The use of aspirin for primary and secondary prevention of colorectal cancer

Aayah Ghassan Nounu

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Life Sciences

School of Cellular and Molecular Medicine

September 2018

Word Count: 75,837

Abstract

Colorectal cancer (CRC) is the third most common cancer worldwide. Observational studies and randomised controlled trials for vascular events have shown aspirin as a possible chemopreventative agent in CRC. Aspirin is a well-established inhibitor of the (COX)/(PGE2) signalling pathway. However, an aspirin derivative (NCX-4016) that does not inhibit COX activity was reported to have a better chemopreventative effect in a mouse model of colon cancer implicating other unknown aspirin anti-cancer effects. More recently aspirin has been reported to induce post-transcriptional modifications, specifically, protein and histone acetylation.

The aim of this PhD thesis was to investigate the effect of short-term and long-term aspirin exposure in colorectal adenoma cells (RG/C2) by looking at the regulation of cell growth as well as changes in DNA methylation, gene and protein expression. Mendelian randomisation (MR) was used to assess whether variations in aspirin metabolite levels are associated with risk of CRC.

Through combining multiple 'omics, short-term aspirin was not found to regulate genome-wide DNA methylation but was implicated in regulating specific gene expression e.g. suppressing *STMN1*. This led to a reduction in cell migration, which may partly explain the anti-metastatic function of aspirin. Conversely, cells exposed long-term to aspirin (≥18 weeks) were less responsive to aspirin's growth inhibition. Long-term treatment resulted in a genome-wide effect on DNA methylation, highlighting a previously unreported role for aspirin in epigenetic regulation. Finally, the MR approach showed a possible association between aspirin metabolism and cancer risk, complementing other epidemiological evidence with regards to aspirin use and cancer risk.

For the first time, combining population-based and laboratory-based methods have identified possible new mechanisms of aspirin's action. Genetic variants may help inform the effect of aspirin use on CRC risk and epigenetic modifications from long-term use could help identify patients who become less responsive to the drug.

.

Acknowledgements

Firstly, I would like to thank both of my supervisors, Professor Ann Williams and Professor Caroline Relton. I would like to thank Ann specifically for always asking me questions that have allowed me to truly understand my work, for always being supportive and inspiring me to think of new directions to take in research. I would also like to thank Caroline for her guidance, without which this research would not have been possible.

I would also like to thank the Medical Research Council for funding my PhD. Thank you to all the kind collaborators- most specifically the GECCO consortium (in particular, Professor Ulrike Peters and Dr. Yi Lin) as well as the Twins UK registry (in particular, Miss Genevieve Lachance and Victoria Vazquez) for allowing me to use their data in order to carry out my analyses. I also acknowledge that this work would not have been possible without the data from the UK Biobank consortium. Furthermore, I recognise that the analysis of the 3D cell growth in this thesis would not have been possible without the help and microscopes from the Wolfson Bioimaging Facility.

I am grateful for the help and guidance from many individuals, specifically: Dr. Gemma Sharp, Dr. Kaitlin Wade, Dr. Philip Haycock, Dr. Ed Mountjoy, Dr. Louise Millard, Dr. Anna Guyatt, Dr. Alex Greenhough and Tracey Collard. I would also like to thank everyone from both offices who have always made me feel welcome, have provided intellectual conversations and distractions when needed.

Finally, I would like to thank all my family members for always encouraging me and allowing me to complete this PhD. I am grateful for having wonderful extended family members (my grandparents, aunts, uncles and cousins) for their limitless love and pride. I specifically would like to thank my sister who listened to my endless rants when things were tough, my parents who were always patient and helped whenever they could and my nieces and nephews for their constant joy who showed me the care-free and innocent side of life.

Author declarations

Comments on the entire thesis were received from my supervisors, Professor Ann Williams and Professor Caroline Relton.

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE:.....

Table of contents

List of tables

List of figures

List of abbreviations

27K array	Infinium HumanMethylation27 BeadChip
3'UTR	3' Untranslated Region
450K array	Illumina Infinium® HumanMethylation450K BeadChip
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
5'UTR	5' Untranslated Region
850K array	MethylationEPIC BeadChip (850K) Infinium
ACSM2B	Acyl-CoA Synthetase Medium-Chain Family Member 2B
ADF	Advanced DMEM/F12 with reduced FBS supplementation
ADP	Adenosine Diphosphate
ADP	Adenosine Diphosphate
AERD	Aspirin-Exacerbated Respiratory Disease
AFPPS	Aspirin/Folate Polyp Prevention Study
AIU	Aspirin-Induced Urticaria
ALS	Amyotrophic Lateral Sclerosis
ANOVA	Analysis of Variance
APACC	Association pour la Prévention par l'Aspirine du Cancer Colorectal
APC	Adenomatous Polyposis Coli
ARRIVE	Aspirin to Reduce Risk of Initial Vascular Events
ASCOLT	Aspirin for Dukes C and High Risk Dukes B Colorectal Cancers
ASPIRED	ASPirin Intervention for the REDuction of colorectal cancer risk
ASPREE	ASPirin in Reducing Events in the Elderly
BCHE	Butyrylcholinesterase
BMI	Body Mass Index
CALGB	Colorectal Adenoma prevention study originated in the cooperative trials group cancer and Leukaemia Group B
CBC cells	Crypt Base Columnar Cells
C-CFR	Colon Cancer Family Registry
CDH1	E-cadherin 1
CDK1	Cyclin Dependent Kinase 1
cGMP-PKG	cyclic Guanosine Monophosphate- Protein Kinase G
CHD	Coronary Heart Disease
CI	Confidence Intervals
CIMP	CpG Island Methylation Pathway
CIN	Chromosomal Instability
COX	Cyclooxygenase
CRC	Colorectal Cancer
CSC	Cancer Stem Cell
CYP2C9	Cytochrome P450
DAG	Directed Acyclic Graph
DAP-kinase	Death-Associated Protein Kinase
DCS cells	Deep Crypt Secretory cells
DMEM	Dulbecco's Modified Eagles Medium
DMSO	Dimethyl Sulfoxide

DNMT	DNA methyltransferases
DRAQ5	1,5-bis{[2-(di-methylamino) ethyl]amino}-4, 8-dihydroxyanthracene-9,10-dione
ECM	Extra-Cellular Matrix
EDTA	Ethylenediaminetetraacetic Acid
EGFR	Epidermal Growth Factor Receptor
EHS	Engelbreth-Holm-Swarm
EMT	Epithelial to Mesenchymal Transition
EP300	E1A-associated Protein p300
EWAS	Epigenome-Wide Association Studies
FACS	Fluorescence-activated Cell Sorting
FBS	Foetal Bovine Serum
FCS	Fetal Calf Serum
FDR	False Discovery Rate
FOB	Faecal Occult Blood
G6PD	Glucose-6-Phosphate Dehydrogenase
GC/MS	Gas Chromatography-Mass Spectrometry
GECCO	Genetics and Epidemiology of Colorectal Cancer Consortium
GWAS	Genome-Wide Association Studies
H3K27	Histone 3 lysine 27
H3K27me3	Histone 3 lysine 27 trimethylation
H3K4me3	Histone 3 lysine 4 trimethylation
H3K9me3	Histone 3 lysine 9 trimethylation
HAT	Histone acetyltransferase
HDAC	Histone deacetyltransferase
HDM	Histone demethylase
hm^5C	5- hydroxymethylcytosine
HMT	Histone methyltransferase
HUVECs	Human Umbilical Vein Endothelial Cells
I	Inosine
IGF	Insulin-like Growth Factor
IKK	IκB kinase
IL-17A	Interleukin-17A
IL-22	Interleukin-22
IL-25	Interleukin-25
IL-6	Interleukin-6
iNOS	inducible Nitric Oxide Synthase
InSIDE	The Instrument Strength Independent of Direct Effect
IVW	Inverse Variance Weighted
J-CAPP	Japan Colorectal Aspirin Polyps Prevention
KAT	Lysine Acetyl Transferase
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LC/MS	Liquid Chromatography-Mass Spectrometry
LD	Linkage Disequilibrium
LGR5	Leucine-rich repeat-containing G-protein-coupled receptor 5
LIMS	Laboratory Information Management System
LOXHD1	Lipoxygenase Homology Domains 1
M cells	Microfold cells
m^5C	5-methylcytosine

m^6A	N^6-methyladenosine
MAPK	Mitogen-Activated Protein Kinase
Mb	Megabase
MCM2	Minichromosome Maintenance Complex Component 2
MCM5	Minichromosome Maintenance Complex Component 5
MMR	Mismatch Repair
MR	Mendelian Randomisation
MSI	Microsatellite Instability
NANOG	Homeobox Protein NANOG
NFκB	Nuclear Factor kappa-light-chain-enhancer of activated B cells
NHS	National Health Service
NSAID	Non-Steroidal Anti-Inflammatory Drug
NSG	Immunodeficient NOD-scid IL2Rγ^{null}
NSMCE1	Non-Structural Maintenance of Chromosomes Element 1
P/S	Penicillin and Streptomycin
PAFAH1b2	Platelet- Activating Factor Acetylhydrolase
PBS	Phosphate Buffered Saline
PCNA	Proliferating Cell Nuclear Antigen
PGE_2	Prostaglandin E2
PGE-M	urinary metabolite of prostaglandin E(2)
PI3K-AKT	Phosphoinositide 3-kinase- protein kinase B
POLD2	DNA Polymerase Delta Subunit 2
PP2A	Protein Phosphatase 2 A
PTM	Post-translational modifications
PVDF	Polyvinylidine Difluoride
QC	Quality Control
qPCR	quantitative Polymerase Chain Reaction
QQ	Quantile-Quantile
QVD	N-(2-Quinolyl)valyl-aspartyl-(2,6-difluorophenoxy)methyl ketone
RCT	Randomised Controlled Trial
RR	Relative Risk
SD	Standard Deviation
SDS-PAGE	Sodium Dodecyl Polyacrylamide Gel Electrophoresis
SE	Standard Error
seAFOod	Systematic Evaluation of Aspirin and Fish Oil
SELECT	Selenium and Vitamin E Cancer Prevention Trial
SILAC	Stable Isotope Labelling of Amino Acids
SLC8A3	Solute Carrier Family 8 Member A3
SNP	Single Nucleotide Polymorphism
STAT3	Signal Transducer and Activator of Transcription 3
STMN1	Stathmin-1
SVA	Surrogate Variable Analysis
TDG	Thymine DNA Glycosylation
TET	Ten-eleven translocation
TF	Transcription Factor
TFB2M	Dimethyladenosine Transferase 2
TGF-β	Transforming Growth Factor Beta
TK1	Thymidine Kinase 1

TLR4	Toll-Like Receptor 4
TNF	Tumour Necrosis Factor
TNF-α	Tumour Necrosis Factor α
TWIST	TWIST family bHLH transcription factor 1
TXA_2	Thromboxane A_2
UGT1A6	UDP-Glucuronosyltransferase 1-6
ukCAP	United Kingdom Colorectal Adenoma Prevention
VASP	Vasodilator-Stimulated Phosphoprotein
vWF	von Willebrand factor
WHO	World Health Organization
X^2	Chi square
ZEB	Zinc finger E-box-binding homeobox 1
β	Beta
λ	Genomic Inflation Factor
Ψ	Pseudouridine

Chapter 1 Introduction

1.1 Thesis overview

1.1.1 Overview

In this thesis, I have applied laboratory and epidemiological methods to interrogate the effect of aspirin on colorectal adenoma cells.

The introduction chapter (Chapter 1) provides an overview of the project as well as the aims and objectives. It describes the molecular biology behind colorectal cancer as well its descriptive statistics. The chapter then introduces aspirin and its current usage before discussing the epidemiological evidence that shows it may be a chemopreventative agent. The chapter follows on from this by describing what is already known about aspirin's mechanism of action. A summary of epigenetics such as DNA methylation and histone modifications follows on from this. Finally, the chapter describes a causal analysis method called Mendelian randomisation which may help to interrogate the effect of aspirin on risk of colorectal cancer.

Chapter 2 provides details of both the laboratory-based methods as well as the epidemiological methods applied in this thesis, although details of the methods are present in their relevant chapters. Chapter 3 investigates the effect of short-term and long-term aspirin treatment on cell growth using 2D and 3D cell models. Chapter 4 describes an approach combining both laboratory-based methods and epidemiological-based methods to identify new targets of aspirin. Chapter 5 assesses the effect of long-term aspirin treatment on DNA methylation of adenoma cells. Chapter 6 investigates the effect of aspirin on risk of colorectal cancer through the use of a Mendelian randomisation approach. Finally, chapter 7 attempts to combine the overall findings from the thesis and the implications and further analyses that may follow-on from these results.

1.1.2 Importance of an inter-disciplinary approach

Epidemiological approaches are powerful in that they have the ability to interrogate multiple questions in a short space of time. On the other hand, laboratory-based methods traditionally take longer for a discovery to be made and can be biased due to the investigator's previous knowledge/opinions. Through combining both approaches, it is possible that epidemiology can be used as a non-biased approach to identify possible genes affected by aspirin use that also alter the risk of colorectal cancer and then validate these results through functional analyses within the laboratory. The combination allows for the identification of new targets not otherwise considered and may facilitate a faster turn around of identifying novel and significant results.

The opposite is also true in that epidemiological approaches may help validate laboratory-based methods. Many a time, the translation of findings from the laboratory into the clinic is unsuccessful (1). It may be that epidemiological approaches can be used to test laboratory findings before these results are tested in the clinic to save both time and money. In this project, we used epidemiological and basic scientific approaches together to identify new targets of aspirin as well as to identify whether long-term aspirin significantly alters DNA methylation in colorectal adenoma cells. These results may help in directing towards generating drugs for new targets of aspirin as well as identifying signature methylation profiles related to long-term aspirin treatment.

1.1.3 Aims and objectives

This thesis aims to address the effect of short-term and long-term aspirin treatment on colorectal adenoma cells. Cell culture was used to investigate the effect of aspirin on cell growth within a laboratory setting exploring cell survival, cell death and other assays. Causal analysis methods, namely Mendelian randomisation, was used to assess whether variations in aspirin metabolites (salicylic acid and salicyluric acid) affected risk of colorectal cancer. More specifically, this thesis aims to address these questions:

1- Does the length of duration of aspirin treatment affect cell survival in *in vitro* settings (Chapter 3)?
2- Are there new targets of aspirin that have not yet been identified in the literature that may explain part of the protective effect of the drug (Chapter 4)?
3- Does the length of duration of aspirin treatment affect the epigenetics, more specifically DNA methylation, of colorectal adenoma cells (Chapter 5)?
4- Can we address the effect of aspirin treatment on risk of colorectal cancer through a Mendelian randomisation approach (Chapter 6)?

I hypothesise that there are new targets of aspirin that may partially explain its chemopreventative properties. I also think that long-term aspirin treated cells become less sensitive to the drug, and this may be through aspirin's effect on epigenetics. Furthermore, I believe that with suitable genetic instruments, I will be able to predict how aspirin metabolites may affect the risk of colorectal cancer.

1.1.4 Publications and presentations

Published

The Telomeres Mendelian Randomization Collaboration. Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases. A Mendelian Randomization Study. JAMA Oncol. 2017;3(5):636-651. doi:10.1001/jamaoncol.2016.5945

In preparation, related to this thesis

Dawe, K., **Nounu, A**., Elsworth, B., Carr, H., Holly, J., Perks, C., Martin, R., Lewis, S. J., Gaunt, T., Williams, A. C., Higgins, J., Vincent, E. E.. Generating mechanistic hypotheses for the association of aspirin with colorectal cancer using text mining.

Nounu, A., Sharp, G. C., Greenhough, A., Bristol Bioresource Laboratories, Relton, C. L., Williams, A. C. Combining 'omics and epidemiology to uncover mechanisms linking aspirin treatment to reduced risk of colorectal cancer. See Chapter 4.

Nounu, A., Falk, L., Sharp, G. C., Bristol Bioresource Laboratories, Williams, A. C., Relton, C. L.. The effect of long-term aspirin treatment on DNA methylation in colorectal cancer. See Chapter 5.

Nounu, A., GECCO consortium, TwinsUK, UKBiobank , Williams, A. C., Relton, C. L. Salicylic acid and risk of colorectal cancer- a Mendelian randomisation approach. See Chapter 6.

Other publications in progress

Kazmi, N., Wade, K. H., **Nounu, A.**, Langdon, R.J., Tan, V. Y., Yarmolinsky, J., Zheng, J., Hemani, G., Torres, R., Johansson, M., Gaunt, T., Hung, R., Amos, C., Lewis S. J., Martin, R. M., Brennan, P., Relton, C. L., Davey Smith, G., Haycock, P. Identifying smoking-independent pathways for risk of lung cancer: a Mendelian randomization study.

Presentations

NCRI conference 2017, Liverpool. Title: Combining 'omics and epidemiology to uncover mechanisms linking aspirin treatment to reduced risk of colorectal cancer; oral presentation.

26th World Cancer Conference 2017, Dubai: Combining 'omics and epidemiology to uncover mechanisms linking aspirin treatment to reduced risk of colorectal cancer; poster presentation

1.1.5 Contributions

All experiments/analysis in this thesis were carried out by myself, except for these datasets which were generated by others and used in my analyses: SILAC proteomic data of aspirin treated RG/C2s (generated by Alexander Greenhough- a member of our laboratory group), methylation data of aspirin treated RG/C2s (the DNA samples were prepared by myself but the samples were run on the methylation arrays by the Bristol Bioresource Laboratory), transcriptomic data was obtained from

dbGAP (published by Sabates-Bellver et al 2007), logistic regression summary data for CRC risk (obtained from GECCO(2)), measured aspirin metabolite and genotype data (provided by Twins UK (3)) and an alternative source for summary data for CRC risk also stratified between aspirin users and non-users (obtained from UK Biobank (4)).

With regards to statistical analyses, I am very grateful to Dr Gemma Sharp for her help with quality control and running an EWAS with the methylation data as well as Dr Edward Mountjoy and Dr David Hughes for both their help in running linear regressions for the genotypes on changes in aspirin metabolite levels.

1.2 An introduction to colorectal cancer

1.2.1 What is colorectal cancer?

Colorectal cancer is a cancer that develops either in the colon (large intestine or bowel) or the rectum. The colon is divided into 4 sections and these are: the ascending colon, the transverse colon, the descending colon and the sigmoid colon. Cancer can develop in any of these or in the rectum which is found at the end of the colon (Figure 1-1) (5). A vast majority of colorectal cancer arise from the epithelium in what is known as the adenoma to carcinoma sequence (Chapter 1.3.2).

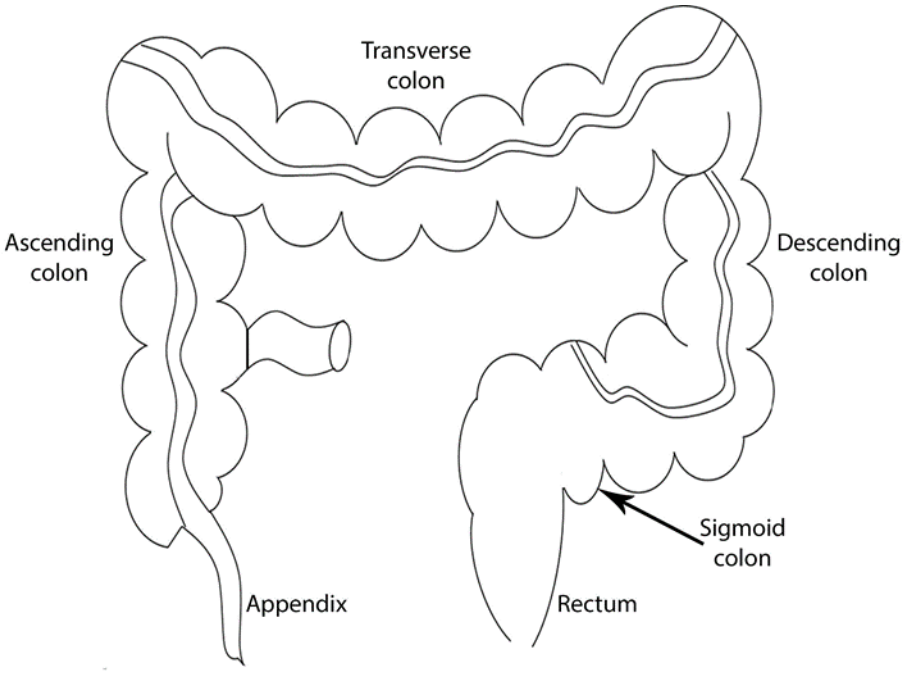


Figure 1-1- Structure of the colon and the rectum

The colon is comprised of 4 sections: the ascending colon, the transverse colon, the descending colon and the sigmoid colon. At the end of the colon is where the rectum is located.

1.2.2 Incidence and mortality patterns

In 2015, colorectal or bowel cancer was the 4th most common cancer in the UK with around 41,700 new cases each year. Bowel cancer incidence rates have remained stable over the last decade and mortality rates have decreased by 42% (6). Worldwide, it accounts for the third highest number of cancer cases and patient prognosis is directly correlated to timing of detection (7).

More than 50% of cases occur in more developed regions of the world with the highest incidence rates in Australia (44.8 per 100,000 for men and 32.2 per 100,000 for women) whereas the lowest incidence for men and women can be found in Western Africa (4.5 and 3.8 per 100,000, respectively) (8).

In terms of mortality, CRC was the second highest cause of cancer-related deaths in 2016 in the UK. For both males and females, it accounts for the 3rd most common cause of death by cancer. Mortality rates have decreased by 14% over the last 10 years and are predicted to fall by 23% by the year 2035 (6). Unlike incidence rates which are higher in developed regions of the world, mortality rates are higher in the less developed areas of the world (8).

Although incidence rates for over 50s has remained relatively stable, rates in the under 50s age group has increased by 33% between 1993-2015 in the UK population (9). This rising incidence trend is also found in the same age group in the United States. In adults aged 20-39, colon cancer incidence trends have risen by 1%-2.4% each year since the 1980s and rectal cancer has risen by 3.2% each year in 20-29 year olds between 1974-2013. In adults over 55, rates have generally been declining for both colon and rectal cancer (10). These statistics show a worrying trend that CRC incidence is increasing in the younger age groups and this highlights a need to find better prevention strategies to reduce risk of cancer.

1.2.3 Factors that increase risk of colorectal cancer

Whilst genetic susceptibility may play a role in a small proportion of CRC cases, most of CRC cases occur in individuals with no genetic predisposition or family history of the disease so are not inherited and are sporadic (11,12). It has been estimated that 48% of colorectal cancer cases in the UK may be due to environmental and lifestyle risk factors (13). Similarly, it has been estimated that 57.5% of CRC cases are attributable to environmental and lifestyle risk factors in the United States (14). These statistics highlight that environmental and modifiable factors play a large role in risk of CRC. Importantly, it is thought that the rising trend in colorectal cancer in the younger age groups is due to changes in diet and lifestyle.

Although a reduction in CRC incidence in older age groups may be attributable to a general reduction in alcohol consumption and smoking (10,15), birth cohorts have also shown a cumulative rising trend in obesity (16) which parallels the trend of CRC incidence. This suggests that obesity and CRC incidence may be linked (10).

There are also many other environmental factors that affect risk of CRC. Many factors have been shown to increase the risk of CRC and these include: processed meat consumption alcohol consumption, tobacco smoke (17), a high BMI and obesity (18) and exposure to radiation (17). On the other hand, physical activity (19,20) and a high fibre diet (21,22) are both associated with a decreased risk. The exact mechanisms by which these factors affect the risk of CRC still remains to be answered although it has been suggested that one plausible route might be through their effect on the gut microbiome (23).

1.2.4 Colorectal cancer diagnosis

To reduce colorectal cancer incidence, screening methods have been adopted to detect and remove adenomas and polyps to prevent tumour progression. Currently, two screening methods are commonly used. These are stool-based screening tests and visual examinations. One of the early signs of CRC is bleeding, therefore stool-based examinations typically involve a faecal occult blood (FOB) test that is used to test for human haemoglobin. On the other hand, visual examinations include endoscopic (colonoscopy and flexible sigmoidoscopy) examinations as well as radiologic examinations (computed tomography colonography). These are often used as a follow-up to positive FOB tests (as the stool test sometimes produces false positive results) and allow direct visualisation of the colon and thus removal of polyps (24).

Within the UK, these two types of screening methods are routinely offered to patients through the National Health Service (NHS). A flexible sigmoidoscopy is offered to all patients over 55 years of age and the FOB test is given every two years to patients between the ages of 60-74 (25).

1.3 Molecular aspects of colorectal cancer

1.3.1 Colonic crypt structure

The main role of the intestines is for nutrient digestion and absorption (26). The luminal surface of the colon is comprised of a single layer of epithelial cells that form a crypt structure otherwise known as the crypts of Lieberkühn (27). The main purpose behind the structure of the crypts is to protect the stem and progenitor cells at the base from damaging factors in the lumen such as microbiota-derived metabolites (28). Unlike the small intestine, the colon has no villi but rather a flat surface epithelium (26).

Intestinal multipotent stem cells are found at the base of the crypt that allow for a 5-day rate of colonic epithelium renewal (Figure 1-2) (27) and are also referred to as crypt base columnar cells (CBC) (26). Through asymmetrical division, they produce transit-amplifying cells that divide 4-5 times before they become terminally differentiated (26). These cells are able to differentiate into 3 types of terminally differentiated cells: colonocytes (also known as enterocytes), goblet cells and enteroendocrine cells (26). These cells are also able to differentiate into Paneth cells that play a role in innate immunity. Paneth cells are mainly found in the small intestine but can be present in the ascending colon (29). The equivalent of Paneth cells in the colon are deep crypt secretory (DCS) cells which are interspersed between the CBC cells. These cells provide a niche for the stem cells, evidenced by the loss of the stem cells when DCS cells were eliminated in colonic organoids (30). DCS cells allow for stem cell maintenance and self-renewal through their expression of the Notch ligands Dll1 and Dll4 (31).

At the bottom of the crypt is what is known as the stem cell niche. It is here that the stem cells are found along with mesenchymal cells called pericryptal myofibroblasts that produce the WNT signal ligands that then bind to the Frizzled receptors on the stem cells (29). The WNT and Notch signalling pathways are important for maintaining an undifferentiated and proliferative state at the base of the crypt. WNT signalling is highest at the base of the crypt and lowest towards the lumen (26).

Stem cells are able to differentiate into 3 cell types. The first type, colonocytes, are found at the top of the crypt and absorb water and electrolytes. The second, goblet cells, secrete mucins which allow for lubrication and provide a layer of protection from foreign particles such as bacteria and also protects from mechanical damage (32). The third type of cells are enteroendocrine cells, of which there are multiple types. These cells secrete hormones such as ghrelin, somatostatin, gastrin, and others (33).

Another rarer epithelial cell type are tuft cells which are involved in activating a type 2 response to parasitic infections. They secrete interleukin-25 (IL-25) into the lamina propria which results in the induction of a Th2 cell response (34). These tuft cells have also been shown to express LGR5 as well as other markers of stem cells that are usually only expressed in the crypt, indicating that these cells may act as reserve stem cells (35). Other cells that also play a role in intestinal immunological responses are microfold cells (M cells). These cells are present in the small and large intestine and transcytose pathogens from the lumen and transport them to the lymph tissues (36). Lastly, it has been reported that there are stem cells that differ to the CBCs and are slow cycling. These cells are found at the '+4' position in the crypt relative to the central cell in the crypt base which is given position 0. It is thought that both can give rise to cells of all lineages but the exact relationship between the active and quiescent stem cells requires further research (37).

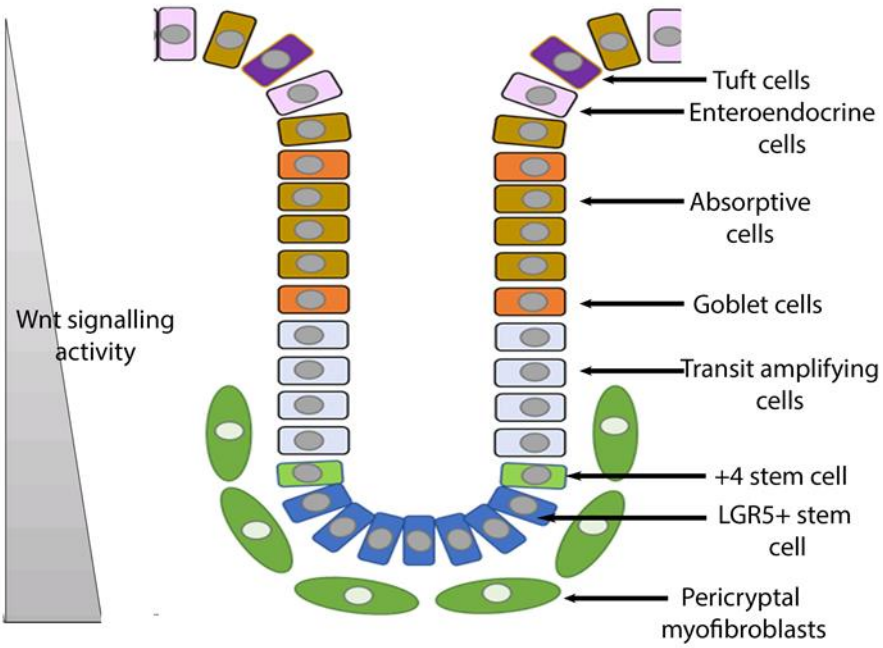


Figure 1-2- Structure of the colonic crypt (image adapted from Humphries and Wright, 2008 and Barker 2014 (29,38)**)**

The base of the crypt contains LGR5+ stem cells which are able to divide to form transit amplifying cells. At the +4 position is another stem cell with a more quiescent phenotype. Transit amplifying cells can terminally differentiate into the 3 main cell types: enterocytes, goblet cells and enteroendocrine cells. A rarer type of cell found in the crypt are tuft cells which are involved in an immune response against parasitic infections. Around the stem cells at the base are mesenchymal cells known as pericryptal myofibroblasts that produce ligands that stimulate an active WNT pathway in the stem cells. Other cells are also present in the intestinal crypt such as DCS cells and M cells but are not pictured here.

A normal colonic epithelium usually has a smooth surface with evenly spaced crypts. However, this regular structure is disrupted in colorectal cancer with branching and irregular crypts and the presence of a high quantity of desmoplastic stroma (39).

1.3.2 Adenoma-carcinoma sequence

In 1990, Vogelstein proposed a model in which multiple genetic mutations may provide epithelial cells with a growth advantage thereby resulting in cancer. In his model, he stated that it was the accumulation of genetic events, and not their order, which is the cause of this outcome. He proposed that the initiating event is a mutation in the gene adenomatous polyposis coli (*APC*) which leads to deregulation of β-catenin signalling and the formation of adenomas. This mutation is then followed by others for tumour progression (40). This is now known as the adenoma-carcinoma sequence which is summarised in Figure 1-3 (41). However, the adenoma-carcinoma sequence is an over-simplification of the alterations required for cancer progression, as not all tumours follow this sequence and other genetic mutations may be present (42).

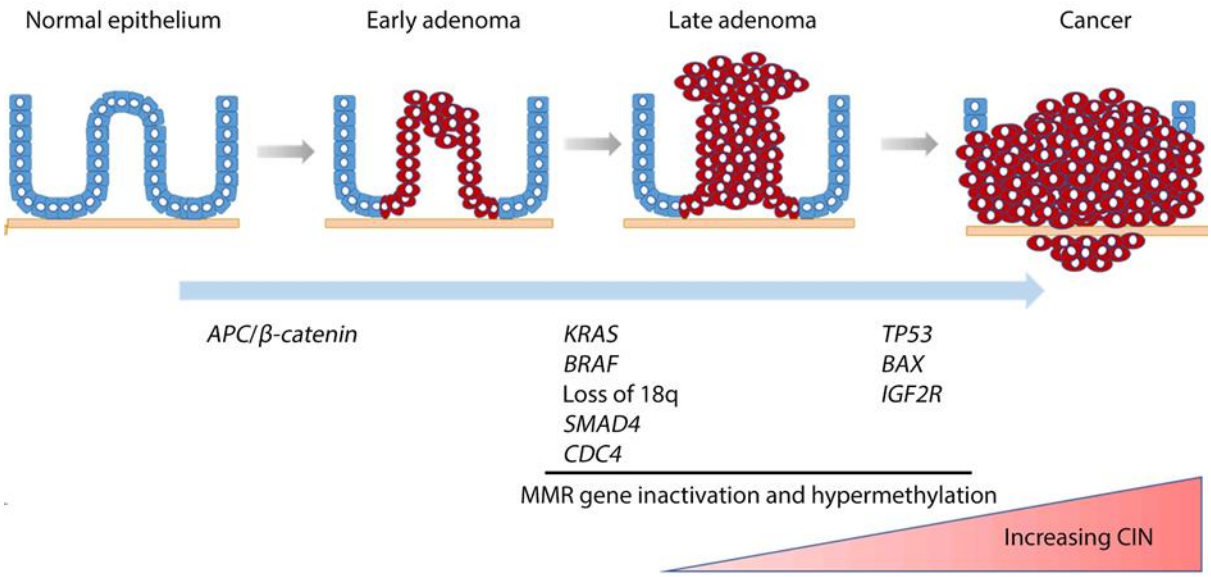


Figure 1-3- Schematic of the adenoma-carcinoma sequence in colorectal cancer (adapted from Walther et. al 2009 (41)**)**

The adenoma-carcinoma sequence presented highlights some of the well-known mutations along the pathway but is an over-simplification of all mutations involved in colorectal cancer progression. One of the first events to occur along the adenoma-carcinoma sequence is the loss of the *APC* gene or mutations in the *β-catenin* gene which leads to adenoma formation (40,43). APC promotes degradation of β-catenin therefore preventing it from entering into the nucleus and inducing transcription of WNT target genes. Loss of APC through mutations results in constitutive transcriptional activation of these WNT target genes (44). Aberrant WNT signalling is a common occurrence in colorectal cancer and is involved in cancer stem cell maintenance and metastasis (45). Further mutations occur along the sequence such as those in *KRAS*, loss of chromosome 18q and *SMAD4,* mutations in *BRAF* and *CDC4*. For progression to cancer, mutations may occur in *TP53* gene, *BAX* and *IGF2R*. Along this sequence, mismatch repair deficiency and increasing CIN may play also role in tumour progression (41).

1.3.3 Molecular pathways involved in CRC

Within the literature, three major molecular pathways contribute to tumour progression leading to CRC. These are:

- The chromosomal instability (CIN) pathway
- The microsatellite instability (MSI) pathway
- The CpG Island Methylation Phenotype (CIMP) pathway

The most common of these pathways is the CIN pathway thought to be the cause of 70% of sporadic colorectal cancers. It is characterised by the accumulation of many chromosomal abnormalities whether numerically (aneuploidy) or structurally resulting in an abnormal karyotype (46,47). This pathway can result in chromosomal rearrangements, loss of heterozygosity at tumour suppressor genes and mutations within oncogenes (such as *APC, P53* and *KRAS*) and tumour suppressor genes (46).

MSI is a condition in which DNA is prone to hypermutations due to the inability to repair the DNA through the mismatch repair (MMR) pathway (48). Microsatellites range from 1-6 base pairs of a repeating nucleotide sequence. They are prone to mutations as DNA polymerases cannot bind to them efficiently resulting in insertions and/or deletions. A consequence of this is frameshift mutations which ultimately can result in protein truncations(46). The MSI pathway is implicated in around 15% of sporadic colorectal cancers (48).

CIMP describes a molecular subtype of cancer whereby there is a high frequency of hypermethylation at some CpG islands (49). One commonly used panel to determine CIMP status is called the Weisenberger panel of genes and includes *CACNA1G, NEUROG1, RUNX3, SOCS1* and *IGF2* (50). Colorectal cancers that exhibit CIMP can be further classified as either CIMP-high or CIMP-low; CIMP-high tumours are hypermethylated in 3 or more of these 5 genes and commonly have an activating mutation in *BRAF* whereas CIMP-low tumours are hypermethylated in 1-3 of these Weisenberger panel of genes and are accompanied with *KRAS* mutations (51).

These 3 pathways are not mutually exclusive and it is possible that tumours may progress through their combined effect (46).

As well as the genetic components that may lead to tumour progression, there is also the contribution of epitranscriptomics and gut microbiomics. Epi-transcriptomics relates to post-translational modifications such as N^6-methyladenosine (m^6A), 5-methylcytosine (m^5C), inosine (I), 5-hydroxymethylcytosine (hm^5C) and pseudouridine (Ψ), that functionally modify RNA molecules (52). Alterations to the epitranscriptome have been implicated in colorectal cancer development (53). Furthermore, it has recently come to light that cancer progression might be affected by bacterial species, for example, colorectal cancer metastasis is associated with the presence of the *Fusobacterium* species of bacteria (54).

Overall, colorectal cancer has multiple pathways for development and these include genetic pathways, epigenetic changes, epitranscriptomic changes as well as changes in the gut microbiota.

1.3.4 Cancer stem cells

Stem cells are cells that have two main qualities: the ability to self-renew and the ability to generate differentiated cells for a specific tissue (multipotency) (27). Recently, it has become increasingly clear that stem cell hierarchies are plastic- this means that differentiated cells have the ability to de-differentiate to take the place of any lost stem cells. This is heavily influenced by the stem cell niche (55).

Much like adult stem cells which give rise to tissue specific cells, a cancer stem cell (CSC) theory was developed that states that tumours originate from hidden tumour stem cells. A small subset of cells within the tumours have the ability to self-renew and produce more differentiated cells (56) indicating a hierarchical organisation. These cells are termed CSCs and are recognisable through specific markers such as leucine-rich repeat-containing G-protein-coupled receptor 5 (LGR5) (57). This explains many observations such as tumour recurrence after chemotherapy or radiotherapy, tumour dormancy and metastasis (55).

Some tumours, such as glioblastoma, follow a unidirectional and hierarchical system whereby CSCs divide and differentiate and not the reverse, therefore targeting CSCs in these tumours prevents tumour regeneration (Figure 1-4) (58). Within other tumours, and much like adult stem cells, tumour cells may exhibit plastic properties in that both CSCs and non-CSCs have the ability to transition between the two states depending on external stimuli (55). The evidence for this was provided by de Sousa e Melo et. al (2017) who showed that ablation of LGR5+ CSC cells in mouse tumours temporarily prevents tumour growth. LGR5- cells then proliferate to replace the LGR5+ pool of cells and therefore allow tumour growth to continue (59). In these types of tumours, the microenvironment and extracellular signals are believed to influence cell plasticity. This is why therapies in these tumours should target the niche rather than the CSCs to prevent tumour regeneration (Figure 1-4) (55).

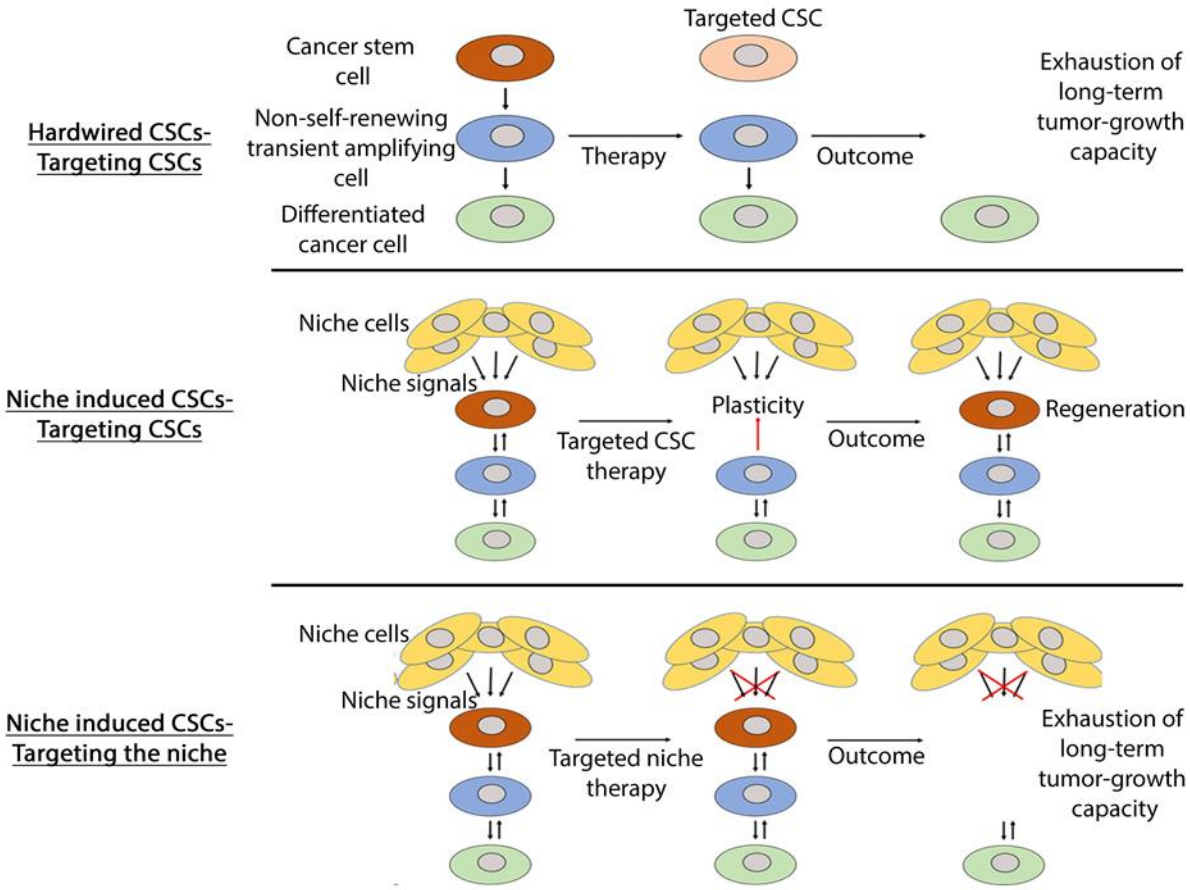


Figure 1-4- Consequences of therapies targeting CSCs (adapted from Batlle and Clevers, 2017 (55)**)**

Some tumour CSCs follow a unidirectional hierarchy whereby the CSCs divide and differentiate into cancer cells. Targeting the CSCs in this case is enough to diminish tumour growth and results in a cure. In other cases, CSCs are more plastic and niche signals can instruct more differentiated cells to de-differentiate after CSC loss. In this case, targeting CSCs only is likely to fail as progeny cells will replace and regenerate the lost CSCs. However, targeting niche signals that are needed for CSC maintenance will be more effective and prevent plasticity therefore resulting in tumour diminution (55).

1.3.5 Colorectal cancer metastasis

The cause of almost 90% of cancer-related mortalities is due to cancer metastasis (60). CRC can spread from a local site and disseminate widely through metastasis. Cancer can metastasise to distant organs, which are commonly the lung and liver for CRC (Figure 1-5) (42). Invasive and motile cells can enter the blood stream and be transported to other organs. Most of these cells don't survive in the circulation for reasons not yet known. Even from the small percentage that do, not all are able to establish a colony due to the foreign hostile environment (61). However, some cells can remain dormant for years before allowing the eventual regeneration of a tumour (62). They are able to survive in a foreign organ usually due to the acquisition of a stem-like phenotype providing them with self-renewal and growth capabilities. (42).

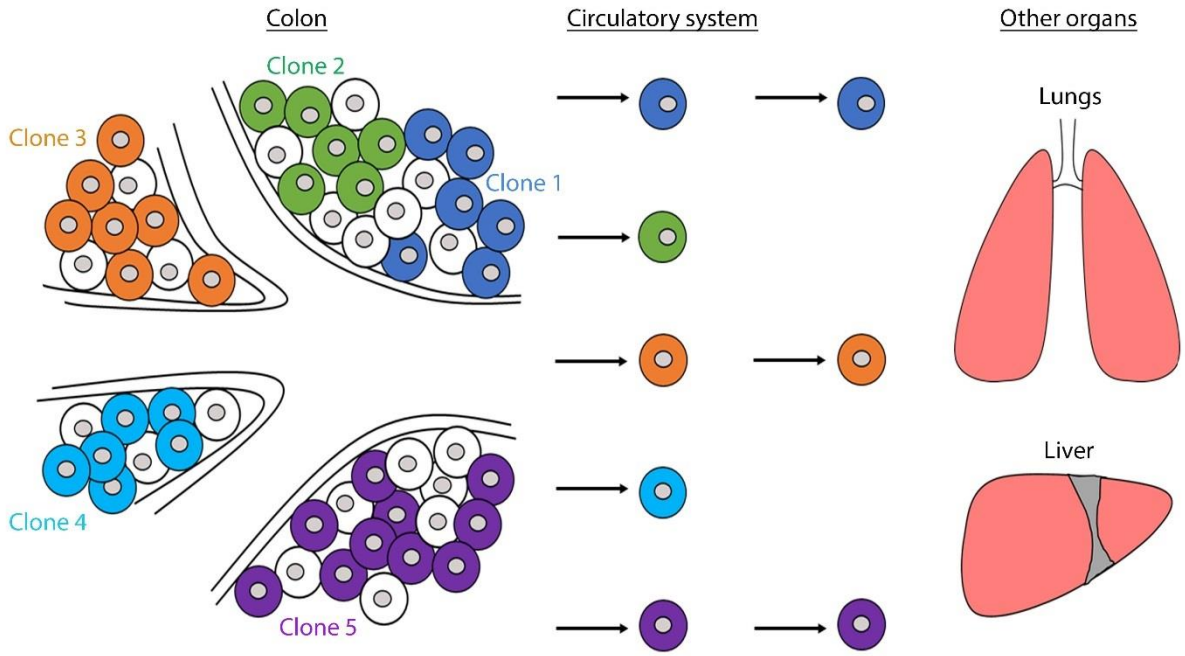


Figure 1-5- Colorectal cancer cell migration (adapted from Tauriello et. al (2017) (42)**)**

The colonic tumour may be comprised of multiple clones of heterogenous cells. These cells disseminate into the blood stream where a few are able to survive and migrate to distant organs such as the liver or the lungs. These cells can lay dormant for a long period of time before allowing for tumour regeneration in distant sites.

Cancer stem cells in CRC are able to self-renew and have the ability to initiate tumours and are thought to give rise to metastatic stem cells (63). It is believed that some of these cells are at the forefront of a tumour in tumour buds which are small clusters of detached cancers cells. These cells typically express genes involved in the breakdown of the extra-cellular matrix as well as so those needed for epithelial-mesenchymal transition (64).

1.3.6 Epithelial to mesenchymal transition (EMT)

One important process that occurs during metastasis is epithelial to mesenchymal transition (EMT). This involves tumour cells from an epithelium derived carcinoma, such as CRC, acquiring a mesenchymal phenotype. Epithelial cells are polarised, immotile and are connected via several types of junctions. They also usually interact with the basement membrane. Once EMT occurs, these cell-cell and cell-basement contacts are disrupted with a remodelling of the cell cytoskeleton allowing them to become more motile (Figure 1-6) (65,66).

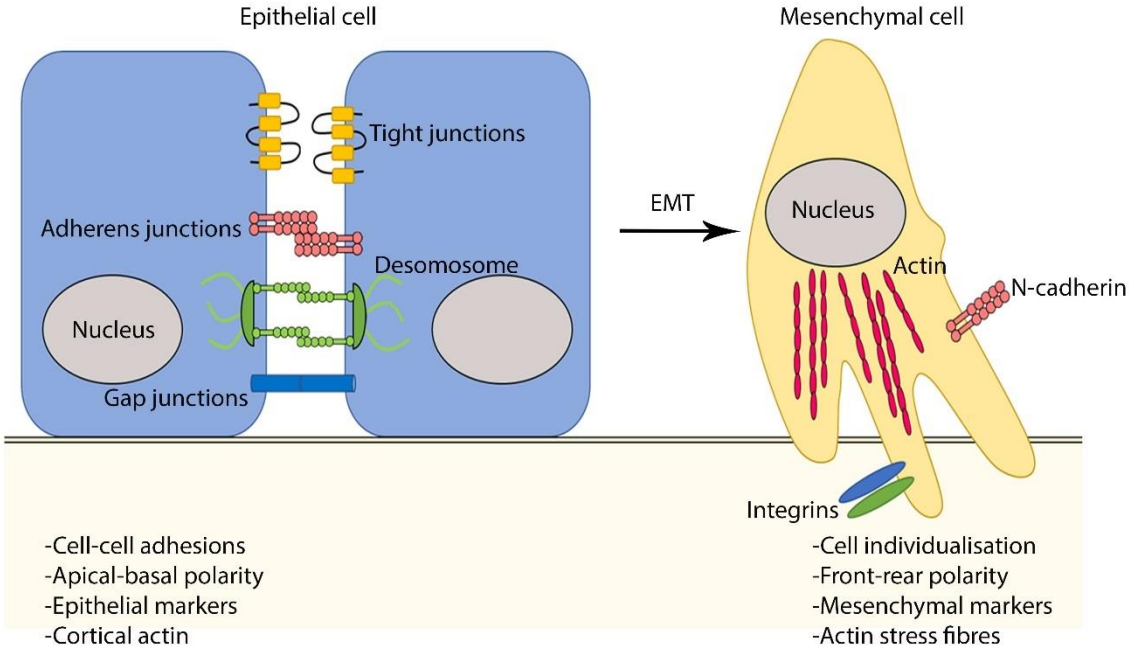


Figure 1-6- Epithelial to mesenchymal transition (adapted from Lamouille et. al (2014) (67)**)**

Epithelial cells are characterised by their cell-cell junctions and properties such as apical-basal polarity and cortical actin. Once these cells undergo EMT, they lose the cell-cell junctions and become more individual as well as having front-rear polarity and expressing actin stress fibres. All of these lead to cell motility.

Genes that encode epithelial cell junction proteins such as E-cadherin are downregulated via transcription factors such as Snail (68). At the same time, genes which promote mesenchymal adhesion are upregulated such as vimentin and fibronectin (69). Three groups of transcription factors (TFs) are activated early on in EMT and these include:

- the SNAIL family of TFs , SNAIL and SLUG (69)
- the Zinc finger E-box-binding homeobox 1 (ZEB) family of TFs, ZEB1 and ZEB2 (69)
- the TWIST family bHLH transcription factor 1 (TWIST) family of TFs, TWIST1 and TWIST2 (69)

As well as upregulation of these transcription factors, many signalling pathways are implicated in EMT. In the early stages of a tumour, the transforming growth factor beta (TGF-β) signalling pathway

possesses tumour suppressive properties but as the tumour advances, it has been shown to promote EMT in CRC (70). The secretion of TGF-β from platelets results in the expression of SNAIL, vimentin and fibronectin which promote EMT (71). The WNT signalling pathway also promotes EMT and has been shown to increase the expression of SNAIL thereby repressing E-cadherin and allowing local invasion (72).

Therefore, through upregulation of transcription factors and signalling pathways that promote EMT, cells become more invasive and motile allowing the cancer to metastasise.

1.4 Aspirin

1.4.1 History of aspirin

Aspirin is the most widely used drug worldwide (73). Salicylic acid (a derivative of aspirin) is found in willow trees and myrtle plants. It was used to treat pain, inflammation and fever as far back as 3000BC (74).

Aspirin as acetylsalicylic acid was initially prepared by Felix Hoffman, a German chemist, in1897. The literature suggests that his motivation was to produce an analgesic that would reduce the pain that his father was experiencing due to rheumatic fever but without the unpleasant effects of sodium salicylate (75). He found that by acetylating salicylic acid it reduced the side effects caused by its prolonged use (76).

In 1899, pure acetylsalicylic acid was then produced by Bayer, a German pharmaceutical company, under the registered trademark name of Aspirin. However, Arthur Eichengrün, claims that he was the one who instructed Hoffman to produce acetylsalicylic acid. However, he could not refute Hoffman's claim that he developed aspirin due to the political nature of the situation as Eichengrün was a Jew and the Nazi party had gained power. Therefore it took Eichengrün 15 years before he was able to contest what Hoffman had published (75).

1.4.2 Use of aspirin

Aspirin has been used clinically for many purposes. Its primary use is in prophylaxis to prevent cardiovascular disease occurrence through inhibiting platelet aggregation (77). Aspirin has also been shown to be useful as an analgesic to treat fever, inflammation and acute pain (78), for protection against cardiovascular disease (79), to reduce the risk of recurrent stroke (80) and to prevent preterm pre-eclampsia (81). A meta-analysis of observational studies published in 2018 has shown that aspirin is protective against many cancers including ovarian, gastric, esophageal, colorectal and breast cancers (82).

1.4.3 Sources of salicylates

As well as intake through aspirin, salicylates are found in natural sources such as fruits and vegetables, herbs and spices and some beverages (83).Two studies have published the salicylic acid content of certain food types from the Scottish population and the Australian population (83,84).

Wood et.al (2011) conducted a systematic review of salicylates in food in a Scottish population. Their results found that the main sources of salicylate intake were alcoholic beverages, other beverages including fruit juices, fruits and vegetables, tomato-based sauces and herbs and spices (Figure 1-7). Wood et.al (2011) were able to measure levels of salicylic acid in foods by using high-performance liquid chromatography (HPLC) with electrochemical detection and then using these results to estimate median salicylate intake in the UK population through the use of semi-quantitative food frequency questionnaires and applying it to a food composition database. They estimate that the male median intake of salicylic acid is 4.4mg/day and the female intake is 3.2mg/day (84). Other sources of salicylate intake include mouth wash, toothpaste and food preservatives (85).

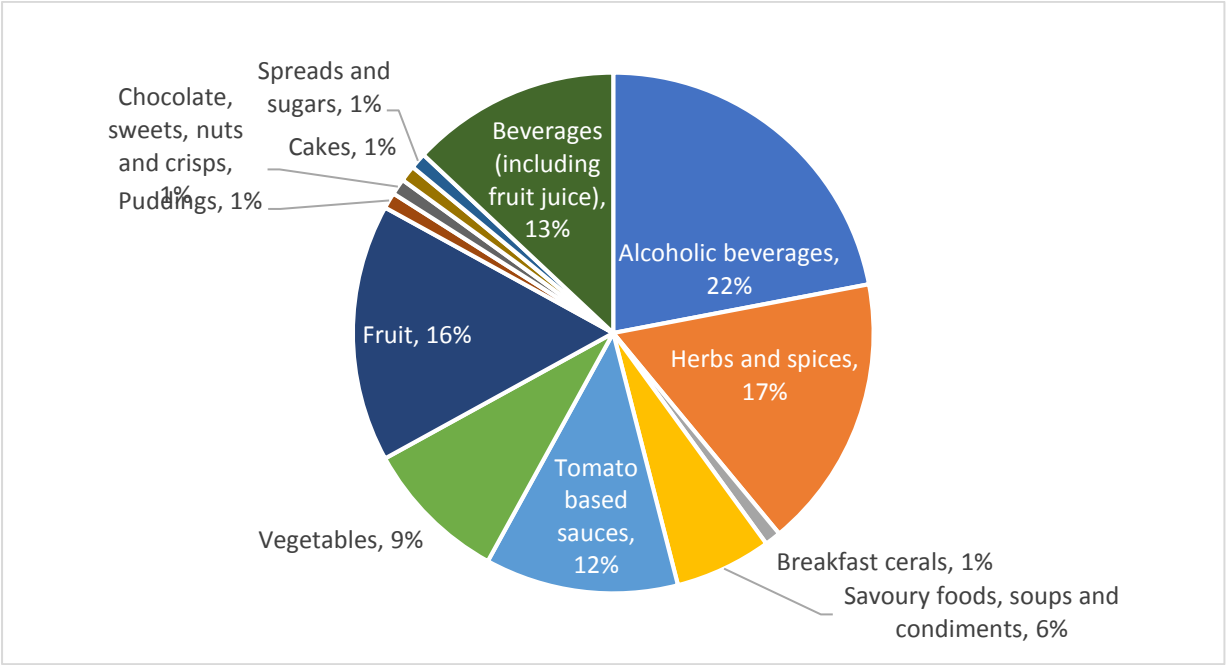


Figure 1-7- Relative amount of food groups contributing to total salicylate intake in Scottish population (adapted from Wood et. al (2011) (84)**)**

In the Australian population, Malakar et. al (2017) conducted a gas chromatography-mass spectrometry (GC/MS) approach to quantify the amount of total salicylic acid (both free and bound forms) in common Australian food items (fruits, vegetables, nuts and beverages). The 5 foods containing the highest amount of total salicylic acid are summarised in Table 1-1. When compared to the results achieved by Wood et.al (2011), there are some differences (83,84). This highlights one of

the problems of measuring dietary intake of salicylic acid which is that the levels vary in different varieties of the same food (84). It is likely that salicylate content of primary foods is affected by many factors such as plant varieties, seasonality, growing conditions and storage (86). This makes it difficult to measure dietary intake and therefore results should be interpreted with caution.

Table 1-1- The top 5 salicylate containing foods in an Australian study (83)

Food	Variety	Salicylic acid content (mg/kg)	
		Malakar et. al (2017)	Wood et. al (2011)
Cumin	powder	604.97	29.76
Coriander	powder	207.18	-
Vanilla bean	extract	130.26	-
Cinnamon	powder	57.24	0.78
Drinking chocolate	powder	51.48	-

A (-) denotes an unavailable/unmeasured result.

1.4.4 Aspirin and its side effects

Unfortunately, aspirin intake is not without some side-effects. Long term aspirin use is associated with upper gastrointestinal bleeding (87), the risk of which increases with age (87). Aspirin use is also associated with haemorrhagic stroke (88), peptic ulcerations (89) as well as nephrotoxicity and tinnitus (78).

The benefit of long-term aspirin use needs to be measured against the risk of side effects. It is estimated that 10-20 years of aspirin use in 1250 individuals would prevent death from CRC for 1 person (90).Cuzick et. al (2015) found that prophylactic use of aspirin at doses between 75-325mg/day for at least 5 years provides a favourable benefit-harm profile where benefits increase with time(91). On the other hand, 4-6 years of aspirin treatment would result in at least one major gastrointestinal haemorrhage or stroke in 800 individuals (92). Clearly, there is a need to develop a safer form of aspirin, or combine it with another drug, or stratify the population by identifying those most likely to benefit so as to improve the risk-benefit balance (93).

Aspirin intolerance can also be found in hypersensitivity diseases. Some patients suffer from salicylic acid hypersensitivity and up to 2.5% of Europeans may be affected by this (94). Aspirin-exacerbated respiratory disease (AERD) is the induction of an asthma attack in asthmatic patients through intake of a non-steroidal anti-inflammatory drug (NSAID) such as aspirin (95,96). The US National Heart, Lung and Blood Institute states that within adult patients, 21% have AERD (97). Aspirin intolerance can also induce urticaria in a condition known as aspirin-induced urticaria (AIU) (95). A study by Sánchez-Borges et. al (2015) found that 12.2% of patients with urticaria had AIU (98). Therefore, care needs to be taken before aspirin to avoid adverse side effects.

1.5 Epidemiological evidence for aspirin use and reduction of colorectal cancer risk

1.5.1 Epidemiological approaches to studying the role of aspirin in CRC- a brief introduction

As defined by the World Health Organization (WHO), "Epidemiology is the study of the distribution and determinants of health-related states or events (including disease), and the application of this study to the control of diseases and other health problems" (99). Therefore, this involves studying the distribution of diseases as well as trying to identify the causes. With regards to the relationship between aspirin and CRC, many epidemiological study designs have been used to address this.

1.5.2 Conventional epidemiological study designs

When faced with a question of whether an exposure results in an outcome, this question may be assessed through employing one of the methods listed in Figure 1-8. This is known as the "hierarchy of evidence" and the study designs are ranked in order of the weight of evidence they provide to infer causality. The methods differ by how much confounding (or internal validity) is present. (100,101).

Systematic reviews aim to systematically and comprehensively locate all information related to a particular question using an organised, transparent method with a replicable procedure. The quantitative results generated from a systematic review are then summarised using meta-analytic techniques to produce an overall summary of knowledge pertaining to a particular topic (102). The main advantage behind systematic reviews and meta-analyses is that they summarise large quantities of data and help to improve the precision in estimating risk by combining multiple pieces of evidence (103) such as the Cochrane reviews (104). However, they are limited by only summarising known interventions and can be influenced by publication bias. Publication bias is the defined as the publication of studies with positive findings or studies in English and a lower frequency of publications with null findings (105). Therefore, results from systematic reviews become limited and influenced by this (106).

RCTs are considered the gold standard for investigating the effect of an exposure on a specific outcome. Whilst observational studies, such as cohort studies and case-control studies, investigate both the exposure and outcome, RCTs observe the outcome of a given and specified exposure. In RCTs, a study sample (often small) is divided into two groups each of which is randomly assigned to either the treatment or a placebo. The major advantage of RCTs is that they allow the investigation of cause-effect relationships whilst minimising bias and avoiding the issue of confounding (107), by virtue of their randomised design.

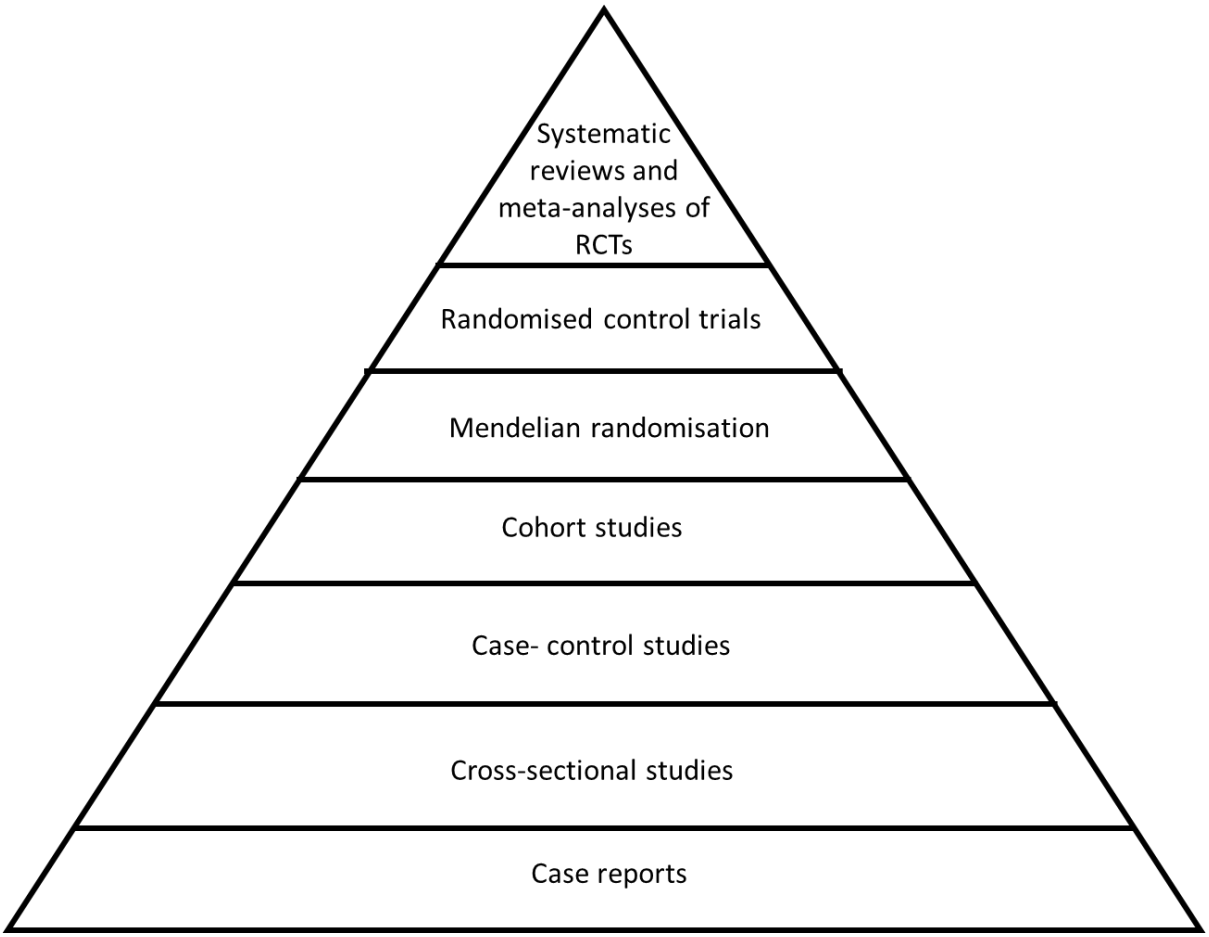


Figure 1-8- A pyramid depicting the hierarchy of evidence based on epidemiological studies. Adapted from Petticrew et al (2003) and Davies et. al (2018) (100,101)

Whilst RCTs are considered the "gold" standard when testing for causality, sometimes it is neither ethical nor possible for them to be carried out. In these instances, the next methods used are well designed observational studies: cohort studies and case-control studies (108). The modern definition of a cohort study refers to a group of people who are followed during their lifecourse to determine incidence or mortality rates for a particular outcome (109). On the other hand, case-control studies involve identifying cases through their outcome status and then following them up. Once cases have been identified, controls are selected as being subjects without the outcome but are from the same population (108). Cases and controls are matched based on age, sex and race as these are typically strong confounders of disease (110). Case-control studies are used to investigate rare outcomes or to investigate outcomes with long latency periods because the cases are selected at the beginning of the study (108). Observational studies allow the identification of possible associations between exposures and outcomes which can then be tested in an RCT setting.

1.5.3 Randomised trials for the prevention of vascular events

The most convincing evidence for aspirin's protective effect against CRC comes from randomised control trials for primary and secondary prevention of vascular events.

Rothwell et. al (2010) used data from 4 trials in the UK and Sweden as these had centralised death certification for secondary prevention data analysis. to analyse the results. Results included 4 randomised trials, 2 for primary prevention (Thrombosis Prevention Trial, British Doctors Aspirin Trial) and 2 for secondary prevention of vascular events (Swedish Aspirin Low Dose Trial, UK-TIA Aspirin Trial). By combining data from these trials, it was shown that aspirin reduced the 20-year risk of colon cancer incidence (HR 0.76; 95% CI: 0.60-0.96, p=0.02) and mortality (HR 0.65, 95% CI: 0.48-0.88, p= 0.005) (111).

Rothwell et. al (2011) then limited their search to trials that had a mean or median treatment duration of 4 years. A meta-analysis of 3 randomised trials with long-term follow up for daily aspirin use (10-20 years) for the prevention of vascular events was undertaken and showed that aspirin reduces the risk of death due to colorectal cancer (pooled hazard ratio(HR) 0.51; 95% CI: 0.35-0.74, p=0.0005) (112).

Using data from five primary prevention trials for vascular events that contained data on weight, Rothwell et. al (2018) also stratified by lower bodyweight (<70 kilograms) and higher bodyweight (>70 kilograms) and found that low-dose aspirin of 75-100mg decreases the 20-year risk of CRC in those with lower bodyweight (HR 0.64, 95% CI: 0.50-0.82, p=0.0004) but had no significant effect on those with higher bodyweight (HR 0.87, 95% CI: 0.71-1.07, p=0.32). Higher doses of aspirin (≥325 mg) was able to significantly decrease the risk of CRC for patients weighing up to 80kg (HR 0.69, 95% CI: 0.55-0.87, p=0.0014) but still showed no apparent protective effect for those weighing ≥80 kg (HR 1.08, 95% CI: 0.83-1.39) (113).

As well as investigating risk of incidence and mortality, Rothwell et. al (2012) also addressed the effect of daily aspirin on cancer metastasis. They included 5 randomised trials of aspirin use for vascular events (British Doctors Aspirin Trial UK Transient Ischaemic Attack Aspirin Trial, Thrombosis Prevention Trial, Prevention of Progression of Arterial Disease and Diabetes Trial and Aspirin for Asymptomatic Atherosclerosis Trialist) and found that aspirin reduced the risk of metastasis in CRC patients that did not present with metastasis initially (HR 0.29, 95% CI: 0.11-0.57, p= 0.0008) (114).

Overall, these results suggest that long-term aspirin use for the prevention of vascular events also reduces the risk of CRC incidence and mortality and this is more pronounced in those with a lower body weight.

1.5.4 Evidence from observational studies

As well as the evidence from randomised trials of aspirin for prevention of vascular events for a reduction in CRC incidence and mortality, observational studies have also shown a reduction in risk of CRC. A meta-analysis of observational studies involving case-control and cohort studies was carried out to identify the effect of aspirin use on the relative risk (RR) of CRC. Qiao et. al (2018) conducted a systematic review and identified 39 observational studies dating from 1989 - 2017 that have included aspirin use as an exposure and CRC as an outcome. After undertaking a random effects meta-analysis, it was shown that aspirin decreases the risk of colorectal cancer (RR: 0.78; 95% CI: 0.71-0.84) (Figure 1-9) (82), which follows with the RCT-derived evidence.

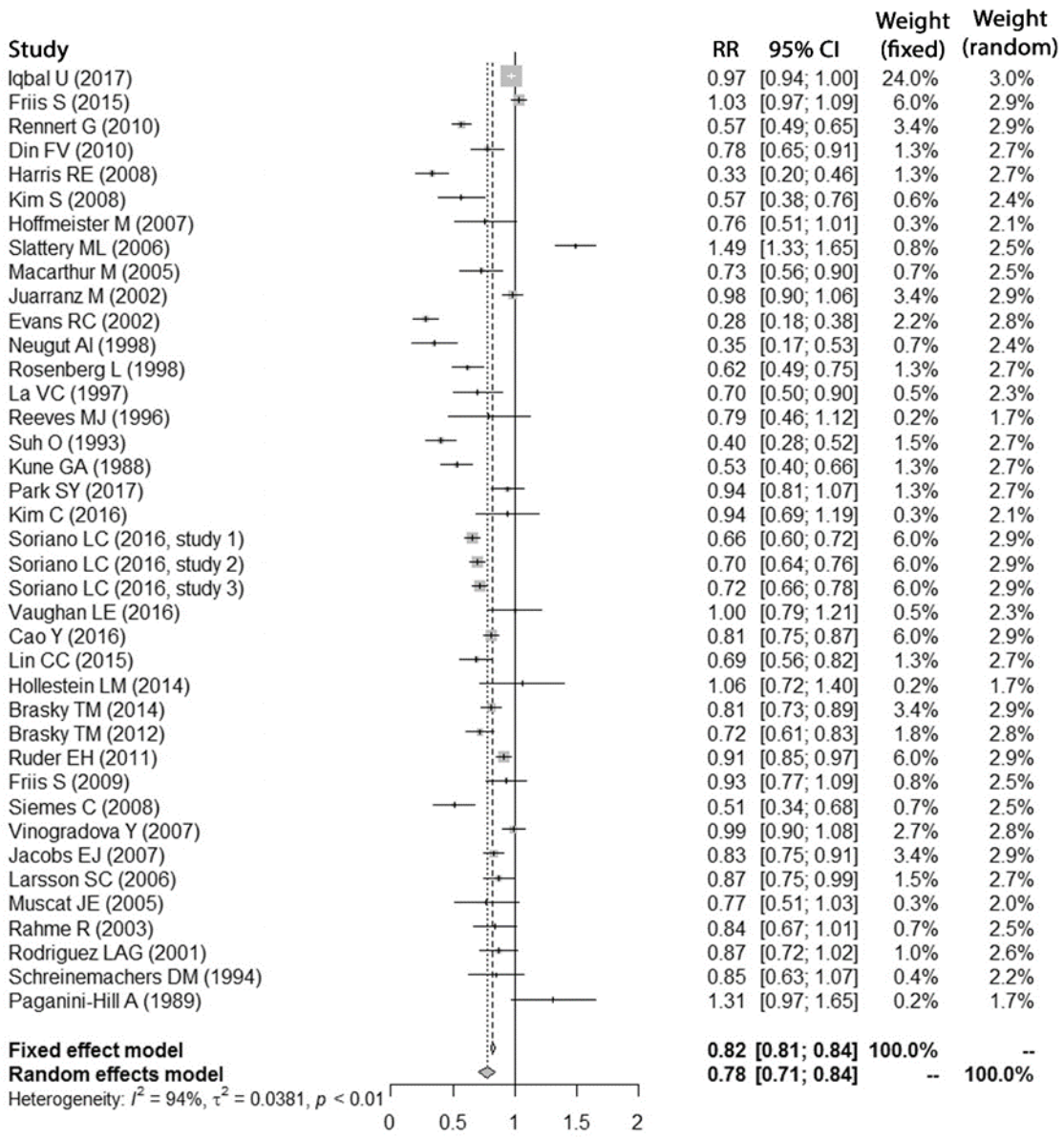


Figure 1-9- Forest plot of aspirin use and the risk of colorectal cancer using data from observational studies (adapted from Qiao et. al (2018) (82)**)**

A systematic review was carried out to identify all case-control and cohort studies that investigated the relative risk of aspirin use on colorectal cancer and results were combined via a meta-analysis. Abbreviations: RR, relative risk; CI, confidence intervals.

1.5.5 Aspirin for the prevention of colorectal adenomas

Adenomas are precursors of most CRCs (115) therefore one of the strategies to prevent CRC incidence is to prevent adenoma progression to cancer. One such method may be through the use of aspirin to prevent this secondary progression. The effect of aspirin on the prevention of colorectal adenoma recurrence has been studied using randomized clinical trials. Some of these studies include:

- Aspirin/Folate Polyp Prevention Study (AFPPS) (116)
- Association pour la Prévention par l'Aspirine du Cancer Colorectal (APACC) (117)
- Colorectal Adenoma prevention study originated in the cooperative trials group cancer and Leukaemia Group B (CALGB) (118)
- United Kingdom Colorectal Adenoma Prevention (ukCAP) (119)
- Japan Colorectal Aspirin Polyps Prevention (J-CAPP) (120)

The results of these studies are summarised in Table 1-2 with significant results in a bold red font. Cole et. al (2009) carried out a meta-analysis combining the results of all these studies excluding the J-CAPP trial. Through a random-effects meta-analysis, they showed that aspirin use reduced the risk of any adenoma recurrence (RR 0.83, 95% CI: 0.72-0.96, p= 0.012) and advanced adenoma recurrence (RR 0.72, 95% CI: 0.57-0.90, p=0.0046). High aspirin doses (300 or 325mg/d) also significantly decreased the risk of advanced adenoma recurrence (RR 0.71, 95% CI: 0.56-0.92, p=0.0089) but did not significantly decrease the risk of any adenoma. Only the AFPPS and APACC trials investigated the effect of low dose aspirin (81mg or 160mg/d) and together they show that low-dose aspirin reduced the risk of any adenoma recurrence (RR 0.83, 95% CI: 0.71-0.96, p=0.012) but not advanced lesions (121).

1.5.6 Drug repurposing

Drug repurposing (also known as drug repositioning, drug re-profiling, therapeutic switching and drug re-tasking (122)) is defined as the use of approved drugs for new diseases (123). It is believed to take around 10-17 years from the conception of a drug to when it becomes a licensed treatment for disease with a success rate of less than 10% (124). There are several advantages with regards to drug repurposing. Since the drug is already being used for other diseases, the safety profile is well-known. Furthermore, it reduces the large cost of creating a new drug and bringing it into the clinic (123).

Examples of drugs that have been repurposed over the years include: thalidomide, lithium, sildenafil, cyclosporine and warfarin. One other such example is aspirin (122). Aspirin is commonly used to treat cardiovascular diseases (77) and as an analgesic to treat fever, inflammation and acute pain

(78). Recent observational and RCT evidence with regards to aspirin's protective effect against colorectal cancer (Chapter 1.5.3 and Chapter 1.5.4) have highlighted how aspirin may be repurposed as a chemopreventative agent.

1.5.7 Current clinical trials

To specifically address the question of how aspirin may be beneficial in preventing cancer progression/recurrence, randomised controlled trials are currently being conducted. Within the UK and India, the "Add-aspirin" trial has recruited patients with lung, gastro-oesophageal, colon and prostate tumours and has randomised patients to either 100mg or 300mg of daily aspirin intake or a matched placebo. These patients have undergone curative treatment for their tumours such as surgery and are then randomised to the treatment with an active follow-up of at least 5 years. The primary outcome for CRC is disease-free survival and overall survival after 5 years with secondary outcomes including toxicity, adherence and cardiovascular events (125).

Similarly, the ASPirin Intervention for the REDuction of colorectal cancer risk (ASPIRED) trial is being conducted within the Massachusetts General Hospital, randomising patients to either daily low-dose of aspirin (81mg) or standard dose (325mg) or placebo. Participants will have had the removal of at least one adenoma during an endoscopy within 9 months of enrolment for the study. The trial duration is a minimum of 8 weeks and maximum of 12 weeks to investigate the primary endpoint of urinary PGE-M (a biomarker of prostaglandin levels) levels as well as various other secondary endpoints such as the expression of WNT signalling genes and aspirin's effect on bacterial populations and microbial pathways associated with CRC within the stool and saliva (126).

Other ongoing randomised controlled trials include:

- Aspirin to Reduce Risk of Initial Vascular Events (ARRIVE) (127)
- Aspirin for Dukes C and High Risk Dukes B Colorectal Cancers (ASCOLT) (128)
- ASPirin in Reducing Events in the Elderly (ASPREE) (129)
- CAPP3 (130)
- Systematic Evaluation of Aspirin and Fish Oil (seAFOod) polyp prevention trial (131)

The results from these trials will better inform whether aspirin should be used as a chemopreventative measure for colorectal cancer alone or in combination with other dietary factors and the effects seen at different dosages. These will help advise clinical recommendations to patients and may reduce the burden of CRC. Aspirin is a cheap intervention so the potential impact of using it to prevent CRC is very high.

Table 1-2- Clinical trials comparing aspirin and placebo for the prevention of colorectal adenomas (adapted from Drew et. al (2016) (132)**)**

Trial Name	Primary end point	Inclusion criteria	Participants (n)	Median duration of follow-up years	Aspirin dose	RR (95% CI)		References
						Adenoma	Advanced Adenoma	
AFPPS	Recurrent adenoma	Recent resection of colorectal adenomas	1,084	3.0	81 mg per day	**0.81 (0.69 - 0.96)**	**0.62 (0.39 - 0.97)**	(116,121)
					325mg per day	0.96 (0.82–1.12)	0.86 (0.58–1.30)	
					Any aspirin	0.88 (0.77–1.02)	0.74 (0.52–1.06)	
APACC	Recurrent adenoma	Recent resection of colorectal adenomas	184	4.0	160mg per day	0.88 (0.65–1.19)	1.19 (0.65–2.21)	(117,121)
					300mg per day	1.03 (0.77–1.37)	0.58 (0.25–1.37)	
					Any aspirin	0.95 (0.75–1.21)	0.91 (0.51–1.60)	
CALGB	Adenoma	Recent resection of Dukes' stage A or B1 CRC or B2 CRC and 5-year disease-free survival	517	1.1	325mg per day	**0.61 (0.44–0.86)**	0.77 (0.29–2.05)	(118,121)
ukCAP	Recurrent adenoma	Recent resection of colorectal adenomas	853	3.0	300mg per day	**0.79 (0.63–0.99)**	**0.63 (0.43–0.91)**	(119,121)
J-CAPP	Recurrent adenoma and CRC	Recent resection of colorectal adenomas and CRCs	311	2.0	100mg per day	**0.60 (0.36–0.98)**, pooled with advanced melanoma		(120)

Significant results are in a bold and red font. Abbreviations: AFPPS, Aspirin/Folate Polyp Prevention Study; APACC, Association pour la Prévention par l'Aspirine du Cancer Colorectal; CALGB, Colorectal Adenoma prevention study originated in the cooperative trials group cancer and Leukaemia Group B; CI, confidence interval; CRC, colorectal cancer; J-CAPP, Japan Colorectal Aspirin Polyps Prevention; RR, relative risk; ukCAP, United Kingdom Colorectal Adenoma Prevention. *Trial had a 2 × 2 factorial design. ukCAP: aspirin with or without folate supplement (0.5 mg per day).

1.6 Aspirin: mechanism of action

1.6.1 An overview of what is known about aspirin's mechanism of action

The evidence provided by RCTs and observational studies have shown a possible chemopreventative property of aspirin. The literature provides many possible methods and mechanisms by which this may be so. These are summarised in Figure 1-10 and further expanded on in the remainder of this Chapter and Chapter 1.7.

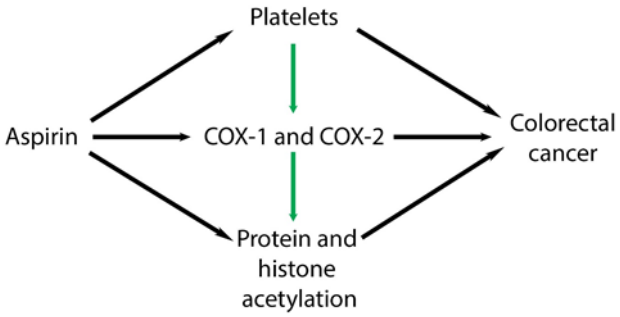


Figure 1-10- Well-known mechanisms of action of aspirin that may affect risk of CRC.

This is a very simplified diagram of the mechanism of action of aspirin. These pathways are the ones that will be focussed on in this thesis.

1.6.2 Aspirin and platelets

Megakaryocytes are cells found within the bone marrow that produce platelets, which are important for maintaining blood vessel integrity, hemostasis and to induce blood clots, all to prevent excessive bleeding through platelet aggregation (76). Platelet aggregation follows a three-step sequence: initiation, extension and stabilization (Figure 1-11).

- Initiation:

  Initially, platelets will bind to von Willebrand factor (vWF) and collagen at injured sites in the endothelium. This adhesion stimulates platelet activation leading to a release of bound ionic calcium to become free inside the platelet (133). This results in the release of adenosine diphosphate (ADP), thromboxin and thromboxane A_2 (TXA_2) from the activated platelets (76).
- Extension:

  The release of ADP, thromboxin and TXA_2 then activates other platelets to bind to the initial platelets bound to the site of injury (76).

- Stabilisation

  Platelets continue to aggregate through binding of surface glycoproteins, fibrin, fibrinogen and vWF. Further signalling events occur to maintain these platelet interactions and prevent them from dispersing due to other external forces (76).

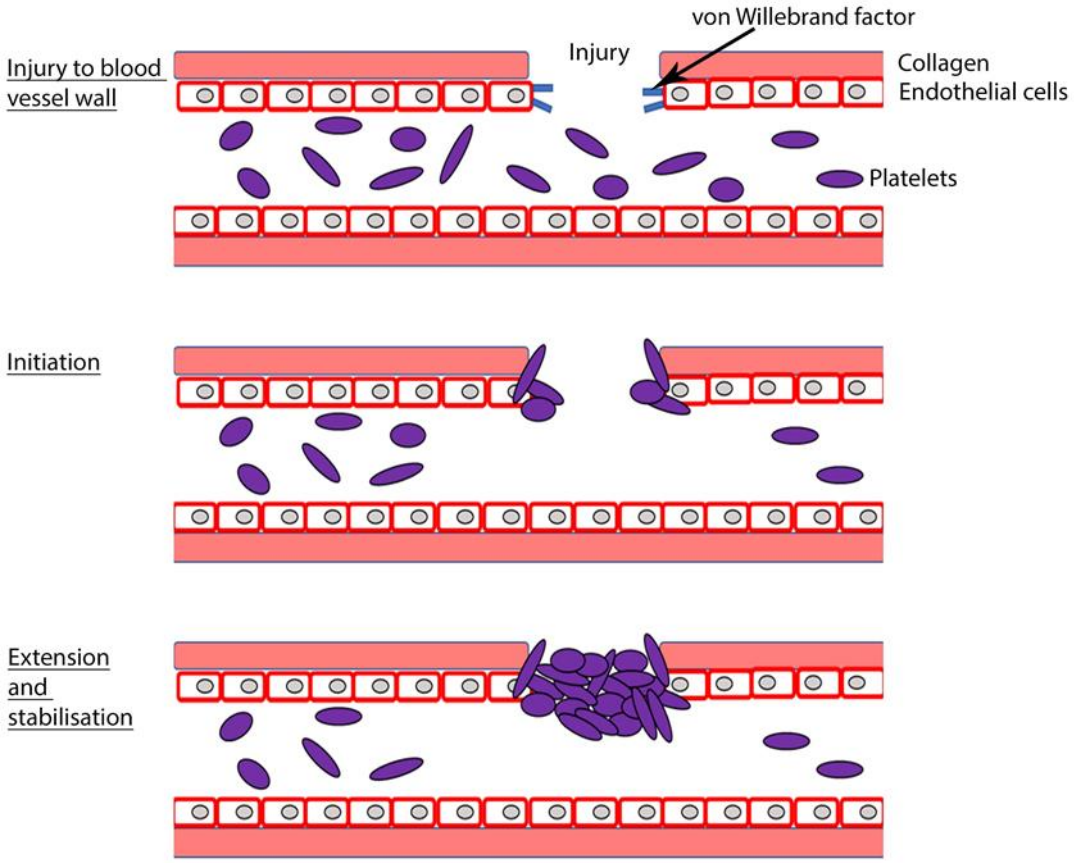


Figure 1-11- The platelet clotting cascade

At the exposure to an injury/cut to the blood vessel wall, platelets circulating in the blood will bind to free collagen and von Willebrand factor (vWF) expressed by endothelial cells. This activates the platelets to release bound calcium ions to become free inside the platelets as well as stimulating platelets to release adenosine diphosphate (ADP), thromboxin and thromboxane A2 (TXA2). These activate other platelets to bind the first layer of platelets bound to the site of injury and they become more stable through expressing surface glycoproteins, fibrin, fibrinogen and vWF. These interactions stabilise platelets against external forces.

Factors such as ADP, TXA_2, thrombin, epinephrine and collagen all activate G protein-coupled receptors to activate phospholipase A2. This enzyme then cleaves phosphatidylcholine to produce arachidonic acid which can be converted to a variety of prostaglandins such as prostaglandin E_2 by the cyclooxygenase enzymes (COX-1 and COX-2) (76). These prostaglandins play a role in the inflammatory response and can also be further broken down into TXA_2 to further mediate platelet activation and aggregation (Figure 1-12) (76,133).

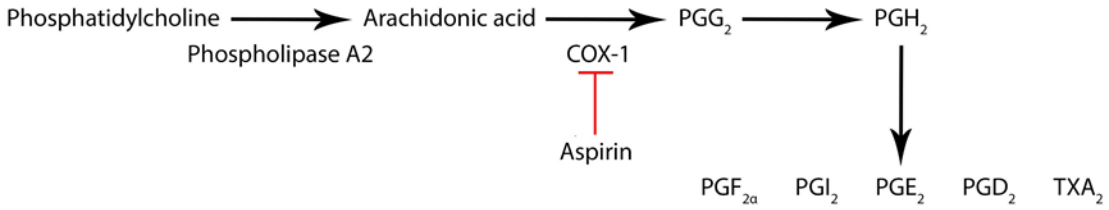


Figure 1-12- The prostaglandin synthesis pathway in platelets

The release of factors such as ADP, TXA_2, thrombin, epinephrine and collagen all activate G protein-coupled receptors to activate phospholipase A2. This enzyme then cleaves phosphatidylcholine into arachidonic acid which is further converted to prostaglandins via the COX-1 enzyme. These prostaglandins further mediate the inflammatory response as well as platelet activation. Aspirin can inhibit the production of prostaglandins through acetylating COX-1 in platelets and thus inactivating them.

1.6.3 Aspirin inhibition of COX-1 and COX-2

COX-1 and COX-2 enzymes are involved in the conversion of arachidonic acid into prostaglandins. Platelets primarily express COX-1. Aspirin can inhibit the COX enzymes through acetylating the serine residues 529 in COX-1 and 516 in COX-2 (134) . These serine residues are found in the active site of the enzyme thereby their acetylation blocks the ability of the enzyme's active site from binding to any substrates (135). This acetylation irreversibly inhibits the enzyme therefore decreasing the production of prostaglandins (Figure 1-12). In platelets, the activity of this enzyme can only be recovered through formation of new platelets (76).

COX-1 is considered to be the constitutive isoform of the two enzymes and is expressed in most tissues. On the other hand COX-2 expression can be induced with low levels detected in normal tissues. However, evidence suggests that it is more complex than previously believed. COX-1 is uniformly expressed in almost all tissues, with most of the protein expression found in blood vessels, platelets, smooth muscle cells, interstitial cells and mesothelial cells. COX-2 expression is more highly variable although expression can be found in almost all tissues and mostly expressed in their parenchymal cells (136).

Increased expression of COX-2 is commonly found in colorectal adenomas and carcinomas (137). Therefore, the cyclooxygenase(COX)/prostaglandin E2(PGE_2) signalling pathway is frequently upregulated in CRC and contributes to cancer progression through upregulating various pathways such as angiogenesis and cell proliferation as well as inhibiting apoptosis (138). Deregulation of the pathway can affect the various hallmarks of cancer such as evading apoptosis, self-sufficiency in growth signals, the ability to replicate continuously, to promote angiogenesis the ability to metastasise and the ability to evade the immune response (139). Part of the chemopreventative

effect of aspirin on CRC incidence and mortality may be explained through its ability to inhibit the COX-2/PGE_2 signalling pathway. The evidence from this comes in the form of COX-2 selective inhibitors that reduce the risk of CRC (140,141).

1.6.4 The crosstalk between platelets, tumour progression and aspirin

The literature has shown that there may exist a feedback loop between platelets and tumours cells in that tumour cells may activate platelets and platelets may stimulate tumors to grow and later metastasise.

Platelets are involved in allowing tumour progression in many ways. These include:

- Platelet aggregation around the tumour cells in the vasculature. This protects tumour cells from immune surveillance (142)
- Platelets induce tumour cell arrest in the vasculature as well as disruption of the endothelium to allow tumour cells to invade tissues and form secondary lesions (142,143)
- Platelets promote angiogenesis by inducing epithelial cells to proliferate and form new blood vessels (142).
- Platelets may promote tumour metastasis through interaction with cancers cells thus activating the TGFβ/Smad and nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB) signalling pathways in the tumour cells and promoting EMT (71).

To further investigate the effect of platelets on cancer metastasis, immunodeficient NOD-scid $IL2R\gamma^{null}$ (NSG) mice were injected with HT29 cells that were either cultured with or without platelets *in vitro* before injection. These mice were used as they are also used as a model for lung metastases. Injection of HT29 cells previously cultured with platelets increases the urinary PGE-M levels (the metabolite of PGE_2) compared to those not previously cultured with platelets. They also treated mice with an aspirin dosage that would only inhibit COX-1 activity in platelets and showed that this reduced the metastatic potential of HT29 cells previously cultured with platelets. Furthermore, pre-treatment of HT29 cells with platelets reduced the protein and mRNA expression of E-cadherin. This effect was prevented when platelets were exposed to aspirin (144). Therefore, their results provide evidence that platelets may play a role in promoting tumour metastasis and that aspirin may reduce risk of metastasis through its effects on platelets.

1.6.5 Aspirin's effect on colorectal cancer cells in cell culture

As well as reducing tumour metastasis through its effect on platelets, aspirin has also been shown to affect cancer cells independently of platelets. Using HCT116 and C26 cells, aspirin reduced the metastatic potential of these cells (145). It was also shown using SW620 and HCT116 cells that

aspirin was able to induce cellular senescence in cell culture (146).Furthermore, aspirin has been shown to inhibit cell cycle progression through arresting cells at G_1 phase of the cell cycle (147). These provide evidence that aspirin's effect on CRC cells is not solely through its effect on platelets.

1.7 Epigenetics

1.7.1 DNA structure

In 1974, Kornberg proposed that 200 base pairs of DNA are wrapped around a histone core (148). X-ray diffraction led to the discovery of a histone octamer consisting of two copies of the histone proteins H2A, H2B, H3 and H4 making up a histone core (149). DNA consisting of 146 base pairs wraps around this histone core. Extended from these histones are histone tails that are prone to post-translational modifications (Figure 1-13 A) (150). This structure is a repeating unit termed a nucleosome which can be further stabilised and compacted by another H1 histone protein (150). The first level of DNA organisation is termed 'beads-on-a-string' which consists of the repeating nucleosome structure with linker DNA between them (Figure 1-13 B) (151).

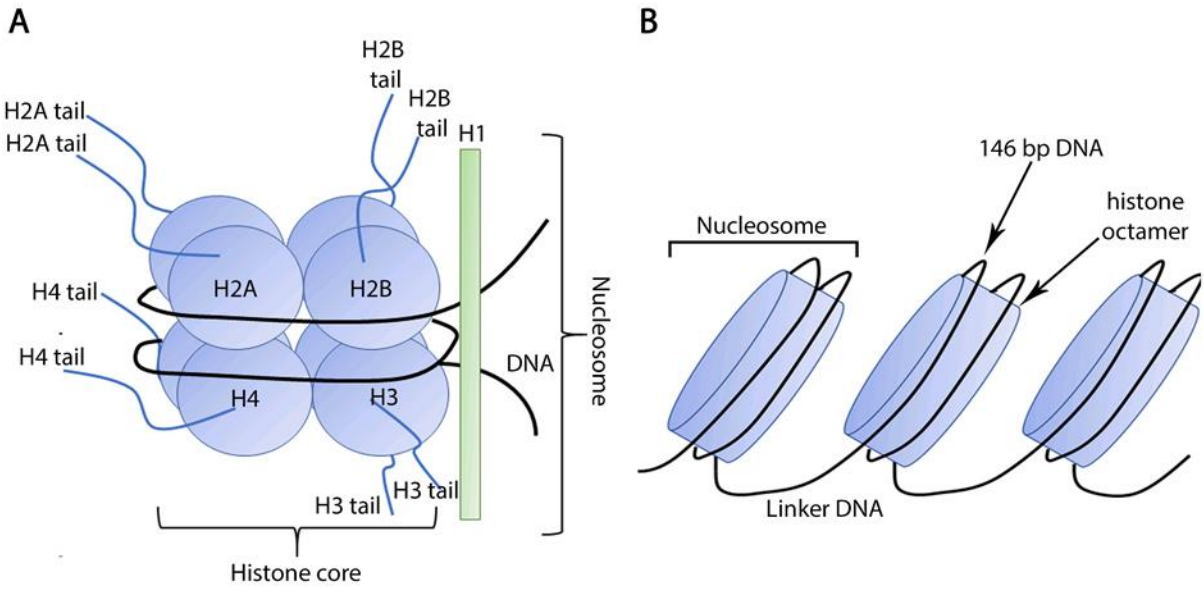


Figure 1-13- Models of the nucleosome and beads on a string

(A)A nucleosome consists of 146 base pairs of DNA that wraps around a histone core comprised of duplicates of 4 histone proteins: H2A, H2B, H3 and H4. Extending from these histones are histone tails that are subject to post-translational modification. This structure is further compacted and stabilised by the histone H1 protein. (B) The nucleosome is a repeating structure with linker DNA in between. This forms the first level of DNA organisation and is termed 'beads-on-a-string'.

1.7.2 What is epigenetics

Epigenetics refers to the changes in gene function that are not due to DNA sequence variations but rather they are due to a system that regulates the expression of DNA which can be inherited during cell division (152). Epigenetics affects the compactness of the DNA thus regulating gene expression. Highly condensed chromatin is called heterochromatin and contains genes that are not expressed. On the other hand, euchromatin is more open and accessible thus allowing for gene expression (Figure 1-14). The structure of the chromatin is affected by histone modifications, DNA methylation and nucleosome positioning (152).

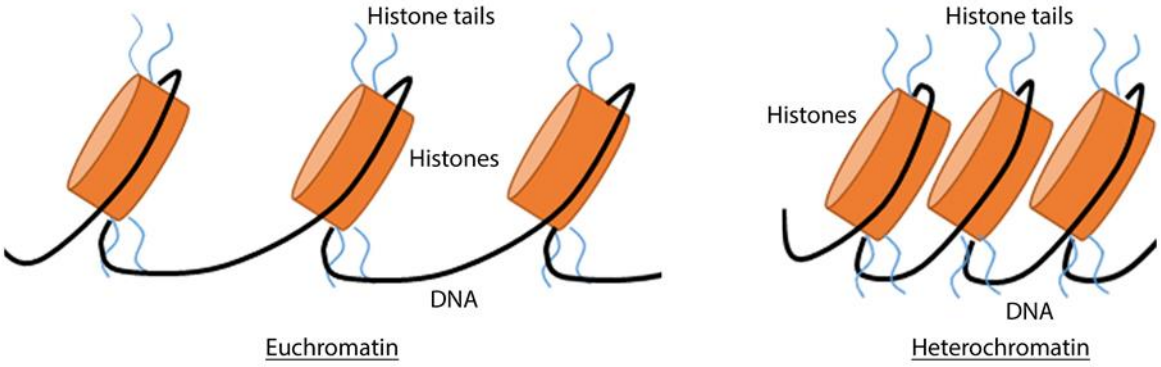


Figure 1-14- Models of euchromatin and heterochromatin

DNA can have an active or open structure whereby allowing gene expression, termed euchromatin. DNA can also contain inactive genes and a compressed structure, termed heterochromatin.

Epigenetics is comprised of:

- Post-translational modifications of proteins such as histone modifications- the main modification that affects accessibility of the DNA for transcription factors is acetylation (153).
- DNA methylation- a methyl group is added to a carbon of a cytosine residue (154). This usually results in silencing of the gene with the removal of acetylation from the histone molecules (155)
- miRNAs- small non-coding RNAs that have the ability to regulate translation of mRNA into proteins (156)

1.7.3 Post-translational modifications of histones

Several amino acids on the histone tails are subject to post-translational modifications. These are summarised in Table 1-3. They include acetylation, methylation, phosphorylation, sumoylation, ubiquitylation, deamination, adenosine diphosphate (ADP)-ribosylation and proline isomerization. These modifications have the ability to affect functions such as transcription, repair and condensation (157). More recently, other modifications have been discovered such as propoionylation and butyrylation (158). Others also include arginine acetylation, lysine ribosylation and phosphorylation and arginine acetylation (152,159). Figure 1-15 summarises the best studied modifications that are present on the N-terminus of histone tails (150,160). Acetylation and phosphorylation both neutralise the positive charge of lysine therefore reducing the electrostatic attractions between the histone tails and the DNA.

Table 1-3- Different types of histone modifications. Adapted from Kouzarides, 2007 ((157)

Chromatin modification	Amino acids modified	Functions affected
Acetylation	Lysine (K-ac)	Transcription, repair, replication, condensation
Methylation	Lysine (K-me1, K-me2, K-me3)	Transcription, repair
Methylation	Arginine (R-me1, R-me2a, R-me2s)	Transcription
Phsophorylation	Serine (S-ph) and threonine (T-ph)	Transcription, repair, condensation
Ubiquitylation	Lysine (K-ub)	Transcription, repair
Sumoylation	Lysine (K-su)	Transcription
ADP ribosylation	Glutamic acid (E- ar)	Transcription
Deimination	Arginine converted to citrulline (R > Cit)	Transcription
Proline Isomerization	Proline cis isomer to trans isomer (P-cis > P-trans)	Transcription

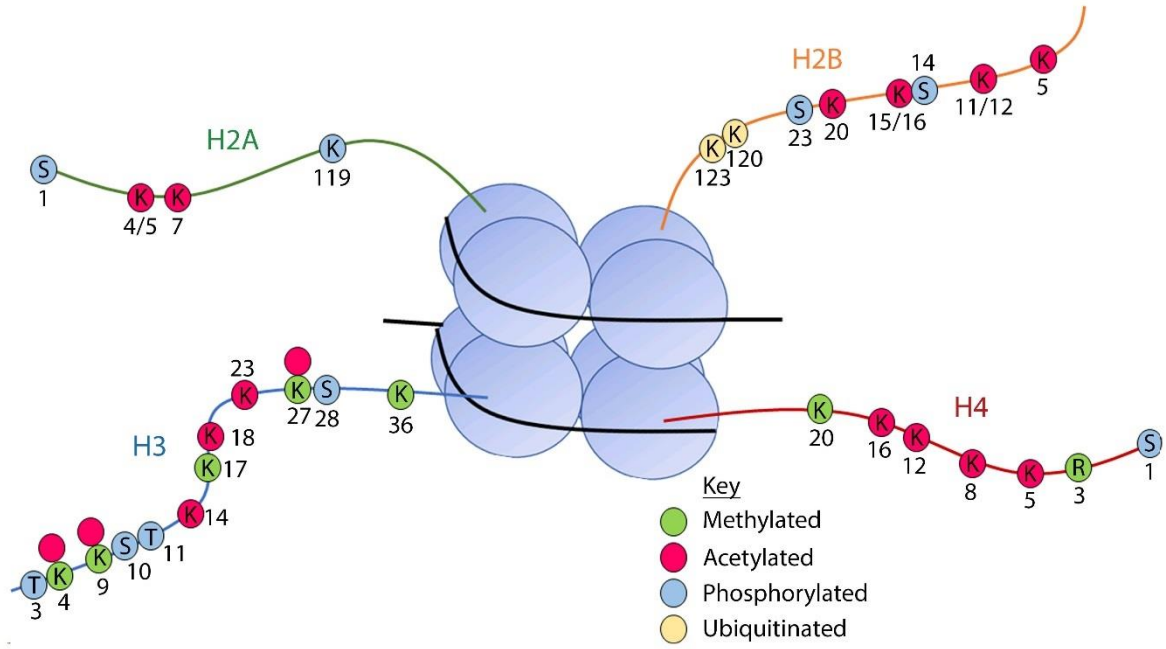


Figure 1-15- Schematic showing the post-translational modifications of histone tails. Adapted from Lawrence et. al 2016 (160)

The figure summarises the locations of 4 types of post-translational modifications on histone tails H2A, H2B, H3 and H4. These include: methylation (green), acetylation (pink), phosphorylation (blue) and ubiquitination (beige). The position of the amino acids is labelled in black for lysine (K), arginine (R), serine (S), threonine (T).

