=0.56, 95%CI 0.40-0.76). These results are consistent with biological knowledge of these genes.

Results were explored further in multivariate modelling. Stepwise model building incorporating SNP-NSAID interactions resulted in a model containing *CYP1A1* rs1799814, *PPARg* rs1801282, and 8q24 rs10956369. The functional significance of rs1799814 is unclear; however, *CYP1A1* is involved in xenobiotic metabolism, and rs1799814 has previously been implicated as a modifier of lung cancer risk in smokers. The functional significance of *PPARg* rs1801282 is also unclear. The functional significance of the 8q24 locus has not been fully elucidated. However, there is some indication that it may be involved in the function of the *MYC* gene, which encodes a transcription factor that may regulate expression of many genes. [2] This means that in addition to its role as a classical transcription factor, MYC also functions to regulate global chromatin structure by regulating histone acetylation both in gene-rich regions and at sites far from any known gene. [2]. These variants may affect transcription or expression of relevant genes, or the observed interactions may be due to chance.

Ideally, predictive models made in one dataset should be validated against new data. This is a necessary step in the development of useful model, and the model developed here would need to be validated in an external dataset. The original ARCTIC study used a case-control dataset from Scotland [87] to validate findings. In order to test the validity of the model developed in this thesis, these data will be sought and the model developed here applied to this data.

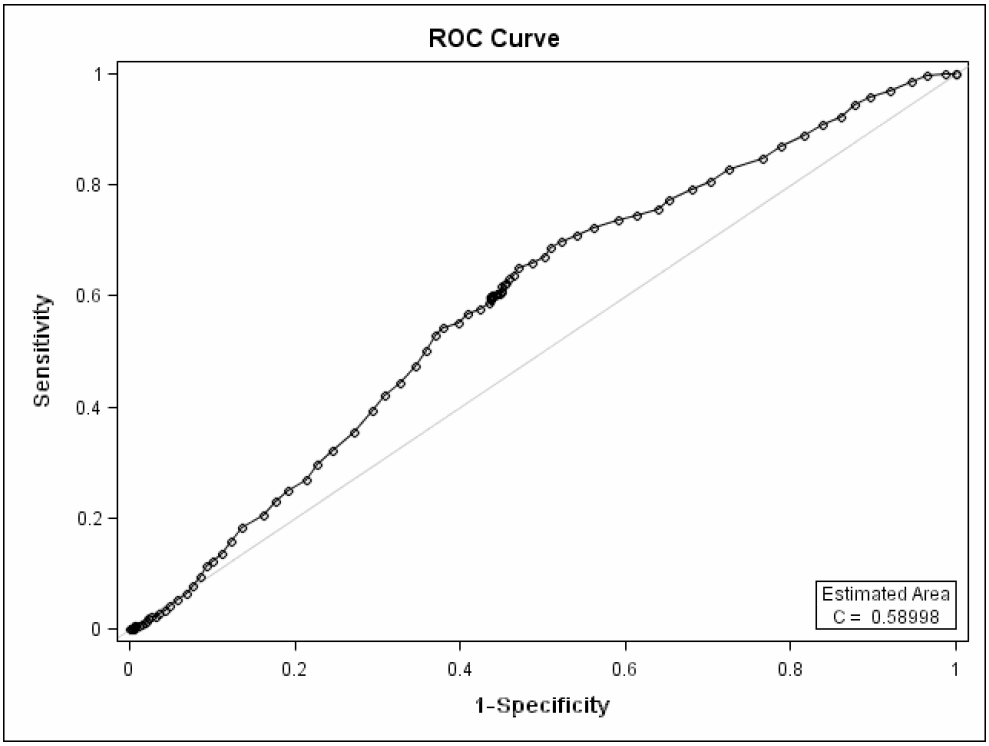

Figure 4.1: ROC curve with AUC for model including age and sex

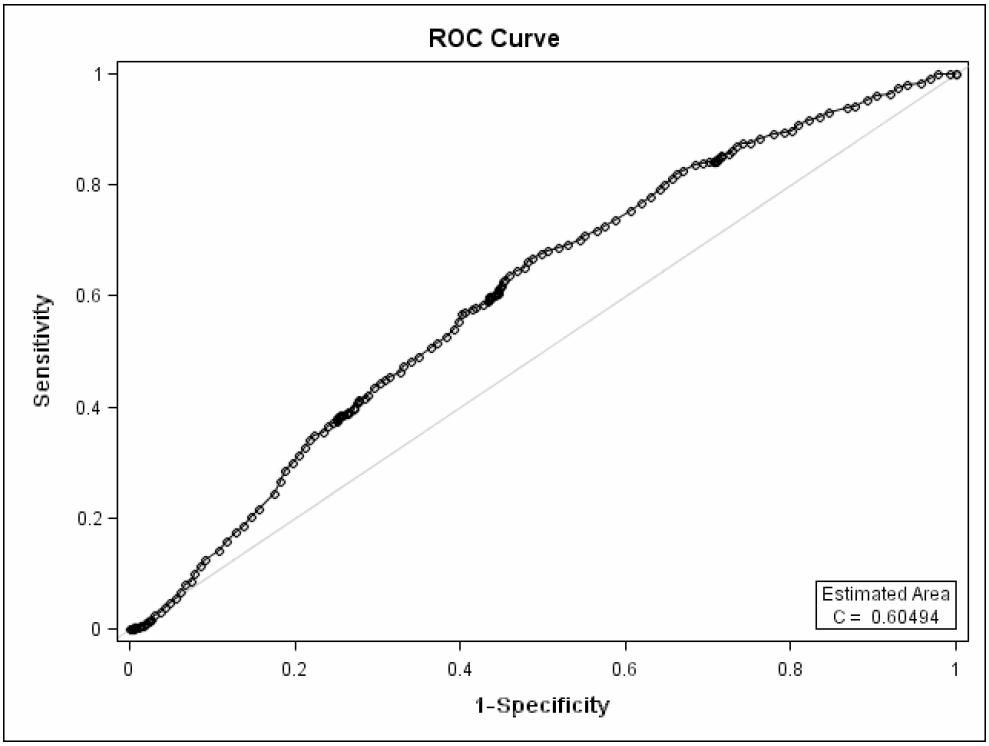

Figure 4.2: ROC curve with AUC for model including age, sex, and NSAID use

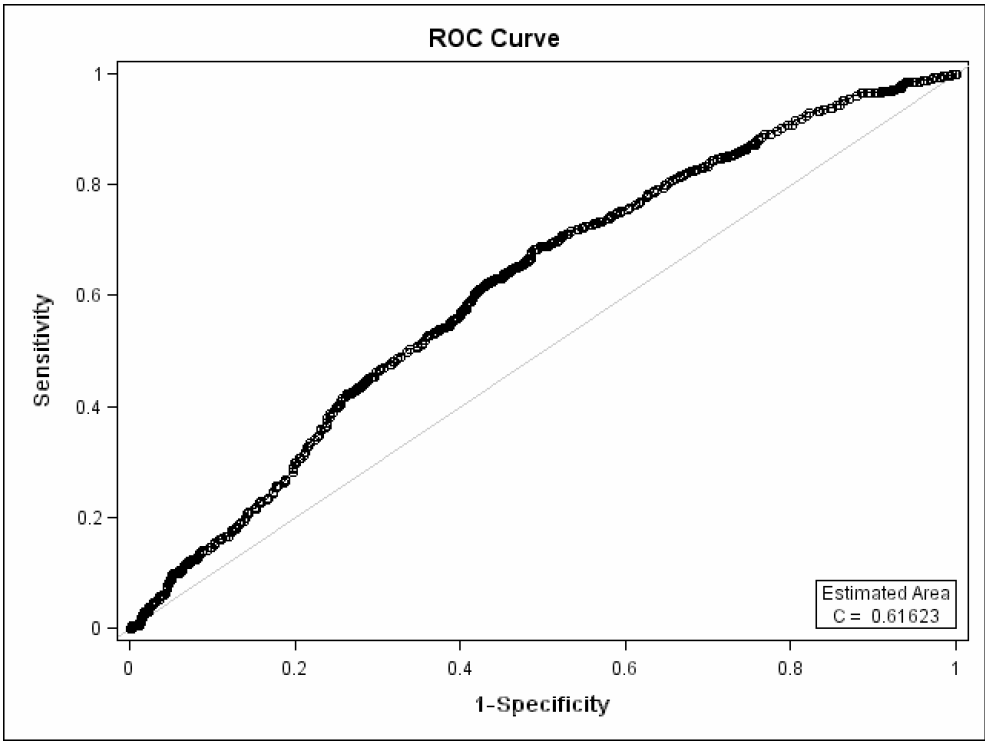

Figure 4.3: ROC curve with AUC for model including age, sex, NSAID use, 3 SNPs, and NSAIDxSNP interactions