1.7.4 Histone acetylation

Crystallography studies in 1997 showed that histone tails extend from histone proteins in the nucleosome (150). These tails contain lysine residues that have a positively charged amine group. Histone acetyltransferases (HATs) and histone deacetyltransferases (HDACs) are enzymes that regulate the acetylation of the lysine residues. HATs are able to transfer an acetyl group from acetyl-

CoA thus removing the positive amine charge (Figure 1-16) (161) thereby weakening the electrostatic interactions between the histone tails and the DNA phosphate backbone which is negatively charged (153). This allows the DNA to be open and accessible to transcription factors (153,159). Conversely, HDACS remove the acetyl group ergo returning the positive charge and allowing for DNA compaction (162). As well as affecting DNA accessibility to transcription factors, histone acetylation may also prevent DNA from compacting into higher order structures allowing for transcription as well (153).

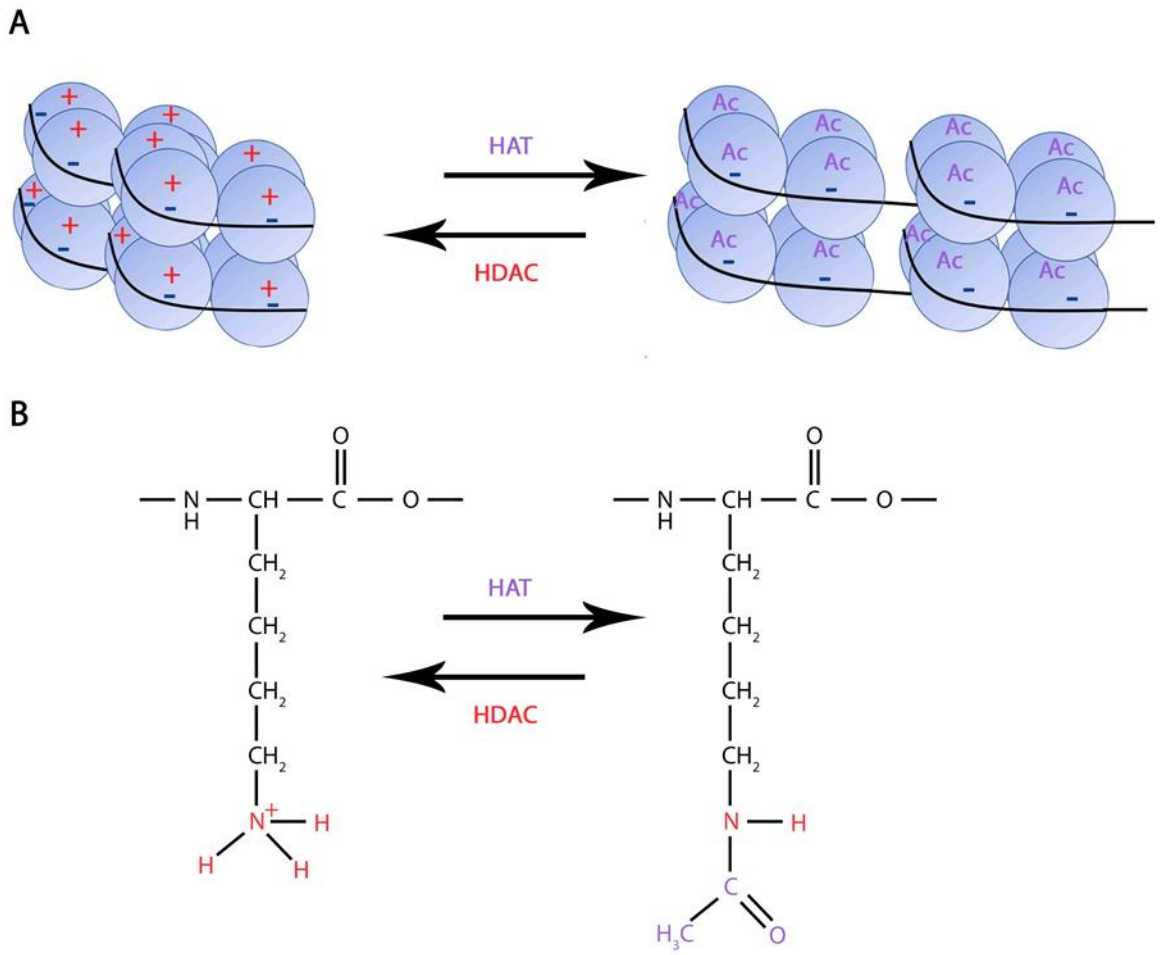


Figure 1-16- The effect of histone acetylation on DNA compaction. Adapted from Cress & Seto, 2000 (153) **and Yang & Seto 2007** (163)

(A) DNA is tightly packed due to the positive charge of the histone tails and the negative charge of the DNA phosphate backbone. HATs remove this positive charge by acetylating the histone tails thereby allowing the DNA to be less compact. HDACs are able to remove the acetylation and thus the histones resume their positive charge. (B) The figure shows lysine residues that normally have positive charge due to the amine group. HATs are able to remove the amine group (red) and replace it with an acetyl group (purple) thereby removing the positive charge. On the other hand, HDACs are able to remove the acetyl group and replace it with a positively charged amine group.

1.7.5 Histone methylation

Another type of histone modification involves the methylation of lysine and arginine residues on histone side chains. Histone methyltransferases (HMTs) are a group of enzymes that catalyse the addition of a methyl group and histone demethylases (HDMs) catalyse the removal of the methyl group (164). Histone methylation does not affect the electrostatic charge of histones therefore it does not affect the structure of chromatin. Instead, methyl groups allow the binding of transcriptional complexes which is why histone methylation can allow both expression and silencing of genes (164). An example of this is that histone 3 lysine 4 trimethylation (H3K4me3) is associated with transcriptional activation whereas histone 3 lysine 9 trimethylation (H3K9me3) and histone 3 lysine 27 trimethylation (H3K27me3) are associated with gene silencing (157,165). As well as affecting gene transcription and DNA replication, histone methylation has also been shown to be associated with various cancer pathways and aberrant methylation has been observed in CRC (164).

1.7.6 Post-translational modifications of other proteins

As well the modification of histones, other proteins with catalytic functions such as enzymes can also be modified through post-translational modifications (PTM). These are covalent modifications that occur after DNA transcription and translation into proteins (166) These include: acetylation, phosphorylation, glycosylation, methylation and ubiquitylation (166). These modifications can affect enzymatic activity, localisation within the cell, interactions with other proteins, the ability to bind DNA and also the stability of the protein. The most well-characterised PTM is protein phosphorylation and over the last 20 years, protein acetylation has also come to the forefront of research (167). Proteins can be acetylated through the addition of an acetyl group to an amino acid. This can be on serine and threonine residues (168) as well lysine residues on histones, of which results in chromatin decondensation and gene transcription (167).

1.7.7 DNA methylation

DNA methylation is the covalent addition of a methyl group ($-CH_3$) to the fifth position on a cytosine residue in a cytosine-guanine dinucleotide (5'-CpG-3'). A cytosine bound to a guanine residue via a phosphate bond is commonly called a CpG dinucleotide. Most of these dinucleotides are methylated. However, some are unmethylated in clusters known as CpG islands which are commonly found in promoter regions of genes (169,170).

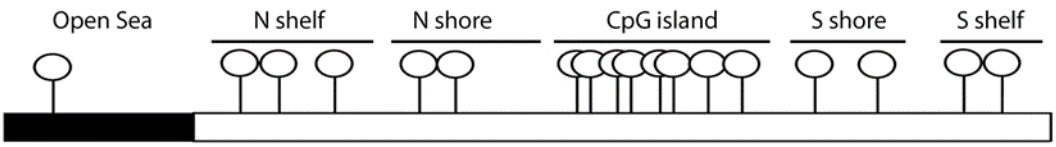


Figure 1-17- Location of CpG dinucleotides in relation to CpG islands

Shores are considered 2-kb sequences either side of the CpG island, shelves are 2-kb sequences next to the shore. Open sea regions are CpG dinucleotides not found near the CpG islands or the shores and shelves. Abbreviations: N, north; S, south.

The location of CpG dinucelotides can also be found in regions around the CpG islands called: north shores, north shelves, south shores and south shelves. Shores are considered 2-kb sequences either side of the CpG island, shelves are 2-kb sequences next to the shore. In addition, CpG dinucleotides not found near the CpG islands or shores and shelves are termed open sea regions (Figure 1-17) (171). Most of the tissue-specific DNA methylation is located in the shores of CpG islands (172). Also, shore methylation may affect splice variant expression as well as inactivating transcription (173).

Methylation of cytosine residues is catalysed by DNA methyltransferases (DNMTs). There are three DNMTs, two of which methylate new sites (DNMT3a and DNMT3b) and one that maintains the current methylation status (DNMT1) (174,175). The removal of DNA methylation is thought to occur either through a passive mechanism or an active mechanism. The passive process involves preventing the maintenance methylases from maintaining current methylated sites during DNA replication (176). As for the active removal of methylation, three mechanisms have been proposed: the removal of the methyl group from the cytosine, the removal of the whole cytosine base or the modification of the methylated cytosine residue (176). It was shown in 2009 that 5-methylcytosine (5mC) can be oxidised into 5-hydroxymethylcytosine (5hmC) (177) and this is catalysed by ten-eleven translocation (TET) enzymes (177,178). It has been proposed that the conversion of 5mC to 5hmC is an intermediate during DNA demethylation (179), however, it may also be involved in transcriptional regulation and act as a unique epigenetic mark (180). Further reactions occur to return the 5hmC into a demethylated cytosine residue as summarised in Figure 1-18.

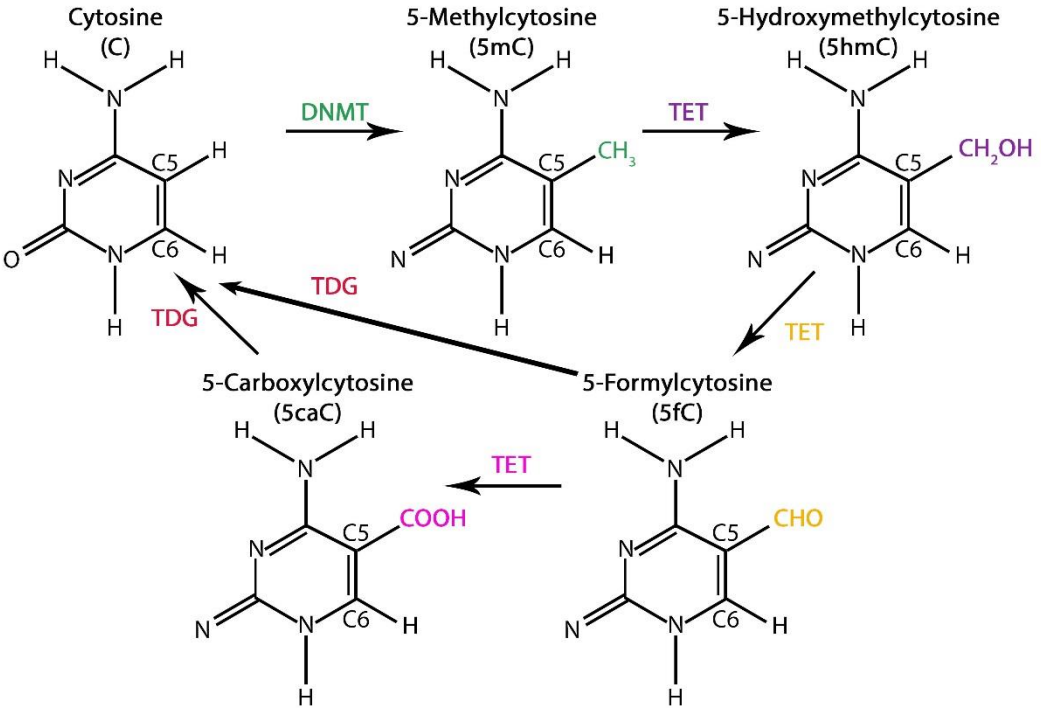


Figure 1-18- The pathway of DNA methylation and demethylation

Cytosine is methylated via a DNA methyltransferase (DNMT) into 5-methylcytosine (5mC). TET enzymes oxidise 5mC into 5-hydroxymethylcytosine (5hmC) and further oxidises this into 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). Thymine DNA glycosylation (TDG) can then cleave 5fC and 5caC to form cytosine (181).

It is believed that DNA methylation can prevent transcription factors from binding to promoter regions and thereby prevent gene transcription. Methylation can attract methyl-binding proteins , of which some can recruit repressor complexes that modify histones resulting in a compact chromatin structure (182,183).

1.7.8 DNA methylation and its role in cancer

Widespread changes in DNA methylation have been observed in multiple forms of cancer including colorectal cancer when compared to normal adjacent tissue. These methylation patterns are used to determine CRC subtypes (184) as well as identifying the CRC from healthy normal tissue (173) and to develop biomarkers to detect CRC (185).

It is commonly believed that methylation may play a causal role in cancer through hypermethylation of gene promoters which results in silencing of tumour suppressor genes and hypomethylation of oncogene promoters resulting in an increase in transcription (186). Whilst aberrant DNA methylation is present in cancer, there are still many questions that are unanswered, such as whether DNA methylation is the cause of cancer or just an effect, the timing of when the genes are silenced/methylated and which genes need to be silenced for cancer progression (186) .

1.7.9 Changes in DNA methylation during adenoma to carcinoma progression

Recent evidence has shown that changes in DNA methylation occur as an early event in the progression from adenoma to carcinoma sequence. The methylation status of 18 adenomas was assessed using the Illumina HumanMethylation450 Beadchip arrays revealing 86460 differentially methylated probes where 40% were hypermethylated and 60% hypomethylated compared to normal colon mucosa. Luo *et al.* carried out cluster analysis on the adenoma samples using the 10,000 most variable CpG probes and their results revealed two distinguishable CpG patterns- adenomas with high methylation termed *adenoma-H* and adenomas with low methylation termed *adenoma-L*. They also found that methylation pattern of adenoma-L polyps was similar to normal colon mucosa, unlike adenoma-H patterns which were more similar to intermediate and highly methylated CRC samples. These results suggest that the epigenome of adenomas may help identify the 90% of polyps that remain benign and those that will progress to CRC (187).

1.7.10 Aspirin and epigenetics

The most well-known effect of aspirin on epigenetics is its role in protein, enzyme and DNA acetylation, shown as far back as 1968 (188). Aspirin has the ability to acetylate proteins directly, for example, serine residues in the COX-1 and COX-2 enzymes (134). Aspirin also acetylates other proteins as well as histones although whether this is directly or indirectly is not yet clear. Guo et. al (2016) showed that aspirin treatment was able to increase the percentage of histone 3 lysine 27 trimethylation acetylation (H3K27ac) positive cells compared to untreated in mice models (189). Tatham et. al (2017) undertook a proteomic approach to look at the effect of aspirin on the lysine acetylome in HeLa cells and showed that aspirin was able to acetylate histone proteins hence having a global role on transcription (78). Evidence for aspirins indirect effect on acetylation is shown through experiments carried out by Passacquale et. al (2015) who showed that aspirin and salicylic acid increased the activity of histone acetyltransferases (HATs) and reduced the activity of histone deacetylases (HDACs) in HUVECs treated with tumour necrosis factor α (TNF-α) (190). Therefore aspirin also histone acetylation indirectly through influencing the activity of the proteins involved in acetylation/deacetylation. Since aspirin is known to influence DNA and histone acetylation, it may be that aspirin is affecting other epigenetic modifications such as DNA methylation. This is discussed in more detail in Chapter 4.1.

1.8 Mendelian randomisation and the study of cancer

1.8.1 What is causality?

As previously stated by Rothman and Greenland, the cause of a disease or event is defined as an event, condition or characteristic that occurred before the initiation of the disease and that was necessary for the disease to occur, or would have resulted in the disease eventually (191). However, researchers now understand that multifactorial diseases exist and that there may be many risk factors small in magnitude associated with the outcome as opposed to one main exposure that leads to disease (192). This is also found within genetic epidemiology, rarely is it that a single genetic polymorphism would result in a disease and it is more likely that a combination of many genetic polymorphisms may contribute to disease risk (193)

Sir Austin Bradford-Hill listed 9 factors to be considered when trying to assess causality and these include: strength of association, consistency, specificity, temporality, biological gradient, plausibility, coherence, experiment and analogy and these are often referred to as the Bradford Hill Criteria (194). Using these nine criteria, the evidence that the exposure associated with the disease may be causal thus informs / decision making (195,196).

1.8.2 Limitations of previous epidemiological studies of the role of aspirin in CRC

Limitations of trials with regards to cancer prevention include the fact that prevention trials focus on a limited time frame as opposed to life-long exposures (197). Often times, a sufficient follow-up period to a trial is not given which means that sometimes the effect of an intervention is undetected. Furthermore, the issue with long term chemoprevention trials can be non-compliance of those taking the intervention and contamination of the control arm (198).

With regards to observational epidemiology such as case-control and cohort studies, these also have their own disadvantages summarised in Figure 1-19. These include:

- Confounding: where the exposure is not directly associated with the disease but is associated with other factors that do alter disease risk (199)
- Reverse causation: where the exposure is instead affected by the outcome rather than the exposure affecting the outcome (199).

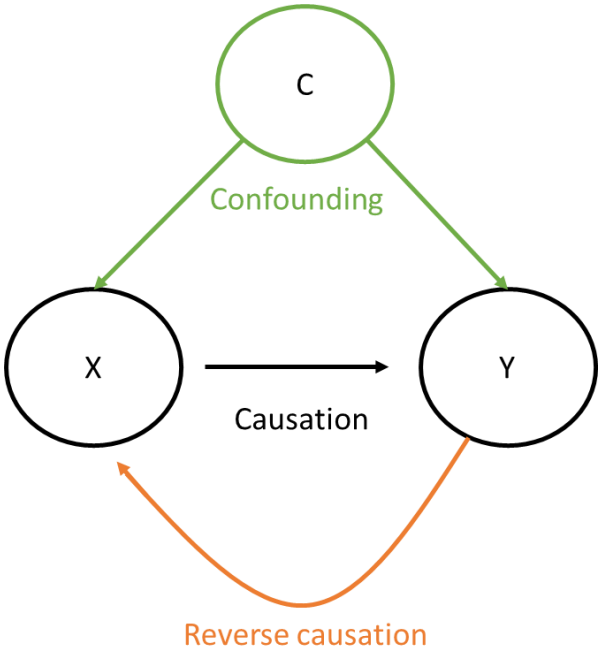


Figure 1-19- Graphical view of causation, reverse causation and confounding.

Abbreviations: X, exposure; Y, outcome; C, confounding.

Two examples of where observational epidemiological results have failed to replicate in RCT settings include the observation that increased dietary β-carotene reduces the risk of smoking-related cancers and the inverse association observed between vitamin E and coronary heart disease (CHD) (200,201). Early observational studies indicated a possible association between increased β-carotene intake from the diet and reduction of risk of certain cancers (202) as well as cardiovascular disease (203). However, when an RCT was conducted, no evidence of an association was found between β-carotene supplementation and neoplasms and cardiovascular disease (200). Observational studies showed evidence that vitamin E reduced the risk of vascular disease incidence (204,205). Based on this evidence, an RCT was conducted but found no association between vitamin E intake and heart attacks and strokes (201).

Whilst RCTs and observational studies have shown that aspirin may reduce the risk of CRC (Chapter 1.5.3 and Chapter 1.5.4), they do not provide us with any mechanistic insights into how it does so. Therefore, whilst these types of studies are useful tools to study association, other analyses need to be undertaken to identify mechanisms of action, including other epidemiological analyses as well as laboratory-based methods.

1.8.3 How can recent developments in epidemiology be applied to understand the role of aspirin in CRC

Epidemiology is an expanding field advanced by new technologies and platforms. These advancements allow the interrogation of multiple questions in a short time frame. Some of these methods are summarised below.

Genome-wide association studies (GWASs) are experiments used to detect associations between hundreds and thousands of single nucleotide polymorphisms (SNPs) and an exposure/trait (206). They are generally conducted to understand disease aetiology to be able to then prevent or treat a disease (207). GWASs are useful as they provide a non-biased approach in assessing which variants may be causal, therefore they are not influenced by prior scientific knowledge (206). GWASs have been useful in identifying mutations that are found in monogenic diseases thereby informing disease aetiology (208). As well as understanding the consequence of a SNP on an exposure/outcome, SNPs can be used as proxies for an exposure to test for an exposure's association with an outcome of interest. This is the basis of Mendelian randomisation which seeks to assess whether a specific exposure is causally associated with a specific outcome (209). In the context of this study, SNPs associated with aspirin metabolites may help show an association between the metabolites and risk of CRC.

Therefore, the strength in using methods such as GWASs and Mendelian randomisation is that they allow the identification of SNPs associated with an exposure and then causally assessing whether it affects an outcome, thereby informing whether an RCT should be conducted or not. RCTs have previously been carried out on the basis of findings from observational data only to find null associations costing the investigators both time and money. This is discussed in more detail in Chapter 6.1.3.

In much the same sense, epigenome-wide association studies (EWASs) detect associations of changes in DNA methylation between hundreds of thousands of CpG sites and an exposure/trait. In the context of cancer, DNA methylation patterns are clinically relevant for 2 reasons: they may provide biomarkers to diagnose and predict disease outcome and they may also be used as drug targets (210). DNA methylation can be used in a mediation analysis, which is where an exposure may affect DNA methylation which may affect an outcome (211). In the context of this study, it may be that aspirin affects DNA methylation and it is these changes in epigenetics that alter the risk of CRC. Therefore, this helps in assessing whether changes in DNA methylation (by aspirin) is causally associated with CRC helping to identify a mechanism of action of aspirin.

1.8.4 Aspirin metabolism pathway

Aspirin (acetylsalicylic acid) is broken down into many metabolites- a summary of this breakdown is shown in Figure 1-20. Aspirin is the acetylated form of salicylic acid. It has a half-life of ~20 minutes in circulating blood and after ingestion, it is rapidly deacetylated to form salicylic acid which has a half-life of 2-4 hours (212,213).

Initially, aspirin is absorbed through the cell membrane in the stomach, however, most of the drug is absorbed in the small intestine due to the higher pH and larger surface area (214,215). The initial step in the metabolism of aspirin is its' hydrolysis to salicylic acid catalysed by serine esterases which can be found in both the intestinal wall, the liver and erythrocytes (216). After absorption into the blood stream, aspirin can be metabolised to salicylic acid in either the plasma or the erythrocytes through similar but slightly different enzymes. In the plasma, these serine esterases are butyrylcholinesterase (BCHE) and a homodimer of the enzyme platelet- activating factor acetylhydrolase (PAFAH1b2). On the other hand, in erythrocytes it is metabolized by BCHE as well as a heterodimer PAFAH1b2/PAFAH1b3 (217). Expression of BCHE protein and mRNA has also been found in the liver and small intestine indicating aspirin hydrolysis into salicylic acid may occur in these tissues as well (218). The most active compounds from the pathway include acetylsalicylic acid (219) and its breakdown product salicylic acid (220).

Most of salicylic acid is metabolised in the liver (221). Salicylic acid is transported to the liver where it is then metabolised through conjugation with glycine to form salicyluric acid. Around 63% of salicylic acid is converted to salicyluric acid through the enzyme Acyl-CoA Synthetase Medium-Chain Family Member 2B (ACSM2B) (221). Salicyluric acid is the major metabolite of salicylic acid that is excreted via the urine (222). Salicylic acid and salicyluric acid can be broken down further into other metabolites as shown in Figure 1-20. The other enzymes involved in the metabolic pathway include: UDP-glucuronosyltransferase 1-6 (UGT1A6), cytochrome P450 (CYP2C9) (221). Genetic variants within these enzymes as well as *ACSM2B*, *BCHE* and *PAFAH1b2/3* will be used to proxy for levels of the aspirin metabolites. Approximately 90% of aspirin is metabolized to salicylic acid and only 10% of unchanged aspirin is secreted in the urine (221). Salicylic acid and its metabolites are excreted mainly by the kidneys (223).

Whilst BCHE is able to convert acetylsalicylic acid into salicylic acid, other esterases also exist such as paraoxonase-1 (PON1) (224). However, the actual contribution of PON1 to this reaction is not yet clear (217). Nonspecific esterases also hydrolyse aspirin to salicylic acid in the liver, therefore other enzymes than those mentioned may also be involved (221). Whilst UGT1A6 is the main enzyme involved in the conjugation of salicylic acid into salicyl ohenolic or acyl glucuronide, other closely

related enzymes are also able to catalyze the same reaction including: UGT1A1, 1A3, 1A7, 1A8, 1A9, 1A10, 2B4 and 2B7 (225). As well as enzymatic metabolism of aspirin, a small percentage of cells can be hydrolysed into salicylic acid non-enzymatically (226).

Concentrations of aspirin achieved in the plasma were measured using a high-performance liquid chromatography assay. Nagelschmitz et. al (2014) found that oral doses of 100mg, 300mg and 500mg of aspirin achieved peak plasma concentrations of 1.01 mg/L, 3.01 mg/L and 4.84 mg/L, respectively (227) but very little is known in the literature with regards to the concentrations of aspirin that are achieved within tissues and it is possible that higher concentrations are attained in some tissues, including colorectal compared to plasma. With regards to the concentrations used in this thesis, 0.5mM aspirin equates to 2.78 mg/L, 2mM equates to 11.1 mg/L and 4mM equates to 22.2 mg/L. Whilst 0.5mM is a clinically relevant concentration, 2mM and 4mM are higher than oral doses of 500mg but may be clinically relevant when investigating the effect of intravenous aspirin administration.

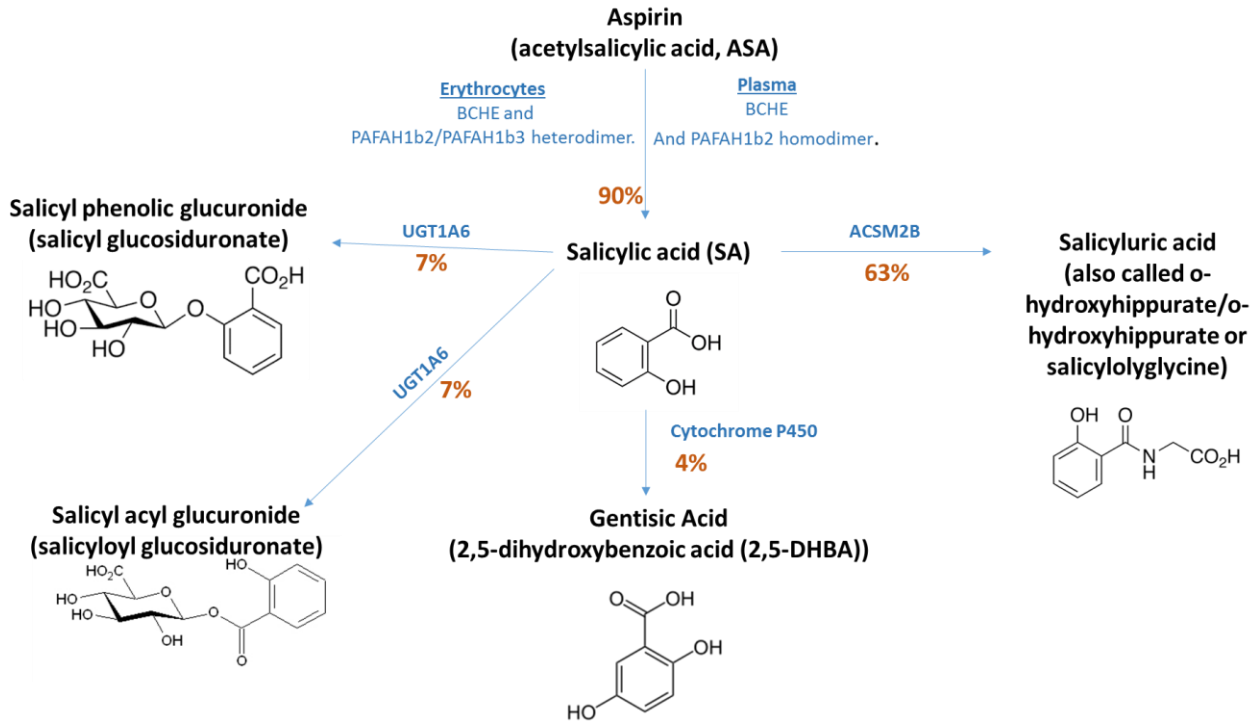


Figure 1-20- Aspirin metabolism pathway

Roughly 10% of aspirin remains unchanged and is excreted in the urine as aspirin. Aspirin is broken down into various metabolites, the most active of them being salicylic acid (220). Various enzymes are involved in the metabolism pathway. The percentages indicate how much of the drug is being metabolised in that pathway. Abbreviations: BCHE, butyrylcholinesterase; PAFAH1b2, platelet- activating factor acetylhydrolase 2; PAFAH1b3, platelet- activating factor acetylhydrolase 3; UGT1A6, UDP-glucuronosyltransferase 1-6 and ACSM2B, Acyl-CoA Synthetase Medium-Chain Family Member 2B.

1.8.5 Mendelian randomisation

Mendelian randomisation (MR) is a method that utilizes Mendel's second law but within the context of an epidemiological setting. This law is the law of random assortment which suggests that traits are inherited independently of each other (199,228). Used in a contemporary context, this refers to the independent inheritance of genotypes, alleles and phenotypes (209). MR seeks to use genetic polymorphisms, most commonly single nucleotide polymorphisms (SNPs), to proxy for modifiable exposures to test for their association with diseases (209).

There are three main assumptions when carrying out an MR. These are:

1- The genetic instrument is associated with the exposure
2- The genetic instrument is independent of any confounders that may affect the relationship between the exposure and outcome
3- The genetic instrument is independent of the outcome given the exposure and confounders. (209)

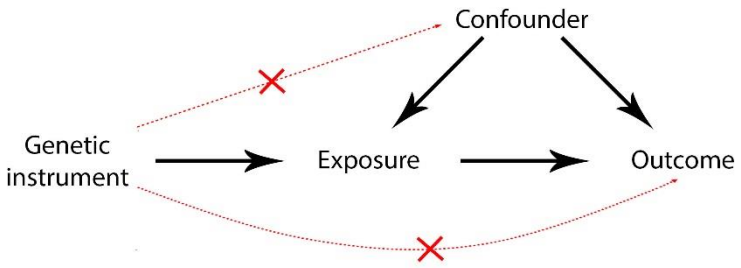


Figure 1-21- Directed acyclic graph (DAG) showing the model used in a MR approach

The genetic instrument is associated with the exposure independently of any confounders associated with the exposure and outcome. The instrument is only affecting the outcome through the exposure. Violations of the main assumptions are highlighted as red arrows.

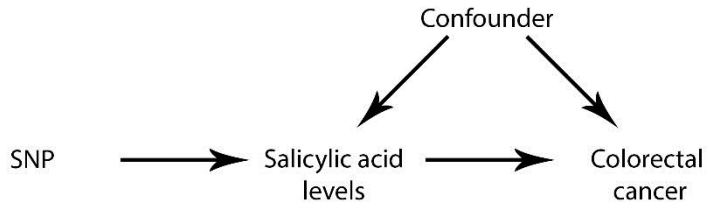


Figure 1-22- Directed acyclic graph (DAG) used in the context of our study using genetic instruments to interrogate the effect of salicylic acid levels on risk of colorectal cancer.

Therefore, genetic instruments can be used to proxy as an exposure and to test for an association with the outcome (Figure 1-21). For example, 11 SNPs strongly associated with selenium levels in the blood predict a 114 μg/L increase in selenium levels. These SNPs are therefore used to "represent" an increase in selenium levels. These SNPs can then be tested for associations with various outcomes such as cancer. If such an association holds true, then the conclusion would be that an increase in selenium levels may be affecting risk of cancer (229). In the context of the current study, single nucleotide polymorphisms (SNP) will be used as genetic instruments to proxy for changes in levels of aspirin metabolites (salicylic acid and salicyluric acid) and to then test for an association with CRC incidence (Figure 1-22). The idea is that salicylic acid is responsible for a large proportion of the anti-cancer activity of aspirin in humans and that plasma levels of this metabolite may correlate with the efficacy of aspirin in humans. Reactions catalysing the formation of other metabolites will reduce the levels of salicylic acid, which would lead to decreased efficacy. Therefore, salicyluric acid plasma concentrations would be expected to inversely correlate with efficacy. MR will be used to proxy for both these metabolites under the hypothesis that SNPs that predict increased levels of salicylic acid also predict a decreased risk of CRC and SNPs that predict an increase in levels of salicyluric acid predict an increased risk of CRC.

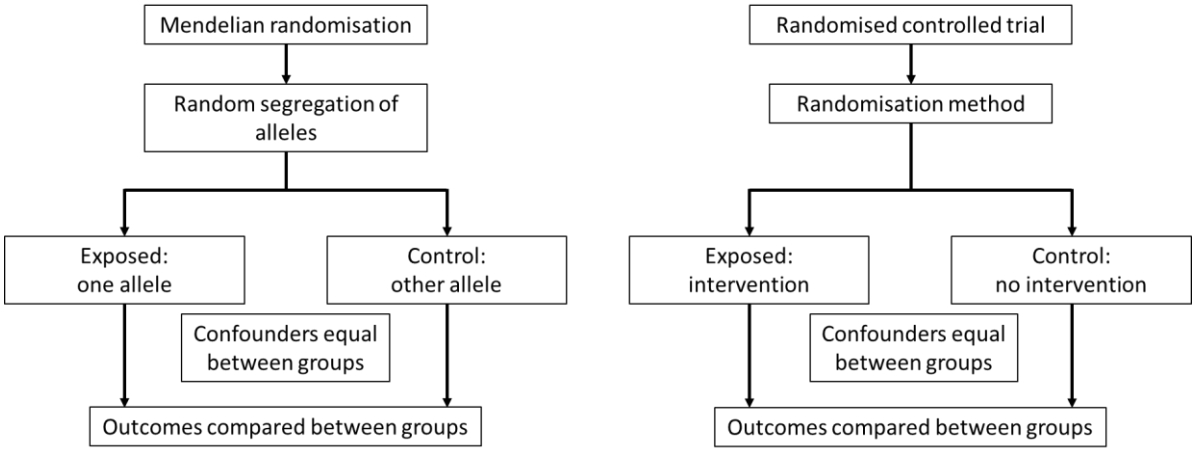


Figure 1-23- A schematic comparing Mendelian randomisation and randomised controlled trials. Adapted from Davey Smith et al, 2005 (230)

Mendelian randomisation has been likened to a randomised controlled trial in that alleles are randomly segregated at conception the same way that an intervention is randomly allocated at the start of a trial (Figure 1-23) (230,231). This lends to many advantages for the use of MR which include:

- Reduced confounding: genetic variants are generally not prone to confounding by factors such as behavioural, physiological and socioeconomic (230).

- Unlikelihood for reverse causation: the onset of disease will not influence the genetic variants and exposure history of individuals with disease will not be differentially reported
- Reduced selection biases: how participants are selected into a study is unlikely to influence the genetic variants
- Reduced attenuation by errors: genetic variants will proxy for exposure level across a lifetime and therefore the associations will not be affected by random imprecise measurements of the exposure

However, this is not to say that MR is without limitations. These are summarised in Table 1-4.

Table 1-4- Limitations that need to be considered when conducting an MR approach

Limitation	Explanation
Genetic instrument	A reliable association between the genetic instrument and exposure needs to be established (230) where enough data needs to be present to establish reliable associations between genotype and trait or phenotype (232). The effect of the genetic instrument on the outcome occurs only through the intermediate phenotype proposed and the effect of the genetic instrument on the outcome is similar to the effect of the intermediate phenotype/environmental exposure on the outcome (232).
Confounding	Linkage disequilibrium can occur where the genetic variant of interest is inherited with other genetic variants that may be the true cause of the disease (230). Another cause of confounding can be from population stratification-i.e. mating non-random (232).
Segregation	Alleles are randomly assigned at conception so segregational distortion at the locus of interest does not occur (232).
Pleiotropy	Genetic variants may affect multiple pathways or exposures leading to the inability to make definitive conclusions (230).
Canalisation	Genetic variants that could potentially be disrupted are dampened or buffered by other processes such genetic redundancy-more than one gene having a similar function (230).
Survival	The genetic variant has not led to a selective survival (232).
Parent-of-origin	There is no parent-of-origin effect in that the phenotypic effect of the genetic variant is not affected by whether it is inherited by the mother or father (232).

1.8.6 Multiple instruments

In genetic epidemiology, genome-wide association studies have allowed the identification of many SNPs associated with various exposures, but all with small effect sizes (206). Therefore to increase statistical power, a genetic risk score can be calculated using multiple SNPs. This genetic risk score represents the additive effect of SNPs in combination to explain more of the variance in the exposure of interest thereby increasing power and avoiding weak instrument bias (233). For

example, the effect of a SNP that is estimated to increase aspirin concentration by 0.02mM on the risk of colorectal cancer may be difficult to detect unless a very large sample size was available. To overcome this, multiple SNPs that predict small changes in aspirin concentration can be combined such that if 8 SNPs predicted an overall increased aspirin concentration of 2mM, the effect of this on colorectal cancer would be more easily detected without needing extremely large sample sizes.

An unweighted method may be used which adds together the effect of various variants but more commonly, a weighted approach is used such as the inverse-variance weighted (IVW) method. The SNPs used to generate this score are generally not in linkage disequilibrium (LD) so are independently inherited, although SNPs that are in low to moderate LD can be combined into a genetic risk score through methods such as the likelihood-based method which is a form of weighted generalised linear regression (234).

One consideration to be made when incorporating many SNPs into the risk score is that any of these SNPs may potentially be pleiotropic- i.e. it is not solely associated with the exposure of interest but also with other risk factors or pathways (horizontal pleiotropy) (233) . This is why it is important to examine the SNP-exposure and SNP-outcome effect estimates, for example, through a scatter plot to identify horizontal pleiotropy. Outliers can be visualised and this will help discern whether the pleiotropic SNP is affecting the slope of association thereby biasing the results (directional pleiotropy). To test for this, a sensitivity analyses termed “leave-one-out” can be adopted whereby one SNP from the variants is left out of the analysis and the tests are carried out again to see whether the SNP dropped is responsible for affecting the exposure-outcome association (235). To detect directional pleiotropy, methods such as MR-Egger regression and weighted median approach may be used (236,237).

An example of detecting pleiotropy through scatter plots and leave-one out analyses is shown in Figure 1-24. An MR approach to look at the effect of caffeine consumption on platelet count was conducted. The scatter plot shows a SNP that is an outlier and appears to be biasing the results of the inverse variance weighted method (light blue line) and the MR egger approach (dark blue line) towards it. The leave-one out analysis shows that the SNP biasing the results is rs2240466 as the removal of this SNP from the analysis produces a more positive association between caffeine consumption and platelet count as well as smaller confidence intervals.

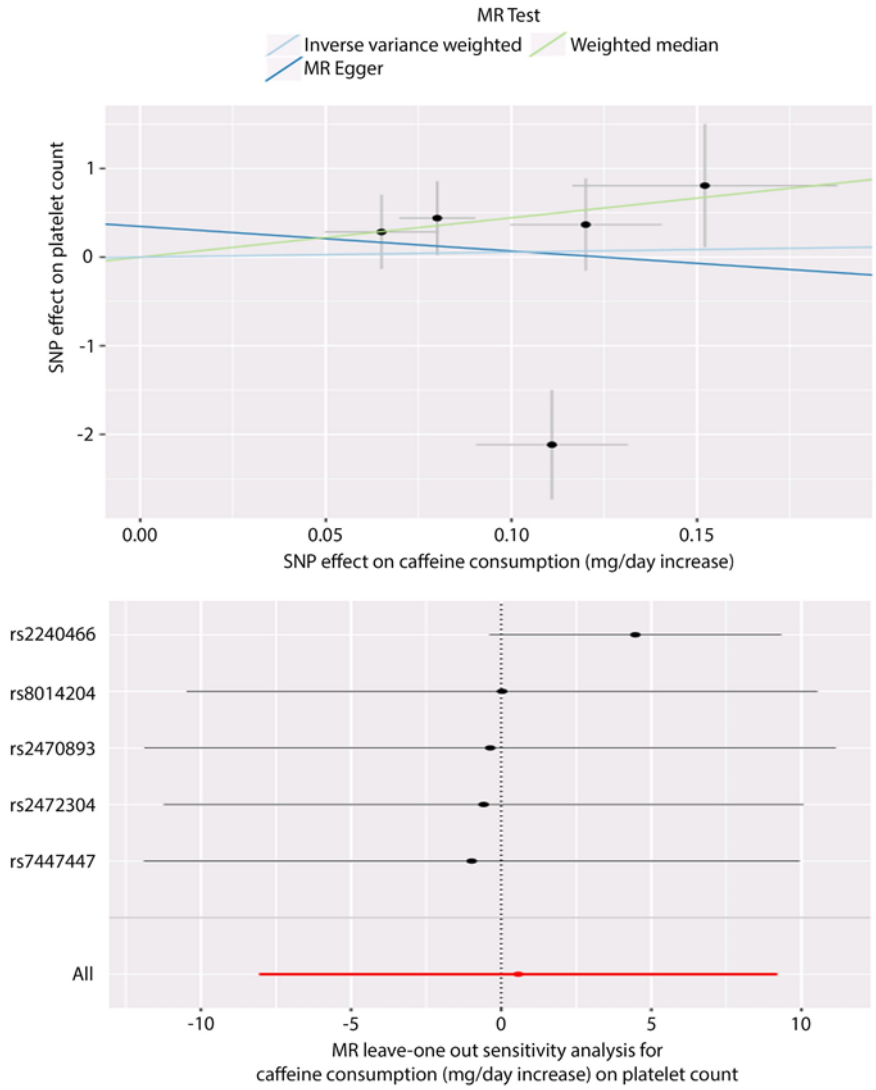


Figure 1-24- Example of identifying pleiotropy through scatter plots and leave-one out analyses

The scatter graph plots the effect of the SNP on caffeine consumption on the x-axis and the effect of the SNP on platelet count on the y-axis. Each dot represents a SNP and the lines through the dot represent the confidence intervals of the SNP effect on caffeine consumption (parallel to the x-axis) and SNP effect on platelet count (parallel to the y-axis). Scatter plot regression lines: light blue, inverse variance weighted method; dark blue, MR egger; light green, weighted median. The leave-one out analysis shows the summary result when each SNP is systematically removed. Each dot represents the combined effect of the SNPs (representing caffeine consumption) on platelet count but excluding the name of the SNP on the y-axis. The lines through the dots represent the confidence intervals. The red line is the overall effect of all the SNPs on the outcome. Both graphs show a SNP (rs2240466) which appears to be an outlier in the scatter graph and the leave-one out analysis thereby biasing the results.

The ideal in an MR study is to use a single instrument with a genetic variant whose function is well-known and measured. However, the only problem with this is that the SNP may have a small effect which is why a genetic risk score combining many SNPs is adopted. However, this may introduce bias as the biological effects of some of the SNPs included may not be known. One approach is to carry out an analysis using a small number of SNPs with known function, and then to expand this in a

secondary analysis with a larger number of SNPs knowing that there may be some biological uncertainty (238).

1.8.7 MR methodology

When there is only one genetic instrument to proxy for an exposure, the method used to quantify the effect of the genetic instrument on the exposure is called the Wald estimator. This is shown in Figure 1-25. In the case of a linear regression of the instrument and exposure, a beta estimate is calculated. When the outcome is binary, a logistic regression is carried out to calculate a log odds ratio (OR). Wald ratios are calculated by dividing the per-allele effect of the SNP (the log OR) on the outcome by the per allele effect of the SNP (beta) on the exposure.

However, it is more common that there is more than one instrument and these need to be regressed to understand the relationship between the exposure and the outcome (209). Many methods currently exist for regressing the exposure on the outcome. The simplest method to use is the two-stage least squares approach. This involves carrying out a least-squares regression of the exposure on the genetic instrument and then carrying out a least-squares regression of the outcome on the value calculated by the first regression (209). This can be either in a one sample setting (where the same sample is used to measure instrument-exposure and instrument-outcome association) or using a two-sample approach where a different sample is used to measure the instrument-exposure and instrument-outcome association (239).

Whilst individual data can be used in two-sample settings, logistical issues (such as data storage) and privacy concerns have made it more common to use summary data instead (240). Two-sample MR methodology has been developed whereby summary level data for gene-exposure data and gene-outcome data from different samples can be used (241). The advantage of using summary data has motivated the development of various methods to test for exposure and outcome association using genetic instruments (240). The most widely used method in a two-sample MR is the inverse variance weighted method (Figure 1-25) but other methods that relax the strict assumptions of MR have also been developed such as maximum likelihood method, mode-based methods and weighted median approach. These are summarised in Table 1-5 .

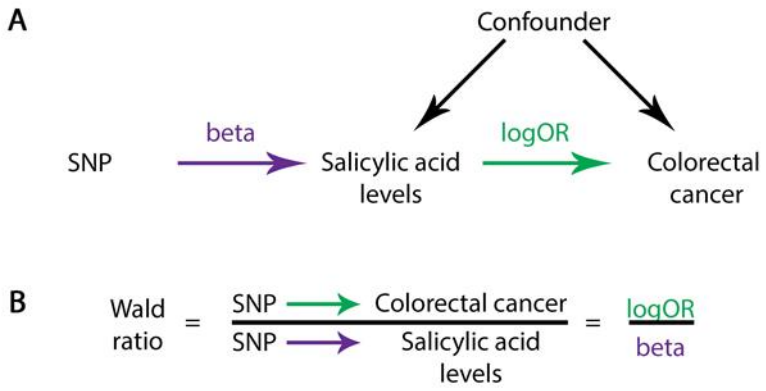


Figure 1-25- Wald ratio and inverse-variance weighted (IVW) method

(A) A directed acyclic graph (DAG) used in the context of our study using genetic instruments to interrogate the effect of salicylic acid levels on risk of colorectal cancer. Linear regressions of the SNP-exposure are conducted to calculate the per allele effect of the SNP on salicylic acid levels (beta) and logistic regressions are carried out to calculate the per allele effect of the SNP on colorectal cancer risk (logOR). (B) Wald ratios are calculated by dividing the per-allele effect of the SNP (the log OR) on the outcome (βZY) by the per allele effect of the SNP (beta) on the exposure(βZX). 95% confidence intervals (95% CIs) are calculated by using the standard errors (SE) of each Wald ratio. These standard errors are calculated by diving the SE of the SNP-outcome association by the SE of the SNP-exposure association. To obtain the inverse variance weighted estimates, the Wald ratios are weighted by the inverse variance. The regression is conditioned to pass through zero implying that when the value of the exposure is 0, the value of the outcome is also 0. The inverse variance weighted method assumes that all the SNPs are valid instrumental variables. To obtain a pooled causal estimate of the exposure-outcome association, individual Wald ratios and 95% CIs are pooled and presented as a forest plot.

Table 1-5- A summary of two-sample MR methods and their advantages and limitations

MR Method	Description	Advantage	Limitation
Inverse variance weighted (IVW)	The Wald ratio estimates the causal effect for different SNPs. The inverse variance weighted estimate of the Wald ratios is combined by a fixed-effects meta-analysis approach (234). The point estimates from this method are akin to a weighted linear regression of the SNP-outcome and SNP-exposure associations (242). The gene-outcome association coefficient is regressed on the gene-exposure association coefficient. The weights are obtained by the inverse variance of the gene-outcome association and the intercept must pass through zero. The slope from this method informs of the effect of the exposure on the outcome (241).	Efficient method when all the genetic proxies are valid instruments (237)	The IVW estimate can be bias even if just one of the genetic instruments is invalid (237)
Maximum likelihood	This method is used to find the linear relationship between the gene-exposure and gene-outcome coefficients (243). The model assumes a that both SNP-exposure and SNP-outcome associations are normally distributed (234). Direct maximisation of the likelihood can be used to estimate the causal effect of the exposure on the outcome, which is assumed to be the same for all SNPs (244).	- Gives appropriately sized confidence intervals when imprecision in the estimate is present unlike the IVW which will provide over-precise causal estimates (243). -Allows for uncertainty in both gene-exposure and gene-outcome coefficients (243). -Useful in overlapping samples as it allows for some correlation between the SNP association with exposure and SNP association with the outcome (234).	The InSIDE assumption is required: strength of the gene-exposure association must not correlate with the strength of bias due to horizontal pleiotropy
MR Egger methods	Just like the IVW approach, the gradient from this method also represents the causal estimate. The only difference is that the intercept is not constrained to 0 (236). This method is used to address the issue of horizontal (directional) pleiotropy: the association of a SNP with multiple traits independently of the exposure being investigated. This violates one of the assumptions of MR (242). Pleiotropy can usually be	-Addresses the issue of horizontal pleiotropy (241) -Can be applied using summary level data (241) -Can provide consistent estimates even when all the genetic instruments are invalid (237)	The Instrument Strength Independent of Direct Effect (InSIDE) assumption is required. The effect of the genetic variants on gene-outcome are independently distributed and not correlated with the gene-exposure associations (236).

	visualised in a funnel plot whereby the MR estimate is plotted against its precision. The presence of symmetry in the plot would indicate the absence of pleiotropy (241). Pleiotropy is also detected through regressing the gene-outcome coefficient on the gene-exposure coefficient but without constraining the intercept to pass through zero (236). The intercept from this test provides an indication of directional pleiotropy- if the gene-exposure association is 0 then so should be the gene-outcome association. Any other result indicates that the effect of the gene on the outcome may not be through the exposure being investigated. The slope from MR-Egger regression adjusts the exposure-outcome association for directional pleiotropy (241).		
MR Egger bootstrap	Bootstrapping is a technique in which you create many smaller datasets from your original dataset to repeat your analysis in. Our samples are only a small representation of the true population and therefore by randomly selecting observations from our sample this will be a more true representation of the original population (245). Therefore, the MR Egger bootstrapping approach is simply carrying out the MR Egger test in multiple smaller samples.	Can be used when assumptions for a normal distribution are not met (245).	Need to have a large sample size (246) to avoid extreme values from the distribution.
Median-based estimator	<u>Simple Median:</u> This is the median of the ratio estimates which provides a consistent causal estimate even when 50% of the genetic instruments are invalid. This method is used to test the violation of all MR assumptions (237). <u>Weighted Median</u> Whilst in the simple approach, 50% of the instruments need to be valid, with the weighted approach, "at least 50% of the weight of the estimator comes from valid IVs". This is beneficial when fewer than 50% of the variants are valid instruments but the weight of the valid instruments contributes to at least 50% of the weighted causal estimate (237).	-Can be applied using summary level data (241) -Provides consistent value of causal effect even when 50% of the genetic instruments are invalid unlike the IVW approach where one invalid SNP can bias the results (237) -Does not suffer from low power like the MR Egger (236)	At least 50% of the genetic instruments need to be valid

Mode-based method	Simple mode: This method involves grouping the SNPs into clusters based on having similar causal effects. The cluster with the most number of SNPs is used to calculate the causal effect (247). Weighted mode: This is similar to the IVW and weighted median approach in that within each cluster, the contribution of the SNP to that cluster is weighted based on the inverse variance of the effect of the SNP on the outcome (248).	If all the SNPs in a cluster are valid instruments then the causal estimate calculated is an unbiased result (247).	If any of the genetic instruments in the cluster are not valid then bias may be introduced (247).

1.9 Thesis overview

In this thesis, epidemiological and laboratory-based approaches will be used to explore causality and potential mechanisms of aspirin on CRC.

A schematic for the aims of each chapter are summarised in Figure 1-26. Based on observational studies that show that timing and duration of aspirin intake have different effects on risk of CRC (Chapter 3.1.1), we aim to assess whether any functional phenotypic differences between short-term and long-term aspirin treatment are present in the context of cell culture (Chapter 3). Aspirin has a well-known effect on protein acetylation (78) and possibly modulates DNA methylation (249). Due to these evidences, we aim to assess whether aspirin's effect on DNA methylation is a potential mechanism for the differences observed in cell growth due to short-term and long-term treatment (Chapter 4, short-term treated cells; Chapter 5, Long-term treated cells). Finally, to causally assess whether aspirin is affecting risk of CRC in the population, we aim to identify SNPs that are associated with aspirin metabolites levels (salicylic acid and salicyluric acid) and to then test the association of these SNPs with risk of CRC incidence (Chapter 6).

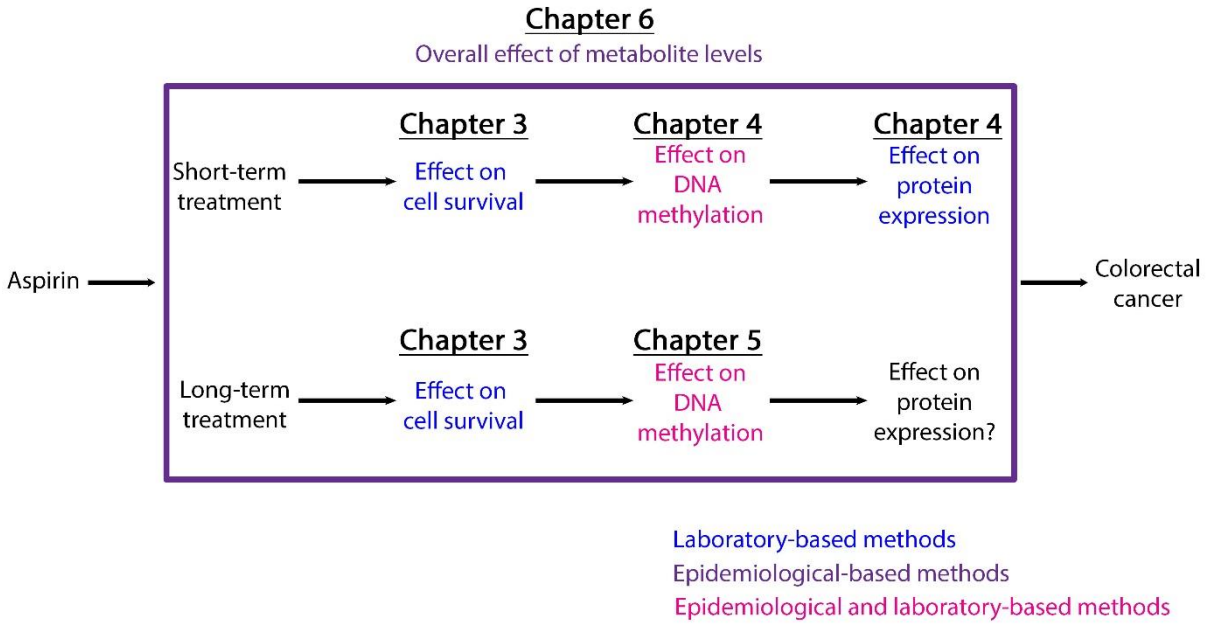


Figure 1-26- Overview of the questions addressed in this thesis to explore causality and potential mechanisms of aspirin on CRC.

Chapter 2 Methods

2.1 Cell culture

2.1.1 Cell lines

2.1.1.1 RG/C2

The cell line used in this project as an *in vitro* model is the human colorectal adenoma cell line S/RG/C2 (referred to as RG/C2 henceforth whereby the prefix "S" denotes that they are from a sporadic tumour) and their origin is parental S/RG cells. This cell line was isolated from a colonic adenoma from a 59-year old female patient with no previous history of colorectal cancer. They were established in this laboratory as being anchorage-dependent cells that are non-tumorigenic when injected into athymic mice (250). RG/C2 cells express WT full length *APC* (251) as well as wild type *KRAS* and *PIK3CA* (252) but express mutant *TP53* (253). These cells also express a one mutant β-catenin allele (Greenhough, University of Bristol, personal communication).

2.1.1.2 LS174T

This cell line was obtained from American Type Culture Collection (ATCC, Rockville, MD, USA). It was obtained from a 58 year old Caucasian female patient with a Duke's type B adenocarcinoma (254). These cells express the wild type forms of APC and p53 proteins (255,256). These cells express mutant β-catenin that are degradation-resistant and a mutated and activated K-RAS (257,258).

2.1.1.3 SW620

This cell line was obtained from American Type Culture Collection (ATCC, Rockville, MD, USA). These cells are obtained from a 51-year-old Caucasian male with a Duke's type B adenocarcinoma. This cell line is derived from a lymph node secondary tumour after the cells metastasised (254). Mutations in this cell line are present in K-RAS, APC and p53 proteins (259).

2.1.2 Conditions for cell maintenance

RG/C2s were cultured in Dulbecco's Modified Eagles Medium (DMEM) (Life Technologies, Paisley, UK) and was supplemented with 20% foetal bovine serum (FBS) (Life Technologies, Paisley, UK), L-glutamine (2mM) (Life Technologies), penicillin (100 units/ml) (Life Technologies), streptomycin (100 ug/ml) (Life Technologies) and insulin (0.2 units/ml) (Sigma, Poole, UK). LS174Ts and SW620s were cultured with 10% FBS (Life Technologies),), L-glutamine (2mM) (Life Technologies), penicillin (100 units/ml) (Life Technologies) and streptomycin (100 ug/ml) (Life Technologies).

Cells were maintained in T25 (25cm^2) flasks and were medium changed every 3-4 days under sterile conditions. They were incubated with 5% CO_2 at 37^{o}C in a non-humidified incubator.

2.1.3 Cell passaging

When cells reached confluence, they were passaged by firstly washing with 1x phosphate buffered saline (PBS pH 7.4; Severn Biotech Ltd, Kidderminster, UK) and then incubated at 37°C for 5-10 minutes with 0.1% trypsin (w/v) (BD Bioscience, Oxford, UK) and 0.1% (w/v) ethylenediaminetetraacetic acid ((EDTA) Sigma). To neutralize the trypsin, fresh medium was added. The suspension was transferred to a universal tube and centrifuged at 3000rpm for a duration of 3 minutes. After centrifugation, the supernatant was aspirated, and fresh medium was added to resuspend the cells. To ensure that the cell clump was resuspended as single cells, cells were repeatedly syringed with a 1.2mm blunt needle. For experiment set up, cells were counted and seeded into T25 flasks or 24 or 48 well plates (Corning, New York, USA) at appropriate densities.

2.1.4 Cell storage and recovery in liquid nitrogen

Cells were passaged within a limit of ~10 passages to maintain cell line characteristics and avoid molecular changes that occur after multiple passages. Therefore, stocks for the cell line were frozen and stored in liquid nitrogen for future use.

Firstly, cells were trypsinized, pelleted and resuspended in growth medium that contains 10% dimethyl sulfoxide ((DMSO) VWR International. Lutterworth, UK) which is a cryoprotectant. Cells were then aliquoted into 1ml cryovials (Thermo Scientific, Waltham, Massachusetts, USA), placed into a cryo-1°C-freezing chamber (Thermo Scientific) and slowly frozen down at a cooling rate of -1°C per minute at -70°C. Once cells had reached -70°C, the cryovials were stored in liquid nitrogen (-196°C) for storage.

To retrieve fresh stocks of cells, cryovials were thawed in an incubator at 37°C and resuspended in normal growth medium. Cells were then subject to centrifugation at 3000rpm for 3 minutes to achieve pellets. The medium containing DMSO was then aspirated and cells were resuspended in fresh growth medium upon which they were seeded into T25 flasks and incubated at 37°C.

2.1.5 Cell counting

It was shown in 1993 that floating cells found in the growth medium of flasks are apoptotic and so can be used to count the number of apoptotic cells (260). To do this, the medium that cells were cultured in was transferred to a universal tube whereby 7.5µl of this medium was loaded onto a Neubauer counting chamber to count the apoptotic cells. To count adherent (live) cells, the same procedure as cell passaging was followed up to the resuspension of the cell clump. Cells were resuspended in 10ml of fresh growth medium and syringed to break the clumps onto single cells. Depending on the cell density, cells were further diluted to a ratio of 1:10 (50µl cell suspension with 450µl of PBS or fresh growth medium) and then counted on the Neubauer counting chamber.

2.1.6 Seeding

Experiments were seeded at the same density per flask. Cell counting was carried out as previously described and the cells were seeded into new T25 flasks according to the desired cell density and then fresh growth medium was added to make up 4mL per flask.

2.2 Cell Treatments in 2D

2.2.1 Acetylsalicylic acid (Aspirin)

Aspirin (Sigma) was dissolved in either ethanol or fresh growth medium for use in experiments. Aspirin was dissolved in fresh growth medium at 12mM by dissolving 0.0432g of aspirin in 20mL of fresh growth medium. For growth as spheroids, 12mM of aspirin was dissolved in advanced DMEM (Life Technologies). For 2D cell culture, flasks were treated with various concentrations of aspirin dissolved in fresh growth medium as shown in Table 2-1. For short term experiments, cells were treated for 24 hours before being lysed. Alongside the short-term aspirin treatments, RG/C2 cells were seeded in various aspirin doses and have been continually treated with aspirin since 29.10.2015 and these are termed the long-term aspirin treated cells.

Table 2-1- Aspirin concentrations used in 2D cell culture

Aspirin concentration (mM)	Total volume of 12mL		Equivalent aspirin concentration (mg/L)
	Volume of normal growth medium (mL)	Volume of aspirin medium (mL)	
0	12	0	0
0.5	11.5	0.5	2.78
2	10	2	11.1
4	8	4	22.2

Nagelschmitz et. al (2014) measured plasma conocentrations of aspirin and found that oral doses of 100mg, 300mg and 500mg of aspirin achieved peak plasma concentrations of 1.01 mg/L, 3.01 mg/L and 4.84 mg/L, respectively. They also showed that intravenous administration of 250mg and 500mg of aspirin achieved peak plasma concentrations of 29.62 mg/L and 54.25 mg/L, respectively (227). With regards to the concentrations used in this thesis, 0.5mM aspirin equates to 2.78 mg/L, 2mM equates to 11.1 mg/L and 4mM equates to 22.2 mg/L (Table 2-1). Whilst 0.5mM is a clinically relevant concentration, 2mM and 4mM are higher than oral doses of 500mg but may be clinically relevant when investigating the effect of intravenous aspirin administration.

2.2.2 N-(2-Quinolyl)valyl-aspartyl-(2,6-difluorophenoxy)methyl ketone (QVD)

QVD is an irreversible, broad spectrum caspase inhibitor that is able to permeate the cell therefore inhibiting all major caspase-mediated apoptosis pathways. QVD was stored as a liquid at 10mM concentration at -20^{o}C. Cells were treated with 10µM QVD (Millipore, Billerica, Massachusetts, USA) for 3 hours before the medium

was changed with fresh medium treated with 4mM aspirin and QVD for 72 hours. Cells were treated in the dark as QVD is light-sensitive.

2.2.3 Vehicle controls

The use of dimethyl sulfoxide (DMSO) was used as a vehicle for QVD therefore this was controlled for in experiments by the addition of DMSO to all cells to ensure treatment conditions contained equal concentrations of DMSO. The final concentration of DMSO was never more than 0.2%.

2.3 Transfections

Cells were medium changed 24 hours before transfection to 10% or 20% DMEM (depending on cell line growth properties) free of penicillin or streptomycin. Trypsin was added to these sub-confluent cells which were counted and seeded at a concentration of 1-1.5 million cells per flask (depending on the growth properties of the cell line) in T12.5 flasks (Corning). To increase the efficiency of transfection and rate of transfection, cells were seeded in 10% or 20% DMEM free of penicillin and streptomycin (P/S-free). Non-targeting (control) siRNA and gene-specific siRNA was added as well as 5µL of RNAiMAX (Life Technologies) in 500µL of Opti-MEM (Life Technologies), as stated in the manufacturer's instructions, to achieve a final concentration of 25nM. Cells were incubated for 16 hours at 37°C before the medium containing the siRNA complexes was aspirated and 10% or 20% P/S free medium (depending on cell line growth properties) added in its place. Cells were lysed at time points of interest relative to the time point of siRNA addition. Details about the siRNA sequence used in this thesis are shown in Table 2-2.

Table 2-2- SiRNA sequence

Target	Supplier	Catalogue number	Sequence (5' - 3')
STMN1	Santa Cruz	sc-36127	Sequence A CCAGAUCCAGACUGUAAGAtt Sequence B GGGAGAAACUGAAAGUGUUtt Sequence C CCUCCUGGUUGAUACUUGUtt

2.4 Spheroids

2.4.1 Cell culture

Spheroids were cultured in Advanced DMEM/F12 with reduced FBS supplementation ((ADF) Life Technologies). The base medium contained 0.1% BSA, 2mM L-glutamine, 10mM Hepes, 100µg/ml streptomycin and 100 units/ml penicillin. Fifty millilitre aliquots were transferred to falcon tubes and stored at -20°C. This medium was further supplemented with these cell nutrients: N2 ((1:100) Life Technologies), B27 ((1:50) Life Technologies) and N-acetyl-cysteine ((1:500) Sigma). The medium was supplemented with further factors as shown in Table 2-3.

Table 2-3- Further supplements added to medium for spheroid culture (261)

Reagent/Growth factor	Concentration	Supplier
hEGF	50ng/mL	PeproTech, Rocky Hill, USA
hNoggin	100ng/mL	RD Systems, Minneapolis, USA
hGastrin	10nM	Sigma, Poole, UK
Nicotinamide	10mM	Sigma, Poole, UK

RG/C2 cells were counted using the protocol described previously but cells were resuspended in PBS. After counting, a cell suspension volume was calculated in order to seed 400 cells per well and using pre-frozen pipette tips, the cell suspension was resuspended in the appropriate volume of Matrigel© (BD Biosciences) to give 400 cells per 20 µL of Matrigel©/well. Matrigel© will set at room temperature therefore it was kept on ice during this process. Forty-eight well plates were incubated at 37°C before use and then 20µL of Matrigel© was aliquoted as a blob into the inner 24 wells. Plates were incubated at 37°C for a time period between 10-30 minutes to allow the Matrigel© to set. To keep the moisture inside the plate, 500µL of sterile PBS was pipetted into the outer wells of the 48 well plate. Matrigel blobs were cultured in 200µL of appropriate medium and this was changed every 4 days. Plates were placed in a sandwich box with a moist tissue-which was changed every 2-4 days- to maintain humidity levels. Spheroids were then monitored and imaged at days 4, 7, 11, 14 and 21 using a widefield microscope in the Wolfson Bioimaging Facility using Leica Application Suite Advanced Fluorescence Lite 2.6.0

Each well was imaged three times firstly at magnifications of 20X, 10X and 5X depending on the size of the spheroids. The images were then run through MATLAB R2015a to automatically calculate the area of the spheroids in each image. The median of the areas were then plotted using GraphPad Prism 5.

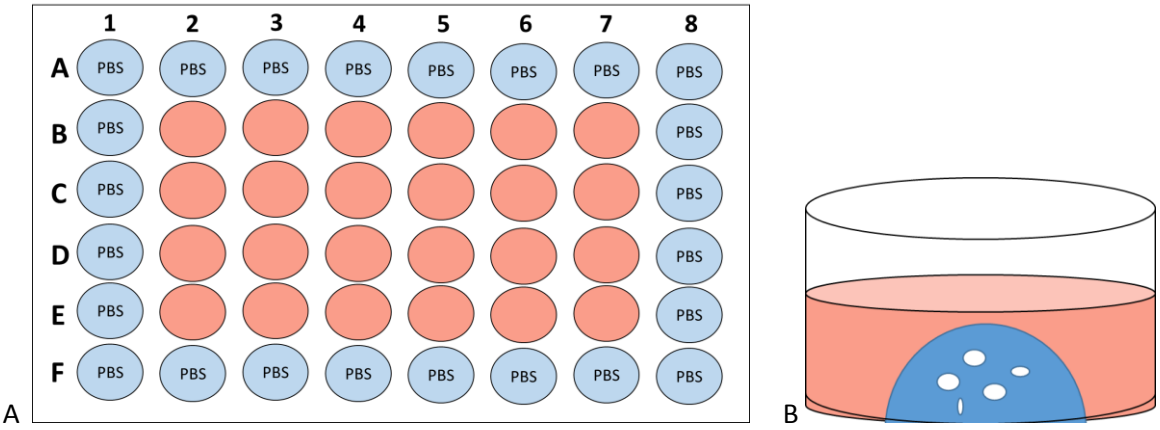


Figure 2-1- Diagram of plate organization for culturing of spheroids

(A) The wells along the outside of the 48-well plate were filled with PBS to retain moisture and prevent medium evaporation whilst the 24 wells inside the plate were used for spheroid growth. (B) This is a diagram of how a Matrigel© blob (blue) with suspended RG/C2 cells (white) looks inside a well filled with growth medium (orange).

2.5 Cell Treatment in 3D

2.5.1 Acetylsalicylic acid

Aspirin was dissolved in fresh growth medium at 12mM by dissolving 0.0432g of aspirin in 20mL ADF. For 3D cell culture, 24 or 48 well plates (Corning) were treated with various concentrations of aspirin dissolved in advanced DMEM (Table 2-4).

Table 2-4- Concentrations of aspirin dissolved in Advanced DMEM used in 3D cell culture

Aspirin concentration (mM)	Total volume of 2000µL		
	Volume of supplemented medium (µL)	Volume of Advanced DMEM medium (µL)	Volume of aspirin medium (µL)
0	1000	1000	0
0.5	1000	83.33	916.67
2	1000	166.67	833.33
4	1000	333.33	666.67
6	1000	666.66	333.33

2.6 Apoptosis and cell cycle assays

2.6.1 Measurement of attached/floating yield number and apoptosis

After treatment of RG/C2 cells in 2D experiments, floating cells were collected and attached cells harvested by trypsinisation. Attached and floating cells were separately counted and floating cells were calculated as a percentage of the total cell population. Previously, it has been shown that >90% of floating cells present apoptotic morphology whereas <3% of attached cells exhibit this morphology (260).

2.6.2 Fluorescence-activated cell sorting (FACS)

RG/C2 cells were treated with 0, 2 and 4mM aspirin for 24 hours. Floating and attached cells were collected and washed with PBS. Cells were fixed with 70% ethanol while vortexing and incubated for at least 1 hour at -20^{o}C. Cells were then spun down, washed with PBS and syringed at least 3 times to dissolve cell clumps. A concentration of 1.5 x10^{6} cells was taken from each experiment and stained with. 1,5-bis{[2-(di-methylamino) ethyl]amino}-4, 8-dihydroxyanthracene-9,10-dione (DRAQ5)(eBioscience, San Diego, CA) at a concentration of 10µM for 30 minutes. DRAQ5 is light sensitive so cells were stained and maintained in the dark. Data was analysed by flow cytometry using the Novocyte 3000 (Acea Biosciences, San Diego, USA).

Data was collected using the Software NovoExpress (Acea Biosciences) and manually analysed using FlowJo (version 7.6.5)

2.6.3 Assessment of apoptosis in 3D cultures using fluorescent dyes

To assess whether aspirin is causing apoptosis, RG/C2s were grown as spheroids for 14 days before being treated with aspirin for 72 hours and stained with calcein AM and ethidium homodimer 1. Calcein AM is a permanent cell dye that is taken up by live cells and undergoes acetoxymethyl ester hydrolysis by intracellular esterases to convert the compound into green fluorescent calcein. Ethidium homodimer is unable to permeate the cell and is weakly fluorescent until bound to DNA whereby it emits red fluorescence.

Spheroids were stained with 10µM calcein AM (ThermoFisher, Newport, UK) and 5µM ethidium homodimer (ThermoFisher) to each well. These compounds are light sensitive therefore the stains were added in the dark and the plate was covered with foil for 30 minutes before being imaged using a confocal microscope.

2.7 Western Blotting

Western blotting is a technique used to detect specific proteins from any given cells/tissues. Firstly, proteins are separated using a technique called sodium dodecyl polyacrylamide gel electrophoresis (SDS-PAGE). The proteins migrate through the gel according to their relative molecular weight. Then proteins are transferred onto a membrane whereby they are incubated with a primary and secondary antibody that is specific for the protein of interest so that the protein can be visualised.

2.7.1 Preparation of cell lysate

The cell lysis procedure is always carried out on ice. Firstly, the medium was aspirated before cells were washed twice with ice-cold PBS. Cells were lysed in 1X cell lysis buffer (Table 2-6) to obtain whole cell extracts. They were incubated with the buffer for 10 minutes before cell lysate was collected using a scraper and transferred to an Eppendorf tube. Lysates were centrifuged at 18500rpm for 10 minutes and then the supernatant was transferred to new Eppendorf tubes and stored at -70^{o}C. The protein was quantified for each sample using the Bio-Rad DC protein assay kit according to the manufacturer's instructions (Bio-Rad, Hemel Hampstead, UK) following which 40µg-100µg of sample was prepared. Laemmli buffer (Table 2-6) was added to the cell lysate and the sample was boiled for 5 minutes at 100^{o}C. These samples were then stored at -20^{o}C.

2.7.2 SDS Page

Western gels with various acrylamide percentages depending on protein size were run using Mini-Protean 3 Electrophoresis apparatus (Bio-Rad). Equal volumes of samples were loaded into the separate wells as well as a Precision Plus Protein Kaleidoscope marker (Bio-Rad). Gels were initially run in running buffer (Table 2-6) at 100V until the samples had passed through the stacking gel upon which they were run at 180-200V for 1 hour.

Table 2-5- Polyacrylamide gels

Resolving gel	Supplier	9%	12.5%	Stacking gel (4.5%)
Resolving Buffer 1.5M Tris pH8.8/0.4% SDS	National Diagnostics, Hessle, UK	4.4mL	4.4mL	-
Stack Buffer 0.5M Tris pH6.8/0.4% SDS	National Diagnostics, Hessle, UK	-	-	2mL
30% acrylamide/1%bis	National Diagnostics, Hessle, UK	5.2mL	7.3mL	1.1mL
ddH_2O	Severn Biotech Ltd, Kidderminster, UK	7.9 mL	5.8mL	4.4mL
Ammonium persulphate (0.5g/ml in distilled water)	Sigma, Poole, UK	110µL	110µL	57µL
TEMED	BioRad, Hemel Hampstead, UK	3.6µL	3.6µL	1.8µL

Table 2-6- Buffers in western blotting

Solution/Buffer	Composition
Cell Lysis Buffer	1 tablet of Protease Inhibitor Cocktail (Roche Diagnostics, Germany) per 10mL of lysis buffer (New England Biolabs, Ipswich, Massachusetts, USA)
5X Laemmli Buffer	62mM Tris-HCL pH 6.8, 10% (v/v) glycerol, 5% (v/v) 2-mercaptoethanol, 0.01% bromophenol blue
Running buffer	14.4g glycine (National Diagnostics) and 3g Tris (National Diagnostics) and 1g SDS (Fisher Scientific, Loughborough, UK) dissolved in 1 litre H_2O
Transfer buffer	14.4g glycine (National Diagnostics) and 3g Tris (National Diagnostics) dissolved in 800mL H_2O plus 200mL methanol
10 X TBS stock solution	1 pack of 20x TBS powder sachet (Severn Biotech Ltd) with 2 litres H_2O.
Milk blocking buffer	50mL of 1X TBS diluted in water with 2.5g powdered milk
TBST	1X TBS with 1% (w/v) tween (Sigma, Poole, UK)
Stripping buffer	1% SDS (w/v), 25mM glycine in H_2O, pH adjusted to 2 using hydrochloric acid

2.7.3 Wet transfer

After gel electrophoresis, gels were soaked in transfer buffer (Table 2-6) to remove any residual SDS. The apparatus was assembled in a transfer cassette (Figure 2-2) and the proteins were transferred on an Immobilon polyvinylidine difluoride (PVDF) membrane (Millipore) using a Transblot Cell (Bio-Rad) at a voltage of 125V for 1.5 hours in transfer buffer (Table 2-6).

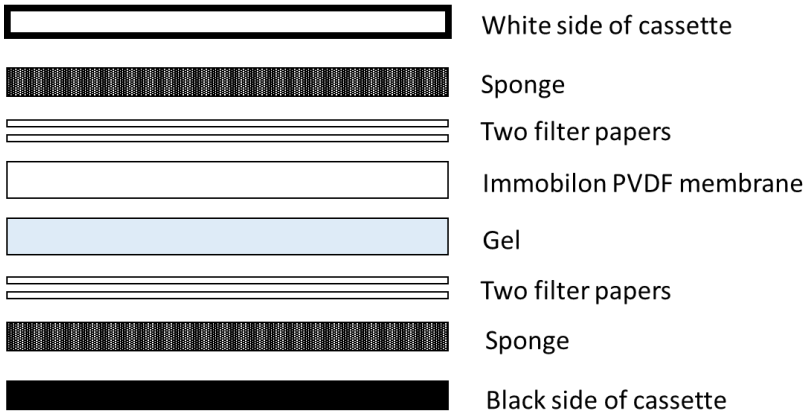


Figure 2-2- Diagram of assembly of apparatus for wet transfer. A sponge and two filter papers were placed on either side of the Immobilon PVDF membrane and gel.

2.7.4 Protein detection

After transfer, the Immobilon PVDF membranes were blocked in milk blocking buffer (Table 2-6) for one hour. They were then incubated at 4^{o}C with the primary antibody diluted in 5mL milk blocking buffer in a sealed plastic bag and left to rotate overnight. The antibodies and their concentrations are listed in Table 2-7.

After incubation with the primary antibody, membranes were washed three times for 10 minutes in TBS-T (Table 2-6). Following this, they were then incubated with horseradish peroxidase conjugated secondary antibodies (dilution 1:1000) in milk buffer (Table 2-6) for 1 hour at room temperature. Membranes were then washed for a further three times at 10 minutes each in TBS-T to remove any unbound secondary antibody. Visualisation of the protein bands was done using LumiGLO chemiluminescent substrate (KPL, Gaithersburg, Maryland, USA). Blots were incubated with chemiluminescence chemicals for 1 minute before being transferred to a cassette with one to three films. The exposure time varied from 1 second to 20 minutes depending on the strength of the antibody signal, and then films were developed using a Compact X4 Film Processor (XO-graph Imaging Systems Ltd., Stone House, UK). Subsequently, films were scanned onto a computer using Epson Perfection V750 Pro and the images were processed using Adobe Photoshop CS6 (64 Bit).

Table 2-7- Western blotting- primary and secondary antibodies

Target	Species	Source	Cat. No.	Dilution
Cleaved PARP	Rabbit	Abcam (Bristol, UK)	Ab32064	1:10000
Cleaved Caspase-3	Rabbit	Cell Signalling (Danvers, Massachusetts, USA)	9664S	1:1000
Tubulin	Mouse	Sigma (Poole, UK)	T9026	1:10000
STMN1	Rabbit	Cell Signalling (Danvers, Massachusetts, USA)	3352S	1:500
Mouse-IgG Peroxidase	Goat	Sigma (Poole, UK)	A4416	1:1000
Rabbit-IgG Peroxidase	Goat	Sigma (Poole, UK)	A6154	1:1000

2.7.5 Stripping western blots and re-probing

Membranes were stripped of previous antibodies if more than one protein was going to be examined. Membranes were incubated at room temperature in stripping buffer (Table 2-6) for 30-60 minutes to remove previous antibodies. Following this, they were incubated in TBS-T twice for 10 minutes and blocked in milk blocking buffer (Table 2-6) for 60 minutes. Membranes were then probed with alternative primary and secondary antibodies as previously described in Chapter 2.7.4.

2.8 Gene expression analysis

2.8.1 RNA sample collection

To assess the levels of mRNA expression, RNA was collected from RG/C2 cells and LS174T cells in response to short term aspirin treatment (24 hours). After cells were treated, they were then washed with PBS and then 1mL of TRI-reagent was added to each flask. After this, cells were scraped and transferred to a sterile and RNAse-free eppendorf. If not used immediately, samples were stored at -80^{o}C.

2.8.2 RNA isolation

To separate the different phases, 0.2mL of chloroform was added to each sample. The eppendorfs were then shaken for 15 seconds before being incubated at room temperature for 2-3 minutes. The samples were then centrifuged at 12,000 *g* at 4^{o}C for 15 minutes. The results in the separation of the sample into 3 separate layers- protein, DNA and RNA. The top layer containing aqueous RNA was transferred to another Eppendorf and 0.5mL of 100% isopropanol was added. Samples were then mixed and incubated at room temperature for 10 minutes so that the RNA precipitates. To pellet the RNA, the samples were centrifuged at 12,000 *g* for 10 minutes at 4^{o}C. The supernatant was removed and 1mL of 75% ethanol was added, after which the samples were vortexed briefly and then centrifuged at 7500 *g* for 5 minutes at 4^{o}C. Following this, the supernatant was discarded and the RNA pellet was left to air dry for 10 minutes. RNA was then resuspended

in 50µL of RNase-free water and incubated in a heat block set at 60°C for 10 minutes. Unless used immediately, samples were frozen at -80°C.

2.8.3 RNA purification and production of cDNA

For RNA purification, a polymerase chain reaction (PCR) hood which contains a ultraviolet (UV) lamp was used. The UV lamp sterilises the workspace by destroying potential contaminants, in particular, DNA. The hood was sterilised with UV light for 30 minutes before any handling of RNA. Before proceeding onto the production of cDNA, samples were cleaned using the QIAGEN RNeasy clean kit (Qiagen, Manchester, UK) as per the manufacturer's instructions.

The concentration and purity of RNA in the samples was quantified using a Nanodrop (Thermo Scientific, Karlsruhe, Germany) and ND-100 software. RNA samples were considered pure if they had an $A_{260/280}$ ratio above 2. For each sample, 2µg of RNA and 1µL of oligo dT primer (Promega, Madison, Wisconsin, USA) were added together to a pair of sterile RNase-free PCR tubes. These samples were then placed in the PCR machine for 5 minutes at 70°C to melt the secondary structure within the template. To prevent the structure from reforming, samples were immediately placed on ice. To each annealing template, 5µL of M0MLV 5x reaction buffer (Promega), 5µL dNTP (Promega) and 0.63µL of Recombinant RNasin Ribonculease Inhibitor (Promega) were added in the order as stated. When a large number of samples was being prepared, a master mix was made up. One PCR tube per pair was used as a control by adding 1µL of ddH_2O whereas the other sample had 1µL of reverse transcriptase, M-MLV RT (Promega), added. The samples were made up to 25µL through the addition of ddH_2O. The tubes were gently mixed before being incubated at 40°C for 60 minutes. After this, cDNA was diluted to a concentration of 0.01µg/µL through the addition of 175µL of ddH_2O to each sample. Unless used immediately, cDNA samples were stored at -20°C.

2.8.4 Quantitative real time polymerase chain reaction (qPCR)

To determine the level of mRNA expression for a particular gene, cDNAs generated from the RNA samples were used for qPCR. Primers were obtained from (Qiagen) and Quantifast SYBR green (QIAGEN) containing dNTPs, SYBR green I dye, ROX dye and hot start Taq polymerase was used. A master mix was made containing Quantifast SYBR green, ddH_2O and primer sets for the gene of interest. In a 96 well polypropylene plate (Agilent Technologies, West Lothian, UK), 10.5µL of the master mix and 2µL of the cDNA was added. For each sample, 3 replicates were carried out and to control for contaminating DNA, each sample had a no RT control. To consolidate the samples, the plate was spun and the qPCR plate was then analysed using a MxPro 3005P Real-time Thermal Cycler (Agilent Technologies) and fluorescence data was generated using the thermal profile shown in Table 2-8, after which the results were viewed on the M-X-Pro software (Stratagene, Stockport, UK). The primers used in this study are shown in Table 2-9. As a reference dye, ROX was used and samples were normalised to a housekeeping gene. Error bars represent the standard deviation of the qPCR data.

Table 2-8- Cycling conditions for qPCR

Step	Temperature (°C)	Time	Number of cycles
Activation	95	10 minutes	1
Denaturation	94	15 seconds	40
Annealing and extension (fluorescence measured at the end of each cycle)	60	1 minute	

Table 2-9- Primers for qPCR

Gene	Manufacturer	Cat. Number/sequence
MCM2	Qiagen, Manchester, UK	QT00070812
MCM5	Qiagen, Manchester, UK	QT00084000
POLD2	Qiagen, Manchester, UK	QT01668912
STMN1	Qiagen, Manchester, UK	QT00039417
TK1	Qiagen, Manchester, UK	QT00083874

To confirm efficient target detection and amplification, standard curves were produced. To confirm that primers were detecting and amplifying a single product, dissociation curves were analysed.

2.9 Migration assays

2.9.1 Cell seeding

Cells were detached from T25 flasks using 0.1% trypsin, centrifuged as previously mentioned in Chapter 2.1.3 to produce a cell pellet and resuspended in FCS free culture medium. Cells were counted and 1.5ml of cell solution was seeded on to the membrane of a 6-well plate transwell (Corning) at a concentration of $1x10^6$/ml. Cells were then incubated for 10 minutes at 37 °C and 5% CO_2 to allow them to settle. Into the bottom of the well, 2.6ml of chemoattractant (10% FCS medium) was carefully added to avoid creating bubbles to produce a chemotactic gradient. Cells were left for 24 hours.

2.9.2 Cell staining

The medium was aspirated from the top of the transwell and any remaining cell and media were carefully removed using a cotton bud. To wash the transwells, they were then inserted into another 6 well plate containing 3mL of PBS. After this, the transwells were transferred to another 6 well plate containing 3mL of 70% ethanol for 10 minutes to fix the cells. The transwells were then removed from the ethanol and allowed to dry.

To stain the cells, 0.5% crystal violet solution (Sigma) was used. The transwells were inserted into a 6 well plate containing 3mL of crystal violet solution and left for 10 minutes. After this, the crystal violet was washed off by dipping the transwells into water carefully so as to avoid washing off the fixed cells. The transwells were then left to dry.

2.9.3 Cell quantification

The number of stained cells that had migrated through the pore were counted at a magnification of x20 at 10 random sites. The number of cells was then averaged for each well. Three images were taken per well using a DS-5 Nikon camera using the software NIS Elements software. Relative change in cell numbers was calculated between 0mM aspirin treated cells and 4mM treated cells as well as control cells and *STMN1* siRNA treated cells.

2.10 Methylation analysis

A short overview of the methodology for methylation analysis is depicted in Figure 2-3. (1) After DNA is extracted from the cells, it is bisulphite converted so that all unmethylated C residues are converted to T bases instead and methylated C residues remain. (2) After bisulphite conversion, samples are loaded onto the array. This example shows the MethylationEPIC BeadChip (850K array) Infinium (210) array. (3) This array contains wells with beads that have 50 base pair probes that are complementary to specific sequences of bisulphite converted DNA (262). Two types of probes are present in this array and these are shown in Figure 2-4. (4) After hybridisation, a fluorescently labelled dNTP is incorporated to the 3' end of the probe to identify whether the C remains (after bisulphite conversion) or whether it has been converted to a T base (263). The fluorescent signal is then measured using an Illumina iScan.

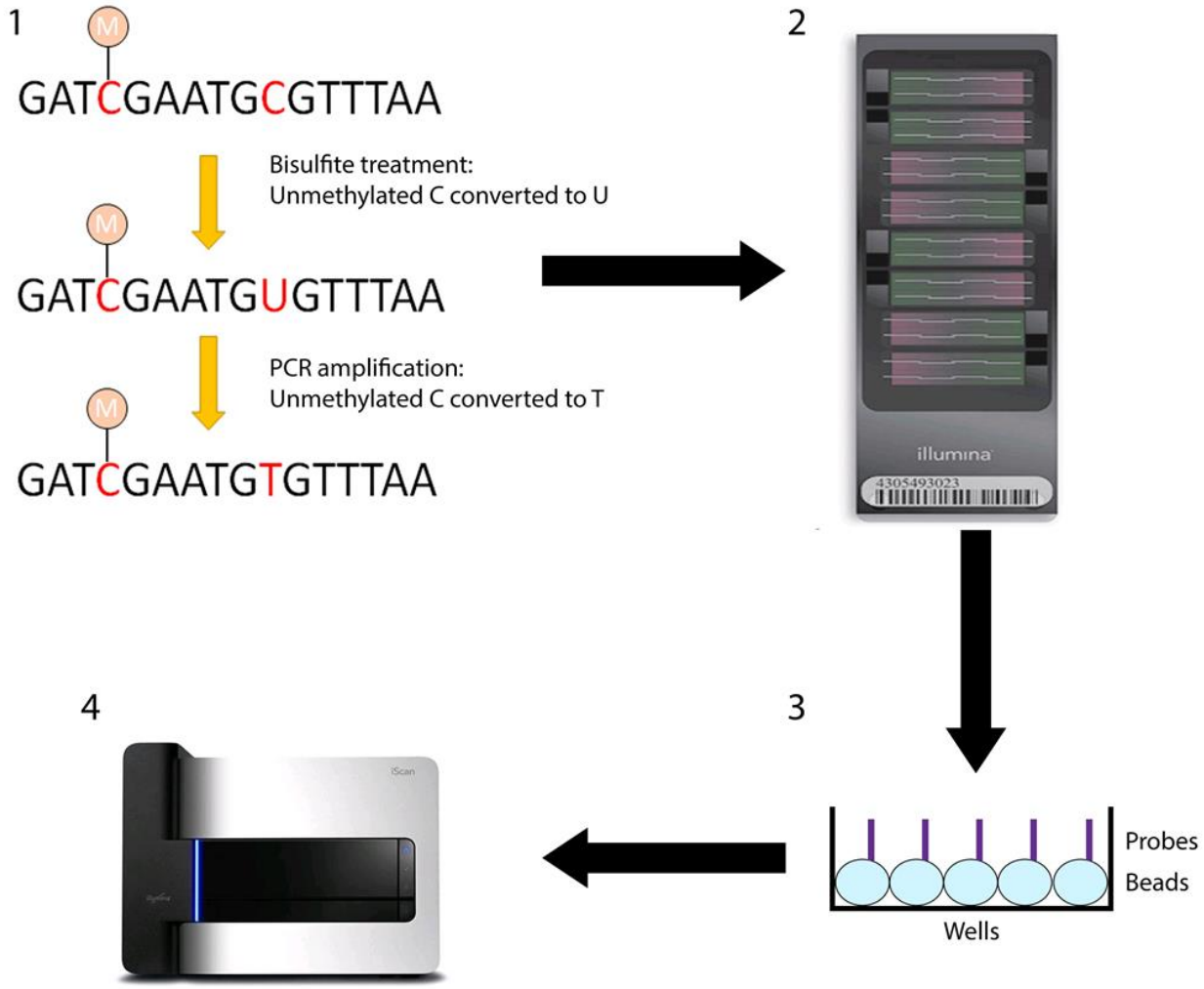


Figure 2-3- Schematic overview of the steps involved in the analysis of DNA methylation using MethylationEPIC BeadChip (850K array) Infinium array.

(1) After DNA extraction, the DNA is bisulfite converted so that only methylated C residues remain and all unmethylated C residues are converted to Ts. (2) The samples are then run over the methylation array. (3) Within the array are wells that contain beads that have 50 base pair probes that are complementary to specific sequences of bisulphite converted DNA. (4) After hybridisation, a fluorescently labelled dNTP is incorporated to the 3' end of the probe to identify whether the C remains (after bisulphite conversion) or whether it has been converted to a T base. The fluorescent signal is then measured using an Illumina iScan.

2.10.1 DNA extraction

DNA was extracted from short term aspirin treated RG/C2 cells (24 hours) and long-term aspirin treated cells (75 weeks) using the TRIzol™ Reagent protocol (ThermoFisher) and the PureLink Genomic DNA Mini Kit (ThermoFisher), respectively, and then stored at -20°C as per the manufacturer's instruction. The concentration of DNA was measured using a NanoDrop (Thermo Scientific) so that the appropriate concentration could be made for bisulphite conversion.

2.10.2 Bisulphite conversion

DNA was diluted to produce 500ng in a volume of 20µL.This was treated with bisulfite using the EZ DNA Methylation-Gold kit (Zymo Research, Irvine, CA). Eluation volume was 12µL. The bisulphite conversion cycling conditions are shown in Table 2-10.

Table 2-10- Bisulphite conversion cycling conditions

Temperature	Time	Number of cycles
98°C	10 minutes	1
64°C	2.5 hours	1
4°C	up to 20 hours (optional)	1

2.10.3 Measuring DNA methylation

After conversion, The Illumina Infinium® HumanMethylation450K (450K array) BeadChip (Illumina Inc. CA, USA) (171) and the MethylationEPIC BeadChip (850K array) Infinium (210) (Illumina) were used to generate methylation data for RG/C2 cells treated with 24 hours aspirin and 75 weeks of aspirin treatment, respectively. More details on the 450K and 850K array can be found in Chapter 4.1.7.1 and Chapter 5.1.3.1, respectively. With regards to the 850K array, there are two types of probes and these are described in more detail in Figure 2-4. An Illumina iScan was used to scan the arrays and initial quality review was evaluated using GenomeStudio (version 2011.1).

In order to minimize confounding by batch effects, samples were randomly distributed across the slides. Batch variables were recorded using the laboratory information management system (LIMS). For each sample, LIMS also records the quality control (QC) metrics from the standard control probes. Samples failing quality control ((>10% probes with a detection P-value ≥ 0.01) were removed from further analysis.

The 450K and 850K assays measure the proportion of molecules methylated at each CpG site on the array. The methylation levels are expressed as a "Beta" value (β-value) which is the ratio of the methylated probe intensity and the overall intensity with values ranging from 0 to 1 where 0 indicates no cytosine methylation and 1 indicates complete cytosine methylation. Methylation data were pre-processed using R (version 3.3.1). Sample quality control and functional normalisation were carried out using the Meffil package (264).

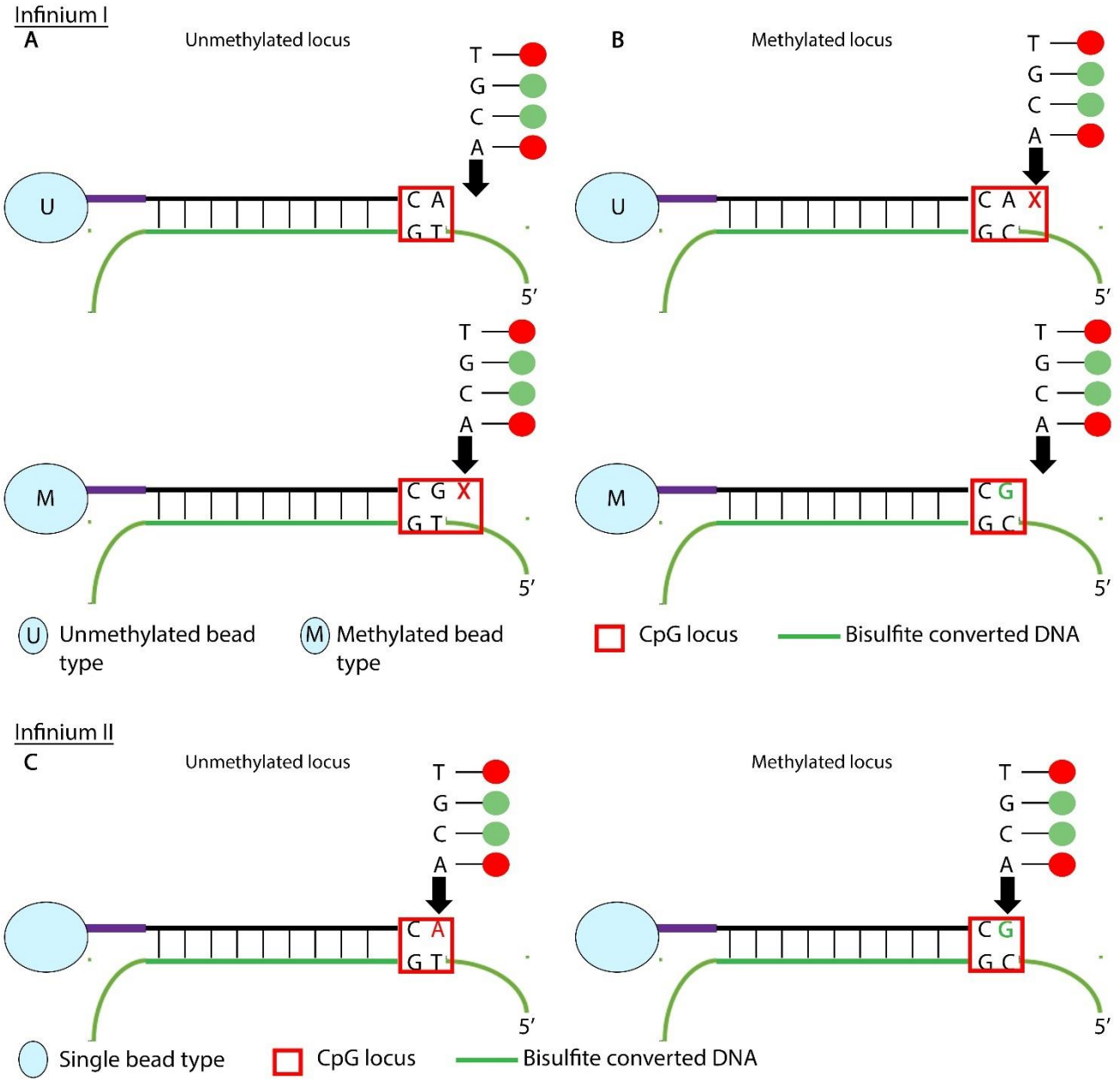


Figure 2-4- The two types of probes found on the MethylationEPIC BeadChip (850K array) Infinium array. Adapted from Infinium® MethylationEPIC BeadChip datasheet (265)**.**

The Infinium I probes have 2 separate sequences for each CpG site. The unmethylated (U) bead measures the unmethylated signal and the methylated (M) bead measures the methylated signal. (A) An unmethylated bisulphite converted sequence hybridises to the probe on the U bead, an A nucleotide is incorporated, and a signal is detected in the *RED channel*. When this same unmethylated bisulphite converted sequence hybridises to the M bead, this results in a mismatch at the 3' end and therefore this inhibits single base extension. (B) A methylated bisulphite converted sequence hybridises to the U bead and mismatch occurs at the 3' end as a result. When this methylated sequence hybridises to the M bead, a G nucleotide is incorporated, and this allows single base extension and a signal is detected in the *GREEN channel*. (C) With regards to the Infinium II probes, the unmethylated and methylated sequences are measures on the same probe. The probe is designed to match the sequence of both the methylated and unmethylated DNA sequences. If the unmethylated bisulphite converted sequence binds to the probe, a labelled A nucleotide binds to the sequence and a signal is detected on the *RED channel*. If the methylated bisulphite converted sequence binds to the probe, a labelled G nucleotide binds to the sequence and a signal is detected on the *GREEN channel* (263).

2.11 Pyrosequencing

After DNA is extracted and bisulphite converted, DNA is amplified using PCR. Through nucleotide incorporation and light emission, sequences of methylated DNA can be detected through pyrosequencing, summarised in Figure 2-5 and Figure 2-6. Details of the method are described below.

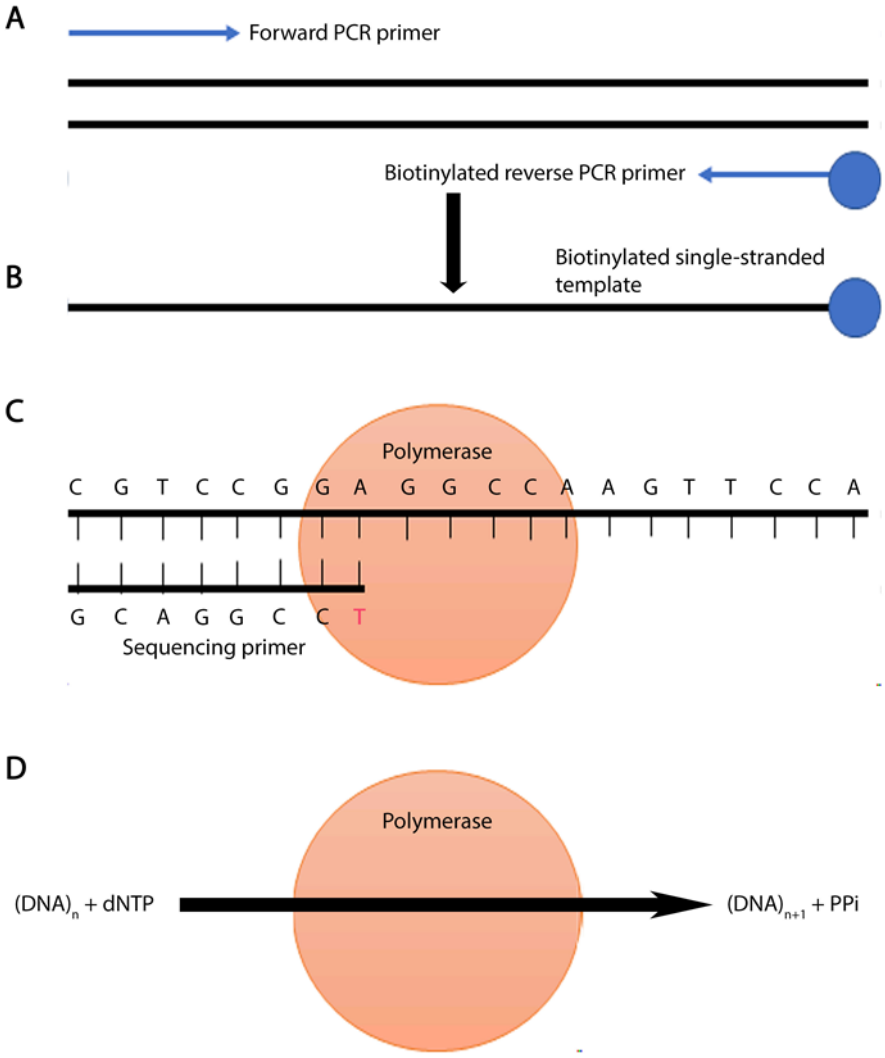


Figure 2-5- The incorporation of the nucleotide to the sequencing primer. Adapted from the Qiagen website (266)**.**

(A) Bisulphite-sequenced DNA is amplified using PCR where one of the primers is biotinylated. (B) After denaturation, the single-stranded PCR amplicon that has been biotinylated is isolated and binds to a sequencing primer. (C) The single-stranded template and the hybridized primer are incubated with 4 enzymes: DNA polymerase, ATP sulfurylase, luciferase and apyrase. They are also incubated with the substrates adenosine 5' phosphosulfate (APS) and luciferin. (D) The first nucleotide (deoxyribonucleotide triphosphate (dNTP)) is added. The DNA polymerase incorporates the complementary nucleotide into the sequencing primer. Each incorporation releases a pyrophosphate (PPi).

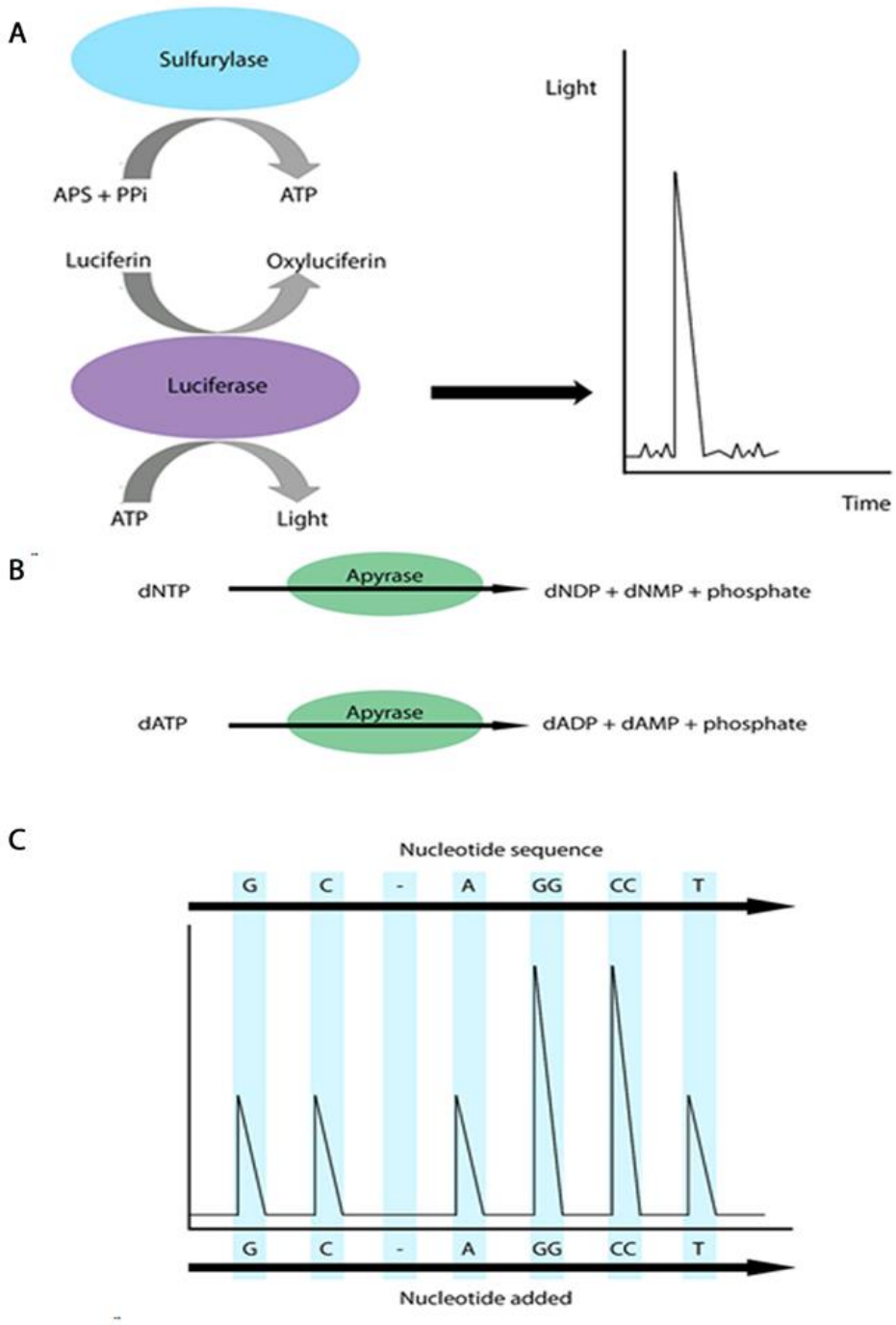


Figure 2-6- Detection of light caused by dNTP incorporation. Adapted from the Qiagen website (266)**.**

(A) The pyrophosphate (PPi) released by the incorporation of the nucleotide is converted to ATP by the enzyme ATP sulurylase in the presence of APS. This ATP then allows the conversion of luciferin to oxyluciferin by the enzyme luciferase. This reaction generates light which is detected by CCD sensors and seen as a peak in the output (Pyrogram). The height of each peak indicates how many nucleotides were incorporated. (B) Apyrase continuously degrades the unincorporated nucleotides. Another nucleotide is added once degradation is complete. (C) The addition of each dNTP is done sequentially. The complementary DNA strand is elongated, and the sequence of the nucleotides is deduced from the peaks in the Pyrogram trace.

2.11.1 DNA extraction

DNA was extracted from short term aspirin treated RG/C2 cells (24 hours) and long-term aspirin treated cells (110 weeks) using the PureLink Genomic DNA Mini Kit (ThermoFisher) and then stored at -20^{o}C as per the manufacturer's instruction. The concentration of DNA was quantified using a Nanodrop (Thermo Scientific) so that the appropriate concentration could be made for bisulphite conversion.

2.11.2 Bisulphite conversion

DNA was diluted to produce 500ng in a volume of 20µL. This was treated with bisulfite using the EZ DNA Methylation-Gold kit (Zymo Research). Eluation volume was 12µL. The bisulphite conversion cycling conditions are shown in Table 2-11.

Table 2-11- Bisulphite conversion cycling conditions

Temperature	Time	Number of cycles
98^{o}C	10 minutes	1
64^{o}C	2.5 hours	1
4^{o}C	up to 20 hours (optional)	1

2.11.3 CpG site primer design

The forward sequences of the CpG sites were obtained from the Illumina library. The forward sequences were then entered into BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat) and 300 base pairs were added on either side of the sequence to allow for primer design. Common SNPs and CpG sites located in this region were marked. The DNA region was then entered into the Zymo website (http://www.zymoresearch.de/tools/bisulfite-primer-seeker) to convert the sequence into the bisulfite modified sequence. This sequence was then copied into PyroMark Assay Design software (Qiagen) and the CpG site of interest was set as the target region. After the software designed multiple primers to analyse the CpG site of interest, primers that were most suitable for the assay were chosen. The software assigns scores to the primers out of 100 calculated according to many factors such as the likelihood of mispriming, the primer length (optimal is 200 base pairs or less) and the chance of primer dimer. We used a threshold score of 80 or more as a cut-off for a suitable primer for the assay.

2.11.4 PCR

To amplify the DNA, PCR was carried out using 96 well plate formats on the DNA samples and EpiTect control DNA, 100% methylated and 0% methylated (Qiagen, Manchester, UK). Four µL of the bisulfite treated DNA was added to 50µL of the PCR reaction containing the reagents listed in Table 2-12. PCR was carried out using a Senquest PCR machine (Genflow, UK) and the thermal cycle was programmed as listed in Table 2-13.

Table 2-12- PCR reagents

Reagent	Supplier	Volume for 1 sample (µL)
PyroMark PCR Master Mix	Qiagen, Manchester, UK	25
CoralLoad concentrate	Qiagen, Manchester, UK	5
Forward primer (0.2µmol/L, biotinylated)	IDT, Leuven, Belgium	0.2
Reverse primer (0.2µmol/L)	IDT, Leuven, Belgium	0.2
dH_2O		19.6
Total volume		50

Table 2-13- PCR cycling conditions

Temperature	Time	Number of cycles
95°C	15 minutes	1
94°C	30 seconds	45
56°C	30 seconds	
72°C	30 seconds	
72°C	10 minutes	1
4°C	hold	1

2.11.5 Gel Electrophoresis

To confirm the PCR products and the presence of DNA, 7µL of the samples were loaded into wells of a 1% (W/V) agarose gel alongside a GeneRuler 100 bp DNA Ladder (Thermo Scientific), in 1 X TBE buffer with 10µL of GelRed™ Nucleic Acid Gel Stain 10,000X (Biotium, Fremont, CA) per 100ml. The bands were then observed using UV light using UVP ChemStudio PLUS (Analytik Jena, Jena, Germany) and images were visualised and captured using the VisionWorks® Software (Analytik Jena).

2.11.6 Pyrosequencing primers

Using the PyroMark Assay Design Software 2.0 (Qiagen), pyrosequencing assays and primers were designed. The oligonucleotides were purchased from IDT (Leuven, Belgium). Target sequences for the PCR and pyrosequencing primers are shown in Figure 2-7. Details on CpG sites and the primers designed for PCR and pyrosequencing are found in Table 2-14 and Table 2-15, respectively.

Illumina probe ID: cg02490185

TTTTGTTGGGTAGAGGAATAGTATATTTAAAGGTTAAGGGTAGGTTTTTTATTGTTGTAGAGTTTTAGAGGGGTTTTCGGGACGGTG
TTATATTAGTTATTTGTTTGTTTTATTTTTTTAGGGTTGGTTATTTGTTAGTAGAGTAGAGGATTTCGGAGGTTTTGGTTTTAAGTTTGT
TTAGATTTTTAATTGATAGTAAGTGGTTATCGAGTGGGTTTGAGTGATCGTTTTATAGGGATGAGATTTGGTTGGGAGTTAGGAG

Illumina probe ID: cg16799926

GGGGAAGGAAAGGGAATTATTATTTGTTGAGTTTGTTGTTTGTGTTTTTTTATGTTATTCGTACGATAATTTAATGTGAATGTTATTAT
TTTTAGGTAATTGTTGAAGAAACGGAAGTTTAAAGAGGTAAGAGATTTGGTTAAGGTTA

Illumina probe ID: cg26432519

AGGAGGGAGGGAATTAGTTATTTTTTGGGAGGAGATGTTTTGTTTAGATAGAAAGCGATATATTGTGGAGAAAATTTGGATTGATT
AG

Forward primer

Reverse primer

Sequencing primer

CpG site of interest

Other CpG sites

Figure 2-7- PCR and pyrosequencing primers target sequence

Nucleotides highlighted in yellow are the target sequences of the forward primer for PCR. Nucleotides highlighted in blue are the target sequences of the reverse primer for PCR. Nucleotides highlighted in grey are target sequences for the pyrosequencing assay. CpG sites highlighted in green are the target CpG site of interest, all other CpG sites in the sequence are highlighted in pink.

Table 2-14- CpG site information required for PCR and pyrosequencing primer design

Illumina probe ID for CpG site	Chr	Genomic Region	Strand	Infinium Design Type (I or II)	Genome Build/Assembly	CpG site Chr	CpG site position
cg02490185	1	26136539-26137260	Forward	II	37	1	26136899
cg16799926	14	70,546,335-70,547,056	Reverse	II	37	14	70546695
cg26432519	18	44,086,996-44,087,717	Forward	II	37	18	44087356

Abbreviations: Chr, chromosome.

Table 2-15- Primers designed for PCR and pyrosequencing

Illumina probe ID for CpG site	Oligonucleotide	Sequence 5' -> 3'	Length of PCR product
cg02490185	Forward oligo	TTTTGTTGGGTAGAGGAATAGTATATTTAA	262
	Reverse oligo	/5BiosG/ATTTGGTTGGGAGTTAGGAG	
	Sequencing oligo	TGTTAGTAGAGTAGAGGAT	
cg16799926	Forward oligo	GGGGAAGGAAAGGGAATTAT	143
	Reverse oligo	AGTTTAAAGAGGTAAGAGATTTGGTTA/3BiosA/	
	Sequencing oligo	ATTATTTTTAGGTAATTGTTGAAG	
cg26432519	Forward oligo	GGAGGGAGGGAATTAGTT	88
	Reverse oligo	TATATTGTGGAGAAAATTTGGATTGATTA/3BiosG/	
	Sequencing oligo	AGGAGATGTTTTGTTTAGATA	

Biotinylation is required so that the amplified sequences can bind to the streptavidin beads required for the pyrosequencing assay. Abbreviations: 5Bios, 5' biotinylated nucleotide; 3'Bios, 3'-biotinylated nucleotide.

2.11.7 Pyrosequencing assay

2.11.7.1 PCR product immobilization to beads

In a 96-plate format, 40µL of amplicon was inserted along with 1.5µL of streptavidin sepharose beads (GE Healthcare, Stockholm, Sweden) and 40µL binding buffer (Qiagen). The plate was sealed and agitated on a shaker for 10 minutes to avoid the sepharose beads from sedimenting.

2.11.7.2 Preparation of the PyroMark Q96 Plate low

In each well of the PyroMark Q96 Plate low (Qiagen), 0.16µL of 100µM sequencing primer (IDT, Leuven, Belgium) was inserted with 40µL of annealing buffer (Qiagen).

2.11.7.3 Strand separation

Filter probes with a vacuum applied were used to transfer the DNA from the PCR plate to the pyrosequencing plate using the PyroMark Q96 Vacuum Workstation (Qiagen) and this was set up as per the manufacturers' instructions. The filter probes were washed in pure water and then lowered into the PCR plate to capture the beads that contain the template DNA. Once all the liquid was aspirated, the filter probes were lowered into 70% ethanol for 5 seconds. The probes were then transferred to the Denaturation Solution (Qiagen) for 5 seconds. After this, the probes were lowered into the Wash Buffer (Qiagen) for a total of 10 seconds. The filter probes were then held over the PyroMark Q96 Plate low at which point the vacuum was turned off. This allows the beads to be released into the plate containing the sequencing primer and the annealing buffer.

2.11.7.4 Annealing of sequencing primer to samples

In order for the sequencing primer to anneal to the samples, the PyroMark Q96 Plate low was placed on a heating block for 2 minutes at 80°C. The plate was then removed and allowed to cool for 10 minutes at room temperature.

2.11.7.5 Reagent preparation

Reagents were prepared according to the PyroMark Gold Q96 Reagents (Qiagen) instructions. The enzyme solution, substrate solution and the dNTP solutions containing A, C, G and T nucleotides were pipetted into

the appropriate PyroMark Q96 Cartridge (Qiagen) compartments as instructed by the manufacturers' instructions.

2.11.7.6 Pyrosequencing

Once the cartridge containing the reagents was loaded, pyrosequencing was carried out using a PyroMark 96 ID pyrosequencer (Qiagen). Results were analysed using the PyroMark Q96 Software (Qiagen).

2.12 Mendelian randomisation

In order to carry out a Mendelian randomisation analysis of the aspirin metabolites and risk of colorectal cancer, datasets with genetic data as well as measured metabolite levels or recorded cancer information were contacted and permission requested to use their data. These datasets are described below.

2.12.1 TwinsUK cohort details and genotyping information

2.12.1.1 Cohort and genotyping platform

TwinsUK is a registry of ~12,000 volunteer twins in the UK aged between 18-103 years with a mean age of 55 years. The percentage of monozygotic twins is 51% and the percentage of dizygotic twins is 49% and around 83% of the registry consists of females. More details on the registry has been published previously (3). TwinsUK dataset genotyping was achieved with 4 types of Illumina arrays: HumanHap300, HumanHap610Q, 1M-Duo and 1.2MDuo 1M.

2.12.1.2 Metabolite measurements

Salicylic acid was measured using gas chromatography-mass spectrometry (GC/MS) and salicyluric acid was measured using liquid chromatography-mass spectrometry (LC/MS). TwinsUK used the Metabolon platform (Metabolon Inc., Durham, USA) which is a non-targeted metabolomic analysis. It uses 2 LC/MS injections (one for acid and for basic species) as well as one GC/MS injection per sample. This allows the detected of 510 small molecules that include salicylic acid and salicyluric acid. Since this is an untargeted approach, the metabolite measures are quantitative values of relative changes as opposed to the targeted approach which would achieve absolute quantification of metabolite concentrations (3,267).

2.12.1.3 Quality control

Sample exclusion criteria included a sample call rate <98%, heterozygosity across all SNPs ≥2 standard deviations from the sample mean, evidence of non-European ancestry assessed by principal components analysis comparison with HapMap3 populations and observed pairwise identical by descent probabilities suggesting sample identity errors.

SNP exclusion criteria included a Hardy-Weinberg equilibrium P-value < 10^{-6} assessed in a set of unrelated samples; a minor allele frequency (MAF) < 1% which was assessed in a set of unrelated samples and a SNP call rate <97% (for SNPs with MAF ≥ 5%) or < 99% (for 1% ≤ MAF < 5%).

2.12.1.4 Imputation

The imputations were performed using the Michigan Imputation Server.

2.12.2 GECCO cohort details and genotyping information

2.12.2.1 Cohort and genotyping platform

We used epidemiological and genetic data from 11895 cases and 14659 population-based controls from 25 studies part of the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) and Colon Cancer Family Registry (C-CFR) (Table 2-16). Full details on the consortium have been previously published (2).

A list of the platforms used for SNP genotyping for each of the studies are summarised in Table 2-16.

Table 2-16- Study populations included in GECCO consortium

Study Name	Platform	Design	Country	N Case	N Control	N Female	N Male	Total*
ASTERISK	Illumina CytoSNP BeadChip	Case-control	France	892	947	763	1076	1839
CCFR Set 1	Illumina Human1M and Human1M-Duo	Case-control	United States, Canada, Australia	1171	980	1075	1076	2151
CCFR Set 2	Illumina Human1M and Human1M-Duo	Case-control	United States, Canada, Australia	787	0	420	358	787
COLO 2&3	Illumina CytoSNP BeadChip	Case-control	United States	87	125	95	117	212
DACHS Set 1	Illumina CytoSNP BeadChip	Case-control	Germany	1710	1708	1395	2023	3418
DACHS Set 2		Case-control	Germany	666	498	435	729	1164
DALS Set 1	Illumina 610K and 550K platforms	Case-control	United States	706	710	615	801	1416
DALS Set 2	Illumina CytoSNP BeadChip	Case-control	United States	410	464	414	460	874
HPFS Set 1	OmniExpress	Cohort	United States	227	230	0	457	457
HPFS Set 2	OmniExpress	Cohort	United States	176	172	0	348	348
HPFS_AD	OmniExpress	Cohort	United States	313	345	0	658	658
MEC	Illumina CytoSNP BeadChip	Cohort	United States	328	346	313	361	674
NHS Set 1	OmniExpress	Cohort	United States	391	774	1165	0	1165
NHS Set 2	OmniExpress	Cohort	United States	158	181	339	0	339
NHS_AD	OmniExpress	Cohort	United States	513	578	1091	0	1091
OFCCR	Affymetrix GeneChip Human	Case-control	Canada	602	522	581	543	1,124

	mapping 100K and 500K Array Set and a 10K non-synonymous SNP chip							
PHS Set 1+2	OmniExpress	Cohort	United States	375	389	0	764	764
PLCO Rematch	Illumina 550K	Cohort	United States	505	522	419	608	1027
PLCO Set 1	Illumina 550K and 610K platforms	Cohort	United States	533	1976	667	1842	2509
PLCO Set 2	Illumina CytoSNP BeadChip	Cohort	United States	486	415	383	518	901
PMH-CCFR	Illumina CytoSNP BeadChip	Case-control	United States, Canada, Australia	280	122	402	0	402
VITAL	Illumina CytoSNP BeadChip	Cohort	United States	285	288	273	300	573
WHI Set 1+Hip	Illumina 550K, 550K duo, and 610K	Cohort	United States	313	1467	1780	0	1780
WHI Set 2	Illumina CytoSNP BeadChip	Cohort	United States	651	816	1467	0	1467
WHI WGS		Cohort	United States	610	309	919	0	919

Abbreviations: ASTERISK, French Association Study Evaluating RISK for sporadic colorectal cancer; C-CFR, Colon-Cancer Family Registry; CPSII, Cancer Prevention Study II; Colo 2&3, A case-control study from the University of Hawai'i;; DACHS, Darmkrebs: Chancen der Verhütung durch Screening; DALS, Diet, Activity, and Lifestyle Study; HPFS_AD, Health Professionals Follow-up Study for colorectal adenoma; MEC, Multiethnic Cohort; NHS, Nurses' Health Study; OFCCR, Ontario Colorectal Cancer Family Registry; PHS, Physicians' Health Study; PLCO, Prostate, Lung, Colorectal, and Ovarian; PMH-CCFR, Post-Menopausal Hormone-CCFR Study; VITAL, CancerScreening Trial; VITamins And Lifestyle cohort; WHI WGS, Women's Health Initiative Whole-Genome Sequencing. *Sample sizes based on GECCO GIGS data.

2.12.2.2 Quality Control

For all the studies included in GECCO and C-CFR, the following quality control checks were implemented. The sample exclusion criteria include: sample call rate ≤ 97%, heterozygosity, unexpected duplicates, unexpected relative pairs, gender discrepancy and PCA outlier of HapMap2 CEU cluster.

The SNP exclusion criteria includes: SNP call rate ≤ 98%, lack of Hardy–Weinberg equilibrium in controls ($P < 1 \times 10^{-4}$), MAF < 5/sample size (CytoSNP) or 5% (WHI Set1, PLCO Set1 and DALS Set1) and SNPs that don't perform consistently across platforms. SNP imputation was performed using MACH for all studies and SNPs were imputed to haplotype consortium reference panel.

2.12.2.3 Imputation

MACH was used to carry out imputation to haplotype consortium reference for all studies.

2.12.3 UK Biobank cohort details and genotyping information

2.12.3.1 Cohort and genotyping platform

UK Biobank is a population-based health research resource that contains ~500,000 people aged recruited between 2006-2010 across the UK aged between 38-78 years. Participants provided a range of information such as lifestyle measures and demographics through questionnaires and interviews. Patients also provided blood, urine and saliva samples as well as blood pressure readings. A full description of the study design and participants has been described previously (4).

The number of successfully genotyped samples is 488,377. Individuals were genotyped using the UK BiLEVE array (n=49,979) and UK Biobank axiom array (n=438,398). The MRC IEU UK Biobank GWAS pipeline has been published (268).

2.12.3.2 Quality control

Samples excluded are those where sex-mismatch (identified by comparing genetic sex and reported sex) was identified or in individuals with sex-chromosome aneuploidy. The sample was also restricted to those of white British ancestry. In addition, kinship was estimated and individuals who were related to other individuals were removed. SNP exclusion criteria included multiallelic SNPs or those with MAF ≤1%. Quality Control filtering of the UK Biobank data has been published (269).

2.12.3.3 Imputation

Genotype imputation was performed using IMPUTE2 algorithms (270) to a reference set combining the UK10K haplotype and Haplotype Reference Consortium (HRC) reference panels (271)

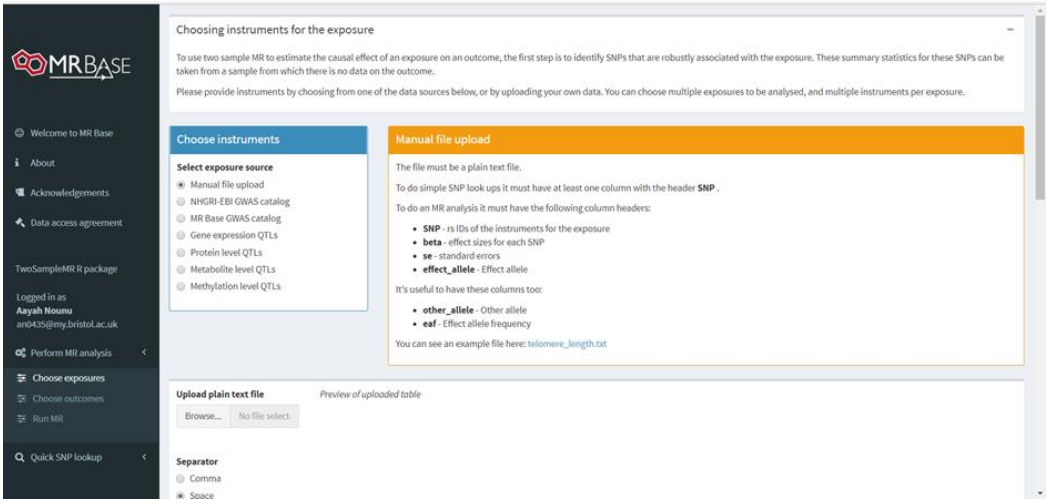
2.12.4 Mendelian randomisation analysis

Mendelian randomisation (MR) is a method that utilizes Mendel's second law but within the context of an epidemiological setting. This law is the law of random assortment which suggests that traits are inherited independently of each other (199,228). MR seeks to use genetic polymorphisms, most commonly single nucleotide polymorphisms (SNPs), to proxy for modifiable exposures to test for their association with diseases (209).

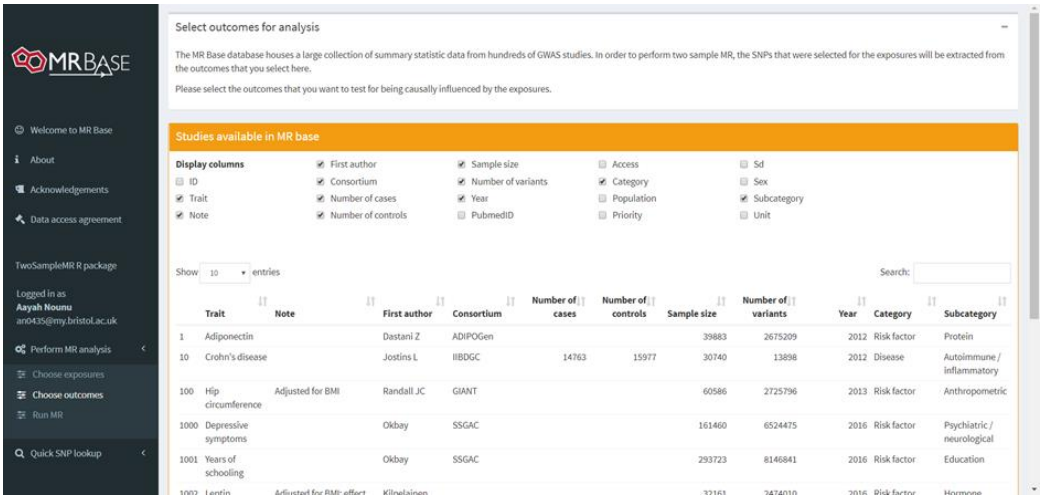
SNPs from enzymes involved in aspirin metabolism and metabolite data was extracted from the TwinsUK registry. Linear regression analyses were carried out to identify SNPs associated with the metabolites salicylic acid and salicyluric acid. SNPs were clumped removing all SNPs in high LD within 10,000kb of the variant of interest and leaving only the SNP with the lowest p-value of association with the exposure. Logistic regression analyses were carried out using the SNPs and colorectal cancer risk as the outcome using data from GECCO and UK Biobank. Alleles were harmonised before Wald ratios were calculated. Results were summarised using the IVW method, maximum likelihood approach, the MR Egger approach, the MR Egger (bootstrap) approach, the simple median approach, the weighted median approach, the simple mode approach and the weighted mode approach. Sensitivity analyses were carried out through leave-one-out analyses and funnel plots. Analyses were carried out in R version 3.2.3 using the "Two-Sample MR" package

(248). This package allows the formatting, allele harmonisation and thus analysis of summary data in an automated manner. A web interface for this package also exists at http://app.mrbase.org/ (Figure 2-8). A more detailed methods description can be found in Chapter 6.1.6.

A



B



C

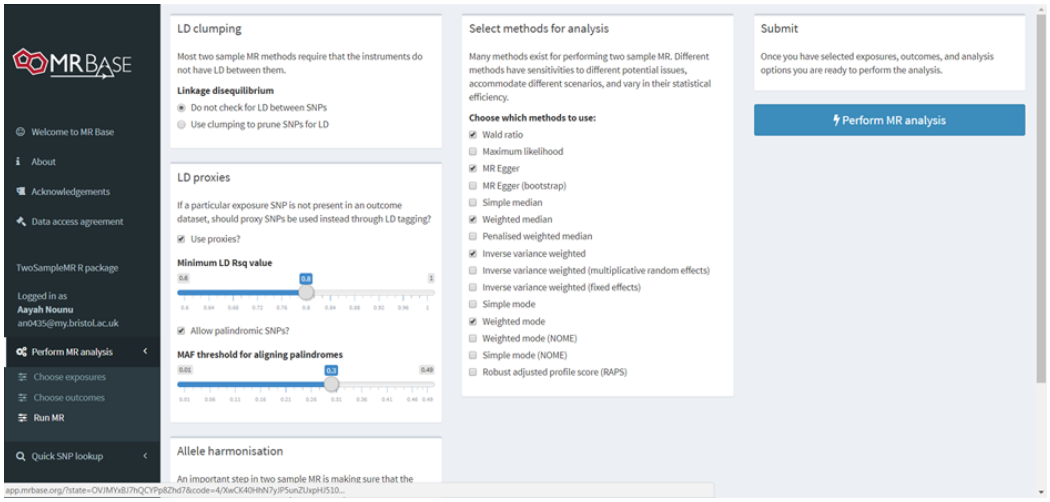


Figure 2-8- Screenshots of MR-Base web platform

(A)A screen shot of choosing SNPs to proxy for the exposure of interest. (B) A screen shot of choosing the outcome of interest. (C) A screen shot for conducting LD clumping, allele harmonisation and choosing the method of analysis desired.

2.13 Statistical analysis

2.13.1 Laboratory-based experiments

Results are mean values of raw or log-transformed data +/- standard deviations (SD) where three independent experiments were undertaken. P-values less than or equal to 0.05 ($p \leq 0.05$) were considered statistically significant (*). Two asterisks indicate $p \leq 0.01$ (**) and three asterisks indicate $p \leq 0.001$ (***). GraphPad Prism (GraphPad Software Inc, California, USA) was used to calculate statistical significance using Analysis of Variance (ANOVA) and Student's T test with a Dunnett's multiple comparison test and Bonferroni multiple comparison test as a post-hoc test. We used a Chi square (X^2) test with Yates' continuity correction to calculate significance of an association between an observed and expected outcome. Statistical tests used in each analysis are stated appropriately.

2.13.2 Population-based significance- adjusting for multiple testing

The Bonferroni correction divides the alpha level by the number of statistical tests being carried out- whilst it is the most conventional and simple method, it tends to be overly conservative, thereby rejecting associations that may be true and also has lower power (272,273). This has been described in Chapter 4.1.7.2.

An alternative method to the Bonferroni correction, is the false discovery rate (FDR) adjusted P-value. The false discovery rate describes the probability at which you discover, for example, a SNP association with an exposure which is untrue and has simply occurred due to chance, therefore it is a false positive (273). This has been described in Chapter 4.1.7.2

Chapter 3 Aspirin's effect on colorectal adenoma cell growth *in vitro*

This chapter aims to address the effect of short-term and long-term aspirin treatment on cell survival using *in vitro* culture, circled in red in Figure 3-1. Firstly, epidemiological evidence that indicates differing effects of aspirin pre and post-diagnosis of cancer are compiled before a description of 3D culture systems is introduced.

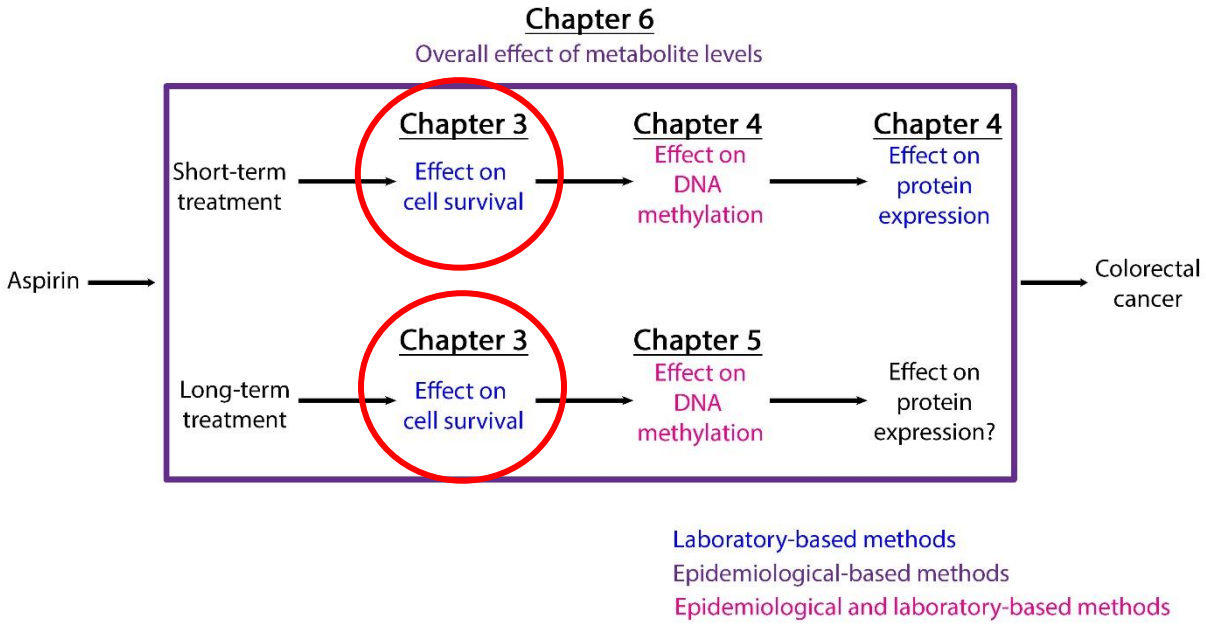


Figure 3-1- Overview of the questions addressed in this thesis to explore causality and potential mechanisms of aspirin on CRC.

3.1 Background

3.1.1 Aspirin pre and post-diagnosis

Although evidence for aspirin's protective effect is clear (Chapter 1.5), studies comparing aspirin's beneficial properties pre-diagnosis and post-diagnosis have shown that the timing of aspirin intake may affect how early the beneficial properties of aspirin are observed and also indicate that different mechanisms of aspirin resistance may be observed possibly explained by the cells being at different stages of cancer.

A large amount of evidence has been published in the literature to support aspirin's beneficial effects at reducing colorectal cancer (CRC) incidence and mortality pre-diagnosis (112,274,275). We focus here on the studies that compared the effect of aspirin both pre and post-diagnosis. Chan et. al were the first to report that aspirin use post-diagnosis was more effective at increasing survival, especially in tumours that expressed high cyclooxygenase 2 (COX-2) (276). They showed that aspirin use post-diagnosis was associated with decreasing the risk of CRC mortality by 47% (HR: 0.53, 95% CI 0.33-0.86) and that aspirin use pre-

diagnosis did not decrease risk of mortality (HR: 0.89, 95% CI 0.59-1.35). Other similar studies are summarised in Table 3-1.

More recently, a meta-analysis approach by Li et. al (2015) was carried out to pool the results of these studies. They combined results from seven studies for pre-diagnosis and seven studies for post-diagnosis aspirin use. The results from their meta-analysis showed that aspirin use post-diagnosis reduced the risk of all-cause mortality (HR: 0.84, 95% CI 0.75-0.94) whereas pre-diagnosis aspirin use was not associated with reducing this risk (HR: 1.01, 95% CI 0.96-1.06). However, their results also showed that neither pre-diagnosis (HR: 0.77, 95% CI 0.52-1.14) nor post-diagnosis aspirin use (HR: 0.93, 95% CI 0.82-1.05) was associated with decreasing the risk of CRC-specific mortality (277).

Since 2015, several other studies have been carried out summarised in Table 3-2. Their results are conflicting as Hua et. al (2017) showed that aspirin post-diagnosis but not pre-diagnosis decreased the risk of CRC specific mortality (278) whereas Bains et. al (2016) showed aspirin intake pre-diagnosis and post-diagnosis was effective at reducing CRC-specific mortality (279) and Gray et. al (2018) showed that neither aspirin pre or post-diagnosis reduced CRC-specific mortality (280). Most strikingly, a recent randomised controlled trial (RCT) for the effect of aspirin on disability-free survival has shown that aspirin increases the risk of CRC-specific mortality in elderly individuals (HR: 1.77, 95% CI 1.02-3.06) (281).

Taken together, these studies suggest that aspirin may be exerting different effects depending on whether it is taken for a short time period (short-term) or whether it is taken for a longer time period (long-term) and we hypothesise that long-term exposure may result in an acquired “resistance”. This raises the possibility that previous long-term aspirin users are less sensitive or more resistant to the protective effects experienced by post-diagnosis users due to acquired phenotypic changes.

Table 3-1- A summary of studies that compared the effect of aspirin pre and post-diagnosis

Author/Year/Reference/ study type	N Aspirin users	N Aspirin non-users	Pre/Post-diagnosis	Median years of follow up	Outcome	Result of aspirin use
Chan 2009 (276)- cohort	549	730	pre-diagnosis	11.8	CRC mortality	HR: 0.89 (95% CI, 0.59-1.35)
			post-diagnosis	11.8	CRC mortality	**HR: 0.53 (95% CI, 0.33-0.86)**
Bastiaannet 2012 (282)- cohort	2086 prescribed aspirin/NSAIDs prediagnosis and postdiagnosis, 1219 prescribed aspirin/NSAIDs postdiagnosis only	1176	pre-diagnosis	3.5	CRC overall survival	RR: 0.96 (95% CI, 0.84-1.09)
			post-diagnosis	3.5	CRC overall survival	**RR: 0.75 (95% CI, 0.62-0.92)**
			frequent users post diagnosis	3.5	CRC overall survival	**RR: 0.69 (95% CI, 0.56-0.86)**
McCowan 2013 (283)- cohort	1340	1650	pre-diagnosis	2.8	CRC mortality	HR: 0.96 (95% CI, 0.84-1.11) (adjusted)
			post-diagnosis	2.8	CRC specific mortality	**HR: 0.58 (95% CI, 0.45-0.75) (adjusted)**
			pre-diagnosis	2.8	all cause mortality	HR: 0.99 (95% CI, 0.90-1.09)
			post-diagnosis	2.8	all cause mortality	**HR: 0.67 (95% CI, 0.57-0.79) (adjusted)**

Abbreviations: NSAIDs, non-steroidal anti-inflammatory drugs; HR, hazard ratio; RR, relative risk. Statistically significant results are in bold, red font.

Table 3-2- A summary of results from recent studies comparing the effect of aspirin pre and post-diagnosis

Study/Year/Reference/ Study type	N Aspirin users	N Aspirin non-users	Pre/Post-diagnosis	Median years of follow up	Outcome	Result of aspirin use
Hua 2017 ((278)-cohort	676	1397	pre-diagnosis	1 year pre-diagnosis,5 years post-diagnosis	CRC-specific survival	HR: 0.45 (95% CI, 0.19-1.08)
			post-diagnosis	5 years post-diagnosis	CRC specific survival	**HR: 0.39 (95% CI, 0.19-0.78)**
Bains 2016 (279)-cohort	1711	2680	pre and post-diagnosis	3 years	CRC-specific survival	**HR: 0.77 (95% CI 0.71-0.84) (adjusted)**
			post-diagnosis	3 years	CRC-specific survival	HR: 1.00 (95% CI 0.87–1.14) (adjusted)
Gray 2018 (280)-cohort	2150	6241	pre-diagnosis	at least 1 year pre-diagnosis, 3.6 years post diagnosis	CRC-specific mortality	HR: 0.96 (95% CI 0.88-1.05) (adjusted)
			post-diagnosis	3.6 years	CRC-specific mortality	HR: 1.17 (95% CI 1.00-1.36) (adjusted)
McNeil 2018 (281) - RCT	9525	9589	pre-diagnosis	4.7 years	CRC-specific mortality	**HR: 1.77 (95% CI 1.02-3.06)**

Abbreviations: RCT, randomised controlled trial; HR, hazard ratio; RR, relative risk. Statistically significant results are in bold, red font.

3.1.2 Aspirin resistance

Genetic aspirin resistance is a well-documented phenomenon with regards to coronary artery disease, but patients respond differently to aspirin- raising the possibility of aspirin resistance. Aspirin resistance follows two definitions for either clinical resistance or laboratory resistance. In terms of clinical resistance this is when aspirin fails to prevent another coronary event occurring, laboratory aspirin resistance is when the drug fails to prevent platelet reactivity/aggregation (133,284,285).

Upon long-term administration patients undergo tachyphylaxis i.e. they become less responsive to the effects of aspirin the longer they take it. (284) This was demonstrated as early as 1999 and 2001 where studies showed that those who took aspirin before treatment for coronary syndromes responded less to aspirin intake and had a higher risk of cardiac events. (286,287). Reasons for genetic aspirin resistance in the

context of cardiovascular diseases include single nucleotide polymorphisms in genes such as *COX* and receptors and integrin complexes on the surface of platelets (288). In the context of *COX*, polymorphisms may occur in the serine residues targeted by aspirin for acetylation thus inhibiting aspirin from acetylating the enzyme and rendering it inactive.

It is possible that the concept of tachyphylaxis may be present in the context of cancer. As shown by the epidemiological studies (Chapter 3.1.1), not everyone benefits from the chemopreventative properties of aspirin thereby indicating that a subset of the population harbours a genetic resistance to aspirin. However, what is currently unclear is the existence of acquired aspirin "resistance" whereby cells become more resistant to aspirin over time. We therefore aimed to investigate the effects of aspirin on cell survival and apoptosis in 2D and 3D cell culture upon short-term (24-168 hours) and long-term (at least 18 weeks) treatment.

3.1.3 Three-dimensional culturing systems

To assess the effect of short-term and long-term aspirin use, we treated human adenoma cells in culture. One limitation of growing cells in 2D culture is that they do not form the cell-cell as well as the cell-extracellular matrix (ECM) interactions and structures that are formed *in vivo* (289). Three-dimensional culture systems are much better at mimicking *in vivo* conditions making them useful for tumour modelling (290,291).

It is becoming increasingly clear that culturing cells in 2D vs 3D result in different gene and protein expression profiles. Takagi et. al (2007) found that the prostate cancer cell line LNCaPs grown in 3D exhibit more similar microarray expression profiles to solid tumour tissues compared to monolayer cultures (292). Yu et. al (2016) also showed using stable isotope labelling of amino acids (SILAC) that the proteome and phosphoproteome of colon carcinoma HT29 cells grown in 2D and 3D culturing systems vary and found that cells grown in 3D were enriched for proteins involved in RNA binding and cytoskeletal organisation (293). It is for these reasons that 3D culture systems are being utilised more to investigate the effects of treatments on cancer cells as evidenced by the increasing number of methods developed (294).

Many 3D culturing systems have emerged, including: multicellular tumour spheroids, suspension culture, device-assisted culture, gel embedding culture and scaffold culture (295). One example of a gel embedding culture is growing cells within Matrigel©. Matrigel© is the commercial trade name for a basement membrane extract derived from the Engelbreth-Holm-Swarm (EHS) cell line of mouse sarcoma cells (296,297). It contains many extracellular matrix proteins such as laminin, nidogen-1 (entactin), collagen and other growth factors (297). In 2011, Sato et. al published a method for culturing organoids from human colon epithelial cells and adenoma cells that exhibited similar architecture and self-renewal capacity of cells within the colonic epithelium using Matrigel© supplemented with various growth factors (298). In the

following experiments, we have used an adapted method based on that published by Sato et. al (2011), to grow human adenoma derived cell lines in 3D.

3.1.4 Hypothesis and aims

Hypothesis:

Long-term aspirin treatment confers resistance in adenoma cells.

Aims:

1- To investigate the short-term effects of aspirin on cell growth and survival in 2D models
2- To investigate the short-term effects of aspirin on cell growth and survival in 3D models
3- To investigate the long-term effects of aspirin on cell growth and survival in 3D models
4- To establish whether exposure to aspirin results in an acquired phenotypic change

3.2 Methods

In order to assess aspirin's effect on the number of attached and floater cells, RG/C2 cells were grown in both 2D culture with 0, 2 and 4mM concentrations of aspirin treatment for a maximum of 96 hours. Aspirin dosages are based on previous experiments from our laboratory and time points were chosen based on when the effect of aspirin was most clear for the outcome to be measured. Details on cell culture, treatment and counting can be found in the methods (Chapter 2.1). To assess whether aspirin was affecting apoptosis, cells were treated with QVD (Chapter 2.2.2). Western blots were carried out to confirm that QVD treatment was inhibiting apoptosis (Chapter 2.7). To assess aspirin's effect on cell cycle progression, cells were subject to FACS analysis (Chapter 2.6.2). RG/C2 cells with short-term aspirin treatment (7 days) were also grown in 3D culture (Chapter 2.4). Using 3D culture, short-term aspirin treated cells (72 hours) were stained with calcein AM and ethidium homodimer 1 to assess aspirin's effect on apoptosis (Chapter 2.6.3).

RG/C2 cells were also grown with long-term aspirin (18 weeks) in 2D culture. Cells were then extracted and grown in 3D culture (Chapter 2.4). Images were taken using widefield microscopy and areas were calculated using MATLAB R2015a (Chapter 2.4). Long-term aspirin treated cells were also removed from aspirin and their growth was measured in 3D culture (Chapter 2.4).

3.3 Results

3.3.1 Short term aspirin treatment (up to 7 days) in 2D and 3D models

Pre- and post-diagnosis aspirin use appear to have differing effects on CRC survival and mortality (Chapter 3.1.1) giving rise to the possibility of acquired "resistance". To investigate, we decided to carry out short-term and long-term aspirin treatment of cells within the laboratory. Initially, we cultured RG/C2 cells with aspirin in 2D culture for 96 hours to confirm aspirin response of adenoma cells in 2D and the effect on cell growth and apoptosis. The same experiments were carried out in 3D culture to compare the response to aspirin in 3D culture with previous 2D culture system results (290,291).

3.3.1.1 Aspirin causes a decrease in attached cells and increase in floating cell numbers in RG/C2 cells in 2D culture

Previously, our laboratory has shown that aspirin is able to induce cell cycle arrest but not apoptosis of RG/C2 cells in 2D culture (299). Before investigating aspirin's effect on 3D culture, we carried out baseline experiments in 2D models first to establish the response.

RG/C2 cells were cultured with 0, 2 and 4mM aspirin for a period of 24, 48, 72 and 96 hours. The number of attached cells and the number of floating cells were counted for each time point and the percentage floating cells were calculated. A one-way ANOVA and Dunnett's multiple comparison test to calculate significance were applied.

Our results show that 2mM aspirin does not significantly decrease the number of attached cells at any of the time points. It also has no significant effect on the percentage floaters. The higher aspirin dose of 4mM treatment significantly decreases the number of attached cells at 48 ($p \leq 0.001$), 72 ($p \leq 0.05$) and 96 hours ($p \leq 0.001$) and significantly increases the percentage floaters at 48 ($p \leq 0.001$), 72 ($p \leq 0.05$) and 96 hours ($p \leq 0.001$). Therefore only 4mM aspirin decreases cell growth and increases the percentage of floaters, previously shown to indicate apoptosis (260).

Figure 3-2 shows that the number of attached cells in 4mM aspirin treated RG/C2s does not change across all time points. At 72 and 96 hours, the average number of cells with 4mM aspirin treatment is 5.16 $x10^6$ (SD 1.47) and 4.3 x 10^6 (SD 1.24) , respectively, and the average number in the untreated is 8.57 x 10^6 (SD 1.32) 9.08 x 10^6 (SD 0.91), respectively. The percentage floaters at 72 and 96 hours for 4mM treatment is 5.78% (SD 1.07) and 7.99% (SD 1.51), respectively, and the untreated is 3.76% (SD 0.84) and 4.97% (SD 1.25), respectively. The increase in percentage floaters is not enough to explain the decrease of the number of attached cells at 4mM compared to the untreated cells suggesting that aspirin is inducing growth arrest rather than inducing apoptosis.

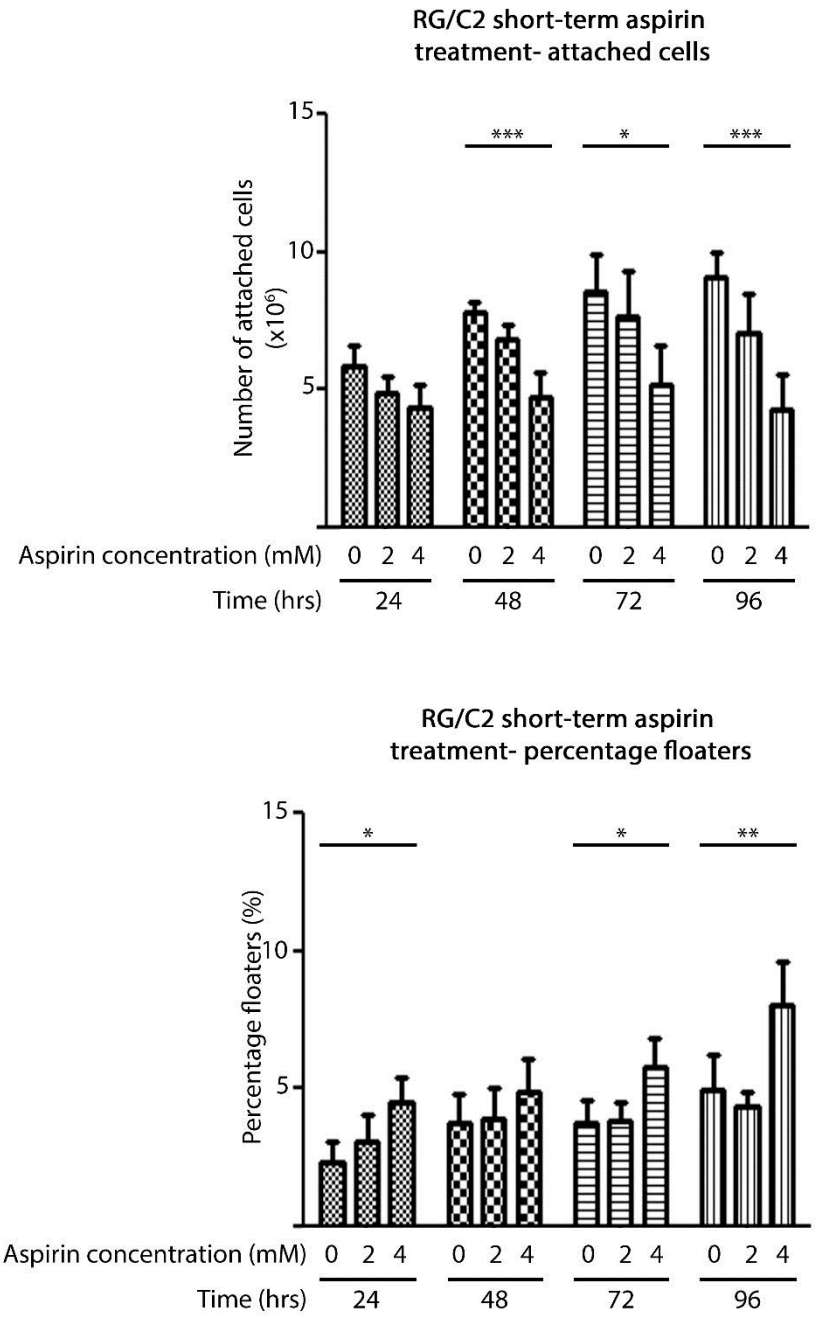


Figure 3-2- Short-term aspirin treatment decreases the number of attached cells and increases the percentage floaters.

RG/C2 adenoma cells were treated with 0, 2 and 4mM aspirin for 24, 48, 72 and 96 hours. The number of attached and floating cells were counted, and percentage floaters was calculated. Aspirin significantly decreases the number of attached cells at 4mM but not 2mM at (48 ($p \leq 0.001$), 72 ($p \leq 0.05$) and 96 hours ($p \leq 0.001$). The percentage floaters also increased with 4mM aspirin treatment at 24 ($p \leq 0.05$), 72 ($p \leq 0.05$) and 96 hours ($p \leq 0.01$). Results are mean values from 4 repeat experiments with standard deviations and a one-way ANOVA and Dunnett's multiple comparison test (*** $p \leq 0.001$, **$p \leq 0.01$ and * $p \leq 0.05$).

3.3.1.2 Aspirin does not induce apoptosis in RG/C2 cells in 2D culture at 72 hours

To identify whether the increase in percentage floating cells was due to apoptosis, RG/C2 cells were treated with 4mM aspirin and QVD (a broad spectrum caspase inhibitor (300))(10µM) for 72 hours. This time point was chosen as a significant decrease in the number of attached cells and a significant increase in the percentage floaters was achieved. A one-way ANOVA and Dunnett's multiple comparison test to calculate significance were applied.

Our results firstly show that 4mM aspirin treatment at 72 hours has no significant effect on the percentage of floating cells in the controls but QVD treatment significantly decreases the percentage floaters in both aspirin untreated ($p \leq 0.001$) and aspirin treated ($p \leq 0.001$) (Figure 3-3 A). Apoptosis inhibition is confirmed through western blotting which shows that QVD treatment decreases the levels of cleaved PARP and cleaved caspase 3 (Figure 3-3 B). With regards to the attached cells, aspirin significantly decreases the number of attached cells which is unaffected by QVD treatment ($p \leq 0.001$); there is no significant change in the number of attached cells between 4mM aspirin treated cells with and without QVD (Figure 3-3 C).

The results show that although QVD inhibits apoptosis (as indicated by the decrease in percentage floaters and the decrease in the expression of proteins involved in the caspase-driven apoptotic pathway), there was no significant change between 4mM QVD treated and untreated RG/C2s indicating that the decrease in cell yield in aspirin treated cells is not due to an induction of apoptosis.

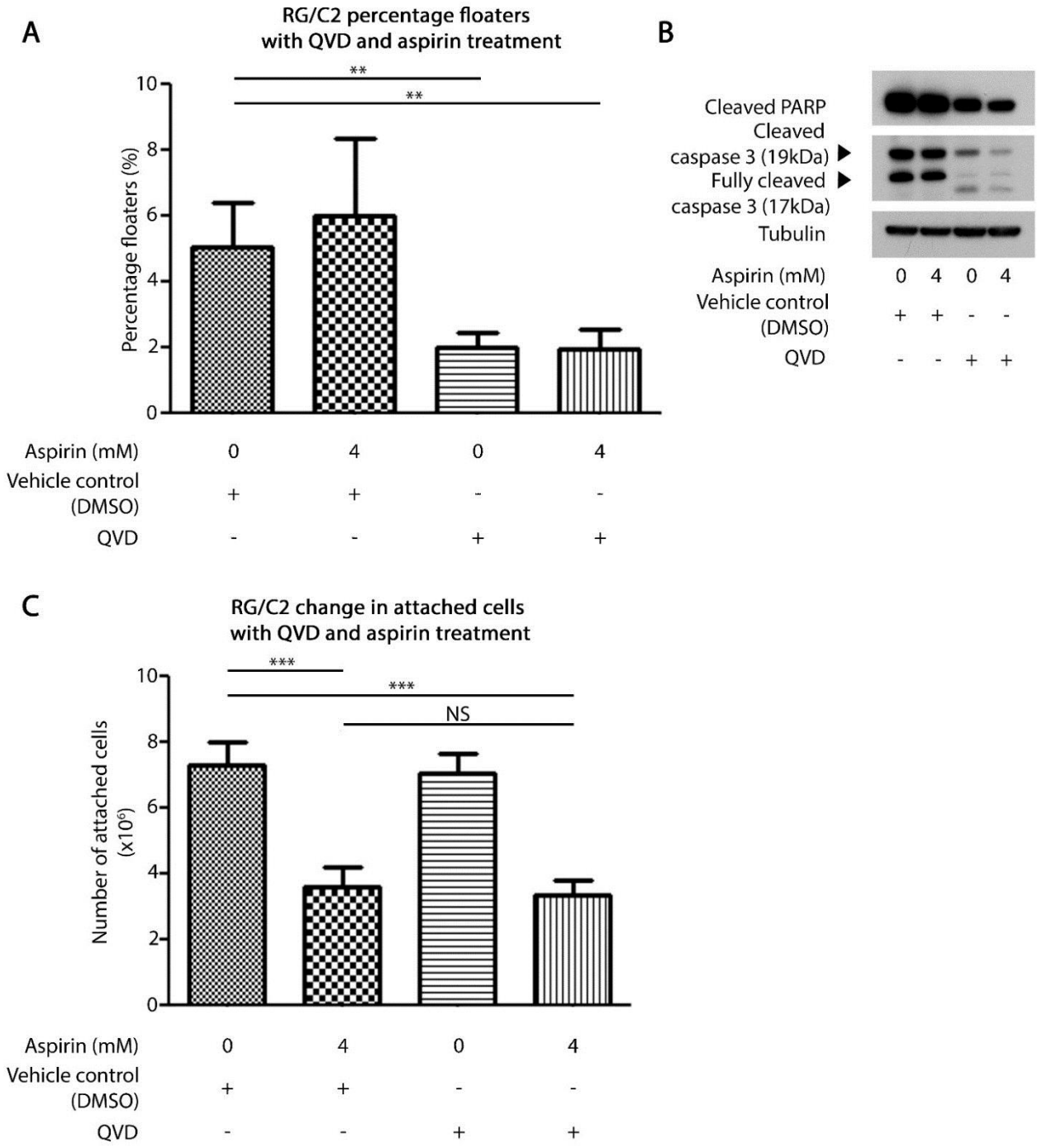


Figure 3-3- Short-term Aspirin does not induce apoptosis in RG/C2 cells in 2D culture at 72 hours

RG/C2 adenoma cells were treated with 4mM aspirin +/- QVD (10µM) for 72 hours. DMSO was used as a vehicle control. The number of attached and floating cells were counted, and percentage floaters was calculated. Cell lysates were extracted to confirm apoptosis inhibition. A – indicates absence of treatment (vehicle or QVD) and a + sign indicates the addition of treatment (DMSO or QVD). Cell count results are mean values from 3 repeat experiments with standard deviations and a one-way ANOVA and Dunnett's multiple comparison test (*** $p \leq 0.001$, **$p \leq 0.01$ and * $p \leq 0.05$, NS= non-significant). Western blotting results are representative of data from 3 independent experiments. (A) 4mM aspirin treatment has no significant effect on the percentage of floating cells in the controls but QVD treatment significantly decreases the percentage floaters in both untreated ($p \leq 0.001$) and aspirin treated cultures ($p \leq 0.001$). (B) Western samples set up in parallel with the experiment were used to investigate cleaved PARP, cleaved caspase 3 (19kDa) and fully cleaved caspase 3 (17kDa). Tubulin was used as a loading control. These results show that QVD treatment reduces cleaved PARP, cleaved caspase 3 and fully cleaved caspase 3. (C) Aspirin significantly decreases the number of attached cells with vehicle ($p \leq 0.001$) and with QVD treatment ($p \leq 0.001$). There is no significant difference in the number of attached cells in aspirin treated cells with and without QVD treatment.

3.3.1.3 *Aspirin significantly increases the percentage of cells in G_0/G_1 phase and decreases the percentage of cells in S phase in 2D culture*

Due to 4mM aspirin doses significantly decreasing cell numbers but not significantly increasing percentage floaters, we investigated whether aspirin causes cell cycle arrest. Adenoma cells were treated with 0, 2 and 4mM aspirin for 24 hours. Cells were then collected, fixed and stained with DRAQ5 before being analysed by flow cytometry. A one-way ANOVA and Dunnett's multiple comparison test to calculate significance were applied.

Our results show that 2mM aspirin treatment has no significant effect on the percentage of cells in G_0/G_1 phase or S phase but 4mM significantly increases the percentage of cells in G_0/G_1 from 60.76% (SD 1.37) in the untreated to 68.41% (SD 0.56) ($p \leq 0.01$) (Figure 3-4 A, Table 3-3). 4mM aspirin treatment also significantly decreases the percentage of cells in S phase from 18.9% (SD 2.58) in the untreated to 11.82% (SD 0.37) ($p \leq 0.01$) (Figure 3-4 B, Table 3-3). Representative images of the cell cycle graphs analysed in FlowJo (version 7.6.5) are shown in Figure 3-4 C. These results confirm that 4mM aspirin reduces cell cycle progression.

Table 3-3- The percentage of cells at each stage in the cell cycle

	Cell cycle					
	G_0/G_1		S		G_2/M	
Aspirin concentration (mM)	Percentage	SD	Percentage	SD	Percentage	SD
0	60.76	1.37	18.9	2.12	18.86	2.59
2	60.63	2.3	20.23	1.88	17.45	3.34
4	68.41	0.56	11.82	1.58	17.27	0.38

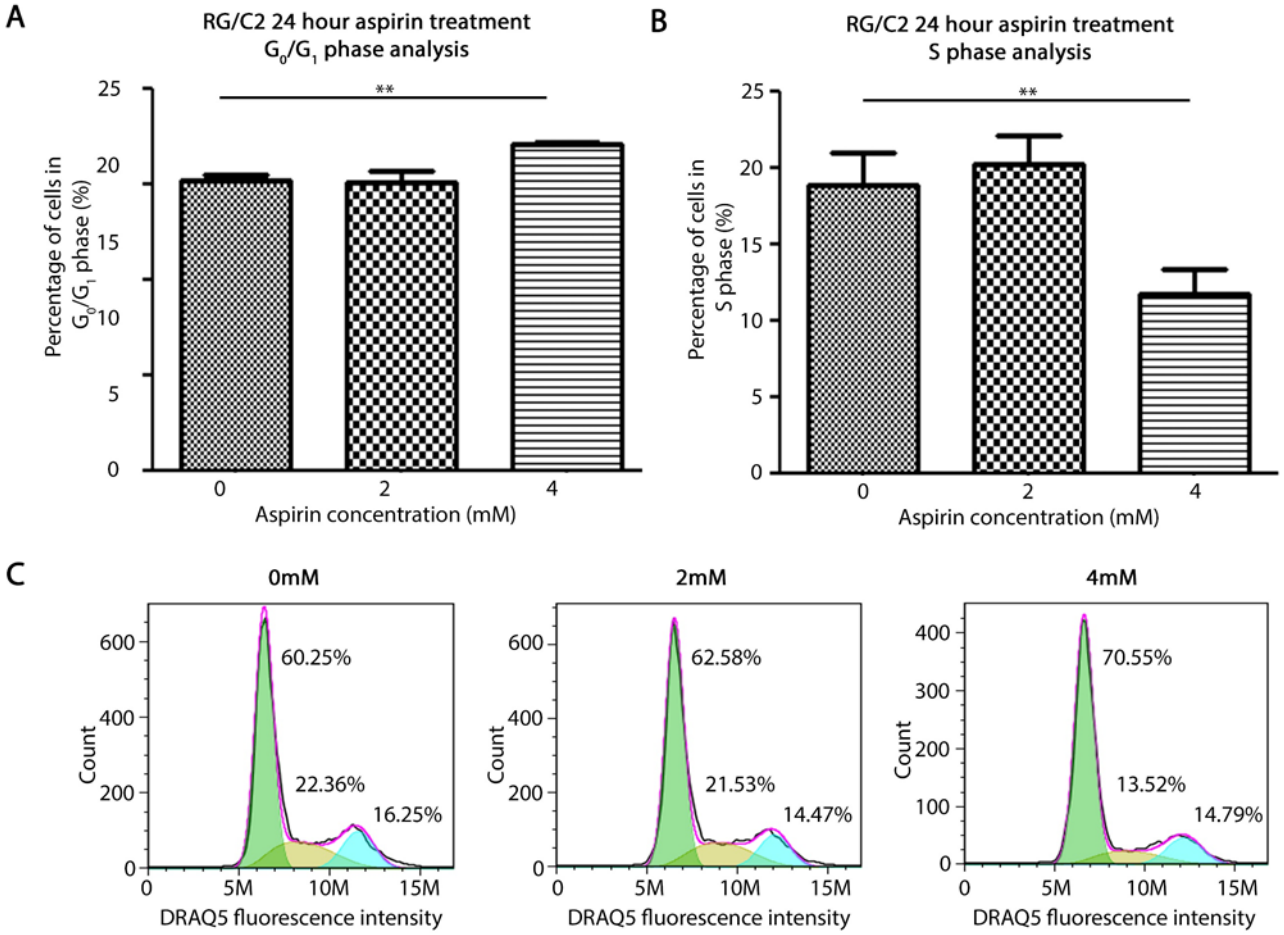


Figure 3-4- 4mM Aspirin significantly increases the percentage of cells in G0/G1 phase and decreases the percentage of cells in S phase.

RG/C2 adenoma cells were treated with 0, 2 and 4mM aspirin for 24 hours. Live and dead cells were extracted and stained with DRAQ5. Data was analysed by flow cytometry using the Novocyte 3000. Results are mean values from 3 repeat experiments with standard deviations and a one-way ANOVA and Dunnett's multiple comparison test (*** $p \leq 0.001$, **$p \leq 0.01$ and * $p \leq 0.05$) (A) 4mM aspirin treatment significantly increases the percentage of cells in G0/G1 phase ($p \leq 0.01$) but 2mM has no significant effect. (B) 4mM aspirin significantly decreases the percentage of cells in S phase ($p \leq 0.01$) but no significant effect is observed with 2mM treatment. (C) Representative images from one FACS experiment. The green peaks indicate the percentage of cells at G_0/G_1, the yellow peaks indicate the percentage of cells in the S phase and blue peaks indicate the number of cells at G_2/M phase. Images were manually analysed in FlowJo (version 7.6.5).

3.3.1.4 Aspirin reduces cell growth in 3D cultures

To investigate the effect of short-term aspirin treatment in 3D models we grew the cells as spheroids in Matrigel because these models allow us to better characterise the phenotypic effect of aspirin on colorectal adenomas (291).

Adenoma cells were grown in a 3D culture system adapted from Sato et. al (2011) (Chapter 2.4) for 2 weeks and then treated with 0, 0.5, 2, 4 and 6mM aspirin for 7 days (298). Images of the spheroids was taken at day 14 and 21. These time points were based on data from previous experiments in the laboratory. Average spheroid sizes for each condition was calculated and the data was converted to the log scale to achieve a normal distribution and therefore carry out statistical tests of significance. The average change in spheroid size from day 14 to day 21 for each aspirin condition was calculated. A one-way ANOVA and Dunnett's multiple comparison test to calculate significance were applied.

The results show that aspirin significantly reduces the growth of RG/C2 cells at 2mM ($p \leq 0.05$), 4mM ($p \leq 0.001$) and 6mM ($p \leq 0.001$) (Figure 3-5 A) and representative images show that increasing the aspirin concentration slows down the growth rate of the spheroids, especially at 4mM and 6mM (Figure 3-5 B). One difference between the results obtained from 2D and 3D experiments is that spheroids are more sensitive to aspirin treatment than cells grown in monolayers. Interestingly there is no change in size between day 14 and day 21 at 4mM and 6mM (0.03 (SD 0.04) and 0.00 (SD 0.03)) indicating that aspirin inhibits cell growth in 3D suggesting maximum inhibition of growth can be achieved at 4mM. However, possible induction of apoptosis needs to be examined in 3D cultures as cells are more sensitive than in 2D.

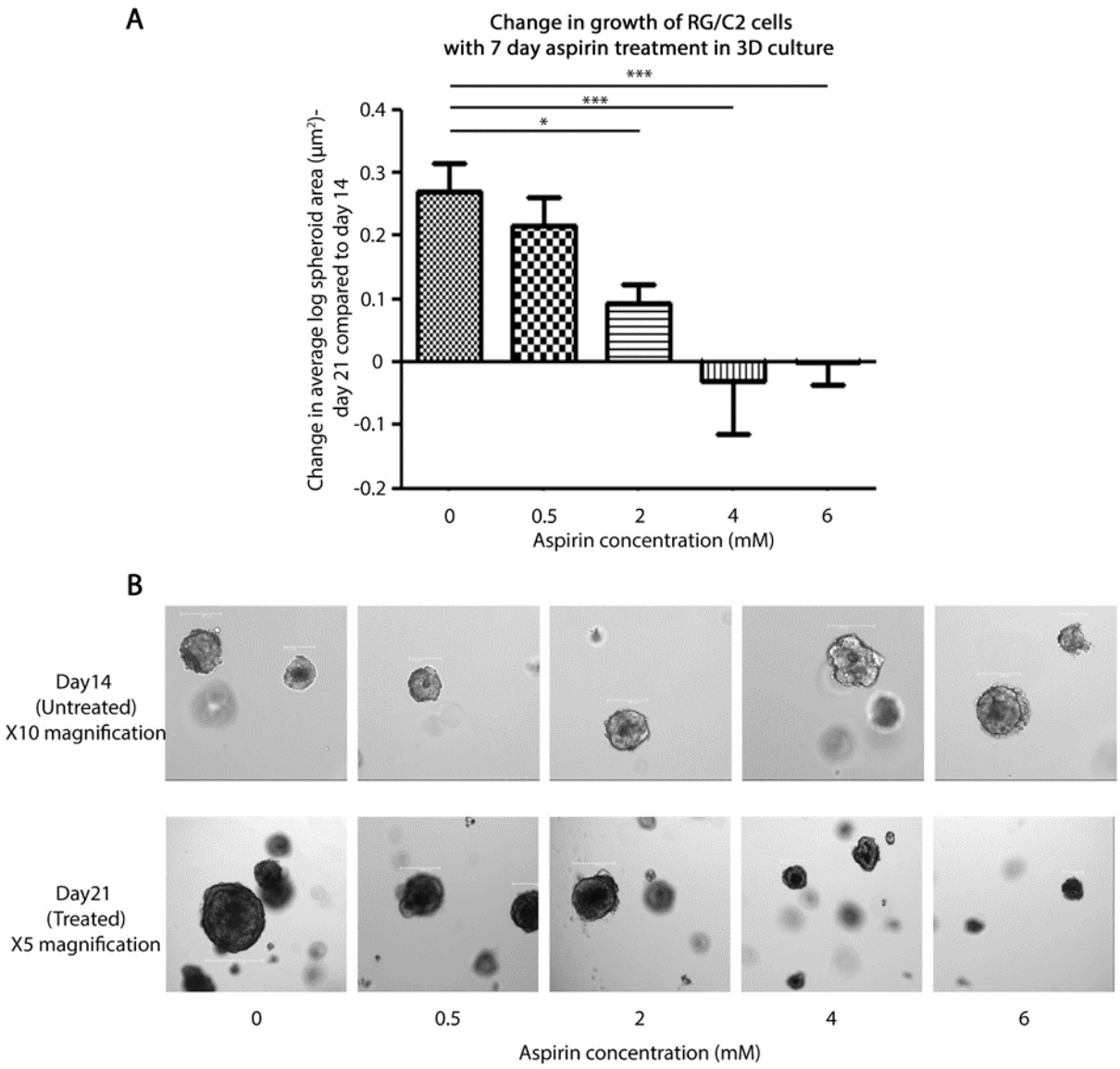


Figure 3-5- Aspirin decreases the growth of RG/C2 cells grown in 3D culture

RG/C2 cells were grown as 3D spheroids for 2 weeks and then treated with 0, 0.5, 2, 4 and 6mM aspirin for 7 days. Images were taken at days 14 and 21 and average spheroid areas were calculated for each condition. In order to carry out statistical tests, data was converted to the log scale to achieve a normal distribution. (A) Aspirin significantly slows the growth of RG/C2 cells at 2mM ($p \leq 0.05$), 4mM ($p \leq 0.001$) and 6mM ($p \leq 0.001$). Results are log mean values from 3 repeat experiments with standard deviations and a one-way ANOVA and Dunnett's multiple comparison test to calculate significance (*** $p \leq 0.001$, **$p \leq 0.01$ and * $p \leq 0.05$). (B) Widefield microscopy images of untreated RG/C2 spheroids at day 14 at x10 magnification and images of spheroids at day 21 treated with aspirin for 7 days at x5 magnification. The images show that increasing the aspirin concentration slows down the growth rate of the spheroids, especially at 4mM and 6mM.

3.3.1.5 Aspirin does not increase the percentage of dead cell staining in 3D cultures

To investigate the effect of short term aspirin on apoptosis in 3D, we adopted a live/dead cell staining approach of 3D cultured cells using calcein-AM to stain live cells and ethidium homodimer to stain dead cells. Live cells emit a green fluorescence and dead cells emit a red fluorescence. Therefore, RG/C2 cells were cultured for 14 days as 3D spheroids and then treated with aspirin for 72 hours before staining. We used 4mM aspirin since it appears to cause maximum inhibition of cell growth (Figure 3-5). Images were taken

using a confocal microscope and then analysed using Imaris x64 version 7.6.5 to calculate the volume of live and dead cells.

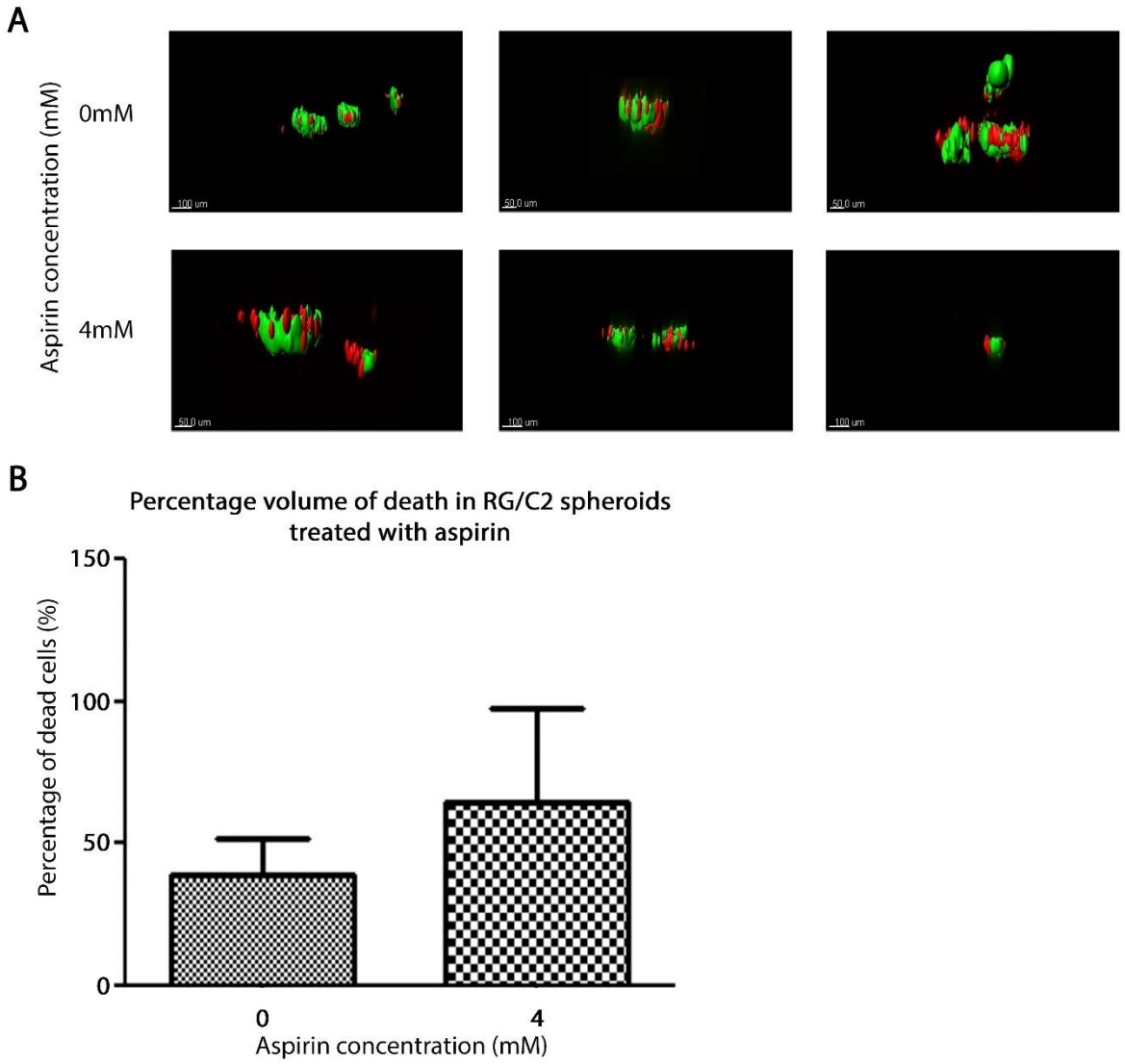


Figure 3-6- Live/dead cell staining of spheroids show that aspirin has no significant effect on apoptosis.

RG/C2 cells were grown as 3D spheroids for 2 weeks and then treated with 0 and 4mM aspirin for 72 hours. (A) Live cells were stained with calcein-AM (green fluorescence) and dead cells were stained with ethidium homodimer (red fluorescence). Images were taken with a confocal microscope and are 3 representative images selected at random for each treatment shown. (B) Volumes of green and red spheroids were analysed and the percentage volume of red (dead cells) within the spheroid was calculated. There is no difference in percentage of dead cells between aspirin treated and untreated spheroids. Results are mean percentages from 3 repeat experiments with standard deviations and Student's t-test.

The images in Figure 3-6 A show that calcein-AM and ethidium homodimer stain live and dead cells in spheroids. In order to better quantify whether there is an increase in cell death with aspirin, volumes of green and red cells were calculated and the percentage of dead cells in the spheroid was obtained by dividing the volume of red cells by the total spheroid volume. The results suggest a potentially clinically important effect size, however, after applying a Student's t-test, our results show that aspirin does not

significantly increase the percentage of apoptotic cells in 3D culture (Figure 3-6 B) similar to that observed in 2D culture(Figure 3-3 C). In the case of this example, it may be that more repeats would have improved the significance of the results, indicating that our results may not have had enough power to identify a significant effect of aspirin on apoptosis in 3D cell culture.

Overall, our results show that both in 2D and 3D cell culture, cell growth was reduced and we confirmed that this was not through apoptosis. We did, however, observe that cells were more sensitive to aspirin treatment as spheroids. For this reason, we decided to continue using 3D models to investigate the effect of long-term aspirin treatment on adenoma cells.

3.3.2 Long -term aspirin treatment and its affect on cell growth in 3D culture

We have shown that short-term aspirin treatment (up to 7 days) is able to significantly reduce cell growth at 4mM aspirin. Epidemiological studies have shown that pre and post-diagnosis aspirin intake may impact how effective the drug is at reducing the risk of CRC (Chapter 3.1.1). This suggests that long-term aspirin exposure may result in a phenotypic change in the cells and raises the possibility that previous long-term aspirin treatment renders the cells less sensitive to its protective effects.

3.3.2.1 Long-term aspirin treatment does not significantly affect spheroid growth in 3D cultures

In order to model long-term exposure, adenoma cells were cultured in 2D for a total of 18 weeks with bi-weekly dosages of aspirin. Cells were then harvested and seeded into Matrigel© for growth as spheroids. Cells were grown for a total of 14 days during which aspirin treatment was continued. Average spheroid area for each condition was calculated and the data were converted to a log scale to achieve a normal distribution and therefore carry out statistical tests which require a normal distribution to avoid violating assumptions. A one-way ANOVA and Dunnett's multiple comparison test to calculate significance were applied.

By looking at average log spheroid area on day 14, the results show that aspirin has no effect regardless of the dose (Figure 3-7 A). Figure 3-7 B also shows that long-term aspirin culture has no effect on the rate of cell growth in 3D culture. There is no difference between the change in average spheroid size between 0mM (0.61, SD 0.14) and 4mM (0.48, SD 0.20) at day 14 compared to day 7. Both results indicate that cells have adapted to aspirin treatment and therefore become "resistant" to its inhibition of cell cycle progression that was observed in experiments with short-term aspirin addition (Chapter 3.2). This suggests that cells become less sensitive to aspirin when cultured long-term either indicating an acquired phenotypic change or that aspirin is selecting for cells that are "resistant" to its effects.

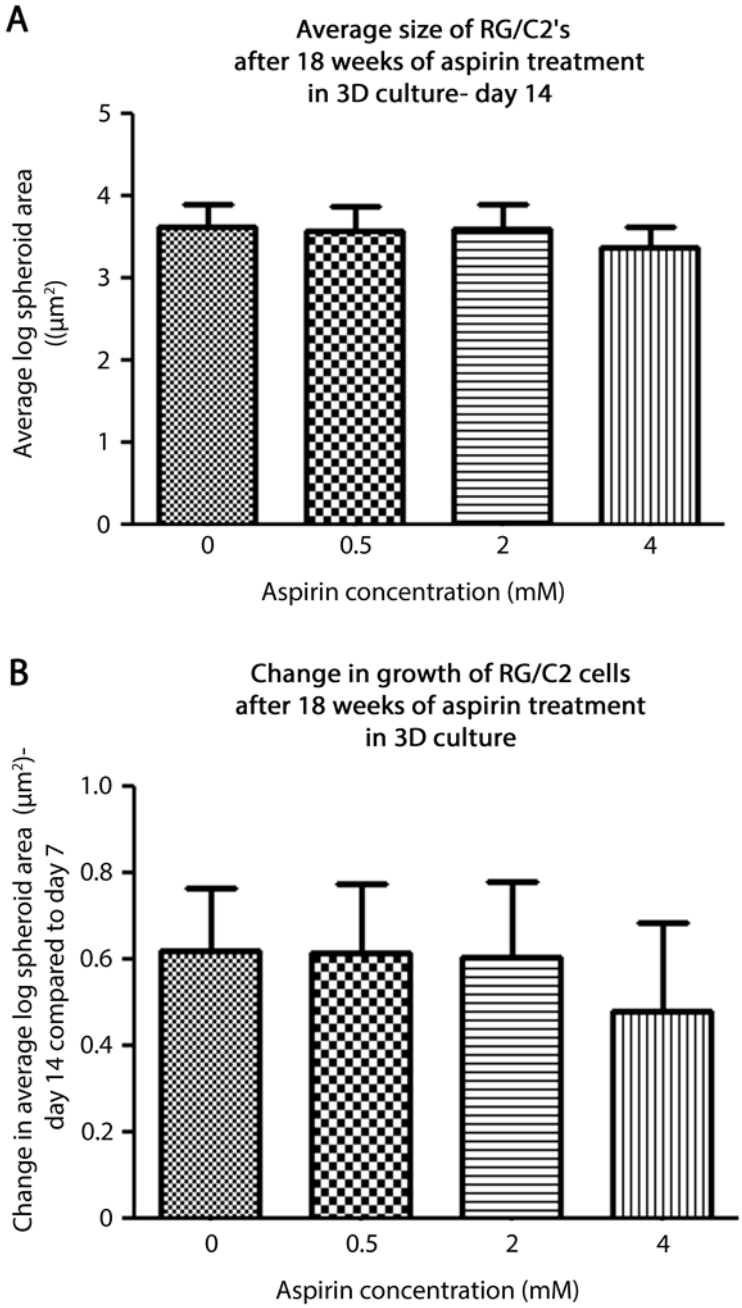


Figure 3-7- Long-term aspirin treatment has no significant effect on cell growth in 3D culture.

RG/C2 cells were grown with 0, 0.5, 2, 4mM aspirin for 18 weeks in 2D culture. Cells were then grown in Matrigel© for 14 days. Images were taken at day 7 and day 14 and areas of the spheroids were calculated. In order to carry out statistical tests, data was converted to the log scale to achieve a normal distribution. Results are log mean values from 3 repeat experiments with standard deviations and a one-way ANOVA and Dunnett's multiple comparison test to calculate significance (*** $p \leq 0.001$, **$p \leq 0.01$ and * $p \leq 0.05$). (A) Aspirin has no significant effect on the average spheroid area at day 14. (B) There is no significant change in growth between the different concentrations therefore aspirin does not affect the rate of spheroid growth.

3.3.2.2 Removal of long-term aspirin treatment from RG/C2s has no significant effect on their growth

Our results indicate that cells adapt to long-term culture with aspirin. To identify whether this is a permanent phenotypic change, we investigated the effect of removing aspirin from cells previously grown with long-term aspirin treatment.

RG/C2 cells were cultured for 28 weeks with 0, 0.5, 2 and 4mM aspirin for 28 weeks in 2D culture. Cells were then harvested and grown aspirin-free in Matrigel © for a total of 21 days. Images were taken at days 4, 7, 11, 14 and 21 and average spheroid areas for each condition was calculated. The data was converted to a log scale to achieve a normal distribution and therefore carry out statistical tests. A one-way ANOVA and Dunnett's multiple comparison test to calculate significance were applied.

Our results show that removing aspirin from previously long-term treated cells has no significant effect on the rate of cell growth in 3D culture (Figure 3-8 A).The results further show that removal of aspirin from cells previously treated for 28 weeks has no effect on the change in spheroid area at day 21 compared to day 4 indicating that the speed of their growth is not affected by the removal of aspirin (Figure 3-8 B). To further interrogate this, we compared the average spheroid area of each previously treated aspirin condition (0.5, 2 and 4mM) to untreated cells at day 21 and found that removing aspirin has no significant effect on spheroid area (Figure 3-8 C).

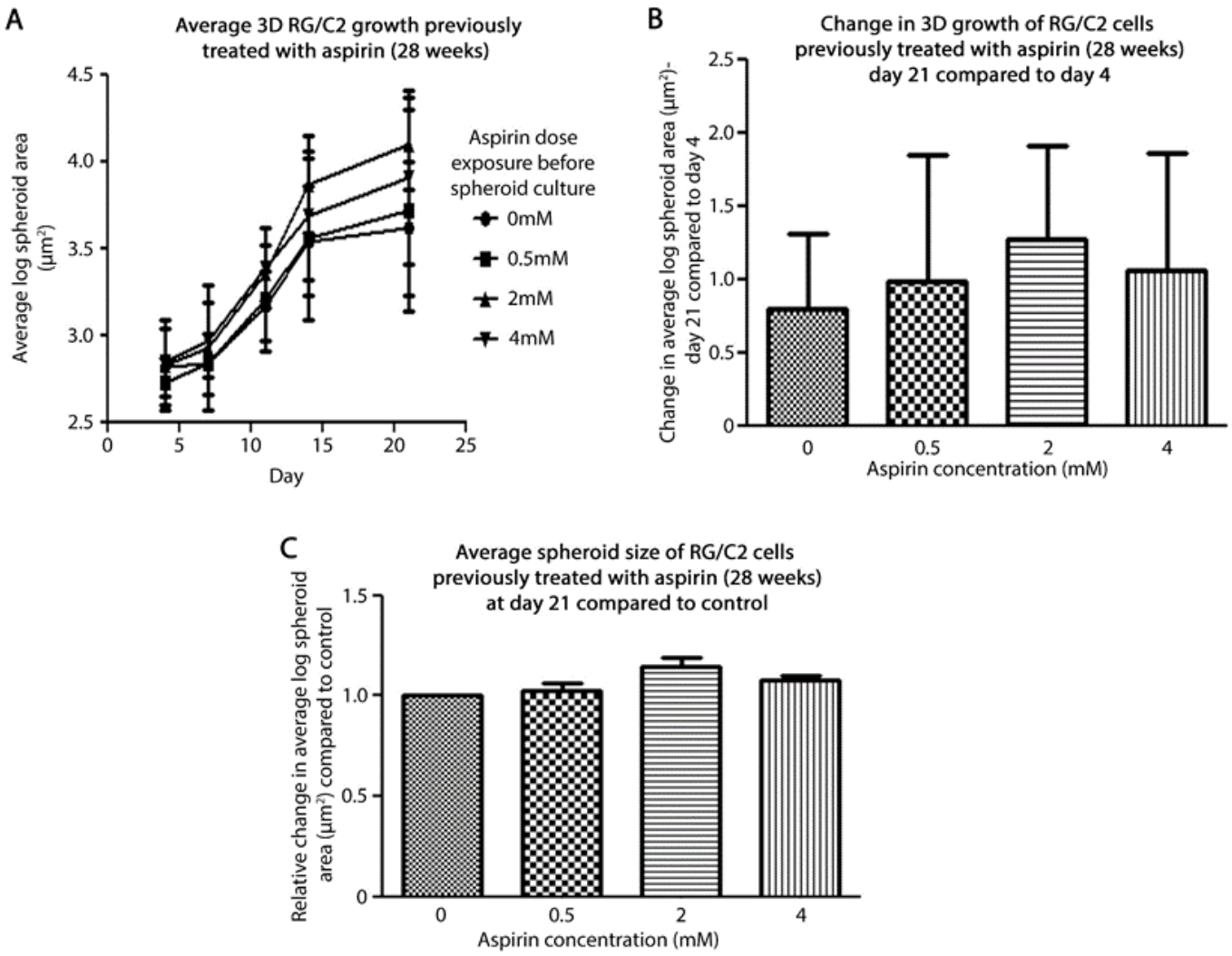


Figure 3-8- The removal of aspirin from RG/C2s previously on 28 weeks of aspirin in 2D has no significant effect on the size of spheroid areas in 3D culture.

RG/C2 cells were cultured with 0, 0.5, 2 and 4mM aspirin for 28 weeks in 2D culture. Cells were then cultured in Matrigel© 3D with no aspirin treatment for a total of 21 days. Images were taken at days 4, 7, 11, 14 and 21 and areas of the spheroids were calculated. To carry out statistical tests, data was converted to the log scale to achieve a normal distribution. Results are log mean values from 3 repeat experiments with standard deviations and a one-way ANOVA and Dunnett's multiple comparison test to calculate significance (*** $p \leq 0.001$, ** $p \leq 0.01$ and * $p \leq 0.05$). (A) The change in average spheroid area between day 21 and day 4 was calculated and shows that removing aspirin has no significant effect on the speed of spheroid growth. (B) A comparison of spheroid areas relative to untreated cells shows that removing aspirin treatment has no significant effect on spheroid area. (C) The average spheroid areas for each aspirin condition was plotted against days 4, 7, 11, 14 and 21. These results show that there is no significant effect of removing aspirin from previous long-term treated cells on their growth in 3D.

Overall, these results indicate two possible/interlinked consequences of long-term aspirin exposure to cells. Either the aspirin treatment has selected for cells "resistant" to its effects or cells have acquired a phenotypic change to adapt to growth with aspirin (Figure 3-9). Both of these phenomena would be part of a process of selection that resulted in an "acquired resistance" of a cell population.

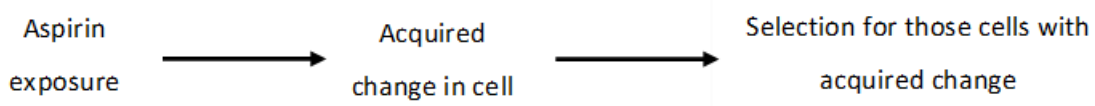


Figure 3-9- Acquisition of aspirin resistance

Long-term aspirin exposure may result in cells acquiring phenotypic changes. The result is the selection of cells with these acquired changes.

3.4 Discussion

Epidemiological studies have shown that aspirin is indeed effective at reducing the risk of CRC. Results of somecohort studies have shown that aspirin intake after diagnosis is more effective than pre-diagnosis indicating that shorter durations of aspirin treatment after diagnosis are more effective than pro-longed aspirin treatment pre-diagnosis (276–278,282,283). However, results from other groups suggested otherwise (279,280). We investigated the short-term and long-term effects of aspirin within *in vitro* cultures to investigate the issue of the potential differing influences of time-scale of drug administration.

Our results show that short-term aspirin treatment in 2D cultures significantly inhibits cell growth through cell cycle arrest but does not induce apoptosis. Luciani et. al showed that aspirin induced cell death mainly in cells that expressed p53 (147).This may explain why induction of cell death by salicylate in RG/C2s was not significant (299), as they are hemizygous at the *p53* locus and their remaining allele is mutant. However Elder et. al also showed that induction of cell death was not significant in PC/AA adenoma cells (299) and these express wild-type p53 (301). Therefore, the available evidence supports the explanation that aspirin induces growth arrest of p53 +/- adenoma cells but not apoptosis.

Our results also show that RG/C2 cells are more sensitive to lower aspirin doses in 3D cultures. Cancer cell response to drug therapy in 3D culture can cause cells to either be more sensitive or more resistant to treatment than 2D cultures (302). Howes et. al (2014) found that microtubule-targeting agents and inhibitors of epidermal growth factor receptor (EGFR) were more cytotoxic to breast cancer cells in 3D spheroids compared to cells grown as monolayers akin to our observation of 3D culture sensitising cells to lower doses of aspirin (303). It is becoming increasingly clear that drugs that require pre-clinical testing should be carried out in 3D as well as 2D models as they better inform of drug potency and therefore may improve the success of new drugs being tested in clinical trials (302).

Long-term aspirin treatment in 3D culture shows that aspirin has no significant effect on spheroid growth or apoptosis. This indicates that cells either adapt to conditions (“acquired resistance”) with aspirin and/or that aspirin is selecting for cells “resistant” to its effects. Drug resistance is not a new phenomenon with regards to cancer. Cree et. al (2017) summarised 6 methods for drug resistance: (1) a mutation in drug targets, (2) removal of drugs through expression of drug pumps, (3) expressing new detoxification methods, (4) reduced

sensitivity to apoptosis, (5) a better ability to repair damage caused to DNA and (6) a change in cell proliferation (304). In the context of aspirin, it may be that cells become "resistant" through adopting any one of these methods or that we are specifically selecting for cells that have better mechanisms of e.g. drug removal through expression of pumps or cells that express COX with genetic polymorphisms that prevent aspirin acetylating and inactivating COX's active site. The effect of long-term aspirin use on adenoma cells requires further experiments and characterisation.

If long-term aspirin treated cells are acquiring "resistance", we would expect a "rebound" effect once the aspirin is removed. We expected cells previously treated long-term with aspirin to grow at an increased rate compared to untreated once the aspirin is removed. Our results showed that removal of aspirin treatment from cells previously treated for 28 weeks shows no significant results in the context of spheroid growth. This suggests that aspirin treatment is selecting for genetically "resistant" cells as opposed to cells acquiring "resistance".

Whilst no results have been published with regards to the effect of discontinued aspirin use on the risk of CRC, it may be that aspirin removal can increase risk of cancer. It may be that we observed no significant effect of discontinued aspirin treatment after previous 28 weeks of treatment as this time scale was not enough for cells to revert to a sensitive phenotype. It may be that the adenoma cells needed to be cultured with aspirin for a longer time period to observe any reversion from the removal of aspirin. In addition, it may be that we had selected for cells that are genetically "resistant" to aspirin and therefore the presence or absence of aspirin has no effect on their growth.

Whilst we postulate that discontinuation of aspirin use may result in an increased risk of cancer, results by McNeil et. al (2018) showed that aspirin increases risk of cancer-specific mortality in elderly individuals in a recent RCT for the effect of aspirin on disability-free survival (HR: 1.77, 95% CI 1.02-3.06) (281). These results contrast with previous literature that aspirin reduces risk of colorectal cancer incidence (Chapter 1.5). These results indicate the exact mechanism of aspirin "resistance" is still unknown and further research is required.

Overall, we have observed that short-term aspirin treatment does not induce apoptosis but rather inhibits cell cycle progression. This observation is no longer significant once cells have been cultured with aspirin long-term indicating that either aspirin is selecting for "resistant" cells or cells have undergone an acquired phenotypic adaptation to withstand the continued aspirin treatment (the removal of aspirin has no effect on cell growth in previously long-term treated adenoma cells). If this result was due to an acquired "resistance", we would expect a rebound effect of cells once aspirin is removed; however, no such effect is seen. This suggests that aspirin is selecting for genetically "resistant" cells. For this reason, there is a need to investigate the phenotypic changes of aspirin treatment to be able to identify whether cells are selected for or whether over time if aspirin removed cells revert their sensitivity. Therefore, markers of resistance or phenotypic changes need to be identified in order to answer this question.

Chapter 4 Combining 'omics to identify new potential targets of aspirin

This chapter (and the next) aim to address one of the possible mechanisms by which aspirin may have differing effects depending on short-term and long-term culture, of which we postulate may be due to DNA methylation. This chapter also seeks to identify possible new targets of aspirin through combining multiple 'omics (methylomic, transcriptomic and proteomic), circled in red in Figure 4-1.

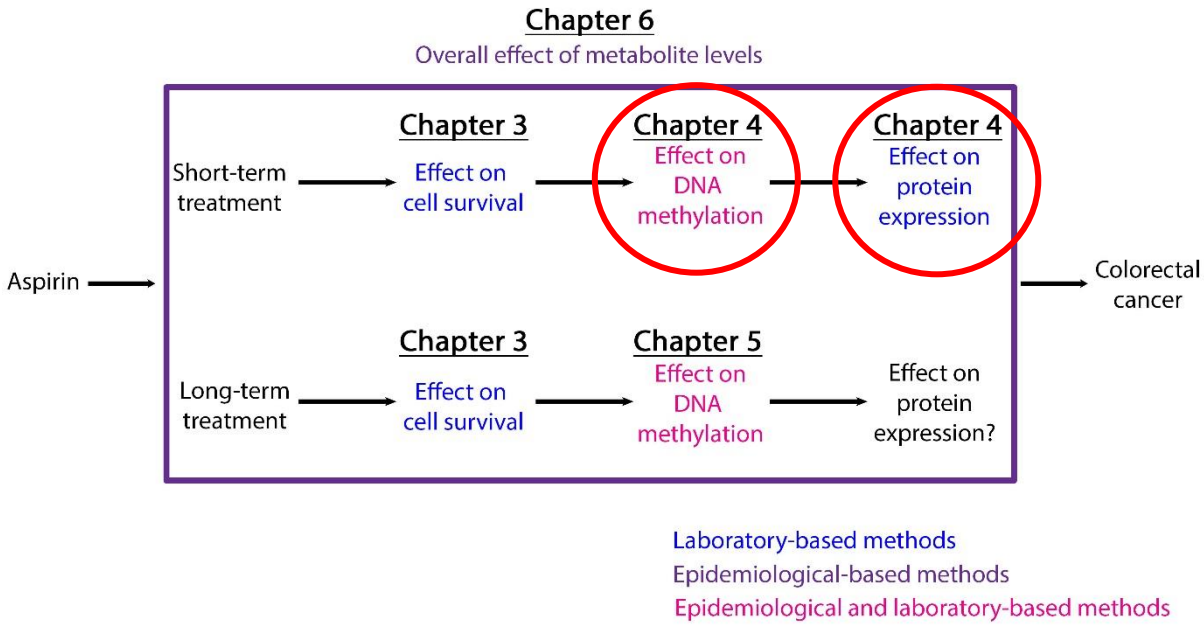


Figure 4-1- Overview of the questions addressed in this thesis to explore causality and potential mechanisms of aspirin on CRC.

4.1 Background

4.1.1 Evidence for COX-independent effects of aspirin

Aspirin is a well-established inhibitor of the pro-tumorigenic cyclooxygenase (COX)/prostaglandin 2 (PGE_2) signalling pathway reviewed in Greenhough et. al (2009) (305) and this is through the irreversible acetylation of the COX enzyme (306).

One of the earliest pieces of evidence for COX-independent effects of aspirin was published by Bak et. al (1998). They showed that NCX-4016, an aspirin derivative that does not inhibit COX, had a higher chemopreventative effect compared to aspirin in a mouse model of colon cancer indicating that aspirin may act through other mechanisms not yet known. This study showed that NCX-4016 reduced the mean number of aberrant crypt foci by 85% whereas aspirin reduced them by 64% (307).