Chapter 5

Conclusions

In this thesis a review of the basic biology of colorectal cancer and methods for investigating GxE interactions, a systematic review of the literature to identify candidate GxE interactions for further analysis, and a secondary analysis of a GWA case-control study were performed. The systematic review identified several potential interactions, but information on the specific genetic variants was not available in the GWA either directly or through imputation. Instead, a set of candidate SNPs were selected based on a separate systematic review and knowledge of NSAID pharmacokinetics. These candidates were screened for potential interactions, and several possible interactions were identified. The joint effects were similar for Aspirin and Ibuprofen, but not acetaminophen. This strengthens the case for these interactions being true biological effects, and not simply the result of chance. SNPs that had demonstrated a differential effect by genotype were investigated further using a stepwise model building procedure. This resulted in a model containing three SNPs, Aspirin or Ibuprofen use, their interactions, sex and age. This model was better able to discriminate cases and controls, demonstrated better calibration, and had a better AIC score than models without the interaction terms. In order to test the validity of this model, data used to validate the original ARCTIC GWA will be sought, and the results of the present study applied to these data.

5.1 Limitations

Several variants of interest that had been identified in the systematic review (Chapter 3) were not available in the ARCTIC dataset. The intent of the systematic review was to increase prior knowledge informing SNP selection, thereby increasing the probability that

genomic profiling based on these results would be true. That these desired SNPs (nor tag-SNPs in LD with the variants) were not available through ARCTIC is a limitation of the present analysis.

The choice of interaction scale (additive versus multiplicative) also impacts the interpretation of any observed interaction. Greenland and others [31] have argued that biological interaction between risk factors is best measured as the deviation from additivity, rather than deviation from a multiplicative model. The argument holds that while multiplicative interaction may be a valid component of model building, the identification of a multiplicative interaction does not imply a biological interaction. Based on these considerations, it was decided to investigate interactions on the additive scale for all analyses presented in this thesis, and strata-specific ORs were presented wherever possible.

5.2 Future directions

Recent randomized trials confirm that NSAIDs are effective in preventing colorectal neoplasia, but there remains discrepancy in the effect of various doses. In one randomized trial, Aspirin led to a statistically significant reduction of adenoma at 3 years only at the 81 mg daily dose, but not at the 325mg daily dose [8]. In another, the 325 mg dose resulted in a significant risk reduction [5]. The heterogeneity in response for the 325 mg dose is not well understood. Before adopting Aspirin as a chemopreventive agent, understanding the sources of such variability in efficacy is warranted.

A 2005 article [38] that explored GxE interactions for human disease laid out a number of requirements for their future study. First, coordination across studies is essential. This would be required to increase the power of analyses for common disease by studying larger cohorts, and extending the study over longer periods of time. A key component would be adding DNA collection and consent to existing prospective studies that do not currently have biobanked tissue or arrangements for ongoing use of the collected data. Second, new prospective studies should be initiated to replace the current set of ongoing investigations using new targets, new study designs, and improved measurements of environmental exposures. With the advantage of hindsight, these studies should endeavor to maximize compatibility of the genetic and environmental information collected. Third, mechanisms for making unpublished results available for study should be strongly encouraged. A large number of null findings are inevitable in the study of GxE interactions in the anticipated long-term studies, and access to

the unpublished data will help ensure that publication bias be minimized. Finally, the ongoing effort to develop better statistical approaches to integrate epidemiological and biological data in fields such as systems biology, biostatistics, and bioinformatics should be encouraged, as finding the optimal methods will be vital in producing robust results.

Two immediate extensions of the approach explored in this thesis are evident. First, expanding to target therapies or preventive advice that target multiple disease, most likely by targeting common pathways that influence multiple disease states. There is accumulating interest in prostate cancer prevention using NSAIDs, in addition to the widespread interest in NSAID prophylaxis for secondary prevention of myocardial infarction. An investigation of the pooled risk vs. benefit of widespread NSAID use for a set of potential disease targets may result in a radically different estimate of this balance than has been determined through the analysis of individual conditions. Second, such investigation should consider both sides of the risk-benefit leger; in the particular case of NSAIDs, it is the harms rather than the lack of a well-described benefit that limit its application as a preventive intervention. Future studies should examine the potential for genetic variation to alter harms associated with NSAID prophylaxis. The ARCTIC study did not examine measures of NSAID intolerance or the common side-effects of NSAID use (such as gastric ulcers and bleeding), but these may be as or more important as measures of preventive effectiveness in modifying the risk-benefit balance of NSAID prophylaxis. The same applies to the gastrointestinal toxicity of nonselective NSAIDs, which have been estimated to cause 25% of all reported drug-related adverse events.