Other evidence for the presence of COX-independent effects is presented by Waskewich et. al (2002) who tested the effect of a COX-2 selective inhibitor, celecoxib, on hematopoietic and epithelial cancer cell lines and found that it is most effective at inhibiting growth in COX-2 negative cells (308).

Together, these results and early studies that investigated aspirin's effect on other pathways provided evidence that aspirin also has COX-independent targets. The other targets of aspirin are summarised in Chapter 4.1.2.

4.1.2 COX-independent targets of aspirin

Other than just targeting the COX enzymes, aspirin can also target many other proteins. It has been shown to inhibit IκB kinase (IKK) and NF-κB *in vivo* and *in vitro* (309,310) as well as downregulating proteins such as c-myc, cyclin D1, cyclin A and proliferating cell nuclear antigen (PCNA) thereby causing cell cycle arrest (311). It has also been implicated in affecting the WNT/β-catenin signalling pathway through increasing the phosphorylation of protein phosphatase 2 A (PP2A). This inhibits PP2A's activity which then leads to an increase in phosphorylation of β-catenin resulting in reduced activity of the Wn/β-catenin pathway (312).

Aspirin has also been shown to increase the permeability of the mitochondrial membrane resulting in the release of cytochrome C thereby activating apoptosis (313–315). Furthermore, it has been shown that the aspirin breakdown product salicylate is able to inhibit the uptake of calcium by the mitochondria causing an anti-proliferative effect (316).

Research to identify new targets explaining aspirin's efficacy has increased. Some of the most recently identified targets are summarised in Table 4-1 along with the consequence of being targeted by aspirin. They include: E1A-associated protein p300 (EP300) (317), toll-like receptor 4 (TLR4) (145), homeobox protein NANOG (NANOG) (318), cyclin dependent kinase 1 (CDK1) (319), glucose-6-phosphate dehydrogenase (G6PD) (320), p53 (321) and signal transducer and activator of transcription 3 (STAT3) (322).

Table 4-1- Proteins recently identified to be targeted by aspirin and the consequence of their targeting.

Protein	Aspirin's effect on protein	Consequence	Reference
E1A-associated protein p300 (EP300)	competitively inhibits the bindning of acetyl coenzyme A (AcCoA)	EP300's acetyltransferase activity is inhibited so this inhibits acetylation of other proteins. The end result is induction of autophagy.	(317)
Toll-like recetor 4 (TLR4)	Aspirin reduces TLR4 expression	Blocks lipopolysaccharide (LPS) from gram-negative bacteria from activating TLR4 which would usually result in cancer cell migration and epithelial-mesenchymal transition (EMT).	(145)
Homeobox protein NANOG (NANOG)	Reduces protein expression and stability	Inhibiting NANOG may inhibit cancer stem cells thereby inhibiting tumour growth and stemness	(318)
Cyclin dependent kinase 1 (CDK1)	Aspirin acetylates Lys34 of this protein	Unknown whether this affects its activity	(319)
Glucose-6-phosphate dehydrogenase (G6PD)	Aspirin acetylates 14 lysine residues in this protein	Inhibition of G6PD activity resulting in a reduction of ribose sugars and NAPDH production	(320)
p53	Aspirin acetylates p53	Unknown effect on protein function and structure	(321)
Signal transducer and activator of transcription 3 (STAT3)	Aspirin induces the transcription of the long non-coding RNA OLA1P2	Blocks phosphorylated STAT3 thereby reducing the risk of cell growth and metastasis	(322)

4.1.3 Aspirin and post-translational modifications: protein lysine acetylation

Protein acetylation involves the transfer of an acetyl group from acetyl coenzyme A to a specific amino acid in the protein resulting in a post-translational modification (323). Protein acetylation occurs in two forms- the first being the acetylation of N-termini from newly formed polypeptide chains (324) and the second being the acetylation of lysine residues either in histones or other proteins (325).

Aspirin has been shown as far back as 1968 to acetylate proteins, enzymes and DNA (188). Aspirin has the ability to acetylate proteins directly, such as the COX-1 and COX-2 enzymes, as well as indirectly influencing acetylation (Chapter 4.1.4). Aspirin is also able to acetylate many proteins that interact with platelets. These include fibrinogen, albumin and haemoglobin (188,326,327). Other proteins also targeted for acetylation include p53 and glucose-6-phosphate dehydrogenase (320,328).

4.1.4 Aspirin and post-translational modifications: histone lysine acetylation

Recent evidence has shown that aspirin has the ability to acetylate histones and therefore alter epigenetics. Passacquale et. al (2015) treated human umbilical vein endothelial cells (HUVECs) with both tumour necrosis factor α (TNF-α) and aspirin to investigate their effect on histone acetylation, more specifically, histone H3 acetylation. They found that TNF-α alone did not change H3 acetylation levels but in the presence of aspirin, there was a higher amount of acetylation compared to controls. They also found that aspirin and salicylic acid increased the activity of histone acetyltransferases (HATs) and reduced the activity of histone

deacetylases (HDACs) in HUVECs treated with TNF-α (190). Their results lend evidence to the idea that as well as directly acetylating proteins, aspirin also affects the activity of enzymes involved in acetylation.

Conversely, Fernandez et. al (2017) found that sodium salicylate decreases the activity of lysine acetyltransferases (KATs) in *in vitro* assays. They also observed that increased sodium salicylate concentrations resulted in a decrease of acetylated histones (329).

Guo et. al (2016) investigated the effect of aspirin on histone H3 lysine 27 acetylation (H3K27ac). In general, histone acetylation opens the chromatin allowing it to be accessible to transcription factors and allow gene transcription (153) therefore H3K27ac is an active chromatin mark. In this study, the mice were treated with dextran sulfate sodium used as an inflammatory stimulus and azoxymethane to initiate colitis-accelerated colon carcinogenesis. It was found that co-treating these mouse models with aspirin perhaps surprisingly resulted in the decrease in protein expression as well as inhibition of HDACs, consequently resulting in a higher percentage of H3K27ac positive cells compared to untreated. In contrast, they also showed that aspirin was able to reduce H3K27ac at the promoters of inflammatory markers such as inducible nitric oxide synthase (iNOS), tumour necrosis factor alpha (TNF-α) and interleukin 6 (IL-6) thereby reducing their gene expression (189).

Tatham et. al (2017) undertook a proteomic approach to look at the effect of aspirin on the lysine acetylome in HeLa cells. They identified 12069 lysine sites acetylated by aspirin in 3763 proteins. Their results show that of the 12069 lysine sites targeted by aspirin only 10% of these were within histone sites. However, the histone sites targeted by aspirin can experience a more than 10 fold increase in acetylation occupancy and some are over 100 fold more acetylated (78). Their results further confirm that aspirin has the ability to acetylate histone proteins hence having a global role on transcription.

Overall, aspirin has the ability to influence protein and histone acetylation directly or through affecting the activity of enzymes involved in acetylation or deacetylation.

4.1.5 Aspirin and DNA methylation

The field of aspirin and its effect on DNA methylation is relatively unknown. Those that have explored this focused their attention on the association rather than the effect of aspirin on a subset of CpG sites from the CpG Island Methylator Phenotype (CIMP). CIMP refers to methylation events that occur in a specific subset of adenomas and CRC therefore being involved in the pathogenesis of the cancer (49). Slattery et. al (2006) examined the association of aspirin with methylation of 5 CIMP CpG sites including *MINT1*, *MINT2*, *MINT31*, *p16* and *hMLH1* and found that aspirin was not associated with CIMP positive tumours (330).

Other studies have focused on how the methylation of certain sites impacts aspirin absorption. Li et.al (2017) identified that hypomethylation of *ABCB1*, a protein involved in aspirin absorption in the intestines, was associated with a lower absorption of aspirin (331).

Whilst few studies have interrogated the effect of aspirin on genome-wide promoter CpG methylation, Noreen et. al (2014) investigated the effect of aspirin on age-associated DNA methylation in normal colon biopsies using the Infinium HumanMethylation27 Beadchip arrays that identifies the methylation state of 27,578 CpGs. They found that aspirin intake suppressed the age-dependent hypermethylation of CpG sites and that the methylation retardation effect of aspirin was more pronounced in individuals with a high body mass index (BMI)/ than those with a normal BMI. They also analysed the sites that were hypermethylated with both age and CRC and found aspirin to suppress their methylation (249).

To our knowledge, there have been no published studies on the epigenome-wide effect of aspirin on CpG methylation using the more recent platforms such as the Infinium HumanMethylation450 BeadChip® and the Infinium MethylationEPIC BeadChip (Illumina, Inc. CA, USA) which interrogate methylation at over 480,000 and 850,000 CpG sites, respectively. The use of these platforms allows the cost-effective interrogation of a large number of CpG sites across the genome therefore allowing the study of aspirin's effect on DNA methylation over 850,000 CpG sites.

4.1.6 Interplay between acetylation and DNA methylation

It has long been believed that actively transcribed regions of CpG islands are unmethylated, lack histone H1 and contain hyperacetylated H3 and H4 (332). Rose and Klose (2014) stated that DNA methylation results in many effects to chromatin of which include histone deacetylation, methylation and a compact chromatin structure (333).

Whilst the exact relationship between DNA methylation and histone modifications is complex, the literature indicates that correlations between DNA methylation and histone acetylation exist. In the context of cancer, Sheffield et. al (2017) investigated the relationship between DNA methylation and 7 histone marks in 3 Ewing sarcoma tumours and found a consistent inverse relationship between DNA methylation and H3K27 acetylation (334). Other evidences for an inverse correlation between these epigenetic modifications is provided by Heyn et.al (2016) who investigated the relationship between DNA methylation and H3K27 acetylation at super-enhancer regions in normal breast tissue and found an inverse correlation- super-enhancers that exhibited low DNA methylation had a high level of H3K27 acetylation and vice versa (335). More recently, González et. al showed that mice treated with methamphetamine (a psychostimulant) resulted in a decrease in levels of histone 3 and 4 acetylation and an increase in CpG methylation (336).

Whilst all of these studies are unrelated to aspirin, their results show an inverse relationship between histone acetylation and DNA methylation. Since Guo et. al (2016) showed that aspirin treatment increased the acetylation of H3K27 in mouse models of colon cancer (189), and studies have shown that aspirin influences protein and histone acetylation, this led us to hypothesise that aspirin may also be altering (reducing) genome-wide DNA methylation.

4.1.7 Methods for assessing DNA methylation

4.1.7.1 Measuring changes in DNA methylation

There are many methods to measure DNA methylation, of which the most comprehensive method is whole genome bisulfite sequencing to understand the complete methylome. However, this approach is only realistic when processing a small number of samples as it is very costly and time-consuming. The Infinium HumanMethylation450 BeadChip® (450K array; Illumina, Inc. CA, USA) allows interrogation of more than 480,000 CpG sites and is useful for many reasons such as reagent costs, time of labour and its ability to interrogate multiple samples (337).

Through the incorporation of labelled nucleotides, the Illumina methylation arrays have been designed so that methylated CpG sites emit a different fluorescence to unmethylated CpG sites. The intensity of the fluorescence is measured and a ratio of the fluorescent signal of the methylated site to the signal from the unmethylated and methylated combined is calculated (338). This ratio is called beta (β). It's values range from 0 to 1 and roughly represents changes in percentage methylation where 0 is equal to 0% and 1 is equal to 100% (339). These beta values are absolute values and when a comparison between samples or groups is made they can be positive to represent an increase in methylation or negative to represent a decrease in methylation. These beta values allow for direct biological interpretation of results, unlike some other methods (340).

Major changes (>60%) in DNA methylation can result in diseases by affecting gene repression or expression. For more complex diseases, a more subtle change in methylation (<10%) can be observed (341). According to a review by Michels et. al*,* it has been suggested that methylation changes of <5% should be interpreted with caution but changes of >10% provide increased confidence in the biological significance of the result (342). Other criteria for identifying methylation changes that affect biological pathways are summarised in Table 4-2. In the context of this study, we aim to identify CpG sites with a change of at least 10% in DNA methylation by aspirin treatment.

Table 4-2- Criteria for identifying methylation changes that affect biological pathways- adapted from Michels et. al (342)

Criteria	Confidence that methylation difference mediates biological pathway	
	Increase confidence	Decrease confidence
Statistical significance	Reaches genome-wide significance	Does not meet predefined significance threshold that takes into account multiple testing
Effect size (difference in methylation)	Large (>10% difference)	Small (<5% difference)
Bias and confounding	Bias and confounding are prevented by design or controlled for in the analyses	Bias or uncontrolled confounding may exist and explain the differences observed
Genomic location	Differential methylation is in a region that may impact regulation of transcription	Current knowledge cannot explain the influence of the observed difference in methylation at that locus on regulation of transcription
Functional relevance	Affects expression	Does not affect expression
Biological relevance	Gene codes for known biological function	Biological relevance of DMR location unknown or unrelated to phenotype
Validation	Findings are replicated in an independent human cohort or animal model using a different technique	No validation of results attempted or results are not replicated in a validation study

4.1.7.2 Adjustment for multiple testing

One of the issues with increasing the number of tests and using the conventional P-value cut-off of 0.05 is that there is an increasing chance of finding false positives. For example, if 10000 CpG sites were tested for association with aspirin treatment, by chance 500 of these sites would be associated with aspirin. However, it is highly unlikely that all of these CpG sites are truly associated with drug treatment. Methods exist to reduce the chance of these false positives and these include the Bonferroni correction and the false discovery rate adjusted P-value. The Bonferroni correction divides the alpha level by the number of statistical tests being carried out, so if 382,486 CpG sites were being tested for significance, the Bonferroni corrected P-value would be $0.05/382,486 = 1.31 \times 10^{-7}$. Therefore, if a CpG site association with aspirin is below this P-value, this provides stronger evidence that this association is not due to chance. Whilst it is the most conventional and simple method, it tends to be overly conservative, thereby rejecting associations that may be true and also has lower power (272,273).

An alternative method to the Bonferroni correction, is the false discovery rate (FDR) adjusted P-value. The false discovery rate describes the probability at which you discover, for example, a SNP association with an exposure which is untrue and has simply occurred due to chance, therefore it is a false positive (273). In the context of multiple comparisons, Benjamini & Hochberg (1995) termed an FDR-adjusted P-value which sets a limit on the false discovery rate (343). When an FDR-adjusted P-value of 0.05, this informs us that the result

actually being a false positive from the null associations in the dataset is 5%. Therefore, the lower the P-value, the less likely the result is a false positive (344).

4.1.8 Hypothesis and aims:

Hypothesis:

Short term aspirin treatment induces phenotypic changes in cells although the underlying mechanisms for this are not yet fully understood. These have the potential to be used to identify anti-cancer function of aspirin.

Aims:

1- To compare genome-wide DNA methylation levels in RG/C2 cells following short-term aspirin treatment
2- To combine methylomic, proteomic and transcriptomic data to identify new targets of aspirin action
3- Validation and functional analyses of targets of aspirin action
4- To confirm aspirin's effect on DNA methylation through pyrosequencing validation

4.2 Methods

4.2.1 SILAC proteomic analysis

Previously in our laboratory, a stable isotope labelling with amino acids in cell culture (SILAC) approach was carried out on RG/C2 cells treated with aspirin to obtain proteomic data. Control cells (0mM aspirin) were cultured with an L-arginine and L-lysine (light labelling), 2mM were cultured with $^{2}H_{4}$-lysine and $_{13}C_{6}$-arginine (medium labelling) and 4mM were cultured with $^{15}N_{2}{}^{13}C_{6}$-lysine and $^{15}N_{4}{}^{13}C_{6}$-arginine (heavy labelling) (Cambridge Isotope Laboratory, Massachusetts, United States). These methods were informed by the SILAC-based mass spectrometry approach by Trinkle-Mulcahy et. al (2008) (345).

Cells were cultured with aspirin and the isotopes for 24 hours before extracting protein lysates. This experiment was carried out in duplicate. Details of SILAC labelling and proteomics have been previously published (346).

4.2.2 Transcriptomic analysis

A published transcriptomic dataset generated by Sabates-Bellver et. al (2007) that compared mRNA levels of normal and colorectal adenoma samples from 32 patients was used (347). This was obtained from GEO DataSets (348).

4.2.3 DNA methylation methylomic analysis

4.2.3.1 Short-term aspirin treatment

The exposure variable in this analysis was the short-term treatment (24 hours) of RG/C2 cells with aspirin. RG/C2 cells were cultured in 2D with 0mM, 2mM and 4mM aspirin in technical duplicates with 5 biological repeats resulting in a total number of 30 samples, of which 22 samples had sufficient DNA for the analysis.

Aspirin was considered a continuous exposure variable when comparing the effect of 0mM, 2mM and 4mM therefore identifying the change in methylation per 2mM increase of aspirin treatment. Aspirin was also considered a binary exposure when comparing untreated (0mM) to treated (2mM and 4mM) cells.

4.2.3.2 DNA methylation

The main outcome measure in this analysis was the level of DNA methylation at each of the CpG sites in the 450K array for RG/C2 cells treated with 0mM, 2mM and 4mM aspirin. Laboratory methods, QC and pre-processing of the methylation assay was described in Chapter 2.10. Of the 22 samples, 11 samples failed quality control (>10% probes with a detection P-value ≥ 0.01); however, these were not excluded due to the small sample size as a smaller sample size reduces power and makes it more likely that a null association is detected when actually a true association exists between aspirin treatment and effect on CpG site methylation (349).

4.2.3.3 Other variables

The bisulphite conversion batch for each sample was the only adjustment made in this analysis. Due to the controlled nature of the experiments it was not necessary to consider any other covariates.

4.2.3.4 DNA methylation analysis- multivariable regression analysis

Using DNA samples obtained from RG/C2 cells, methylation at 485,512 CpG sites was measured using the 450K array. We conducted multiple linear regressions with the exposure defined as aspirin treatment as a continuous variable (0mM, 2mM and 4mM) or aspirin treatment as a binary variable (untreated vs. treated) and the outcome defined as methylation (untransformed β-values). Analyses were run with adjustment for batch effects only.

DNA methylation sites were annotated using the Meffil package (264). A thorough assessment by Naeem et. al (2014) identified probes that may not be accurate (350). We therefore removed these sites as well as sites on sex chromosomes from the analysis. Therefore, the total number of CpG sites assessed in our analysis was 382,486. CpGs considered to be "EWAS-significant" were those which were below the Bonferroni correction calculated using the total number of tests performed ($P=0.05/382,486 = 1.31x10^{-7}$) (Chapter 4.1.7.2). CpGs with values below this were considered true positives and worthy of further investigation. The limitation of the Bonferroni correction is that it assumes that all CpG sites are independent of each other when this may not be true. For this reason, we also applied a less conservative statistical threshold called the

false discovery rate (FDR) which is based on the method by Benjamini-Hochberg also used to adjust for multiple testing (351). Using this method, CpG sites with FDR below a 0.05 threshold were considered EWAS-significant.

4.2.4 Power calculations

Power calculations were carried out using the statistical software G*Power 3.1.9.2 (352). Our study had 80.8% power to detect a mean methylation difference between aspirin treated and untreated samples of 0.1 beta (or 10% change in methylation) with a standard deviation of 0.1 for 20 treated samples vs 10 untreated samples at an α of 0.05.

4.3 Results

4.3.1 Linear regression results of aspirin use as a binary variable showed a higher effect on DNA methylation than aspirin use as a continuous variable

Our experiments have shown that aspirin inhibits cell growth and does not induce apoptosis. Whilst aspirin is a well-established inhibitor of COX2 (134), it is becoming increasingly obvious that it is a pleiotropic drug that has many targets. In order to identify targets that may explain the anti-cancer effect of aspirin, we decided to carry out a multi-omic approach combining methylomic data through DNA extraction of RG/C2 cells treated with aspirin for 24 hours, proteomic data previously available from our laboratory and publicly available transcriptomic data.

DNA was extracted from RG/C2 cells treated with 0, 2 and 4mM aspirin for 24 hours. DNA methylation was measured using the Illumina Infinium HumanMethylation 450K BeadChip assay. Given that aspirin treatment was administered at 3 doses and we had no prior evidence for a continuous vs. a threshold effect on aspirin DNA methylation, we initially compared the use of exposure data in a binary and continuous way. Linear regressions of the effect of aspirin as a continuous variable (0, 2 and 4mM) and as a binary variable (aspirin treated vs untreated) on the effect of DNA methylation were run. Results were plotted as waterfall plots to present the highest changes (both increase and decrease) in DNA methylation associated with treatment (Figure 4-2). Results were also plotted as Manhattan plots using the qqman package (353) to show the P-values of each CpG site aspirin association (Figure 4-3).

The results from our methylation analysis show that the changes in methylation β values from aspirin use as a binary variable (treated vs untreated) have a more pronounced effect on methylation levels compared to the regression analysis that compares methylation levels per 2mM aspirin. The waterfall plots in Figure 4-2 show that the highest change in methylation per 2mM increase of aspirin ranges between a decrease of 0.20 (20%) and an increase of 0.2 (20%). The range of β values from treated vs untreated is higher showing the highest increase in methylation as 0.35 (35%) and the highest decrease as 0.37 (37%).

In general statistics, for a result to be considered statistically significant, an alpha level of 0.05 is adopted, although many consider significance to be an arbitrary cut-off and other parameters should be used to judge the weight of evidence of any association. Therefore, if the P-value is less than 0.05 then it is below the alpha level and considered to be statistically significant and the null hypothesis is rejected. For results of an EWAS to be considered statistically significant, they should pass a correction threshold for multiple testing. When carrying out multiple statistical tests, the chance of false positives results or results that are associated when actually the null hypothesis is true (termed type I error) increases. Therefore adjustments for multiple testing needs to be made (272).

The Manhattan plots in Figure 4-3 show that aspirin does not enrich for methylation at a specific genomic locus but appears to be have more widespread effects. To correct for testing of multiple CpG sites on the effect of methylation, we applied a Bonferroni correction and calculated an FDR-adjusted P-value (Chapter 4.1.7.2). After calculating a Bonferroni corrected P-value, none of the CpG sites' P-values was below the threshold of 1.31×10^{-7}. We also applied an FDR adjusted P-value whereby this needs to be below 0.05 for results to be considered significant (343). Whilst neither of these analyses reach this threshold, the results from aspirin treated vs untreated achieved lower p-values than the change in DNA methylation per 2mM aspirin increase. The Manhattan plots in Figure 4-3 show that the lowest P-value identified from our linear regression where aspirin was included as a continuous variable was 2.56×10^{-5} whereas the lowest from our linear regression where aspirin was considered a binary variable was 7.75×10^{-6}. Each dot represents a CpG site therefore the higher the dot, the lower the P–value. Continuous variables tend to be better powered statistically to discern an effect than binary variables. However, aspirin use as a binary variable in the model has produced more statistically significant results, indicative of a threshold effect.

Despite the fact that our EWAS analysis did not identify clear and robust associations between aspirin treatment and changes in DNA methylation at specific CpG sites, we did observe quite widespread differences associated with treatment that fell below the conventional threshold for statistical significance ($P \leq 0.05$). We therefore opted to use the data in an integrated approach by combining it with proteomic and transcriptomic data to establish whether collective evidence from all three 'omic platforms might help to highlight potential pathways of aspirin action.

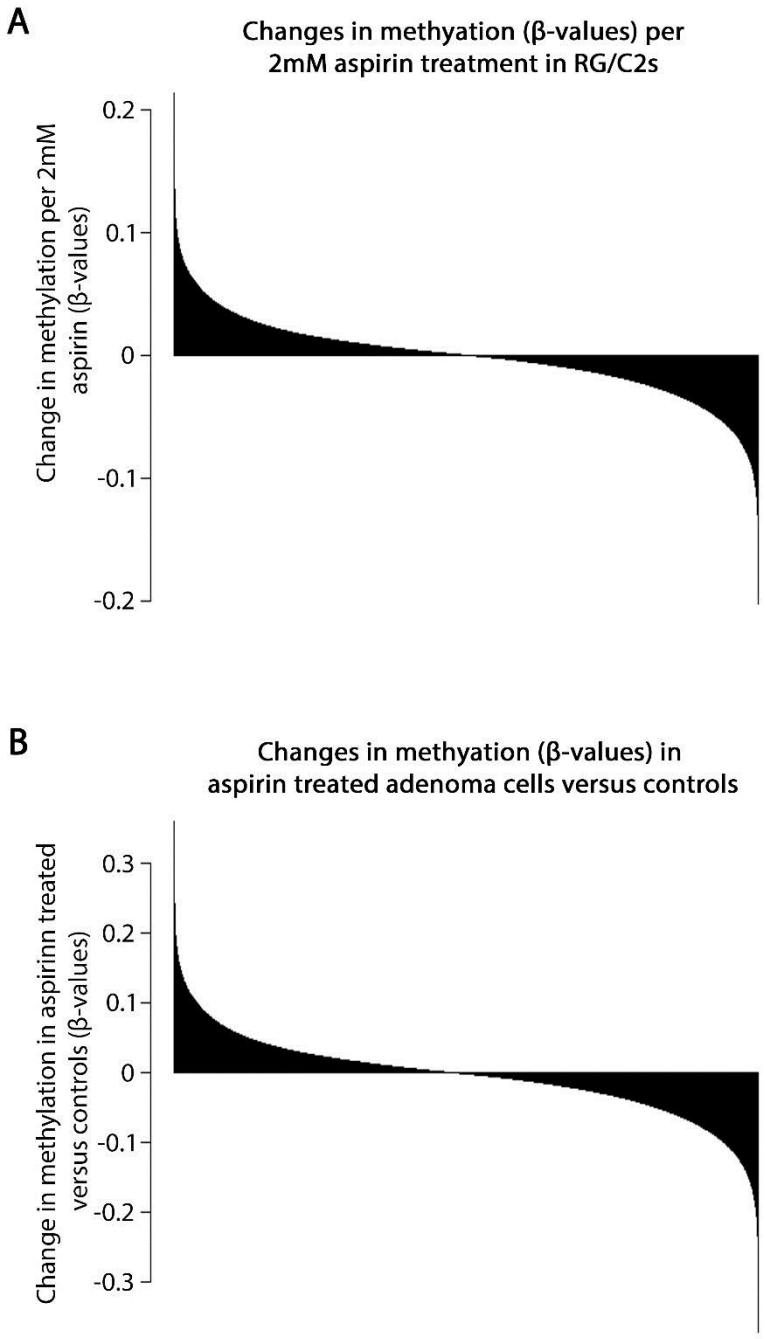


Figure 4-2- Aspirin exposure as a binary variable in the EWAS regression model shows a larger effect on methylation when compared to aspirin as a continuous variable

RG/C2 cells were treated with 0, 2 and 4mM aspirin for 24 hours before DNA was extracted. Samples were then bisulfite converted and run over the Illumina Infinium HumanMethylation 450K BeadChip assay. The data generated from the assay was processed using the Meffil package in R (version 3.3.1) and the methylation at each CpG site was quantified. Methylation levels are expressed as a "Beta" value (β-value) with values ranging from 0 to 1 where 0 indicates no cytosine methylation and 1 indicates complete cytosine methylation. (A) Waterfall plot of the linear regression results of aspirin use as a continuous variable (0, 2 and 4mM) show that per 2mM aspirin treatment, the highest decrease in DNA methylation is -0.20 (20%) and the highest increase is 0.20 (20%). The beta values denote the change in methylation per 2mM aspirin increase in concentration. (B) Waterfall plot of the linear regression results of aspirin use as a binary variable (treated vs untreated) show that aspirin's highest effect on DNA methylation is a

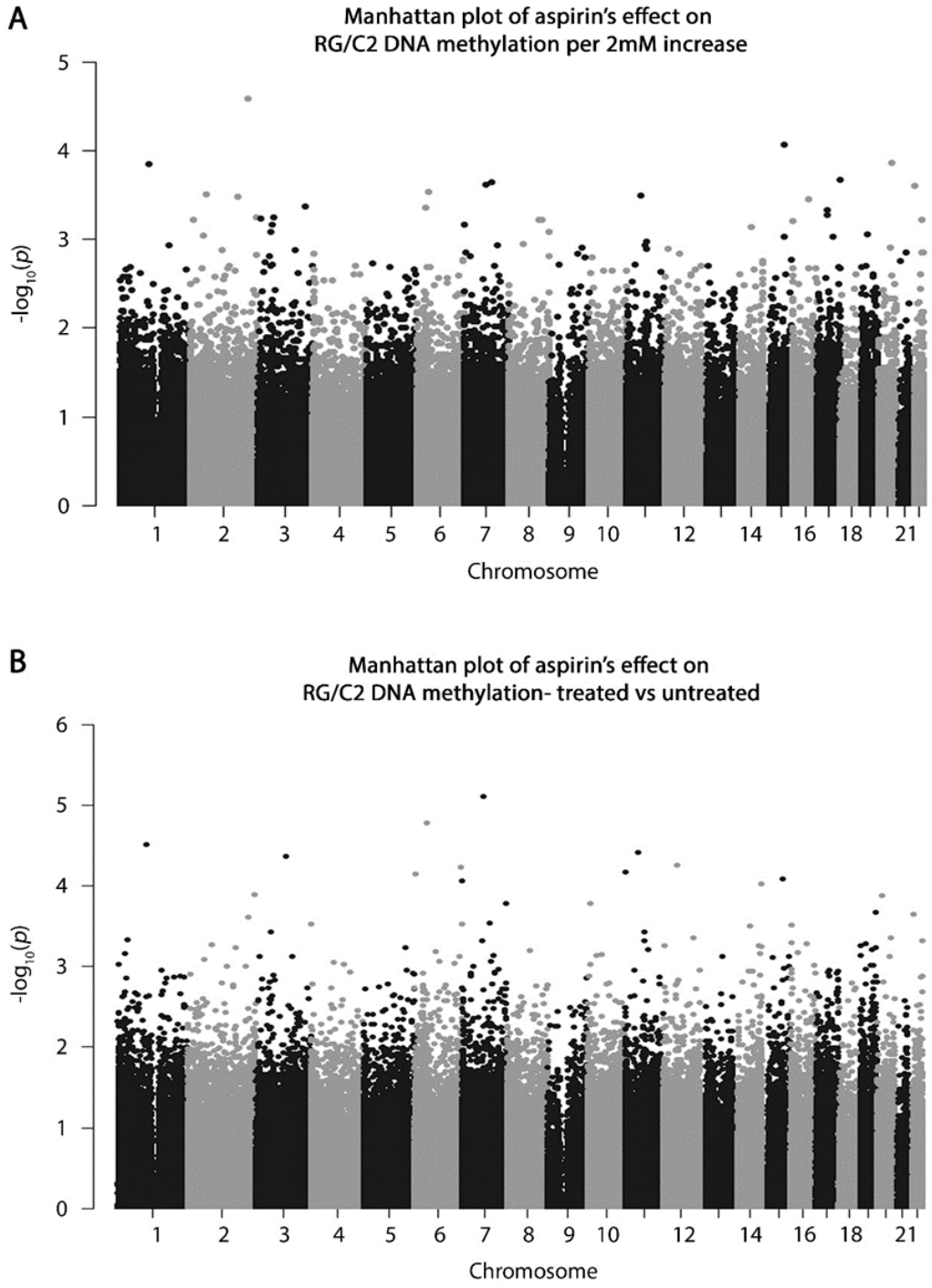


Figure 4-3- Aspirin exposure as a binary variable in the EWAS regression model produces more discernible differences in DNA methylation when compared to aspirin considered as a continuous variable.

RG/C2 cells were treated with 0, 2 and 4mM aspirin for 24 hours before DNA was extracted. Samples were then bisulfite converted and run over the Illumina Infinium HumanMethylation 450K BeadChip assay. The data generated from the assay was processed using the Meffil package in R (version 3.3.1) and the methylation at each CpG site and their P-value for significance was quantified. There is no epigenome-wide significance line on the plots as this would equate to 1.31×10^{-7} which is not present on the y-axis of either graph. (A) Manhattan plot showing the location of each CpG site (x-axis) against the P-value of the change in DNA methylation (y-axis) when aspirin use is a continuous variable (0, 2 and 4mM). Each dot represents a P-value for a particular CpG site. The smallest P-value achieved per 2mM aspirin treatment is 2.65×10^{-5}. (B) Manhattan plot showing the location of each CpG site (x-axis) against the P-value of the change in DNA methylation (y-axis) between treated and untreated cells. Each dot represents a P-value for a particular CpG site. The smallest P-value achieved between aspirin treated and untreated is 7.75×10^{-6}.

4.3.2 4mM aspirin treatment causes a decrease in protein expression

To identify novel targets of aspirin, we investigated the effect of aspirin on multiple 'omics, one of which is the proteome. The intention was to combine results from the methylomic data with the proteomic data to find new targets of aspirin that may be influenced by epigenetic regulation.

RG/C2 cells were cultured for 24 hours with 0, 2 and 4mM aspirin using a SILAC approach. Control cells were cultured with a light amino acid, 2mM were cultured with a medium amino acid and 4mM were cultured with a heavy amino acid. When cells divide in culture, eventually the normal amino acids are replaced by the isotope labelled ones. These produce different molecular weights for proteins which are then detected using mass spectrometry (354). This data was previously available from our laboratory and results were analysed.

The SILAC approach detected 5886 proteins in total. To identify changes in protein expression, we used a 1.4-fold change as a threshold between 4mM/control and 2mM/control as previously used by another group (355). We also limited our results to protein fold changes that had <100% variability (to remove results with extreme values of variability) and a count of at least 2 or more. By applying this threshold, we identified that aspirin affects the expression of 131 proteins (Figure 4-4 A), further details on these proteins can be found in the Appendix 1. Of these, 129 proteins were found to decrease in expression and only 2 proteins increased in expression indicating that 4mM aspirin treatment decreased expression of 98.5% of the proteins measured (Figure 4-4 B).

As short-term aspirin appears to mainly decrease protein expression, we decided to focus on these proteins. The next step was to combine these protein expression changes with the relevant loci with DNA methylation data so a protein that decreases expression would be expected to have an increase in its' CpG site methylation. This would provide more evidence to the gene expression being a target of aspirin.

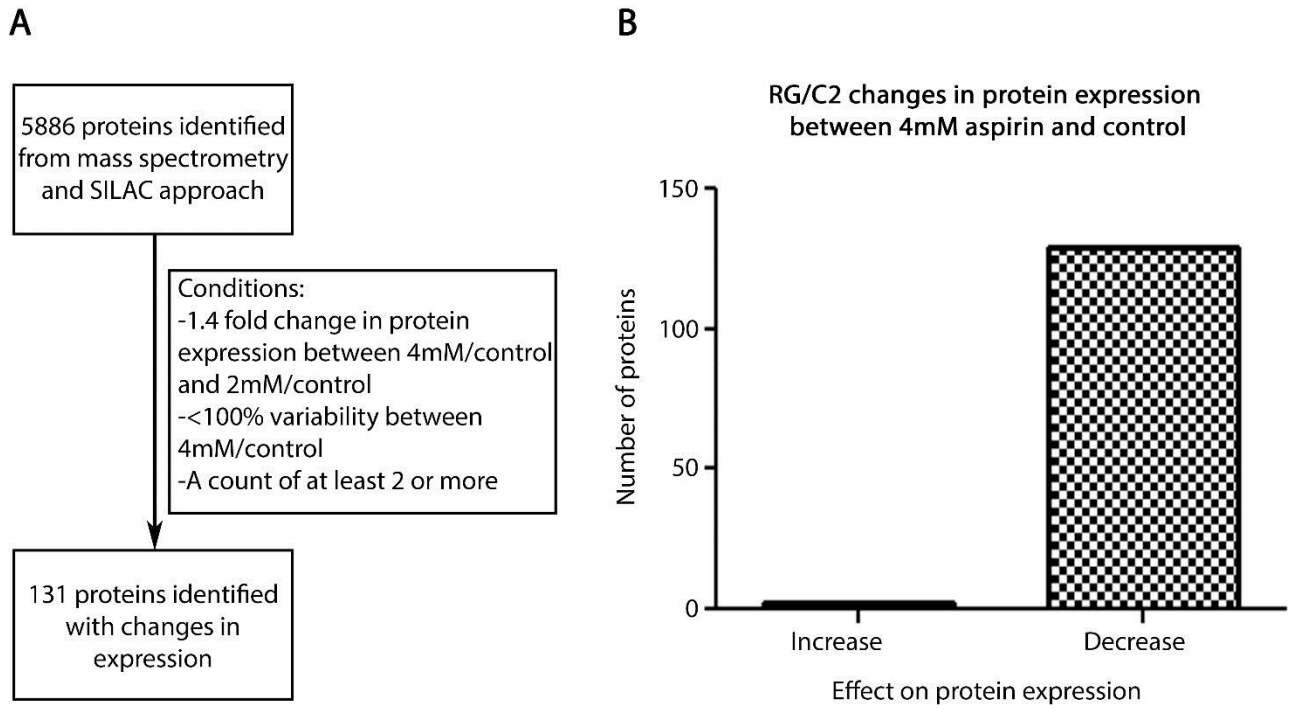


Figure 4-4- Aspirin treatment alters the expression of 131 proteins.

RG/C2s were cultured with 0, 2 and 4mM aspirin for 24 hours using a SILAC approach. Control cells were cultured with a light amino acid, 2mM were cultured with a medium amino acid and 4mM were cultured with a heavy amino acid. The proteins were then detected using mass spectrometry and fold changes between the different treatments were calculated. (A) Data was filtered by setting a threshold of 1.4-fold change in protein expression between 4mM/control and 2mM/control, <100% variability and a count of at least 2 or more. This showed that aspirin affects the expression of 131 proteins. (B) Results were stratified between proteins that increase or decrease expression with 4mM aspirin treatment. From the 131 proteins identified, 129 proteins experienced a decrease in expression.

4.3.3 Combining proteomic, methylomic and transcriptomic data identifies 7 genes targeted by aspirin

Since aspirin decreased expression of 98.5% of the proteins assayed using the SILAC approach, we then identified genes within this group that also showed changes in methylation. To do this, we set a threshold of ≥0.1 change in β as this denotes increased methylation which has previously been associated with inhibiting access by transcription factors causing decreased protein expression (170).

The 129 proteins that showed decreased protein expression were matched to the methylomic data. Proteins were matched to the methylome by searching for the lowest P-value cis CpG site (defined as 1Mb either side of the gene start site (Figure 4-5). This yielded a match of 125 sites (Figure 4-7 A).

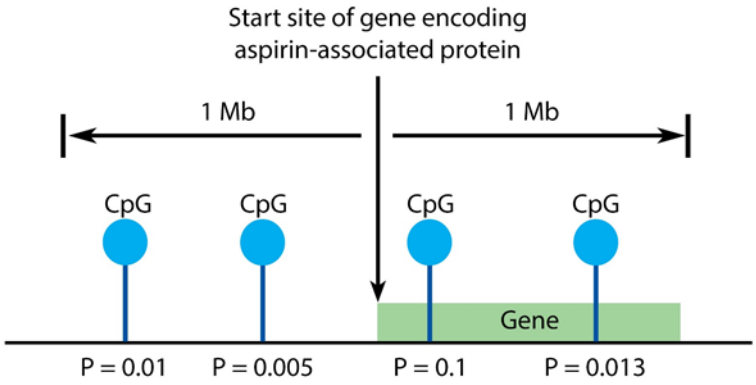


Figure 4-5- Schematic of choosing a cis CpG site that is ≤ 1 megabase (Mb) from the gene start site.

To try and identify new targets of aspirin, we used a published transcriptomic dataset generated by Sabates-Bellver et. al (2007) that compared mRNA levels of normal and colorectal adenoma samples from 32 patients (347). We sought to establish whether the proteins that showed a decrease in expression with aspirin were involved along the pathway from normal to adenoma progression (Figure 4-6). Therefore, the transcriptomic data was matched to the 129 proteins identified via the SILAC approach, and gene expression data from normal to adenoma progression was available for 114 of these genes (Figure 4-7 A).

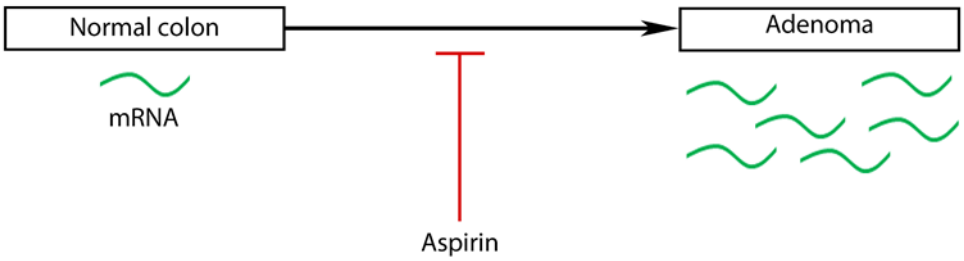
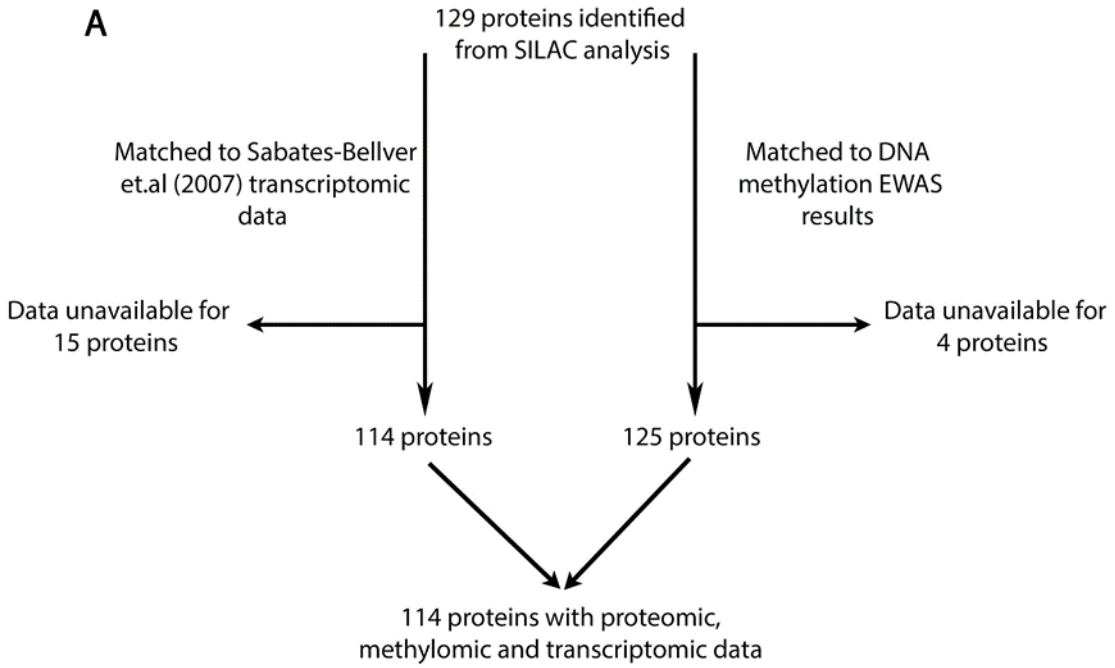


Figure 4-6- Schematic of the increase in mRNA from normal to adenoma progression.



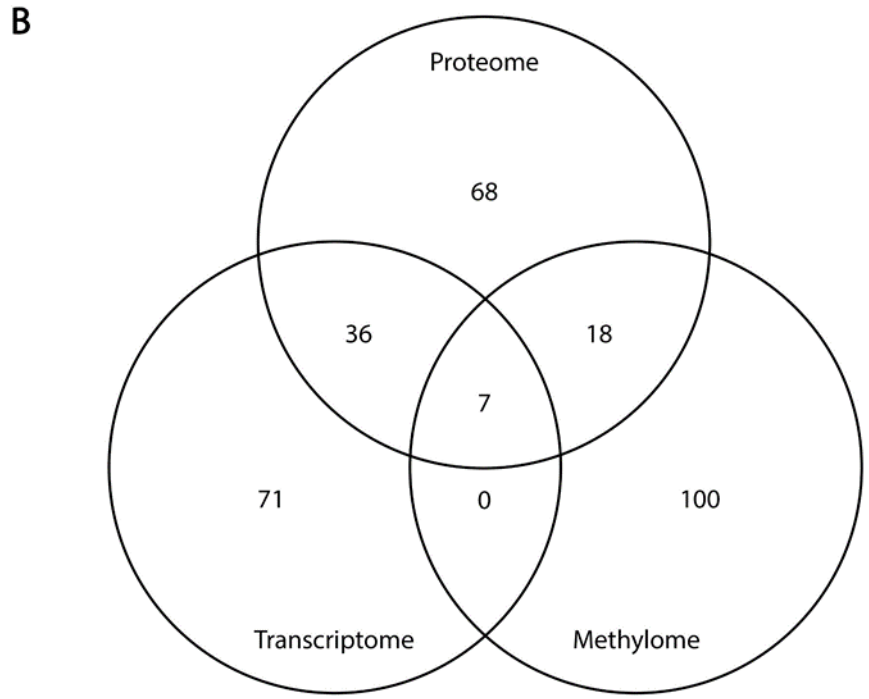


Figure 4-7- Combining 'omics data identifies aspirin as potentially targeting the expression of 7 new proteins

Proteomic data using the SILAC approach was combined with changes in methylation between aspirin treatment (2mM and 4mM) and control as well as a published transcriptomic dataset by Sabates-Bellver et. al (2007) of changes in gene expression between normal colorectal and adenoma samples. (A) Methylomic and transcriptomic data were matched to the proteomic data. Overall, information from all 3 'omics was available for 114 proteins. (B) After setting a threshold of a 1.4 fold decrease in protein expression, a 10% increase in DNA methylation and a 1.4 fold decrease in gene expression, 7 proteins are potentially targeted for decreased expression by aspirin treatment.

Table 4-3- The thresholds used to identify proteins that decrease in expression with aspirin treatment

Overall change in gene and protein expression	Fold change in Protein expression (4mM asp treated vs untreated)	Change in Methylation (β) (asp treated vs untreated)	Fold change in Gene expression (adenoma vs normal)
Decreased	≤0.714	≥0.1	≥1.4

Once data for the 114 genes was collated, a threshold was set for each dataset. The thresholds for identifying genes that have decreased protein and gene expressions are summarised in Table 4-3. The first threshold set was applied to the proteomic data. Boulon *et. al* showed that a >1.4 fold change in protein expression would eliminate most of the contaminants, low affinity and low abundance interacting partners (356). Therefore, our inclusion criteria of a >1.4 fold decrease in protein expression was applied to the SILAC proteomic data, as well as a count of at least 2 or more and <100% variability to remove results with extreme values of variability.

A threshold of ≥0.1 change in methylation β values was set for the 125 CpG sites matched to the SILAC data. This is guided by a review by Michels et. al who suggested that methylation changes of <5% should be interpreted with caution but changes of >10% provide increased confidence in the biological significance of the result (342). As for the fold changes in gene expression, 1.4 was used as a threshold due to results by Morey *et. al* who showed that these results were shown to be highly correlated with qPCR results (357).

Overall, we set a 1.4 fold decrease (≤0.714 fold change in protein expression in 4mM treated compared to untreated) in protein expression, an increase in methylation of at least 10% and a 1.4 fold increase in gene expression as our threshold for identifying new proteins targeted by aspirin (Table 4-3). After the thresholds were defined and set, we identified seven proteins that were potentially targeted by aspirin. The combination of the proteomic, methylomic and transcriptomic results are shown as a Venn diagram in Figure 4-7 B. The diagram shows that 7 genes survive the thresholds set for all 3 datasets. These are: minichromosome maintenance complex component 2 (*MCM2*), minichromosome maintenance complex component 5 (*MCM5*), Non-structural maintenance of chromosomes element 1 (*NSMCE1*), DNA polymerase delta subunit 2 (*POLD2*), stathmin 1 (*STMN1*), dimethyladenosine transferase 2 (*TFB2M*) and thymidine kinase 1 (*TK1*). The fold change in protein and gene expression as well as the change in methylation are summarised in Table 4-4.

Table 4-4- A summary of the changes in expression/methylation of the 7 identified targets

Gene Name	GenBank Accession number	CpG site			Change in gene expression (fold change) of normal to adenoma samples	Change in protein expression (fold change) of 0mM aspirin treated adenoma cells vs 4mM	Change in methylation (β) of aspirin treated vs untreated adenoma cells
		Illumina Probe ID	Chromosome	Genome location			
MCM2	NM_004526	cg03255953	3	127056972	2.69	0.61	0.34
MCM5	NM_006739	cg27116232	22	35775959	1.84	0.62	0.20
MCM5	NA	cg27116232	22	35775959	2.01	0.62	0.20
NSMCE1	AF161451	cg03631519	16	28081266	1.43	0.70	0.25
POLD2	NM_006230	cg26390609	7	44579467	1.96	0.35	0.16
STMN1	NM_005563	cg02490185	1	26136899	1.84	0.61	0.26
TFB2M	NM_022366	cg03896752	1	245727658	1.50	0.56	0.18
TK1	BC007986	cg22386583	17	78753756	2.09	0.46	0.16
TK1	NM_003258	cg22386583	17	78753756	2.42	0.46	0.16

GenBank Accession number refers to the transcript variant of the mRNA. Abbreviations: MCM2, minichromosome maintenance complex component 2; MCM5, minichromosome maintenance complex component 5; NSMCE1, non-structural maintenance of chromosomes element 1; POLD2, DNA polymerase delta subunit 2 ; STMN1, stathmin 1; TFB2M, dimethyladenosine transferase 2;, TK1, thymidine kinase 1; NA= missing values, not found from Sabates-Bellver et al's raw data.

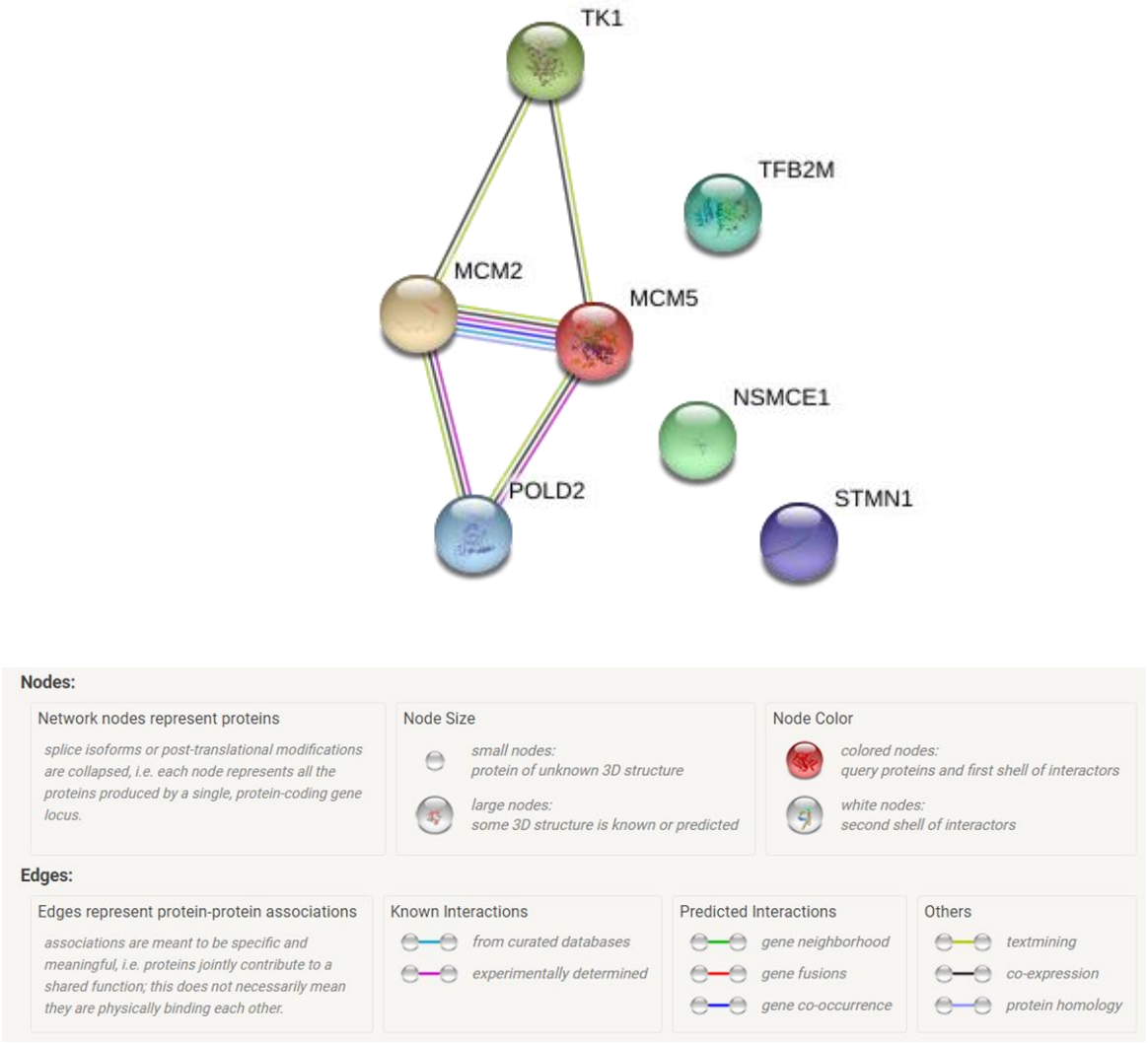


Figure 4-8- STRING diagram showing the interactions between the 7 proteins identified from combining the 3 datasets.

Using the STRING resource, an experimental relationship of protein-protein association between MCM2, MCM5 and POLD2 was found, which have also been found to be co-expressed. Through text mining approaches, a protein-protein association was found between TK1, MCM2, MCM5 and POLD2. The proteins TFB2M, NSMCE1 and STMN1 are independent with no evidence of interaction with the other proteins.

Using the STRING resource, protein interactions were investigated to identify whether there were common pathways that were regulated by aspirin treatment (358).The diagram shows that 4 of the 7 proteins interact and are co-expressed or identified as interacting through text mining. The proteins MCM2, MCM5 and POLD2 have further evidence of protein interaction via published experimental data which also shows that they are co-expressed. Text mining approaches have also shown a protein-protein interaction between these three proteins along with TK1(Figure 4-8). The proteins TFB2M, NSMCE1 and STMN1 are independent with no evidence of interaction with the other proteins.

4.3.4 Short term aspirin decreases *STMN1* mRNA in RG/C2s and LS174Ts but not SW620s

By combining the multiple 'omics, we have identified 7 possible targets of aspirin of relevance to colorectal cancer. Since there were no publicly available transcriptomic datasets for aspirin's effect on colorectal adenomas, we opted to use one that only compared normal and adenoma samples to see if our targets experience an increase in gene expression along the tumour progression pathway. This meant that we needed to carry out quantitative polymerase chain reaction (qPCR) for our identified targets to verify whether they were indeed targets of aspirin.

RG/C2 cells were cultured for 24 hours with 0, 2 and 4mM aspirin. Cells were then treated with TRIZOL, RNA was extracted and samples were then subject to qPCR reaction. This was carried out for the following genes: *MCM2*, *MCM5*, *POLD2*, *TK1* and *STMN1*. These genes were chosen as *MCM2*, *MCM5*, *POLD2*, *TK1* have been shown to be functionally related and the literature has shown protein-protein interaction as well as their co-expression. We also chose *STMN1* as this protein has been shown to be associated with metastasis (359). Results are mean values from 3 repeat experiments with standard deviations and a one-way ANOVA and Dunnett's multiple comparison test to calculate significance.

The results show that relative to control, aspirin has no significant effect on the expressions of *MCM2, POLD2* and *TK1* at either 2 or 4mM (Figure 4-9 A, C and E). Aspirin does significantly affect the gene expression of *MCM5* at 4mM ($p \leq 0.05$) but not at 2mM. Aspirin also significantly affects the gene expression of *STMN1* at both 2mM ($p \leq 0.05$) and 4mM ($p \leq 0.01$).

The results show that aspirin significantly decreases the mRNA expression of *MCM5* at 4mM ($p \leq 0.01$) aspirin and *STMN1* both at 2mM ($p \leq 0.05$) and 4mM ($p \leq 0.01$) (Figure 4-9). However, aspirin does not significantly affect the mRNA expression of *MCM2, POLD2,* and *TK1*. Our proteomic results show that aspirin decreases the protein expression of MCM5 and STMN1 at 4mM by 0.62 and 0.61, respectively. Our qPCR results also show that aspirin decreases their gene expression to 0.54 (SD 0.17) and 0.45 (SD 0.23), respectively. Therefore, with regards to these 2 genes, aspirin is consistently decreasing their gene expression.

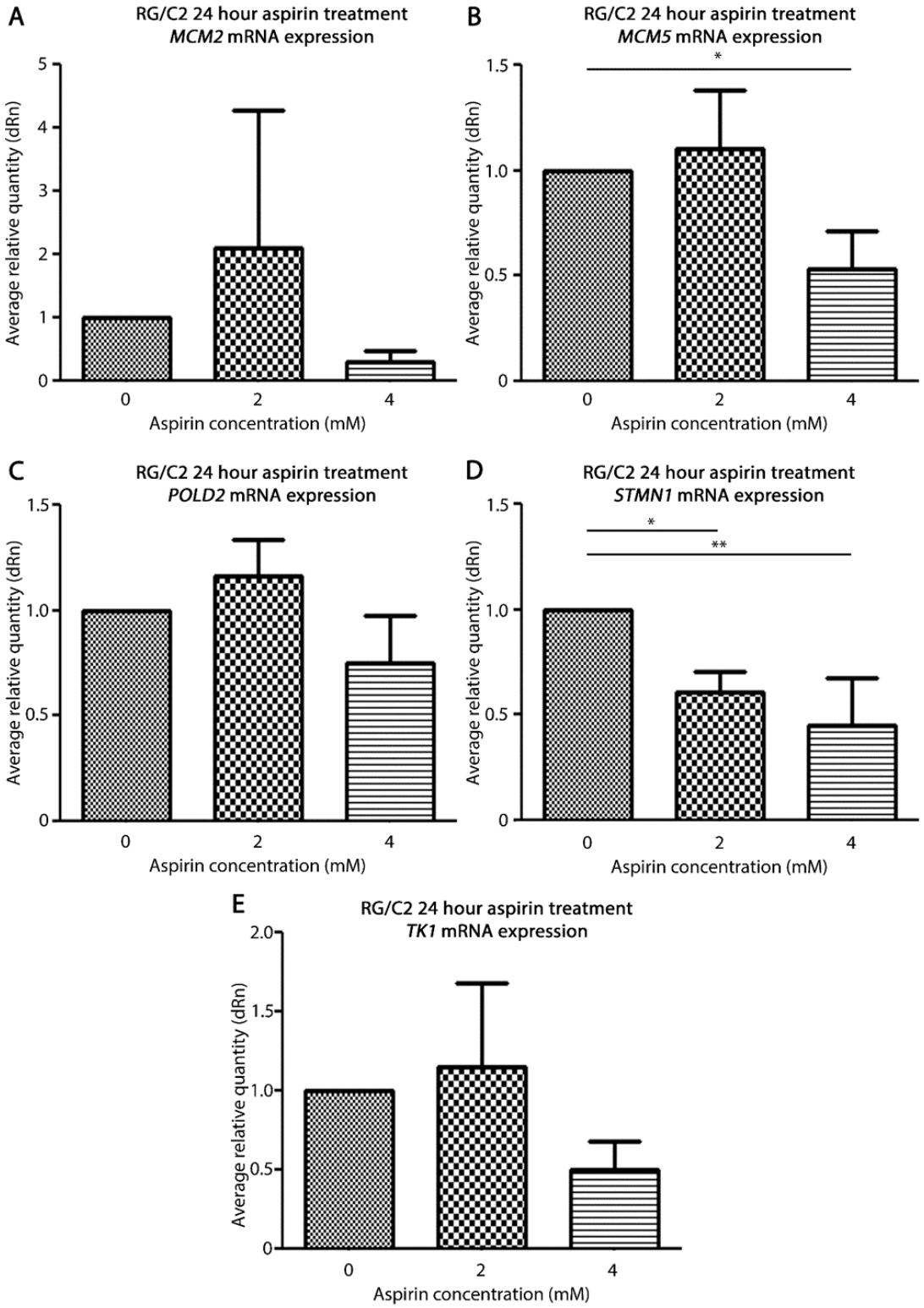


Figure 4-9- Graphs of RG/C2 qPCR results showing the average relative quantity of the mRNA expression for 5 of the 7 targets: *MCM2*, *MCM5*, *POLD2*, *STMN1* and *TK1*.

RG/C2 cells were cultured for 24 hours with 0, 2 and 4mM aspirin. Cells were then treated with TRIZOL and RNA was extracted. Samples were then subject to qPCR reaction. Results show the average relative quantity to control. Results are mean values from 3 repeat experiments with standard deviations and a one-way ANOVA and Dunnett's multiple comparison test to calculate significance (*** $p \leq 0.001$, **$p \leq 0.01$ and * $p \leq 0.05$). (A) Aspirin treatment had no significant effect on *MCM2* gene expression at 2 and 4mM. (B) Aspirin treatment significantly decreased *MCM5* gene expression at 4mM ($p \leq 0.05$) but not 2mM. (C) Aspirin treatment had no significant effect on *POLD2* gene expression at 2 and 4mM. (D) Aspirin significantly decreased *STMN1* gene expression at both 2mM ($p \leq 0.05$) and 4mM ($p \leq 0.01$). (E) Aspirin treatment has no significant effect on *TK1* gene expression at 2mM and 4mM.

Since *STMN1*'s gene expression was regulated at both 2 and 4mM aspirin, we investigated whether it was also regulated in cancer derived cells LS174T and SW620 cells (derived from a lymph node metastatic patient). This is because the protein expression of STMN1 has been significantly associated with TNM staging and lymph-node metastasis (359). Therefore we wanted to test if aspirin can also decrease its mRNA expression at other stages in the tumour progression pathway and whether this may partly explain how aspirin reduces the risk of cancer metastasis (114). We cultured LS174T and SW620 cells for 24 hours with 0, 2 and 4mM aspirin. Cells were treated with TRIZOL and RNA was extracted. Samples were then subject to a qPCR reaction. Results are mean values from 3 repeat experiments with standard deviations and a one-way ANOVA and Dunnett's multiple comparison test to calculate significance.

Our results show that aspirin significantly decreases the mRNA expression of *STMN1* at 4mM to 0.64 average relative quantity (SD 0.09) ($p \leq 0.01$) but not 2mM aspirin in LS174T cells (Figure 4-10 A). Aspirin does not significantly affect *STMN1* gene expression in SW620 cells at 2mM or 4mM (Figure 4-10 B). These results suggest that stathmin-1 is less regulated in carcinoma cells compared to adenoma derived cells.

Overall, our results show that 4mM aspirin decreases the gene expression of *STMN1* at both the adenoma stage and the carcinoma stage but has no significant effect on *STMN1* expression in metastatic cells indicating that aspirin is most effective on pre-metastatic tumours.

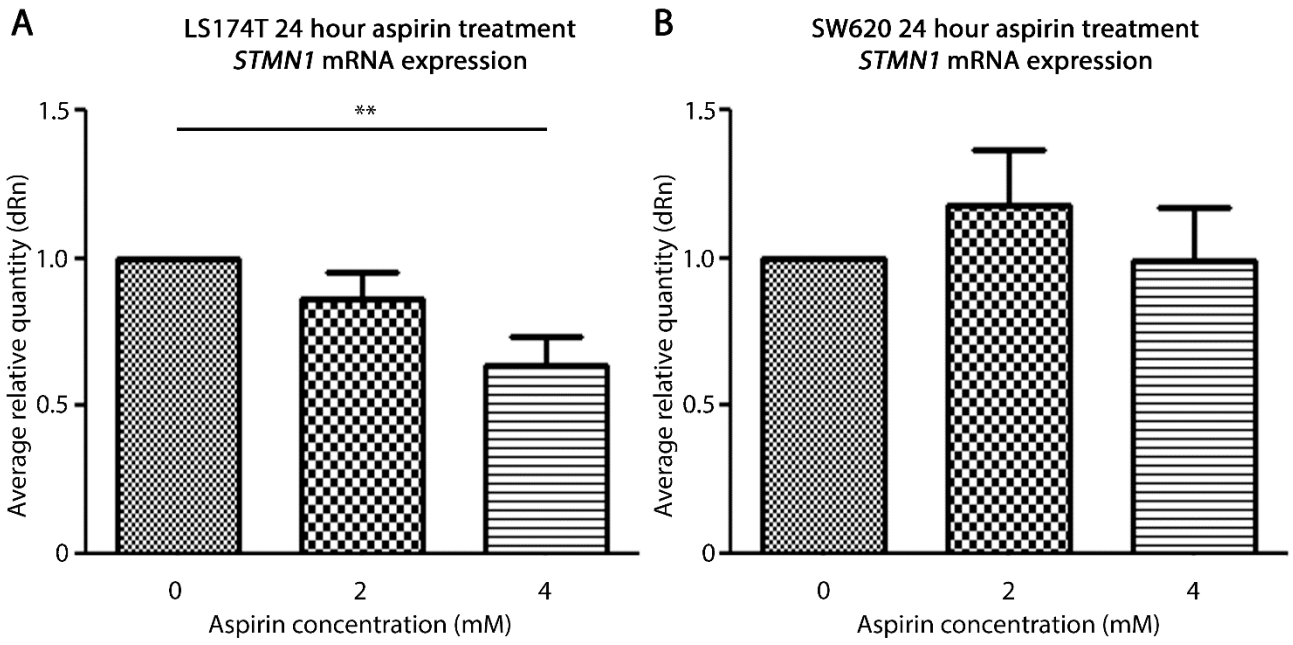


Figure 4-10- Graphs of LS174T and SW620 qPCR results showing the average relative quantity of the mRNA expression of *STMN1*.

LS174Ts and SW620s were cultured for 24 hours with 0, 2 and 4mM aspirin. Cells were then treated with TRIZOL and RNA was extracted. Samples were then subject to qPCR reaction. Results show the average relative quantity to control. Results are mean values from 3 repeat experiments with standard deviations and a one-way ANOVA and Dunnett's multiple comparison test to calculate significance (*** $p \leq 0.001$, **$p \leq 0.01$ and * $p \leq 0.05$). (A) Aspirin treatment significantly decreased the gene expression of *STMN1* at 4mM ($p \leq 0.01$) but not 2mM aspirin. (B) Aspirin treatment had no significant effect on the gene expression of *STMN1 at 2mM or 4mM.*

4.3.5 Aspirin decreases the protein expression of Stathmin-1 in RG/C2s and LS174Ts

In order to confirm that aspirin decreased STMN1 protein expression, we carried out western blotting experiments. RG/C2 and LS174T cells were cultured in 2D with 0, 2 and 4mM aspirin for 24, 48, 72 and 96 hours. Cell lysates were obtained, and samples were used in a western blotting experiment.

Firstly, we showed striking changes in stathmin-1 expression over time, noticeably it increases with increasing confluence in RG/C2 cells. Our results show that aspirin is indeed reducing the protein expression of STMN1 at 48, 72 and 96 hours with 2 and 4mM aspirin in RG/C2s although it has no effect on protein expression at 24 hours (Figure 4-11 A). Our results also show that aspirin decreases STMN1 expression in LS174Ts at 24 and 48 hours with 4mM and at 72 and 96 hours with both 2 and 4mM (Figure 4-11 B).

Therefore, we confirmed that aspirin reduces both the gene expression and protein expression of STMN1 in both adenoma and carcinoma cells. This result led us to investigate the functional importance of this protein and the consequence of reducing its expression and how it affects tumour progression.

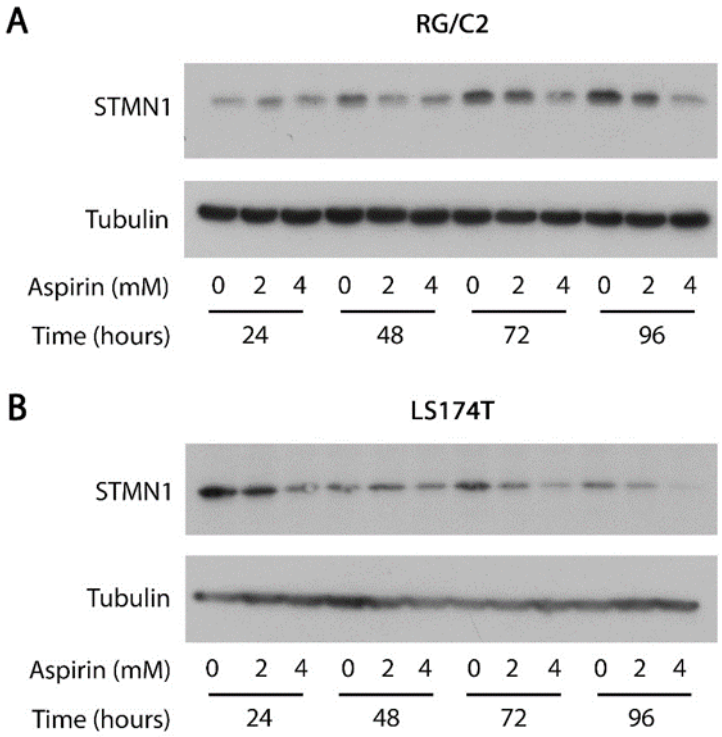


Figure 4-11- Aspirin decreased the protein expression of STMN1 in RG/C2s and LS174Ts

RG/C2 and LS174T cells were cultured with 0, 2 and 4mM aspirin for 24, 48, 72 and 96 hours. Cell lysates were extracted, and western blotting was carried out. Tubulin was used as a loading control. Results are representative of data from 3 separate experiments. (A) Aspirin decreased STMN1 protein expression from 48 hours onwards at both 2 and 4mM in RG/C2 cells. (B) Aspirin decreased STMN1 protein expression at 48 hours with 4mM and at 72 hours onwards with 2 and 4mM in LS174Ts.

4.3.6 STMN1 knockdown has no effect on the number of attached and floating cells

In order to determine the functional effect of STMN1 expression, we investigated the effect of knocking down STMN1 expression on the number of attached and floating cells, in order to see if the effect of aspirin treatment could be recapitulated (Chapter 3.3.1.1).

To determine which time points and concentration of *STMN1* siRNA we would use, we carried out a dose response experiment. RG/C2 cells were cultured for 24, 48, 72 and 96 hours with 25nM and 50nM *STMN1* siRNA. Cell lysates were extracted and western blotting was carried out.

The results show that *STMN1* siRNA knocked down STMN1 expression from 48 hours onwards with both 25 and 50nM concentrations (Figure 4-12 A). Therefore, we used a concentration of 25nM and a starting time point for measurements at 48 hours (for suppression of stathmin-1 protein expression) to set up an experiment investigating the effect of STMN1 knockdown on the number of attached and floating cells.

RG/C2s were cultured with 25nM of *STMN1* siRNA for 48, 72 and 96 hours. Experiments were also set up with control siRNA to compare results of STMN1 knockdown vs control. The number of attached and floating cells were counted, and percentage floaters was calculated. Cell count results are mean values from 3 repeat experiments with standard deviations and a one-way ANOVA and Dunnett's multiple comparison test. Western samples were set up in parallel with the experiment to confirm *STMN1* knockdown and tubulin was used as a loading control.

Our results show that *STMN1* knockdown has no significant effect on the number of attached cells in all the time points (Figure 4-12 B) and it also has no significant effect on the percentage floating cells (Figure 4-12 C). We confirmed via western blotting that STMN1 was indeed knocked down at 48, 72 and 96 hours (Figure 4-12 D).

In summary, *STMN1* expression has no significant effect on the number of attached and floating cells. This indicates that aspirin's effect on reducing the number of attached cells is independent to its effect on *STMN1* expression. This suggests suppression of stathmin-1 does not mediate the effect of aspirin on cell survival.

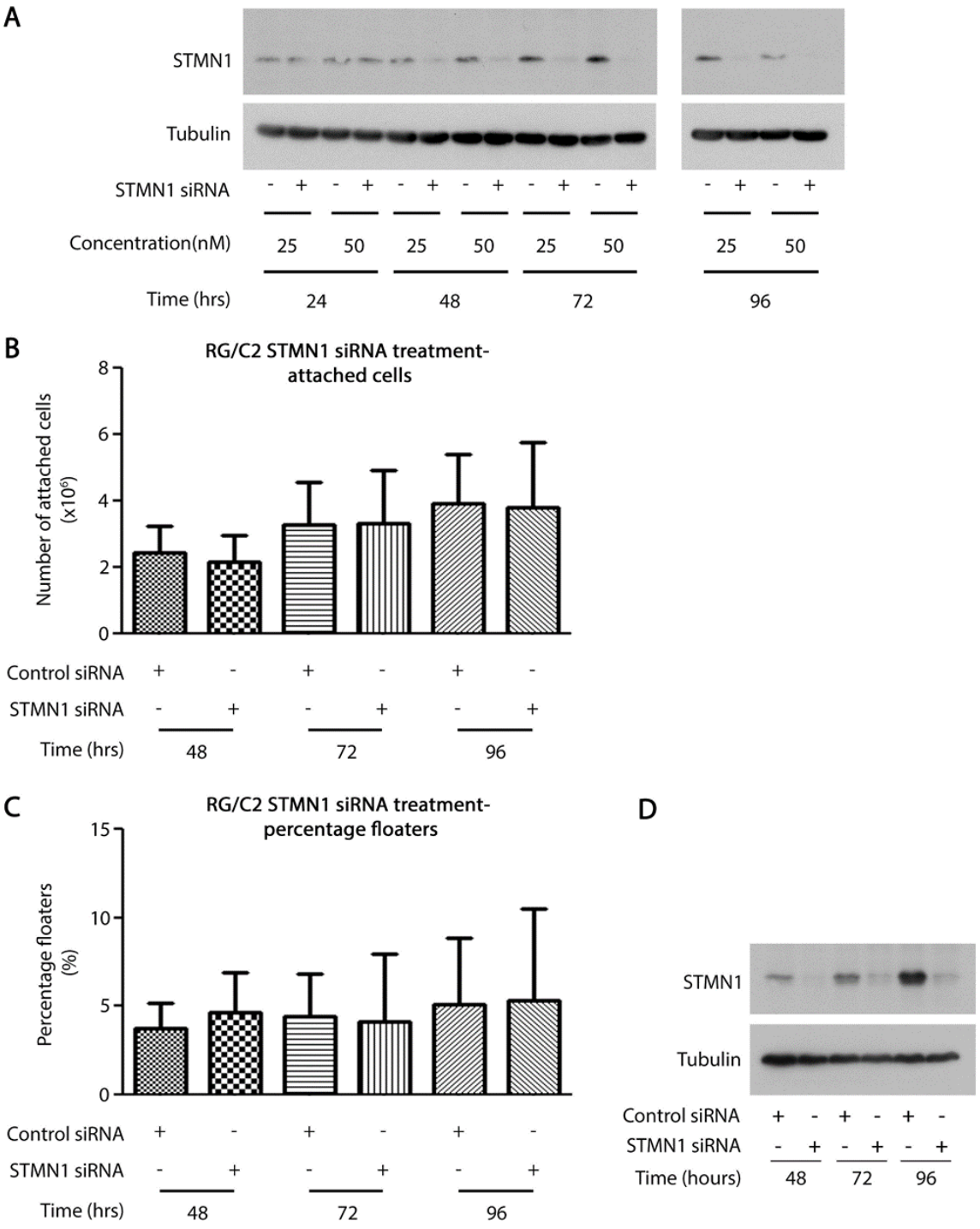


Figure 4-12- STMN1 knockdown has no significant effect on cell counts in RG/C2 cells

RG/C2 cells were cultured with 25nM of *STMN1* siRNA for 48, 72 and 96 hours. The number of attached and floating cells were counted, and percentage floaters was calculated. Cell lysates were extracted to confirm STMN1 knockdown. A – indicates the absence of the control or *STMN1* siRNA and a + indicates the presence of the siRNA. Cell count results are mean values from 3 repeat experiments with standard deviations and a one-way ANOVA and Dunnett's multiple comparison test (*** $p \leq 0.001$, **$p \leq 0.01$ and * $p \leq 0.05$). Western blotting results are representative of data from 3 separate experiments. (A) *STMN1* siRNA was added to RG/C2 cells for 24, 48, 72 and 96 hours at 25nM and 50nM to find out which time point and concentration of siRNA was best to use for STMN1 knockdown experiments. *STMN1* siRNA at 25nM successfully knocked down STMN1 protein expression at 48 hours and onwards. (B) STMN1 knockdown had no significant effect on the number of attached cells at 48, 72 and 96 hours. (C) STMN1 knockdown had no significant effect on the percentage floaters at 48, 72 and 96 hours. (D) Western samples set up in parallel with the experiment were used to confirm STMN1 knockdown. Tubulin was used as a loading control. The results show that STMN1 was successfully knocked down from 48 hours onwards.

4.3.7 STMN1 knockdown increases the percentage of cells in G0/G1 phase and decreases the percentage of cells in G2 phase in RG/C2 cells

Since the knockdown of *STMN1* has no effect on cell viability, we investigated its effect on cell cycle progression.

RG/C2s were cultured with 25nM of *STMN1* siRNA or negative control siRNA for 48 hours. Cells were then collected, fixed and stained with DRAQ5 before being analysed by flow cytometry (Figure 4-13). Western samples were set up alongside the experiment to confirm STMN1 knockdown and tubulin was used as a loading control (Figure 4-13 C).

The results show that STMN1 knockdown causes a small but significant increase in the percentage of cells at G_0/G_1 phase of the cell cycle from 53.43% (SD 0.71) to 56.43% (SD1.93) ($p \leq 0.05$) (Figure 4-13 A, Table 4-5). STMN1 knockdown also significantly decreases the percentage of cells at G_2 from 23.16% (SD 2.16) to 17.56% (SD 2.11) (Figure 4-13 B, Table 4-5).

Although these results show that STMN1 knockdown has a significant effect on cell cycle progression, it only increases the percentage of cells in G_0/G_1 phase by 4% and decreases the percentage of cells in G_2 phase by 5.6%. We previously showed that 24-hour treatment of 4mM aspirin increases the percentage of cells in G_0/G_1 phase by 7.65% (Chapter 3.3.1.3). Whilst the decreased expression of *STMN1* may partly explain this effect, these results suggest that a decrease in STMN1 expression is not causing the arrest observed in aspirin treated cells.

Table 4-5- The percentage of cells at each stage in the cell cycle.

Aspirin concentration (mM)	Experiment	Cell cycle					
		G_0/G_1		S		G_2/M	
		Percentage	SD	Percentage	SD	Percentage	SD
Neg siRNA	1	51.95	0.82	24.2	0.79	24.69	1.72
	2	53.25	9.48	26.01	2.77	20.69	0.98
	3	52.10	1.21	21.71	0.33	24.10	1.27
	average	52.43	0.71	23.97	2.16	23.16	2.16
STMN1 siRNA	1	55.71	0.76	21.67	0.50	19.87	1.52
	2	54.97	1.52	23.35	1.92	15.71	0.62
	3	58.61	2.28	20.51	1.79	17.10	1.88
	average	56.43	1.93	21.84	1.43	17.56	2.11

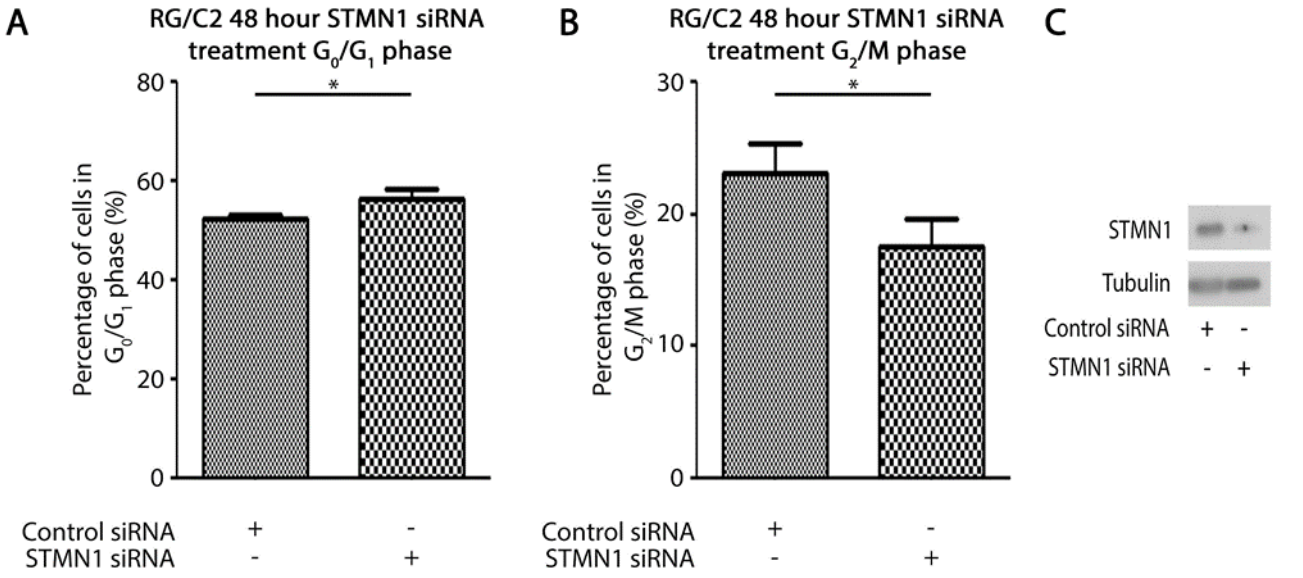


Figure 4-13- STMN1 knockdown significantly increases the percentage of cells in G_0/G_1 and decreases the percentage of cells in G2.

RG/C2 cells were cultured with 25nM of *STMN1* siRNA for 48 hours. Cells were collected, fixed and stained with DRAQ5 and then analysed using flow cytometry. Cell lysates were extracted to confirm STMN1 knockdown. A – indicates the absence of the control or *STMN1* siRNA and a + indicates the presence of the siRNA. Results are mean percentages from 3 repeat experiments with standard deviations and Student's t-test. (*** $p \leq 0.001$, **$p \leq 0.01$ and * $p \leq 0.05$). Western blotting results are representative of data from 3 separate experiments. (A) STMN1 knockdown increased the percentage of cells in G0/G1 phase ($p \leq 0.05$). (B) STMN1 knockdown significantly decreased the percentage of cells in G2 phase ($p \leq 0.05$).(C) Western samples set up in parallel with the experiment were used to confirm STMN1 knockdown. Tubulin was used as a loading control. The results show that STMN1 was knocked down at 48 hours.

4.3.8 Aspirin decreases LS174T cell migration

As the literature states that STMN1 is involved in CRC metastasis (360), we carried out transwell migration assays to investigate the effect of both aspirin and STMN1 knockdown on the ability of cells to migrate.

RG/C2 cells are adenoma cells therefore are unable to migrate through transwells. Since we previously showed that aspirin decreases the expression of STMN1 in LS174T cells (Chapter 4.3.4), we decided to use this carcinoma cell line for the migration assays.

LS174T cells were treated with 4mM aspirin in 2D cell culture for 48 hours. Cells were then extracted, counted and seeded into transwells in FCS-free medium. To create a chemotactic gradient, wells were filled with FCS supplemented media. After 24 hours, cells on the bottom of the transwell were fixed, stained with crystal violet and counted. Western samples were set up alongside the experiment to confirm STMN1 was regulated by aspirin and tubulin was used as a loading control.

The results show that 48 hours of aspirin treatment reduces the amount of crystal violet stained cells (purple) that migrate through the transwell pores (Figure 4-14 A). Relative to untreated, aspirin significantly reduces the number of cells to 0.7 (SD 0.14) ($p \leq 0.01$) (Figure 4-14 B). The western blots also confirm a decrease in STMN1 expression when treated with siRNA (Figure 4-14 C).

These results show that aspirin significantly decreases the number of cells that migrate through the transwell pores. Although we have confirmed through western samples that there is a reduced protein level of STMN1, it is unclear whether the reduction in cell migration is through this mechanism.

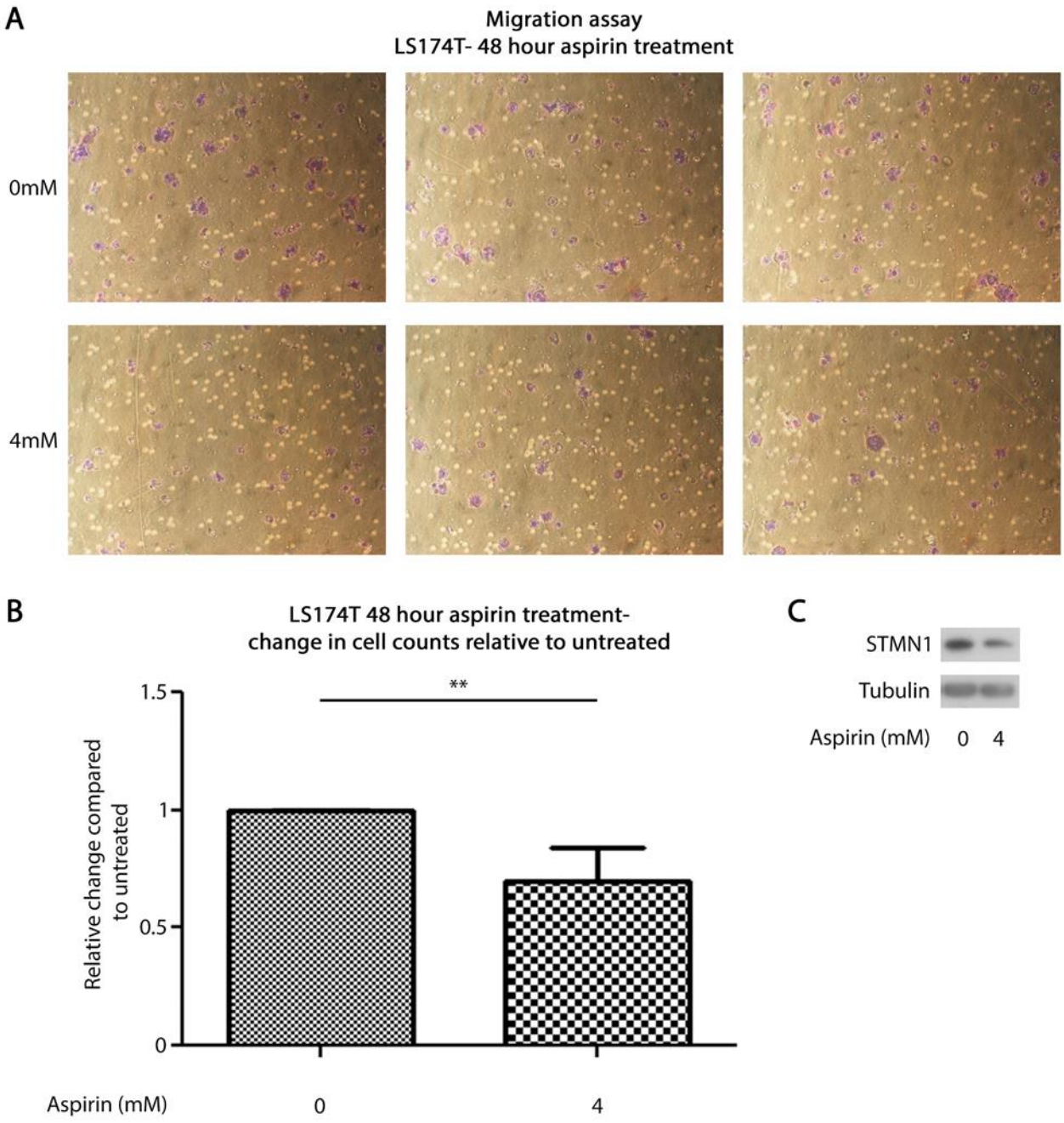


Figure 4-14- Aspirin reduces the number of LS174T cells that migrate through transwell pores.

LS14T cells were treated with 4mM aspirin for 48 hours in 2D cell culture. Cells were then extracted and seeded into transwells in FCS free medium. Wells were filled with FCS supplemented media to create a chemotactic gradient. After 24 hours, cells were fixed and stained using crystal violet before being counted. Results are mean number of cells from 4 repeat experiments with standard deviations and Student's t-test (*** $p \leq 0.001$, **$p \leq 0.01$ and * $p \leq 0.05$).. (A) Transwell images for one of the four experiments showing the pores (white) and the crystal violet stained cells that have moved through the pores (purple). These images show that there were less cells that migrated through from the 4mM aspirin treated samples. (B) Aspirin significantly reduced the amount of cells that migrated through the transwell pores ($p \leq 0.01$). (C) Western samples set up in parallel with the experiment were used to confirm STMN1 knockdown. Tubulin was used as a loading control. The results show that 4mM aspirin treatment reduced the protein levels of STMN1 at 48 hours.

4.3.9 STMN1 knockdown decreases cell migration in LS174T cells

To address whether a reduction in STMN1 protein levels reduces migration in LS1174Ts, we measured migration after knocking down *STMN1* expression.

LS174T cells were treated with 25nM of STMN1 siRNA in 2D cell culture for 48 hours. Cells were then extracted, counted and seeded into transwells in FCS-free medium. To create a chemotactic gradient, wells were filled with FCS supplemented media. After 24 hours, cells on the bottom of the transwell were fixed, stained with crystal violet and counted. As before, western samples were set up alongside the experiment to confirm STMN1 knockdown and tubulin was used as a loading control.

The results show that 48 hours of STMN1 siRNA treatment reduces the amount of cells that migrate through the transwell pores (stained in crystal violet) (Figure 4-15 A). Relative to untreated, STMN1 siRNA reduces the number of cells to 0.26 (SD 0.14) ($p \leq 0.001$) (Figure 4-15 B). The western blots also confirm a decrease in STMN1 protein levels when treated with siRNA (Figure 4-15 C).

Our results confirm that STMN1 is required for cell migration and suggest that the reduction in metastasis in patients who take aspirin may be partially explained by aspirin's effect of decreasing STMN1 levels.

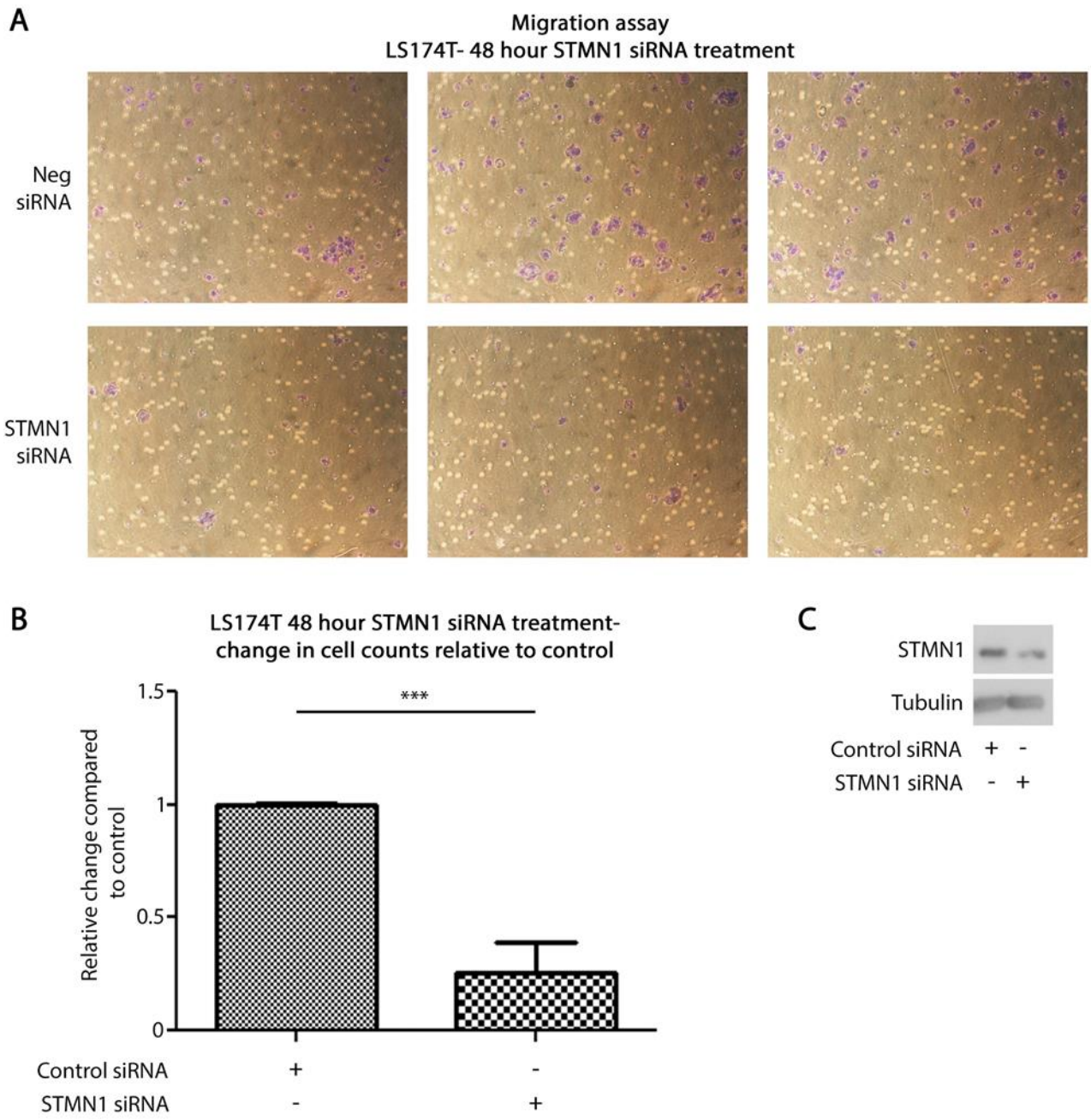


Figure 4-15- STMN1 knockdown reduces the number of LS174T cells that migrate through transwell pores.

LS14T cells were treated with 25nM STMN1 siRNA for 48 hours in 2D cell culture. Cells were then extracted and seeded into transwells in FCS free medium. Wells were filled with FCS supplemented media to create a chemotactic gradient. After 24 hours, cells were fixed and stained using crystal violet before being counted. Results are mean number of cells from 3 repeat experiments with standard deviations and Student's t-test (*** $p \leq 0.001$, ** $p \leq 0.01$ and * $p \leq 0.05$).. (A) Transwell images for one of the 3 experiments, showing the pores (white) and the crystal violet stained cells that have moved through the pores (purple). These images show that there were less cells that migrated through when STMN1 was knocked down. (B) STMN1 knockdown significantly reduced the amount of cells that migrated through the transwell pores ($p \leq 0.001$). (C) Western samples set up in parallel with the experiment were used to confirm STMN1 knockdown. Tubulin was used as a loading control. The results show that STMN1 was knocked down at 48 hours with STMN1 siRNA.

4.3.10 Aspirin's effect on the methylation site of STMN1

So far we have shown that aspirin reduces the protein level of STMN1 which may explain the decreased migration capability of aspirin treated cells. One question that we wanted to answer was how aspirin is reducing the protein level of STMN1. Our methylomic data shows that aspirin increases the methylation of the most significant cis CpG site (1mB either side of *STMN1*'s gene start site) by 26%. Therefore, we hypothesise that aspirin may be increasing the CpG methylation at the promoter of STMN1.

To address this question, we cultured RG/C2 cells for 24 hours with 0, 2 and 4mM aspirin in 2D cell culture. DNA was extracted and then quantified before being subject to bisulphite modification, PCR and pyrosequencing. Details on primer design and methods can be found in Chapter 2.11. We then applied a one-way ANOVA and Dunnett's multiple comparison test to calculate significance.

In order to carry out pyrosequencing, we first needed to check that the PCR amplified samples indeed contain DNA and that the primer is amplifying the desired product. The results show that DNA is present in all samples but there appears to be DNA contamination in the bisulphite modified control sample but no contamination in the control sample for PCR (Figure 4-16 A). The gel also shows that the primers are successfully amplifying the desired PCR product only.

After carrying out pyrosequencing, 24 hour aspirin treatment had no discernible effect on DNA methylation. At 0mM, the percentage methylation at the Illumina probe cg02490185 for the CpG site at chromosome 1 position 26136899 is 74.27% (SD 1.87) and the percentage methylation at 2mM and 4mM is 71.46% (SD 3.89) and 74.69% (SD 3.01), respectively (Figure 4-16 B and Table 4-6). The results of pyrosequencing differ to the results using the Infinium Human Methylation 450K BeadChip only for the effect of 0mM aspirin treatment on the methylation of the CpG site from the Illumina probe ID cg02490185. The pyrosequencing results show that 0mM aspirin treatment is associated with methylation at the CpG site at the level of 74.27% (SD1.87) whereas the Infinium Human Methylation 450K BeadChip shows that 0mM aspirin is associated with the methylation level 47.96% (SD 20.14). The results for 2mM and 4mM using both methods are similar.

Firstly, the pyrosequencing experiment requires replication as the agarose gel indicates DNA contamination in the bisulphite modified control and therefore we cannot be sure that the rest of the samples have not also been contaminated. If in fact the pyrosequencing results that we have observed are true, they indicate that short term aspirin treatment has no significant effect on the methylation of the CpG site measured by the Illumina probe cg02490185. However, it is worth noting that this CpG site may not have any effect on STMN1 expression therefore the methylation of other CpG sites associated with STMN1 expression need to be measured before any conclusions can be made about aspirin's effect on DNA methylation and STMN1 expression. There is also a need to assay more CpG sites across the gene locus and its regulatory elements.

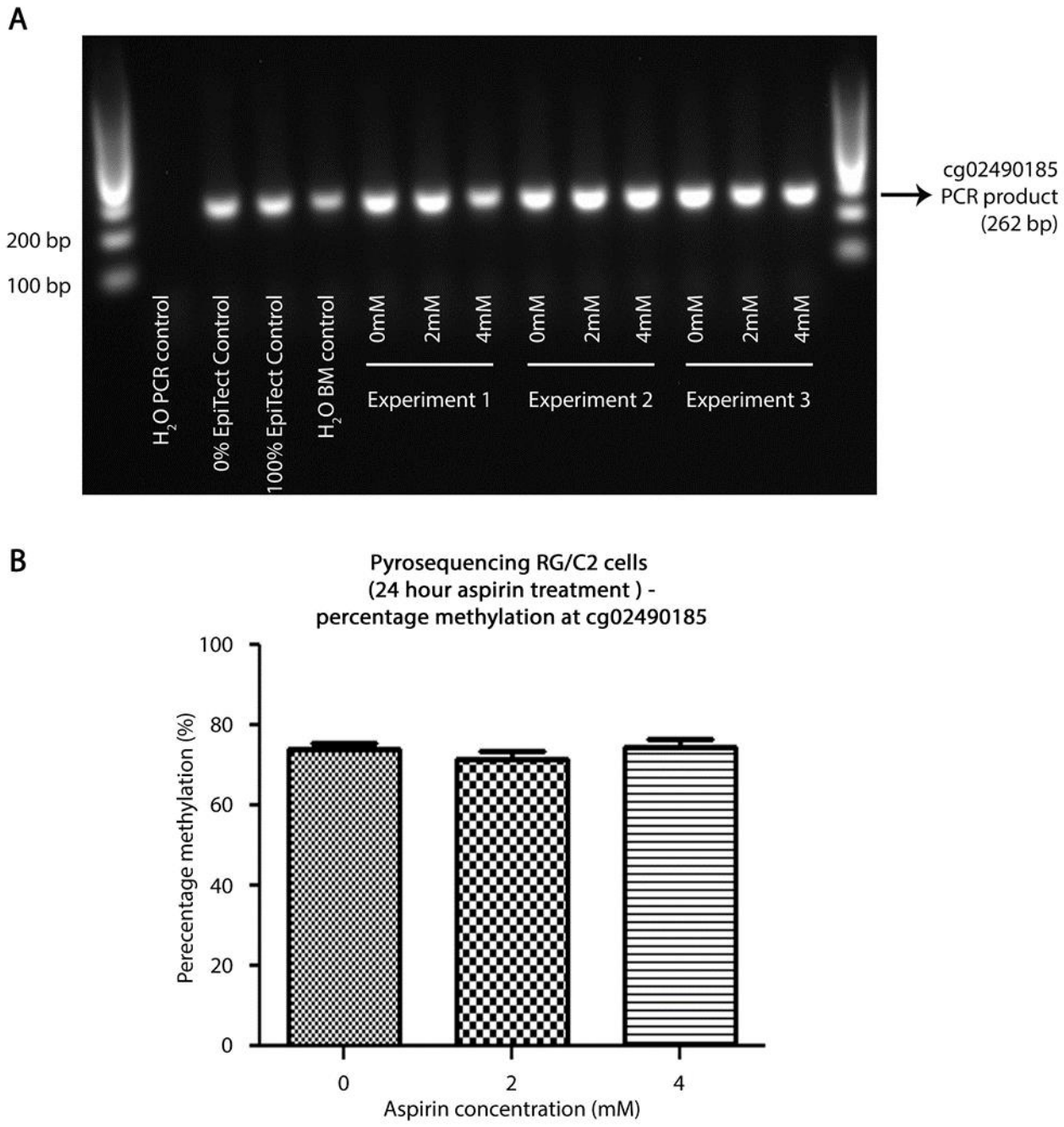


Figure 4-16- Short term aspirin treatment has no significant effect on the CpG site from the Illumina probe cg02490185 methylation

RG/C2 cells were cultured for 24 hours with 0, 2 and 4mM aspirin in 2D cell culture. DNA was extracted and then quantified before being subject to bisulphite modification, PCR and pyrosequencing. Results are mean values from 3 repeat experiments with standard deviations and a one-way ANOVA and Dunnett's multiple comparison test to calculate significance (*** $p \leq 0.001$, **$p \leq 0.01$ and * $p \leq 0.05$). (A) The agarose gel shows that the primers successfully produce the desired PCR product and that there is DNA present in all samples. The gel also shows that there appears to be DNA contamination in the bisulphite modified control (H_2O BM control) but not the PCR control (H_2O PCR control). (B) Aspirin has no significant effect on the methylation of the CpG site at the Illumina probe ID cg02490185.