The study of GxE interactions may be profitable along a number of other avenues. First, if could help to target therapies by identifying genes in pathways that increase the effectiveness of interventions, decrease the potential for harm, or both. Additionally, such studies may help determine the specific compounds in mixtures that prevent or cause disease. Knowledge of the GxE interactions could inform individuals on the most productive preventive activities given their unique set of susceptibility and resistance alleles, thereby empowering individuals to undertake the lifestyle modifications most likely to be of benefit, and avoid efforts that are unlikely to have an impact.

This prospect seems to conflict with the view expressed by Rose [67] that population-based interventions are more effective means of improving health and reducing disease incidence than targeted interventions aimed at high-risk individuals. However, while this view is likely reasonable for presumably universally-beneficial actions (e.g. physical exercise), or avoidance of major risk factors (e.g. tobacco smoke), it is not necessarily so for interventions with a heterogeneous result. In the current example of NSAID prophy-

laxis, it is possible that a sub-population would be expected to suffer serious harm from the intervention, while a separate subgroup could benefit substantially. Understanding these differences could thus contribute to a substantial improvement in the effectiveness of public health interventions with variable effects.

Future prospects for preventive public health action involve screening populations for large number of genetic variants. Using this information, which would be widely available, would be used to tailor advice on preventive practices, as well as secondary and tertiary medical interventions. This vision requires robust knowledge of the relevant GxE interactions; in fact, this may be the crucial component of the genomics effort. It is possible that future histories will consider the multi-billion dollar sequencing of the human genome as the elementary component of the field, with the ensuing development of a credible database of interactions as the major challenge.

Appendix A

Appendices

Appendix B

Glossary of Terms

Adenoma A benign neoplasm of glandular epithelium.

Adenocarcinoma Carcinoma arising from glandular epithelium.

Carcinoma Invasive tumor arising from an epithelium.

CRA Colorectal adenoma

CRC Colorectal carcinoma

Dysplasia The irregular appearance of cells reflecting altered growth control mechanisms.

GWA Genome-wide association

HWE Hardy-Weinberg Equilibrium; states that both allele and genotype frequencies in a population remain in equilibrium from generation to generation unless specific disturbing influences are introduced, such as non-random mating, mutation, selection, or limited population size.

Neoplasia Processes resulting in the formation of new, abnormal growths.

Neoplasm An abnormal mass which grows faster than surrounding tissue, is uncoordinated with normal tissues, and persists in its abnormal growth after the removal of the stimulus that provoked the change.

NSAID Nonsteroidal anti-inflammatory drug

Polymorphism a location in the human genome where at least two sequence variants exist with an appreciable frequency, often distinguished from mutation by the stipulation that it be present in >1% of a given population.

SNP Single-nucleotide polymorphism; polymorphism that differs from the reference sequence by a single base substitution.

Bibliography

[1] A. Ahlbom and L. Alfredsson. Interaction: A word with two meanings creates confusion. *European journal of epidemiology*, 20(7):563–564, 2005. id: 1; LR: 20071115; JID: 8508062; ppublish.

[2] N. Ahmadiyeh, M. M. Pomerantz, and C. Grisanzio. 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with myc. *Proceedings of the National Academy of Sciences of the United States of America*, 107(21):9742–9746, May 25 2010.

[3] I. U. Ali, B. T. Luke, M. Dean, and P. Greenwald. Allellic variants in regulatory regions of cyclooxygenase-2: association with advanced colorectal adenoma. *British journal of cancer*, 93(8):953–959, Oct 17 2005.

[4] N. Arber, C. J. Eagle, and J. Spicak. Celecoxib for the prevention of colorectal adenomatous polyps. *The New England journal of medicine*, 355(9):885–895, Aug 31 2006.

[5] J. A. Baron, B. F. Cole, and R. S. Sandler. A randomized trial of aspirin to prevent colorectal adenomas. *The New England journal of medicine*, 348(10):891–899, Mar 6 2003.

[6] J. A. Baron, R. S. Sandler, and R. S. Bresalier. A randomized trial of rofecoxib for the chemoprevention of colorectal adenomas. *Gastroenterology*, 131(6):1674–1682, Dec 2006.