Table 4-6- Average percentage methylation of the CpG site at the Illumina probe ID cg02490185 using the Infinium Human Methylation 450K BeadChip and pyrosequencing after 24 hours of aspirin treatment in RG/C2 cells

	Infinium Human Methylation 450K BeadChip		Pyrosequencing	
Aspirin concentration	Average percentage methylation	SD	Average percentage methylation	SD
0mM	47.96	20.14	74.27	1.87
2mM	73.13	4.63	71.46	3.89
4mM	72.41	16.82	74.69	3.01

4.4 Discussion

By combining multiple 'omics, we set out to identify novel targets to explain the anti-cancer effect of aspirin. Aspirin has many targets, but these alone do not fully explain aspirin's effect on cancer cells (Chapter 4.1.1 and Chapter 4.1.2). The combination of omics has many uses, one of which is to identify new potential therapeutic targets and predictive biomarkers for cancer as well as improving classification of cancer tissue (361). Through combining proteomic, methylomic and transcriptomic datasets, we identified 7 potentially new targets of aspirin. After carrying out qPCR, we confirmed that aspirin decreases gene expression of *MCM5* and *STMN1*. We focused on *STMN1* gene and protein expression as its' mRNA and protein levels have previously been reported to be higher in colon tumours compared to adjacent normal tissue ($p<0.05$) in Chinese patients and that the expression of this protein was significantly associated with TNM staging and lymph node metastasis ($p<0.05$) (359).

Whilst we showed that *STMN1* gene and protein expression was reduced in adenoma (RG/C2s) and carcinoma cells (LS174Ts) when treated with aspirin, this was not the case in metastatic cells (SW620s) indicating that aspirin's effect on *STMN1* expression is most effective in pre-metastatic cells. Our results indicate that aspirin is most effective on adenoma cells as it significantly reduces *STMN1* gene expression at both 2 and 4mM aspirin whereas it only reduces *STMN1* gene expression in LS174Ts at 4mM. Therefore, these results suggest that aspirin is a more effective chemopreventative agent for early stage lesions as the more advanced the tumour is, the less sensitive it is to aspirin treatment with regards to *STMN1* regulation. To confirm this *in vitro* observation, the expression of this protein needs to be investigated *in vivo* using CRC patient biopsies of aspirin-users at different stages of the tumour progression pathway and non-users as well as normal colon samples.

Stathmin-1 is a microtubule destabilizer and plays a role in cell cycle progression, cell survival, segregation of chromosomes and cell motility (362). The stathmin-1 protein has two domains to allow it to bind to tubulin and to promote microtubule catastrophe, as shown in Figure 4-17 (363). The microtubule-destabilising activity of stathmin was identified back in 1996. It was shown that stathmin has the ability to promote the microtubule catastrophe directly (364) and it also has the ability to form a stable T_2S ternary complex by sequestering the free αβ-tubulin heterodimers therefore preventing the incorporation of tubulin into the growing microtubules (365).

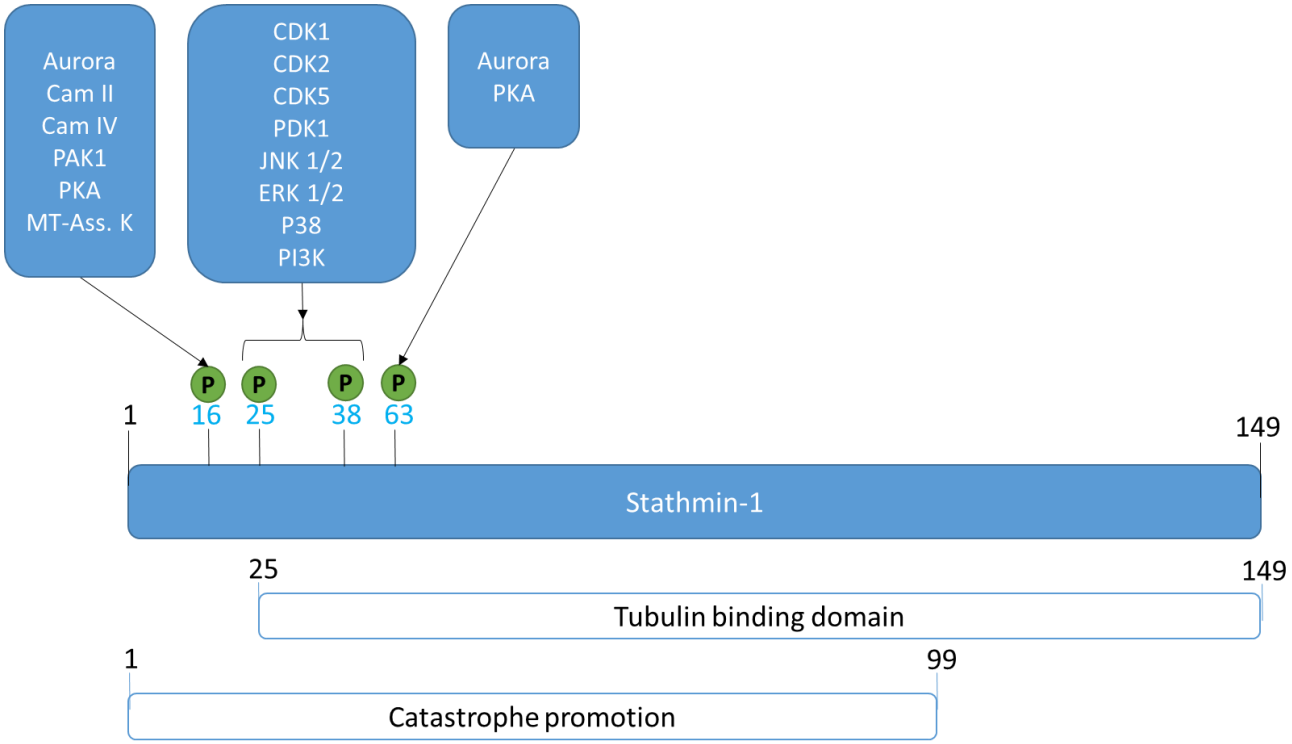


Figure 4-17- A diagram of the stathmin-1 protein

It has a sequence of 149 amino acids (aa) and two main domains: a catastrophe promoting domain at the N-terminus consisting of aa 1-99 that allows the protein to bind to microtubules and a tubulin binding domain at the C-terminus from aa 25-149 that allows the protein to sequester free αβ-tubulin heterodimers. The diagram also shows the four sites of phosphorylation (blue font) and the kinases responsible for this.

Stathmin-1 is also required during cell cycle progression. It undergoes phosphorylation during prophase to inhibit its activity and allow the formation of the mitotic spindle. As cells continue on to anaphase and telophase, stathmin is dephosphorylated so that the spindle can be disassembled through microtubule catastrophe (Figure 4-18) (363). Stathmin overexpression induces cells to accumulate in the G2/M phase of the cell cycle and may also partially inhibit the transition from prophase to metaphase. Also, substitution of the serine phosphorylation sites with alanine caused an accumulation of cells in mitosis before metaphase (366).

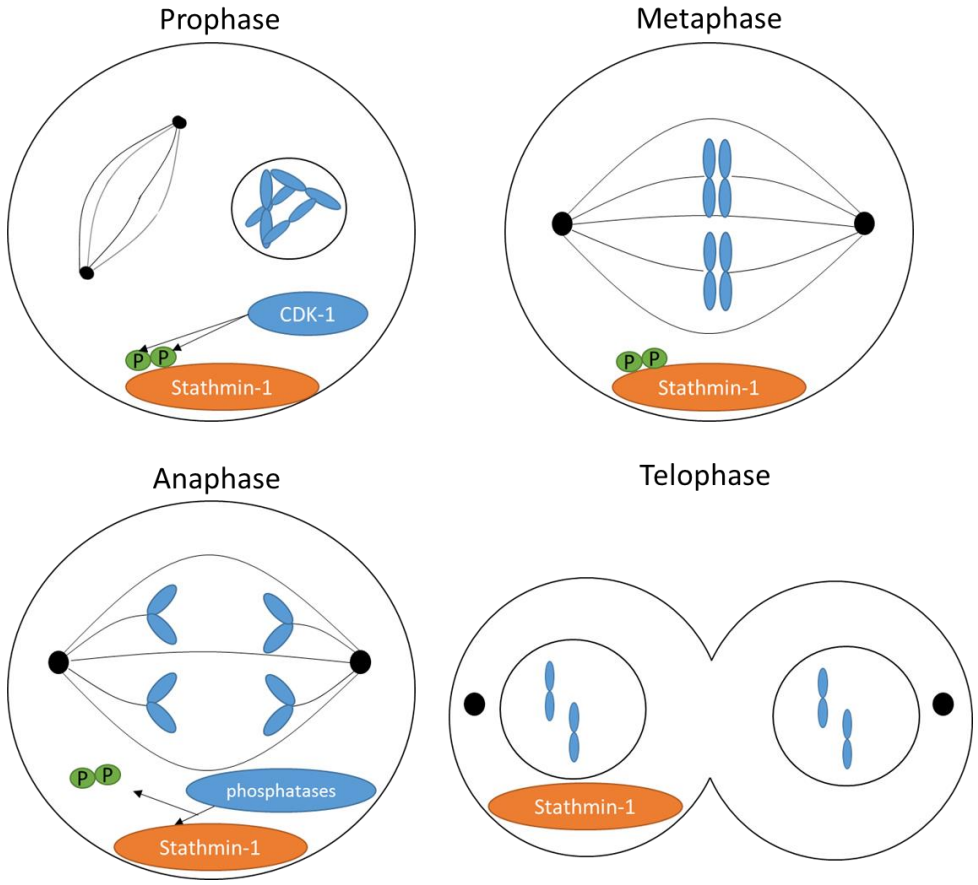


Figure 4-18- Diagram showing the involvement of stathmin during cell cycle progression.

During prophase, stathmin-1 is phosphorylated on Ser-25 and Ser-38 by CDK-1 to inhibit its activity. After chromosome segregation, the mitotic spindle is no longer needed, therefore stathmin-1 is dephosphorylated and promotes microtubule catastrophe.

Stathmin is required for correct formation of the mitotic spindle and to allow segregation of the sister chromatids. Increased expression/activity or depletion of stathmin affects the integrity of the spindle formation therefore results in chromosome misalignment. The consequence of this is the delay or prevention of M phase progression (363). For this reason, we investigated the effect of *STMN1* knockdown on cell proliferation and cell cycle progression. We found that knockdown had no significant effect on proliferation, although it did actually reduce the percentage of cells in G2/M phase of the cell cycle. This is consistent with stathmin-1 overexpression causing cell accumulation in G2/M (366). Our observation is in line with research involving other cancer cell lines that showed stathmin depletion resulted in a delay in the G2 phase and apoptosis (367). Carney et. al (2012) demonstrated that a reduction in stathmin expression delayed cells in G2 which they suggested maybe through microtubule stabilisation (367).

STMN1 has also been shown to be associated with metastasis in cancers such as pancreatic and breast cancer as well as colon (359,368,369). This is of particular interest as aspirin reduces the odds of colorectal adenocarcinoma metastasis by 64% (OR:0.36 (CI: 0.18-0.74)) (114). Therefore, we hypothesised that reduction in expression of STMN1 caused by aspirin may be playing a role in reducing cancer migration and hence metastasis. Our results show that aspirin reduces the migration of colorectal carcinoma cells. A more

apparent reduction in migration compared to aspirin treated cells was observed when *STMN1* expression is knocked down, leading us to hypothesise that aspirin reduces *STMN1* expression which reduces cell migration. However, to confirm that aspirin is reducing cell migration through decreasing STMN1 protein expression, cell migration experiments need to be repeated whereby aspirin treated cells are also transfected with *STMN1* to rescue the phenotype. If the reduction in cell migration is reversed, then this gives more evidence that aspirin is reducing cell migration through decreasing *STMN1* expression.

Although STMN1 was considered to only affect microtubule stabilisation, results by Tan et. al (2018) revealed that STMN1 affects the expression and phosphorylation of the actin cytoskeleton regulator vasodilator-stimulated phosphoprotein (VASP) (370)- these proteins are involved in promoting actin filament elongation (371). VASP can be phosphorylated at Ser-157 and Ser-239. Phosphorylation at Ser-157 results in the binding of phosphorylated VASP to tubulin and may play a role in assembling and stabilising the spindle fibres and promoting cancer metastasis (370,372). Tan et. al (2018) showed that silencing of *STMN1* in HCT-116 cells results in a decrease of VASP Ser-157 phosphorylation suggesting a mechanism by which aspirin may reduce cell migration and therefore metastasis (370). It is also becoming increasingly clear that microtubules may play a role in cancer migration as evidenced by Panopoulos et. al (2011) who showed that glioblastoma motility occurred through microtubules in the absence of actin polymers (373). The direct mechanism of how STMN1 may be affecting cell migration needs to be investigated further.

The method by which aspirin affects STMN1 expression is unclear. We carried out pyrosequencing of the lowest P-value cis CpG site (defined as 1Mb either side of the gene start site) to the start site of *STMN1* (Illumina probe ID cg02490185) but the effect of aspirin on DNA methylation of this site was not significant. It may be that this site does not affect *STMN1* expression and aspirin may be acting on alternative CpG sites. One other factor to consider is that the cell line used to look at the effect of DNA methylation is RG/C2s and their doubling time is 24 hours. Changes or maintenance of DNA methylation occur during DNA replication (374). Replicating DNA can be actively methylated and maintained through the enzyme DNA (cytosine-5)-methyltransferase 1 (DNMT1) (175,375) or passively demethylated through the absence of the DNA methylation maintenance enzymes during replication (376). It may be that 24 hours is not a long enough time period to observe significant changes in DNA methylation caused by aspirin. Therefore, in order to obtain more conclusive results, the ideal would be to repeat this experiment in more than one cell line at different time points and/or measure DNA methylation levels at alternative or additional CpG sites. In the future, it would be important to obtain methylomic data from CRC patients who take aspirin and compare them to patients who are non-aspirin users.

Whilst we did not investigate it further, aspirin treatment also reduced *MCM5* gene expression. Minichromosome maintenance (MCM) proteins work together to form a helicase complex consisting of 6 homologous proteins. It consists of MCM2-MCM7 and this hexamer is involved in unwinding double stranded DNA during DNA replication (377). Downregulation of MCM proteins has been reported to induce

G1 and G2 arrest in colon adenocarcinoma cells (378). This correlates with our results (Chapter 3.3.1.3) that aspirin induces cell cycle arrest in adenoma cells and it may be through the downregulation of *MCM5* expression.

To identify new targets of aspirin, we combined both laboratory (proteome and cell culture results) and epidemiological (EWAS) approaches. To our knowledge, we have shown for the first time that aspirin affects the genetic and protein expression of *STMN1* and this may be one of the mechanisms by which aspirin reduces the risk of cancer metastasis. Further investigations on *STMN1* expression needs to be carried out *in vitro* to confirm this effect. These results may also have implications therapeutically, as new drugs may be designed to target STMN1 expression to reduce the risk of CRC metastasis and therefore avoiding the unwanted side effects of aspirin. Our results show the strength of combining different techniques which in combination serve to reduce or eliminate the biases of each individual approach used to address the hypothesis. In order for further questions in the aspirin and cancer field to be answered, approaches such as ours need to be undertaken such as elucidating the effect of long-term aspirin treatment on colorectal cancer.

Chapter 5 Long-term aspirin treatment and its effect on DNA methylation of adenoma cells

This chapter aims to address one of the possible mechanisms by which aspirin may have differing effects depending on short-term and long-term culture, this time focusing on the effect of long-term aspirin on DNA methylation, circled in red in Figure 5-1. Any significant targets identified from this analysis may help identify how or why patients become "resistant" to aspirin.

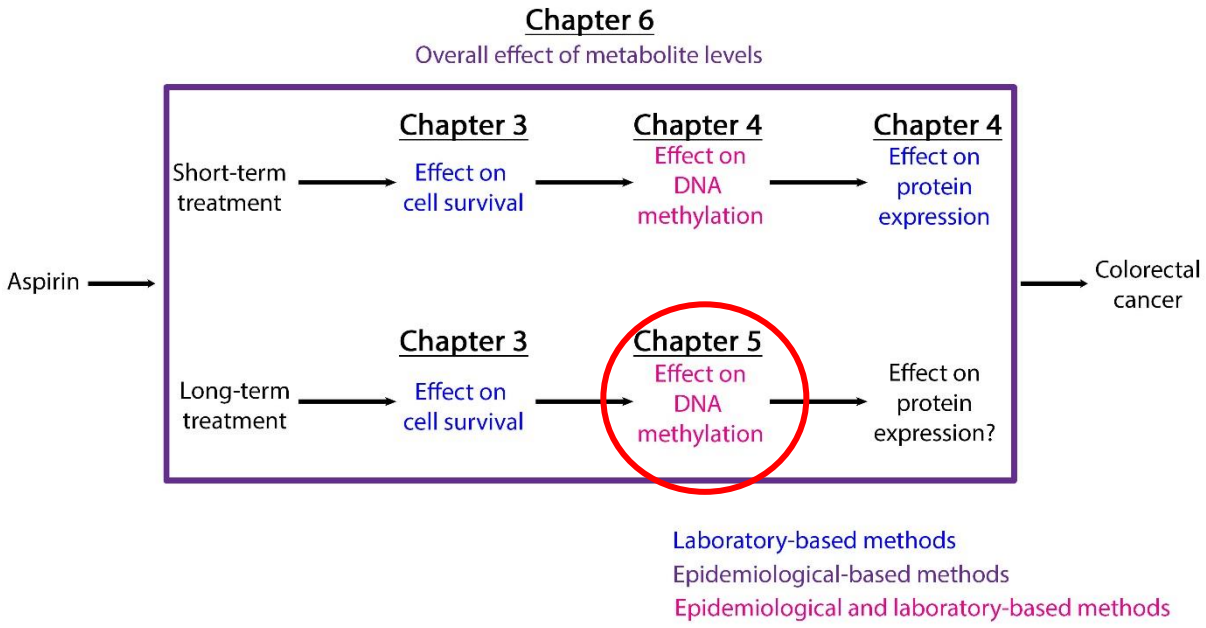


Figure 5-1- Overview of the questions addressed in this thesis to explore causality and potential mechanisms of aspirin on CRC.

5.1 Background

The observational studies summarised in Chapter 3.1.1 suggest that aspirin may be exerting different effects depending on whether it is taken for a short time period (short-term) or whether it is taken for a longer time period (long-term) and we hypothesise that long-term exposure may result in an acquired "resistance". For the purpose of this chapter, aspirin "resistance" denotes a decreased sensitivity to the protective effect of aspirin. Aspirin plays a role in epigenetic regulation, of which its effect on acetylation is the most well-known (Chapter 4.1.3 and Chapter 4.1.4). We have previously shown that short term aspirin treatment (24 hours) has a relatively small effect on DNA methylation in colorectal adenoma cells (Chapter 4.3.1). Since aspirin appears to have a different effect when taken short-term or long-term, we set out to identify whether this was through its effect on DNA methylation.

5.1.1 Colorectal cancer causal pathway

There are many potential mechanistic pathways that may contribute to the colorectal cancer development. These include genetics, epigenetics and the microbiome. With regards to genetics, it is well understood that sporadic tumours may occur through the CIN pathway which leads to an abnormal karyotype (46,47), and/or through the microsatellite instability (MSI) pathway, whereby inefficient repair of insertions and/or deletions leads to DNA frameshifts and possibly truncated proteins (Chapter 1.3.3) (46,48).

As well as genetic contributions to disease progression, there is also an epigenetic component. Some sporadic colorectal cancers exhibit CpG island methylation phenotype (CIMP) pathway resulting in hypermethylation at some CpG sites (49). Furthermore, it has recently come to light that cancer progression can be affected by bacterial species, for example, colorectal cancer metastasis is associated with the presence of the *Fusobacterium* species of bacteria (379). Not only that, the intestinal microbiota has a major effect on patient response to immunotherapy (54).

It is thought that a combination of all of these factors may affect both risk and progression of cancer.

5.1.2 Long term aspirin and DNA methylation

Despite evidence to support aspirin in epigenetic regulation (Chapter 4.1), investigation into the effect of long-term aspirin use on DNA methylation in colorectal cancer is lacking. One evidence for the effect of long-term treatment on healthy colon comes from Noreen et. al (2014) who explored the effect of long-term aspirin (at least 2 years or more) use in 546 healthy women and showed that long-term aspirin use was associated with a >50% suppression of rate of methylation at promoter for *hMLH1* and *MGMT* when compared to non-users (249).

In human gastric mucosa, Tahara et. al (2010) showed that use of aspirin for around 2 years was associated with reduced E-cadherin 1 (*CDH1*) methylation although state that the exact mechanism by which aspirin is able to affect methylation is unclear (380). Michigami et. al (2017) took this one step further and investigated the effect of long-term aspirin use (3 years or more) on the methylation status of seven genes in patients with atrophic gastritis and gastric cancer. They concluded that aspirin does affect certain methylations sites in patients with chronic gastritis who have precancerous conditions; aspirin was shown to reduce *CDH1* methylation in the atrophic mucosa as well as reducing CIMP methylation in intestinal metaplasia (381).

Evidence for the effect of other non-steroidal anti-inflammatory drugs (NSAIDs) on DNA methylation was also carried out by Tahara et. al (2009) who investigated the effect of aspirin, loxoprofen sodium, lornoxicam and etodolac on four promoters in non-neoplastic gastric mucosa. These promoters are: p14 (ARF), p16 (INK4a), death-associated protein kinase (*DAP-kinase*), and E-cadherin (*CDH1*). Their results show that chronic NSAID use (at least 3 months) was associated with a reduction in methylation of these four sites (382).

Wilson et. al (2015) examined the effect of regular NSAID use (defined as taking an NSAID at least triweekly for 3 months or more) on DNA methylation of 27589 CpG sites using the Infinium HumanMethylation27 BeadChip (27K array) in whole blood from 871 women as a discovery sample and then examined 485512 CpG sites using the Illumina Infinium HumanMethylation450 BeadChip 450,000 CpG site platform (450K array) in 187 women as a replication sample. Unfortunately, they could not replicate the results from their discovery sample in the smaller replication set and concluded that NSAID usage may only have a small effect on DNA methylation (383).

Overall, it is evident that the literature lacks information on the epigenome-wide effect of long-term aspirin treatment, and this is not only limited to colorectal cancer. There is also a lack of suitable models (cell lines or mouse models) to study the effect of long-term aspirin treatment on DNA methylation.

5.1.3 Methods for assessing DNA methylation

5.1.3.1 MethylationEPIC BeadChip (Infinium) microarray

Whilst bisulfite genomic sequencing of multiple clones is the most comprehensive method to determine the methylation status of CpG sites in a DNA sequence, this method is limited because it is difficult to use at a large scale for monetary and time reasons. For this reason, DNA methylation arrays were developed and are more cost-effective and labour-efficient (210). The Illumina Infinium HumanMethylation450 BeadChip 450,000 CpG site platform (450K array; Illumina Inc., CA, USA) has been widely used to look at changes in DNA methylation (171). However, a new array was developed, MethylationEPIC BeadChip Infinium (850K array), to include over 850,000 CpG sites. The purpose behind this was to incorporate 333,265 CpG sites that are located in enhancer regions (i.e. gene regulatory regions) (210) that were identified by the projects FANTOM5 (384) and ENCODE (385,386) which are both online resources for DNA data (such as enhancer regions, promoters, transcription factors) in the human genome. A summary of the location of the CpG sites from the 850K array in relation to gene regions and CpG context as well as the percentage of sites associated with RNA transcripts is summarised in Table 5-1, Table 5-2 and Table 5-3. If aspirin is affecting DNA methylation, it is possible that it targets specific sites based on genetic regions or relative to CpG islands.

Table 5-1- Percentage of CpG sites for specific gene regions in the 850K array

Gene region	Percentage of CpG sites
Proximal promoters: -200 bp upstream of transcription start site -1500 bp upstream of transcription start site -Exon 1 -5'UTR -3'UTR	25.8% 5.4% 10.6% 0.4% 9.4% 0.9%
Gene bodies	40%
Intergenic regions	33.3%

Table 5-2- Percentage of CpG sites located in relation to CpG islands in the 850K array

CpG context	Percentage of CpG sites
North shelf	9.63%
North shore	3.69%
Island	18.65%
South shore	8.22%
South shelf	3.43%
Open sea	56.38%

Shores are considered 2-kb sequences either side of the CpG island, shelves are 2-kb sequences next to the shore. Open sea regions are CpG dinucleotides not found near the CpG islands or the shores and shelves. Abbreviations: N, north; S, south.

Table 5-3- Percentage of CpG sites associated with RNA transcripts in the 850K array

RNA transcript	Percentage of CpG sites
coding messenger RNA	74.4%
non-coding RNA (miRNA and long non-coding RNA)	0.85%
no associated transcript	24.6%

5.1.3.2 Inflation in EWAS

Inflation refers to an overestimation of statistical significance meaning that CpG sites that are not truly associated with aspirin treatment are incorrectly calculated as significant. Inflation of test statistics results in an overestimation of findings that are statistically significant and therefore increases the number of findings that are false positives (387). In the context of aspirin, this means that methylation of CpG sites that are not actually significantly altered by aspirin treatment are incorrectly calculated as being significantly associated. This phenomenon is common and a major concern in GWAS, but is also present in EWAS (388).

It is thought that the main cause of inflation and bias is through unmeasured biological and technical confounding (389). The term "confounding factors" refers to an unobserved variable that is correlated with both the variable of interest and the outcome variable, and this phenomenon is commonly known as confounding (390). This means that a variable other than our exposure of interest (that may not be accounted for) is also affecting the results thereby inflating the P-values to be more significant. For example, another variable that we did not control for might be truly influencing CpG site methylation rather than our exposure of interest (aspirin) itself. Confounding factors can be known, e.g. batch effects, or unknown and are therefore considered in statistical models in a variety of ways including as a named covariate in a statistical model (391) or as a hidden covariate, e.g. surrogate variables (392) or a latent factor (393).

Batch effects are measurements that may behave differently across different conditions and are not related to the scientific/biological variable in a study. Examples of this include the day each experiment was run or whether it was a different technician carrying out each experiment. It can also include whether different chips/reagents/instruments were used for each experimental subset. Whilst batch effects often play a role in high-through put genomics research, they can also affect experiments such as northern blots or quantitative PCR but is harder to detect in these assays (391).

Quantile-quantile (QQ) plots are plots that allow the visualisation of whether all the data points are more significantly associated with an exposure than is due to chance (Figure 5-2). These plot expected P-values (from a standard normal distribution) versus observed P-values (from our experiment) (389). If we take an example of the association of aspirin with 100 CpG sites, we can plot the P-values of our results against the P-values we would expect from a standard normal distribution. The standard normal distribution is standardised according to the z-scores. This score represents how many standard deviations away from the mean the value is. Therefore standardization involves converting the raw scores from your distribution into standard deviation units (394). In this case, the standard normal distribution would be divided into the same number of quantiles as our distribution of interest (CpG site methylation) so 100 quantiles. The corresponding z-score for each quantile can then be used to calculate P-values for each quantile. P-values from our test statistics are plotted against the P-values from the standard normal distribution. If the P-values from the methylation results follow a normal distribution, then a diagonal line can be drawn at a 45^{o} angle (Figure 5-2 A) (395). If there is a variable that is influencing all the results (systematic error) to be more

significant than we would expect according to a standard normal distribution, then all of the points would shift at an angle higher than 45^{o} indicating inflation (Figure 5-2 B) (389). Ideally, most of the P-values should follow a standard normal distribution but only a couple may be more associated with the exposure than is due to chance/what we would expect (C). Any points outside this line indicate that the theoretical distribution isn't correct or that some other factor is affecting the P-value such as a true association with the exposure (396)

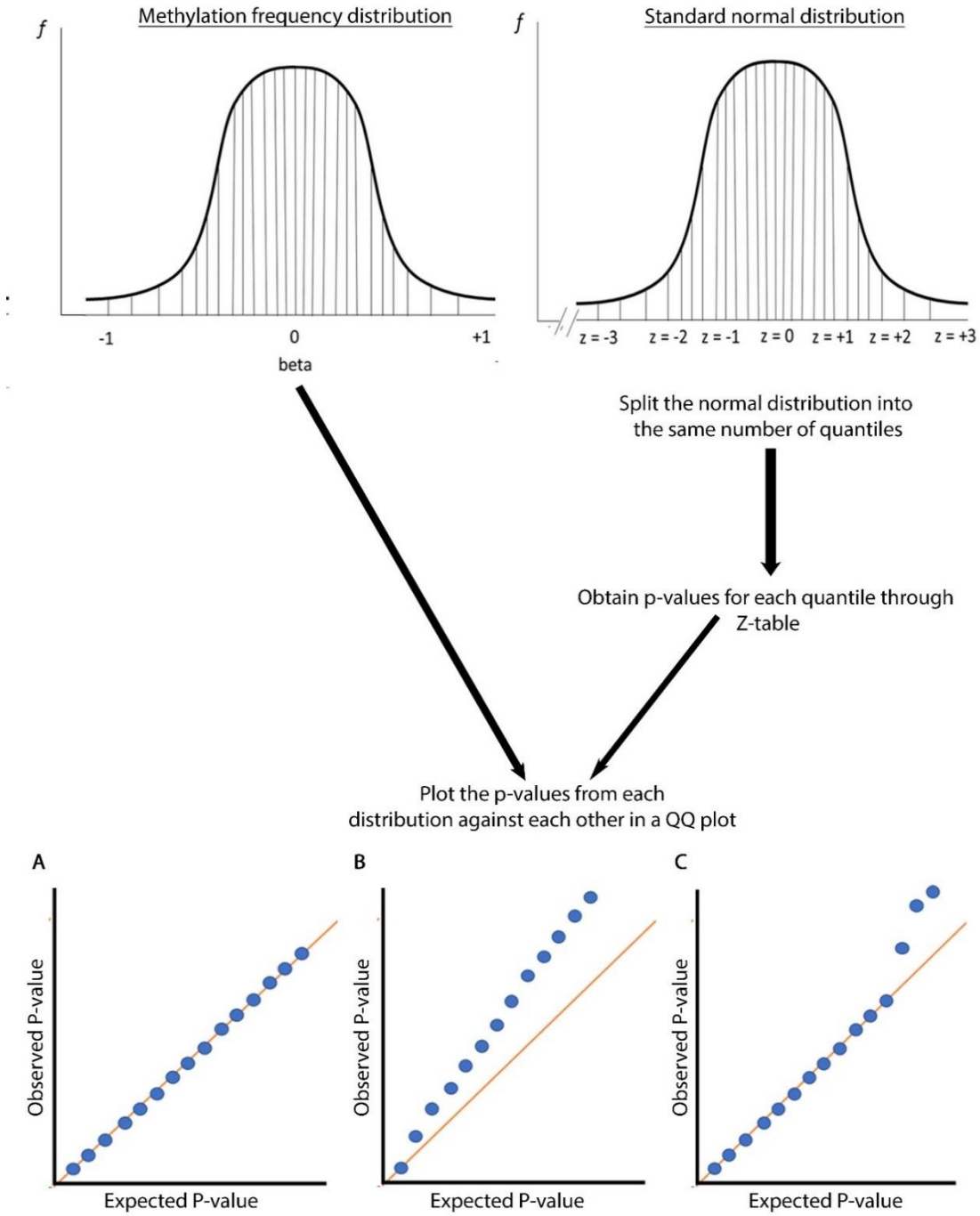


Figure 5-2- QQ plot explanation

QQ plots are plots that help visualise whether data follows a predicted P-value or whether the P-value of our CpG site methylation is more associated with the exposure than we would expect. Methylation results are split into quantiles which means they are split into the total number of CpG sites measured, e.g. 100. The standard normal distribution is standardised according to the z-score. This score represents how many standard deviations away from the mean the value is. The standard normal distribution is also split according to the same number of quantiles (100) as the CpG methylation beta values. These z-scores can then be used to calculate P-values for each quantile. P-values from our distribution of test statistics are plotted against the P-values from a standard normal distribution. If the P-values from the methylation results follow a normal distribution, then a diagonal line can be drawn at a 45^{o} angle (A). If there is another factor shifting all the P-values away from the standard normal distribution, then we can deduce that there may be a variable not adjusted for that is confounding all the results (B). Ideally, most of the P-values should follow a standard normal distribution but only a couple may be more associated with the exposure than is due to chance/what we would expect.

5.1.3.3 *Surrogate variable analysis*

Confounding is a common issue within epidemiological studies. Even a well-designed study may experience confounding due to logistical reasons or variations in sample quality. One major issue is when it is not known what these confounding factors are (392). A possible consequence of confounding is that it may hide true associations turning them from true positives to false negatives or the opposite (397).

DNA methylation variation may be influenced by other factors other than the primary variable of interest. These sources can be categorised as 3 factors:

- Unmodelled factors: these are other variables that have been measured but just not included in the statistical model due to reasons such as sample size restrictions.
- Unmeasured factors: these are variables that may cause variation in DNA methylation but were not measured
- Gene-specific noise: random fluctuations in the expression of genes independent of one another.

To deal with these confounders, Leek and Storey (2007) developed an algorithm called Surrogate Variable Analysis (SVA) (392). This method seeks to identify groups of genes that are affected by each unobserved factor and then estimates the effect of this factor based on the expression of these genes. (392) Whilst SVA was developed for gene expression studies, the same basic ideas can apply to DNA methylation studies. We used SVA to adjust for confounding in our results.

5.1.3.4 *Adjusting for multiple testing*

A Bonferroni and FDR-adjusted P-value were both calculated to adjust for the issue of multiple testing. These have been described in more detail in Chapter 4.1.7.2

5.1.3.5 *Interpreting EWAS results*

Through the incorporation of labelled nucleotides, the Illumina methylation arrays have been designed so that methylated CpG sites emit a different fluorescence to unmethylated CpG sites. The intensity of the fluorescence is measured and a ratio of the fluorescent signal of the methylated site to the signal from the unmethylated and methylated combined is calculated (338). This ratio is called beta (β). It's values range from 0 to 1 and roughly represents changes in percentage methylation where 0 is equal to 0% and 1 is equal to 100% (339). These beta values can be positive to represent an increase in methylation or negative to represent a decrease in methylation. These beta values allow for direct biological interpretation of results, unlike other methods (340). According to a review by Michels *et. al,* they suggested that methylation changes of <5% (β=0.05) should be interpreted with caution but changes of >10% (β=0.10) provide increased confidence in the biological significance of the result (342).

5.1.4 Hypothesis and aims

Hypothesis:

Long-term aspirin treatment is associated with differences in DNA methylation and hence gene expression.

Aims:

1- To identify whether long-term aspirin treatment is associated with differences in DNA methylation in colorectal adenoma cells.
2- To validate differentially methylated CpG sites using pyrosequencing

5.2 Methods

5.2.1 DNA methylation analysis- variables

5.2.1.1 Long-term aspirin treatment

The exposure variable in this analysis was the long-term treatment (75 weeks) of RG/C2 cells with aspirin. RG/C2 cells were cultured in 2D with 0mM, 0.5mM, 2mM and 4mM aspirin with 3 biological repeats resulting in a total number of 12 samples.

Aspirin treatment was considered a binary exposure variable to compare untreated (0mM) vs treated (0.5mM, 2mM and 4mM). This was because initial exploratory analysis showed no indication of a dosage effect.

5.2.1.2 DNA methylation

The main outcome measure in this analysis was the level of DNA methylation at each of the CpG sites in the 850K array for RG/C2 cells treated with 0mM, 0.5mM, 2mM and 4mM aspirin. Laboratory methods, QC and pre-processing of the methylation assay was described in Chapter 2.10. All 12 samples passed quality control measures (>10% probes with a detection P-value ≥ 0.01).

5.2.2 DNA methylation analysis- Multivariable regression analysis

Using DNA samples obtained from RG/C2 cells, methylation at 865,859 CpG sites was investigated using the 850K array. Samples were distributed randomly on the arrays.

We conducted multiple linear regressions with the exposure defined as aspirin treatment as a binary variable (untreated (0mM) vs treated (0.5mM, 2mM and 4mM)) and the outcome defined as methylation (untransformed β-values). Analyses were run with adjustment for batch effects only.

DNA methylation sites were annotated using the Meffil package (264). CpGs considered to be "EWAS-significant" were those which were below the Bonferroni corrected threshold using the total number of tests performed ($P=0.05/865,859 = 5.77x10^{-8}$). CpGs with values below this are considered true positives and worth investigating. The limitation of the Bonferroni correction is that it assumes that all CpG sites are independent of each other when this may not be true. For this reason, we also applied a less conservative test called the false discovery rate (FDR) which is based on the method by Benjamini-Hochberg also used to

adjust for multiple testing (351). Using this method, CpG sites with an FDR adjusted P -value below a 0.05 threshold were considered EWAS-significant.

5.2.3 DNA methylation- Surrogate variable analysis

To try and adjust for unknown confounders, we adopted the Surrogate Variable Analysis developed by Leek and Storey (392). This method seeks to identify groups of loci that are affected by each unobserved/unknown factor and then estimates the effect of this factor based on DNA methylation. We included 6 surrogate variables in the linear regression model as these were the maximum number of variables that could be calculated using the sva package developed by Leek et. al (2012) (398) calculated in the software: R (version 3.3.1).

Therefore the multiple variable regression was repeated with the exposure defined as aspirin treatment as a binary variable (untreated (0mM) vs treated (0.5mM, 2mM and 4mM) and the outcome defined as methylation (untransformed β-values) but this time adjusting for batch effects and 6 surrogate variables.

5.2.4 Power calculations

Power calculations were carried out using the statistical software G*Power 3.1.9.2 (352). Our study had 80.6% power to detect a mean methylation difference between aspirin treated and untreated samples of 0.18 beta (or 18% change in methylation) with a standard deviation of 0.1 for 9 treated samples vs 3 untreated samples at an α of 0.05.

5.3 Results

5.3.1 Multivariable regression

The literature has alluded to the fact that aspirin can affect protein and histone acetylation (Chapter 4.1). However, it's effect on DNA methylation is unclear. The literature suggests an inverse association between histone acetylation and CpG site methylation (Chapter 4.1.6). For this reason, we sought to investigate the effect of long-term aspirin on DNA methylation.

RG/C2 cells were cultured with 0mM, 0.5mM, 2mM and 4mM aspirin continuously for 75 weeks. DNA was extracted, and bisulphite converted. Each repeat was set up on separate dates and DNA was extracted from cells when they reached a certain confluency. DNA methylation was measured using the MethylationEPIC BeadChip Infinium array and samples were randomly distributed on the array. Multivariable regression was carried out to compare the effect of aspirin treatment (0.5mM, 2mM and 4mM) versus no treatment on DNA methylation, adjusting only for batch. Results were plotted as Manhattan plots using the qqman package (353) to show the P-values of each CpG site. We used the Bonferroni method to adjust for multiple testing (Chapter 5.1.3.4) therefore we divided a P-value of 0.05 by the total number of CpG sites on the array (865859) to obtain a Bonferroni adjusted P-value of 5.77×10^{-8}. This was done to reduce the number of results that are falsely associated with aspirin treatment (previously explained in Chapter 5.1.3.4). The

results show that 1562 CpG sites fell below the Bonferroni P-value threshold for significance of 5.77×10^{-8}. Since a high number of CpG sites fell below this threshold, we decided to use the more stringent threshold of 5×10^{-8} as is currently used for genome wide association studies of which 1492 CpG sites fell below this P-value.

We also used an FDR adjusted P-value which informs us of the probability of a result being a false positive when it is truly from the null associations (Chapter 5.1.3.4). We identified that 27709 CpG sites fell below the FDR value of 0.05 which means that the probability of these CpG sites being false positive is 5% or less (Figure 5-3 A).

Since a large quantity of CpG sites fell below the more stringent P-value threshold for significance (5×10^{-8}), we plotted a QQ plot to identify whether inflation (too many CpG sites associated with aspirin) due to possible confounders of results was present. Quantile-quantile (QQ) plots are plots that allow the visualisation of whether all the data points are more significantly associated with an exposure than is due to chance. These plot expected P-values (from a standard normal distribution) versus observed P-values (from our experiment) (389). Quantile-quantile plots are explained in more detail in Chapter 5.1.3.2. In this graph, the minus log10-transformed P-values of the results are plotted against P-values expected from a null distribution. The genomic inflation factor (λ) was used to compare the distribution of the P-values from the epigenome-wide distribution with those from the expected null distribution. If there is no inflation, the value of λ would be equal to 1. Our results show that there is strong inflation (λ=2.27) (Figure 5-3 B) indicating that there may be some confounding not adjusted for in our multivariable regression model.

Overall, it appears that aspirin is affecting the methylation of a large number of CpG sites, however, our QQ plot indicates that this association may be due to confounders systematically influencing methylation of all the CpG sites.

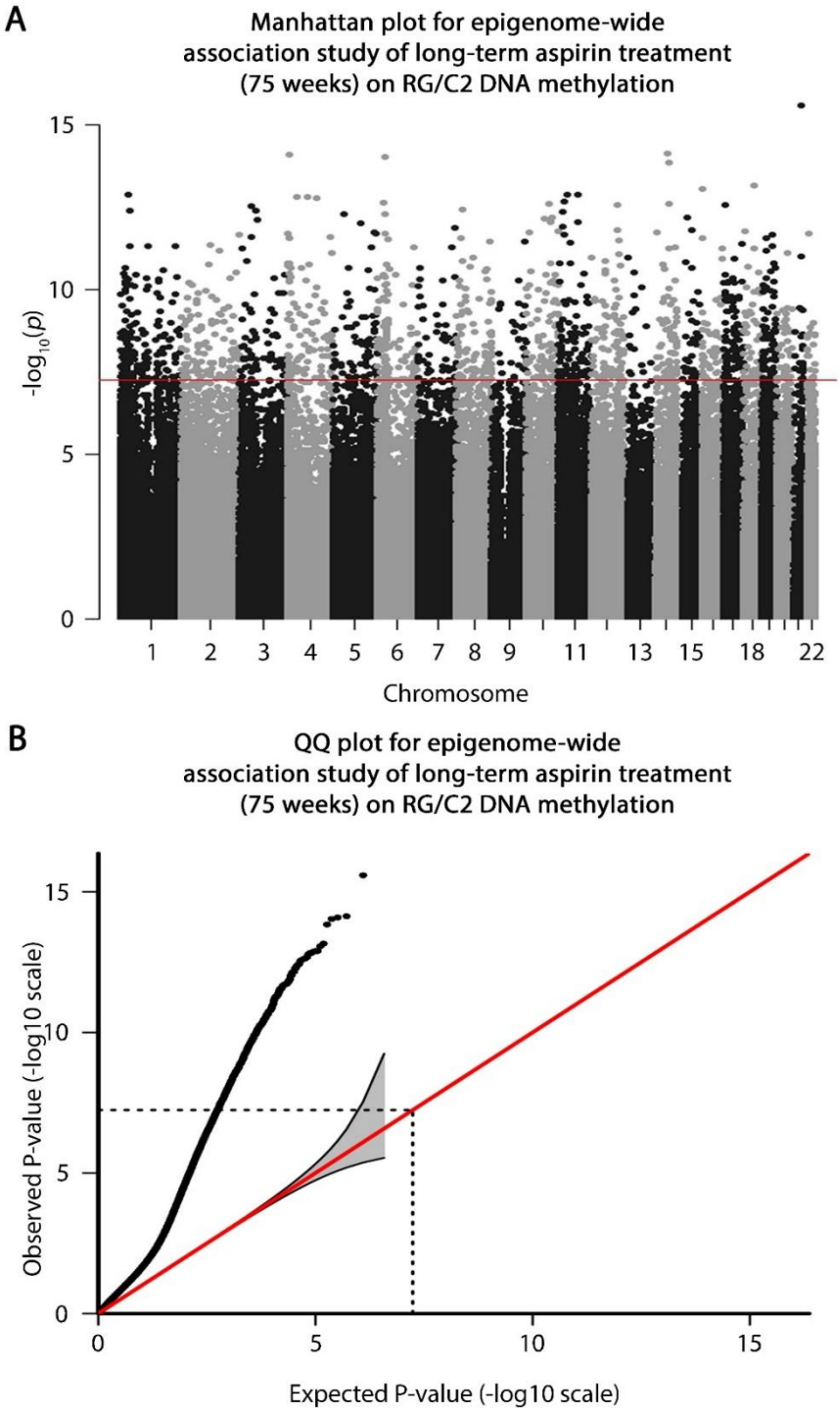


Figure 5-3- Manhattan and QQ plot for epigenome-wide association study of long-term aspirin treatment on RG/C2 cells DNA methylation.

RG/C2 cells were treated with 0, 0.5, 2 and 4mM aspirin for 75 weeks. DNA was then extracted, bisulphite converted, and methylation was measured using the MethylationEPIC BeadChip Infinium array. Multivariable regression was carried out adjusting only for batch. (A) The Manhattan plot shows each CpG site, displayed as a dot, plotted according to its genomic co-ordinates along the X axis and plotted as the negative logarithm of the P-value for association of aspirin with CpG site methylation along the Y axis. The plot shows that 1492 CpG sites fell below the more stringent P-value threshold for significance (5×10^{-8}) (red line) and 27709 fell below the FDR adjusted P-value cut-off of 0.05. (B) The QQ plot shows the minus log10-transformed P-values of long-term aspirin treatment on CpG site methylation corrected for batch effects against quantiles from the null distribution. The red line indicates the line of equality whereby the P-values from our study should be plotted if they followed a null distribution. The 95% confidence intervals are shaded in grey and the black lines indicate the 0.05 P-value significance level. Strong inflation (inflation factor $\lambda = 2.27$) was observed.

5.3.2 Multivariable regression with surrogate variable analysis

Our results suggest that aspirin's effect on DNA methylation is confounded by unknown factors as indicated by the QQ plot (Figure 5-4 B) and inflation factor of 2.27. To try and adjust for these unknown confounders so that we identify the effect of aspirin only on DNA methylation, we adopted the Surrogate Variable Analysis developed by Leek and Storey (392). Through mathematical modelling, this method seeks to identify loci that are affected by each unobserved/unknown factor and then estimates the effect of this factor based on DNA methylation.

Multivariable regression was carried out to compare the effect of aspirin treatment (0.5mM, 2mM and 4mM) versus no treatment on DNA methylation, adjusting for batch and 6 surrogate variables. The results show that no CpG sites fell below the more stringent P-value threshold for significance (5×10^{-8}) and the FDR adjusted P-value cut-off of 0.05 (Figure 5-4 A). Our results also show that there is no evidence of inflation (λ=0.96) (Figure 5-4B) indicating that there may have been some confounding previously present that is now adjusted for in our multivariable regression model by including these surrogate variables.

To identify which results make more biological sense, a comparison was made of the raw average methylation of 0mM, 0.5mM, 2mM and 4mM aspirin treated cells of the top 10 most significant CpG sites from the models adjusted for batch only and the models adjusted for batch and 6 surrogate variables.

The results of the model adjusted for batch only show that the top 10 aspirin-associated CpG sites show a clear change in raw DNA methylation of each aspirin treatment. The most robustly associated CpG site cg05514510 (P-value-= 2.58E-16) showed that aspirin decreases CpG site methylation by 33% in 0.5mM, 2mM and 4mM treated cells when comparing the raw averages of DNA methylation (Table 5-4).

On the other hand, when fully adjusting for SVs, the results are greatly attenuated with no sites being FDR significant. The effect sizes predicted by fully adjusting are also difficult to interpret, for example, SVA predicts a change of -0.87 which equates to a decrease in 87% methylation at the Illumina probe ID cg16799926. However, the raw methylation values suggest a decrease of 38% from 0mM to any aspirin treatment (0.5mM, 2mM and 4mM). In this case, fully adjusting for SVs may be too stringent and may be removing the real signal from aspirin treatment. When calculating the correlation coefficient between the 6 SVs and aspirin concentration, the first SV is highly correlated with aspirin treatment with an R^2 of 0.96 indicating that this SV is simply adjusting for the effect of aspirin treatment.

Furthermore, the changes in average raw methylations of the most statistically significant CpG sites when including surrogate variables are not as obvious. The most significant CpG site after applying SVA is cg21232203 (P-value=2.01E-06) whereby aspirin decreases the percentage of methylation by 7% in 0.5mM treated cells and 5% in 2mM and 4mM treated cells when comparing the raw averages of DNA methylation (Table 5-5). In addition, the CpG site cg07001545 shows no difference in raw DNA methylation at all concentrations.

It may be that surrogate variable analysis is inappropriate to be applied to experiments carried out in a laboratory setting. Laboratory experiments are controlled environments where it is unlikely that unknown confounding will occur. In the case of this experiment, each set of repeat was set up on the same day, treated with aspirin at the same time and DNA was extracted when they reached a certain confluency. The only difference is the time point at which the cells were lysed due to the 2mM and 4mM treated cells growing much slower than the control and 0.5mM aspirin treated cells. Whilst controlling for unknown factors is important in conventional EWAS, it is unlikely that this is needed in the laboratory setting. To confirm whether aspirin is affecting levels of DNA methylation, applying a method that assays fewer sites and that is less susceptible to batch effects such as methylation-specific PCR or pyrosequencing needs to be used to study a subset of CpG sites. The level of inflation observed in the QQ plot remains a concern and may be indicative of technical issues with the analysis. Reassurance could be found by repeating the entire experiment and randomizing the samples to different arrays.

For the purpose of further analyses, the multivariable regression adjusting only for batch (i.e. results from the minimally adjusted model) will be used onwards as the controlled environment of the experiment makes it unlikely that external confounding factors are present. Therefore, we believe that aspirin may have a genome-wide effect on DNA methylation thereby influencing many CpG sites.

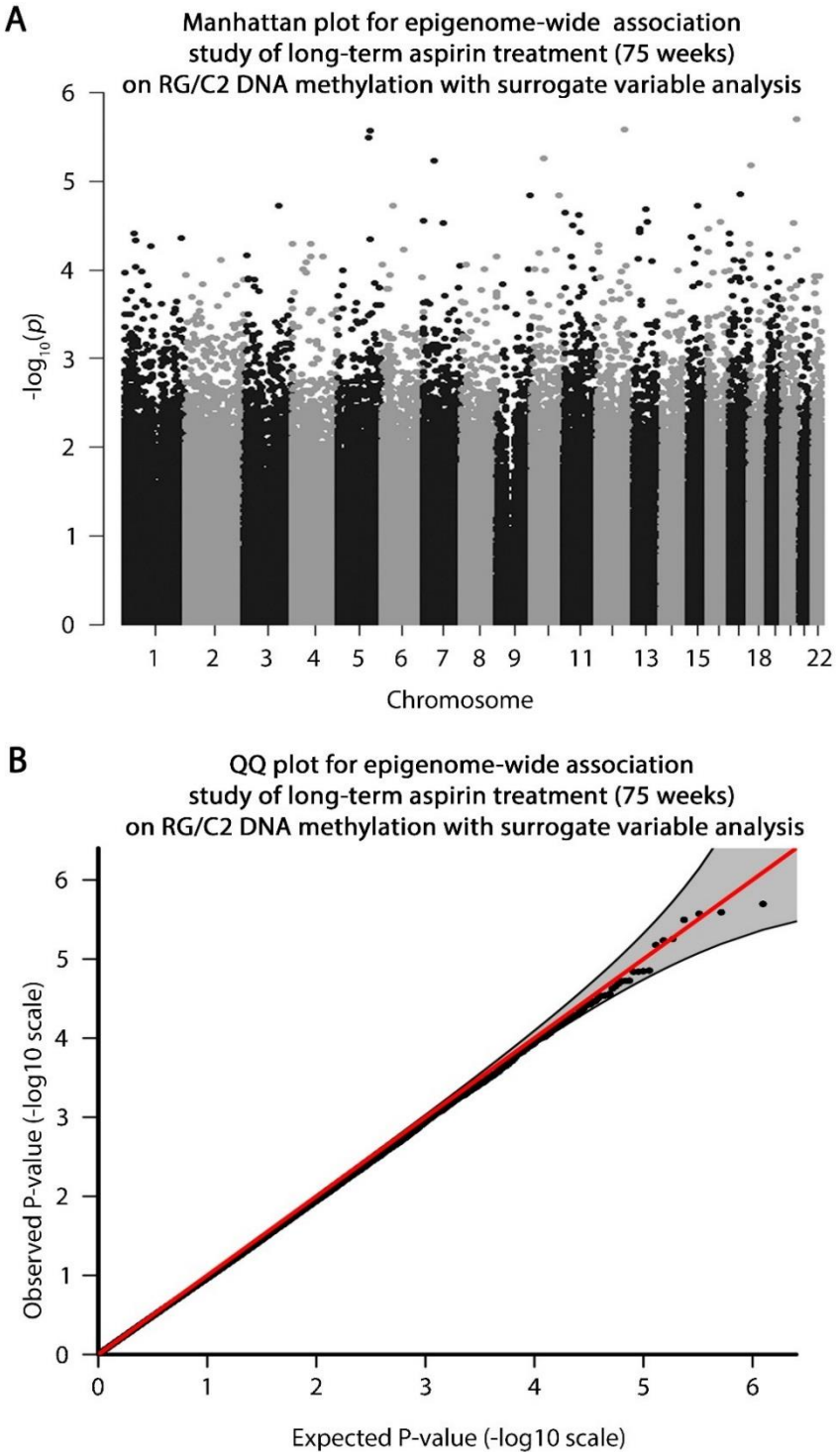


Figure 5-4- Manhattan and QQ plot for epigenome-wide association study of long-term aspirin treatment on RG/C2 cells DNA methylation with surrogate variable analysis.

RG/C2 cells were treated with 0, 0.5, 2 and 4mM aspirin for 75 weeks. DNA was then extracted, bisulphite converted, and methylation was measured using the MethylationEPIC BeadChip Infinium array. Multivariable regression was carried out adjusting for batch and 6 surrogate variables. (A) The Manhattan plot shows each CpG site, displayed as a dot, plotted according to its genomic co-ordinates along the X axis and plotted as the negative logarithm of the P-value for association of aspirin with CpG site methylation along the Y axis. The plot shows that no CpG sites fell below the P-value threshold for significance (5×10^{-8}) (red line) and the FDR adjusted P-value cut-off of 0.05. (B) The QQ plot shows the minus log10-transformed P-values of long-term aspirin treatment on CpG site methylation corrected for batch effects against quantiles from the null distribution. The red line indicates the line of equality whereby the P-values from our study should be plotted if they followed a null distribution. The 95% confidence intervals are shaded in grey and the black lines indicate the 0.05 P-value significance level. No evidence of inflation (inflation factor λ=0.96) was observed.

Table 5-4- Results of lowest 10 P-values of multivariable variable regression (adjusting only for batch) on the effect of aspirin on DNA methylation in RG/C2 cells

Illumina probe ID for CpG site	Chr	Gene Region	Position*	Multivariable linear regression comparing methylation of aspirin treated with untreated (adjusted for batch) (N=9 (treated) vs N=3 (untreated))				Multivariable linear regression comparing methylation of aspirin treated with untreated (adjusted for batch and 6 SVs) (N=9 (treated) vs N=3 (untreated))				Average fluorescent signal of methylation for each treatment individually (N= 3 for each concentration)			
				Effect size§ (β)	SE	P-value	FDR adjusted P-value	Effect size§ (β)	SE	P-value	FDR adjusted P-value	0mM	0.5mM	2mM	4mM
cg05514510	21	-	37987644	-0.33	0.0024	2.58E-16	2.23E-10	-0.14	0.22	0.565	0.9999	0.35	0.02	0.02	0.02
cg07139509	14	*C14orf162*	70038717	-0.20	0.0021	7.37E-15	1.99E-09	-0.43	0.11	0.028	0.9999	0.22	0.02	0.02	0.02
cg16438832	4	*SORCS2*	7341811	-0.35	0.0036	8.24E-15	1.99E-09	-0.56	0.23	0.098	0.9999	0.38	0.04	0.03	0.03
cg22064581	6	-	34113293	-0.30	0.0033	9.17E-15	1.99E-09	-0.28	0.20	0.263	0.9999	0.33	0.02	0.02	0.02
cg16799926	14	*SLC8A3*	70546695	-0.38	0.0043	1.46E-14	2.52E-09	-0.87	0.16	0.012	0.9999	0.43	0.05	0.04	0.04
cg26432519	18	*LOXHD1*	44087356	-0.84	0.0113	6.98E-14	1.01E-08	-1.51	0.39	0.030	0.9999	0.89	0.03	0.06	0.04
cg07427438	16	-	4819304	-0.19	0.0026	8.59E-14	1.06E-08	0.10	0.08	0.300	0.9999	0.2	0.01	0.02	0.02
cg07736530	11	-	33479526	-0.41	0.0059	1.27E-13	1.10E-08	-0.71	0.34	0.129	0.9999	0.46	0.04	0.05	0.04
cg08670313	11	*TENM4*	79082422	-0.25	0.0036	1.28E-13	1.10E-08	-0.40	0.30	0.274	0.9999	0.3	0.04	0.04	0.05
cg17504584	1	*IQCC*	32671645	-0.21	0.0030	1.34E-13	1.10E-08	-0.35	0.17	0.141	0.9999	0.26	0.06	0.05	0.05

*Chromosomal position based on NCBI human reference genome assembly Build 37.3 § Effect size = ratio of methylated fluorescence signal over total fluorescence signal of methylated and unmethylated (β). It's values range between 0 and 1 and roughly represents changes in percentage methylation where 0 is equal to 0% and 1 is equal to 100%. A positive value denotes an increase in methylation and a negative value denotes a decrease in methylation. (-) CpG site not mapped to a gene region.

Table 5-5- Results of lowest 10 P-values of multivariable variable regression (adjusting for batch and 6 surrogate variables) on the effect of aspirin on DNA methylation in RG/C2 cells

Illumina probe ID for CpG site	Chromosome	Gene Region	Position*	Multivariable linear regression comparing methylation of aspirin treated with untreated including surrogate variables ((N=9 (treated) vs N=3 (untreated))				Average fluorescent signal of methylation for each treatment individually (N= 3 for each concentration)			
				Effect size§ (β)	SE	P-value	FDR adjusted P-value	0mM	0.5mM	2mM	4mM
cg21232203	20	*CTSZ*	57574714	-1.11	0.011	2.01E-06	0.6907	0.87	0.8	0.82	0.82
cg16581782	12	*RPH3A*	1.13E+08	-4.64	0.049	2.56E-06	0.6907	0.42	0.4	0.33	0.24
cg15897840	5	-	1.35E+08	-5.20	0.055	2.67E-06	0.6907	0.56	0.54	0.48	0.79
cg06928924	5	*ISOC1*	1.28E+08	0.37	0.004	3.19E-06	0.6907	0.03	0.03	0.03	0.03
cg07001545	10	*ERCC6*	50747328	-0.08	0.001	5.57E-06	0.8228	0.01	0.01	0.01	0.01
cg00725654	7	*HECW1*	43314146	-2.57	0.036	5.81E-06	0.8228	0.92	0.93	0.89	0.83
cg25899954	18	-	8660164	0.99	0.014	6.65E-06	0.8228	0.94	0.94	0.92	0.91
cg00916973	17	-	43580461	1.30	0.024	1.40E-05	0.9999	0.77	0.74	0.75	0.78
cg17743249	X	*PHKA2*	18923628	1.19	0.022	1.43E-05	0.9999	0.87	0.9	0.89	0.9
cg03124708	9	*RALGDS*	1.36E+08	0.90	0.017	1.45E-05	0.9999	0.03	0.03	0.05	0.04

*Chromosomal position based on NCBI human reference genome assembly Build 37.3 § Effect size = ratio of methylated fluorescence signal over total fluorescence signal of methylated and unmethylated (β). It's values range between 0 and 1 and roughly represents changes in percentage methylation where 0 is equal to 0% and 1 is equal to 100%. A positive value denotes an increase in methylation and a negative value denotes a decrease in methylation. (-) CpG site not mapped to a gene region.

.

5.3.3 Aspirin favours decreasing the methylation of CpG sites

Initially, we analysed the effect of aspirin on over 850,000 CpG sites from the 850K array through Manhattan plots and QQ plots to identify the statistically significant CpG sites. We investigated the general effect of aspirin on DNA methylation on the sites that were below the P-value threshold for significance (5×10^{-8}) to eliminate any results which would otherwise be falsely associated with aspirin treatment (previously explained in Chapter 5.1.3.4). This means that only CpG sites with P-value s below 5×10^{-8} are considered to be associated with aspirin

After carrying out multivariable regression to compare the effect of aspirin treatment (0.5mM, 2mM and 4mM) versus no treatment on DNA methylation adjusting only for batch, a P-value threshold for significance (5×10^{-8}) was set. This yielded 1492 CpG sites that fell below this cut-off (Figure 5-3 A).

Of these 1492 CpG sites, the highest increase in DNA methylation was a beta of 0.7953 which equates to 79.53% increase in CpG site methylation in treated vs untreated cells. The highest decrease in methylation was a beta of -0.8439 which is equal to a decrease of 84.39% in CpG site methylation (Figure 5-5 A). Furthermore, long-term aspirin treatment decreases DNA methylation of 999 CpG sites and increases methylation of 493 sites (Figure 5-5 B). Therefore, long-term aspirin treatment favours a decrease in CpG site methylation as 66.96% of the sites were demethylated.

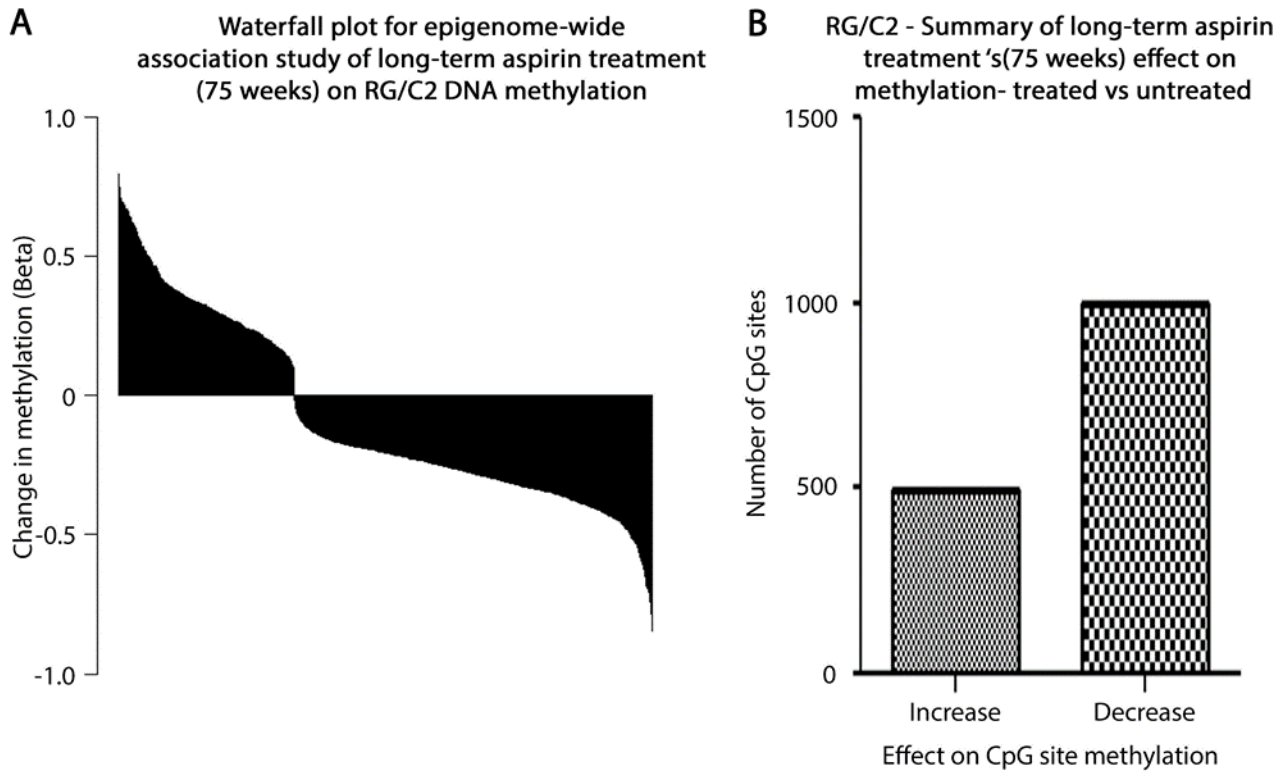


Figure 5-5- Waterfall plot and summary bar chart of the effect of long-term aspirin treatment on CpG site methylation changes in RG/C2 cells.

RG/C2 cells were treated with 0, 0.5, 2 and 4mM aspirin for 75 weeks. DNA was then extracted, bisulphite converted, and methylation was measured using the MethylationEPIC BeadChip Infinium array. Multivariable regression was carried out adjusting only for batch. (A) The waterfall plot shows that the highest beta of increase in methylation as a result of aspirin treatment is 0.7953 which equates to a 79.53% increase in CpG site methylation. The largest decrease in methylation is -0.8439 which equates to a 84.39% reduction in CpG site methylation. (B) Of the 1492 CpG sites that fell below the P-value threshold for significance (5×10^{-8}), aspirin increased the methylation of 493 CpG sites (33.04%) and decreased the methylation of 999 CpG sites (66.96%).

5.3.4 Methylation site enrichment analysis- genomic regions

In order to further understand the effect of long-term aspirin treatment on DNA methylation, we investigated whether aspirin was targeting specific genomic regions and/or specific locations relative to the CpG island. The rationale being that this may provide insights in to the mode of action of functional pathways via which aspirin is exerting downstream effects on "resistance" or CRC risk.

After applying a stringent P-value threshold for significance (5×10^{-8}), 1492 CpG sites remained as being statistically significant. We further separated these into sites that increase in DNA methylation with aspirin treatment and those that decrease in methylation. We then categorised these CpG sites based on their genomic regions thereby calculating the percentage of them found in the proximal promoter (200 base pairs upstream of the transcription start site, 1500 base pairs upstream of the transcription start site, exon 1 and the 5' untranslated region (UTR)), the 3' UTR, the gene bodies

(exons) and the intergenic regions (introns) (Figure 5-6 A). After that, we compared these to percentage values of CpG sites found in these regions for the entire array (N=865859) to be able to identify whether aspirin treatment enriches for changes in DNA methylation at specific genomic regions. We used a Chi square (X^2) test with Yates' continuity correction to compare the percentage of CpG sites in these genomic regions for sites below the P-value threshold for significance (5×10^{-8}) that decrease/increase methylation (observed data) compared to percentages from the entire array (expected data).

Table 5-6- A Comparison of the percentage of sites in genomic regions targeted by aspirin compared to the total percentage of sites from the array

	All sites (N= 865859)	P-value ≤ 5×10^{-8} significant sites of increased methylation (N=493)			P-value ≤ 5×10^{-8} significant sites of decreased methylation (N=999)		
Gene region	Percentage	Percentage	X-squared	P-value	Percentage	X-squared	P-value
Proximal promoter	35.20%	**27.38%**	6.481	0.0109	31.13%	3.461	0.0628
Exon 1	1.20%	1.62%	0.414	0.5201	**4.60%**	89.131	3.7×10^{-21}
TSS200	9.40%	**3.45%**	17.291	3.21×10^{-5}	7.81%	2.328	0.1271
TSS1500	13.90%	11.36%	1.873	0.1712	**7.81%**	24.264	8.4×10^{-7}
5'UTR	10.70%	10.95%	0.008	0.9276	10.91%	0.020	0.8867
3'UTR	1.90%	**5.07%**	23.203	1.46×10^{-6}	2.00%	0.013	0.9080
Gene body (exons)	35.90%	42.19%	3.654	0.0559	33.83%	0.829	0.3625
Intergenic region	26.90%	**34.69%**	7.972	0.0048	**35.84%**	21.490	3.6×10^{-6}

Abbreviations: TSS200, 200 base pairs upstream of the transcription start; TSS1500, 1500 base pairs upstream of the transcription start site; 5'UTR, 5' untranslated region; 3' UTR, 3' untranslated region. Statistically significant results are shown in red and bold font.

Aspirin increases the methylation of 493 sites and decreases the methylation of 999 sites from the 1492 CpG sites below the P-value threshold for significance (5×10^{-8}). The results of the X^2 test are summarised in Table 5-6. Of the sites that increase in methylation, an increase in methylation is significantly enriched in the 3' UTR (5.07% of the 493 sites are within the 3' UTR compared to 1.9% of the array, P-value = 1.46×10^{-6}) and the intergenic regions (34.69% of the 493 sites are found in intergenic regions compared to 26.9% of the array, P-value= 0.0048) whereas aspirin does not target promoter regions for methylation (27.38% of the 493 sites are found in promoter regions compared to 35.2% of the array, P-value = 0.0109) and the regions 200 base pairs upstream of the transcription start site (3.45% of the 493 sites are found in the regions 200 base pairs upstream of the transcription start site compared to 9.45% of the array, P-value = 3.21×10^{-5}) (Figure 5-6 B). There is no significant difference in the percentage of CpG sites targeted for increased methylation between

the 493 sites and the array for exon 1 regions, sites 1500 base pairs upstream of the transcription start site, the 5' UTR and the gene body.

With regards to sites of decreased methylation, aspirin demethylates intergenic regions more than expected (35.84% of the 999 sites are found in intergenic regions compared to 26.9% found in the array, P-value = $3.6x10^{-6}$) and demethylates exon 1 sites more than expected (4.60% of the 999 sites are found in exon 1 compared to 1.2% found in the array, P-value = $3.7x10^{-21}$). Aspirin does not target regions 1500 base pairs upstream of the transcription start site exon 1 sites for demethylation (7.81% of the 999 sites are demethylated compared to 13.9% of the array, P-value = $8.4x10^{-7}$) (Figure 5-6 C). There is no significant change in the percentage of CpG sites targeted for demethylation between the 999 sites and the array for the proximal promoter, regions 200 base pairs upstream of the transcription start site, the 5' UTR, the 3' UTR and the gene body.

These results indicate that aspirin appears to target the 3'UTR and intergenic regions for increased methylation more than chance but does not target the proximal promoter and regions 200 base pairs upstream of the transcription start site. Aspirin also demethylates CpG sites in exon 1 and intergenic regions more than chance but does not target sites 1500 base pairs upstream of the transcription start site for demethylation.

Overall, the results suggest that aspirin demethylates promoter regions exon 1 and 1500 base pairs upstream of the transcription start site and methylates 3'UTR regions. The demethylation of promoter sites would indicate an increase in transcription and gene expression (399) and 3' UTR sites are known for regulating the fate of mRNAs such as localization, translation and degradation (400).

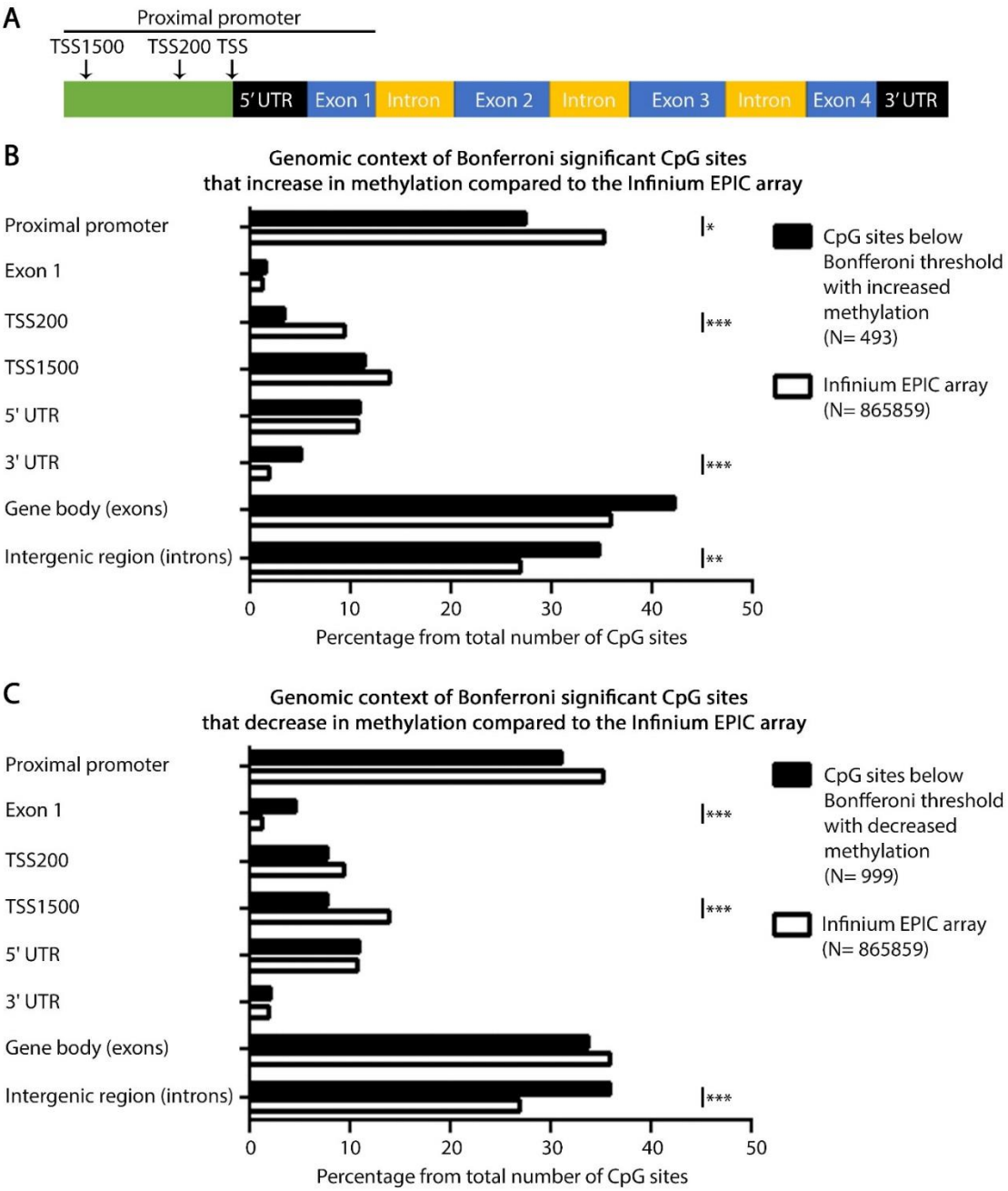


Figure 5-6- A comparison of CpG sites enriched for methylation and demethylation in different genomic regions after long-term aspirin treatment (75 weeks) in RG/C2 cells

The percentage of methylated and demethylated CpG sites found in specific genomic regions were compared to the percentage of CpG sites found in these regions from the 865859 sites of the MethylationEPIC BeadChip Infinium array. We used a Chi square (X^2) test with Yates' continuity correction to identify whether aspirin targets specific regions for methylation/demethylation more or less than we would expect (*** $p \leq 0.001$, **$p \leq 0.01$ and * $p \leq 0.05$). Abbreviations: TSS200, 200 base pairs upstream of the transcription start; TSS1500, 1500 base pairs upstream of the transcription start site; 5'UTR, 5' untranslated region; 3' UTR, 3' untranslated region. (A) A schematic of the genomic regions analysed and their locations within the genome. (B) Aspirin methylates CpG sites in the 3'UTR ($p \leq 0.001$) and intergenic regions ($p \leq 0.01$) more than expected but does not methylate sites in the proximal promoter ($p \leq 0.05$) or 200 base pairs upstream of the transcription start site ($p \leq 0.001$). (C) Aspirin demethylates CpG sites 1500 base pairs upstream of the transcription start site ($p \leq 0.001$) and within intergenic regions ($p \leq 0.001$) more than chance but does not demethylate sites in exon 1 ($p \leq 0.001$).

5.3.5 Methylation site enrichment analysis- CpG context

To further understand which sites/regions are targeted for methylation/demethylation by aspirin, we analysed the location of sites relative to CpG islands. CpG islands are clusters of CpG dinucleotides commonly found in promoter regions of genes (169,170).

Statistically significant sites were calculated as above (Chapter 5.3.4) and separated into sites that increase in DNA methylation with aspirin treatment and those that decrease in methylation. CpG sites were categorised based on their CpG context by calculating the percentage of sites found in the north shelf, north shore, CpG island, south shore, south shelf and open sea regions (CpG sites located outside of these regions) (Figure 5-7 A). The annotation of the CpG islands was as follows: shores are considered 2-kb sequences either side of the CpG island, shelves are 2-kb sequences next to the shore and open sea sites are CpG dinucleotides not found in the CpG islands or the shores and shelves next to it (171).

We compared these to percentage values of CpG sites found in these regions for the entire array (N=865859) to identify enrichment for changes in DNA methylation at certain locations relative to the CpG island. We used a Chi square (X^2) test with Yates' continuity correction to compare the percentage of CpG sites in these genomic regions.

Table 5-7- A Comparison of the percentage of sites in the CpG content and neighbourhood context by aspirin compared to the total percentage of sites from the array

Relation to Island	All sites (N= 865859)	P-value ≤ 5 x 10^{-8} significant sites of increased methylation (N=493)			P-value ≤ 5 x 10^{-8} significant sites of decreased methylation (N=999)		
	Percentage	Percentage	X-squared	P-value	Percentage	X-squared	P-value
North shelf	9.63%	10.14%	0.075401	0.7836	11.11%	1.9025	0.1687
north shore	3.69%	4.26%	0.27028	0.6031	2.90%	1.437	0.2306
Island	18.65%	**2.03%**	70.514	4.5 x 10^{17}	**14.61%**	7.365	0.00665
South shore	8.22%	9.13%	0.34884	0.5548	**11.61%**	12.1	5.041x10^{-4}
South shelf	3.43%	4.46%	1.1825	0.2769	**1.90%**	6.226	0.0126
Open Sea	56.38%	**69.98%**	9.2859	2.31 x10^{-3}	57.86%	0.21923	0.6396

The annotation of the CpG islands was as follows: shores are considered 2-kb sequences either side of the CpG island, shelves are 2-kb sequences next to the shore and open sea sites are CpG dinucleotides not found in the CpG islands or the shores and shelves next to it. Statistically significant results are shown in red and bold font.

Aspirin targets open sea regions for methylation more than expected (69.98% of the 493 sites are methylated compared to 56.38% of the array, P-value= 2.31 x10^{-3}) but aspirin does not target CpG islands for methylation (2.03% of the 493 sites are methylated compared to 18.65% of the array, P-value= 4.5 x 10^{-17}) (Table 5-7). There is no significant change in the percentage of CpG sites that are methylated between the 493 sites and the array for CpG sites in the north shelf, north shore, south shore and south shelf (Figure 5-7 B).

On the other hand, aspirin significantly demethylates south shore regions more than expected (11.61% of the 999 CpG sites compared to 8.22% of the array, P-value= 5.041x10^{-4}). Long-term aspirin treatment does not target islands for demethylation (14.61% of the 999 CpG sites compared to 18.65% of the array, P-value= 0.00665) nor does it target south shelf regions (1.90% of the 999 CpG sites compared to 3.43% of the array, P-value= 0.0126). There is no significant change in the percentage of CpG sites that are demethylated between the 999 sites and the array for CpG sites in the north shelf, north shore and open sea.

Overall, these results suggest that aspirin methylates open sea regions and demethylates south shore regions. The consequence of methylation/demethylation of these sites is still largely unknown.

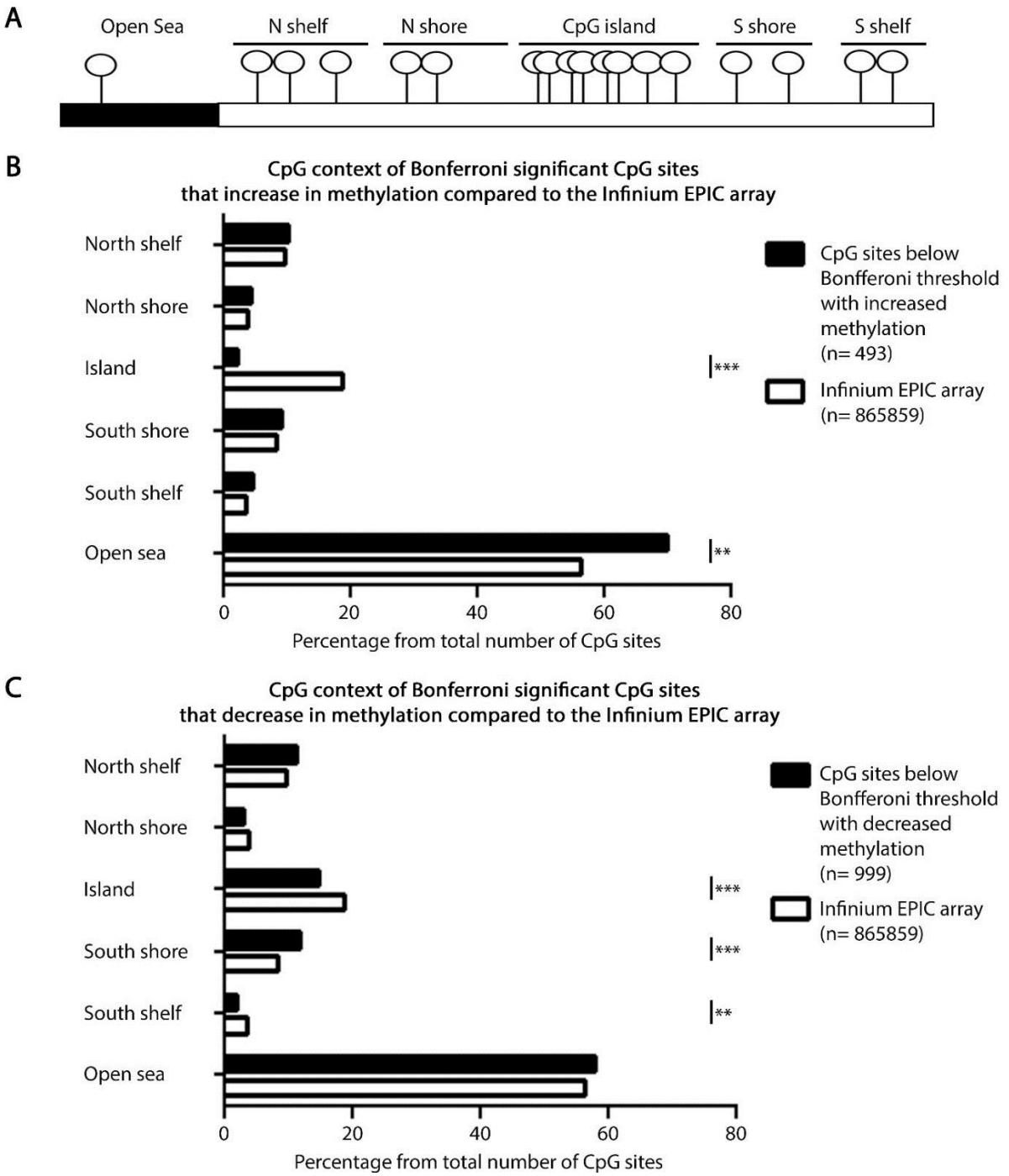


Figure 5-7- A comparison of CpG sites enriched for methylation and demethylation relative to the CpG island after long-term aspirin treatment (75 weeks) in RG/C2 cells

The percentage of methylated and demethylated CpG sites found in specific regions relative to CpG islands were compared to the percentage of CpG sites found in these regions from the 865859 sites of the MethylationEPIC BeadChip Infinium array. Shores are considered to be 2-kb sequences either side of the CpG island, shelves are 2-kb sequences next to the shore and open sea sites are CpG dinucleotides not found in the CpG islands or the shores and shelves next to it. We used a Chi square (X^2) test with Yates' continuity correction to identify whether aspirin targets specific regions for methylation/demethylation more or less than we would expect (*** $p \leq 0.001$, **$p \leq 0.01$ and * $p \leq 0.05$). A) A schematic of CpG sites relative to the CpG island. (B) Aspirin methylates CpG dinucleotides in open sea regions more than expected ($p \leq 0.01$) but does not target CpG islands ($p \leq 0.001$). (C) Aspirin demethylates CpG dinucleotides in south shore regions more than expected ($p \leq 0.001$) but does not target CpG islands ($p \leq 0.001$) and south shelf regions ($p \leq 0.01$).

5.3.6 Gene ontology pathway

After identifying how many sites aspirin targets for increased methylation and decreased methylation, we investigated whether the epigenetic effects of aspirin were targeting specific pathways.

CpG site annotations (chromosome, position, gene etc.) were extracted using The Meffil package in R (264). CpG sites were not necessarily only in the promoter regions but could also be located in other coding and non-coding regions within the genomic co-ordinates of the gene. Gene names associated with the 493 CpG sites of increased methylation and 999 sites of decreased methylation were extracted. Sites were tested for ontology enrichment using the Kyoto Encyclopedia of Genes and Genomes (KEGG) collection of pathways in R version 3.2.3 using the missMethyl package. Pathways were considered statistically significant when the P-value using the FDR method was 0.05 for over-representation of the KEGG term. This means that any pathways with an FDR-adjusted P-value of less than 0.05 have a probability of 5% or less of being a false positive when actually they are a null association.

For an increase in methylation, aspirin targets 7 CpG sites in genes involved in the pathways in cancer (P-value= 1.24×10^{-5}, FDR adjusted P-value = 0.004) and 5 sites in the pathways of Human T-lymphotropic virus infection (P-value = 1.29×10^{-4}, FDR adjusted P-value = 0.0207) and human papillomavirus infection (P-value= 4.10×10^{-4}, FDR= 0.0437) (Figure 5-8 A, Table 5-8).

For a decrease in methylation, aspirin targets 29 pathways ranging from a total of 3 genes to 17 genes. The 5 most over-represented pathways are the Mitogen-Activated Protein Kinase (MAPK) signalling pathway (P-value =3.66×10^{-8}, FDR adjusted P-value= 1.17×10^{-5}), protein digestion and absorption pathway (P-value= 8.24×10^{-8}, FDR adjusted P-value= 1.32×10^{-5}), cyclic guanosine monophosphate-protein kinase G (cGMP-PKG) signalling pathway (P-value = 7.77×10^{-7}, FDR adjusted P-value = 8.29×10^{-5}), metabolic pathways (P-value= 3.45×10^{-6}, FDR adjusted P-value= 2.76×10^{-4}) and cytokine-cytokine receptor interaction pathway (P-value = 1.06×10^{-5}, FDR adjusted P-value= 6.75×10^{-4}) (Figure 5-8 B, Table 5-9). Aspirin also demethylates genes in the Human T-lymphotropic virus infection pathway (P-value= 1.10×10^{-3}, FDR adjusted P-value= 0.018).

Aspirin demethylates CpG sites in genes involved in the MAPK signalling pathway, cytokine-cytokine receptor interactions, insulin secretion and PI3K-AKT signalling pathway. The MAPK signalling pathways is over activated and expressed in colorectal cancer progression (401). Cytokines such as interleukin-6 (IL-6) and tumour necrosis factor (TNF) as well as interleukin 17A (IL-17A) and interleukin 22 (IL-22) have been shown to be associated with the hallmarks of cancer such as resistance to apoptosis, proliferation, angiogenesis as well as invasion and metastasis (402) . Also,

several proteins involved in the insulin/insulin-like growth factor (IGF) signalling pathway are involved in the transformation of normal colon epithelial cells (403). As for the phosphoinositide 3-kinase- protein kinase B (PI3K-AKT) signalling pathway, it is involved in reducing apoptosis as well as stimulating cell proliferation (404). These results suggest that long-term aspirin treatment targets the demethylation of pathways involved in cancer progression. In general, demethylation would result in an increase in gene expression and therefore increase the signalling of these pathways. This demethylation and thus activation of signalling pathways involved in cancer progression may be one of the mechanisms by which patients acquire "aspirin resistance" since changes in DNA methylation are not immediate (as seen from short-term aspirin treatment which did not significantly affect DNA methylation (Chapter 4.3.1)) and this suggests that aspirin's effect on DNA methylation takes place over a longer period of time.

Therefore, aspirin targets genes for increased methylation in the cancer pathways and viral infections. Aspirin demethylates genes involved in many signalling and metabolic pathways, of which some are involved in colorectal cancer progression. Overall, these results suggest that aspirin targets a plethora of different genes involved in many pathways through epigenetic mechanisms and aspirin's effect on DNA methylation may be one of the mechanisms through which patients could acquire "resistance" aspirin if taken long-term prior to exposure (276).

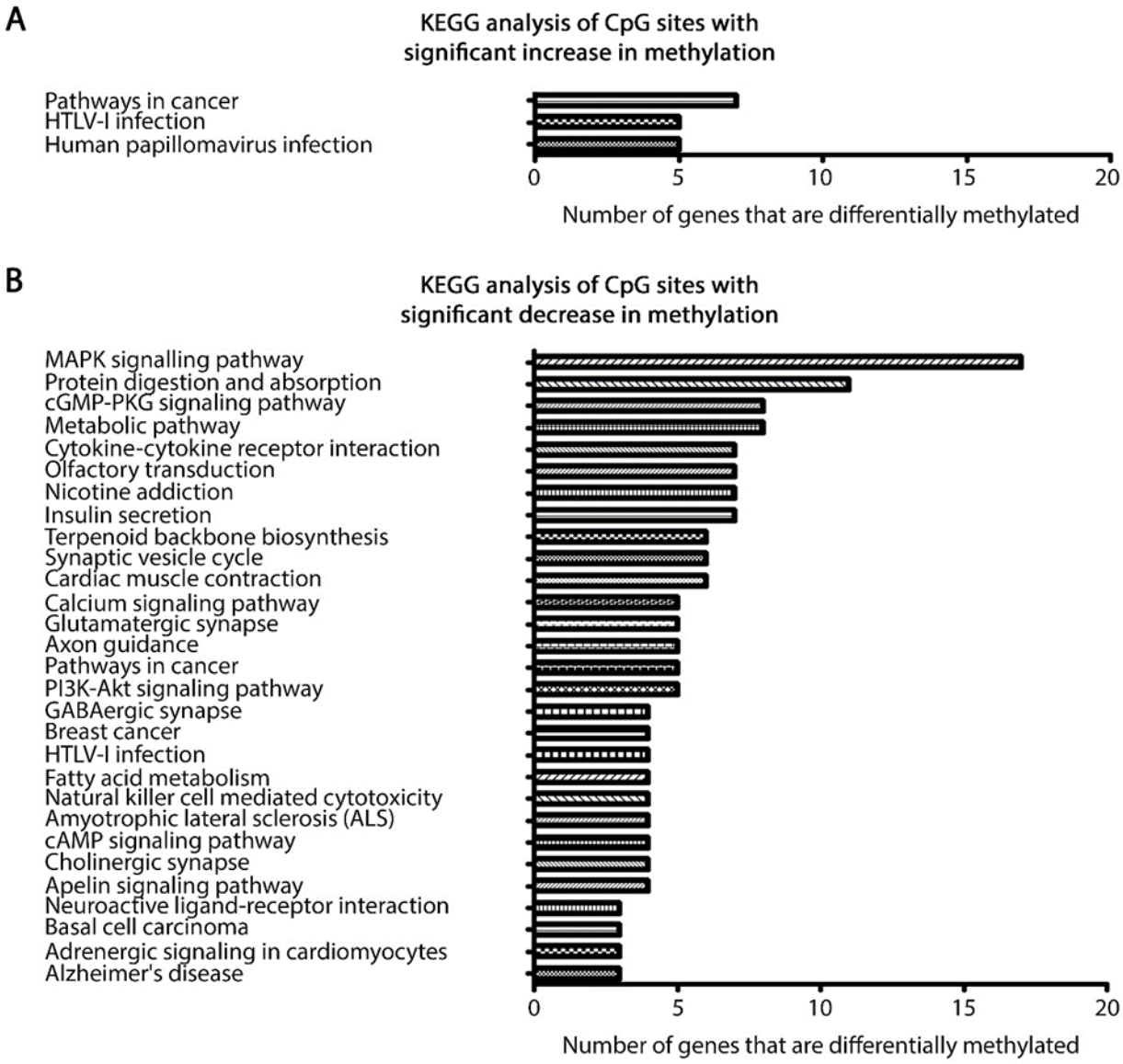


Figure 5-8- Pathways targeted by aspirin through a methylation increase or decrease of their genes.

After applying a more stringent P-value threshold for significance (5 x 10^{-8}), aspirin increases methylation of 493 CpG sites and demethylates 999 sites. Gene names associated with these CpG sites were extracted and were tested for ontology enrichmnent for the KEGG collection of pathways using the missMethyl package in R version 3.2.3. Pathways were considered to be statistically significant when the P-value using the FDR method was 0.05 for over-representation of the KEGG term. (A) Aspirin methylates genes in the pathways of cancer, HTLV-1 infection and human papillomavirus infection pathways. (B) Aspirin demethylates genes in 29 pathways ranging from 3 genes to 17.

Table 5-8- Over-represented pathways of genes that increase in methylation with long-term aspirin treatment (75 weeks) in RG/C2 cells

Pathway	Number of genes in the KEGG term	Number of genes that are differentially methylated	P-value for over-representation of the KEGG term	FDR adjusted P-value
Pathways in cancer	394	7	1.24×10^{-5}	0.0040
Human T-lymphotropic virus infection	252	5	1.29×10^{-4}	0.0207
Human papillomavirus infection	309	5	4.10×10^{-4}	0.0437

Abbreviations: FDR, false discovery rate.

Table 5-9- Over-represented pathways of genes that decrease in methylation with long-term aspirin treatment (75 weeks) in RG/C2 cells

Pathway	Number of genes in the KEGG term	Number of genes that are differentially methylated	P-value for over-representation of the KEGG term	FDR adjusted P-value
MAPK signalling pathway	254	11	3.66×10^{-8}	1.17×10^{-5}
Protein digestion and absorption	89	7	8.24×10^{-8}	1.32×10^{-5}
cGMP-PKG signalling pathway	162	8	7.77×10^{-7}	8.29×10^{-5}
Metabolic pathways	1240	17	3.45×10^{-6}	2.76×10^{-4}
Cytokine-cytokine receptor interaction	254	7	1.06×10^{-5}	6.75×10^{-4}
Olfactory transduction	395	7	1.57×10^{-5}	8.35×10^{-4}
Nicotine addiction	40	4	3.68×10^{-5}	1.68×10^{-3}
Insulin secretion	85	5	5.41×10^{-5}	2.16×10^{-3}
Terpenoid backbone biosynthesis	22	3	6.77×10^{-5}	2.41×10^{-3}
Synaptic vesicle cycle	63	4	1.66×10^{-4}	5.27×10^{-3}
Cardiac muscle contraction	74	4	1.81×10^{-4}	5.27×10^{-3}
Calcium signalling pathway	180	6	1.98×10^{-4}	5.29×10^{-3}
Glutamatergic synapse	114	5	2.62×10^{-4}	6.46×10^{-3}
Axon guidance	174	6	3.51×10^{-4}	8.02×10^{-3}
Pathways in cancer	394	8	5.76×10^{-4}	0.0123
PI3K-AKT signalling pathway	329	7	6.85×10^{-4}	0.0137
GABAergic synapse	88	4	8.31×10^{-4}	0.0157
Breast cancer	151	5	9.76×10^{-4}	0.0173
Human T-lymphotropic virus infection	252	6	1.10×10^{-3}	0.0180
Fatty acid metabolism	47	3	1.13×10^{-3}	0.0180
Natural killer cell mediated cytotoxicity	116	4	1.24×10^{-3}	0.0190
Amyotrophic lateral sclerosis (ALS)	51	3	1.67×10^{-3}	0.0243
cAMP signalling pathway	198	5	1.96×10^{-3}	0.0273
Cholinergic synapse	112	4	2.42×10^{-3}	0.0323
Apelin signalling pathway	136	4	3.15×10^{-3}	0.0403
Neuroactive ligand-receptor interaction	273	5	3.50×10^{-3}	0.0430
Basal cell carcinoma	63	3	3.81×10^{-3}	0.0449
Adrenergic signalling in cardiomyocytes	143	4	3.96×10^{-3}	0.0449
Alzheimer's disease	164	4	4.07×10^{-3}	0.0449

Abbreviations: MAPK, Mitogen-Activated Protein Kinase; cGMP-PKG, cyclic guanosine monophosphate-protein kinase G; PI3K- AKT, Phosphoinositide 3-kinase- protein kinase B; FDR, false discovery rate.

5.3.7 Aspirin has the same effect on CpG site methylation regardless of dose

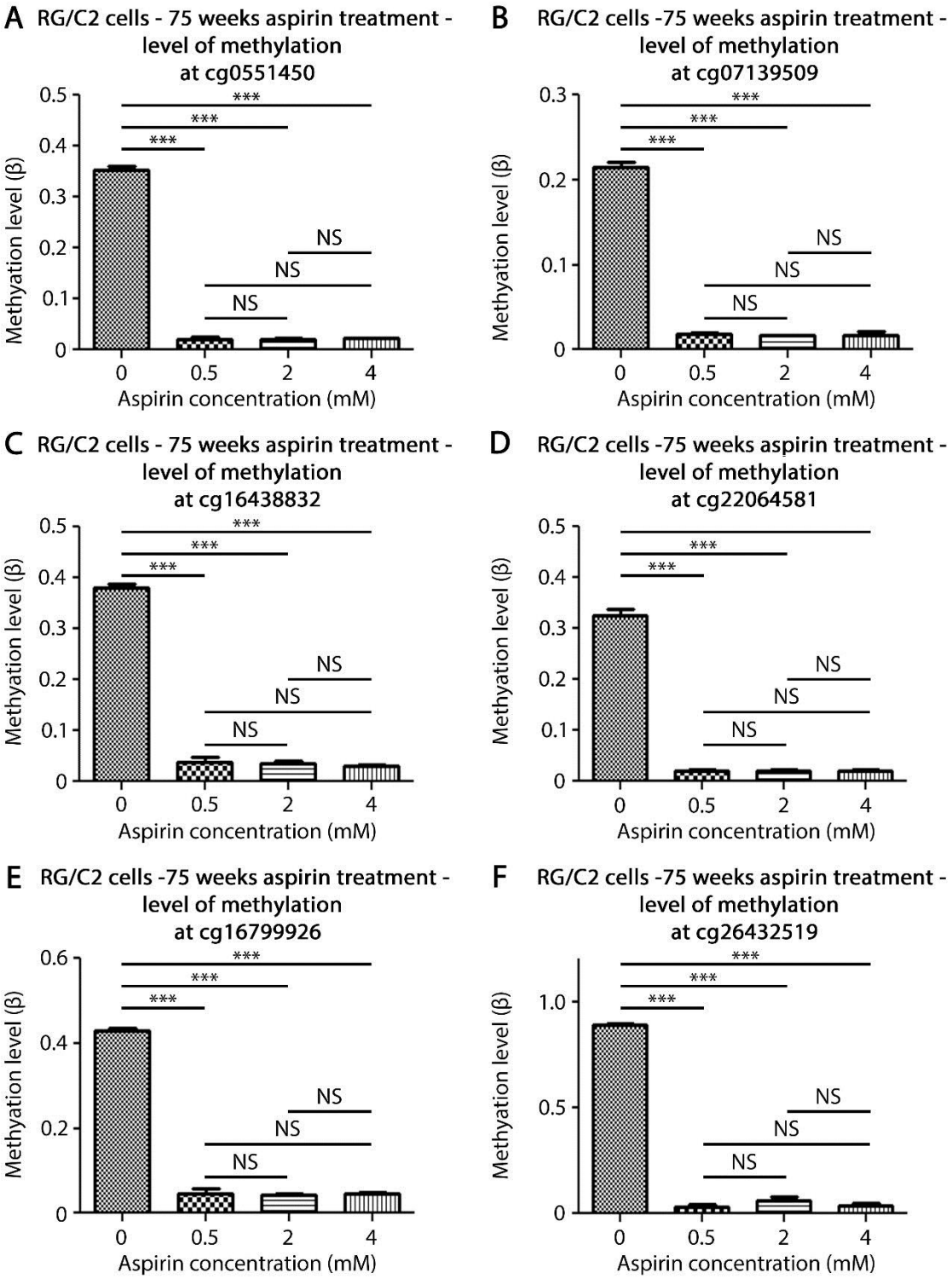
The results from Chapter 5.3.2 suggest that aspirin may be having a similar effect on DNA methylation regardless of dose. For this reason, we decided to investigate the 10 most statistically significant CpG sites from the multivariable regression analysis of aspirin as a binary exposure variable and adjusting for batch only. Raw methylation levels for RG/C2 cells with 0mM, 0.5mM, 2mM and 4mM aspirin treatment was extracted. A one-way ANOVA and Bonferroni multiple comparison test was used to calculate significance between each of the treatments.

The levels of DNA methylation, including standard deviations, for each aspirin condition in the 10 CpG sites is summarised in Table 5-10, where 0 denotes unmethylated and 1 denotes complete methylation. The results show that there is a significant decrease in methylation between 0mM aspirin treatment and all aspirin doses ($p \leq 0.001$) in the 10 CpG sites (Figure 5-9). The results also show that there is no significant difference between levels of methylation between 0.5mM vs 2mM, 0.5mM vs 4mM and 2mM vs 4mM. This is consistent for all CpG sites. It is worth to consider that the nature of the model run (treated vs untreated) was best powered to select for these results. Therefore a linear regression should be run comparing aspirin concentrations as a continuous exposure variable rather than binary.

These results are important as they indicate that long-term aspirin treatment may exhibit similar effects on DNA methylation regardless of the dose of aspirin. Doses as low as 0.5mM affect DNA methylation of a CpG site to a similar magnitude as the higher doses. For example, 0.5mM aspirin decreases the DNA methylation of cg05514510 by 0.33 and this is the same as 2mM and 4mM. This could indicate that there is a threshold for aspirin's effect on DNA methylation that is met by doses as small as 0.5mM.

Table 5-10- Average level of methylation for each aspirin dose for the 10 most statistically significant CpG sites in the multivariable regression

Aspirin concentration (mM)	Average fluorescent signal of methylation for each treatment individually							
	0mM		0.5mM		2mM		4mM	
Illumina probe ID for CpG site	Average β	SD	Average β	SD	Average β	SD	Average β	SD
cg05514510	0.35	0.007	0.02	0.003	0.02	0.002	0.02	0.001
cg07139509	0.22	0.006	0.02	0.001	0.02	0.000	0.02	0.003
cg16438832	0.38	0.007	0.04	0.008	0.03	0.007	0.03	0.001
cg22064581	0.33	0.012	0.02	0.002	0.02	0.004	0.02	0.003
cg16799926	0.43	0.008	0.05	0.010	0.04	0.001	0.04	0.006
cg26432519	0.89	0.005	0.03	0.008	0.06	0.017	0.04	0.011
cg07427438	0.20	0.005	0.01	0.001	0.02	0.005	0.02	0.005
cg07736530	0.46	0.005	0.04	0.005	0.05	0.017	0.04	0.005
cg08670313	0.30	0.001	0.04	0.004	0.04	0.005	0.05	0.009
cg17504584	0.26	0.005	0.06	0.003	0.05	0.002	0.05	0.006


A RG/C2 cells - 75 weeks aspirin treatment - level of methylation at cg0551450
B RG/C2 cells -75 weeks aspirin treatment - level of methylation at cg07139509
C RG/C2 cells - 75 weeks aspirin treatment - level of methylation at cg16438832
D RG/C2 cells -75 weeks aspirin treatment - level of methylation at cg22064581
E RG/C2 cells -75 weeks aspirin treatment - level of methylation at cg16799926
F RG/C2 cells -75 weeks aspirin treatment - level of methylation at cg26432519
Methyation level (β)
Aspirin concentration (mM)
0
0.5
2
4
***
NS

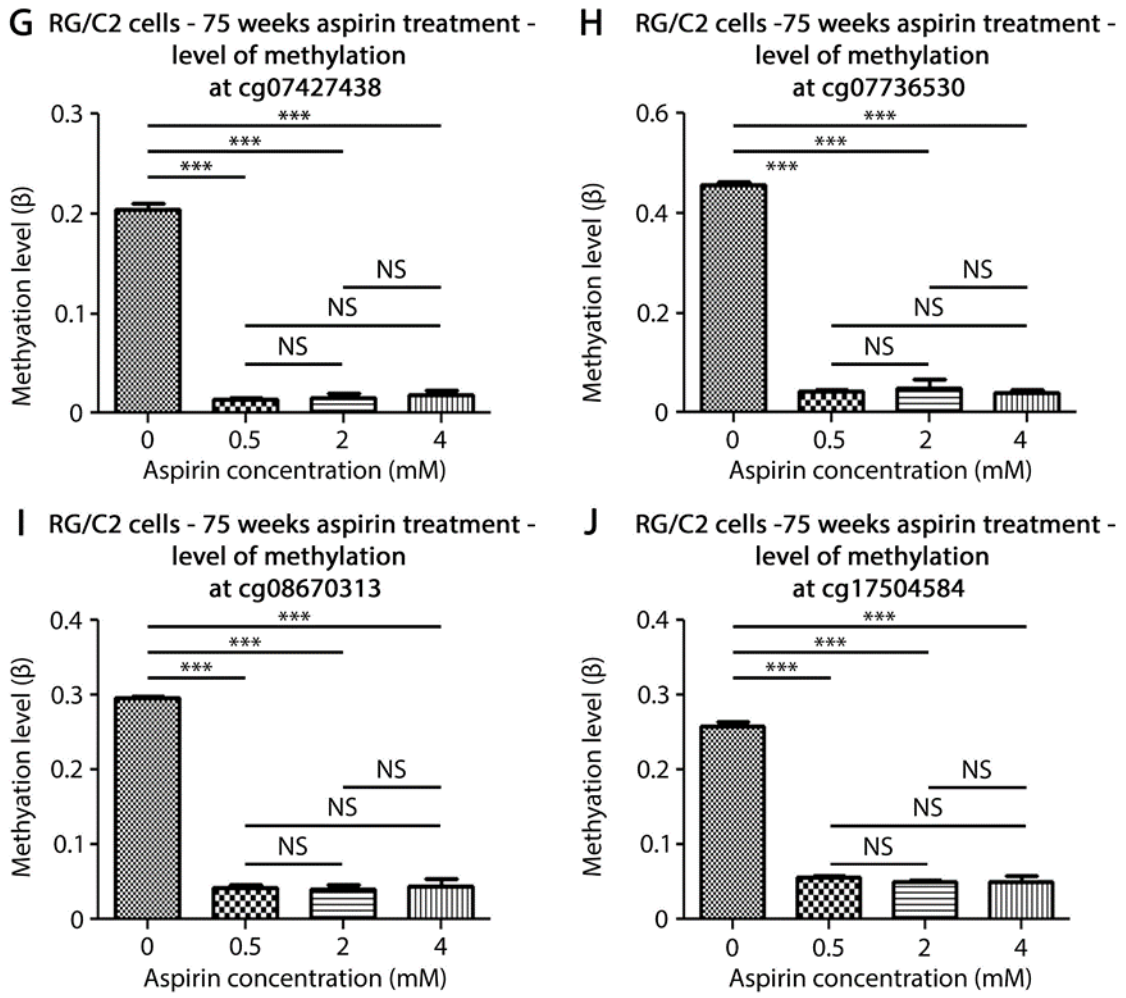


Figure 5-9- The effect of long-term aspirin treatment on the 10 most statistically significant CpG sites stratified by aspirin dose.

RG/C2 cells were treated with 0, 0.5, 2 and 4mM aspirin for 75 weeks. DNA was then extracted, bisulphite converted, and methylation was measured using the MethylationEPIC BeadChip Infinium array. Multivariable regression was carried out adjusting only for batch. A one-way ANOVA and Bonferroni multiple comparison test was used to calculate significance between each of the treatments (*** $p \leq 0.001$, **$p \leq 0.01$ and * $p \leq 0.05$, NS= non-significant). Bar charts of the average level of DNA methylation for the 10 most statistically significant Illumina probe IDs for CpG site methylation are shown; (A) cg05514510, (B) cg07139509, (C) cg16438832, (D) cg22064581, (E) cg16799926, (F) cg26432519, (G) cg07427438, (H) cg07736530, (I) cg08670313 and (J) cg17504584. Aspirin (0.5mM, 2mM and 4mM) significantly decreases the methylation for all 10 CpG sites ($p \leq 0.001$) but there is no significant difference between 0.5mM and 2mM, 0.5mM and 4mM or 2mM and 4mM aspirin treatment on CpG site methylation levels.

5.3.8 Pyrosequencing

To confirm that aspirin was directly affecting CpG methylation, validation of these results through pyrosequencing was attempted (171,405).

At this point, RG/C2 cells had been cultured for 108 weeks with 0mM, 0.5mM, 2mM and 4mM aspirin in 2D cell culture. DNA was extracted and then quantified before being subject to bisulphite modification. PCR primers were designed using the PyroMark Assay Design software around the CpG site of interest (previously explained in Chapter 2.11). The software assigns scores to the primers out of 100 calculated according to many factors such as the likelihood of mispriming, the primer length (optimal is 200 base pairs or less) and the chance of primer dimer. We used a threshold score of 80 or more as a cut-off for a suitable primer for the assay. We chose the first two Illumina probe numbers for CpG sites with the most significant P-value of association with aspirin treatment that we could design a primer for with a score of 80 or more. These are summarised in Table 5-11. Once suitable primers were designed, bisulphite modified samples were subject to PCR and pyrosequencing. We then applied a one-way ANOVA and Dunnett's multiple comparison test to calculate statistical significance.

Table 5-11- Primer assay scores for pyrosequencing assays of the 6 most statistically significant CpG sites associated with aspirin treatment

Illumina probe ID for CpG site	Assay score	Gene name	P-value
cg05514510	75	*C14orf162 NR_024630*	2.58×10^{-16}
cg07139509	Assay not possible	None annotated	7.37×10^{-15}
cg16438832	79	None annotated	8.24×10^{-15}
cg22064581	Assay not possible	None annotated	9.17×10^{-15}
cg16799926	82	*SLC8A3*	1.46×10^{-14}
cg26432519	87	*LOXHD1*	6.98×10^{-14}

Primer scores were assessed starting with the most statistically significant CpG site from aspirin treatment until primers with an assay score of 80 or more (which indicates a higher likelihood of a successful PCR assay) were found. The CpG sites that had the highest scoring primers were the sites measured by the Illumina probe IDs cg16799926 and cg26432519 located in the genes solute carrier

family 8 member A3 (*SLC8A3)* and lipoxygenase homology domains 1 (*LOXHD1)*, respectively (Table 5-11), which had a primer assay score of 82 and 87. These CpG sites are at chromosome 14 position 70546695 and chromosome 18 position 44087356, respectively. For this reason, we tested for changes in DNA methylation in these 2 CpG sites due to aspirin treatment to validate our EWAS results.

In order to carry out pyrosequencing, we first checked that the PCR amplified samples contained DNA and that the primer was amplifying the desired product ((Figure 5-10 A). The results show that DNA is present in all samples but there appears to be DNA contamination in the bisulphite modified control sample and the PCR control sample for both CpG sites at the Illumina probe IDs cg16799926 (Figure 5-10 A) and cg26432519 (Figure 5-11 A). The gel also shows that the primers are successfully amplifying the desired PCR product only.

Table 5-12- Average percentage methylation of CpG at the Illumina probe ID cg16799926 using the MethylationEPIC BeadChip Infinium (850K) array and pyrosequencing after 108 weeks of aspirin treatment in RG/C2 cells

SLC8A3	MethylationEPIC BeadChip Infinium (850K) array		Pyrosequencing	
Aspirin concentration	Average percentage methylation	SD	Average percentage methylation	SD
0mM	42.80	0.0079	21.75	3.61
0.5mM	4.66	0.0100	7.91	6.89
2mM	4.34	0.0006	2.44	4.23
4mM	4.42	0.0057	3.53	3.11

Due to the contamination of the PCR control and the bisulphite modified control, it is difficult to derive any conclusions from the pyrosequencing results, although both methods do allude to a decrease in CpG site methylation with aspirin (Table 5-12 and Figure 5-10 B). Table 5-12 summarises the percentage methylation with each aspirin treatment for CpG site at the Illumina probe ID cg16799926 comparing the results from the EWAS and the results from pyrosequencing. However, this assay needs to be repeated before any further conclusions can be made.

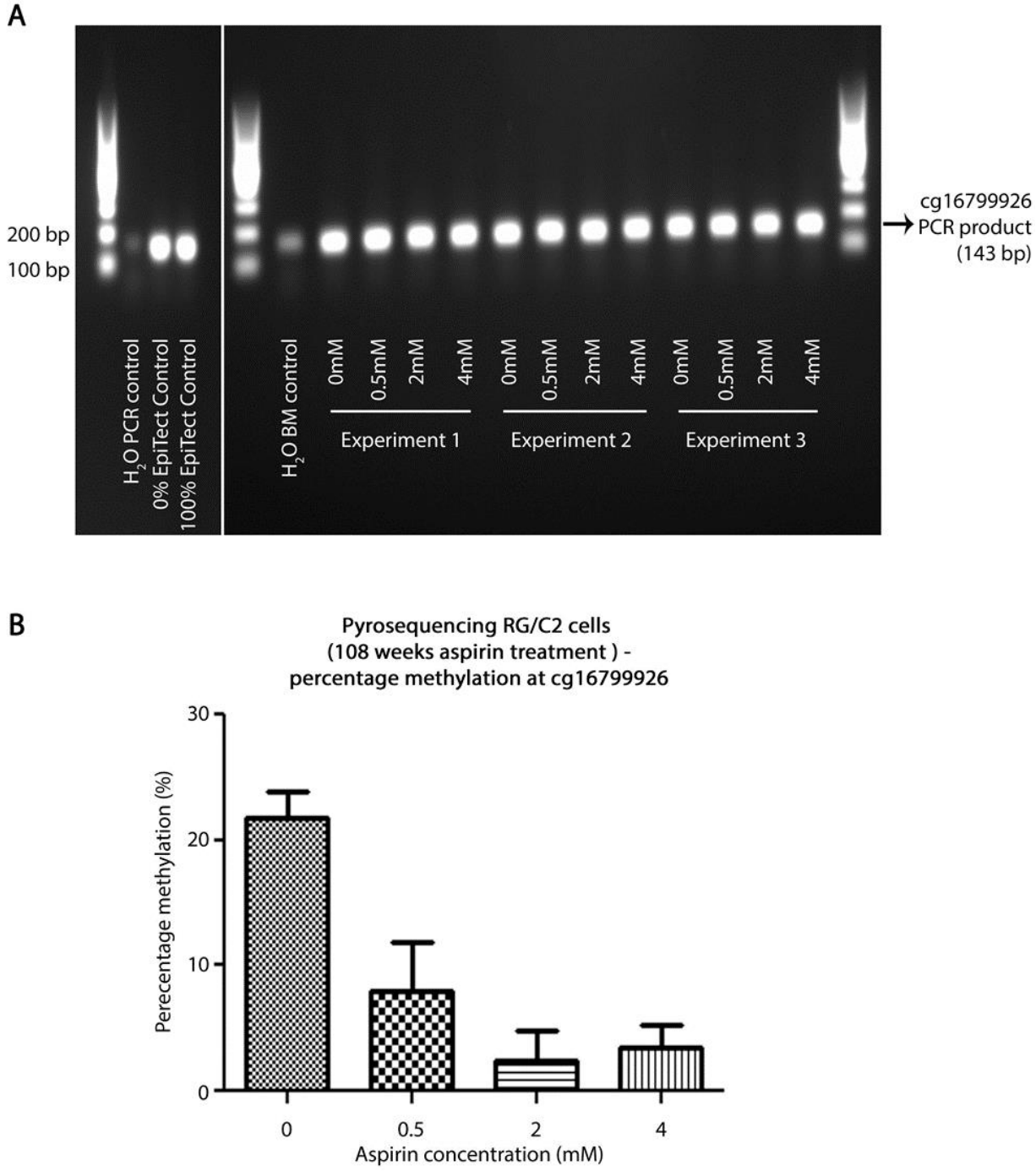


Figure 5-10- Long term aspirin treatment (108 weeks) effect on the CpG site from the Illumina probe cg16799926 methylation

RG/C2 cells were cultured for 108 weeks with 0, 0.5, 2 and 4mM aspirin in 2D cell culture. DNA was extracted and then quantified before being subject to bisulphite modification, PCR and pyrosequencing. Results are mean values from 3 repeat experiments with standard deviations and a one-way ANOVA and Dunnett's multiple comparison test to calculate significance (*** $p \leq 0.001$, **$p \leq 0.01$ and * $p \leq 0.05$). (A) The agarose gel shows that the primers successfully produce the desired PCR product and that there is DNA present in all samples. The gel also shows that there appears to be DNA contamination in the bisulphite modified control (H_2O BM control) and PCR control (H_2O PCR control). (B) Aspirin may be decreasing the percentage methylation of the CpG site from the Illumina probe cg16799926 at 0.5mM, 2mM and 4mM. However, the quality of the samples is not high therefore these results need to be repeated and checked.

Table 5-13- Average percentage methylation of the CpG site at the Illumina probe ID cg26432519 using the MethylationEPIC BeadChip Infinium (850K) array and pyrosequencing after 108 weeks of aspirin treatment in RG/C2 cells

LOXHD1	MethylationEPIC BeadChip Infinium (850K) array		Pyrosequencing	
Aspirin concentration	Average percentage methylation	SD	Average percentage methylation	SD
0mM	88.79	0.0053	41.40	2.67
0.5mM	3.30	0.0076	0.00	0.00
2mM	6.26	0.0174	0.00	0.00
4mM	3.64	0.0111	2.70	4.68

Table 5-13 also shows summary results of percentage methylation with each aspirin treatment for the CpG site of the Illumina probe cg26432519 comparing the results from the EWAS and the results from pyrosequencing. Again, due to the contamination of the PCR control and the bisulphite modified control, it is difficult to derive any conclusions from the pyrosequencing results, although both methods indicate a decrease in CpG site methylation with aspirin (Table 5-13 and Figure 5-11 B).

Overall, whilst both methods allude to a similar effect of long-term aspirin treatment on decreasing DNA methylation, conclusions cannot be made before the pyrosequencing assay is repeated due to the presence of DNA contamination in both the bisulphite modified control and the PCR control as well as poor sample quality which have provided unreliable results.

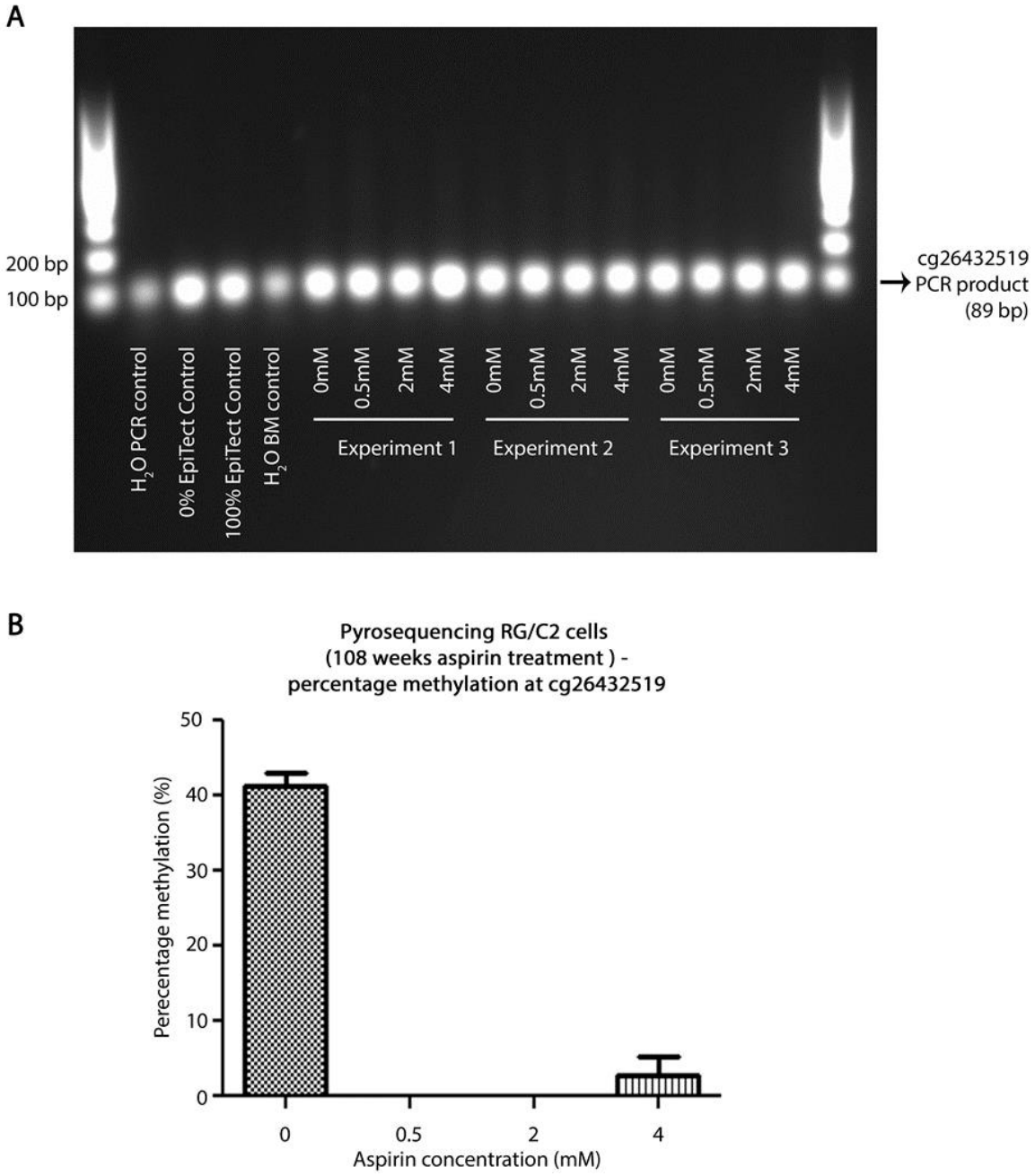


Figure 5-11- Long term aspirin treatment (108 weeks) effect on the CpG site from the Illumina probe cg26432519 methylation

RG/C2 cells were cultured for108 weeks with 0, 0.5, 2 and 4mM aspirin in 2D cell culture. DNA was extracted and then quantified before being subject to bisulphite modification, PCR and pyrosequencing. Results are mean values from 3 repeat experiments with standard deviations and a one-way ANOVA and Dunnett's multiple comparison test to calculate significance (*** $p \leq 0.001$, **$p \leq 0.01$ and * $p \leq 0.05$). (A) The agarose gel shows that the primers successfully produce the desired PCR product and that there is DNA present in all samples. The gel also shows that there appears to be DNA contamination in the bisulphite modified control (H_2O BM control) and PCR control (H_2O PCR control). (B) No peaks were detected in the 0.5mM and 2mM treated samples indicating poor sample quality. Aspirin significantly decrease the percentage methylation of the CpG site from the Illumina probe cg26432519 at 4mM ($p \leq 0.001$). However, the quality of the samples is not high therefore these results need to be repeated and checked.

5.4 Discussion

We have previously shown that short term aspirin treatment has no significant effect on genome-wide DNA methylation (Chapter 4.3.1). However, many observational studies summarised in Chapter 3.1.1 suggest that the duration of aspirin may affect how it alters risk of colorectal cancer. Since aspirin appears to have a different effect when taken short-term or long-term, we set out to identify whether long-term exposure to aspirin had a different effect on DNA methylation. Power calculations were carried out using the statistical software G*Power 3.1.9.2 (352). Our study had 80.6% power to detect a mean methylation difference between aspirin treated and untreated samples of 0.18 beta (or 18% change in methylation) with a standard deviation of 0.1 for 9 treated samples vs 3 untreated samples.

To our knowledge, aspirin's effect on DNA methylation is largely unknown as most research has focused on its effect on protein and histone acetylation (78). Overall, we have shown that long-term aspirin affects genome-wide levels of DNA methylation. It did not specifically methylate multiple CpG sites in one or two chromosomes/genomic regions only, rather it significantly affected the methylation of multiple CpG sites across all the chromosomes. This is similar to aspirin's effect on other epigenetic modifications as evidenced by Tatham et. al (2017) who found that aspirin acetylated proteins globally and did not enrich for a specific functional group when compared to a control list of proteins from untreated HeLa cells (78). However, when we compared our results to a normal distribution through a QQ plot, it indicated the presence of confounding which was systematically affecting the significance of all the results. Whilst we adjusted for batch only, it may be that there are other factors that may be biasing the results, for example, we did not control for the time at which DNA was extracted from cells as this was done when cells reached a certain confluency in the flasks. Since 2mM and 4mM grew at a much slower rate than 0mM and 0.5mM aspirin treated cells, these were cultured for a longer time before DNA was extracted.

We attempted to adjust for this confounding using surrogate variable analysis which mathematically calculates surrogate or potentially confounding variables and includes them in the logistic regression model. After carrying out SVA, the observed inflation in the test statistics was removed but the most statistically significant CpG sites were those with very small changes in average methylation fluorescence. For example, the most pronounced site using the SVA model was the CpG site from the Illumina probe number cg21232203 whereby aspirin decreased the percentage of methylation by 7% in 0.5mM treated cells and 5% in 2mM and 4mM treated cells when comparing the raw averages of DNA methylation. The SVA results indicate that the most significant CpG sites are those where aspirin hardly affects DNA methylation leading to the possible conclusion that surrogate

variable analysis is removing the effect of aspirin itself on CpG methylation which suggests over-adjustment. Jaffe et al. (2015) stated that one must be careful when 'cleaning' biological data e.g. by carrying out surrogate variable analysis, as it may result in the removal of a true biological signal or otherwise retain unwanted variability (406). After calculating a correlation coefficient between the 6 SVs and aspirin concentration, one of the SVs was indeed highly correlated with aspirin treatment with an R^2 of 0.96 indicating that this SV is simply adjusting for the effect of aspirin treatment. For this reason, and due to the nature of the controlled laboratory environment making it unlikely for confounding to occur, we did not think results from surrogate variable analysis were representative and appropriate to use.

After adjusting for multiple testing using a Bonferroni correction, 1562 CpG sites were shown to be associated with aspirin treatment. Since a high number of CpG sites fell below this threshold, we decided to use the more stringent threshold of 5×10^{-8} as is currently used for genome wide association studies of which 1492 CpG sites fell below this P-value, of which it methylated 493 and demethylated 999 CpG dinucleotides.

Overall, aspirin treatment favours a decrease in DNA methylation which is consistent with aspirin's association with increased acetylation of histone proteins as indicated by the literature (78,188–190). The literature suggests an inverse relationship with active histone marks and DNA methylation as previously summarised in Chapter 4.1.6. For this reason, as previous literature suggests aspirin treatment increases acetylation, we would expect to observe a decrease in DNA methylation.

In terms of genetic regions targeted by aspirin, demethylation occurs within promoter regions and methylation occurs in the 3' UTR. Hypermethylation and hypomethylation both occur in intergenic regions. It is known that CpG islands containing many CpG sites exist in promoter regions. The methylation of these dinucleotides determines whether a gene is expressed or not. CpG site methylation represses gene transcription and demethylation allows for gene transcription (399). Therefore, the demethylation of promoter sites may be increasing expression of certain genes. As for methylation of the 3' UTR, these sites are known for regulating the fate of mRNAs such as localization, translation and degradation (400). Choi et. al (2009) stated that methylation at the coding boundaries (5' UTR and 3' UTR) plays a role in controlling RNA splicing and transcription elongation (407). It is also well-known that microRNAs (miRNAs) target these 3' UTRs and their binding inhibits the initiation of mRNA translation (408). The consequence of 3' UTR methylation on the ability of miRNAs to bind to them is still unclear but it is possible that the increase in methylation at this coding boundary may affect miRNA binding and thus gene translation.

We tried to identify whether aspirin targets specific genetic regions or CpG sites within a specific CpG context. Of the most interesting results, aspirin may be demethylating promoter sites which would indicate an increase in transcription and gene expression (399). With regards to CpG context, our results show that aspirin targets open sea regions for methylation. Open sea regions are CpG sites outside of the CpG islands. Blood samples from healthy Dutch patients show that these regions are mostly hypermethylated unlike CpG sites in islands that are mostly unmethylated (409). Klett et. al (2018) analysed the methylation status of 77 CRC tissues and 108 adjacent mucosa tissues using the 450K array. They observed the opposite association in CRC tissues whereby CpG islands were hypermethylated and open sea CpG sites were hypomethylated (410). Aspirin is hypermethylating open sea CpG sites therefore, it may be that aspirin is preventing cancer progression by reversing epigenetic changes that occur along the cancer development pathway suggesting a possible mechanism by which aspirin is therapeutic.

CpG island shores are sequences that are 2-kb either side of the CpG island. They are important for cell differentiation and are targeted for aberrant DNA methylation in colorectal cancer. Irizarry et. al (2009) showed that most DNA methylation changes occur in the shore regions of CpG islands. They analysed DNA methylation from 13 colon cancers and their adjacent mucosal tissue and observed both hypomethylation and hypermethylation of CpG island shores (172). McInnes et. al found that compared to normal mucosa, 12% of hypomethylated CpGs in colorectal tumour samples were found in shore regions (411). We also found a similar effect with aspirin treatment whereby the percentage of hypomethylated CpGs found in north and south shores totalled 11.91%. The exact interpretation of this biological effect is not yet clear.

After testing CpG sites for ontology enrichment using the KEGG collection of pathways, our results showed that aspirin targets many pathways, especially in the sites targeted for a decrease in methylation: aspirin demethylates CpG sites in pathways associated with colorectal cancer progression such as the MAPK signalling pathway (401), the cytokine-cytokine pathway such as IL-6 and TNF (402), the insulin and IGF signalling pathway (403) and the PI3K-AKT signalling pathway (404). CpG site methylation (particularly in promoter regions) represses gene transcription and demethylation allows for gene transcription (399). Surprisingly, as demethylation suggests an increase in gene expression, demethylation of CpG sites related to these pathways results in the over-activation of these pathways and hence cancer progression, contrary to its role as an anti-cancer agent. However, evidence suggesting “aspirin-resistance” comes from epidemiological studies (Chapter 3.1.1) and from our own experiments whereby long-term aspirin treated cells are no longer sensitive to the growth inhibitory effect of aspirin (Chapter 3.3.2.1). The demethylation and over-activation of pathways involved in cancer progression may be one of the mechanisms by

which patients become less sensitive to aspirin treatment. Short-term aspirin treatment has no discernible effect on DNA methylation (Chapter 4.3.1) suggesting that long-term treatment is required for epigenetic modifications.

Mansoori et. al (2017) stated that one of the mechanisms through which cancers become resistant to a drug is through epigenetic alterations, whether histone alterations or DNA methylation changes (412). The phenomenon of drug resistance through changes in DNA methylation is not unique to aspirin. Chen et. al (2015) found that treatment of the cervical cancer cell line S3 with oxaliplatin lead to resistance and increased global DNA methylation. Through monitoring of DNA methylation, they found that they could identify at what point cervical cancer cells became resistant to cisplatin and taxol (413).

Terai et. al (2013) developed gefitinib resistant PC9 lung cancer cell lines through culturing the cells with the epidermal growth factor receptor (EGFR) inhibitor drug for approximately 6 months (414). They then interrogated whether long-term treatment with gefitinib affected DNA methylation through the use of the Infinium HumanMethylation27 Bead Arrays which measures methylation for approximately 27,578 sites and found that methylation was increased in 640 genes of the resistant cells compared to the parental cells. After combining their results with cDNA microarray results for gene expression they identified 29 genes affected by long-term exposure to the drug, of which 2 were involved in drug resistance (415). Whilst they did not investigate mechanisms involved in epigenetic changes through long-term drug exposure, our findings are similar to their results in that long-term treatment with a drug may induce drug resistance through epigenetic alterations.

Whilst we have not yet explored whether any of the CpG sites methylated are involved in drug resistance, it may be worth obtaining cDNA expression results for our samples as well to be able to identify genes where DNA methylation and leads to a change in mRNA expression, and then carry out functional experiments such as the induction or silencing of genes to find out if this rescues the cells from an “aspirin-resistant” phenotype.

One other finding from our results that may be important for informing clinical practice is the apparent uniform effect of aspirin on DNA methylation regardless of dose. There is very little in the literature with regards to aspirin or NSAID dosage and its consequence on DNA methylation. Our results show that 0.5mM, 2mM and 4mM have similar effects on DNA methylation suggesting that there may be a concentration threshold to obtain maximum effect of aspirin on DNA methylation. Our results are important therapeutically as they may help clinicians recommend smaller doses of aspirin to patients. This is especially important to avoid the unwanted side effects of long-term aspirin treatment, as previously summarised in Chapter 1.4.4.

We attempted to validate results from an EWAS by measuring percentage methylation of specific CpG sites using pyrosequencing which showed a similar effect of aspirin on CpG site methylation to the 850K array. However, we did experience contamination in our control samples. Due to time constraints, it was not possible to repeat these experiments, therefore these experiments need to be repeated before any conclusions can be made. When the pyrosequencing method is successful, the advantage of it is the quantitative nature of the result- it is known to detect differences in methylation as small as 5%. However, in the case of our experiments, it may also be useful to validate our EWAS results through other methods such as digestion-based assays followed by PCR or qPCR or methylation specific PCR of bisulphite converted samples (416).

With regards to chemoprevention, our results indicate that aspirin has similar effects on CpG site methylation regardless of dose, although we acknowledge that this needs to be shown through other methods such as pyrosequencing assays. If this is indeed true, it may help clinicians advise patients to take small doses of aspirin for chemoprevention.

Our results indicate that aspirin has similar effects on CpG site methylation regardless of dose, although we acknowledge that this needs to be shown through other methods such as pyrosequencing assays. If this is indeed true, it may help clinicians advise patients to take small doses of aspirin for chemoprevention. Evidence for aspirin's therapeutic effect comes from its hypermethylation of open sea CpG sites, possibly reversing CpG site hypomethylation in CRC tissue (410). However, our pathway analysis indicates that long-term aspirin may possibly demethylate cancer-associated pathways and may possibly be involved in "aspirin-resistance". The consequence of long-term aspirin on DNA methylation for therapeutic use is still unclear.

To our knowledge, we are the first to show that long-term aspirin treatment has a genome-wide effect on DNA methylation in colorectal adenoma cells. The consequence of the changes in DNA methylation caused by aspirin and how this drug is able to affect methylation patterns is not yet known and requires further investigation. It is important to note that these results are still preliminary and needs to be replicated in other cell lines or with DNA samples from adenomas in patients who take aspirin regularly. The identification of CpG sites that confer "aspirin resistance" may help in monitoring patients and inform whether the patients should continue taking the drug or not. For this to happen, a gene expression microarray of the long-term aspirin treated adenoma cells needs to take place to identify concurrent changes in DNA methylation and mRNA expression. Furthermore, these markers need to be validated in patients who are on long-term aspirin treatment to confirm the *in vitro* results are also present *in vivo*.

Chapter 6 Using SNPs from enzymes in the aspirin metabolism pathway to predict metabolite levels and risk of CRC – a Mendelian randomisation approach

This chapter aims to apply a Mendelian randomisation (MR) approach to a biochemical pathway to identify whether SNPs can be used to proxy for aspirin metabolites and identify whether aspirin metabolite levels are associated with risk of colorectal cancer, circled in red in Figure 6-1. Considerations of using MR for biochemical pathways and cancer is first discussed before a detailed description of how MR is undertaken is explained.

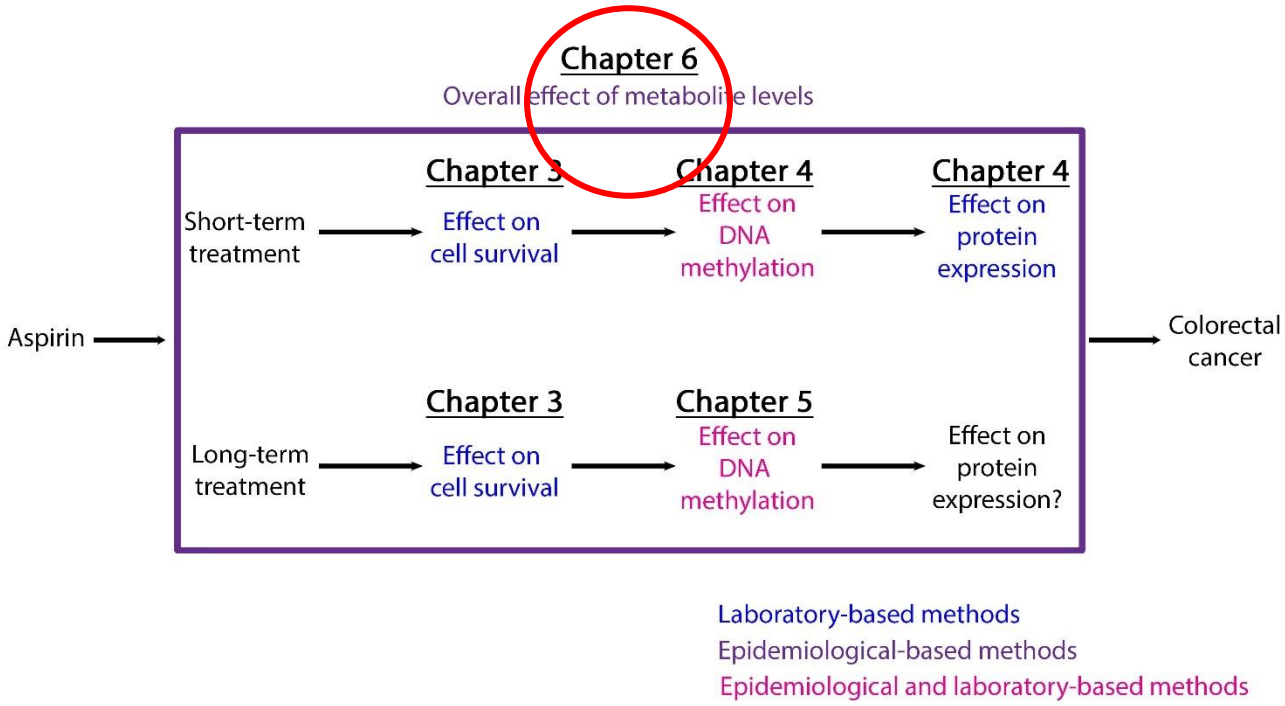


Figure 6-1- Overview of the questions addressed in this thesis to explore causality and potential mechanisms of aspirin on CRC.

6.1 Background

6.1.1 Using Mendelian randomisation to gain causal evidence in a biochemical pathway

Mendelian randomisation (MR) has and is conventionally used to assess the causal relationship between a modifiable exposure (e.g. alcohol intake, smoking, dietary factors) and a disease outcome (209). Thanks to the increase in technologies that allow the measurement of intermediate

biomarkers at a large scale and generate GWAS data of these predictive biomarker traits (e.g. lipids, amino acids, c-reactive protein), MR has extended its application in this area (417).

In this chapter, we demonstrate a novel approach that begins with knowledge of a biochemical pathway and seeks to apply MR to the enzymes controlling the rate limiting steps of that pathway. Biological knowledge of the pathway allows the informing of which genetic variants to use an instrumental variable in the analysis.

A background on the aspirin metabolic pathway has been previously described (Chapter 1.8.4). This chapter discusses how MR has or can be used in cancer studies. Details of how MR has been carried out has been outlined in general (Chapter 6.1.6) before the methods and results of the analysis are presented.

6.1.2 Aspirin metabolism pathway considerations

In this thesis we focus on aspirin metabolites measured in the blood plasma (Chapter 1.8.4). With regards to aspirin metabolite measurements, these blood plasma measurements are dependent on a variety of factors, such as whether aspirin was taken just before blood was taken from the study participant, or whether aspirin was taken on an empty stomach or whether the participants take any other medications. The timing of aspirin ingestion and the other factors mentioned are important as they may affect the concentration of the metabolites detected. As well as through the ingestion of aspirin, it is also important to consider that salicylates can also be sourced from the diet, albeit at lower concentrations (Chapter 1.4.3). In our analysis, the metabolite measurements are likely to represent dietary salicylates rather than aspirin taken for medicinal purposes.

6.1.3 The use of Mendelian randomisation to inform cancer trials

MR is increasingly being applied in the cancer field. With respect to randomised controlled trials, an MR approach is useful in that it can test results from observational studies before a large amount of money is spent on conducting randomised controlled trials. One such example where MR would have been useful is with regards to The Selenium and Vitamin E Cancer Prevention Trial (SELECT) trial (418). Observational data had previously shown that increasing selenium levels decreases the risk of prostate cancer (419–421). For this reason, the SELECT trial was set up. The results from this trial found that selenium may increase the risk of high grade prostate cancer and risk of type 2 diabetes and was therefore abandoned before being completed (418,422). Yarmolinsky et. al (2018) used an MR approach using genetic instruments that reported the same uptake of selenium as the trial (a median of 114 lg/L increase in circulating selenium) and their findings mimic the results of the trial. They found that selenium was associated with advanced prostate cancer and type 2 diabetes, albeit weakly (229). This MR analysis highlights how results from this approach may have been useful

to inform whether to conduct a trial that cost $114 million and that was actually weakly associated with increasing high grade prostate cancer risk (229).

MR has been used in other instances to confirm or provide evidence for/against associations obtained in observational studies. A meta-analysis of observational studies for vitamin D levels and CRC risk showed an inverse association whereby low levels of vitamin D increases CRC risk (423). To test this association, He et. al (2018) used genetic instruments to proxy for 25-hydroxyvitamin D and actually found no causal relationship between this metabolite and risk of CRC (424).

Alternatively, MR can be used to further confirm results from observational studies. One such example is the effect of high body mass index (BMI) on increasing the risk of CRC (425) Jarvis et. al (2016) later confirmed this through using genetic instruments associated with BMI, waist-hip ratio and childhood obesity and showed that these factors all increase risk of CRC (426).

Therefore, MR is a quick and efficient way to test the effect of lifestyle or modifiable factors and assess their causal relationship with cancer incidence before spending large quantities of money to conduct clinical trials.

6.1.4 Considerations when conducting Mendelian randomisation for exposures and cancer risk

The MR approach provides some advantages over randomised controlled trials. Limitations of trials with regards to cancer prevention include the fact that prevention trials focus on a limited time frame as opposed to life-long exposures (197). Often, long follow-up period to a trial is not possible which means that sometimes the effect of an intervention remains undetected. Furthermore, the issue with long term chemoprevention trials can be non-compliance of those taking the intervention and contamination of the control arm (198). MR is useful in this sense because it works by the principal that a person's genotype is determined at conception. Therefore the exposure that a particular genotype is associated with proxies for a lifelong exposure to that factor (427)

Also, timing of exposure to prevent cancer may be most effective in the earlier stages of life and would therefore not be tested in a randomised controlled trial (428). However, addressing the long-term effect of an exposure on cancer risk is not feasible as carrying out intervention trials in children and adolescents would not be ethical or practical (198). Again, MR is useful as it proxies for lifelong exposures (427).

An important consideration when addressing the effect of an exposure on an outcome is whether the said exposure affects cancer incidence or cancer progression, or both. For example, dietary folate may reduce CRC risk before neoplasia but may promote tumour progression through

increased proliferation and invasion after the development of the tumour (429). It is important to interrogate both questions as results will affect whether the exposure/treatment should be used for prevention of incidence or for reducing progression. In addition, it is important to note that when conducting MR studies, cancers must be divided into their subtypes as exposures may have different effects depending on the cancer types (430).

6.1.5 Mendelian randomisation and drug development or re-purposing

Drug discovery is a huge business in the field of medicine. However, 85-90% of new approved drugs since mid-1990s provide little or no clinical advantage (431). Pre-clinical trial experiment/data focus solely on the drugs intended targets. This, however, is not a holistic representation of the drugs' effects. Drugs may have off-target adverse effects or off-target protective/beneficial effects allowing for drug repurposing (123).

It is believed to take around 10-17 years from the conception of a drug to when it becomes a licensed treatment for disease with a success rate of less than 10% (124). MR has been most useful in drug development when it provides evidence against any causal association between a biomarker and diseases, therefore reducing monetary investment and time in the development of the drug (432).

MR has the potential to address drug association with disease through using biomarkers of the drugs targets. It can also test for association with off-target effects as well as other diseases for the purpose of drug repositioning (433).

As an established pharmaceutical agent, there is no lead time to 'discover' a new therapeutic. However, MR would be useful in assessing whether aspirin may have any off-target effects. This would highlight added benefits or competing effects that would need to be considered if this drug was to be used as a prophylactic for CRC prevention. Applying MR in this context could potentially highlight other benefits and therefore uses of aspirin.

6.1.6 Mendelian randomisation methodology

MR is based on the premise of using genetic variants to proxy for a specific factor of interest. The MR field has progressed rapidly in recent years (242) and multiple considerations need to be taken before the method is implemented. Some key considerations are outlined in the following sections.

6.1.6.1 Genetic instrument selection

These are SNPs that are found to be strongly associated with a specific exposure, commonly through a genome wide association study (GWAS)(206). Alternatively, a biological understanding of the exposure may help in formulating the question. In this case, we hypothesise that SNPs from the

enzymes involved in aspirin metabolism may be associated with changes in levels of the aspirin metabolites salicylic acid and salicyluric acid.

6.1.6.2 Test for SNP association with exposure- salicylic acid

Once a hypothesis has been established, the next step is to test the SNPs' association with the exposure. If the exposure is continuous, then a linear regression is carried out. Essentially, this involves averaging the mean value of metabolite level for each of the 3 genotypes for a single SNP (e.g. AA, AC and CC). Then, a linear regression is carried out using the equation $Y=\beta x+c+\varepsilon$ whereby β is the slope of the line, c is the y intercept and ε is the error term. The value of most interest is β as we are interested in how much the levels of salicylic acid changes per additional allele.

Considering the example shown in Figure 6-2, for every C allele change, the predicted level of salicylic acid in these individuals increases by 0.4mM. Therefore, on average, individuals with an AA genotype for this particular SNP are predicted to have 0.2mM salicylic acid levels whereas individuals with a CC genotype are predicted to have 1mM salicylic acid levels.

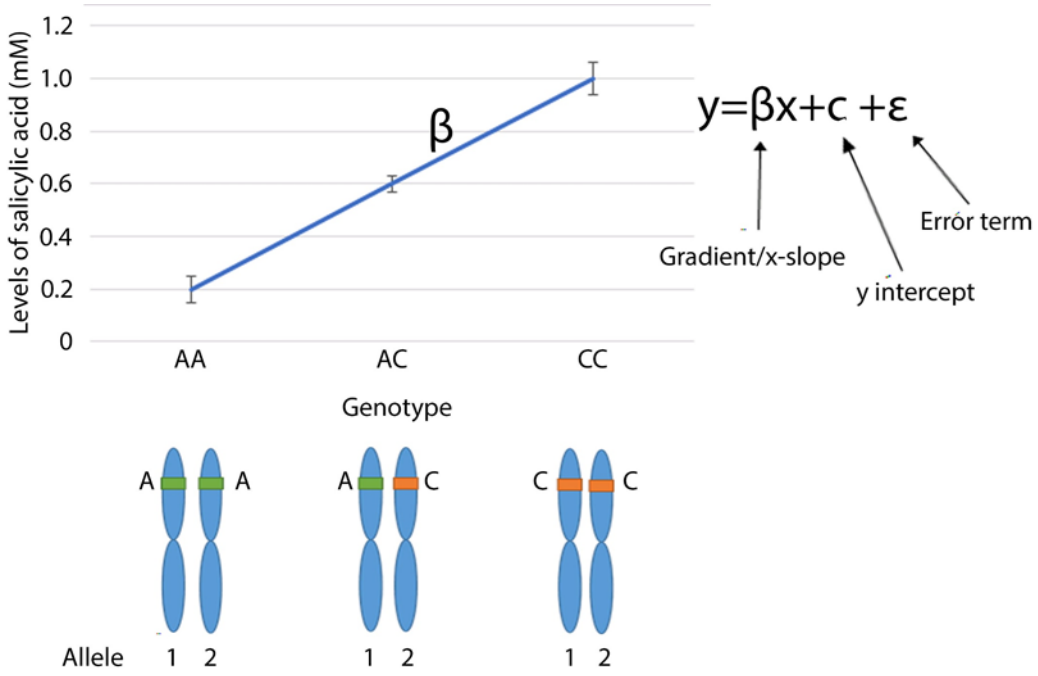


Figure 6-2- Testing for SNP association with exposure using linear regression

Linear regression is used to test for a SNP association with a continuous outcome, in this case, salicylic acid levels. Average levels of salicylic acid for each genotype of a SNP are calculated including all the study participants. A gradient/slope (termed β) is then calculated to identify whether a statistically significant effect is present for each allele change. The graph shows an increase in salicylic acid levels by 0.4mM for every A > C allele change. Included in the regression model is a y intercept(c) and an error term (ε) and represents any residual error between the independent variable (SNP genotype) and the dependent variable (salicylic acid levels).

6.1.6.3 SNP clumping

One of the main assumptions of Mendelian randomisation is that SNPs are inherited independently, however, this is rarely so. Linkage disequilibrium (LD) is the non-random association of SNPs due to their close proximity resulting in a decreased likelihood of meiotic recombination (Figure 6-3) (434). It is a measure of whether there is any difference between the observed frequency of the two loci being inherited and the expected frequency. One of the common methods to measure this is through calculating r^2 which ranges from 0-1 where 1 denotes the 2 SNPs to always be inherited together, thus the value of this number describes the degree to which one SNP is inherited with another SNP (435). SNPs in high LD are removed so as to use one SNP to represent the association for the other SNPs e.g. if 5 SNPs are in high LD and have similar effect sizes, then there is no need to use all 5 as they are essentially "replicates" of each other, therefore one is chosen to represent all 5. This also helps when adjusting the results for multiple testing as it reduces the extent of corrections needed to account for the statistical burden of multiple testing. To choose SNPs that are truly independent of each other, we used an r^2 cut-off of 0.001. An r^2 of 1 indicates perfect LD, the closer the value is to 1, the higher the LD is (436).

SNP clumping is the process of removing all SNPs in high LD within 10,000kb of the variant of interested and leaving only the SNP with the lowest P-value of association with the exposure.

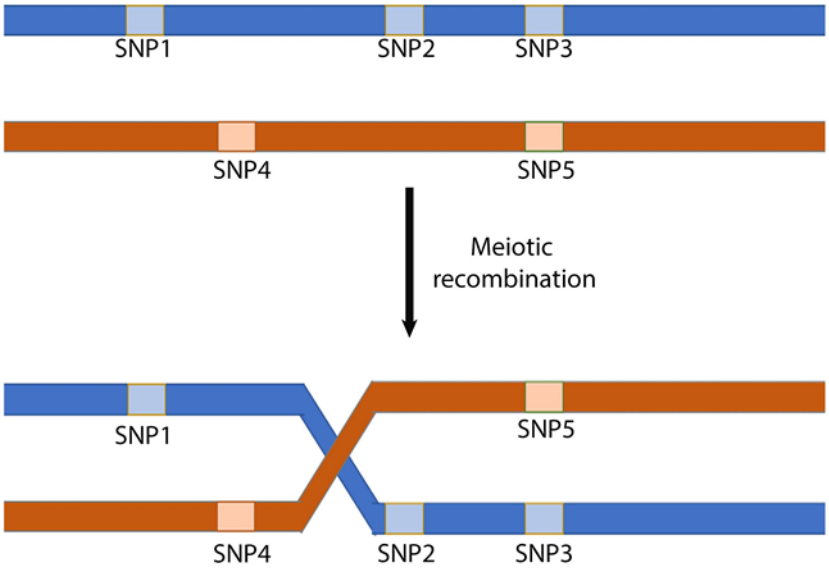


Figure 6-3- Linkage disequilibrium

SNPs may be co-inherited due to their close proximity decreasing the likelihood of meiotic recombination(434). The diagram shows how SNP2 and SNP3 in close proximity are not affected by meiotic recombination and so are inherited together.

6.1.6.4 Test for SNP association with outcome- colorectal cancer

After identifying which SNPs are associated with the exposure, these SNPs are then tested for their association with an outcome, in this case, CRC. Since incidence of CRC is a binary variable (either a case or control), then a logistic regression is carried out. Cases and controls are assigned numerical values, 1 and 0, respectively, to be able to carry out statistical analyses. One of the issues with using a linear regression for a binary outcome, is that the probability of the outcome may be less than 0 or more than 1 which is just not possible (Figure 6-4). Therefore, two adjustments need to be made to the linear regression model so that the probability of the outcome being a control is never less than 0 and the probability of it being a case is never more than 1 and this is done through fitting a logistic model (437). So we start off with the original linear regression model:

$$P = \beta x + c + \varepsilon$$

Exponents are taken of the equation so that the probability is always above 0

$$P = e^{\beta x + c}$$

The equation is then divided by itself +1 so that the probability is always less than 1

$$P = \frac{e^{\beta x + c}}{e^{\beta x + c} + 1}$$

Then the equation is rearranged as such to provide the logistic regression model.

$$\ln \frac{p}{1-p} = \beta x + c$$

Therefore the overall aim of a logistic regression function is to transform an S-shaped curve into a more straight line (437).

Typically, when looking at an outcome, an odds ratio (OR) is calculated. This ratio represents the odds that the outcome will occur with the exposure against the odds that the outcome will occur without the exposure present (438). In the logistic regression model, since we have taken logs in the model, the regression coefficient is the increase in the log OR of the disease per allele change in the SNP (241).

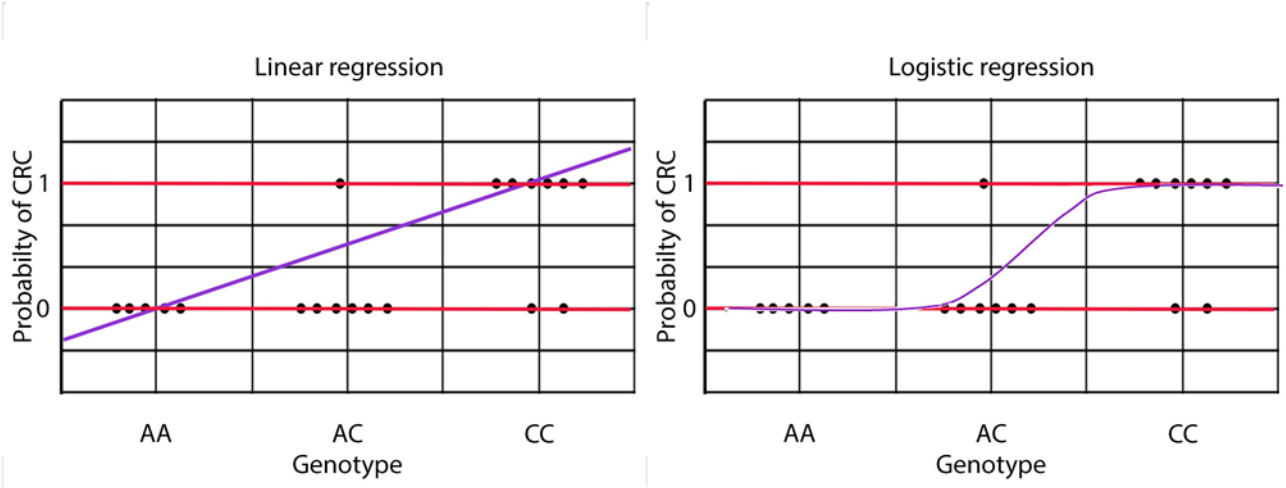


Figure 6-4- Linear and logistic regressions

The outcome of cases and controls has been converted to binary numerical values of 1 and 0, respectively. Each dot represents an individual. The red lines indicate the maximum upper and lower limits, 0 and 1. Therefore, in theory, the line of best fit should not go above or below these red lines. On the left is the general linear regression model that shows a continuous line of best fit (purple) above and below the maximum limits. On the right is the logistic regression model whereby the same equation of the linear regression model is log-transformed so that the line of best fit (purple) and outcome can never be more than 1 or less than 0.

6.1.6.5 Allele harmonisation

In a two-sample MR when data are utilised from more than one source, summary level data for gene-exposure and gene-outcome are from different samples (241). For this reason, it is important to harmonise the alleles; i.e. make sure that they both considered the same allele to be the effect allele (allele of interest) and that they have the same allele frequencies.

For example, if for a particular SNP, the exposure dataset (TwinsUK) termed the effect allele as G and the other allele as A, then ideally, we would want the outcome dataset (GECCO) to have also labelled the effect allele as G and the other allele as A. If GECCO had chosen the effect allele as A and the other as G, as long as the allele frequencies are similar to TwinsUK, then the effect of the outcome can be flipped e.g. from -0.05 to 0.05 to correspond with TwinsUK. A list of examples for allele harmonisations and any issues that can arise are summarised in Table 6-1.

Table 6-1- Allele harmonisation examples

Description		Effect allele	Other allele	Effect size	Minor allele frequency	Action
Correct reference, unambiguous	Exposure	G	A	0.5	-	Keep
	Outcome	G	A	-0.05	-	
Incorrect reference, unambiguous	Exposure	A	G	0.5	-	Flip outcome effect size
	Outcome	C	T	-0.05	-	
Ambiguous	Exposure	A	G	0.5	-	These don't correspond to the same SNP so drop
	Outcome	A	C	-0.05	-	
Palindromic SNP, inferable	Exposure	A	T		0.09	Each dataset has chosen a different forward strand. So flip the outcome effect size
	Outcome	A	T		0.91	
Palindromic SNP, not inferable	Exposure	A	T		0.5	Unlike above, the allele frequency here gives no information so can not infer whether it is forward strand or reverse strand
	Outcome	A	T		0.5	

6.1.6.6 Wald ratios

The Two Sample MR package (248,439) automatically changes the SNP association with the exposure so that they are all positive associations. This is done as follows: if a SNP with effect allele A and frequency of 0.1, with the other allele G and association with salicylic acid as -0.3, then the package will then simply assign allele G as the effect allele with an allele frequency of 0.9 and SNP effect on salicylic acid as 0.3. This process allows for ease of interpreting the results.

Once the alleles have been harmonised and SNPs all have a positive association with the exposure of interest, a Wald ratio is calculated. This is calculated using the formula: logOR/β.

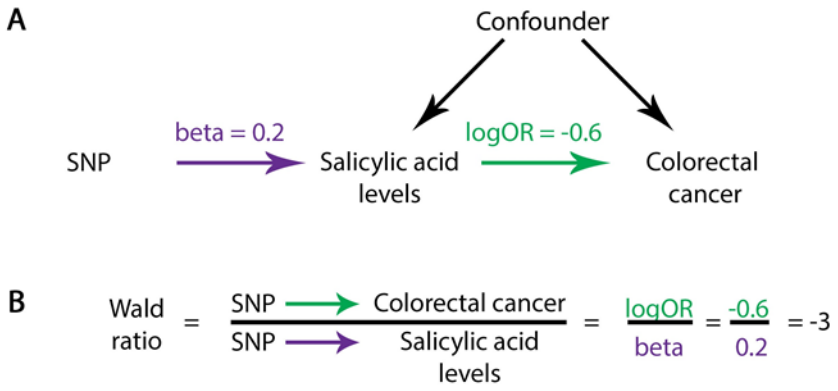


Figure 6-5- Example of Wald ratio calculation

(A) **A d**irected acyclic graph (DAG) showing genetic instruments to interrogate the effect of salicylic acid levels on risk of colorectal cancer. A linear regression of the SNP-exposure is conducted and identifies a 0.2 increase in salicylic acid levels per allele change. A logistic regression of the SNP-outcome is conducted and identifies a -0.6 logOR in risk of colorectal cancer per allele change. (B) A Wald ratio is calculated by dividing the per-allele effect of the SNP (the log OR) on the outcome by the per allele effect of the SNP (beta) on the exposure which in this case is -0.6/0.2 which results in a value of 3.

This ratio can be interpreted as the log odds ratio of the disease per unit increase/decrease to exposure due to the SNP (241). If for example we look at the SNP rs10882183, if the A allele in SNP predicts a 0.2 increase in salicylic acid and a log OR of CRC as -0.6, we would then deduce that for every A allele change in the SNP, the Wald ratio is - 3 (Figure 6-5). Therefore, for every A allele, the level of salicylic acid increases and the risk of CRC decreases.

6.1.6.7 Summarising results

Since more than one SNP can be used to proxy for an exposure (e.g. more than 5 SNPs may be associated with levels of salicylic acid and risk of CRC), then the results of the Wald ratios and their standard errors can be summarised using various method which include the inverse variance weighted (IVW) method, the maximum likelihood approach, the MR Egger approach, median-based estimates and mode-based approach. These have been previously described (Chapter 1.8.7). Essentially, these are just different methods and mathematical formulae to draw a line of best fit, each with different assumptions/relaxations of certain restrictions, an example of which is shown in Figure 6-6. The idea is to use as many methods as possible, each with different assumptions, to be able to assess how reliable the results are including evaluating the strength of the genetic instruments (SNPs) used (235,241).

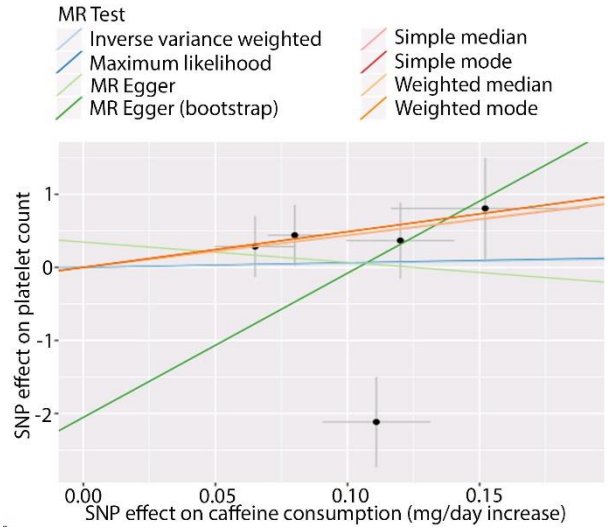


Figure 6-6- Examples of the methods used to summarise results

Various methods are used to summarise the Wald ratio estimates and their standard errors for the 5 SNPs associated with caffeine consumption and platelet count. Scatter plot regression lines: light blue, inverse variance weighted; dark blue, maximum likelihood; light green, MR Egger; dark green, MR Egger (bootstrap); light pink, simple median; dark pink, simple mode; light orange, weighted median; dark orange, weighted mode. These are just different methods and mathematical formulae to draw a line of best fit, each with different assumptions/relaxations of certain restrictions. The idea is to use as many methods as possible, each with different assumptions, to be able to assess how reliable the results are including evaluating the strength of the genetic instruments (SNPs) used.

6.1.6.8 Sensitivity analyses

Sensitivity analysis explore whether the observed relationships seen in an MR study are due to outliers having a disproportionate influence on the results. These analyses commonly include visualising the data to assess the distribution or re-running specific statistical tests in a stratified way or leaving specific data points out.

Leave-one-out analyses involve systematically dropping one SNP from the analysis to see if dropping them results in a dramatic change in the estimate. If so, this SNP may be considered an outlier and may be horizontally pleiotropic i.e. it is associated with another exposure/pathway. Leave-one-out analyses have been previously explained (Chapter 1.8.6) (248).

Funnel plots can also be used to visualise horizontal pleiotropy. The estimate of a SNP is plotted against its precision. Ideally, funnel plots should be symmetrical, meaning that there is balanced pleiotropy. If the plot is imbalanced, it may bias the IVW estimate(248) towards the pleiotropic SNPs, unlike the MR Egger estimate which is unaffected (Figure 6-7).

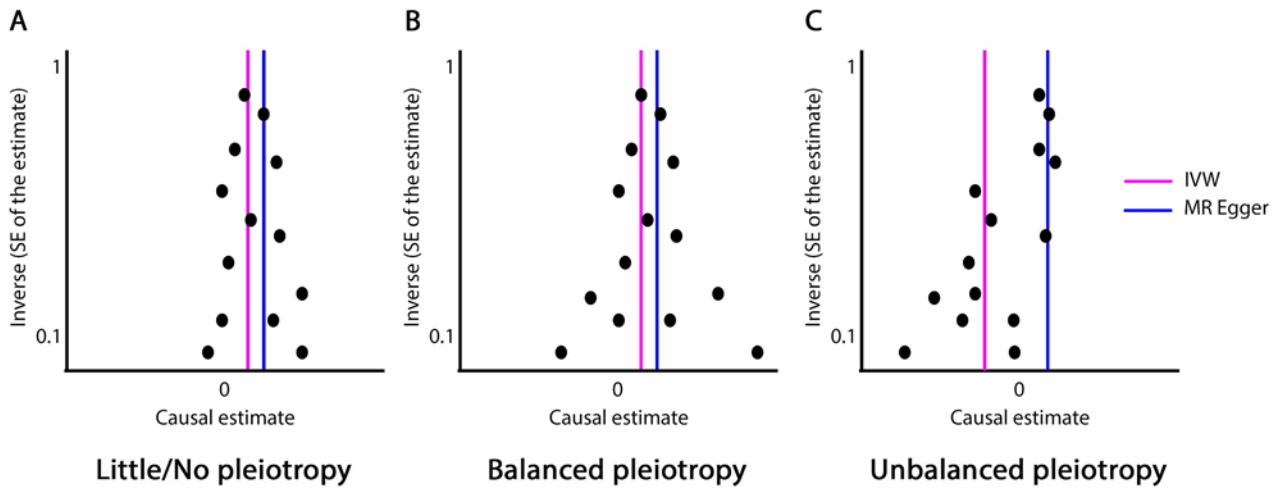


Figure 6-7- Examples of funnel plots and how they can be used to detect pleiotropy

These are 3 examples of plots that could be obtained from an MR analysis. The plots show the causal estimate of each SNP on the x-axis plotted against the inverse variance of the estimate on the y-axis. Each black dot represents a SNP and the summary causal estimates using the IVW and MR Egger methods are represented as pink and blue vertical lines, respectively. (A) The causal estimates for each SNP are symmetrical about the causal estimate and the data are not spread out very much which indicates little pleiotropy. (B) The causal estimates for each SNP are more spread out but there is still symmetry indicating balanced pleiotropy. (C) The causal estimates for each SNP are no longer symmetrical and this biases the IVW estimate towards the pleiotropic SNPs. On the other hand, the MR Egger estimate is unaffected. This figure is based on a slide from a presentation delivered by Dr Jack Bowden (440).

6.1.7 Limitations of Mendelian randomisation studies

Whilst MR has many advantages, it does have some limitations which need to be considered before interpreting results. These are summarised in Table 6-2.

Table 6-2- Considerations when interpreting Mendelian randomisation results

Limitations of MR studies	Explanation
Difficulty in finding genetic instruments	Although genome-wide association studies (GWASs) have allowed the discovery of many SNP-exposure associations, these have mostly been carried out on inexpensive or easy traits to measure (241). There is still a lack of GWASs for more specific traits in particular biochemical intermediates that are not often measured at a scale needed for GWAS.
Low power	Genetic polymorphisms usually inform us of a small percentage of the variation in a trait. Due to this, MR studies require large sample sizes to detect these small effects. An example: to detect a SNP that explains 1% of the variation of a trait, one would need around 9500 cases and 9500 controls to achieve a power of 80% to detect an OR of 1.5 per SD change in the exposure with a P-value of 0.05 threshold of significance. Multiple SNPs may be combined into an allele score to improve power but this should be done with caution otherwise bias may be introduced through the inclusion of invalid instruments (441).
Winner's curse	This involves using only the most significant SNP found in a GWAS despite the presence of other SNPs found within the same genomic region or loci. This SNP-trait effect may be an grossly inflated and may even be independent of the true effect size and is called winner's curse or Beavis effect (442). This correlation may occur due to chance. If the two samples of gene-exposure association and gene-outcome association are different, then the winner's curse will not affect the size of the causal estimate but it will tend towards a null value. However, if one was to use the same sample for discovery and MR analysis, the effect of the causal estimate will be overestimated (241).
Poor biological understanding	Plenty of GWASs have allowed the discovery of many SNP-trait associations. However, little is often known about these SNPs and their underlying mechanisms of association which may lead to counter-intuitive results (241).
Trait heterogeneity	SNPs can be associated with more than one trait or multiple aspects of the same trait, e.g. it may affect the size of the protein and the concentration of it (241).
Canalization	Genetic polymorphisms during development may allow alternative compensatory processes to reduce the phenotype impact of the SNP or genetic variant being used as a proxy for an exposure (443). This has the potential to lead to null or reduced associations between the genetic instrument and the outcome (241).
Estimating associations for binary outcomes	If the outcome is binary and the genetic associations with the outcome are estimated in a case-control setting, then the exposure effect on outcome is only approximate (444). If exposure data was obtained in a case-control setting after disease diagnosis, then gene-exposure associations may be subject to reverse causation unless the gene-exposure associations are obtained from controls only (445).

Having described the methodological principles, strengths and limitations of the MR approach, the remainder of this chapter describes its application to find out if there is a causal effect of the different aspirin metabolites on risk of CRC, the reasoning being that whilst there is evidence for a possible protective effect of aspirin use on CRC, it is still unclear as to which individuals would most benefit.

6.1.8 Hypothesis and aims

Hypothesis:

Increased levels of salicylic acid (the active metabolite (220)) confers a protective effect against CRC incidence and increasing levels of salicyluric acid increase the risk of CRC through decreasing the levels of salicylic acid.

Aims:

1- To identify SNPs that can be used to proxy for salicylic acid and salicyluric acid levels
2- To test for a causal relationship between SNPs that proxy for the metabolites and their association with CRC incidence in a MR framework
3- To test for a causal relationship between SNPs that proxy for the metabolites and their association with CRC incidence when stratifying between aspirin users and non-users, using a MR framework.

6.2 Methods

6.2.1 Genetic instrument selection

Through a literature search, it was established that the enzymes involved in aspirin metabolism in the blood are butylrylcholinesterase (BCHE) (217) and platelet- activating factor acetylhydrolase (PAFAH1b2) (217) and the liver are: UDP-glucuronosyltransferase 1-6 (UGT1A6) (221), cytochrome P450 2C9 (CYP2C9) (221) and acyl-CoA Synthetase Medium-Chain Family Member 2B (ACSM2B) (221) (Figure 1-20). The reasoning behind their choice is described in more detail in Chapter 1.8.4. We identified the start and end sites for these genes using the Ensembl resource (446) (Table 6-3). Using these genome coordinates, genotyped SNPs found in these regions were extracted from TwinsUK genotype data.

Whilst BCHE is able to convert acetylsalicylic acid into salicylic acid, other esterases also exist such as paraoxonase-1 (PON1) (224). However, the actual contribution of PON1 to this reaction is not yet clear (217). Nonspecific esterases also hydrolyse aspirin to salicylic acid in the liver, therefore other enzymes than those mentioned may also be involved (221). Whilst UGT1A6 is the main enzyme involved in the conjugation of salicylic acid into salicyl phenolic or acyl glucuronide, other closely related enzymes are also able to catalyze the same reaction including: UGT1A1, 1A3, 1A7, 1A8, 1A9, 1A10, 2B4 and 2B7 (225).

Table 6-3- Chromosomal regions of the genes for the enzymes involved in the metabolism of aspirin

Gene	Chromosome	Start site	End site
BCHE	3	165,772,904	165,837,472
PAFAH1B2	11	117,144,267	117,176,894
UGT1A6	2	233,691,607	233,773,300
CYP2C9	10	94,938,658	94,989,390
ACSM2B	16	20,536,226	20,576,427

6.2.2 SNP associations with metabolites

To calculate SNP associations with the aspirin metabolites salicylic acid and salicyluric acid, we used genotype and metabolite data provided by the TwinsUK registry- full details of the study are described in Chapter 2.12.1 (3). An inverse rank transformation was applied to the metabolites to makes sure there is adequate control of type I error rates. This reduced the effect of deviations and outliers from normality on statistical analyses (447). SNP associations with metabolites were tested using a linear mixed model approach implemented in BOLT-LMM (v2.2) (448) which involves calculating the BOLT-LMM-inf statistic (standard (infinitesimal) mixed model association) (448) to adjust for population structure, relatedness and repeated measures. We included age and year of measurement as quantitative covariates and included sex and visit (1 -3) as binary/categorical covariates in the regression model. Salicylic acid was measured using gas chromatography-mass spectrometry (GC/MS) and salicyluric acid was measured using liquid chromatography-mass spectrometry (LC/MS). TwinsUK used the Metabolon platform (Metabolon Inc., Durham, USA) which is a non-targeted metabolomic analysis. It uses 2 LC/MS injections (one for acid and for basic species) as well as one GC/MS injection per sample. This allows the detected of 510 small molecules that include salicylic acid and salicyluric acid. Since this is an untargeted approach, the original units of measurement are so-called "ion counts" assessed as the area under the curve of the corresponding peak in the mass spectrum therefore these measurements are quantitative values of relative changes as opposed to the targeted approach which would achieve absolute quantification of metabolite concentrations (3,267). Therefore a linear mixed model approach was used to test for associations between the SNPs and the metabolites, adjusting for age, year of measurement, sex and visit.

6.2.3 Stratification by aspirin use

During a 5-year period starting from the year 2005, around 500,000 participants aged between 45-64 took part in UK Biobank. An online questionnaire and interviews were used to collect data on medication use such as aspirin . Using the UK Biobank data, we extracted genotype data for individuals who take aspirin 75mg (Field ID 20003, coding 1140868226), nu-seals aspirin 75mg enteric-coated tablet (Field ID 20003, coding 1140864860) and aspirin (Field ID 20003, coding 1140861806). These were classified as aspirin users. We also obtained genotype data for individuals who did not report taking any treatment or medications under the Field ID 20003. These were classified as the aspirin non-users. For each aspirin users and non-users, CRC cases and controls were assigned numerical values, 1 and 0, respectively.

6.2.4 SNP associations with CRC

Firstly, SNP associations with CRC were calculated using the GECCO consortium (2). Single variant analyses using unconditional logistic regression for CRC was conducted adjusting for sex, age, genotyping phase, batch effect for ASTERISK and 3 principal components using data provided by GECCO. Since a logistic regression was carried out, GECCO results are calculated as log odds ratios.

As a replication study, UK Biobank data was also used (449). SNP associations with CRC were tested using a linear mixed model approach in BOLT-LMM (v2.2) (448) which involves calculating the BOLT-LMM-inf statistic using UK Biobank data for CRC incidence adjusting for sex, genotyping array and the first 10 principal components. Therefore, a linear mixed model approach was conducted to test for SNP association with CRC incidence adjusting for age, sex and genotyping array and the first 10 principal components). Since a linear mixed model approach was carried out using UK Biobank data, these results are interpreted as the change in prevalence of the outcome per copy of the effect allele or unit change in exposure. This means that e.g. if beta is 0.01, this is the proportionate change in prevalence so equates to a 1% increase in prevalence. The analysis was also repeated but stratifying between aspirin users and non-users in UK Biobank (Chapter 6.2.3).

More details for the GECCO consortium and UK Biobank data are described in Chapter 2.12.

6.2.5 Statistical methods

Analyses were carried out in R version 3.2.3 using the “Two-Sample MR” package (248). This package allows the formatting, allele harmonisation and thus analysis of summary data in an automated manner.

6.2.5.1 Data formatting and LD clumping

Data was formatted and labels renamed to the requirements of the package. Firstly, a P-value threshold of ≤0.05 was set on SNPs from the TwinsUK regression analysis. Using the clump_data() function in the package, data was LD-clumped to remove SNPs in linkage disequilibrium at an R^2 of more than 0.001. The SNPs identified as associated with the metabolite of interest (salicylic acid or salicyluric acid) were then extracted from GECCO and UK Biobank datasets. In order to calculate the Wald ratio and estimate causal associations, genotype data for the exposure and outcome was harmonised using the harmonise_data() function using option 2 so that the strand of palindromic SNPs was inferred using the allele frequencies.

6.2.5.2 Calculating significance - adjusting for multiple testing

Where relevant, a Bonferroni and FDR-adjusted P-value were both calculated to adjust for the issue of multiple testing. These have been described in more detail in Chapter 4.1.7.2

6.2.5.3 Power calculations

Based on the 8 salicylic acid genetic instruments that we are identified in TwinsUK (Chapter 6.3.2), these variants explain approximately 0.25% of the variation in salicylic acid levels. With the GECCO sample size of N=28059, we have 80% power at α level 0.05 to detect an effect of salicylic acid on CRC risk of an odds ratio ≤0.51. In the reciprocal direction, we have 80% power to detect an effect of salicylic acid on CRC risk of an odds ratio ≥1.9. With the UK Biobank sample size of 462933 (of which 641 were cases), we had 80% power at α level 0.05 to detect an effect of salicylic acid on CRC risk of an odds ratio ≥3.2 only.

From TwinsUK, we also used 9 genetic instruments (Chapter 6.3.4) that explained 0.25% of the variation in salicyluric acid levels. Using GECCO data, we have 80% power at α level 0.05 to detect an effect of salicyluric acid on CRC risk of an odds ratio ≤0.51. In the reciprocal direction, we have 80% power to detect an effect of salicylic acid on CRC risk of an odds ratio ≥1.9. With the UK Biobank sample size of 462933 (of which 641 were cases), we had 80% power at α level 0.05 to detect an effect of salicylic acid on CRC risk of an odds ratio ≥3.2 only. The mRnd online calculator was used to calculate statistical power(450) (available: http://cnsgenomics.com/shiny/mRnd/).

6.2.5.4 MR methods

Using the mr() function in the Two-Sample MR package, we calculated the causal estimates using 8 methods, all of which have been summarised in Chapter 1.8.7:

1. Inverse variance weighted (IVW) method
2. Maximum likelihood approach
3. MR Egger
4. MR Egger bootstrap
5. Simple mode
6. Weighted mode
7. Simple median
8. Weighted median

6.2.5.5 Sensitivity analyses

The Two-Sample MR package allows for sensitivity analyses, three of which were used in this analysis (248). These include:

1- Single-SNP analysis

The causal estimate of each SNP on exposure and outcome is calculated using the Wald ratio method. The estimate of each SNP can then be visualised using a forest-plot.

2- Leave-one out analysis

This is where the summary causal estimate is calculated but dropping one SNP from the analysis each time to see whether the SNP dropped is predominantly responsible for affecting the exposure-outcome association (235). By default, the IVW method is used to calculate the summary causal estimate.

3- Funnel plots

Funnel plots can also be used to visualise horizontal pleiotropy. The estimate of a SNP is plotted against its precision. Ideally, funnel plots should be symmetrical, meaning that there is balanced pleiotropy. If the plot is imbalanced, it may bias the IVW estimate (248) towards the pleiotropic SNPs, unlike the MR Egger estimate which is unaffected. A diagrammatical representation of this is shown in Figure 6-7.

6.2.5.6 Representation of results

Using the TwoSample MR package, results were presented as scatter plots, forest plots showing single SNP ratio estimates, funnel plots and “leave-one-out” plots (248).

6.3 Results

6.3.1 Study participants basic characteristics

Data on a total of 3135 individuals was used for identifying SNP associations with salicylic acid and salicyluric acid using TwinsUK (3). Blood samples were taken for more than one session and these are summarised as visits in Table 6-4. The vast majority of the participants were female in each of the visits ranging from 73.11% to 96.61%. The average measures of salicylic acid and salicyluric acid levels varied across the different visits ranging from -0.542-0.186 and -0.259-0.260, respectively. Average BMI measures remained consistent ranging from 25.41 to 26.48 with the date of the first participant measure in October 1996 and the last participant measure in June 2015.

Table 6-4- Baseline characteristics of TwinsUK participants

Visit	N	Average age	Age range	N Males	N Females	Percentage female (%)	Average BMI	Date of first participant measurement	Date of last participant measurement	Average measure of salicylic acid	Average measure of salicyluric acid
v1	2058	51.50	32.8-73.7	70	1988	96.60	25.41	17/10/1996	06/12/2005	-0.115	-0.259
v2	2069	58.59	35.8-80.7	70	1999	96.61	26.48	15/10/1998	15/02/2012	0.015	0.010
v3	2069	64.71	42.4-87.8	70	1999	96.61	26.47	30/01/2006	03/11/2014	0.186	0.260
b1	111	38.68	18.8-70.8	14	97	87.39	25.54	24/04/1997	07/06/2001	-0.542	-0.105
b2	424	41.18	16.1-76.4	114	310	73.11	25.73	08/03/2004	28/07/2009	-0.150	-0.037
b3	590	47.25	18.2-79.7	125	465	78.81	25.18	20/09/2010	08/06/2015	-0.089	0.009

6.3.1.1 *Outcome number of cases and controls*

We used epidemiological and genetic data from 11895 cases and 14659 population-based controls of European descent from 25 studies that were part of the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) and Colon Cancer Family Registry (C-CFR). Full details on the consortium have been previously published (2). We also used epidemiological and genetic data from UK Biobank of 2816 colorectal cancer cases and 460197 controls (Table 6-5). Individuals aged between 38-73 years were recruited between 2006 and 2010 (451) and provided self-reported information on CRC incidence. A full description of the study design, participants and quality control (QC) methods have been described elsewhere (4). Further information for both studies is provided in Chapter 2.12. The number of CRC cases and controls stratified according to aspirin use are summarised in Table 6-5.

Table 6-5- Number of cases and controls used in the analysis

Study13175	N cases	N controls
GECCO	13175	14884
UK Biobank	641	462292
UK Biobank- aspirin users	7	4127
UK Biobank- aspirin non-users	112	130751

6.3.2 Salicylic acid genetic instruments

To test for a causal relationship between the SNPs that proxy for salicylic acid with CRC incidence, we used genotype and metabolite data provided by TwinsUK (3). Using the genomic locations of the enzymes involved in aspirin metabolism, a total of 4713 SNPs were available from the dataset. We excluded SNPS that were hemizygous or homozygous and ran a linear mixed model regression for the remaining SNPs on salicylic acid adjusting for age and year of measurement, sex and visit. We then removed SNPs that were not associated with salicylic acid at a P-value >0.05 therefore 162 SNPs remained for the analysis. We then clumped these at an r^2 of more than 0.001 to remove SNPs in high linkage disequilibrium (Figure 6-8). Therefore, we identified 8 independent SNPs that are associated with salicylic acid at a P-value of ≤0.05. These are summarised in Table 6-6.

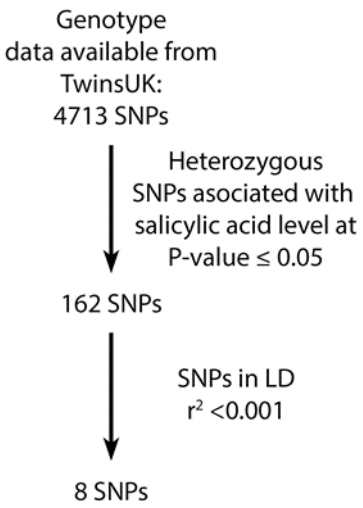


Figure 6-8- Flow diagram of the number of SNPs used in the analysis for salicylic acid

Using the TwinsUK dataset, a total of 4713 SNPs were genotyped for the genomic locations of the enzymes from the aspirin metabolism pathway. Homozygous and hemizygous SNPs were excluded from the analysis. After running the linear mixed model regression, we excluded SNPs not associated with salicylic acid at a P-value threshold of > 0.05. We then carried out LD clumping to remove SNPs in high linkage disequilibrium at an r^2 of more than 0.001. The result was 8 SNPs independently associated with salicylic acid levels at a P-value of ≤0.05.

Table 6-6- SNPs used to proxy for salicylic acid metabolite levels from TwinsUK

SNP	Gene	Chr	Position	Effect allele	Other allele	EAF	β	SE	P-value
rs10882183	*CYP2C9*	10	94974083	G	A	0.177	-0.165	0.039	2.90E-05
rs9922093	*ACSM2B*	16	20568843	T	C	0.963	-0.234	0.078	0.0026
rs58395451	*ACSM2B*	16	20557990	A	T	0.019	0.309	0.110	0.0049
rs62191775	*UGT1A6*	2	233745022	C	T	0.018	0.297	0.116	0.01
rs67095238	*BCHE*	3	165830829	T	A	0.080	-0.137	0.056	0.015
rs11604395	*PAFAH1B2*	11	117162272	A	G	0.028	-0.215	0.093	0.021
rs138780562	*UGT1A6*	2	233765484	ACAGT	A	0.967	0.177	0.084	0.036
rs74765243	*PAFAH1B2*	11	117147180	G	C	0.010	-0.295	0.149	0.049

Abbreviations: Chr, chromosome; EAF, effect allele frequency; SE, standard error.

6.3.3 Mendelian randomisation- salicylic acid and colorectal cancer-

6.3.3.1 GECCO

After identifying 8 independent SNPs associated with salicylic acid, we then extracted logistic regression results for these SNPs from the GECCO dataset, of which only 2 were available. These are rs9922093 and rs58395451, both found in *ACSM2B*.

To identify the effect of salicylic acid on CRC, we carried out maximum likelihood and inverse variance weighted approaches. Since only 2 SNPs were available, the other 6 methods could not be carried out as they require more than 2 SNPs.

For an association between exposure and outcome to be present, the MR P-value needs to be ≤0.05. Our results show that there is no evidence of a causal relationship between salicylic acid levels and CRC as the P-values for the maximum likelihood method and the inverse variance weighted method are 0.885 and 0.968, respectively (Table 6-7).

Table 6-7- Mendelian randomisation of salicylic acid on colorectal cancer using GECCO data.

Method	N SNP	β	SE	95% confidence intervals		P-value
				LCI	UCI	
Maximum likelihood	2	0.029	0.200	-0.364	0.421	0.885
Inverse variance weighted	2	0.016	0.400	-0.769	0.801	0.968

Abbreviations: SE, standard error; LCI, lower confidence interval; UCI, upper confidence interval.

This can be visualised in the scatter plots (Figure 6-9 A) where no clear evidence of a causal relationship between salicylic acid and CRC risk can be seen. The Two Sample MR package automatically re-assigns effect alleles and other alleles so that SNP associations with exposure are positive (Chapter 6.1.6.6). For example, with regards to the SNP rs9922093, whilst the T allele is associated with a -0.234 decrease in salicylic acid, the C allele is associated with an increase of 0.234 in salicylic acid levels. The Two Sample MR package then harmonises the alleles with the outcome data (GECCO) so that the same allele is assigned as the effect allele in exposure and outcome datasets. Therefore, the C allele in GECCO would also be assigned as the effect allele and flipped the association accordingly (e.g. positive to negative).

The forest plot of single SNP associations (Wald ratios) shows opposing effects of an increase in salicylic acid on risk of CRC. Therefore, the summarised IVW result has wide confidence intervals and overlaps with the null (Figure 6-9 B). Also, since there are very few SNPs, it is very difficult to detect pleiotropy using a funnel plot (Figure 6-9. C).

Overall, these SNPs that represent an increase in salicylic acid show no evidence of a causal relationship between the metabolite and CRC. However, it is difficult to detect pleiotropy since only 2 SNPs were used in the analysis therefore this MR needs to be repeated in other CRC datasets that have more than these 2 SNPs genotyped.

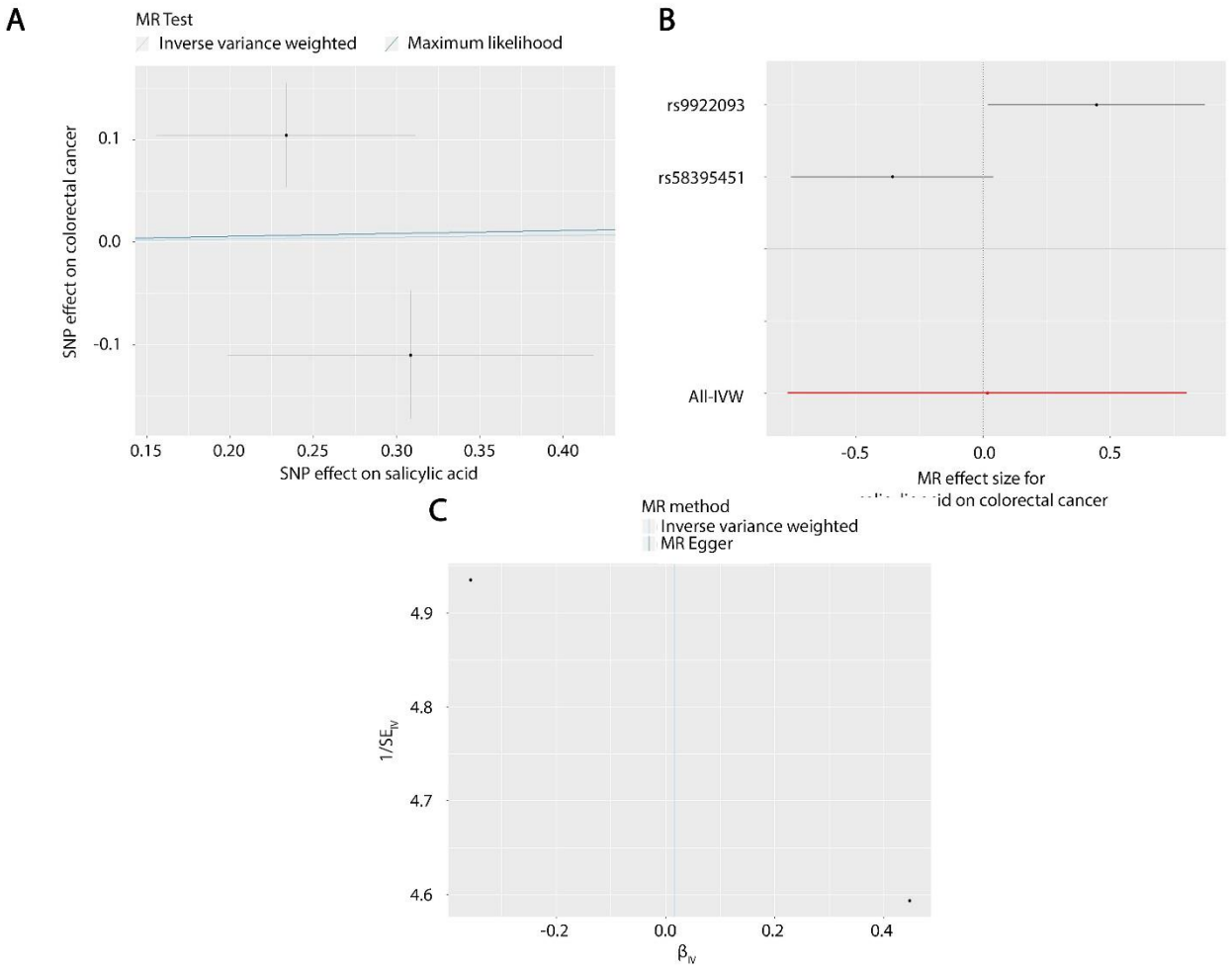


Figure 6-9- Mendelian randomisation of salicylic acid on colorectal cancer using data from GECCO

Only 2 of the 8 SNPs associated with salicylic acid levels were available from GECCO. Using these SNPs, we conducted a Mendelian randomisation of salicylic acid on CRC risk. (A) Scatter plot plotting the SNP effect on salicylic acid (x-axis) and SNP effect on risk of CRC (y-axis). The black dots represent each SNP and the lines either side parallel to the x-axis and y-axis represent the 95% confidence intervals for SNP effect on salicylic acid and SNP effect on cancer risk, respectively. Two methods were used to summarise the results: inverse variance weighted (light blue) and maximum likelihood (dark blue). The results show little evidence of a causal relationship between salicylic acid and CRC risk. (B) Forest plot of single SNP associations (Wald ratios) for the effect of salicylic acid on CRC. The overall effect of the SNPs using an IVW method is shown in red. The dotted line represents a null. The two SNPs are having opposing effects on risk of CRC. The overall effect is a null association as the confidence intervals overlap with the dotted line of null. (C) Funnel plot showing the causal estimate of each SNP on the x-axis plotted against the inverse variance of the estimate (light blue line) on the y-axis. Each black dot represents a SNP and the summary causal estimates using the IVW method. The MR Egger method could not be used due to an insufficient number of SNPs. It is difficult to determine whether there is pleiotropy due to the small number of SNPs involved in the analysis.

6.3.3.2 UK Biobank- all

Since we only had genetic information and association results with CRC for 2 of our SNPs, we sought alternative datasets that had genotyped more of our 8 SNPs of interest. To test the effect of the 8 SNPs associated with salicylic acid on risk of CRC, we used an alternative dataset with genetic information on CRC incidence through UK Biobank (449). After a linear mixed model approach adjusting for sex, genotyping array and the first 10 principal components was performed for the SNPs on risk of CRC, we extracted results for 6 of the 8 SNPs associated with salicylic acid.

To identify the effect of 1 standard deviation increase in salicylic acid (predicted by the genetic instruments) on CRC incidence, we used the 8 methods shown in Table 6-8. Unfortunately, the effect sizes (β) of salicylic acid on CRC were very small, the highest being 0.000901 which means that for every standard deviation increase in salicylic acid, the prevalence of CRC increases by 0.09% (beta: 0.000901, 95% CI: -0.00059-0.00239). However, all the methods produced null associations as the P-values were all above 0.05.

The results can be visualised in the scatter plot (Figure 6-10 A) which shows all the methods with various slopes of lines of best fit. For results to be convincing, ideally all of the regression lines should have similar directions and slopes as this would increase the reliability of the results (235,241).

Table 6-8- Mendelian randomisation of salicylic acid on colorectal cancer using UK Biobank data.

Method	N SNP	β	SE	95% confidence intervals		P-value
				LCI	UCI	
Inverse variance weighted	6	1.58E-04	3.47E-04	-5.23E-04	8.39E-04	0.649
Maximum likelihood	6	1.69E-04	3.60E-04	-5.36E-04	8.75E-04	0.638
MR Egger	6	-3.57E-04	1.16E-03	-2.62E-03	1.91E-03	0.773
MR Egger (bootstrap)	6	9.01E-04	7.61E-04	-5.90E-04	2.39E-03	0.098
Simple median	6	-1.60E-04	5.26E-04	-1.19E-03	8.70E-04	0.760
Simple mode	6	-4.83E-04	7.38E-04	-1.93E-03	9.62E-04	0.541
Weighted median	6	3.27E-04	4.63E-04	-5.81E-04	1.23E-03	0.481
Weighted mode	6	6.14E-04	5.64E-04	-4.92E-04	1.72E-03	0.326

Abbreviations: SE, standard error; LCI, lower confidence interval; UCI, upper confidence interval.

The forest plot of single SNP associations (Wald ratios) shows that all 6 SNPs have large confidence intervals and overlap with the null. When summarised using an IVW and MR Egger approach, the overall association between salicylic acid and CRC also overlaps with the null due to the very large confidence intervals (Figure 6-10 B).

The funnel plot shows asymmetry indicating that there may be pleiotropy (Figure 6-10 C). The leave-one-out sensitivity analysis shows that the SNP rs10882183 may be biasing the results, however, the large confidence intervals of the results make it difficult to be certain (Figure 6-10 D).

Overall, our results indicate that the SNPs we are using to proxy for salicylic acid may be pleiotropic and associated with other exposures. It also shows that there appears to be no significant association between

an increase in salicylic acid and an association with CRC incidence. We hypothesise that the true causal relationship of salicylic acid and CRC may not be clear in these results as salicylic acid obtained from the diet is very small in comparison to the amount that could be achieved through aspirin ingestion (Chapter 1.4.3). Therefore, this analysis needs to be repeated in CRC cases and controls stratifying between aspirin users and non-users as this may allow clearer visualisation of salicylic acid's effect on cancer risk.

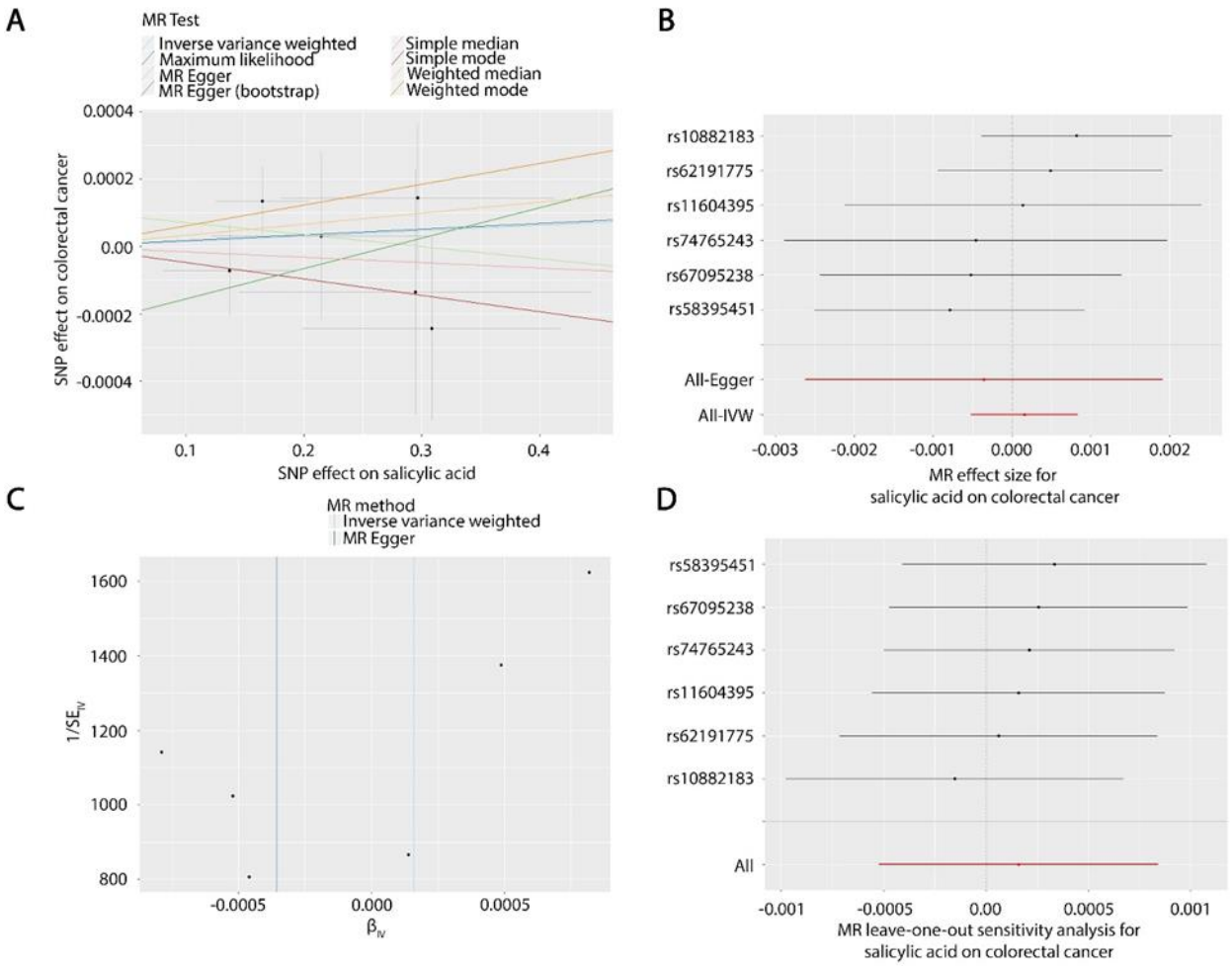


Figure 6-10- Mendelian randomisation of salicylic acid on colorectal cancer using data from UK Biobank

Six of the 8 SNPs associated with salicylic acid levels were available from UK Biobank. Using these SNPs, we conducted a Mendelian randomisation of salicylic acid on CRC risk. (A) Scatter plot plotting the SNP effect on salicylic acid (x-axis) and SNP effect on risk of CRC (y-axis). The black dots represent each SNP and the lines either side parallel to the x-axis and y-axis represent the 95% confidence intervals for SNP effect on salicylic acid and SNP effect on cancer risk, respectively. Eight methods in total were used to summarise the results: inverse variance weighted (light blue), maximum likelihood (dark blue), MR Egger (light green), MR Egger (bootstrap) (dark green), simple median (light pink), simple mode (dark pink), weighted median (light gold) and weighted mode (dark gold). The results show little evidence of a causal relationship between salicylic acid and CRC risk. (B) Forest plot of single SNP associations (Wald ratios) for the effect of salicylic acid on CRC. The overall effect of the SNPs using an IVW method and MR Egger approach is shown in red. The dotted line represents a null. All 6 SNPs have large confidence intervals and overlap with the null. Overall, the results show no causal relationship between salicylic acid and CRC risk as the confidence intervals overlap with the null. (C) Funnel plot showing the the causal estimate of each SNP on the x-axis plotted against the inverse variance of the estimate on the y-axis. Each black dot represents a SNP and the summary causal estimates using the IVW method (light blue line) and MR Egger (dark blue line). The plot indicates pleiotropy as there is no symmetry. (D) Leave-one-out analysis shows the summary IVW estimate when you remove each SNP systematically from the analysis. The SNP excluded is shown on the left side of the graph. The summary IVW estimate is shown in red. The results show that the SNP rs10882183 may be biasing the results, however, the large confidence intervals of the results make it difficult to be certain.