[7] E. L. Barry, J. A. Baron, S. Bhat, M. V. Grau, C. A. Burke, R. S. Sandler, D. J. Ahnen, R. W. Haile, and T. G. O'Brien. Ornithine decarboxylase polymorphism modification of response to aspirin treatment for colorectal adenoma prevention. *Journal of the National Cancer Institute*, 98(20):1494–1500, Oct 18 2006.

[8] R. Benamouzig, J. Deyra, and A. Martin. Daily soluble aspirin and prevention of colorectal adenoma recurrence: one-year results of the apacc trial. *Gastroenterology*, 125(2):328–336, Aug 2003.

[9] M. M. Bertagnolli, C. J. Eagle, and A. G. Zauber. Celecoxib for the prevention of sporadic colorectal adenomas. *The New England journal of medicine*, 355(9):873–884, Aug 31 2006.

[10] J. Bigler, J. Whitton, and J. W. Lampe. Cyp2c9 and ugt1a6 genotypes modulate the protective effect of aspirin on colon adenoma risk. *Cancer research*, 61(9):3566–3569, 2001.

[11] J. Bigler, J. Whitton, and J. W. Lampe. Cyp2c9 and ugt1a6 genotypes modulate the protective effect of aspirin on colon adenoma risk. *Cancer research*, 61(9):3566–3569, May 1 2001.

[12] S. R. Browning and B. L. Browning. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics*, 81(5):1084–1097, Nov 2007.

[13] M. J. Campbell, S. J. Hawken, and A. Demarsh. *Colorectal cancer*. Human Genome Epidemiology: Building the evidence for using genetic information to improve health and prevent disease. Oxford University Press, New York, 2 edition, 2010.

[14] S. C. Canadian Cancer Society, Public Health Agency of Canada. *Canadian Cancer Statistics*. 2010.

[15] J. M. carethers and C. R. Boland. *Neoplasia of the gastrointestinal tract*, page 557. Textbook of Gastroenterology. Lippincott, Williams & Wilkins, Philadelphia, 4 edition, 2003.

[16] A. T. Chan, G. J. Tranah, and E. L. Giovannucci. Genetic variants in the ugt1a6 enzyme, aspirin use, and the risk of colorectal adenoma. *Journal of the National Cancer Institute*, 97(6):457–460, Mar 16 2005.

[17] H. Clevers. Colon cancer - understanding how nsaids work. *New England Journal of Medicine*, 354(7):761–763, 2006.

[18] M. Cotterchio, G. McKeown-Eyssen, and H. Sutherland. Ontario familial colon cancer registry: Methods and first-year response rates. *Chronic diseases in Canada*, 21(2):81–86, 2000.

[19] S. Cotton, L. Sharp, and J. Little. The adenoma-carcinoma sequence and prospects for the prevention of colorectal neoplasia. *Critical reviews in oncogenesis*, 7(5-6):293–342, 1996.

[20] D. G. Cox, C. Pontes, and E. Guino. Polymorphisms in prostaglandin synthase 2/cyclooxygenase 2 (ptgs2/cox2) and risk of colorectal cancer. *British journal of cancer*, 91(2):339–343, Jul 19 2004.

[21] S. Dubey and C. A. Powell. Update in lung cancer 2007. *American journal of respiratory and critical care medicine*, 177(9):941–946, 2008.

[22] E. R. Fearon and B. Vogelstein. A genetic model for colorectal tumorigenesis. *Cell*, 61(5):759–767, 1990.

[23] W. D. Flanders, C. A. Lally, B. P. Zhu, S. J. Henley, and M. J. Thun. Lung cancer mortality in relation to age, duration of smoking, and daily cigarette consumption: results from cancer prevention study ii. *Cancer research*, 63(19):6556–6562, Oct 1 2003.

[24] E. Flossmann, P. M. Rothwell, B. D. A. Trial, and the UK-TIA Aspirin Trial. Effect of aspirin on long-term risk of colorectal cancer: consistent evidence from randomised and observational studies. *Lancet*, 369(9573):1603–1613, May 12 2007.

[25] M. Frattini, D. Balestra, and S. Suardi. Different genetic features associated with colon and rectal carcinogenesis. *Clinical Cancer Research*, 10(12 I):4015–4021, 2004.

[26] P. H. Gann, J. E. Manson, and R. J. Glynn. Low-dose aspirin and incidence of colorectal tumors in a randomized trial. *Journal of the National Cancer Institute*, 85(15):1220–1224, Aug 4 1993.

[27] D. M. Gertig and D. J. Hunter. Genes and environment in the etiology of colorectal cancer. *Seminars in cancer biology*, 8(4):285–298, Aug 1998.

[28] F. M. Giardiello, V. W. Yang, and L. M. Hylind. Primary chemoprevention of familial adenomatous polyposis with sulindac. *The New England journal of medicine*, 346(14):1054–1059, Apr 4 2002.

[29] G. A. Green. Understanding nsaids: from aspirin to cox-2. *Clinical cornerstone*, 3(5):50–60, 2001.

[30] E. R. Greenberg, J. A. Baron, D. H. F. Jr, J. S. Mandel, and R. Haile. Reduced risk of large-bowel adenomas among aspirin users. the polyp prevention study group. *Journal of the National Cancer Institute*, 85(11):912–916, Jun 2 1993.

[31] S. Greenland. Interactions in epidemiology: relevance, identification, and estimation. *Epidemiology (Cambridge, Mass.)*, 20(1):14–17, Jan 2009.

[32] J. A. Hanley and B. J. McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36, 1982.

[33] S. J. Hawken, C. M. Greenwood, T. J. Hudson, R. Kustra, J. McLaughlin, Q. Yang, B. W. Zanke, and J. Little. The utility and predictive value of combinations of low penetrance genes for screening and risk prediction of colorectal cancer. *Human genetics*, 128(1):89–101, Jul 2010.

[34] P. Hewitson, P. Glasziou, L. Irwig, B. Towler, and E. Watson. Screening for colorectal cancer using the faecal occult blood test, hemoccult. *Cochrane database of systematic reviews (Online)*, (1)(1):CD001216, Jan 24 2007.

[35] T. Higuchi, T. Iwama, K. Yoshinaga, M. Toyooka, M. M. Taketo, and K. Sugihara. A randomized, double-blind, placebo-controlled trial of the effects of rofecoxib, a selective cyclooxygenase-2 inhibitor, on rectal polyps in familial adenomatous polyposis patients. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 9(13):4756–4760, Oct 15 2003.

[36] I. M. Hisamuddin and V. W. Yang. Molecular genetics of colorectal cancer: An overview. *Current colorectal cancer reports*, 2(2):53–59, Apr 2006.

[37] R. A. Hubner, K. R. Muir, J. F. Liu, R. F. Logan, M. Grainge, N. Armitage, V. Shepherd, S. Popat, R. S. Houlston, and U. K. C. A. P. Consortium. Genetic variants of ugt1a6 influence risk of colorectal adenoma recurrence. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 12(21):6585–6589, Nov 1 2006.

[38] D. J. Hunter. Gene-environment interactions in human diseases. *Nature reviews.Genetics*, 6(4):287–298, Apr 2005.

[39] W. G. IARC. *IARC Handbooks on Cancer Prevention: Non-Steroidal Anti-Inflammatory Drugs*, volume 1.

[40] J. P. A. Ioannidis. Why most published research findings are false. *PLoS Medicine*, 2(8):0696–0701, 2005.

[41] J. P. A. Ioannidis, T. A. Trikalinos, and M. J. Khoury. Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. *American Journal of Epidemiology*, 164(7):609–614, 2006.

[42] J. R. Jass. Molecular heterogeneity of colorectal cancer: Implications for cancer control. *Surgical oncology*, 16 Suppl 1:S7–9, Dec 2007.

[43] F. E. H. Jr, R. M. Califf, and D. B. Pryor. Evaluating the yield of medical tests. *JAMA : the journal of the American Medical Association*, 247(18):2543–2546, May 14 1982.

[44] J. S. Kaufman. Interaction reaction. *Epidemiology (Cambridge, Mass.)*, 20(2):159–160, Mar 2009.

[45] M. J. Khoury, R. Davis, and M. Gwinn. Do we need genomic research for the prevention of common diseases with environmental causes? *American Journal of Epidemiology*, 161(9):799–805, 2005.

[46] M. J. Khoury and M. Gwinn. What role for public health in genetics and vice versa? [1]. *Community Genetics*, 9(4):282, 2006.

[47] M. J. Knol, M. Egger, and P. Scott. When one depends on the other: reporting of interaction in case-control and cohort studies. *Epidemiology (Cambridge, Mass.)*, 20(2):161–166, Mar 2009.