6.3.3.3 *UK Biobank- stratified between aspirin users and non-users*

To identify whether the effect of salicylic acid on CRC incidence is clearer in aspirin users, we stratified CRC cases and controls between aspirin users and non-users.

To identify the effect of a one standard deviation increase in salicylic acid levels (predicted by the genetic instruments) on risk of CRC incidence, we calculated summary statistics of the effect of salicylic acid on CRC risk in aspirin users (Table 6-9) and aspirin non-users (Table 6-10). Unfortunately, the effect sizes (β) of salicylic acid on CRC in aspirin users and non-users were very small regardless of the method used to summarise the results. However, all the methods returned results that showed a null causal relationship (P-value >0.05).

These can be visualised in the scatter plot for aspirin users (Figure 6-11 A) and non-users (Figure 6-11 B) which shows all the methods with various slopes of lines of best fit. For results to be convincing, ideally all of the regression lines should have similar directions and slopes as this would increase the reliability of the results (235,241).

Table 6-9- Mendelian randomisation of salicylic acid on colorectal cancer using UK Biobank data in aspirin users

Method	N SNP	β	SE	95% confidence intervals		P-value
				LCI	UCI	
Inverse variance weighted	6	0.00235	0.00378	-0.00504	0.00975	0.533
Maximum likelihood	6	0.00254	0.00392	-0.00514	0.01023	0.517
MR Egger	6	0.01293	0.01273	-0.01203	0.03789	0.367
MR Egger (bootstrap)	6	0.00257	0.00845	-0.01399	0.01912	0.378
Simple median	6	0.00488	0.00530	-0.00551	0.01527	0.357
Simple mode	6	0.00620	0.00638	-0.00630	0.01871	0.376
Weighted median	6	0.00469	0.00498	-0.00507	0.01446	0.346
Weighted mode	6	0.00577	0.00616	-0.00630	0.01784	0.392

Abbreviations: SE, standard error; LCI, lower confidence interval; UCI, upper confidence interval.

Table 6-10- Mendelian randomisation of salicylic acid on colorectal cancer using UK Biobank data in aspirin non-users

Method	N SNP	β	SE	95% confidence intervals		P-value
				LCI	UCI	
Inverse variance weighted	6	-0.00027	0.00053	-0.00131	0.000767	0.609
Maximum likelihood	6	-0.00028	0.00054	-0.00134	0.000779	0.605
MR Egger	6	0.000443	0.001763	-0.00301	0.0039	0.814
MR Egger (bootstrap)	6	-0.00092	0.001143	-0.00316	0.001322	0.191
Simple median	6	-0.00031	0.000692	-0.00166	0.00105	0.658
Simple mode	6	-0.00056	0.000822	-0.00217	0.001051	0.526
Weighted median	6	-0.00052	0.000716	-0.00192	0.000882	0.466
Weighted mode	6	-0.00057	0.000784	-0.00211	0.000967	0.501

Abbreviations: SE, standard error; LCI, lower confidence interval; UCI, upper confidence interval.

The forest plot of single SNP associations (Wald ratios) for aspirin users (Figure 6-11 C) and non-users (Figure 6-11 D) shows a null causal relationship for all 6 SNPs. The funnel plots for aspirin users (Figure 6-12 A) and non-users (Figure 6-12 B) both show asymmetry indicating there may be pleiotropy and this biases the IVW estimate. Also, the leave-one out analysis for aspirin users (Figure 6-12 C) and non-users (Figure 6-12 D) do not show any SNP clearly biasing the results.

Overall, stratifying between aspirin users and non-users showed no clear causal relationship between salicylic acid and risk of CRC. However, the results do indicate that the SNPs used to proxy for salicylic acid may be pleiotropic. Therefore, better instruments need to be found in order to better interrogate the causal relationship of salicylic acid with risk of CRC.

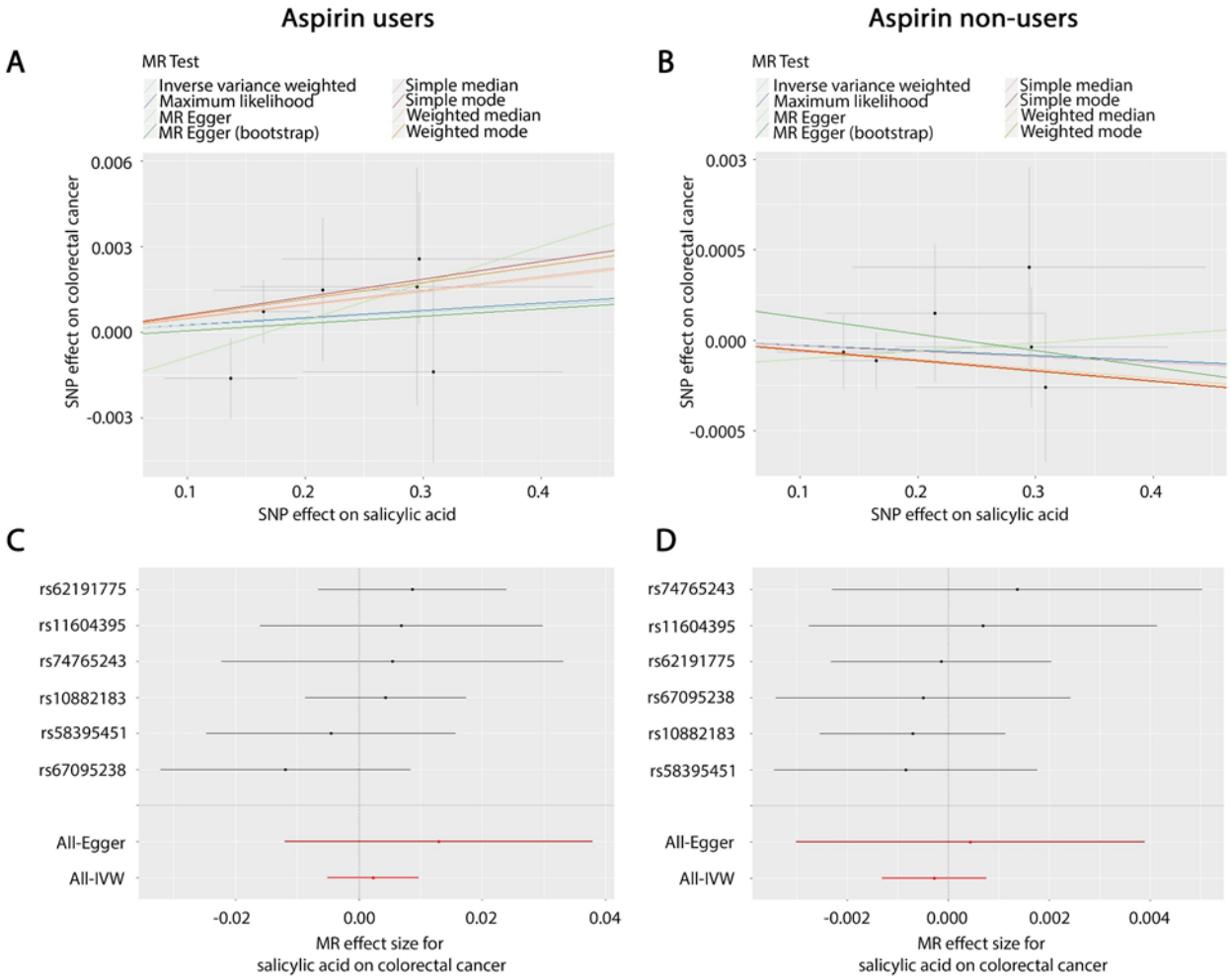


Figure 6-11- Scatter plots and forest plots comparing the effect of salicylic acid on risk of colorectal cancer between aspirin users and non-users

Six of the 8 SNPs associated with salicylic acid levels were available from UK Biobank. Using these SNPs, we conducted a Mendelian randomisation of salicylic acid on CRC risk stratifying between aspirin users and non-users. (A) The scatter graph shows the SNP effect on salicylic acid (x-axis) and SNP effect on risk of CRC (y-axis) in aspirin users. The black dots represent each SNP and the lines either side parallel to the x-axis and y-axis represent the 95% confidence intervals for SNP effect on salicylic acid and SNP effect on cancer risk, respectively. Eight methods in total were used to summarise the results: inverse variance weighted (light blue), maximum likelihood (dark blue), MR Egger (light green), MR Egger (bootstrap) (dark green), simple median (light pink), simple mode (dark pink), weighted median (light gold) and weighted mode (dark gold). The results show little evidence of a causal relationship between salicylic acid and CRC incidence in aspirin users. (B) Scatter plot showing the association between salicylic acid and CRC incidence in aspirin non-users. The results show little evidence of a causal relationship between salicylic acid and CRC risk in aspirin-non users (C) Forest plot of single SNP associations (Wald ratios) for the effect of salicylic acid on CRC in aspirin users. The overall effect of the SNPs using an IVW method and MR Egger approach is shown in red. The dotted line represents a null. All 6 SNPs have large confidence intervals and overlap with the null. Overall, the results show no causal relationship between salicylic acid and CRC risk in aspirin-users as the confidence intervals overlap with the null. (D) Forest plot of single SNP associations (Wald Ratios) for the effect of salicylic acid on CRC in aspirin non-users. Again, all 6 SNPs have large confidence intervals and overlap with the null. Overall, the results show no causal relationship between salicylic acid and CRC risk in aspirin non-users as the confidence intervals overlap with the null.

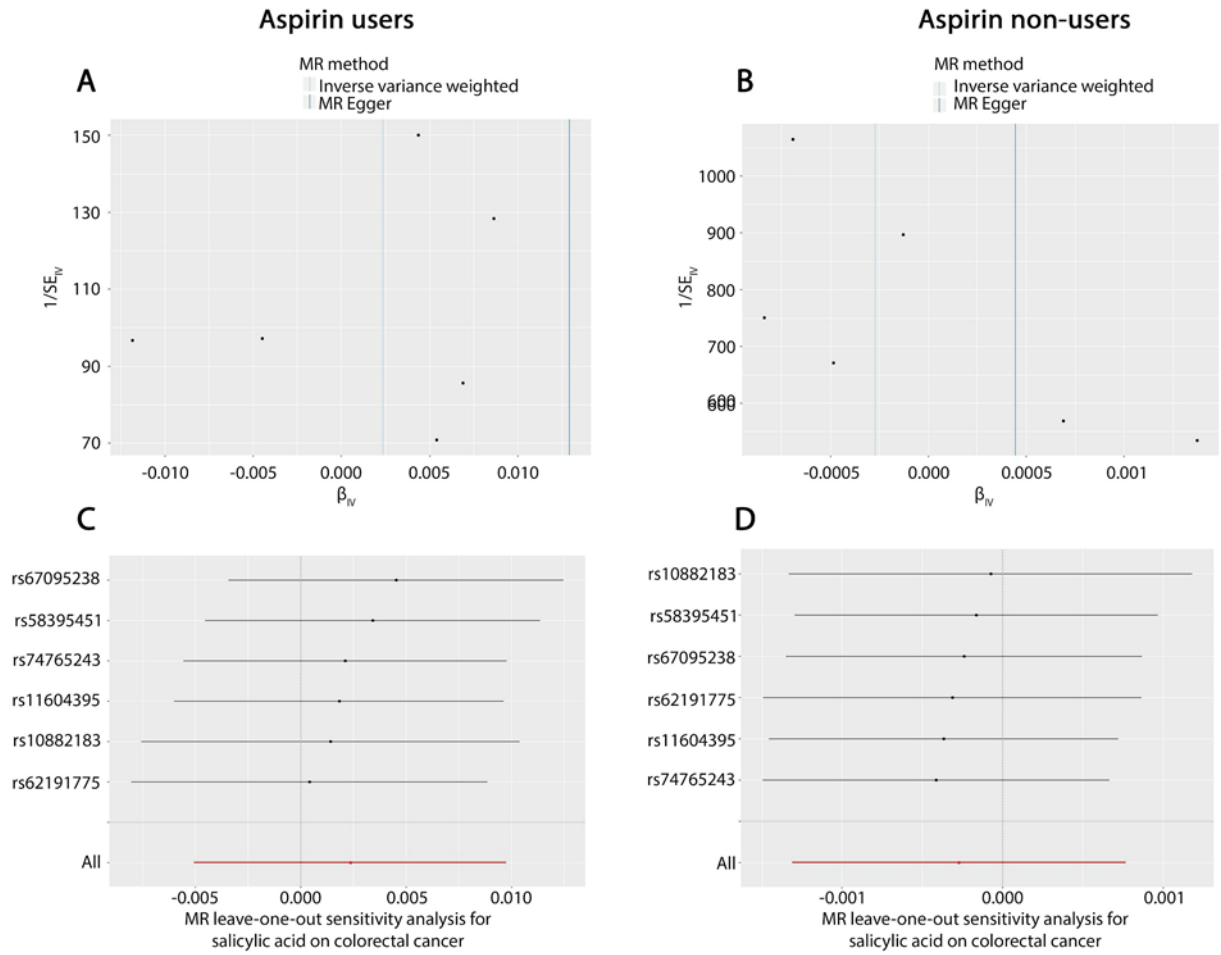


Figure 6-12- Funnel plots and leave-one-out analysis comparing the effect of salicylic acid on risk of colorectal cancer between aspirin users and non-users.

Six of the 8 SNPs associated with salicylic acid levels were available from UK Biobank. Using these SNPs, we conducted a Mendelian randomisation of salicylic acid on CRC risk stratifying between aspirin users and non-users. (A) Funnel plot showing the causal estimate of each SNP on the x-axis plotted against the inverse variance of the estimate on the y-axis. Each black dot represents a SNP and the summary causal estimates using the IVW method (light blue line) and MR Egger (dark blue line). The plot indicates that the SNPs used investigate the effect of salicylic acid on CRC risk in aspirin users are pleiotropic as there is no symmetry and there is a large difference between the IVW estimate and the MR Egger estimate. (B) Funnel plot showing that the SNPs proxying for salicylic acid levels on the risk of CRC for aspirin non-users may be pleiotropic due to the lack of symmetry and due to the difference in the IVW estimate and MR Egger estimate. (C) Leave-one-out analysis shows the summary IVW estimate when you remove each SNP systematically from the analysis. The SNP excluded is shown on the left side of the graph. The summary IVW estimate is shown in red. In aspirin users, the results show that there appears to be no SNP biasing the results with an obvious change. (D) Leave-one-out analysis of salicylic acid on CRC risk in aspirin non-users does not indicate that any of the 6 SNPs is biasing the results.

6.3.4 Salicyluric acid genetic instruments

To test for a causal relationship between the SNPs that proxy for salicyluric acid with CRC incidence, we used genotype and metabolite data provided by TwinsUK (3). Using the genomic locations of the enzymes involved in aspirin metabolism, a total of 4713 SNPs were available from the dataset. We excluded SNPS that were hemizygous or homozygous and ran a linear mixed model regression for the remaining SNPs on salicyluric acid adjusting for age and year of measurement, sex and visit. We then removed SNPs that were not associated with salicyluric acid at a P-value >0.05 therefore 88 SNPs remained for the analysis. We then clumped these at an r^2 of more than 0.001 to remove SNPs in high linkage disequilibrium (Figure 6-13). Therefore, we identified 9 independent SNPs that are associated with salicyluric acid at a P-value of ≤0.05. These are summarised in Table 6-11.

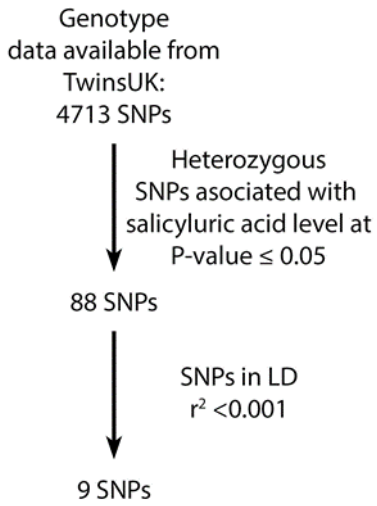


Figure 6-13- Flow diagram of the number of SNPs used in the analysis for salicyluric acid

Using the TwinsUK dataset, a total of 4713 SNPs were genotyped for the genomic locations of the enzymes from the aspirin metabolism pathway. Homozygous and hemizygous SNPs were excluded from the analysis. After running the linear mixed model regression, we excluded SNPs not associated with salicyluric acid at a P-value threshold of > 0.05. We then carried out LD clumping to remove SNPs in high linkage disequilibrium at an r^2 of more than 0.001. The result was 8 SNPs independently associated with salicyluric acid levels at a P-value of ≤0.05.

Table 6-11- SNPs used to proxy for salicyluric acid metabolite levels from TwinsUK

SNP	Gene	Chr	Position	Effect allele	Other allele	EAF	β	SE	P-value
rs17108282	*CYP2C9*	10	94971442	T	C	0.176004	-0.11082	0.038811	0.0043
rs13072970	*BCHE*	3	1.66E+08	T	A	0.390431	-0.0805	0.030409	0.0081
rs74765243	*PAFAH1B2*	11	1.17E+08	G	C	0.010412	-0.36397	0.146387	0.013
rs56354035	*ACSM2B*	16	20560037	C	T	0.231036	-0.08205	0.035307	0.02
rs116987336	*PAFAH1B2*	11	1.17E+08	A	G	0.025781	-0.21211	0.092755	0.022
rs62035044	*ACSM2B*	16	20541107	T	A	0.206743	-0.07662	0.036429	0.035
rs11604395	*PAFAH1B2*	11	1.17E+08	A	G	0.027516	-0.18568	0.090855	0.041
rs138622483	*UGT1A6*	2	2.34E+08	T	C	0.002231	-0.62891	0.31367	0.045
rs62191775	*UGT1A6*	2	2.34E+08	C	T	0.0176	0.225517	0.113362	0.047

Abbreviations: Chr, chromosome; EAF, effect allele frequency; SE, standard error

6.3.5 Mendelian randomisation- salicyluric acid and colorectal cancer

6.3.5.1 GECCO

After identifying 9 independent SNPs associated with salicyluric acid, we then extracted logistic regression results for these SNPs from the GECCO dataset, of which results for only 1 SNP was available. This was rs56354035, found in *ACSM2B*.

The Two Sample MR package automatically re-assigns effect alleles and other alleles so that SNP associations with exposures are positive (Chapter 6.1.6.6). Therefore, for every C allele change in rs56354035, this would predict a 0.08205 decrease in salicyluric acid (Table 6-11). On the other hand, for every T allele change, this would predict a 0.08205 increase in salicyluric acid. To understand the effect of an increase of salicyluric acid levels on CRC risk, the positively associated allele of the SNPs was used. The Two Sample MR package then harmonises the alleles with the outcome data (GECCO) so that the same allele is assigned as the effect allele in exposure and outcome datasets.

To identify the effect of salicyluric acid on risk of CRC, we calculated the Wald ratio which showed that an increase in salicyluric acid increases the risk of CRC. Therefore, for every T allele change resulting in a 0.08205 increase in levels of salicyluric acid, the log OR of CRC increases by 0.0177, giving a Wald ratio of 0.216 (Table 6-12). However, the P-value of the result for the SNP rs56354035 is >0.05 (P-value: 0.424) therefore there is no causal relationship between salicyluric acid and CRC.

if the A allele in SNP predicts a 0.2 increase in salicylic acid and a log OR of CRC as -0.6, we would then deduce that for every A allele change in the SNP, the Wald ratio is- 0.3. Therefore, for every A allele, the level of salicylic acid increases and the risk of CRC decreases.

Table 6-12- Wald ratio results for salicyluric acid levels and risk of colorectal cancer

SNP	β	SE	P-value
rs56354035	0.216	0.270	0.424

Abbreviations: SE, standard error.

Since only 1 SNP was genotyped in the GECCO consortium, this analysis needs to be repeated in an alternative dataset with genetic and outcome risk data for more SNPs

6.3.5.2 UK Biobank- all

Since we only had genetic information and association results with CRC for 1 of our SNPs, we sought alternative datasets that had genotyped more of our 9 SNPs of interest. We used UK Biobank as an alternative consortium (449) and after a linear mixed model approach adjusting for sex, genotyping array and the first 10 principal components was performed for the SNPs on risk of CRC, we extracted results for 7 of the 9 SNPs associated with salicyluric acid.

To identify the effect of 1 standard deviation increase in salicyluric acid levels (predicted by the genetic instruments) on CRC incidence, we used the 8 methods shown in Table 6-13. Unfortunately, the effect sizes (β) of salicyluric acid on CRC were very small, the highest being 0.000813 which means that our 7 SNPs that proxy for an increase in salicyluric acid levels increase the prevalence of CRC increases by 0.0813% by a beta value of (beta:0.000813, 95% CI: -0.000028 -0.001654) using the maximum likelihood method. Whilst none of the methods achieved a P-value significance of ≤0.05, the IVW method and the maximum likelihood method were only just higher than this threshold, P-value =0.052 and P-value =0.058, respectively. The results of both of these methods indicate that an increase in salicyluric acid levels may potentially be associated with an increase in risk of CRC and warrant further analysis. Since a linear mixed model approach was carried out using UK Biobank data, the results are interpreted as the change in prevalence per standard deviation increase in the metabolite. For example, with regards to the IVW estimate, this effect says that for 1 standard deviation increase in salicyluric acid, the prevalence of CRC increased by 0.0759%.

Table 6-13- Mendelian randomisation of salicyluric acid on colorectal cancer using UK Biobank data

Method	N SNP	β	SE	95% confidence intervals		P-value
				LCI	UCI	
Inverse variance weighted	7	0.000759	0.000391	-0.000007	0.001524	0.052
Maximum likelihood	7	0.000813	0.000429	-0.000028	0.001654	0.058
MR Egger	7	0.00028	0.000883	-0.001450	0.002010	0.764
MR Egger (bootstrap)	7	0.000826	0.000849	-0.000838	0.002489	0.148
Simple median	7	0.000787	0.000596	-0.000382	0.001956	0.187
Simple mode	7	0.000807	0.000692	-0.000550	0.002165	0.288
Weighted median	7	0.000777	0.000556	-0.000313	0.001867	0.163
Weighted mode	7	0.000835	0.000745	-0.000626	0.002295	0.306

Abbreviations: SE, standard error; LCI, lower confidence interval; UCI, upper confidence interval.

These results can be visualised in Figure 6-14. The scatter plot shows that all 8 methods consistently show a causal relationship between increasing levels of salicyluric acid and increasing levels of risk of CRC (Figure 6-14 A).

The forest plots of single SNP associations (Wald ratios) shows that all 7 SNPs have large confidence intervals and overlap with the null. When summarised using an IVW and MR Egger approach, the overall association between salicyluric acid and CRC also overlaps with the null due to the very large confidence intervals (Figure 6-14 B). Although the lower confidence interval of the IVW estimate is only just below the null (-0.000007), it does indicate a possible causal relationship between salicyluric acid increasing the risk of CRC.

The funnel plot shows asymmetry indicating that there may be pleiotropy (Figure 6-14 C). The leave-one-out sensitivity analysis does not show any SNP clearly biasing the result (Figure 6-14 D).

Overall, our results indicate that some of the SNPs in the analysis used to proxy for salicyluric acid may be pleiotropic and associated with other exposures. The results of the IVW method and the maximum likelihood method indicate that an increase in salicyluric acid levels may potentially be associated with an increase in risk of CRC and warrant further analysis.

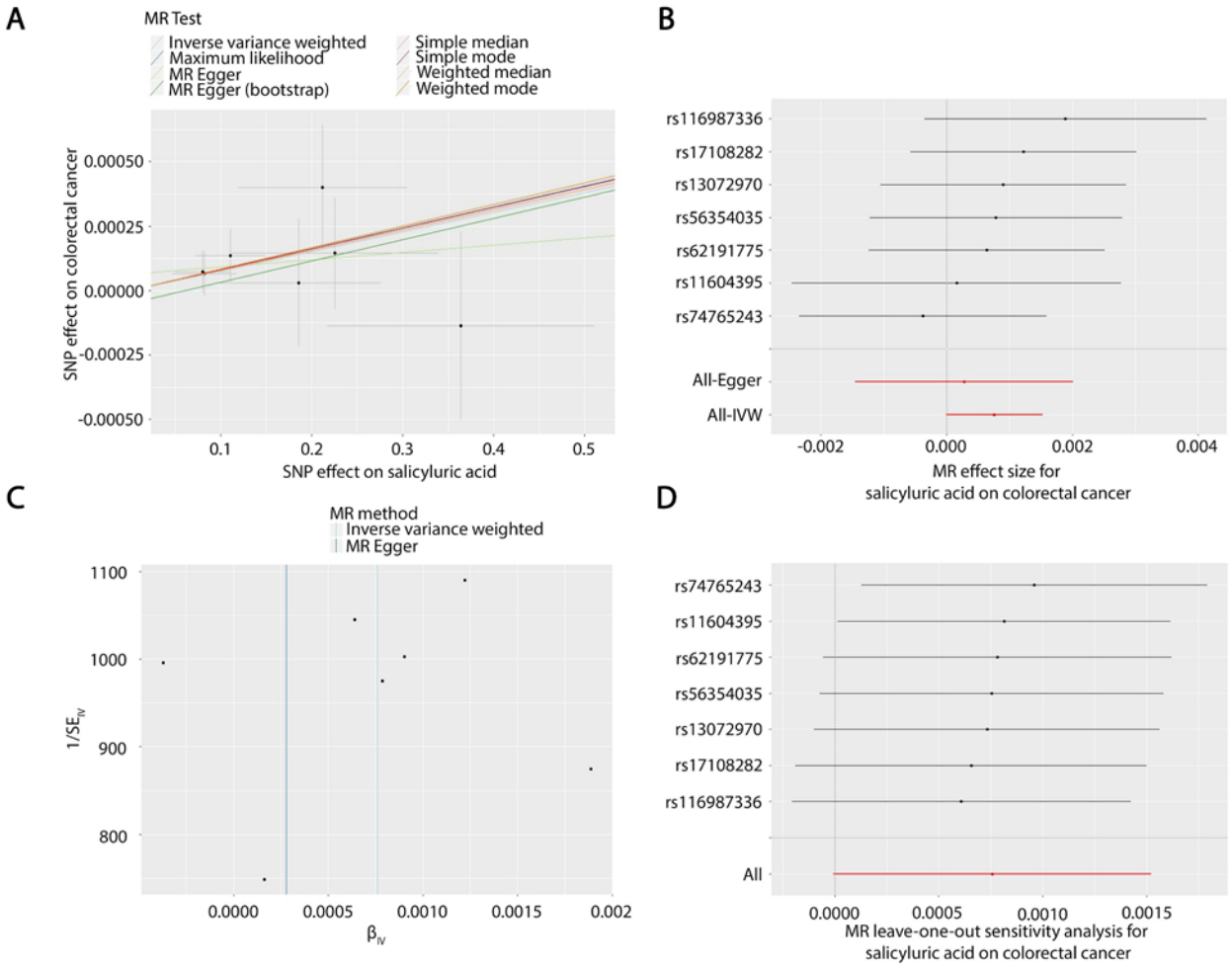


Figure 6-14- Mendelian randomisation of salicyluric acid on colorectal cancer using data from UK Biobank

Seven of the 9 SNPs associated with salicyluric acid levels were available from UK Biobank. Using these SNPs, we conducted a Mendelian randomisation of salicyluric acid on CRC risk. (A) Scatter plot plotting the SNP effect on salicyluric acid (x-axis) and SNP effect on risk of CRC (y-axis). The black dots represent each SNP and the lines either side parallel to the x-axis and y-axis represent the 95% confidence intervals for SNP effect on salicyluric acid and SNP effect on cancer risk, respectively. Eight methods in total were used to summarise the results: inverse variance weighted (light blue), maximum likelihood (dark blue), MR Egger (light green), MR Egger (bootstrap) (dark green), simple median (light pink), simple mode (dark pink), weighted median (light gold) and weighted mode (dark gold). One SNP has been circled in red as it may be biasing the results. Overall, the results show that increased levels of salicyluric acid may increase risk of CRC (B) Forest plot of single SNP associations (Wald Ratios) for the effect of salicyluric acid on CRC. The overall effect of the SNPs using an IVW method and MR Egger approach is shown in red. The dotted line represents a null. All 7 SNPs have large confidence intervals and overlap with the null. Overall, the IVW estimate shows that increased salicyluric acid may be increasing the risk of CRC, however, the estimate still overlaps with the null. The MR Egger test also shows a null association. (C) Funnel plot showing the causal estimate of each SNP on the x-axis plotted against the inverse variance of the estimate on the y-axis. Each black dot represents a SNP and the summary causal estimates using the IVW method (light blue line) and MR Egger (dark blue line). The plot indicates pleiotropy as there is no symmetry and the IVE estimate may be biased towards the pleiotropic SNPs. (D) Leave-one-out analysis shows the summary IVW estimate when you remove each SNP systematically from the analysis. The SNP excluded is shown on the left side of the graph. The summary IVW estimate is shown in red.

6.3.5.3 UK Biobank -Stratifying between aspirin users and non-users

To identify whether the effect of salicyluric acid on CRC incidence is clearer in aspirin users, we stratified CRC cases and controls between aspirin users and non-users.

To identify the effect of 1 standard deviation increase in salicyluric acid levels (predicted by the genetic instruments) on CRC incidence, we calculated summary statistics of the effect of salicyluric acid on CRC risk in aspirin users (Table 6-14) and aspirin non-users (Table 6-15). Unfortunately, the effect sizes (β) of salicyluric acid on CRC in aspirin users and non-users were very small regardless of the method used to summarise the results. However, none of the methods produced results showing a causal relationship between the metabolite and CRC as the P-values were above 0.05. These can be visualised in the scatter plot for aspirin users (Figure 6-15 A) and non-users (Figure 6-15 B) which shows all the methods with various slopes of lines of best fit. For results to be convincing, ideally all of the regression lines should have similar directions and slopes as this would increase the reliability of the results (235,241).

Table 6-14- Mendelian randomisation of salicyluric acid on colorectal cancer using UK Biobank data in aspirin users

Method	N SNP	β	SE	95% confidence intervals		P-value
				LCI	UCI	
Inverse variance weighted	7	0.0014	0.0045	-0.0074	0.0102	0.75
Maximum likelihood	7	0.0017	0.0046	-0.0073	0.0107	0.71
MR Egger	7	0.0055	0.0111	-0.0162	0.0272	0.64
MR Egger (bootstrap)	7	-0.0011	0.0112	-0.0231	0.0209	0.45
Simple median	7	0.0065	0.0065	-0.0062	0.0192	0.32
Simple mode	7	0.0078	0.0084	-0.0086	0.0242	0.39
Weighted median	7	0.0066	0.0059	-0.0050	0.0182	0.27
Weighted mode	7	0.0075	0.0074	-0.0071	0.0220	0.36

Abbreviations: SE, standard error; LCI, lower confidence interval; UCI, upper confidence interval.

Table 6-15- Mendelian randomisation of salicyluric acid on colorectal cancer using UK Biobank data in aspirin non-users

Method	N SNP	β	SE	95% confidence intervals		P-value
				LCI	UCI	
Inverse variance weighted	7	0.00023	0.00060	-0.00093	0.00140	0.70
Maximum likelihood	7	0.00024	0.00061	-0.00095	0.00143	0.69
MR Egger	7	0.00078	0.00134	-0.00186	0.00341	0.59
MR Egger (bootstrap)	7	-0.00033	0.00134	-0.00296	0.00230	0.39
Simple median	7	0.00037	0.00081	-0.00121	0.00195	0.65
Simple mode	7	0.00066	0.00111	-0.00152	0.00284	0.57
Weighted median	7	0.00034	0.00080	-0.00123	0.00191	0.67
Weighted mode	7	0.00052	0.00103	-0.00150	0.00253	0.63

Abbreviations: SE, standard error; LCI, lower confidence interval; UCI, upper confidence interval.

As previously seen with salicylic acid, Wald ratios of salicyluric acid with CRC of all 7 SNPs overlaps with the null for aspirin users (Figure 6-15 C) and non-users (Figure 6-15 D) and the funnel plots indicate pleiotropy due to the asymmetry present for aspirin users (Figure 6-16 A) and non-users (Figure 6-16 B). The leave-one out analysis for aspirin users (Figure 6-16C) shows that 2 SNPs (rs13072970 and rs116987336) may be biasing the results but no SNPs appear to bias the results for non-users (Figure 6-16 D).

Overall, stratifying between aspirin users and non-users showed no clear causal relationship between salicyluric acid and risk of CRC. However, the results do indicate that the SNPs used to proxy for salicyluric acid may be pleiotropic. Therefore, better instruments need to be found in order to better interrogate the causal relationship of salicyluric acid with risk of CRC.

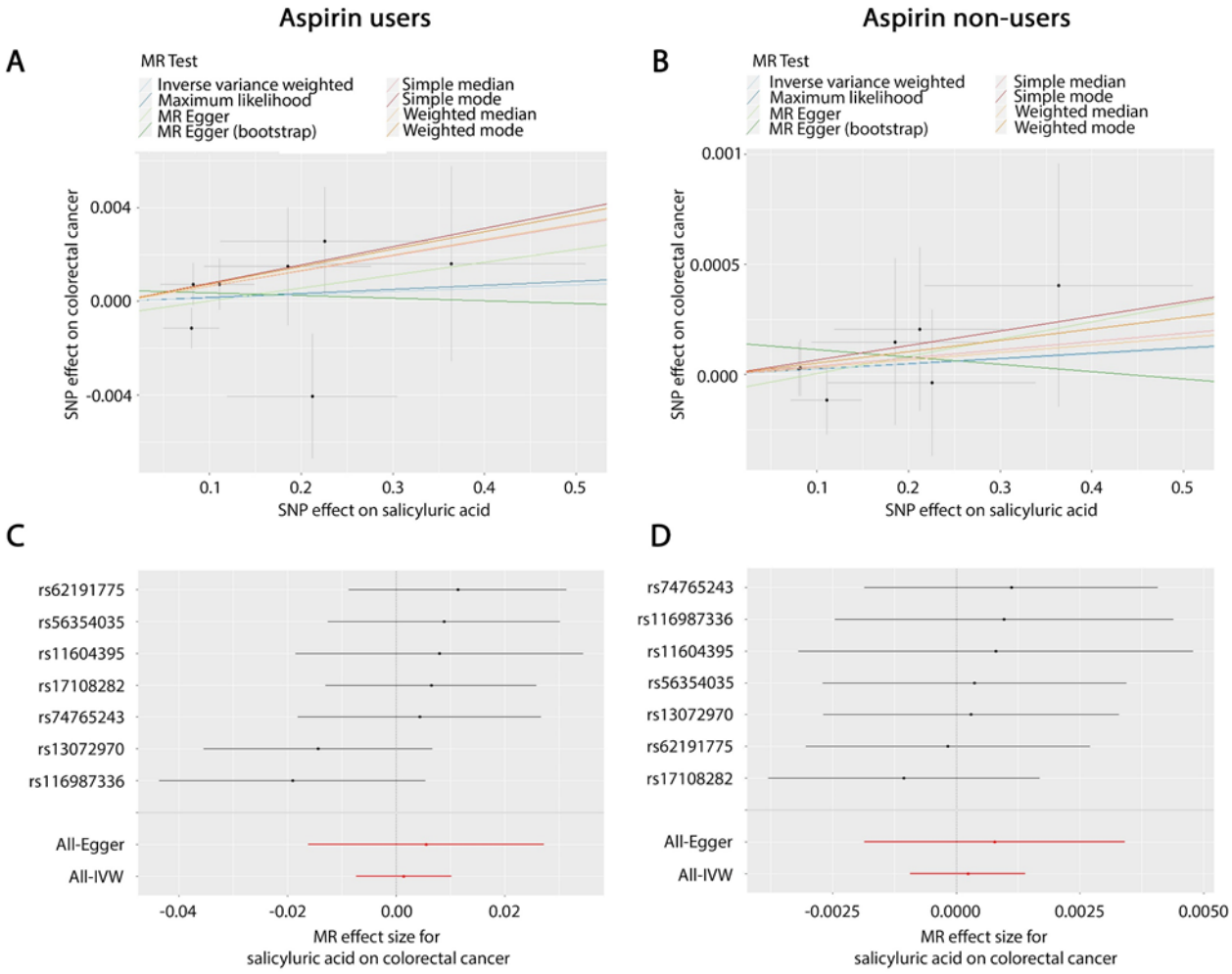


Figure 6-15- Scatter plots and forest plots comparing the effect of salicyluric acid on risk of colorectal cancer between aspirin users and non-users

Seven of the 9 SNPs associated with salicyluric acid levels were available from UK Biobank. Using these SNPs, we conducted a Mendelian randomisation of salicyluric acid on CRC risk stratifying between aspirin users and non-users. (A) The scatter graph shows the SNP effect on salicyluric acid (x-axis) and SNP effect on risk of CRC (y-axis) in aspirin users. The black dots represent each SNP and the lines either side parallel to the x-axis and y-axis represent the 95% confidence intervals for SNP effect on salicyluric acid and SNP effect on cancer risk, respectively. Eight methods in total were used to summarise the results: inverse variance weighted (light blue), maximum likelihood (dark blue), MR Egger (light green), MR Egger (bootstrap) (dark green), simple median (light pink), simple mode (dark pink), weighted median (light gold) and weighted mode (dark gold). The results show little evidence of a causal relationship between salicyluric acid and CRC incidence in aspirin users. (B) Scatter plot showing the causal relationship between salicyluric acid and CRC incidence in aspirin non-users. The results show little evidence of a causal relationship between salicyluric acid and CRC risk in aspirin-non users (C) Forest plot of single SNP associations (Wald Ratios) for the effect of salicyluric acid on CRC in aspirin users. The overall effect of the SNPs using an IVW method and MR Egger approach is shown in red. The dotted line represents a null. All 7 SNPs have large confidence intervals and overlap with the null. Overall, the results show no causal relationship between salicyluric acid and CRC risk in aspirin-users as the confidence intervals overlap with the null. (D) Forest plot of single SNP associations (Wald Ratios) for the effect of salicyluric acid on CRC in aspirin non-users. Again, all 7 SNPs have large confidence intervals and overlap with the null. Overall, the results show no causal relationship between salicyluric acid and CRC risk in aspirin non-users as the confidence intervals overlap with the null.

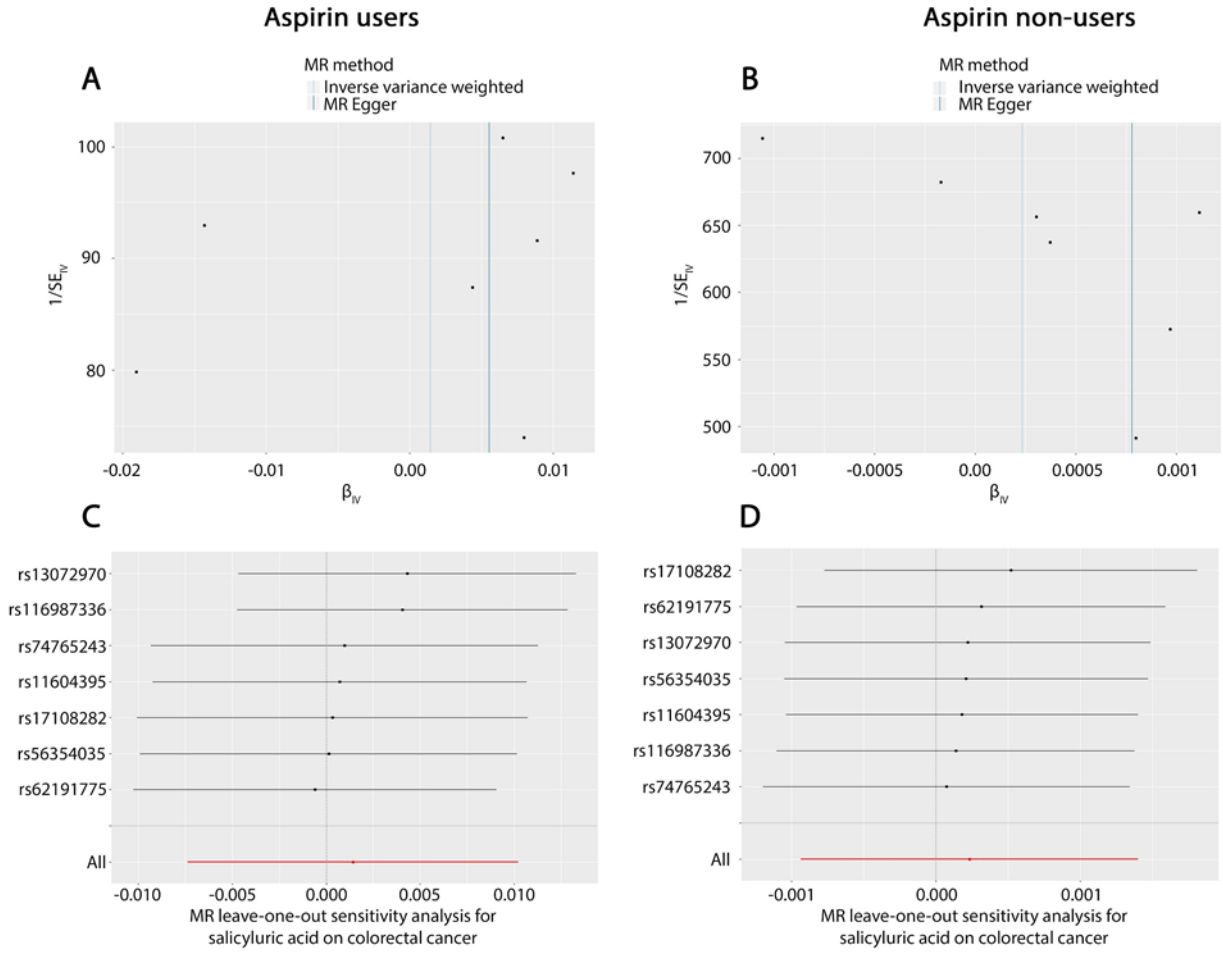


Figure 6-16- Funnel plots and leave-one-out analysis comparing the effect of salicyluric acid on risk of colorectal cancer between aspirin users and non-users.

Seven of the 9 SNPs associated with salicyluric acid levels were available from UK Biobank. Using these SNPs, we conducted a Mendelian randomisation of salicyluric acid on CRC risk stratifying between aspirin users and non-users. (A) Funnel plot showing the causal estimate of each SNP on the x-axis plotted against the inverse variance of the estimate on the y-axis. Each black dot represents a SNP and the summary causal estimates using the IVW method (light blue line) and MR Egger (dark blue line). The plot indicates that the SNPs used investigate the effect of salicyluric acid on CRC risk in aspirin users are pleiotropic as there is no symmetry and there is a large difference between the IVW estimate and the MR Egger estimate. (B) Funnel plot showing that the SNPs proxying for salicyluric acid levels on the risk of CRC for aspirin non-users may be pleiotropic due to the lack of symmetry and due to the difference in the IVW estimate and MR Egger estimate. (C) Leave-one-out analysis shows the summary IVW estimate when you remove each SNP systematically from the analysis. The SNP excluded is shown on the left side of the graph. The summary IVW estimate is shown in red. In aspirin users, the results show that 2 SNPs may be possibly biasing the results: rs13072970 and rs116987336. However, the effect of this is unclear as the summary estimate after removing each SNP still overlaps with the null. (D) Leave-one-out analysis of salicyluric acid on CRC risk in aspirin non-users does not indicate that any of the 6 SNPs is biasing the results.

6.4 Discussion

Overall, we have identified 8 SNPs independently associated with salicylic acid and 9 SNPs independently associated with salicyluric acid at a P-value ≤0.05. These SNPs were used to test for a causal relationship between these metabolites and CRC risk in both GECCO and UK Biobank, as well as stratifying between aspirin users and non-users using UK Biobank.

Using the SNPs that proxy for salicylic acid levels, we found no evidence of a causal relationship between the metabolite and risk of CRC incidence in both GECCO and UK Biobank. However, only 2 of the SNPs' data was available in GECCO therefore this limited the strength of our genetic risk score (the summarised effect of all the SNPs combined). In addition, whilst we identified 8 SNPs associated with salicylic acid at a P-value of ≤0.05, these SNPs did not survive the P-value threshold for multiple testing.

One of the issues with multiple testing is that unless the result is adjusted for, then the probability of rejecting a null hypothesis (so accepting that a SNP is associated with a metabolite) when actually there is no association is inflated (452). The Bonferroni correction divides the alpha level by the number of statistical tests being carried out- whilst it is the most conventional and simple method, it tends to be overly conservative (272). In the case of salicylic acid, this would be calculated as follows: 0.05/4713 (Figure 6-8) =0.000011. Since none of the SNPs were associated with salicylic acid below a P-value of 0.000011, it may be that the genetic instruments used to proxy for the metabolite were weak instruments and therefore to truly test for a causal relationship between the metabolite and CRC risk, more robust instruments need to be found.

Using the SNPs that proxy for salicyluric acid, we found a possible causal relationship between the metabolite and risk of CRC using the UK Biobank dataset (IVW beta: 0.000759, 95% CI: -0.000007- 0.001524, P-value=0.052). Whilst a causal relationship between salicyluric acid and CRC risk was found, the beta estimate was very small (0.000759) indicating the effect of increasing salicyluric acid levels on CRC risk may be very small. Since a linear mixed model approach was carried out using UK Biobank data, the results are interpreted as the change in prevalence per standard deviation increase in the metabolite. The IVW estimate predicts that for every 1 standard deviation increase in salicyluric acid, the prevalence of CRC increases by 0.0759%. If absolute values of the metabolites were measured, it may be possible to multiply the allele 'effect size' to recapitulate the trial dose of aspirin as was previously done with selenium and risk of prostate cancer (229). However, in this case, the results from the MR analysis can be interpreted as the change in log OR or prevalence per standard deviation increase in the inverse ranked measurements of the metabolites. We hypothesise that the SNPs proxying for salicyluric acid levels may cause a more rapid hydrolysis of salicylic acid (the active metabolite (220)) therefore increasing levels of salicyluric acid. Therefore, it may not be salicyluric acid itself that is increasing the risk of CRC but rather it is the decreased time exposed to the protective effect of salicylic acid. Experimental studies would be required to explore this hypothesis further.

Our genetic variants explained approximately 0.25% of the variation in salicylic acid levels and 0.25% of the variation in salicyluric acid levels. Our power calculations show that we had 80% power to detect an association of salicylic acid on CRC risk of an odds ratio ≤0.51 and ≥1.9. in the reciprocal direction and 80% power to detect an association of salicyluric acid on CRC risk of an odds ratio ≤0.51 and ≥1.9. With the UK Biobank sample size of 462933 (of which 641 were cases), we had 80% power at α level 0.05 to detect an effect of salicylic acid on CRC risk of an odds ratio ≥3.2 only. Unfortunately, the genetic instruments explain a small variation of the metabolites, therefore large sample sizes would be needed to detect the effect of this variation on CRC risk. This indicates the need to find genetic instruments that explain more of the variation or to identify studies with a larger number of cases and controls and carry out the MR using this data.

Upon stratifying between aspirin users and non-users, we found no causal relationship between the metabolites salicylic acid and salicyluric acid and their association with CRC incidence. Whilst we anticipated a null causal relationship between the metabolites and CRC in aspirin non-users, we postulated that there would be a causal relationship between the metabolites and cancer risk in aspirin-users. However, the number of CRC cases who documented taking aspirin in the data was only 7, indicating that out study was severely underpowered. This means that the study lacked a large enough sample size to detect a true causal relationship therefore accepting the null hypothesis (type II error). It would also explain the wide confidence intervals observed in the results (242).

The difference in magnitude of causal effect estimates using GECCO or UK Biobank may be explained by the methods used to identify an association between the SNPs and incidence of CRC. GECCO carried out logistic regression analyses between the SNPs and CRC incidence. On the other hand, a linear mixed model approach was carried out using the UK Biobank data. These results are interpreted as the change in prevalence of the outcome per copy of the effect allele or unit change in exposure. This means that e.g. if beta is 0.01, this is the proportionate change in prevalence so equates to a 1% increase in prevalence. This explains why the causal effect estimates calculated using UK Biobank data appear 2 orders of magnitude smaller than using GECCO summary effect estimates.

Associations between variants in enzymes in the aspirin metabolism pathway have previously been shown to be associated with aspirin hydrolysis. Zhou et. al (2013) had shown that a SNP in *BCHE* (rs6445035) was associated with the rate of hydrolysis of aspirin. For every A allele in this SNP, aspirin hydrolysis was decreased by around 1.2 nmol/ml/min (217) Unfortunately, this SNP was not within the gene region of BCHE (Table 6-3) so was not used in our analysis. This highlights a limitation in the study design. It may be that SNPs surrounding the genes may be affecting gene expression and hence activity or transcription factor binding such as *cis*-regulatory elements (453) and long non-coding RNAs (lncRNAs) which can affect gene expression (454) thus impacting upon the metabolism of the very metabolites being investigated. Despite an association between this SNP and aspirin hydrolysis having been found, the effect of this SNP on CRC risk has yet to be tested.

Genetic variation within other enzymes have also been identified to be associated with aspirin usage. Sheth et. al (2018) identified SNPs within *UGT1A6* gene that was associated with aspirin usage (rs2070959 and rs1105879). When they stratified between aspirin users and non-users, they showed a decreased risk of CRC for aspirin users with the wildtype SNP rs2070959 (455). They also showed that a variant in *CYP2C9* (rs1799853) which codes for a variant of the enzyme with reduced activity and metabolic activity (456) reduced the risk of colon cancer, however their result did not reach statistical significance once adjusted for multiple testing (455). These SNPs were not analysed in our approach due to them being outside the gene co-ordinates that were determined for the enzymes.

Whilst these studies' analyses point to an association between the variants and aspirin usage and therefore risk of CRC, not all studies have shown this association between *CYP2C9* and *UGT1A6* genotypes and response to aspirin treatment. Barry et. al (2013) initially showed an association between *CYP2C9* variants and risk of adenoma recurrence, however, this association was no longer present once patients were stratified according to aspirin use (457). Two variants within *UGT1A6* (rs1105879 (R184S) and rs2070959 (T181A)) have previously been predicted to reduce the enzyme's activity by 30-50% (458). These variants were tested for association with risk of CRC by Thompson et. al (2009). No association was found and the null association remained even after stratifying between aspirin users and non-users (459).

The opposing results from these studies highlight the importance of identifying better genetic instruments to proxy for aspirin metabolite levels and address the question of aspirin usage and risk of CRC. In order to find more robust genetic instruments, a genome wide association study of salicylic acid and salicyluric acid is needed. Any SNPs that survive correction for multiple testing can then be used as proxies for the metabolites. It is worth noting that the idea of canalisation may exist when studying biochemical pathways whereby alternative metabolic routes can lead to the same phenotypic endpoint if one enzyme is compromised from the original pathway (199).

Also, the enzymes involved in the metabolism of aspirin are pleiotropic enzymes, in that they are also involved in metabolising other drugs and intermediates in the aspirin metabolism pathway. For example, other substrates of CYP2C9 include: ibuprofen, warfarin, tamoxifen, fluvastatin and irbesartan (460). Therefore, it is plausible to think that SNPs in these enzymes are associated with metabolism of these drugs, as evidenced by GWAS's carried out for warfarin (461) and clopidogrel (462). This means that it may not be appropriate to use SNPs in these enzymes to proxy for aspirin metabolites as their association with other exposures violates one of the assumptions of Mendelian randomisation in that the SNP should only be associated with the outcome through the exposure under investigation (242). In this case, being able to stratify by aspirin use and having data on use of other drugs would help as individuals who take only aspirin and no other drugs could be used in the analysis to partially control for SNP associations with other drugs.

To our knowledge, this is the first MR study exploring the effect of aspirin metabolites on the risk of CRC. Few studies have applied the basic principles of MR to study a biochemical pathway. We propose that due to the increased risk of pleiotropy with regards to drug hydrolysis, the use of multiple SNPs to proxy for the metabolites is strongly recommended as this is the best way to identify and adjust for pleiotropy. Furthermore, a detailed understanding of the pathway is crucial to avoid incorrect assumptions about roles of the enzymes involved and to better interpret the results.

In order for MR to be applied to biochemical pathways, there is a need for GWASs of metabolic intermediates in general. We were limited in this study in that only salicylic acid and salicyluric acid were measured but not the other metabolites (Figure 1-20). However, the challenge in measuring metabolic intermediates is that these metabolites are dynamic and change with time, disease state and many other factors, so interpreting the results from a GWAS may be difficult. We were also limited in interpreting the results as the metabolites used in this study were measured through gas chromatography-mass spectrometry and liquid chromatography-mass spectrometry through an untargeted approach. This means that the values used in the analysis are relative changes rather than absolute values (267). In order to better interpret the results and transfer them to the clinic for patient advice on dosage intake, a more targeted approach to measuring the metabolite levels is needed.

The advantage to including the other metabolites in our MR approach is that by limiting our analysis to a single linear pathway, results from this may be misleading due to the presence of horizontal pleiotropy (such as enzymes being involved in other pathways) (463) which was observed in our results. Using the other pathways, a multivariable MR approach can be conducted whereby MR is conducted on all the pathways involved in salicylic acid hydrolysis into the other metabolites as well. This may help to overcome the issue of horizontal pleiotropy as the results involve combining multiple pathways related to salicylic acid metabolism. Ideally, this can only be done if there are population measures of the metabolites gentisic acid, salicylacyl glucuronide and salicyl phenolic glucuronide (Figure 1-20). To our knowledge, these metabolites have not yet been measured in large studies, therefore limiting this approach.

Our study was also limited in that whilst no data was available with regards to aspirin use, these metabolite measurements are likely to represent dietary salicylates rather than aspirin taken for medicinal purposes. The amount of salicylate obtained from the diet is much less than would be obtained for aspirin intake for medicinal purposes (Chapter 1.4.3). Another limitation of our study is that we tested for a causal relationship between aspirin metabolites and incidence of CRC but did not test for a causal relationship with tumour progression or metastasis. In order to fully investigate aspirin's protective effect, then association tests between salicylic acid and salicyluric acid associations with tumour progression and metastasis need to be carried out.

Overall, our study has shown a possible causal relationship between salicyluric acid and increasing risk of CRC, although our study was severely underpowered to detect this and a causal relationship between salicylic acid and salicyluric acid. What is evident is that the instruments used in our MR approach may be weak highlighting the need to repeat this analysis after identifying more robust genetic instruments to proxy for the metabolites, through adopting a GWAS approach. In order to identify a clearer effect of aspirin intake on risk of CRC, a larger cancer consortium containing both aspirin usage data that can be stratified between users and non-users is required, and this will increase power to identify any potential association.

Chapter 7 General Discussion

To address the effect of aspirin on colorectal cancer (CRC), population and laboratory-based approaches were combined to expand on the current knowledge in the literature. This chapter summarises the main research findings and the possible implications of the results as well as future directions.

Results presented in this thesis indicate that short-term and long-term aspirin treatment have different effects on cell proliferation and survival (Chapter 3). By combining multiple 'omics, we found that short-term aspirin treatment may target *STMN1* expression as a possible new target of aspirin (Chapter 4) but no effect on genome-wide methylation. Results from long-term aspirin treatment showed a genome-wide effect on DNA methylation (Chapter 5). To address the effect of aspirin on risk of CRC, a Mendelian randomisation (MR) approach was undertaken of which an increase in salicyluric acid levels may be associated with an increased risk of CRC (Chapter 6).

7.1 Summary of main findings

Due to the literature indicating different effects of aspirin depending on timing and duration of treatment (Chapter 3.1.1), we set out to address the effect of duration of treatment in a laboratory setting. Our main finding was that whilst short-term (7 days) aspirin treatment may slow cell growth, no such effect was observed in long-term (18 weeks) aspirin treatment in 3D culture suggesting developing "resistance to aspirin".

The literature has shown that one of the main mechanisms of action of aspirin is the irreversible acetylation and thus inactivation of the cyclooxygenase (COX) enzyme (306). However, a study showed that NCX-4016, an aspirin derivative that does not inhibit COX, was more chemopreventative than aspirin (307) indicating alternative mechanisms of action. To address this question, a combined 'omics approach was undertaken. Through this, we were able to identify that aspirin may be decreasing the expression of both stathmin-1 (*STMN1*) and minichromosome maintenance complex component 5 (*MCM5*). Stathmin-1 is a microtubule destabilizer and plays a role in cell motility (362). We found that decreasing the expression of *STMN1* through siRNA knockdown reduced both cell cycle progression as well as cell migration.

In studying the effect of aspirin on the epigenome, our epigenome-wide association study (EWAS) results show that 24-hour aspirin treatment did not significantly affect CpG site methylation after adjustment for multiple testing (chapter 4). However, our long-term (75 weeks) results showed a genome-wide effect on DNA methylation, identifying 1492 CpG sites across the genome that are influenced by aspirin treatment. Aspirin decreased the methylation of 66.96% of sites, of which these were found in pathways associated with cancer such as the MAPK signalling pathway and PI3K-AKT signalling pathways. We also found that the effect of aspirin on DNA methylation seems to occur when a threshold of aspirin concentration has been reached and is not influenced by further increasing concentrations.

To help identify whether we can predict the effect of aspirin on risk of CRC, we adopted an MR approach. SNPs to proxy for salicylic acid and salicyluric acid were tested for an association with CRC incidence, stratifying between aspirin-users and non-users. No association was observed between salicylic acid and risk of CRC. Using the SNPs that proxy for salicyluric acid, we found a possible causal relationship between the metabolite and risk of CRC using the UK Biobank dataset (IVW beta: 0.000759, 95% CI: -0.000007- 0.001524, P-value=0.052). Since a linear mixed model approach was carried out using UK Biobank data, the results are interpreted as the change in prevalence per standard deviation increase in the metabolite. The IVW estimate predicts that for every 1 standard deviation increase in salicyluric acid, the prevalence of CRC increases by 0.0759%.

7.2 Interpretation of results

Evidence from observational studies (82) and randomised trials (111,112,114) for vascular events have both shown aspirin to be beneficial as a chemopreventative agent thus preventing colorectal cancer incidence, providing evidence for aspirin's use pre-diagnostically. Evidence for aspirin's beneficial properties post-diagnosis has also been shown (Chapter 3.1.1). However, some individuals who take aspirin pre-diagnostically still develop colorectal cancer and others who take it post-diagnostically do not benefit from aspirin's chemopreventative effect (279,280). These highlight the possibility that there may be individuals who are resistant to aspirin's effect. The mechanism through which aspirin works is still not completely understood, as is the reason for how/why individuals develop aspirin resistance.

We propose that there may be two mechanisms through which an individual may be resistant to aspirin. The first is through their genotype which may affect enzyme activity and therefore influence the length of time an individual is exposed to the protective metabolite salicylic acid. The second is an acquired resistance through long-term use. The changes observed from our short-term treatment experiments on adenoma cells provide mechanisms for how aspirin may function as a therapeutic agent. Furthermore, the changes we observed in long-term treatment may explain why some studies found no association between aspirin-use and risk of CRC (Chapter 3.1.1).

7.2.1 Genetic resistance and response to aspirin

We propose that an individual's response to aspirin may be dependent on their genotype, allowing individuals to be more or less responsive to the drug's effects.

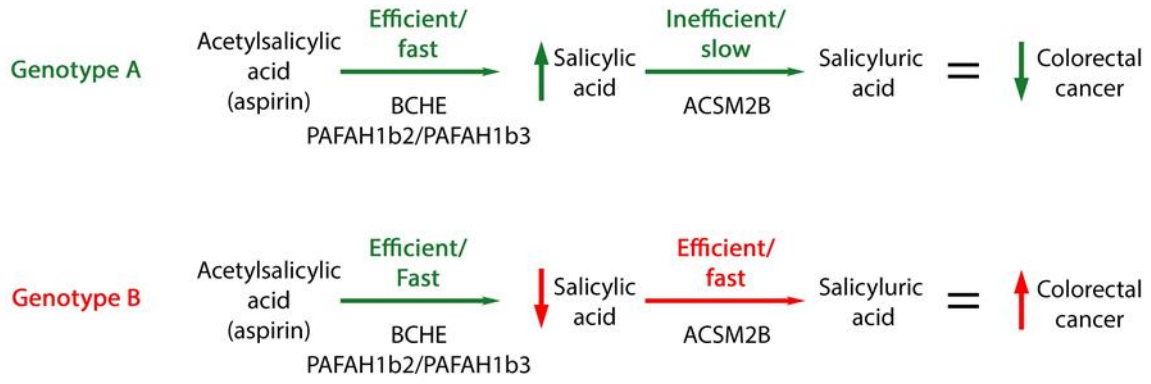


Figure 7-1- Proposed mechanism of genetic "resistance" regulating levels of salicylic acid in the cells

Due to the availability of metabolite data, we only looked at the effect of salicylic acid and salicyluric acid on the risk of CRC. As stated in Figure 7-1, a subset of patients may have SNPs within the enzymes BCHE and PAFAH1b2/PAFAH1b3 that allow efficient hydrolysis of aspirin to salicylic acid. Salicylic acid is the protective metabolite (220) therefore increased exposure time to this may be protective against cancer. Individuals may have genetic polymorphisms in the enzyme ACSM2B involved in breaking down salicylic acid to salicyluric acid which either allow the enzyme to be more efficient or less efficient in this process. If the enzyme is inefficient in its role, this would allow salicylic acid to remain for a longer time before being broken down, thus decreasing the risk of CRC. These individuals are termed as having genotype A. On the other hand, if the genetic polymorphisms allow the enzyme to be more efficient, the result is increased conversion of salicylic acid to salicyluric acid thus reducing the exposure time to salicylic acid and resulting in an increased risk of CRC due to the absence of aspirin's protective effect. These individuals are termed as having genotype B.

The idea that different genotypes and genetic polymorphisms affect enzyme efficiency has been mentioned in the literature. Zhou et. al (2013) had shown that a SNP in *BCHE* (rs6445035) was associated with the rate of hydrolysis of aspirin. For every A allele in this SNP, aspirin hydrolysis was decreased by around 1.2 nmol/ml/min (217). Also, two variants within *UGT1A6* (rs1105879 (R184S) and rs2070959 (T181A)) have previously been predicted to reduce the enzyme's activity by 30-50% (458). Furthermore, a SNP in *CYP2C9* (rs1799853) codes for a variant of the enzyme with reduced activity and metabolic activity (456).

An important implication from these results is the possibility that patients who will most respond to aspirin treatment can be identified based on their genotype or genetic variants and thus advised to take aspirin to reduce their risk of CRC incidence. In addition, those who have a genotype that reduces their sensitivity to

the drug will be advised for other treatment options and advised against long-term aspirin use (to avoid any unwanted side-effects of a drug from which they would not benefit).

7.2.2 Acquired resistance and response to aspirin

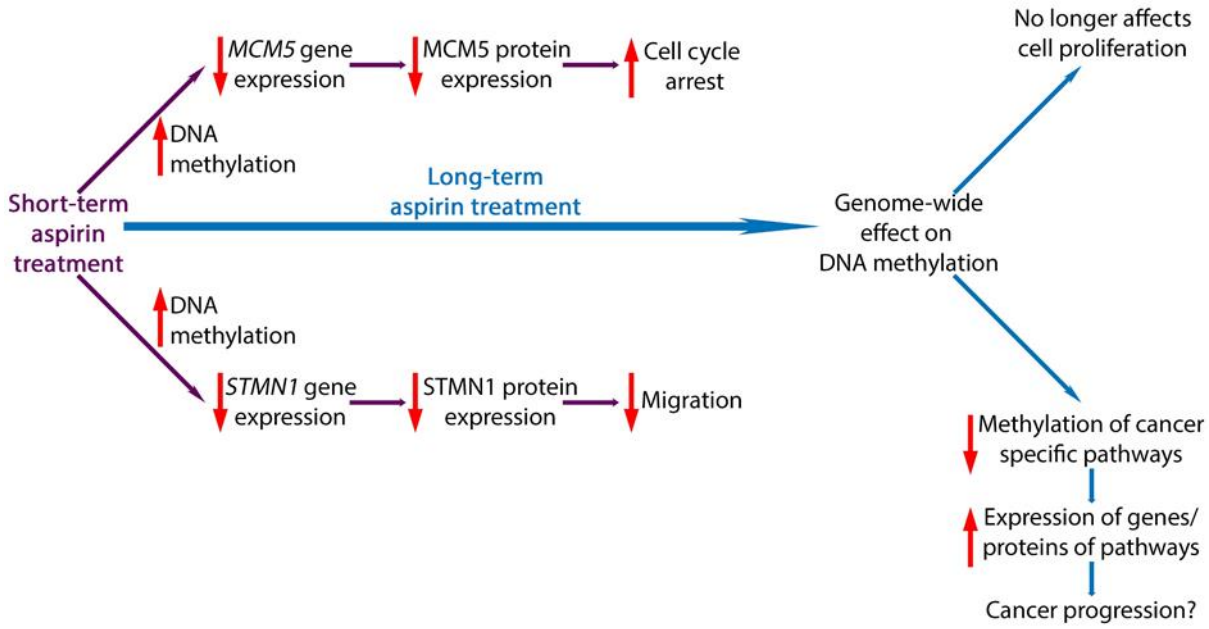


Figure 7-2- Proposed mechanism of action of aspirin upon short-term and long-term treatment

Our short-term aspirin treatment experiments have provided possible mechanisms through which aspirin acts therapeutically to prevent cancer progression. We showed that short-term aspirin treatment may be inducing an increase in DNA methylation of CpG sites found in genes such as *MCM5* and *STMN1*. We confirmed that aspirin was reducing the gene expression of both genes and the protein expression of STMN1. MCM proteins work together to form a helicase complex consisting of MCM2-MCM7 which are involved in DNA unwinding during DNA replication (377). Aspirin has been shown to induce cell cycle arrest previously through various targets (147,319). Downregulation of MCM proteins has been reported to induce G1 and G2 arrest in colon adenocarcinoma cells (378) therefore an alternative mechanism through which aspirin induces cell cycle arrest may be through the downregulation of *MCM5* as was observed in our results (Chapter 3.3.1.3).

STMN1 has also been shown to be associated with metastasis in cancers such as pancreatic and breast cancer as well as colon (359,368,369). This is of particular interest as aspirin reduces the odds of colorectal adenocarcinoma metastasis by 64% (OR:0.36 (CI: 0.18-0.74)) (114). We found that short-term aspirin treatment results in *STMN1* downregulation and found that decreasing its protein expression reduced cell migration. This suggests a mechanism by which aspirin reduces cancer cell migration/metastasis.

Overall, short-term aspirin treatment may be effective as a therapeutic agent by reducing cell migration and cell cycle progression (Figure 7-2).

Our long-term aspirin treatment experiments provide possible reasons why patients become less sensitive to the drug post-diagnosis. Firstly, our initial experiments showed that long-term aspirin treated cells were less sensitive to the growth inhibitory effects of aspirin. To investigate why this may be, we compared the DNA methylation changes of cells treated with short-term and long-term aspirin. Whilst results from short-term treatment were not statistically significant after adjusting for multiple testing, long-term aspirin culture resulted in a striking genome-wide effect on DNA methylation. Aspirin also appeared to reduce methylation of genes within cancer specific pathways such as the MAPK signalling pathway and PI3K-AKT signalling pathway which may increase the expression of genes and activity of these pathways therefore leading to cancer progression.

We postulate that overall, the difference in short-term effect and long-term effect of aspirin on cell growth and response to aspirin may be due to the gradual changes in epigenetics, more specifically, DNA methylation (Figure 7-2). A more pronounced effect on DNA methylation is observed long-term and it may be this that allows cells to be less sensitive to aspirin treatment.

Mansoori et. al (2017) stated that one of the mechanisms through which cancers become resistant to a drug is through epigenetic alterations, whether histone alterations or DNA methylation changes (412). Long-term drug use and development of resistance through genome wide DNA methylation changes have been observed with regards to other drugs such as oxaliplatin (413) and cervical cancer as well as an EGFR inhibitor drug and lung cancer (415). The exact mechanism by which long-term drug treatment may result in drug resistance through changes in DNA methylation is still unknown.

The impact of our observations is the possibility that epigenetic signatures can be identified to determine patients that have become less-sensitive or “resistant” to aspirin after long-term use. This is important as it may be used to inform the next treatment steps for these patients to ensure therapeutic efficacy.

It is possible that length of exposure to aspirin is not the reason for decreased sensitivity to the drug. We identified that long-term aspirin treatment affects epigenetics of adenoma cells. It may be that the protectiveness of aspirin post-diagnosis depends on a specific subtype of adenomas/tumours determined by the tumour’s methylation status. Lou et. al (2014) had previously identified adenoma methylation signatures that could be used to identify adenomas that remain benign or progress to CRC (187). It may be that aspirin epigenetically modifies a specific methylation signature thereby reducing the risk of cancer progression in some patients and having no effect in others.

The idea that aspirin’s protective effect is dependent on an individual’s methylation status may explain the contrasting results of McNeil et. al (2018) who showed that in healthy, older patients, a higher mortality due to cancer was found in those who took aspirin compared to placebo (281). Changes in DNA methylation have been found in both blood and tissue with age (464). This contrasting result of aspirin increasing risk of cancer mortality in elderly patients may possibly be explained by the age-related changes in DNA methylation, since

our results indicate that aspirin can also affect DNA methylation when taken long-term. In this case, it may be that elderly patients are advised against taking aspirin for this reason.

7.3 General strengths and limitations

One of the main strengths behind the methods used here was the ability to undertake an unbiased approach to identify *STMN1* as a potential new target of aspirin, which has not previously been stated in the literature. Whilst we postulate that the mechanism by which aspirin decreases the expression of *STMN1* may be through its effect on DNA methylation, we were limited in the interpretation of the pyrosequencing results due to the contamination of the controls. Also, we acknowledge that this analysis was primarily carried out in cell lines and that the effect of aspirin on *STMN1* expression needs to be investigated within patient samples.

Our approach allowed untargeted and unbiased approach of interrogating multiple CpG sites as opposed to testing for methylation changes in a select few. The nature of culturing cells in a controlled environment reduced the risk of confounding giving more strength to believe that the results observed are only due to the treatment of aspirin. Whilst this is a strength to our approach, we also acknowledge that this analysis was carried out in one cell line and cannot be used to make deductions of the effect of aspirin on all CRC patients. This is a huge limitation in interpreting the results as different adenoma's have different methylation profiles (187). Therefore, the effect of aspirin on adenomas needs to be investigated in patient samples as well. Furthermore, experiments need to be carried out to validate aspirin's effect on DNA methylation. Also, we attempted to adjust for surrogate variables although we find one in particular to be highly correlated with aspirin treatment, therefore in this case, over-adjustment is an issue that needs to be considered, particularly when using results from laboratory-based experiments. One further limitation in our analysis is the fact that we only adjusted for batch. There may have been other factors that may have confounded the results such as the length of aspirin treatment as DNA was extracted based on cell confluency and not length of treatment, as cells grew at different rates under different concentrations.

We also attempted to apply MR to a biochemical pathway to predict drug effect against a disease/outcome. However, a number of limitations were present in our study. By limiting the SNPs used in this study to only those that are within the genomic coordinates of the enzymes, we removed other potential SNPs that may be regulating enzyme expression outside of these regions. Therefore, we propose that a GWAS approach to identify other SNPs associated with the metabolites is necessary. In addition, one major limitation in this study is that when stratified between aspirin-users and non-users, there was only 7 CRC patients who took aspirin meaning that there was low power to detect the effect of the metabolites on risk of CRC. Furthermore, the effect of aspirin metabolites on risk of cancer progression was not addressed in this study.

Another limitation of our MR approach was the presence of metabolite data for salicylic acid and salicyluric acid only, despite salicylic acid being hydrolysed to many other metabolites. By only using a single linear

pathway, results from this may be misleading due to the presence of horizontal pleiotropy (such as enzymes being involved in other pathways) (463) which was observed in our results. However, using a multivariable MR approach whereby MR is conducted on all the pathways involved in salicylic acid hydrolysis into the other metabolites as well, may help to overcome the issue of horizontal pleiotropy as the results involve combining multiple pathways related to salicylic acid metabolism. Ideally, it would be useful to obtain metabolite data on the other metabolites such as gentisic acid and salicylacyl glucuronide (Figure 1-20) and run a multivariable MR approach. To our knowledge, these metabolites have not yet been measured in large studies, therefore this approach is currently not possible.

7.4 Strength and limitations of combining population-based and laboratory-based approaches

For the advancement of research in the cancer field, the combination of different approaches and methods is beneficial. In this thesis, both population and laboratory-based approaches were adopted.

One advantage of this approach is that population-based approaches allowed the interrogation of a large amount of data, for example, the interrogation of thousands of CpG sites. The laboratory-based approach then allowed investigation on a much smaller scale through the use of pyrosequencing to validate changes in DNA methylation at specific CpG sites. Whilst conclusions from pyrosequencing could not be made in this thesis, it shows the possibility that findings from epidemiology can then be validated through smaller and more direct approaches, providing further evidence for/against the association and increasing strength for/against a conclusion. Once there is evidence for these findings, validation can be carried out using patient samples of those who take aspirin against those who are non-users.

One of the most common problems in epidemiology is the possibility of confounding due to the inability to control all variables. Epigenome-wide association studies are subject to confounding by batch effects and biological factors such as cellular heterogeneity (391,465). The advantage of a laboratory-based approach is that experiments can be designed so that the only variable that changes is the exposure of interest, allowing direct conclusions to be made of treatment effect. This was useful with regards to the effect of long-term aspirin treatment on DNA methylation- whilst surrogate variable analysis (SVA) suggested there was confounding, this was unlikely to be true due to the controlled nature of sample generation. It is important that the researcher undertaking the methods has prior and sound knowledge of the strength and limitations of population and laboratory-based methods. In a conventional interpretation of SVA, this would have indicated our results as being null, when an effect on DNA methylation may exist.

Population-based methods involve unbiased approaches therefore not being influenced by prior knowledge or opinions, unlike laboratory-based methods. We adopted an unbiased approach to identify new targets of aspirin through the use of proteomic and methylomic data. This allowed the identification of at least two new targets of aspirin not previously linked to the drug in the literature. Population-based methods are

advantageous as they bypass the need to spend a long amount of time in the laboratory and can be used to provide guidance on the direction of research within the laboratory, therefore saving both time and money.

For MR to be useful and applied in the contexts of drugs and their effect on diseases/outcomes, knowledge of the molecular pathway of drug metabolism is essential. We were able to conduct an MR approach through an understanding of the biochemical pathway of aspirin metabolism (Figure 1-20). The understanding of both laboratory and population-based approaches allowed the strength to interrogate and apply MR to a biochemical pathway, of which has not been done before. Our results would indicate that the aspirin metabolite salicylic acid, the active metabolite of aspirin (220), is not associated with CRC incidence although this is untrue due to evidence provided by observational data and randomised trials (82,111,112,114). Understanding the strengths and limitations of the approach allowed us to be more cautious in how results were interpreted, and the future work needed to be carried out.

In the case of our MR approach, genotype data as well as metabolite and CRC risk data were already available thereby allowing a quick approach to address the question of aspirin metabolites and their association with CRC. However, the limitation of population-based methods includes when the exposure of interest has not yet been measured, such as the other metabolites from the aspirin hydrolysis pathway (Figure 1-20) or when sample sizes are too small (such as only having 7 CRC patients who are aspirin users) therefore there is low power to detect the effect of the aspirin metabolite on risk of CRC. To generate new data, such as the measurement of the other metabolites, a huge cost may be involved. Therefore, population-based methods are advantageous if the data is already present and a large enough sample size is present, otherwise, time and money are both factors that need to be considered before new data is generated. Since laboratory-based approaches involve a much smaller number of mouse models or cell lines, the costs incurred to generate data are not as high.

Overall, population and laboratory-based approaches both provide advantages and limitations on their own, but together they can be used to bypass some of these limitations providing strength to form conclusions of associations. Time to test hypothesis, time to generate data, sample sizes and costs incurred are all factors to be considered over which approach is best for the questions to be addressed. Finally, knowledge and understanding of both fields is crucial to avoid making incorrect conclusions.

7.5 Future directions

Work presented in this thesis has led to many novel questions that need to be addressed. With regards to aspirin's effect on *STMN1* expression, further research using CRC patients is needed. This can be done through immunohistochemistry of patient biopsies to investigate whether STMN1 protein expression is indeed reduced in CRC patients who take aspirin compared to those who are non-users. The effect of *STMN1* gene and protein expression on risk of CRC may also be tested through an MR approach, whereby SNPs or genetic variants to proxy for *STMN1* gene and protein expression can be used to then test whether they

affect risk of CRC. Ideally, testing for an association between *STMN1* gene and protein expression with their effect on CRC progression would be more beneficial as this protein is linked to CRC metastasis (359). Furthermore, to address whether the effect of aspirin on cell migration is indeed through its decrease in *STMN1* expression, migration assays of aspirin-treated cells induced to express *STMN1* is necessary. In addition, assays in mice with CRISPR/Cas9 knockout of *STMN1* can be used to look at tumorigenicity and metastasis *in vivo*. This will help provide further evidence for our hypothesis.

Our long-term aspirin treated cells showed a genome-wide effect of aspirin on DNA methylation. We hypothesise that long-term aspirin treatment can cause methylation changes that may lead to decreased sensitivity or "resistance" to the drug. Experiments to validate aspirin's effect on DNA methylation are necessary. It may be best to combine multiple methods to assess DNA methylation such as digestion-based assays followed by PCR or qPCR or methylation specific PCR of bisulphite converted samples (416). These validation experiments are also important as our analysis showed that 0.5mM aspirin has the same effect on DNA methylation as 4mM aspirin, indicating a threshold for aspirin's effect on DNA methylation which may guide future dose recommendations. Also, the effect of aspirin on DNA methylation needs to be investigated in patient samples as our analysis only used one cell line which cannot be used to represent all tumours.

Although our methylation analysis shows aspirin has a genome-wide effect on DNA methylation, it may be that only one or two of the targeted CpG sites are causally associated with drug resistance. In the case of the EGFR inhibitor drug and lung cancer, it was found that 2 of the genes affected by long-term exposure to the drug were involved in drug resistance (415). A Mendelian randomisation approach involving the use of SNPs to predict changes in DNA methylation may be used to identify which CpG sites influence the risk of CRC and therefore may help identify which sites are involved in aspirin "resistance".

Finally, with regards to our MR approach, a GWAS approach to identify more robust proxies for salicylic acid and salicyluric acid needs to be undertaken, as we only limited our SNPs to those within the genomic regions of the enzymes in the pathway, despite evidence in the literature that suggests *cis* genomic regions and long non-coding RNAs (lncRNAs) outside of the gene of interest itself may regulate gene expression thus impacting upon the metabolism of the very metabolites being investigated (453,454). In addition, one major limitation in this study is that when stratified between aspirin-users and non-users, there was only 7 CRC patients who took aspirin therefore indicating that there was low power to detect the effect of the metabolites on risk of CRC. Therefore, this MR approach needs to be conducted in a larger sample size with information on both aspirin use and CRC incidence.

7.6 Final conclusions

Data from this thesis have suggested different effects of aspirin depending on short-term or long-term treatment. Whilst many aspirin targets are well-known in the literature, evidence for its effect on other proteins lead us to identify *STMN1* as a potentially new target. The risk of colorectal cancer metastasis in aspirin users is reduced (114) and this investigation suggests this may be through the reduction of *STMN1* expression.

For the first time, results from long-term aspirin treatment have indicated that aspirin may have a genome-wide effect on DNA methylation which could explain the reduced effect of aspirin on cell growth in 3D cultures suggesting a mechanism of decreased sensitivity to aspirin or "acquired resistance". Whilst results from our MR analysis were limited by many factors, repeating this approach with a larger sample size and more robust genetic instruments may help to identify aspirin's role in CRC incidence.

These findings are important as they help clarify possible mechanisms through which aspirin is chemopreventative pre and post-diagnosis, as well as discussing the possibility that aspirin resistance may occur due to genetic polymorphisms or due to methylation changes that develop through long-term use. These results may be important in helping decide who would most benefit from the drug for chemoprevention therefore possibly being used to help guide clinical recommendations of use.

References

1. Simon R. Lost in translation: Problems and pitfalls in translating laboratory observations to clinical utility. Eur J Cancer [Internet]. Elsevier Ltd; 2008;44(18):2707–13. Available from: http://dx.doi.org/10.1016/j.ejca.2008.09.009

2. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology. 2013;144(4):799–807.

3. Moayyeri A, Hammond CJ, Hart DJ, Spector TD. The UK Adult Twin Registry (TwinsUK Resource). Twin Res Hum Genet. 2013;16(1):144–9.

4. Collins R. What makes UK Biobank special? Lancet. 2012 Mar 31;379(9822):1173–4.

5. What is bowel cancer? | Cancer Research UK [Internet]. [cited 2018 Aug 9]. Available from: https://www.cancerresearchuk.org/about-cancer/bowel-cancer/about-bowel-cancer

6. Bowel cancer statistics | Cancer Research UK [Internet]. [cited 2018 Apr 13]. Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer#heading-Zero

7. De Rosa M, Pace U, Rega D, Costabile V, Duraturo F, Izzo P, et al. Genetics, diagnosis and management of colorectal cancer (Review). Oncol Rep [Internet]. 2015;1087–96. Available from: http://www.spandidos-publications.com/10.3892/or.2015.4108

8. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359–86.

9. Cancer Research UK. Bowel cancer incidence trends over time [Internet]. [cited 2018 Aug 7]. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/incidence#heading-Two

10. Siegel RL, Fedewa SA, Anderson WF, Miller KD, Ma J, Rosenberg PS, et al. Colorectal Cancer Incidence Patterns in the United States, 1974-2013. J Natl Cancer Inst. 2017;109(8):djw322.

11. Drewes JL, Housseau F, Sears CL. Sporadic colorectal cancer: microbial contributors to disease prevention, development and therapy. Br J Cancer [Internet]. Nature Publishing Group; 2016;115(3):273–80. Available from: http://dx.doi.org/10.1038/bjc.2016.189

12. Yamagishi H, Kuroda H, Imai Y, Hiraishi H. Molecular pathogenesis of sporadic colorectal cancers. Chin J Cancer. BioMed Central; 2016 Jan 6;35:4.

13. Parkin DM, Boyd L, Walker LC. 16. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. Br J Cancer [Internet]. Nature Publishing Group; 2011;105(Suppl 2):S77–81. Available from: http://dx.doi.org/10.1038/bjc.2011.489

14. Islami F, Goding Sauer A, Miller KD, Siegel RL, Fedewa SA, Jacobs EJ, et al. Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. CA Cancer J Clin [Internet]. 2017;68(1):31–54. Available from: http://doi.wiley.com/10.3322/caac.21440

15. Ezzati M, Riboli E. Behavioral and Dietary Risk Factors for Noncommunicable Diseases. N Engl J Med [Internet]. 2013;369(10):954–64. Available from: http://www.nejm.org/doi/10.1056/NEJMra1203528

16. Lee JM, Pilli S, Gebremariam a., Keirns CC, Davis MM, Vijan S, et al. Getting heavier, younger: Trajectories of obesity over the life course. Int J Obes [Internet]. Nature Publishing Group; 2010;34(4):614–23. Available from: http://dx.doi.org/10.1038/ijo.2009.235

17. International Agency for Research on Cancer. Agents Classified by the IARC Monographs, Volumes 1–122 – IARC [Internet]. [cited 2018 Aug 6]. Available from: https://monographs.iarc.fr/agents-classified-by-the-iarc/

18. Secretan BL, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K, et al. Body Fatness and Cancer — Viewpoint of the IARC Working Group. N Engl J Med. 2016;375(8):794–8.

19. Wolin KY, Yan Y, Colditz GA, Lee I-M. Physical activity and colon cancer prevention: A meta-analysis. Br J Cancer. 2009;100(4):611–6.

20. Cong YJ, Gan Y, Sun HL, Deng J, Cao SY, Xu X, et al. Association of sedentary behaviour with colon and rectal cancer: A meta-analysis of observational studies. Br J Cancer. 2014;110(3):817–26.

21. Aune D, Chan DSM, Lau R, Vieira R, Greenwood DC, Kampman E, et al. Dietary fibre, whole grains, and risk of colorectal cancer: systematic review and dose-response meta-analysis of prospective studies. Br Med J [Internet]. 2011;343(d6617):d6617. Available from: http://www.bmj.com/cgi/doi/10.1136/bmj.d6617

22. Ben Q, Sun Y, Chai R, Qian A, Xu B, Yuan Y. Dietary Fiber Intake Reduces Risk for Colorectal Adenoma: A Meta-analysis. Gastroenterology [Internet]. Elsevier; 2014;146(3):689–99. Available from: http://dx.doi.org/10.1053/j.gastro.2013.11.003

23. Song M, Chan AT. Environmental factors, gut microbiota, and colorectal cancer prevention. Clin Gastroenterol Hepatol [Internet]. The American Gastroenterological Association; 2018;In press. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1542356518307122

24. Wolf AMD, Fontham ETH, Church TR, Flowers CR, Guerra CE, LaMonte SJ, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society [Internet]. CA: A Cancer Journal for Clinicians. 2018. p. 250–81. Available from: http://doi.wiley.com/10.3322/caac.21457

25. Bowel cancer screening - NHS UK [Internet]. [cited 2018 Aug 8]. Available from: https://www.nhs.uk/conditions/bowel-cancer-screening/

26. Van der Flier LG, Clevers H. Stem Cells, Self-Renewal, and Differentiation in the Intestinal Epithelium. Annu Rev Physiol [Internet]. 2009;71(1):241–60. Available from: http://www.annualreviews.org/doi/10.1146/annurev.physiol.010908.163145

27. Ricci-Vitiani L, Fabrizi E, Palio E, De Maria R. Colon cancer stem cells. J Mol Med. 2009;87(11):1097–104.

28. Kaiko GE, Ryu SH, Koues OI, Collins PL, Solnica-Krezel L, Pearce EJ, et al. The Colonic Crypt Protects Stem Cells from Microbiota-Derived Metabolites. Cell [Internet]. Elsevier Inc.; 2016;165(7):1708–20. Available from: http://dx.doi.org/10.1016/j.cell.2016.05.018

29. Humphries A, Wright NA. Colonic crypt organization and tumorigenesis. Nat Rev Cancer. 2008;8(6):415–24.

30. Sasaki N, Sachs N, Wiebrands K, Ellenbroek SIJ, Fumagalli A, Lyubimova A, et al. Reg4+ deep crypt secretory cells function as epithelial niche for Lgr5+ stem cells in colon. Proc Natl Acad Sci [Internet]. 2016;113(37):E5399–407. Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.1607327113

31. Pellegrinet L, Rodilla V, Liu Z, Chen S, Koch U, Espinosa L, et al. Dll1- and Dll4-mediated notch signaling are required for homeostasis of intestinal stem cells. Gastroenterology [Internet]. Elsevier Inc.; 2011;140(4):1230–40. Available from: http://dx.doi.org/10.1053/j.gastro.2011.01.005

32. Birchenough GMH, Johansson MEV, Gustafsson JK, Bergström JH, Hansson GC. New developments in goblet cell mucus secretion and function. Mucosal Immunol. 2015;8(4):712–9.

33. Engelstoft MS, Egerod KL, Lund ML, Schwartz TW. Enteroendocrine cell types revisited. Curr Opin Pharmacol [Internet]. Elsevier Ltd; 2013;13(6):912–21. Available from: http://dx.doi.org/10.1016/j.coph.2013.09.018

34. Gerbe F, Sidot E, Smyth DJ, Ohmoto M, Matsumoto I, Dardalhon V, et al. Intestinal epithelial tuft cells initiate type 2 mucosal immunity to helminth parasites. Nature [Internet]. Nature Publishing Group; 2016;529(7585):226–30. Available from: http://dx.doi.org/10.1038/nature16527

35. Itzkovitz S, Lyubimova A, Blat IC, Maynard M, van Es J, Lees J, et al. Single-molecule transcript counting of stem-cell markers in the mouse intestine. Nat Cell Biol [Internet]. Nature Publishing Group; 2011;14(1):106–14. Available from: http://dx.doi.org/10.1038/ncb2384

36. Kraehenbuhl J-P, Neutra MR. EPITHELIAL M CELLS: Differentiation and Function. Annu Rev Cell Dev Biol [Internet]. 2000;16(1):301–32. Available from: http://www.annualreviews.org/doi/abs/10.1146/annurev.cellbio.16.1.301\npapers2://publication/doi/10.1146/annurev.cellbio.16.1.301

37. Takeda N, Jain R, LeBoeuf MR, Wang Q, Lu MM, Epstein JA. Interconversion Between Intestinal Stem Cell Populations in Distinct Niches. Science (80-). 2011;334(6061):1420–4.

38. Barker N. Adult intestinal stem cells: Critical drivers of epithelial homeostasis and regeneration. Nat Rev Mol Cell Biol [Internet]. Nature Publishing Group; 2014;15(1):19–33. Available from: http://dx.doi.org/10.1038/nrm3721

39. Munro MJ, Wickremesekera SK, Peng L, Tan ST, Itinteang T. Cancer stem cells in colorectal cancer: A review. J Clin Pathol. 2018;71(2):110–6.

40. Fearon ER, Vogelstein B. A Genetic Model for Colorectal Tumorigenesis. Cell. 1990;61(5):759–67.

41. Walther A, Johnstone E, Swanton C, Midgley R, Tomlinson I, Kerr D. Genetic prognostic and predictive markers in colorectal cancer. Nat Rev Cancer [Internet]. 2009;9(7):489–99. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19536109

42. Tauriello DVF, Calon A, Lonardo E, Batlle E. Determinants of metastatic competency in colorectal cancer. Mol Oncol. 2017;11(1):97–119.

43. Sparks AB, Morin PJ, Vogelstein B, Kinzler KW. Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer. [Internet]. Cancer research. 1998. p. 1130–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9515795

44. Korinek V, Barker N, Morin PJ, van Wichen D, de Weger R, Kinzler KW, et al. Constitutive transcriptional activation by a b-catenin-Tcf complex in APC-/- colon carcinoma. Science (80-). 1997;275(5037):1784–7.

45. Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. Oncogene [Internet]. Nature Publishing Group; 2017;36(11):1461–73. Available from: http://dx.doi.org/10.1038/onc.2016.304

46. Bogaert J, Prenen H. Molecular genetics of colorectal cancer. Ann Gastroenterol [Internet]. 2014;27(1):9–14. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3959535&tool=pmcentrez&rendertype=abstract

47. Lengauer C, Kinzler KW, Vogelstein B. Genetic instability in colorectal cancers. Nature. 1997. p. 623–7.

48. Chen L, Pan X, Hu X, Zhang Y-H, Wang S, Huang T, et al. Gene expression differences among different MSI statuses in colorectal cancer. Int J Cancer [Internet]. 2018;Epub ahead. Available from: http://doi.wiley.com/10.1002/ijc.31554

49. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa J-PJ. CpG island methylator phenotype in colorectal cancer. Med Sci. 1999;96(July):8681–6.

50. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nat Genet. 2006;38(7):787–93.

51. Ogino S, Kawasaki T, Kirkner GJ, Loda M, Fuchs CS. CpG island methylator phenotype-low (CIMP-low) in colorectal cancer: Possible associations with male sex and KRAS mutations. J Mol Diagnostics [Internet]. American Society for Investigative Pathology and Association for Molecular Pathology; 2006;8(5):582–8. Available from: http://dx.doi.org/10.2353/jmoldx.2006.060082

52. Song J, Yi C. Chemical Modifications to RNA: A New Layer of Gene Expression Regulation. ACS Chem Biol. 2017;12(2):316–25.

53. Porcellini E, Laprovitera N, Riefolo M, Ravaioli M, Garajova I, Ferracin M. Epigenetic and epitranscriptomic changes in colorectal cancer: Diagnostic, prognostic, and treatment implications. Cancer Lett [Internet]. Elsevier Ltd; 2018;419:84–95. Available from: https://doi.org/10.1016/j.canlet.2018.01.049

54. Kroemer G, Zitvogel L. Cancer immunotherapy in 2017: The breakthrough of the microbiota. Nat Rev Immunol [Internet]. Nature Publishing Group; 2018;18(2):87–8. Available from: http://dx.doi.org/10.1038/nri.2018.4

55. Batlle E, Clevers H. Cancer stem cells revisited. Nat Med [Internet]. Nature Publishing Group; 2017;23(10):1124–34. Available from: http://dx.doi.org/10.1038/nm.4409

56. Ricci-Vitiani L, Lombardi DG, Pilozzi E, Biffoni M, Todaro M, Peschle C, et al. Identification and expansion of human colon-cancer-initiating cells. Nature. 2007;445(7123):111–5.

57. Barker N, Ridgway RA, Van Es JH, Van De Wetering M, Begthel H, Van Den Born M, et al. Crypt stem cells as the cells-of-origin of intestinal cancer. Nature [Internet]. Nature Publishing Group; 2009;457(7229):608–11. Available from: http://dx.doi.org/10.1038/nature07602

58. Chen J, Li Y, Yu TS, McKay RM, Burns DK, Kernie SG, et al. A restricted cell population propagates glioblastoma growth after chemotherapy. Nature. 2012;488(7412):522–6.

59. De Sousa E Melo F, Kurtova A V., Harnoss JM, Kljavin N, Hoeck JD, Hung J, et al. A distinct role for Lgr5 + stem cells in primary and metastatic colon cancer. Nature [Internet]. Nature Publishing Group; 2017;543(7647):676–80. Available from: http://dx.doi.org/10.1038/nature21713

60. Chaffer CL, Weinberg RA. A Perspective on Cancer Cell Metastasis. Science (80-). 2011;331(6024):1559–64.

61. Massagué J, Obenauf AC. Metastatic colonization by circulating tumour cells. Nature. 2016;529(7586):298–306.

62. Sosa MS, Bragado P, Aguirre-Ghiso JA. Mechanisms of disseminated cancer cell dormancy: an awakening field. Nat Rev Cancer [Internet]. Nature Publishing Group; 2014;14(9):611–22. Available from: http://dx.doi.org/10.1038/nrc3793

63. Oskarsson T, Batlle E, Massagué J. Metastatic stem cells: Sources, niches, and vital pathways. Cell Stem Cell. 2014;14(3):306–21.

64. Zlobec I, Lugli A. Epithelial mesenchymal transition and tumor budding in aggressive colorectal cancer: Tumor budding as oncotarget. Oncotarget [Internet]. 2010;1(7):651–61. Available from: http://oncotarget.com/abstract/199

65. Boyer B, Thiery JP. Epithelium-mesenchyme interconversion as example of epithelial plasticity. APMIS. 1993 Apr;101(4):257–68.

66. Cao H, Xu E, Liu H, Wan L, Lai M. Epithelial-mesenchymal transition in colorectal cancer metastasis: A system review. Pathol Res Pract [Internet]. Elsevier GmbH.; 2015;211(8):557–69. Available from: http://dx.doi.org/10.1016/j.prp.2015.05.010

67. Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. Nat Rev Mol Cell Biol [Internet]. Nature Publishing Group; 2014;15(3):178–96. Available from: http://dx.doi.org/10.1038/nrm3758

68. Cano A, Pérez-Moreno MA, Rodrigo I, Locascio A, Blanco MJ, del Barrio MG, et al. The transcription factor Snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. Nat Cell Biol. 2000;2(2):76–83.

69. Vu T, Datta PK. Regulation of EMT in Colorectal Cancer: A Culprit in Metastasis. Cancers (Basel). 2017;9(12):1–22.

70. Roberts AB, Wakefield LM. The two faces of transforming growth factor beta in carcinogenesis. Proc Natl Acad Sci U S A [Internet]. 2003;100(15):8621–3. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=166359&tool=pmcentrez&rendertype=abstract

71. Labelle M, Begum S, Hynes RO. Direct Signaling between Platelets and Cancer Cells Induces an Epithelial-Mesenchymal-Like Transition and Promotes Metastasis. Cancer Cell [Internet]. Elsevier Inc.; 2011;20(5):576–90. Available from: http://dx.doi.org/10.1016/j.ccr.2011.09.009

72. Jong IY, Li X-Y, Ota I, Fearon ER, Weiss SJ. Wnt-dependent Regulation of the E-cadherin Repressor Snail. J Biol Chem. 2005;280(12):11740–8.

73. Vane JR, Botting RM. The mechanism of action of aspirin. Thromb Res. 2003;110(5-6):255–8.

74. Norn S, Permin H, Kruse PR, Kruse E. [From willow bark to acetylsalicylic acid]. Dan Medicinhist Arbog. 2009;37:79–98.

75. Sneader W. The discovery of aspirin: a reappraisal. Pharm J. 2000;321(7276):1591–4.

76. Ornelas A, Zacharias-Millward N, Menter DG, Davis JS, Lichtenberger L, Hawke D, et al. Beyond COX-1: the effects of aspirin on platelet biology and potential mechanisms of chemoprevention. Cancer Metastasis Rev. Cancer and Metastasis Reviews; 2017;36(2):289–303.

77. Gravinese C, Bianco M, Cerrato E, Destefanis P, Luciano A, Bernardi A, et al. Is Aspirin Still the Cornerstone of Antiplatelet Therapy in Patients With Coronary Artery Disease? An Historical and Practical Narrative Review. Hosp Pract Res [Internet]. 2017;2(4):94–101. Available from: http://jhpr.ir/article_49556.html

78. Tatham MH, Cole C, Scullion P, Wilkie R, Westwood NJ, Stark L a., et al. A Proteomic Approach to Analyze the Aspirin-mediated Lysine Acetylome. Mol Cell Proteomics [Internet]. 2017;16(2):310–26. Available from: http://www.mcponline.org/lookup/doi/10.1074/mcp.O116.065219

79. Xie M, Shan Z, Zhang Y, Chen S, Yang W, Bao W, et al. Aspirin for primary prevention of cardiovascular events: Meta-analysis of randomized controlled trials and subgroup analysis by sex and diabetes status. PLoS One. 2014;9(10).

80. Rothwell PM, Algra A, Chen Z, Diener H-C, Norrving B, Mehta Z. Effects of aspirin on risk and severity of early recurrent stroke after transient ischaemic attack and ischaemic stroke: time-course analysis of randomised trials. Lancet. Rothwell et al. Open Access article distributed under the terms of CC BY; 2016;388(10042):365–75.

81. Roberge S, Bujold E, Nicolaides KH. Aspirin for the prevention of preterm and term preeclampsia: systematic review and metaanalysis. Am J Obstet Gynecol [Internet]. Elsevier Inc.; 2018;218(3):287–93.e1. Available from: https://doi.org/10.1016/j.ajog.2017.11.561

82. Qiao Y, Yang T, Gan Y, Li W, Wang C, Gong Y, et al. Associations between aspirin use and the risk of cancers : a meta-analysis of observational studies. BioMed Cent Cancer. BMC Cancer; 2018;18(1):1–57.

83. Malakar S, Gibson PR, Barrett JS, Muir JG. Naturally occurring dietary salicylates: A closer look at common Australian foods. J Food Compos Anal [Internet]. Elsevier Inc.; 2017;57:31–9. Available from: http://dx.doi.org/10.1016/j.jfca.2016.12.008

84. Wood A, Baxter G, Thies F, Kyle J, Duthie G. A systematic review of salicylates in foods: Estimated daily intake of a Scottish population. Mol Nutr Food Res. 2011;55(Suppl. 1):S7–14.

85. Ash M, Ash I. Handbook of preservatives. Synapse Information Resources; 2009. 1565 p.

86. Duthie GG, Wood AD. Natural salicylates: foods, functions and disease prevention. Food Funct. 2011;2(9):515–20.

87. Li L, Geraghty OC, Mehta Z, Rothwell PM. Age-specific risks, severity, time course, and outcome of bleeding on long-term antiplatelet treatment after vascular events: a population-based cohort study. Lancet [Internet]. The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license; 2017;390(10093):490–9. Available from: http://dx.doi.org/10.1016/S0140-6736(17)30770-5

88. Gorelick PB, Weisman SM. Risk of Hemorrhagic Stroke With Aspirin Use: An Update. Stroke. 2005;36(8):1801–7.

89. Malfertheiner P, Chan FKL, McColl KEL. Peptic ulcer disease. Lancet [Internet]. Elsevier Ltd; 2009;374(9699):1449–61. Available from: http://dx.doi.org/10.1016/S0140-6736(09)60938-7

90. Wakeman C, Keenan J, Eteuati J, Hollington P, Eglinton T, Frizelle F. Chemoprevention of colorectal neoplasia. ANZ J Surg [Internet]. 2015; Available from: papers2://publication/doi/10.1053/j.gastro.2008.06.092

91. Cuzick J, Thorat MA, Bosetti C, Brown PH, Burn J, Cook NR, et al. Estimates of benefits and harms of prophylactic use of aspirin in the general population. Ann Oncol. 2015;26(1):47–57.