[48] O. Kronborg and C. Fenger. Clinical evidence for the adenoma-carcinoma sequence. *European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP)*, 8 Suppl 1:S73–86, Dec 1999. id: 1; LR: 20051116; JID: 9300837; RF: 155; ppublish.

[49] T. Kudo, T. Narisawa, and S. Abo. Antitumor activity of indomethacin on methylazoxymethanol-induced large bowel tumors in rats. *Gann = Gan*, 71(2):260–264, Apr 1980.

[50] G. A. Kune, S. Kune, and L. F. Watson. Colorectal cancer risk, chronic illnesses, operations, and medications: case control results from the melbourne colorectal cancer study. *Cancer research*, 48(15):4399–4404, Aug 1 1988.

[51] D. Labayle, D. Fischer, P. Vielh, F. Drouhin, A. Pariente, C. Bories, O. Duhamel, M. Trousset, and P. Attali. Sulindac causes regression of rectal polyps in familial adenomatous polyposis. *Gastroenterology*, 101(3):635–639, Sep 1991.

[52] E. S. Lander, L. M. Linton, and B. Birren. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, 2001.

[53] H. J. Lin, K. M. Lakkides, and T. O. Keku. Prostaglandin h synthase 2 variant (val511ala) in african americans may reduce the risk for colorectal neoplasia. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 11(11):1305–1315, Nov 2002.

[54] J. Little, M. J. Khoury, and L. Bradley. The human genome project is complete. how do we develop a handle for the pump? *American Journal of Epidemiology*, 157(8):667–673, 2003.

[55] J. Little and L. Sharp. *Epidemiology of colorectal cancer*, chapter 2, pages 43–75. Colorectal Cancer. Informa Healthcare, New York, 2007.

[56] M. Lundberg, P. Fredlund, and J. Hallqvist. A sas program calculating three measures of interaction with confidence intervals. *Epidemiology (Cambridge, Mass.)*, 7(6):655–656, Nov 1996.

[57] M. Macarthur, L. Sharp, and G. L. Hold. The role of pro- and anti-inflammatory cytokine gene polymorphisms in colorectal cancer in the north east of scotland. *Gut*, 53 (Suppl 3):A42, 2004.

[58] M. E. Martinez, T. G. O'Brien, and K. E. Fultz. Pronounced reduction in adenoma recurrence associated with aspirin use and a polymorphism in the ornithine decarboxylase gene. *Proceedings of the National Academy of Sciences of the United States of America*, 100(13):7859–7864, Jun 24 2003.

[59] A. D. Michelson, M. I. Furman, and P. Goldschmidt-Clermont. Platelet gp iiia pl(a) polymorphisms display different sensitivities to agonists. *Circulation*, 101(9):1013–1018, 2000.

[60] T. Mizushima. Molecular mechanism for various pharmacological activities of nsaids. *Pharmaceuticals*, 3(5):1614–1636, 2010.

[61] K. L. Moore and A. F. Dalley. *Clinically oriented anatomy*. Lippincott Williams & Wilkins, Philadelphia, 1999.

[62] T. Narisawa, M. Sato, M. Tani, T. Kudo, T. Takahashi, and A. Goto. Inhibition of development of methylnitrosourea-induced rat colon tumors by indomethacin treatment. *Cancer research*, 41(5):1954–1957, May 1981.

[63] D. M. Parkin, P. Pisani, and J. Ferlay. Global cancer statistics. *Ca-A Cancer Journal for Clinicians*, 49(1):33–64, 1999.

[64] M. Pollard and P. H. Luckert. Indomethacin treatment of rats with dimethylhydrazine-induced intestinal tumors. *Cancer treatment reports*, 64(12):1323–1327, 1980.

[65] U. S. PSFT. Routine aspirin or nonsteroidal anti-inflammatory drugs for the primary prevention of colorectal cancer: recommendation statement. *American Family Physician*, 76(1):109–113, Jul 1 2007.

[66] L. Rabeneck and L. F. Paszat. A population-based estimate of the extent of colorectal cancer screening in ontario. *American Journal of Gastroenterology*, 99(6):1141–1144, 2004.

[67] G. Rose. Sick individuals and sick populations. *Bulletin of the World Health Organization*, 79(10):990–996, 2001.

[68] A. Rostom, C. Dube, and G. Lewin. Nonsteroidal anti-inflammatory drugs and cyclooxygenase-2 inhibitors for primary prevention of colorectal cancer: a systematic review prepared for the u.s. preventive services task force. *Annals of internal medicine*, 146(5):376–389, 2007.

[69] K. J. Rothman, S. Greenland, and T. L. Lash. *Introduction to regression modeling*. Modern Epidemiology. 3 edition, 2008.

[70] W. S. Samowitz, R. K. Wolff, and K. Curtin. Interactions between cyp2c9 and ugt1a6 polymorphisms and nonsteroidal anti-inflammatory drugs in colorectal cancer prevention. *Clinical Gastroenterology and Hepatology*, 4(7):894–901, Jul 2006.

[71] L. B. Sansbury, A. W. Bergen, and K. L. Wanke. Inflammatory cytokine gene polymorphisms, nonsteroidal anti-inflammatory drug use, and risk of adenoma polyp recurrence in the polyp prevention trial. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 15(3):494–501, Mar 2006.

[72] O. Shpilberg, J. S. Dorman, and R. E. Ferrell. The next stage: molecular epidemiology. *Journal of clinical epidemiology*, 50(6):633–638, Jun 1997.

[73] C. L. Siezen, M. J. Tijhuis, and N. R. Kram. Protective effect of nonsteroidal anti-inflammatory drugs on colorectal adenomas is modified by a polymorphism in peroxisome proliferator-activated receptor delta. *Pharmacogenetics and genomics*, 16(1):43–50, Jan 2006.

[74] M. L. Slattery, W. Samowitz, M. Hoffman, K. N. Ma, T. R. Levin, and S. Neuhausen. Aspirin, nsaids, and colorectal cancer: possible involvement in an insulin-related pathway. *Cancer Epidemiology, Biomarkers and Prevention*, 13(4):538–545, 2004. ID: 13087.

[75] K. Smigel. Arthritis drug approved for polyp prevention blazes trail for other prevention trials. *Journal of the National Cancer Institute*, 92(4):297–299, 2000.

[76] G. Smith, F. A. Carey, and J. Beattie. Mutations in apc, kirsten-ras, and p53 - alternative genetic pathways to colorectal cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 99(14):9433–9438, 2002.

[77] G. Steinbach, P. M. Lynch, and R. K. Phillips. The effect of celecoxib, a cyclooxygenase-2 inhibitor, in familial adenomatous polyposis. *The New England journal of medicine*, 342(26):1946–1952, Jun 29 2000.

[78] O. Suh, C. Mettlin, and N. J. Petrelli. Aspirin use, cancer, and polyps of the large bowel. *Cancer*, 72(4):1171–1177, Aug 15 1993.

[79] D. Thomas. Methods for investigating gene-environment interactions in candidate pathway and genome-wide association studies. *Annual Review of Public Health*, 31:21–36, Apr 21 2010.

[80] M. J. Thun, E. E. Calle, and M. M. Namboodiri. Risk factors for fatal colon cancer in a large prospective study. *Journal of the National Cancer Institute*, 84(19):1491–1500, Oct 7 1992.

[81] O. Tribut, Y. Lessard, and J. M. Pharmacogenomics. *Medical Science Monitor*, 8(7):RA152–RA163, 2002.

[82] C. Ulrich, J. Bigler, and R. Sparks. Polymorphisms in ptgs1 (=cox-1) and risk of colorectal polyps. *Cancer Epidemiology Biomarkers & Prevention*, 13(5):889–893, 2004.

[83] C. M. Ulrich, J. Bigler, and J. D. Potter. Non-steroidal anti-inflammatory drugs for cancer prevention: Promise, perils and pharmacogenetics. *Nature Reviews Cancer*, 6(2):130–140, 2006.

[84] C. M. Ulrich, J. Whitton, and J. H. Yu. Ptgs2 (cox-2) -765g ¿ c promoter variant reduces risk of colorectal adenoma among nonusers of nonsteroidal anti-inflammatory drugs. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 14(3):616–619, Mar 2005.

[85] J. C. Venter, M. D. Adams, and E. W. Myers. The sequence of the human genome. *Science*, 291(5507):1304–1351, 2001.

[86] S. J. Winawer, A. G. Zauber, and M. N. Ho. Prevention of colorectal cancer by colonoscopic polypectomy. the national polyp study workgroup. *The New England journal of medicine*, 329(27):1977–1981, Dec 30 1993.

[87] B. W. Zanke, C. M. Greenwood, and J. Rangrej. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nature genetics*, 39(8):989–994, Aug 2007.