92. Thiagarajan P, Jankowski JA. Aspirin and NSAIDs; Benefits and harms for the gut. Best Pract Res Clin Gastroenterol. 2012;26(2):197–206.

93. Elwood PC, Gallagher AM, Duthie GG, Mur LAJ, Morgan G. Aspirin, salicylates, and cancer. Lancet [Internet]. Elsevier Ltd; 2009;373(9671):1301–9. Available from: http://dx.doi.org/10.1016/S0140-6736(09)60243-9

94. Baenkler H-W. Salicylate intolerance: Pathophysiology, Clinical Spectrum, Diagnosis and Treatment. Dtsch Ärzteblatt Int [Internet]. 2008;105(8):137–42. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2696737&tool=pmcentrez&rendertype=abstract

95. Yamashita M. Aspirin Intolerance: Experimental Models for Bed-to-Bench. Curr Drug Targets. 2016;17(16):1963–70.

96. Samter M, Beers RF. Concerning the nature of intolerance to aspirin. J Allergy. 1967;40(5):281–93.

97. NHLBI. National Asthma Education and Prevention Program Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma Full Report 2007 [Internet]. 2007 [cited 2018 Aug 16]. Available from: https://www.nhlbi.nih.gov/sites/default/files/media/docs/asthgdln_1.pdf

98. Sánchez-Borges M, Caballero-Fonseca F, Capriles-Hulett A, González-Aveledo L. Aspirin-exacerbated cutaneous disease (AECD) is a distinct subphenotype of chronic spontaneous urticaria. J Eur Acad Dermatology Venereol. 2015;29(4):698–701.

99. WHO | Epidemiology. WHO [Internet]. World Health Organization; 2014 [cited 2018 Sep 14]; Available from: http://www.who.int/topics/epidemiology/en/

100. Petticrew M, Roberts H. Evidence, hierarchies, and typologies: horses for courses. J Epidemiol Community Health. 2003;57(7):527–9.

101. Davies NM, Holmes M V., Davey Smith G. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601.

102. Littell JH, Corcoran J, Pillai V. Systematic Reviews and Meta-Analysis. Oxford University Press, Inc; 2008. 1-201 p.

103. Gopalakrishnan S, Ganeshkumar P. Systematic reviews and meta-analysis: Understanding the best evidence in primary healthcare. J Fam Med Prim Care [Internet]. 2013;2(1):9–14. Available from: http://www.jfmpc.com/text.asp?2013/2/1/9/109934

104. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions [Internet]. Version 5.1.0. Higgins J, Green S, editors. The Cochrane Collaboration, 2011; 2011. Available from: www.handbook.cochrane.org.

105. Sterne JAC, Egger M, Davey Smith G. Systematic Reviews In Health Care: Investigating And Dealing With Publication And Other Biases In Meta-Analysis. BMJ [Internet]. 2001;323(7304):101–5. Available from: http://www.jstor.org/stable/25467370

106. Egger M, Dickersin K, Davey Smith G. Problems and Limitations in Conducting Systematic Reviews. In: Matthias Egger, George Davey Smith DGA, editor. Systematic Reviews in Health Care. 2nd ed. London, UK: BMJ Publishing Group; p. 43–68.

107. Spieth PM, Kubasch AS, Penzlin AI, Illigens BM-W, Barlinn K, Siepmann T. Randomized controlled trials - a matter of design. Neuropsychiatr Dis Treat [Internet]. 2016;12:1341–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27354804%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4910682%0Ahttps://www.dovepress.com/randomized-controlled-trials-nbspa-matter-of-design-peer-reviewed-article-NDT

108. Song JW, Chung KC. Observational Studies: Cohort and Case-Control Studies. Plast Reconstr Surg [Internet]. 2010;126(6):2234–42. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2998589/

109. Morabia A. A History of Epidemiologic Methods and Concepts. Basel: Birkhauser Verlag; 2004. 1-405 p.

110. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. I. Principles. Am J Epidemiol. 1992;135(9):1019–28.

111. Rothwell PM, Wilson M, Elwin CE, Norrving B, Algra A, Warlow CP, et al. Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials. Lancet. Elsevier Ltd; 2010;376(9754):1741–50.

112. Rothwell PM, Fowkes FGR, Belch JF, Ogawa H, Warlow CP, Meade TW. Effect of daily aspirin on long-term risk of death due to cancer: Analysis of individual patient data from randomised trials. Lancet [Internet]. Elsevier Ltd; 2011;377(9759):31–41. Available from: http://dx.doi.org/10.1016/S0140-6736(10)62110-1

113. Rothwell PM, Cook NR, Gaziano JM, Price JF, Belch JFF, Roncaglioni MC, et al. Effects of aspirin on risks of vascular events and cancer according to bodyweight and dose: analysis of individual patient data from randomised trials. Lancet [Internet]. 2018;392:387–99. Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(18)31133-4/fulltext

114. Rothwell PM, Wilson M, Price JF, Belch JFF, Meade TW, Mehta Z. Effect of daily aspirin on risk of cancer metastasis: A study of incident cancers during randomised controlled trials. Lancet [Internet]. Elsevier Ltd; 2012;379(9826):1591–601. Available from: http://dx.doi.org/10.1016/S0140-6736(12)60209-8

115. Muto T, Bussey HJ, Morson BC. The evolution of cancer of the colon and rectum. Cancer. 1975 Dec;36(6):2251–70.

116. Baron JA, Cole BF, Sandler RS, Haile RW, Ahnen D, Bresalier R, et al. A Randomized Trial of Aspirin to Prevent Colorectal Adenomas. N Engl J Med. 2003;348(10):891–9.

117. Benamouzig R, Deyra J, Martin A, Girard B, Jullian E, Piednoir B, et al. Daily soluble aspirin and prevention of colorectal adenoma recurrence: One-year results of the APACC trial. Gastroenterology. 2003;125(2):328–36.

118. Sandler S, Halabi S, Baron JA, Budinger S, Paskett E, Keresztes R, Petrelli N, Pipas JM, Karp DD, Loprinzi CL, Steinbach G SR. A Randomized Trial of Aspirin to Prevent Colorectal Adenomas in Patients with Previous Colorectal Cancer. N Engl J Med. 2003;348(10):883–90.

119. Logan RF a, Grainge MJ, Shepherd VC, Armitage NC, Muir KR. Aspirin and Folic Acid for the Prevention of Recurrent Colorectal Adenomas. Gastroenterology. 2008;134(1):29–38.

120. Ishikawa H, Mutoh M, Suzuki S, Tokudome S, Saida Y, Abe T, et al. The preventive effects of low-dose enteric-coated aspirin tablets on the development of colorectal tumours in Asian patients: A randomised trial. Gut. 2014;63(11):1755–9.

121. Cole BF, Logan RF, Halabi S, Benamouzig R, Sandler RS, Grainge MJ, et al. Aspirin for the chemoprevention of colorectal adenomas: Meta-analysis of the randomized trials. Journal of the National Cancer Institute. 2009. p. 256–66.

122. Jafari RM, Sheibani M, Nezamoleslami S, Shayesteh S, Jand Y, Dehpour AR. Drug Repositioning : A Review. J Iran Med Counc. 2018;1(1):7–10.

123. Corbett A, Pickett J, Burns A, Corcoran J, Dunnett SB, Edison P, et al. Drug repositioning for Alzheimer's disease. Nat Rev Drug Discov [Internet]. Nature Publishing Group; 2012;11(11):833–46. Available from: http://dx.doi.org/10.1038/nrd3869

124. Ashburn TT, Thor KB. Drug repositioning: Identifying and developing new uses for existing drugs. Nat Rev Drug Discov. 2004;3(8):673–83.

125. Coyle C, Cafferty FH, Rowley S, MacKenzie M, Berkman L, Gupta S, et al. ADD-ASPIRIN: A phase III, double-blind, placebo controlled, randomised trial assessing the effects of aspirin on disease recurrence and survival after primary therapy in common non-metastatic solid tumours. Contemp Clin Trials [Internet]. The Authors; 2016;51:56–64. Available from: http://dx.doi.org/10.1016/j.cct.2016.10.004

126. Drew DA, Chin SM, Gilpin KK, Parziale M, Pond E, Schuck MM, et al. ASPirin Intervention for the REDuction of colorectal cancer risk (ASPIRED): A study protocol for a randomized controlled trial. Trials. Trials; 2017;18(1):1–11.

127. A Study to Assess the Efficacy and Safety of Enteric-Coated Acetylsalicylic Acid in Patients at Moderate Risk of Cardiovascular Disease (ARRIVE) [Internet]. [cited 2018 Aug 15]. Available from: https://clinicaltrials.gov/ct2/show/NCT00501059

128. Ali R, Toh H-C, Chia W-K, Investigators TA trial. The utility of Aspirin in dukes C and high risk dukes B colorectal cancer - The ASCOLT study: study protocol for a randomized controlled trial. Trials. BioMed Central Ltd; 2011;12(261):1–8.

129. ASPREE Investigator Group. Study design of ASPirin in Reducing Events in the Elderly (ASPREE): A randomized, controlled trial. Contemp Clin Trials. 2013 Nov;36(2):555–64.

130. Burn J, Mathers JC, Bishop DT. Chemoprevention in Lynch syndrome. Fam Cancer. 2013 Dec 24;12(4):707–18.

131. Hull M a, Sandell AC, Montgomery A a, Logan RF, Clifford GM, Rees CJ, et al. A randomized controlled trial of eicosapentaenoic acid and/or aspirin for colorectal adenoma prevention during colonoscopic surveillance in the NHS Bowel Cancer Screening Programme (The seAFOod Polyp Prevention Trial): study protocol for a

randomized cont. Trials [Internet]. Trials; 2013;14(1):237. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3733694&tool=pmcentrez&rendertype=abstract

132. Drew DA, Cao Y, Chan AT. Aspirin and colorectal cancer: the promise of precision chemoprevention. Nat Rev Cancer. Nature Publishing Group; 2016;16:173–86.

133. Hankey GJ, Eikelboom JW. Aspirin resistance. Lancet. 2006;367(9510):606–17.

134. Sciulli MG, Filabozzi P, Tacconelli S, Padovano R, Ricciotti E, Capone ML, et al. Platelet activation in patients with colorectal cancer. Prostaglandins Leukot Essent Fat Acids. 2005;72(2):79–83.

135. Loll PJ, Picot D, Garavito RM. The structural basis of aspirin activity inferred from the crystal structure of inactivated prostaglandin H2 synthase. Nat Struct Biol [Internet]. 1995 Aug [cited 2018 Jun 13];2(8):637–43. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7552725

136. Zidar N, Odar K, Glavač D, Jerše M, Zupanc T, Štajerc D. Cyclooxygenase in normal human tissues - is COX-1 really a constitutive isoform, and COX-2 an inducible isoform? J Cell Mol Med. 2009;13(9b):3753–63.

137. Eberhart CE, Coffey RJ, Radhika A, Giardiello FM, Ferrenbach S, Dubois RN. Up-regulation of cyclooxygenase 2 gene expression in human colorectal adenomas and adenocarcinomas. Gastroenterology. 1994;107(4):1183–8.

138. Gurpinar E, Grizzle WE, Piazza GA. COX-independent mechanisms of cancer chemoprevention by anti-inflammatory drugs. 2013;3(July):1–18.

139. Greenhough A, Smartt HJM, Moore AE, Roberts HR, Williams AC, Paraskeva C, et al. The COX-2/PGE2 pathway: key roles in the hallmarks of cancer and adaptation to the tumour microenvironment. Carcinogenesis [Internet]. 2009;30(3):377–86. Available from: http://www.carcin.oxfordjournals.org/cgi/doi/10.1093/carcin/bgp014

140. North GL. Celecoxib as adjunctive therapy for treatment of colorectal cancer. Ann Pharmacother. 2001;35(12):1638–43.

141. Ng K, Meyerhardt J a., Chan a. T, Sato K, Chan J a., Niedzwiecki D, et al. Aspirin and COX-2 Inhibitor Use in Patients With Stage III Colon Cancer. JNCI J Natl Cancer Inst [Internet]. 2014;107(1):dju345–dju345. Available from: http://jnci.oxfordjournals.org/cgi/doi/10.1093/jnci/dju345

142. Bambace NM, Holmes CE. The platelet contribution to cancer progression. J Thromb Haemost. 2011;9(2):237–49.

143. Gay LJ, Felding-Habermann B. Contribution of platelets to tumour metastasis. Nat Rev Cancer [Internet]. Nature Publishing Group; 2011;11(2):123–34. Available from: http://dx.doi.org/10.1038/nrc3004

144. Guillem-llobat P, Dovizio M, Bruno A, Ricciotti E, Cufino V, Sacco A, et al. Aspirin prevents colorectal cancer metastasis in mice by splitting the crosstalk between platelets and tumor cells. Oncotarget. 2016;[Epub haea.

145. Ying J, Zhou H, Liu P, You Q, Kuang F, Shen Y, et al. Aspirin inhibited the metastasis of colon cancer cells by inhibiting the expression of toll-like receptor 4. Cell Biosci [Internet]. BioMed Central; 2018;8(1):1–9. Available from: https://cellandbioscience.biomedcentral.com/articles/10.1186/s13578-017-0198-7

146. Jung YR, Kim EJ, Choi HJ, Park J-J, Kim H-S, Lee Y-J, et al. Aspirin Targets SIRT1 and AMPK to Induce Senescence of Colorectal Carcinoma Cells. Mol Pharmacol [Internet]. 2015;88(4):708–19. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26219912

147. Luciani MG, Campregher C, Gasche C. Aspirin blocks proliferation in colon cells by inducing a G1 arrest and apoptosis through activation of the checkpoint kinase ATM. Carcinogenesis. 2007;28(10):2207–17.

148. Kornberg R. Chromatin Structure : A Repeating Unit of Histones and DNA Chromatin structure is based on a repeating unit of eight. Science (80-). 1974;184:868–71.

149. Kornberg RD, Lorch Y. Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome. Cell. 1999;98(3):285–94.

150. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature. 1997;389(6648):251–60.

151. Li G, Reinberg D. Chromatin higher-order structures and gene regulation. Curr Opin Genet Dev [Internet]. Elsevier Ltd; 2011;21(2):175–86. Available from: http://dx.doi.org/10.1016/j.gde.2011.01.022

152. Bennett RL, Licht JD. Targeting Epigenetics in Cancer. Annu Rev Pharmacol Toxicol [Internet]. 2018;58:187–207. Available from: http://www.annualreviews.org/doi/10.1146/annurev-pharmtox-010716-105106

153. Cress WD, Seto E. Histone deacetylases, transcriptional control, and cancer. J Cell Physiol [Internet]. 2000;184(1):1–16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10825229

154. Cheng X, Blumenthal RM. Mammalian DNA Methyltransferases: A Structural Perspective. Structure. 2008;16(3):341–50.

155. Ng H-H, Bird A. DNA methylation and chromatin modification. Curr Opin Genet Dev. 1999;9(2):158–63.

156. Kanwal R, Gupta S. Epigenetic modification in cancer. Clin Genet [Internet]. 2012;81(4):303–11. Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1399-0004.2011.01809.x/full

157. Kouzarides T. Chromatin Modifications and Their Function. Cell. 2007;128(4):693–705.

158. Kebede AF, Schneider R, Daujat S. Novel types and sites of histone modifications emerge as players in the transcriptional regulation contest. FEBS J. 2015;282(9):1658–74.

159. Chrun ES, Modolo F, Daniel FI. Histone modifications: A review about the presence of this epigenetic phenomenon in carcinogenesis. Pathol Res Pract [Internet]. Elsevier GmbH.; 2017;213(11):1329–39. Available from: https://doi.org/10.1016/j.prp.2017.06.013

160. Lawrence M, Daujat S, Schneider R. Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends Genet [Internet]. Elsevier Ltd; 2016;32(1):42–56. Available from: http://dx.doi.org/10.1016/j.tig.2015.10.007

161. Carrozza MJ, Utley RT, Workman JL, Côté J. The diverse functions of histone acetyltransferase complexes. Trends Genet. 2003;19(6):321–9.

162. Marks PA, Rifkind RA, Richon VM, Breslow R, Miller T, Kelly WK. Histone deacetylases and cancer: Causes and therapies. Nat Rev Cancer. 2001;1(3):194–202.

163. Yang XJ, Seto E. HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. Oncogene. 2007;26(37):5310–8.

164. Huang T, Lin C, Zhong LLD, Zhao L, Zhang G, Lu A, et al. Targeting histone methylation for colorectal cancer. Therap Adv Gastroenterol. 2017;10(1):114–31.

165. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Res. Nature Publishing Group; 2011;21(3):381–95.

166. Walsh CT, Garneau-Tsodikova S, Gatto GJ. Protein posttranslational modifications: The chemistry of proteome diversifications. Angew Chemie - Int Ed. 2005;44(45):7342–72.

167. Verdin E, Ott M. 50 years of protein acetylation: From gene regulation to epigenetics, metabolism and beyond. Nat Rev Mol Cell Biol [Internet]. Nature Publishing Group; 2015;16(4):258–64. Available from: http://dx.doi.org/10.1038/nrm3931

168. Mukherjee S, Hao Y-H, Orth K. A newly discovered post-translational modification - the acetylation of serine and threonine residues. Trends Biochem Sci. 2007;32(5):210–6.

169. Bird AP. CpG-rich islands and the function of DNA methylation. Nature. 1986;321(6067):209–13.

170. Derek K Lim AH, Maher ER. DNA methylation: a form of epigenetic control of gene expression. Obstet Gynaecol [Internet]. 2010;12(1):37–42. Available from: http://onlinetog.org

171. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics. 2011;6(6):692–702.

172. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet. 2009;41(2):178–86.

173. Naumov VA, Generozov E V., Zaharjevskaya NB, Matushkina DS, Larin AK, Chernyshov S V., et al. Genome-scale analysis of DNA methylation in colorectal cancer using Infinium HumanMethylation450 BeadChips. Epigenetics. 2013;8(9):921–34.

174. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999;99(3):247–57.

175. Hermann A, Goyal R, Jeltsch A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J Biol Chem. 2004;279(46):48350–9.

176. Piccolo FM, Fisher AG. Getting rid of DNA methylation. Trends Cell Biol [Internet]. Elsevier Ltd; 2014;24(2):136–43. Available from: http://dx.doi.org/10.1016/j.tcb.2013.09.001

177. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science (80-). 2009;324(5929):930–5.

178. Ito S, D'Alessio AC, Taranova O V., Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature [Internet]. Nature Publishing Group; 2010;466(7310):1129–33. Available from: http://dx.doi.org/10.1038/nature09303

179. Song C-X, He C. Potential functional roles of DNA demethylation intermediates. Trends Biochem Sci [Internet]. Elsevier Ltd; 2013;38(10):480–4. Available from: http://dx.doi.org/10.1016/j.tibs.2013.07.003

180. Bachman M, Uribe-Lewis S, Yang X, Williams M, Murrell A, Balasubramanian S. 5-Hydroxymethylcytosine is a predominantly stable DNA modification. Nat Chem [Internet]. Nature Publishing Group; 2014;6(12):1049–55. Available from: http://dx.doi.org/10.1038/nchem.2064

181. Branco MR, Ficz G, Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nat Rev Genet [Internet]. Nature Publishing Group; 2011;13(1):7–13. Available from: http://dx.doi.org/10.1038/nrg3080

182. Nan X, Ng H-H, Johnson CA, Laherty CD, Turner BM, Eisenman RN, et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature. 1998;393(6683):386–9.

183. Fournier A, Sasai N, Nakao M, Defossez PA. The role of methyl-binding proteins in chromatin organization and epigenome maintenance. Brief Funct Genomics. 2012;11(3):251–64.

184. Hinoue T, Weisenberger DJ, Lange CPE, Shen H, Byun H-M, Berg D Van Den, et al. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. Genome Res [Internet]. 2012;22(2):271–82. Available from: http://genome.cshlp.org/cgi/doi/10.1101/gr.117523.110\npapers3://publication/doi/10.1101/gr.117523.110

185. Lam K, Pan K, Linnekamp J, Medema JP, Kandimalla R. DNA methylation based biomarkers in colorectal cancer: A systematic review. Biochim Biophys Acta [Internet]. Elsevier B.V.; 2016;1866(1):106–20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27385266

186. Kim MS, Lee J, Sidransky D. DNA methylation markers in colorectal cancer. Cancer Metastasis Rev. 2010;29(1):181–206.

187. Luo Y, Wong CJ, Kaz AM, Dzieciatkowski S, Carter KT, Morris SM, et al. Differences in DNA methylation signatures reveal multiple pathways of progression from adenoma to colorectal cancer. Gastroenterology [Internet]. Elsevier, Inc; 2014;147(2):418–29.e8. Available from: http://dx.doi.org/10.1053/j.gastro.2014.04.039

188. Pinckard RN, Hawkins D, Farr RS. In vitro acetylation of plasma proteins, enzymes and DNA by aspirin. Nature. 1968;219(5149):68–9.

189. Guo Y, Liu Y, Zhang C, Su Z-Y, Li W, Huang M-T, et al. The epigenetic effects of aspirin: The modification of histone H3 lysine 27 acetylation in the prevention of colon carcinogenesis in azoxymethane- and dextran sulfate sodium-treated CF-1 mice. Carcinogenesis. 2016;37(6):616–24.

190. Passacquale G, Phinikaridou A, Warboys C, Cooper M, Lavin B, Alfieri A, Andia ME, Botnar RM FA. Aspirin-induced histone acetylation in endothelial cells enhances synthesis of the secreted isoform of netrin-1 thus inhibiting monocyte vascular infiltration.pdf. Br J Pharmacol. 2015;172(14):3548–64.

191. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. Am J Public Health. 2005;95(SUPPL. 1):144–50.

192. Fedak KM, Bernal A, Capshaw ZA, Gross S. Applying the Bradford Hill criteria in the 21st century: How data integration has changed causal inference in molecular epidemiology. Emerg Themes Epidemiol [Internet]. BioMed Central; 2015;12(14):1–9. Available from: "http://dx.doi.org/10.1186/s12982-015-0037-4

193. Dudbridge F. Polygenic Epidemiology. Genet Epidemiol. 2016;40(4):268–72.

194. Bradford-Hill A. The Environment and Disease: Association or Causation? J R Soc Med. 1965;58(5):295–300.

195. Phillips C V., Goodman KJ. The missed lessons of Sir Austin Bradford Hill. Epidemiol Perspect Innov. 2004;4(1):1–5.

196. Glass TA, Goodman SN, Hernán MA, Samet JM. Causal Inference in Public Health. Annu Rev Public Health. 2013;34:61–75.

197. Colditz GA, Taylor PR. Prevention Trials : Their Place in How We Understand the Value of Prevention Strategies. Annu Rev Public Health. 2010;31:105–20.

198. Yarmolinsky J, Wade KH, Richmond RC, Langdon RJ, Bull CJ, Tilling KM, et al. Causal Inference in Cancer Epidemiology: What Is the Role of Mendelian Randomization? Cancer Epidemiol Biomarkers Prev [Internet]. 2018;995–1011. Available from: http://cebp.aacrjournals.org/lookup/doi/10.1158/1055-9965.EPI-17-1177

199. Davey Smith G, Ebrahim S. “Mendelian randomization”: Can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32(1):1–22.

200. Hennekens CH, Buring JE, Manson JE, Stampfer M, Rosner B, Cook NR, et al. Lack of Effect of Long-Term Supplementation with Beta Carotene on the Incidence of Malignant Neoplasms and Cardiovascular Disease. N Engl J Med [Internet]. 1996;334(18):1145–9. Available from: http://www.nejm.org/doi/abs/10.1056/NEJM199605023341801

201. Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of antioxidant vitamin supplementation in 20 536 high risk individuals: a randomised placebo controlled trial. Lancet. 2002;360(9326):23–33.

202. Peto R, Doll R, Buckley JD, Sporn MB. Can dietary beta-carotene materially reduce human cancer rates? Nature. 1981;290:201–8.

203. Gaziano JM, Manson JE, Buring JE, Hennekens CH. Dietary antioxidants and cardiovascular disease. Ann N Y Acad Sci. 1992;669:249–59.

204. Jha P, Flather M, Lonn E, Farkouh M, Yusuf S. The antioxidant vitamins and cardiovascular disease. A critical review of epidemiologic and clinical trial data. Ann Intern Med [Internet]. 1995;123(11):860–72. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7486470

205. Kushi LH, Folsom AR, Prineas RJ, Mink PJ, Ying W, Bostick RM. Dietary antioxidant vitamins and death from coronary heart disease in postmenopausal women. N Engl J Med. 1996;334(18):1156–62.

206. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 2014;42(Database issue):D1001–6.

207. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet [Internet]. ElsevierCompany.; 2017;101(1):5–22. Available from: http://dx.doi.org/10.1016/j.ajhg.2017.06.005

208. Pranavchand R, Reddy BM. Genomics era and complex disorders: Implications of GWAS with special reference to coronary artery disease, type 2 diabetes mellitus, and cancers. J Postgrad Med. Wolters Kluwer -- Medknow Publications; 2016;62(3):188–98.

209. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008;27(8):1133–63.

210. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics [Internet]. 2016;8(3):389–99. Available from: http://www.futuremedicine.com/doi/10.2217/epi.15.114

211. Richmond RC, Hemani G, Tilling K, Davey Smith G, Relton CL. Challenges and novel approaches for investigating molecular mediation. Hum Mol Genet. 2016;25(R2):R149–56.

212. Wu K. Aspirin and salicylate: An old remedy with a new twist. Circulation. 2000;102(17):2022–3.

213. Paterson JR, Srivastava R, Baxter GJ, Graham AB, Lawrence JR. Salicylic acid content of spices and its implications. J Agric Food Chem. 2006;54(8):2891–6.

214. Hollander D, Dadufalza VD, Fairchild PA. Intestinal absorption of aspirin. Influence of pH, taurocholate, ascorbate, and ethanol. J Lab Clin Med. 1981 Oct;98(4):591–8.

215. Needs CJ, Brooks PM. Clinical Pharmacokinetics of the Salicylates. Clin Pharmacokinet. 1985;10(2):164–77.

216. IARC Working Group on the Evaluation of Cancer-preventive Agents., International Agency for Research on Cancer. Non-steroidal anti-inflammatory drugs. International Agency for Research on Cancer; 1997. 202 p.

217. Zhou G, Marathe GK, Hartiala J, Hazen SL, Allayee H, Wilson Tang WH, et al. Aspirin hydrolysis in plasma is a variable function of butyrylcholinesterase and platelet-activating factor acetylhydrolase 1b2 (PAFAH1b2). J Biol Chem. 2013;288(17):11940–8.

218. Masson P, Lockridge O. Butyrylcholinesterase for protection from organophosphorus poisons: catalytic complexities and hysteretic behavior. Arch Biochem Biophys. 2010 Feb 15;494(2):107–20.

219. Marimuthu S, Chivukula RS V, Alfonso LF, Moridani M, Hagen FK, Bhat GJ. Aspirin acetylates multiple cellular proteins in HCT-116 colon cancer cells: Identification of novel targets. Int J Oncol. 2011;39(5):1273–83.

220. Rang H., Dale MM. H.P Rang, M. Maureen Dale. In: Rang & Dale's Pharmacology. 7th ed. Edinburgh: Elsevier Churchill Livingstone; 2012. p. 113.

221. Agúndez JAG, Martínez C, Pérez-Sala D, Carballo M, Torres MJ, García-Martín E. Pharmacogenomics in aspirin intolerance. Curr Drug Metab [Internet]. 2009;10(9):998–1008. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20214591

222. Lawrence JR, Peter R, Baxter GJ, Robson J, Graham AB, Paterson JR. Urinary excretion of salicyluric and salicylic acids by non-vegetarians, vegetarians, and patients taking low dose aspirin. J Clin Pathol [Internet]. 2003;56(9):651–3. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1770047&tool=pmcentrez&rendertype=abstract

223. Levy G, Tsuchiya T. Salicylate Accumulation Kinetics in Man. N Engl J Med. 1972;287(9):430–2.

224. Santanam N, Parthasarathy S. Aspirin is a substrate for paraoxonase-like activity: Implications in atherosclerosis. Atherosclerosis. 2007 Apr;191(2):272–5.

225. Kuehl GE, Bigler J, Potter JD, Lampe JW. Glucuronidation of the aspirin metabolite salicylic acid by expressed UDP-glucuronosyltransferases and human liver microsomes. Drug Metab Dispos. 2006;34(2):199–202.

226. Testa B, Mayer JM. Hydrolysis in Drug and Prodrug Metabolism : Chemistry, Biochemistry, and Enzymology. VHCA; 2003. 780 p.

227. Nagelschmitz J, Blunck M, Kraetzschmar J, Ludwig M, Wensing G, Hohlfeld T. Pharmacokinetics and pharmacodynamics of acetylsalicylic acid after intravenous and oral administration to healthy volunteers. Clin Pharmacol Adv Appl. 2014;6(1):51–9.

228. Gregor Mendel. Experiments in Plant Hybridization (1865) [Internet]. [cited 2017 Jun 8]. Available from: http://www.mendelweb.org/Mendel.html

229. Yarmolinsky J, Bonilla C, Haycock PC, Langdon RJQ, Lotta LA, Langenberg C, et al. Circulating Selenium and Prostate Cancer Risk: A Mendelian Randomization Analysis. J Natl Cancer Inst [Internet]. 2018;110(9):1–4. Available from: https://academic.oup.com/jnci/advance-article/doi/10.1093/jnci/djy081/4996958

230. Davey Smith G, Ebrahim S. What can Mendelian randomisation tell us about modifiable behavioural and environmental exposures? Br Med J [Internet]. 2005;330(7499):1076–9. Available from: http://www.bmj.com/cgi/doi/10.1136/bmj.330.7499.1076

231. Nitsch D, Molokhia M, Smeeth L, DeStavola BL, Whittaker JC, Leon DA. Limits to Causal Inference based on Mendelian randomization: A Comparison with Randomized Controlled Trials. Am J Epidemiol. 2006;163(5):397–403.

232. Bochud M, Chiolero A, Elston RC, Paccaud F. A cautionary note on the use of Mendelian randomization to infer causation in observational epidemiology. Int J Epidemiol. 2008;37(2):414–6.

233. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. Int J Epidemiol. 2013;42(4):1134–44.

234. Burgess S, Butterworth A, Thompson SG. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data. Genet Epidemiol. 2013;37(7):658–65.

235. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. Epidemiology. 2017;28(1):30–42.

236. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512–25.

237. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40(4):304–14.

238. Burgess S, Thompson SG. Mendelian randomization : Methods for Using Genetic Variants in Causal Estimation. Taylor & Francis Group, editor. CRC Press; 210 p.

239. Lawlor D a. Commentary: Two-sample Mendelian randomization: opportunities and challenges. Int J Epidemiol [Internet]. 2016;45(3):908–15. Available from: https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/dyw127

240. Pasaniuc B, Price AL. Dissecting the genetics of complex traits using summary association statistics. Nat Rev Genet [Internet]. Nature Publishing Group; 2017;18(2):117–27. Available from: http://dx.doi.org/10.1038/nrg.2016.142

241. Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Davey Smith G. Best (but oft-forgotten) practices : the design , analysis , and interpretation of Mendelian randomization studies. Am J Clin Nutr. 2016;103(May):965–78.

242. Zheng J, Baird D, Borges M-C, Bowden J, Hemani G, Haycock P, et al. Recent Developments in Mendelian Randomization Studies. Curr Epidemiol Reports [Internet]. Current Epidemiology Reports; 2017;4(4):330–45. Available from: http://link.springer.com/10.1007/s40471-017-0128-6

243. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, EPIC-InterAct Consortium. Using published data in Mendelian randomization: A blueprint for efficient identification of causal risk factors. Eur J Epidemiol [Internet]. Springer Netherlands; 2015;30(7):543–52. Available from: http://dx.doi.org/10.1007/s10654-015-0011-z

244. Thompson JR, Minelli C, Abrams KR, Tobin MD, Riley RD. Meta-analysis of genetic studies using Mendelian randomization - A multivariate approach. Stat Med. 2005;24(14):2241–54.

245. Bland JM, Altman DG. Statistics notes: Bootstrap resampling methods. BMJ. 2015;350:2–3.

246. Rochowicz Jr JA. Bootstrapping Analysis, Inferential Statistics and EXCEL. Spreadsheets Educ [Internet]. 2010;4(3). Available from: http://epublications.bond.edu.au/ejsie/vol4/iss3/4

247. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46(6):1985–98.

248. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife [Internet]. 2018;7:e34408. Available from: https://elifesciences.org/articles/34408

249. Noreen F, Röösli M, Gaj P, Pietrzak J, Weis S, Urfer P, et al. Modulation of age- and cancer-associated DNA methylation change in the healthy colon by aspirin and lifestyle. J Natl Cancer Inst. 2014;106(7):1–9.

250. Paraskeva C, Finerty S, Mountford RA, Powell SC. Specific cytogenetic abnormalities in two new human colorectal adenoma-derived epithelial cell lines. Cancer Res. 1989;49(5):1282–6.

251. Browne SJ, Williams AC, Hague A, Butt AJ, Paraskeva C. Loss of APC protein expressed by human colonic epithelial cells and the appearance of a specific low-molecular-weight form is associated with apoptosis in vitro. Int J Cancer. 1994;59(1):56–64.

252. Greenhough A, Wallam CA, Hicks DJ, Moorghen M, Williams AC, Paraskeva C. The proapoptotic BH3-only protein Bim is downregulated in a subset of colorectal cancers and is repressed by antiapoptotic COX-2 / PGE 2 signalling in colorectal adenoma cells. Oncogene [Internet]. 2010;29(23):3398–410. Available from: http://dx.doi.org/10.1038/onc.2010.94

253. Baker SJ, Preisinger AC, Jessup JM, Paraskeva C, Markowitz S, Willson JK V, et al. P53 Gene Mutations Occur in Combination With 17P Allelic Deletions As Late Events in Colorectal Tumorigenesis. Cancer Res. 1990;50(23):7717–22.

254. Leibovitz A, Stinson JC, McCombs WB, McCoy CE, Mazur KC, Mabry ND. Classification of human colorectal adenocarcinoma cell lines. Cancer Res. 1976 Dec;36(12):4562–9.

255. Ilyas M, Tomlinson IP, Rowan A, Pignatelli M, Bodmer WF. Beta-catenin mutations in cell lines established from human colorectal cancers. Proc Natl Acad Sci U S A. 1997 Sep 16;94(19):10330–4.

256. Cottu PH, Muzeau F, Estreicher A, Fléjou JF, Iggo R, Thomas G, et al. Inverse correlation between RER+ status and p53 mutation in colorectal cancer cell lines. Oncogene. 1996 Dec 19;13(12):2727–30.

257. Li VSW, Ng SS, Boersema PJ, Low TY, Karthaus WR, Gerlach JP, et al. Wnt Signaling through Inhibition of β-Catenin Degradation in an Intact Axin1 Complex. Cell. 2012 Jun 8;149(6):1245–56.

258. Casares S, Ionov Y, Ge HY, Stanbridge E, Perucho M. The microsatellite mutator phenotype of colon cancer cells is often recessive. Oncogene. 1995 Dec 7;11(11):2303–10.

259. Rodrigues NR, Rowan A, Smith ME, Kerr IB, Bodmer WF, Gannon J V, et al. p53 mutations in colorectal cancer. Proc Natl Acad Sci U S A. 1990 Oct;87(19):7555–9.

260. Hague A, Manning AM, Hanlon KA, Huschtscha LI, Hart D, Paraskeva C. Sodium butyrate induces apoptosis in human colonic tumour cell lines in a p53-independent pathway: Implications for the possible role of dietary fibre in the prevention of large-bowel cancer. Int J Cancer. 1993;55(3):498–505.

261. Sato T, Vries RG, Snippert HJ, van de Wetering M, Barker N, Stange DE, et al. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature [Internet]. Nature Publishing Group; 2009;459(7244):262–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19329995

262. Bibikova M, Le J, Barnes B, Saedinia-melnyk S, Zhou L, Shen R, et al. Genome-wide DNA methylation profiling using Infinium(R) assay. Epigenomics. 2009;1(1):177–200.

263. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol [Internet]. Genome Biology; 2016;17(1):208. Available from: http://dx.doi.org/10.1186/s13059-016-1066-1

264. Min J, Hemani G, Davey Smith G, Relton CL, Suderman M. Meffil: efficient normalisation and analysis of very large DNA methylation samples. Bioinformatics. 2018;1–27.

265. Illumina. Infinium® MethylationEPIC BeadChip [Internet]. [cited 2018 Dec 27]. Available from: https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/humanmethylationepic-data-sheet-1070-2015-008.pdf

266. Pyrosequencing Technology and Platform Overview - QIAGEN [Internet]. [cited 2018 Dec 27]. Available from: https://www.qiagen.com/de/resources/technologies/pyrosequencing-resource-center/technology-overview/

267. Kapoore RV, Vaidyanathan S. Towards quantitative mass spectrometry-based metabolomics in microbial and mammalian systems. Philos Trans R Soc A Math Phys Eng Sci. 2016;374(2079):20150363.

268. Elsworth BL, Mitchell R, Raistrick C, Paternoster L, Hemani G, Gaunt T. MRC IEU UK Biobank GWAS pipeline version 1 [Internet]. 2017. Available from: https://data.bris.ac.uk/data/dataset/2fahpksont1zi26xosyamqo8rr

269. Mitchell R, Hemani G, Dudding T, Paternoster L. UK Biobank Genetic Data: MRC-IEU Quality Control, Version 1 [Internet]. 2017. Available from: https://data.bris.ac.uk/data/dataset/3074krb6t2frj29yh2b03x3wxj

270. Howie B, Marchini J, Stephens M. Genotype Imputation with Thousands of Genomes. Genes|Genomes|Genetics [Internet]. 2011;1(6):457–70. Available from: http://g3journal.org/lookup/doi/10.1534/g3.111.001198

271. Huang J, Howie B, McCarthy S, Memari Y, Walter K, Min JL, et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. Nat Commun [Internet]. Nature Publishing Group; 2015;6:1–9. Available from: http://dx.doi.org/10.1038/ncomms9111

272. Sinclair JK, Taylor PJ, Hobbs SJ. Alpha Level Adjustments for Multiple Dependent Variable Analyses and Their Applicability – A Review. Int J Sport Sci Eng. 2013;07(01):17–20.

273. Colquhoun D. An investigation of the false discovery rate and the misinterpretation of P-values. R Soc Open Sci. 2014;1(3):140216.

274. Bosetti C, Rosato V, Gallus S, Cuzick J, La Vecchia C. Aspirin and cancer risk: A quantitative review to 2011. Ann Oncol. 2012;23(6):1403–15.

275. Cook N. Aspirin and cancer: Evidence from randomised trials. In: NCRI Conference [Internet]. Liverpool; 2012. Available from: http://conference.ncri.org.uk/abstracts/2012/abstracts/Para89.html

276. Chan AT, Ogino S, Fuchs CS. Aspirin use and survival after diagnosis of colorectal cancer. JAMA. 2009;302(6):649–58.

277. Li P, Wu H, Zhang H, Shi Y, Xu J, Ye Y, et al. Aspirin use after diagnosis but not prediagnosis improves established colorectal cancer survival: a meta-analysis. Gut. 2015;64(9):1419–25.

278. Hua X, Phipps AI, Burnett-Hartman AN, Adams S V., Hardikar S, Cohen SA, et al. Timing of Aspirin and Other Nonsteroidal Anti- Inflammatory Drug Use Among Patients With Colorectal Cancer in Relation to Tumor Markers and Survival. J Clin Oncol. 2017;35(24):2806–13.

279. Bains SJ, Mahic M, Myklebust TÅ, Småstuen MC, Yaqub S, Dørum LM, et al. Aspirin As Secondary Prevention in Patients With Colorectal Cancer: An Unselected Population-Based Study. J Clin Oncol. 2016;34(21):2501–8.

280. Gray RT, Coleman HG, Hughes C, Murray LJ, Cardwell CR. Low-dose aspirin use and survival in colorectal cancer: Results from a population-based cohort study. BioMed Cent Cancer. BMC Cancer; 2018;18(1):1–8.

281. McNeil JJ, Woods RL, Nelson MR, Reid CM, Kirpach B, Wolfe R, et al. Effect of Aspirin on Disability-free Survival in the Healthy Elderly. N Engl J Med [Internet]. 2018;NEJMoa1800722. Available from: http://www.nejm.org/doi/10.1056/NEJMoa1800722

282. Bastiaannet E, Sampieri K, Dekkers OM, Craen AJM De, Herk-sukel MPP Van, Lemmens V. Use of Aspirin postdiagnosis improves survival for colon cancer patients. Br J Cancer [Internet]. Nature Publishing Group; 2012;106(9):1564–70. Available from: http://dx.doi.org/10.1038/bjc.2012.101

283. McCowan C, Munro AJ, Donnan PT, Steele RJC. Use of aspirin post-diagnosis in a cohort of patients with colorectal cancer and its association with all-cause and colorectal cancer specific mortality. Eur J Cancer [Internet]. 2013;49(5):1049–57. Available from: http://dx.doi.org/10.1016/j.ejca.2012.10.024

284. Kasmeridis C, Apostolakis S, Lip GYH. Aspirin and aspirin resistance in coronary artery disease. Curr Opin Pharmacol [Internet]. Elsevier Ltd; 2013;13(2):242–50. Available from: http://dx.doi.org/10.1016/j.coph.2012.12.004

285. Tran HA, Anand SS, Hankey GJ, Eikelboom JW. Aspirin resistance. Thromb Res. 2007;120(3):337–46.

286. Alexander JH, Harrington RA, Tuttle RH, Berdan LG, Lincoff AM, Deckers JW, et al. Prior Aspirin Use Predicts Worse Outcomes in Patients with Non – ST- Elevation Acute Coronary Syndromes. Am J Cardiol. 1999;83(8):1147–51.

287. Santopinto J, Gurfinkel EP, Torres V, Marcos E, Bozovich GE, Mautner B, et al. Prior aspirin users with acute non-ST-elevation coronary syndromes are at increased risk of cardiac events and benefit from enoxaparin. Am Heart J. 2001;141(4):566–72.

288. Du G, Lin Q, Wang J. A brief review on the mechanisms of aspirin resistance. Int J Cardiol. Elsevier Ireland Ltd; 2016;220:21–6.

289. Riedl A, Schlederer M, Pudelko K, Stadler M, Walter S, Unterleuthner D, et al. Comparison of cancer cells in 2D vs 3D culture reveals differences in AKT–mTOR–S6K signaling and drug responses. J Cell Sci [Internet]. 2017;130(1):203–18. Available from: http://jcs.biologists.org/lookup/doi/10.1242/jcs.188102

290. Dolznig H, Rupp C, Puri C, Haslinger C, Schweifer N, Wieser E, et al. Modeling Colon Adenocarcinomas in Vitro: A 3D Co-Culture System Induces Cancer-Relevant Pathways upon Tumor Cell and Stromal Fibroblast Interaction. Am J Pathol [Internet]. Elsevier Inc.; 2011;179(1):487–501. Available from: http://dx.doi.org/10.1016/j.ajpath.2011.03.015

291. Artegiani B, Clevers H. Use and application of the 3D-organoid technology. Hum Mol Genet. 2018;0(0):1–9.

292. Takagi A, Watanabe M, Ishii Y, Morita J, Hirokawa Y, Matsuzaki T, et al. Three-dimensional Cellular Spheroid Formation Provides Human Prostate Tumor Cells with Tissue-like Features. Anticancer Res. 2007;27(1 A):45–54.

293. Yue X, Lukowski JK, Weaver EM, Skube SB, Hummon AB. Quantitative Proteomic and Phosphoproteomic Comparison of 2D and 3D Colon Cancer Cell Culture Models. J Proteome Res. 2016;15(12):4265–76.

294. Breslin S, O'Driscoll L. Three-dimensional cell culture: The missing link in drug discovery. Drug Discov Today [Internet]. Elsevier Ltd; 2013;18(5-6):240–9. Available from: http://dx.doi.org/10.1016/j.drudis.2012.10.003

295. Lv D, Hu Z, Lu L, Lu H, Xu X. Three-dimensional cell culture: A powerful tool in tumor research and drug discovery. Oncol Lett. 2017;14(6):6999–7010.

296. Kleinman HK, Martin GR. Matrigel: Basement membrane matrix with biological activity. Semin Cancer Biol. 2005;15(5 SPEC. ISS.):378–86.

297. Knight E, Przyborski S. Advances in 3D cell culture technologies enabling tissue-like structures to be created in vitro. J Anat. 2015;227(6):746–56.

298. Sato T, Stange DE, Ferrante M, Vries RGJ, Van Es JH, Van Den Brink S, et al. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. Gastroenterology [Internet]. Elsevier Inc.; 2011;141(5):1762–72. Available from: http://dx.doi.org/10.1053/j.gastro.2011.07.050

299. Elder DJE, Hague A, Hicks DJ, Paraskeva C. Differential Growth Inhibition by the Aspirin Metabolite Salicylate in Human Colorectal Tumor Cell Lines : Enhanced Apoptosis in Carcinoma and in Vitro -transformed Adenoma Relative to Adenoma Cell Lines. Cancer Res. 1996;56:2273–6.

300. Caserta TM, Smith AN, Gultice AD, Reedy MA, Brown TL. Q-VD-OPh, a broad spectrum caspase inhibitor with potent antiapoptotic properties. Apoptosis. 2003;8(4):345–52.

301. Williams AC, Miller JC, Collard TJ, Bracey TS, Cosulich S, Paraskeva C. Mutant p53 is not fully dominant over endogenous wild type p53 in a colorectal adenoma cell line as demonstrated by induction of MDM2 protein and retention of a p53 dependent G1 arrest after gamma irradiation. Oncogene [Internet]. 1995 Jul 6 [cited 2017 Jun 9];11(1):141–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7624121

302. Sant S, Johnston PA. The production of 3D tumor spheroids for cancer drug discovery. Drug Discov Today Technol [Internet]. Elsevier Ltd; 2017;23:27–36. Available from: http://dx.doi.org/10.1016/j.ddtec.2017.03.002

303. Howes AL, Richardson RD, Finlay D, Vuori K. 3-Dimensional Culture Systems for Anti-Cancer Compound Profiling and High-Throughput Screening Reveal Increases in EGFR Inhibitor-Mediated Cytotoxicity Compared to Monolayer Culture Systems. PLoS One. 2014;9(9):1–11.

304. Cree I a., Charlton P. Molecular chess? Hallmarks of anti-cancer drug resistance. BMC Cancer [Internet]. BMC Cancer; 2017;17(10):1–8. Available from: http://dx.doi.org/10.1186/s12885-016-2999-1

305. Greenhough A, Smartt HJM, Moore AE, Roberts HR, Williams AC, Paraskeva C, et al. The COX-2/PGE2 pathway: key roles in the hallmarks of cancer and adaptation to the tumour microenvironment.pdf. Carcinogenesis. 2009;30(3):377–86.

306. Tosco P, Lazzarato L. Mechanistic insights into cyclooxygenase irreversible inactivation by aspirin. ChemMedChem. 2009;4(6):939–45.

307. Bak AW, McKnight W, Li P, Soldato P Del, Calignano A, Cirino G, et al. Cyclooxygenase-independent chemoprevention with an aspirin derivative in a rat model of colonic adenocarcinoma. Life Sci. 1998;62(23):PL 367–73.

308. Waskewich C, Blumenthal RD, Li H, Stein R, Goldenberg DM, Burton J. Celecoxib exhibits the greatest potency amongst cyclooxygenase (COX) inhibitors for growth inhibition of COX-2-negative hematopoietic and epithelial cell lines. Cancer Res. 2002;62(7):2029–33.

309. Kopp E, Ghosh S. Inhibition of NF-kappa B by sodium salicylate and aspirin. Science (80-) [Internet]. 1994;265(5174):956–9. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.8052854

310. McCarty MF, Block KI. Preadministration of high-dose salicylates, suppressors of NF-kappaB activation, may increase the chemosensitivity of many cancers: an example of proapoptotic signal modulation therapy. Integr Cancer Ther. 2006;5(3):252–68.

311. Law BK, Waltner-Law ME, Entingh AJ, Chytil A, Aakre ME, Nørgaard P, et al. Salicylate-induced Growth Arrest Is Associated with Inhibition of p70s6k and Down-regulation of c-Myc, Cyclin D1, Cyclin A, and Proliferating Cell Nuclear Antigen. J Biol Chem. 2000;275(49):38261–7.

312. Bos CL, Kodach LL, van den Brink GR, Diks SH, van Santen MM, Richel DJ, et al. Effect of aspirin on the Wnt/b-catenin pathway is mediated via protein phosphatase 2A. Oncogene. 2006;25(49):6447–56.

313. Bellosillo B, Piqué M, Barragán M, Castaño E, Villamor N, Colomer D, et al. Aspirin and salicylate induce apoptosis and activation of caspases in B-cell chronic lymphocytic leukemia cells. Blood [Internet]. 1998;92(4):1406–14. Available from: http://www.bloodjournal.org/content/92/4/1406.abstract

314. Zimmermann KC, Waterhouse NJ, Goldstein JC, Schuler M, Green DR. Aspirin Induces Apoptosis through Release of Cytochrome c from Mitochondria. Neoplasia [Internet]. 2000;2(6):505–13. Available from: http://www.sciencedirect.com/science/article/pii/S147655860080041X

315. Dikshit P, Chatterjee M, Goswami A, Mishra A, Jana NR. Aspirin Induces Apoptosis through the Inhibition of Proteasome Function. J Biol Chem. 2006;281(39):29228–35.

316. Núñez L, Valero RA, Senovilla L, Sanz-Blasco S, García-Sancho J, Villalobos C. Cell proliferation depends on mitochondrial Ca2+ uptake: inhibition by salicylate. J Physiol. 2006;571(Pt 1):57–73.

317. Castoldi F, Pietrocola F, Maiuri MC, Kroemer G. Aspirin induces autophagy via inhibition of the acetyltransferase. Oncotarget. 2018;9(37):24574–5.

318. Wang H, Liu B, Wang J, Li J, Gong Y, Li S, et al. Reduction of NANOG Mediates the Inhibitory Effect of Aspirin on Tumor Growth and Stemness in Colorectal Cancer. Cell Physiol Biochem. 2017;44(3):1051–63.

319. Dachineni R, Kumar DR, Callegari E, Kesharwani SS, Sankaranarayanan R, Seefeldt T, et al. Salicylic acid metabolites and derivatives inhibit CDK activity: Novel insights into aspirin's chemopreventive effects against colorectal cancer. Int J Oncol. 2017;51(6):1661–71.

320. Ai G, Dachineni R, Kumar DR, Alfonso LF, Marimuthu S, Bhat GJ. Aspirin inhibits glucose-6-phosphate dehydrogenase activity in HCT 116 cells through acetylation: Identification of aspirin-acetylated sites. Mol Med Rep. 2016;14(2):1726–32.

321. Ai G, Dachineni R, Kumar DR, Marimuthu S, Alfonso LF, Bhat GJ. Aspirin acetylates wild type and mutant p53 in colon cancer cells: identification of aspirin acetylated sites on recombinant p53. Tumor Biol [Internet]. Tumor Biology; 2016;37(5):6007–16. Available from: http://dx.doi.org/10.1007/s13277-015-4438-3

322. Guo H, Liu J, Ben Q, Qu Y, Li M, Wang Y, et al. The aspirin-induced long non-coding RNA OLA1P2 blocks phosphorylated STAT3 homodimer formation. Genome Biol [Internet]. Genome Biology; 2016;17(24):1–15. Available from: http://genomebiology.com/2016/17/1/24

323. Drazic A, Myklebust LM, Ree R, Arnesen T. The world of protein acetylation. Biochim Biophys Acta - Proteins Proteomics [Internet]. The Authors; 2016;1864(10):1372–401. Available from: http://dx.doi.org/10.1016/j.bbapap.2016.06.007

324. Aksnes H, Van Damme P, Goris M, Starheim KK, Marie M, Støve SI, et al. An Organellar Nα-Acetyltransferase, Naa60, Acetylates Cytosolic N Termini of Transmembrane Proteins and Maintains Golgi Integrity. Cell Rep. 2015;10(8):1362–74.

325. Glozak MA, Sengupta N, Zhang X, Seto E. Acetylation and deacetylation of non-histone proteins. Gene. 2005;363(1-2):15–23.

326. Bjornsson TD, Schneider DE, Berger H. Aspirin acetylates fibrinogen and enhances fibrinolysis. Fibrinolytic effect is independent of changes in plasminogen activator levels. J Pharmacol Exp Ther. 1989;250(1):154–61.

327. Bridges KR, Schmidt GJ, Jensen M, Cerami A, Bunn HF. The Acetylation of Hemoglobin by Aspirin. In vitro and in vivo. J Clin Invest. 1975;56(1):201–7.

328. Alfonso LF, Srivenugopal KS, Arumugam T V., Abbruscato TJ, Weidanz JA, Bhat GJ. Aspirin inhibits camptothecin-induced p21CIP1 levels and potentiates apoptosis in human breast cancer cells. Int J Oncol. 2009;34(3):597–608.

329. Fernandez HR, Lindén SK. The aspirin metabolite salicylate inhibits lysine acetyltransferases and MUC1 induced epithelial to mesenchymal transition. Sci Rep. 2017;7(1):1–12.

330. Slattery ML, Wolff RK, Herrick J, Caan BJ, Samowitz W. Tumor markers and rectal cancer: Support for an inflammation-related pathway. International Journal of Cancer. 2009. p. 1698–704.

331. Li X, Zhao K, Ma N, Sun S, Miao Z, Zhao Z. Association of ABCB1 promoter methylation with aspirin exposure, platelet function, and clinical outcomes in Chinese intracranial artery stenosis patients. Eur J Clin Pharmacol. European Journal of Clinical Pharmacology; 2017;73(10):1261–9.

332. Tazi J, Bird A. Alternative chromatin structure at CpG islands. Cell. 1990;60(6):909–20.

333. Rose NR, Klose RJ. Understanding the relationship between DNA methylation and histone lysine methylation. Biochim Biophys Acta - Gene Regul Mech [Internet]. Elsevier B.V.; 2014;1839(12):1362–72. Available from: http://dx.doi.org/10.1016/j.bbagrm.2014.02.007

334. Sheffield NC, Pierron G, Klughammer J, Datlinger P, Schönegger A, Schuster M, et al. DNA methylation heterogeneity defines a disease spectrum in Ewing sarcoma. Nat Med [Internet]. Nature Publishing Group; 2017;23(3):386–95. Available from: http://dx.doi.org/10.1038/nm.4273

335. Heyn H, Vidal E, Ferreira HJ, Vizoso M, Sayols S, Gomez A, et al. Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer. Genome Biol [Internet]. Genome Biology; 2016;17(11):1–16. Available from: http://dx.doi.org/10.1186/s13059-016-0879-2

336. González B, Jayanthi S, Gomez N, Torres O V., Sosa MH, Bernardi A, et al. Repeated methamphetamine and modafinil induce differential cognitive effects and specific histone acetylation and DNA methylation profiles in the mouse medial prefrontal cortex. Prog Neuro-Psychopharmacology Biol Psychiatry. 2018;82:1–11.

337. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. Epigenomics. 2011;3(6):771–84.

338. Weisenberger DJ, Berg D Van Den, Pan F, Berman BP, Laird PW. Comprehensive DNA Methylation Analysis on the Illumina® Infinium® Assay Platform. Appl Note Illumina Epigenetic Anal. 2008;

339. Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, et al. High-throughput DNA methylation profiling using universal bead arrays. Genome Res. 2006;16(3):383–93.

340. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics [Internet]. 2010;11(587):1–9. Available from: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-587

341. Leenen F a. D, Muller CP, Turner JD. DNA methylation: conducting the orchestra from exposure to phenotype? Clin Epigenetics [Internet]. Clinical Epigenetics; 2016;8(92). Available from: http://clinicalepigeneticsjournal.biomedcentral.com/articles/10.1186/s13148-016-0256-8

342. Michels KB, Binder AM, Dedeurwaerder S, Epstein CB, Greally JM, Gut I, et al. Recommendations for the design and analysis of epigenome-wide association studies. Nat Methods [Internet]. 2013;10(10):949–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24076989

343. Benajmini Y, Hochberg Y. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B. 1995;57(1):289–300.

344. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci [Internet]. 2003;100(16):9440–5. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1530509100

345. Trinkle-Mulcahy L, Boulon S, Lam YW, Urcia R, Boisvert F-M, Vandermoere F, et al. Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. J Cell Biol. 2008;183(2):223–39.

346. Greenhough A, Bagley C, Heesom KJ, Gurevich DB, Gay D, Bond M, et al. Cancer cell adaptation to hypoxia involves a HIF-GPRC5A-YAP axis. EMBO Mol Med [Internet]. 2018;10(9):e8699. Available from: http://embomolmed.embopress.org/content/embomm/early/2018/08/23/emmm.201708699.full.pdf

347. Sabates-Bellver J, Van der Flier LG, de Palo M, Cattaneo E, Maake C, Rehrauer H, et al. Transcriptome profile of human colorectal adenomas. Mol Cancer Res. 2007;5(12):1263–75.

348. Home - GEO DataSets - NCBI [Internet]. [cited 2018 Sep 26]. Available from: https://www.ncbi.nlm.nih.gov/gds

349. Jones SR, Carley S, Harrison M. An introduction to power and sample size estimation. Emerg Med J [Internet]. 2003;20(5):453–8. Available from: http://emj.bmj.com/cgi/doi/10.1136/emj.20.5.453

350. Naeem H, Wong NC, Chatterton Z, Hong MKH, Pedersen JS, Corcoran NM, et al. Reducing the risk of false discovery enabling identification of biologically significant genome-wide methylation status using the HumanMethylation450 array. BMC Genomics [Internet]. 2014;15(51):1–15. Available from: http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-51

351. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. Behav Brain Res. 2001;125(1-2):279–84.

352. Universität Düsseldorf: G*Power [Internet]. [cited 2018 Dec 29]. Available from: http://www.gpower.hhu.de/

353. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv [Internet]. 2014;1–2. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0002929707613524

354. SILAC - Stable isotope labeling by amino acids in cell culture [Internet]. [cited 2015 Sep 21]. Available from: http://www.silac.org/

355. Yang W, Chung YG, Kim Y, Kim T-K, Keay SK, Zhang C-O, et al. Quantitative proteomics identifies a beta-catenin network as an element of the signaling response to Frizzled-8 protein-related antiproliferative factor. Mol Cell Proteomics [Internet]. 2011;10(6):M110.007492. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3108841&tool=pmcentrez&rendertype=abstract

356. Boulon S, Ahmad Y, Trinkle-Mulcahy L, Verheggen C, Cobley A, Gregor P, et al. Establishment of a Protein Frequency Library and Its Application in the Reliable Identification of Specific Protein Interaction Partners. Mol Cell Proteomics [Internet]. 2010;9(5):861–79. Available from: http://www.mcponline.org/cgi/doi/10.1074/mcp.M900517-MCP200

357. Morey JS, Ryan JC, Dolah FM. Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR. Biol Proced Online [Internet]. 2006;8(1):175–93. Available from: http://www.springerlink.com/index/10.1251/bpo126

358. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43(Database issue):D447–52.

359. Zhang H-Q, Guo X, Guo S-Q, Wang Q, Chen X-Q, Li X-N, et al. STMN1 in colon cancer: expression and prognosis in Chinese patients. Eur Rev Med Pharmacol Sci [Internet]. 2016 May [cited 2017 Jun 9];20(10):2038–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27249602

360. Tan HT, Wu W, Ng YZ, Zhang X, Yan B, Ong CW, et al. Proteomic analysis of colorectal cancer metastasis: Stathmin-1 revealed as a player in cancer cell migration and prognostic marker. J Proteome Res. 2012;11(2):1433–45.

361. Buescher JM, Driggers EM. Integration of omics: more than the sum of its parts. Cancer Metab [Internet]. Cancer & Metabolism; 2016;4(1):4. Available from: http://www.cancerandmetabolism.com/content/4/1/4

362. Machado-Neto JA, Saad STO, Traina F. Stathmin 1 in normal and malignant hematopoiesis. BMB Rep. 2014;47(12):660–5.

363. Belletti B, Baldassarre G. Stathmin: a protein with many tasks. New biomarker and potential target in cancer. Expert Opin Ther Targets. 2011;15(11):1249–66.

364. Belmont LD, Mitchison TJ. Identification of a Protein That Interacts with Tubulin Dimers and Increases the Catastrophe Rate of Microtubules. Cell. 1996;84(4):623–31.

365. Jourdain L, Curmi P, Sobel A, Pantaloni D, Carlier M-F. Stathmin: A Tubulin-Sequestering Protein Which Forms a Ternary T2S Complex with Two Tubulin Molecules. Biochemistry. 1997;36(36):10817–21.

366. Larsson N, Marklund U, Gradin HM, Brattsand G, Gullberg M. Control of microtubule dynamics by oncoprotein 18: dissection of the regulatory role of multisite phosphorylation during mitosis. Mol Cell Biol. 1997;17(9):5530–9.

367. Carney BK, Caruso Silva V, Cassimeris L. The microtubule cytoskeleton is required for a G2 cell cycle delay in cancer cells lacking stathmin and p53. Cytoskeleton. 2012;69(5):278–89.

368. Suzuki K, Watanabe A, Araki K, Yokobori T, Harimoto N, Gantumur D, et al. High STMN1 Expression Is Associated with Tumor Differentiation and Metastasis in Clinical Patients with Pancreatic Cancer. Anticancer Res. 2018 Jan 20;38(2):939–44.

369. Li Y, Liang Y, Sang Y, Song X, Zhang H, Liu Y, et al. MiR-770 suppresses the chemo-resistance and metastasis of triple negative breast cancer via direct targeting of STMN1. Cell Death Dis [Internet]. Springer US; 2018;9(1):1–12. Available from: http://dx.doi.org/10.1038/s41419-017-0030-7

370. Tan HT, Chung MCM. Label-Free Quantitative Phosphoproteomics Reveals Regulation of Vasodilator-Stimulated Phosphoprotein upon Stathmin-1 Silencing in a Pair of Isogenic Colorectal Cancer Cell Lines. Proteomics. 2018;18(8):1–13.

371. Dominguez R. Actin filament nucleation and elongation factors structure function relationships. Crit Rev Biochem Mol Biol. 2009;44(6):351–66.

372. Tao Y, Chen YC, Wang Y, Zhang ZJ, Xu WR. Phosphorylated vasodilator-stimulated phosphoprotein is localized on mitotic spindles of the gastric cancer cell line SGC-7901. World J Gastroenterol. 2006;12(46):7478–81.

373. Panopoulos A, Howell M, Fotedar R, Margolis RL. Glioblastoma motility occurs in the absence of actin polymer. Mol Biol Cell [Internet]. 2011;22(13):2212–20. Available from: http://www.molbiolcell.org/cgi/doi/10.1091/mbc.E10-10-0849

374. Probst A V., Dunleavy E, Almouzni G. Epigenetic inheritance during the cell cycle. Nat Rev Mol Cell Biol. 2009;10(3):192–206.

375. Chen T, Hevi S, Gay F, Tsujimoto N, He T, Zhang B, et al. Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. Nat Genet. 2007;39(3):391–6.

376. Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. Nature. 2013;502(7472):472–9.

377. Costa A, Onesti S. The MCM complex: (just) a replicative helicase? Biochem Soc Trans [Internet]. 2008;36(Pt 1):136–40. Available from: http://biochemsoctrans.org/lookup/doi/10.1042/BST0360136

378. Kwon HJ, Hong YK, Park C, Choi YH, Yun HJ, Lee EW, et al. Widdrol induces cell cycle arrest, associated with MCM down-regulation, in human colon adenocarcinoma cells. Cancer Lett [Internet]. Elsevier Ireland Ltd; 2010;290(1):96–103. Available from: http://dx.doi.org/10.1016/j.canlet.2009.09.003

379. Bullman S, Pedamallu CS, Sicinska E, Clancy TE, Zhang X, Cai D, et al. Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer. Science (80-) [Internet]. 2017;358(6369):1443–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29170280%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5823247

380. Tahara T, Shibata T, Nakamura M, Yamashita H, Yoshioka D, Okubo M, et al. Chronic aspirin use suppresses CDH1 methylation in human gastric mucosa. Dig Dis Sci. 2010;55(1):54–9.

381. Michigami Y, Watari J, Ito C, Hara K, Yamasaki T, Kondo T, et al. Effects of long-term aspirin use on molecular alterations in precancerous gastric mucosa in patients with and without gastric cancer. Sci Rep. 2017;7(13384):1–9.

382. Tahara T, Shibata T, Yamashita H, Nakamura M, Yoshioka D, Okubo M, et al. Chronic nonsteroidal anti-inflammatory drug (NSAID) use suppresses multiple CpG islands hyper methylation (CIHM) of tumor suppressor genes in the human gastric mucosa. Cancer Science. 2009. p. 1192–7.

383. Wilson LE, Kim S, Xu Z, Harlid S, Sandler DP, Taylor JA. Non-Steroidal Anti-Inflammatory Drug Use and Genomic DNA Methylation in Blood. PLoS One. 2015;10(9):1–8.

384. Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol [Internet]. 2015;16(22). Available from: http://genomebiology.com/2015/16/1/22

385. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature [Internet]. 2012;489(7414):57–74. Available from: http://www.nature.com/doifinder/10.1038/nature11247

386. Siggens L, Ekwall K. Epigenetics, chromatin and genome organization: Recent advances from the ENCODE project. J Intern Med. 2014;276(3):201–14.

387. Devlin B, Roeder K. Genomic Control for Association Studies. Biometrics [Internet]. 1999;55(4):997–1004. Available from: http://doi.wiley.com/10.1111/j.0006-341X.1999.00997.x

388. Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan S-T, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. Genome Biol [Internet]. 2015;16(37):1–12. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0600-x

389. Van Iterson M, van Zwet EW, the BIOS Consortium, Heijmans BT. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. Genome Biol [Internet]. Genome Biology; 2017;18(19):1–13. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-1131-9

390. Wang J, Zhao Q, Hastie T, Owen AB. Confounder Adjustment in Multiple Hypothesis Testing. arXiv:150804178 [Internet]. 2015;1–52. Available from: http://arxiv.org/abs/1508.04178

391. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet [Internet]. Nature Publishing Group; 2010;11(10):733–9. Available from: http://www.nature.com/doifinder/10.1038/nrg2825

392. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. 2007;3(9):1724–35.

393. Friguet C, Kloareg M, Causeur D. A Factor Model Approach to Multiple Testing Under Dependence. J Am Stat Assoc. 2009;104(488):1406–15.

394. Urdan TC. Standardization and z Scores. In: Statistics in Plain English. 2016. p. 43–55.

395. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: A guide for non-statisticians. Int J Endocrinol Metab. 2012;10(2):486–9.

396. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661–78.

397. Teschendorff AE, Zhuang J, Widschwendter M. Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. Bioinformatics. 2011;27(11):1496–505.

398. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28(6):882–3.

399. Herman JG, Baylin SB. Gene Silencing in Cancer in Association with Promoter Hypermethylation. N Engl J Med [Internet]. 2003;349(21):2042–54. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMra023075

400. Mayr C. Regulation by 3 `-Untranslated Regions. Annu Rev Genet Vol 51. 2017;51:171–94.

401. Fang JY, Richardson BC. The MAPK signalling pathways and colorectal cancer. Lancet Oncol. 2005;6(5):322–7.

402. West NR, McCuaig S, Franchini F, Powrie F. Emerging cytokine networks in colorectal cancer. Nat Rev Immunol [Internet]. Nature Publishing Group; 2015;15(10):615–29. Available from: http://dx.doi.org/10.1038/nri3896

403. Vigneri PG, Tirrò E, Pennisi MS, Massimino M, Stella S, Romano C, et al. The Insulin/IGF System in Colorectal Cancer Development and Resistance to Therapy. Front Oncol [Internet]. 2015;5:1–7. Available from: http://journal.frontiersin.org/Article/10.3389/fonc.2015.00230/abstract

404. Danielsen SA, Eide PW, Nesbakken A, Guren T, Leithe E, Lothe R a. Portrait of the PI3K/AKT pathway in colorectal cancer. Biochim Biophys Acta - Rev Cancer [Internet]. Elsevier B.V.; 2015;1855(1):104–21. Available from: http://dx.doi.org/10.1016/j.bbcan.2014.09.008

405. Tost J, Gut IG. DNA methylation analysis by pyrosequencing. Nat Protoc. 2007;2(9):2265–75.

406. Jaffe AE, Hyde T, Kleinman J, Weinberger DR, Chenoweth JG, McKay RD, et al. Practical impacts of genomic data "cleaning" on biological discovery using surrogate variable analysis. BMC Bioinformatics [Internet]. BMC Bioinformatics; 2015;16(372):1–10. Available from: http://dx.doi.org/10.1186/s12859-015-0808-5

407. Choi JK, Bae J-B, Lyu J, Kim T-Y, Kim Y-J. Nucleosome deposition and DNA methylation at coding region boundaries. Genome Biol. 2009;10(9):R89.

408. Suzuki H, Maruyama R, Yamamoto E, Kai M. DNA methylation and microRNA dysregulation in cancer. Mol Oncol [Internet]. Elsevier B.V; 2012;6(6):567–78. Available from: http://dx.doi.org/10.1016/j.molonc.2012.07.007

409. Ma X, Wang Y, Zhang MQ, Gazdar AF. DNA methylation data analysis and its application to cancer research. Epigenomics. 2013;5(3):301–16.

410. Klett H, Balavarca Y, Toth R, Gigic B, Habermann N, Scherer D, et al. Robust prediction of gene regulation in colorectal cancer tissues from DNA methylation profiles. Epigenetics [Internet]. Taylor & Francis; 2018;1–12. Available from: https://doi.org/10.1080/15592294.2018.1460034

411. McInnes T, Zou D, Rao DS, Munro FM, Phillips VL, McCall JL, et al. Genome-wide methylation analysis identifies a core set of hypermethylated genes in CIMP-H colorectal cancer. BioMed Cent Cancer. BMC Cancer; 2017;17(1):1–11.

412. Mansoori B, Mohammadi A, Davudian S, Shirjang S, Baradaran B. The different mechanisms of cancer drug resistance: A brief review. Adv Pharm Bull [Internet]. 2017;7(3):339–48. Available from: http://dx.doi.org/10.15171/apb.2017.041

413. Chen C-C, Lee K-D, Pai M-Y, Chu P-Y, Hsu C-C, Chiu C-C, et al. Changes in DNA methylation are associated with the development of drug resistance in cervical cancer cells. Cancer Cell Int [Internet]. BioMed Central; 2015;15:98. Available from: http://www.cancerci.com/content/15/1/98

414. Terai H, Soejima K, Yasuda H, Nakayama S, Hamamoto J, Arai D, et al. Activation of the FGF2-FGFR1 Autocrine Pathway: A Novel Mechanism of Acquired Resistance to Gefitinib in NSCLC. Mol Cancer Res [Internet]. 2013;11(7):759–67. Available from: http://mcr.aacrjournals.org/cgi/doi/10.1158/1541-7786.MCR-12-0652

415. Terai H, Soejima K, Yasuda H, Sato T, Naoki K, Ikemura S, et al. Long-term exposure to gefitinib induces acquired resistance through DNA methylation changes in the EGFR-mutant PC9 lung cancer cell line. Int J Oncol. 2015;46(1):430–6.

416. Kurdyukov S, Bullock M. DNA Methylation Analysis: Choosing the Right Method. Biol [Internet]. 2016;5(1):3. Available from: http://www.mdpi.com/2079-7737/5/1/3

417. Swerdlow DI, Kuchenbaecker KB, Shah S, Sofat R, Holmes M V, White J, et al. Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. Int J Epidemiol [Internet]. 2016;45(5):1600–16. Available from: https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/dyw088

418. Klein EA, Thompson IM, Tangen CM, Crowley JJ, Lucia MS, Goodman PJ, et al. Vitamin E and the Risk of Prostate Cancer: The Selenium and Vitamin E Cancer Prevention Trial (SELECT). Jama [Internet]. 2011;306(14):1549–56. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2011.1437

419. Helzlsouer KJ, Huang HY, Alberg AJ, Hoffman S, Burke A, Norkus EP, et al. Association between alpha-tocopherol, gamma-tocopherol, selenium, and subsequent prostate cancer. J Natl Cancer Inst [Internet]. 2000 Dec 20;92(24):2018–23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11121464

420. Nomura AMY, Lee J, Stemmermann GN, Combs, Gerald F J. Serum selenium and subsequent risk of prostate cancer. Cancer Epidemiol Biomarkers Prev. 2000;9(9):883–7.

421. Li H, Stampfer MJ, Giovannucci EL, Morris JS, Willett WC, Gaziano JM, et al. A prospective study of plasma selenium levels and prostate cancer risk. J Natl Cancer Inst [Internet]. 2004 May 5;96(9):696–703. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15126606

422. Lippman SM, Klein EA, Goodman PJ, Lucia MS, Thompson IM, Ford LG, et al. Effect of Selenium and Vitamin E on Risk of Prostate Cancer and Other Cancers: The Selenium and Vitamin E Cancer Prevention Trial (SELECT). JAMA. 2009;301(1):39–51.

423. Ma Y, Zhang P, Wang F, Yang J, Liu Z, Qin H. Association between vitamin D and risk of colorectal cancer: A systematic review of prospective studies. J Clin Oncol. 2011;29(28):3775–82.

424. He Y, Timofeeva M, Farrington SM, Vaughan-Shaw P, Svinti V, Walker M, et al. Exploring causality in the association between circulating 25-hydroxyvitamin D and colorectal cancer risk: A large Mendelian randomisation study. BMC Med. BMC Medicine; 2018;16(142):1–11.

425. Bardou M, Barkun AN, Martel M. Republished: Obesity and colorectal cancer. Postgrad Med J. 2013 Sep 16;89(1055):519–33.

426. Jarvis D, Mitchell JS, Law PJ, Palin K, Tuupanen S, Gylfe A, et al. Mendelian randomisation analysis strongly implicates adiposity with risk of developing colorectal cancer. Br J Cancer. 2016;115(2):266–72.

427. Holmes M V, Ala-Korpela M, Davey Smith G. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. Nat Rev Cardiol [Internet]. 2017;14(10):577–90. Available from: http://dx.doi.org/10.1038/nrcardio.2017.78\nhttp://10.1038/nrcardio.2017.78

428. Colditz G a. Overview of the epidemiology methods and applications: Strengths and limitations of observational study designs. Crit Rev Food Sci Nutr. 2010;50(S1):10–2.

429. Kim Y. Role of Folate in Colon Cancer Development and Progression. J Nutr. 2003;133(11):3731S – 3739S.

430. Pierce BL, Kraft P, Zhang C. Mendelian Randomization Studies of Cancer Risk: a Literature Review. Curr Epidemiol Reports. 2018 Jun 18;5(2):184–96.

431. Light DW, Lexchin JR. Pharmaceutical research and development: what do we get for all that money? BMJ [Internet]. 2012;345:e4348. Available from: http://www.bmj.com/cgi/doi/10.1136/bmj.e4348

432. Mokry LE, Ahmad O, Forgetta V, Thanassoulis G, Richards JB. Mendelian randomisation applied to drug development in cardiovascular disease: A review. J Med Genet. 2015;52(2):71–9.

433. Walker VM, Davey Smith G, Davies NM, Martin RM. Mendelian randomization: A novel approach for the prediction of adverse drug events and drug repurposing opportunities. Int J Epidemiol. 2017;46(6):2078–89.

434. J. Machiela M, J. Chanock S. LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015;31(21):3555–7.

435. Bush WS, Moore JH. Chapter 11 : Genome-Wide Association Studies. PLoS Comput Biol. 2012;8(12):e1002822.

436. Ardlie KG, Kruglyak L, Seielstad M. Patterns of Linkage Disequilibrium in the Human Genome. Nat Rev Genet [Internet]. 2002;3(4):299–309. Available from: http://www.nature.com/doifinder/10.1038/nrg777

437. Bewick V, Cheek L, Ball J. Statistics review 14: Logistic regression. Crit Care. 2005;9(1):112–8.

438. Szumilas M. Explaining Odds ratios. J Can Acad Child Adolesc Psychiatry. 2010;19(3):227–9.

439. MR-Base: A platform for Mendelian randomisation using summary data from genome-wide association studies [Internet]. [cited 2018 Sep 18]. Available from: http://app.mrbase.org/

440. Bowden J. School of Social and Community Medicine Short Course: Mendelian Randomisation 7. 2016.

441. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. Int J Epidemiol. 2014;43(3):922–9.

442. Göring HHH, Terwilliger JD, Blangero J. Large Upward Bias in Estimation of Locus-Specific Effects from Genomewide Scans. Am J Hum Genet [Internet]. 2001;69(6):1357–69. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0002929707612646

443. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet [Internet]. 2014;23(R1):R89–98. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4170722&tool=pmcentrez&rendertype=abstract

444. Dai JY, Zhang XC. Mendelian Randomization Studies for a Continuous Exposure Under Case-Control Sampling. Am J Epidemiol. 2015;181(6):440–9.

445. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. Stat Methods Med Res [Internet]. 2007;16(4):309–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17715159

446. Ensembl genome browser 88 [Internet]. [cited 2017 May 3]. Available from: http://www.ensembl.org/index.html

447. Jain M. Heritability of Cardiovascular and Personality Traits in 6,148 Sardinians. PLoS Genet. 2006;2(8):1207–23.

448. Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet [Internet]. Nature Publishing Group; 2015;47(3):284–90. Available from: http://dx.doi.org/10.1038/ng.3190

449. Allen N, Sudlow C, Downey P, Peakman T, Danesh J, Elliott P, et al. UK Biobank: Current status and what it means for epidemiology. Heal Policy Technol [Internet]. Elsevier; 2012;1(3):123–6. Available from: http://dx.doi.org/10.1016/j.hlpt.2012.07.003

450. Brion M-JA, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. Int J Epidemiol. Oxford University Press; 2013 Oct 1;42(5):1497–501.

451. Allen NE, Sudlow C, Peakman T, Collins R, UK Biobank. UK Biobank Data: Come and Get It. Sci Transl Med. 2014 Feb 19;6(224):224ed4.

452. Chen SY, Feng Z, Yi X. A general introduction to adjustment for multiple comparisons. J Thorac Dis. 2017;9(6):1725–9.

453. Li Y, Chen CY, Kaye AM, Wasserman WW. The identification of cis-regulatory elements: A review from a machine learning perspective. BioSystems [Internet]. Elsevier Ireland Ltd; 2015;138:6–17. Available from: http://dx.doi.org/10.1016/j.biosystems.2015.10.002

454. Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature [Internet]. Nature Publishing Group; 2016;539(7629):452–5. Available from: http://dx.doi.org/10.1038/nature20149

455. Sheth H, Northwood E, Ulrich CM, Scherer D, Elliott F, Barrett JH, et al. Interaction between polymorphisms in aspirin metabolic pathways, regular aspirin use and colorectal cancer risk: A case-control study in unselected white European populations. PLoS One. 2018;13(2):e0192223.

456. Crespi CL, Miller VP. The R144C change in the CYP2C9*2 allele alters interaction of the cytochrome P450 with NADPH:cytochrome P450 oxidoreductase. Pharmacogenetics. 1997 Jun;7(3):203–10.

457. Barry EL, Poole EM, Baron JA, Makar KW, Mott LA, Sandler RS, et al. CYP2C9 variants increase risk of colorectal adenoma recurrence and modify associations with smoking but not aspirin treatment. Cancer Causes Control. 2013;24(1):47–54.

458. Ciotti M, Marrone A, Potter C, Owens IS. Genetic polymorphism in the human UGT1A6 (planar phenol) UDP-glucuronosyltransferase: pharmacological implications. Pharmacogenetics. 1997 Dec;7(6):485–95.

459. Thompson CL, Plummer SJ, Merkulova A, Cheng I, Tucker TC, Casey G, et al. No association between cyclooxygenase-2 and uridine diphosphate glucuronosyltransferase 1A6 genetic polymorphisms and colon cancer risk. World J Gastroenterol. 2009;15(18):2240–4.

460. Booven D Van, Marsh S, McLeod H, Carrillo MW, Sangkuhl K, Klein TE, et al. Cytochrome P450 2C9-CP2C9. Pharmacogenet Genomics. 2010;20(4):277–81.

461. Parra EJ, Botton MR, Perini JA, Krithika S, Bourgeois S, Johnson TA, et al. Genome-wide association study of warfarin maintenance dose in a Brazilian sample. Pharmacogenomics. 2015;16(11):1–11.

462. Backman JD, O'Connell JR, Tanner K, Peer CJ, Figg WD, Spencer SD, et al. Genome-wide analysis of clopidogrel active metabolite levels identifies novel variants that influence antiplatelet response. Pharmacogenet Genomics. 2017;27(4):159–63.

463. Burgess S, Thompson SG. Multivariable Mendelian randomization: The use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol. 2015;181(4):251–60.

464. Slieker RC, Relton CL, Gaunt TR, Slagboom PE, Heijmans BT. Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception. Epigenetics and Chromatin [Internet]. BioMed Central; 2018;11(1):1–11. Available from: https://doi.org/10.1186/s13072-018-0191-3

465. Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. Nat Rev Genet [Internet]. Nature Publishing Group; 2013;14(8):585–94. Available from: http://dx.doi.org/10.1038/nrg3405

Appendix 1 131 Proteins regulated by aspirin treatment

Protein	Gene name	Fold change (4mM vs 0mM)	Variability (4mM vs 0mM)	Count (4mM vs 0mM)
28S ribosomal protein S5, mitochondrial OS=Homo sapiens GN=MRPS5 PE=1 SV=2 - [RT05_HUMAN]	MRPS5	0.510	44.1	2
39S ribosomal protein L47, mitochondrial OS=Homo sapiens GN=MRPL47 PE=1 SV=2 - [RM47_HUMAN]	MRPL47	0.678	47.5	3
Alpha-2-HS-glycoprotein OS=Homo sapiens GN=AHSG PE=1 SV=1 - [FETUA_HUMAN]	AHSG	0.010	0.0	2
Alpha-methyl-acyl-CoA racemase IIAs OS=Homo sapiens PE=2 SV=1 - [Q6VRU3_HUMAN]	-	0.674	0.4	2
ARF GTPase-activating protein GIT1 OS=Homo sapiens GN=GIT1 PE=4 SV=1 - [J3QRU8_HUMAN]	GIT1	0.621	2.2	2
Beta adrenergic receptor kinase 1 variant (Fragment) OS=Homo sapiens PE=2 SV=1 - [Q59F73_HUMAN]	-	0.642	4.8	2
Beta-2-microglobulin form pI 5.3 (Fragment) OS=Homo sapiens GN=B2M PE=3 SV=1 - [H0YLF3_HUMAN]	B2M	0.528	0.2	3
Biogenesis of lysosome-related organelles complex 1 subunit 3 (Fragment) OS=Homo sapiens GN=BLOC1S3 PE=4 SV=1 - [K7EN58_HUMAN]	BLOC1S3	0.452	62.2	2
Cation-independent mannose-6-phosphate receptor OS=Homo sapiens GN=IGF2R PE=1 SV=3 - [MPRI_HUMAN]	IGF2R	0.526	16.5	5
CCR4-NOT transcription complex subunit 1 OS=Homo sapiens GN=CNOT1 PE=1 SV=2 - [CNOT1_HUMAN]	CNOT1	0.440	36.5	5
CD151 antigen OS=Homo sapiens GN=CD151 PE=4 SV=1 - [E9PL82_HUMAN]	CD151	0.624	7.7	2
cDNA FLJ10654 fis, clone NT2RP2005901, highly similar to FAST kinase domains-containing protein 2 OS=Homo sapiens PE=2 SV=1 - [B3KMB8_HUMAN]	-	0.559	63.0	3
cDNA FLJ12454 fis, clone NT2RM1000555, highly similar to UNR PROTEIN OS=Homo sapiens PE=2 SV=1 - [Q9GZV0_HUMAN]	-	0.622	17.3	10
cDNA FLJ14980 fis, clone Y79AA1000181, highly similar to Homo sapiens CGI-01 protein mRNA OS=Homo sapiens PE=2 SV=1 - [B3KNM8_HUMAN]	-	0.476	5.9	3
cDNA FLJ31217 fis, clone KIDNE2004095, highly similar to RAC-beta serine/threonine-protein kinase (EC 2.7.11.1) OS=Homo sapiens PE=2 SV=1 - [B3KP61_HUMAN]	-	0.348	16.6	2
cDNA FLJ37346 fis, clone BRAMY2021310, highly similar to Transcriptional repressor p66 beta OS=Homo sapiens PE=2 SV=1 - [B3KSZ4_HUMAN]	-	0.510	7.2	2
cDNA FLJ37836 fis, clone BRSSN2010587, weakly similar to Homo sapiens MDN1, midasin homolog (yeast) (MDN1), mRNA OS=Homo sapiens PE=2 SV=1 - [B3KT81_HUMAN]	-	0.673	11.9	2

cDNA FLJ38173 fis, clone FCBBF1000053, highly similar to HYDROXYMETHYLGLUTARYL-COA SYNTHASE, CYTOPLASMIC OS=Homo sapiens PE=2 SV=1 - [Q8N995_HUMAN]	-	0.594	17.0	17
cDNA FLJ46429 fis, clone THYMU3014372, highly similar to DNA replication licensing factor MCM2 OS=Homo sapiens PE=2 SV=1 - [B3KXZ4_HUMAN]	-	0.611	17.8	19
cDNA FLJ51203, highly similar to Thrombospondin-1 OS=Homo sapiens PE=2 SV=1 - [B4E3J7_HUMAN]	-	0.010	0.0	4
cDNA FLJ51266, highly similar to Vitronectin OS=Homo sapiens PE=2 SV=1 - [B7Z553_HUMAN]	-	0.017	95.3	2
cDNA FLJ52285, highly similar to Vesicular integral-membrane protein VIP36 OS=Homo sapiens PE=2 SV=1 - [B4DWN1_HUMAN]	-	0.481	38.0	12
cDNA FLJ52523, highly similar to Arfaptin-2 OS=Homo sapiens PE=2 SV=1 - [B4DX86_HUMAN]	-	0.692	3.2	3
cDNA FLJ53449, highly similar to rRNA methyltransferase 3 (EC 2.1.1.-) OS=Homo sapiens PE=2 SV=1 - [B4DKU3_HUMAN]	-	0.010	0.0	2
cDNA FLJ53478, highly similar to Galectin-3-binding protein OS=Homo sapiens PE=2 SV=1 - [B4DVE1_HUMAN]	-	0.579	83.6	55
cDNA FLJ54048, highly similar to 55 kDa erythrocyte membrane protein (p55) OS=Homo sapiens PE=2 SV=1 - [B4E325_HUMAN]	-	0.460	6.1	2
cDNA FLJ54365, highly similar to DNA replication licensing factor MCM4 OS=Homo sapiens PE=2 SV=1 - [B4DLA6_HUMAN]	-	0.665	17.9	10
cDNA FLJ56163, moderately similar to Mus musculus sushi, von Willebrand factor type A, EGF and pentraxin domain containing 1, mRNA OS=Homo sapiens PE=2 SV=1 - [B4DWM4_HUMAN]	-	0.682	32.8	2
cDNA FLJ56541, highly similar to Ubiquilin-2 OS=Homo sapiens PE=2 SV=1 - [B4DZF1_HUMAN]	-	0.456	88.7	2
cDNA FLJ59967 OS=Homo sapiens PE=2 SV=1 - [B4DS67_HUMAN]	-	0.674	3.8	2
cDNA FLJ75871, highly similar to Homo sapiens staufen, RNA binding protein (STAU), transcript variant T3, mRNA OS=Homo sapiens PE=2 SV=1 - [A8K622_HUMAN]	-	0.682	12.0	7
cDNA FLJ76072, highly similar to Homo sapiens GIPC PDZ domain containing family, member 1 (GIPC1), transcript variant 1, mRNA OS=Homo sapiens PE=2 SV=1 - [A8K2I7_HUMAN]	-	0.250	33.5	3
cDNA FLJ78157 OS=Homo sapiens PE=2 SV=1 - [A8K2J6_HUMAN]	-	0.657	22.9	3
cDNA FLJ78441, highly similar to Homo sapiens nucleolar and spindle associated protein 1 (NUSAP1),mRNA OS=Homo sapiens PE=2 SV=1 - [A8K4B4_HUMAN]	-	0.623	16.6	2
cDNA FLJ90454 fis, clone NT2RP3001560, highly similar to Cleft lip and palate transmembrane protein 1 OS=Homo sapiens PE=2 SV=1 - [B3KQH2_HUMAN]	-	0.562	7.4	5

cDNA, FLJ93255, highly similar to Homo sapiens solute carrier family 7 (cationic amino acidtransporter, y+ system), member 1 (SLC7A1), mRNA OS=Homo sapiens PE=2 SV=1 - [B2R728_HUMAN]	-	1.499	25.0	3
cDNA, FLJ93378, highly similar to Homo sapiens MCM3 minichromosome maintenance deficient 3 (S. cerevisiae) (MCM3), mRNA OS=Homo sapiens PE=2 SV=1 - [B2R7C5_HUMAN]	-	0.599	16.6	19
cDNA, FLJ93645, highly similar to Homo sapiens chromatin assembly factor 1, subunit B (p60) (CHAF1B),mRNA OS=Homo sapiens PE=2 SV=1 - [B2R7X3_HUMAN]	-	0.637	5.8	2
cDNA, FLJ94888, highly similar to Homo sapiens suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin) (ST14), mRNA OS=Homo sapiens PE=2 SV=1 - [B2RAF9_HUMAN]	-	0.349	22.5	3
cDNA, FLJ95518 OS=Homo sapiens PE=2 SV=1 - [B2RBI0_HUMAN]	-	0.010	0.0	2
Claudin-3 OS=Homo sapiens GN=CLDN3 PE=1 SV=1 - [CLD3_HUMAN]	CLDN3	0.557	14.7	3
Creatine kinase B-type OS=Homo sapiens GN=CKB PE=1 SV=1 - [KCRB_HUMAN]	CKB	0.652	1.7	3
Cytochrome c oxidase protein 20 homolog OS=Homo sapiens GN=COX20 PE=1 SV=2 - [COX20_HUMAN]	COX20	0.681	7.1	2
Desmoglein-2 OS=Homo sapiens GN=DSG2 PE=1 SV=2 - [DSG2_HUMAN]	DSG2	0.507	84.3	3
Dihydrofolate reductase OS=Homo sapiens GN=DHFR PE=1 SV=2 - [DYR_HUMAN]	DHFR	0.313	18.1	3
Dihydropyrimidinase-related protein 2 OS=Homo sapiens GN=DPYSL2 PE=1 SV=1 - [DPYL2_HUMAN]	DPYSL2	0.495	4.0	3
Dimethyladenosine transferase 2, mitochondrial OS=Homo sapiens GN=TFB2M PE=1 SV=1 - [TFB2M_HUMAN]	TFB2M	0.556	27.5	2
Dipeptidase 1 OS=Homo sapiens GN=DPEP1 PE=1 SV=3 - [DPEP1_HUMAN]	DPEP1	0.513	69.2	2
DNA polymerase delta subunit 2 (Fragment) OS=Homo sapiens GN=POLD2 PE=4 SV=1 - [C9JHC7_HUMAN]	POLD2	0.351	5.1	2
DNA replication licensing factor MCM5 OS=Homo sapiens GN=MCM5 PE=1 SV=5 - [MCM5_HUMAN]	MCM5	0.622	17.4	20
E3 ubiquitin-protein ligase UBR5 OS=Homo sapiens GN=UBR5 PE=4 SV=1 - [E7EMW7_HUMAN]	UBR5	0.313	6.9	3
Eukaryotic translation initiation factor 1b OS=Homo sapiens GN=EIF1B PE=1 SV=2 - [EIF1B_HUMAN]	EIF1B	0.291	42.2	2
Fatty acid desaturase 2 OS=Homo sapiens GN=FADS2 PE=1 SV=1 - [FADS2_HUMAN]	FADS2	0.258	28.8	2
Flotillin 1 variant (Fragment) OS=Homo sapiens PE=2 SV=1 - [Q53HQ0_HUMAN]	-	0.713	39.0	7
Fragile X mental retardation syndrome-related protein 1 OS=Homo sapiens GN=FXR1 PE=2 SV=1 - [B4DXZ6_HUMAN]	FXR1	0.502	45.6	2
Glutamate--cysteine ligase regulatory subunit OS=Homo sapiens GN=GCLM PE=1 SV=1 - [GSH0_HUMAN]	GCLM	0.326	15.6	3

Golgi membrane protein 1 OS=Homo sapiens GN=GOLM1 PE=1 SV=1 - [GOLM1_HUMAN]	GOLM1	0.345	24.5	7
Golgin subfamily A member 2 OS=Homo sapiens GN=GOLGA2 PE=1 SV=3 - [GOGA2_HUMAN]	GOLGA2	0.436	8.7	3
Grancalcin (Fragment) OS=Homo sapiens GN=GCA PE=4 SV=1 - [H7C2Z6_HUMAN]	GCA	0.669	19.0	2
Heat shock protein beta-1 OS=Homo sapiens GN=HSPB1 PE=1 SV=2 - [HSPB1_HUMAN]	HSPB1	0.281	8.1	7
Hephaestin (Fragment) OS=Homo sapiens GN=HEPH PE=2 SV=1 - [Q5JZ08_HUMAN]	HEPH	0.649	26.2	6
Heterogeneous nuclear ribonucleoproteins A2/B1 OS=Homo sapiens GN=HNRNPA2B1 PE=1 SV=2 - [ROA2_HUMAN]	HNRNPA2B1	0.603	13.7	42
Histone acetyltransferase type B catalytic subunit OS=Homo sapiens GN=HAT1 PE=1 SV=1 - [HAT1_HUMAN]	HAT1	0.468	2.2	2
Importin subunit alpha OS=Homo sapiens PE=2 SV=1 - [A8K7D9_HUMAN]	-	0.455	2.0	9
Integrin-linked kinase associated phosphatase OS=Homo sapiens GN=ILKAP3 PE=2 SV=1 - [F8SNU7_HUMAN]	ILKAP3	0.706	11.6	2
Junction plakoglobin OS=Homo sapiens GN=JUP PE=1 SV=3 - [PLAK_HUMAN]	JUP	0.331	28.6	7
Keratin, type I cytoskeletal 10 OS=Homo sapiens GN=KRT10 PE=1 SV=6 - [K1C10_HUMAN]	KRT10	0.010	0.0	23
Keratin, type I cytoskeletal 9 OS=Homo sapiens GN=KRT9 PE=1 SV=3 - [K1C9_HUMAN]	KRT9	0.010	0.0	8
Keratin, type II cytoskeletal 2 epidermal OS=Homo sapiens GN=KRT2 PE=1 SV=2 - [K22E_HUMAN]	KRT2	0.010	0.0	15
Keratinocyte proline-rich protein OS=Homo sapiens GN=KPRP PE=1 SV=1 - [KPRP_HUMAN]	KPRP	0.010	0.0	3
Kunitz-type protease inhibitor 1 OS=Homo sapiens GN=SPINT1 PE=4 SV=1 - [F5H877_HUMAN]	SPINT1	0.443	3.2	3
Lymphoid specific helicase variant9 OS=Homo sapiens GN=HELLS PE=2 SV=1 - [Q6I7N8_HUMAN]	HELLS	0.299	19.3	2
Lysosome-associated membrane glycoprotein 2 OS=Homo sapiens GN=LAMP2 PE=2 SV=1 - [B7Z2R9_HUMAN]	LAMP2	0.610	12.5	7
Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial OS=Homo sapiens GN=MCCC2 PE=1 SV=1 - [MCCB_HUMAN]	MCCC2	0.594	32.0	6
MHC class I antigen (Fragment) OS=Homo sapiens GN=HLA-A PE=3 SV=1 - [E5BBI6_HUMAN]	HLA-A	0.641	0.6	2
MHC class I antigen (Fragment) OS=Homo sapiens GN=HLA-B PE=3 SV=1 - [B7VU68_HUMAN]	HLA-B	0.401	12.6	2
Mitochondrial import inner membrane translocase subunit Tim23 OS=Homo sapiens GN=TIMM23 PE=1 SV=1 - [TIM23_HUMAN]	TIMM23	0.683	30.8	2
Mitogen-activated protein kinase 14 isoform 4 OS=Homo sapiens GN=MAPK14 PE=2 SV=1 - [B5TY32_HUMAN]	MAPK14	0.612	83.6	2
MKI67 FHA domain-interacting nucleolar phosphoprotein OS=Homo sapiens GN=MKI67IP PE=1 SV=1 - [MK67I_HUMAN]	MKI67IP	0.483	0.3	2

Mucin-12 OS=Homo sapiens GN=MUC12 PE=4 SV=1 - [F5GWV9_HUMAN]	MUC12	0.010	0.0	4
Multidrug resistance associated protein OS=Homo sapiens GN=MRP3 PE=2 SV=1 - [Q96QA9_HUMAN]	MRP3	0.698	65.5	5
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 4 OS=Homo sapiens GN=NDUFA4 PE=1 SV=1 - [NDUA4_HUMAN]	NDUFA4	0.360	33.7	3
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 6 OS=Homo sapiens GN=NDUFA6 PE=1 SV=3 - [NDUA6_HUMAN]	NDUFA6	0.437	10.7	2
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 4 OS=Homo sapiens GN=NDUFB4 PE=1 SV=3 - [NDUB4_HUMAN]	NDUFB4	0.478	5.3	3
Non-structural maintenance of chromosomes element 1 homolog (Fragment) OS=Homo sapiens GN=NSMCE1 PE=4 SV=1 - [I3L1I3_HUMAN]	NSMCE1	0.702	30.5	2
Nox1 NAD(P)H-oxidase subunit OS=Homo sapiens GN=NOX1 PE=2 SV=1 - [Q8TCT3_HUMAN]	NOX1	0.533	7.7	5
NUCB1 protein OS=Homo sapiens PE=2 SV=1 - [Q96BA4_HUMAN]	-	0.290	89.9	2
Olfactomedin-4 OS=Homo sapiens GN=OLFM4 PE=1 SV=1 - [OLFM4_HUMAN]	OLFM4	0.388	16.4	34
Oxysterol-binding protein (Fragment) OS=Homo sapiens GN=OSBPL8 PE=3 SV=1 - [F8VUA7_HUMAN]	OSBPL8	0.634	15.3	2
Palmitoyl-protein thioesterase 1 OS=Homo sapiens GN=PPT1 PE=1 SV=1 - [PPT1_HUMAN]	PPT1	0.605	22.0	3
Phosphatidate phosphatase LPIN1 OS=Homo sapiens GN=LPIN1 PE=2 SV=1 - [B4DET9_HUMAN]	LPIN1	0.642	5.5	2
Pleiotropic regulator 1 (Fragment) OS=Homo sapiens GN=PLRG1 PE=4 SV=1 - [H0YA24_HUMAN]	PLRG1	0.490	13.3	2
PNKP protein OS=Homo sapiens PE=2 SV=2 - [Q9BUL2_HUMAN]	-	0.532	3.6	2
POLDIP3 protein (Fragment) OS=Homo sapiens GN=POLDIP3 PE=2 SV=2 - [Q96DI9_HUMAN]	POLDIP3	0.572	10.6	3
Polymeric immunoglobulin receptor OS=Homo sapiens GN=PIGR PE=1 SV=4 - [PIGR_HUMAN]	PIGR	0.413	75.6	11
Probable palmitoyltransferase ZDHHC20 OS=Homo sapiens GN=ZDHHC20 PE=2 SV=1 - [B4DRN8_HUMAN]	ZDHHC20	0.650	4.7	2
Prolow-density lipoprotein receptor-related protein 1 OS=Homo sapiens GN=LRP1 PE=1 SV=2 - [LRP1_HUMAN]	LRP1	0.492	50.1	3
Protein AATF OS=Homo sapiens GN=AATF PE=1 SV=1 - [AATF_HUMAN]	AATF	0.012	25.3	2
Protein BUD31 homolog OS=Homo sapiens GN=BUD31 PE=4 SV=1 - [C9JNV2_HUMAN]	BUD31	0.522	17.0	2
Protein FAM134C OS=Homo sapiens GN=FAM134C PE=1 SV=1 - [F134C_HUMAN]	FAM134C	0.706	3.7	2
Protein syndesmos (Fragment) OS=Homo sapiens GN=NUDT16L1 PE=4 SV=1 - [K7EIN2_HUMAN]	NUDT16L1	0.591	11.4	2
Protein-methionine sulfoxide oxidase MICAL1 OS=Homo sapiens GN=MICAL1 PE=1 SV=2 - [MICA1_HUMAN]	MICAL1	0.555	30.9	3
PTGFRN protein (Fragment) OS=Homo sapiens GN=PTGFRN PE=2 SV=1 - [Q1WWL2_HUMAN]	PTGFRN	0.593	26.2	5

Putative uncharacterized protein DKFZp686K06216 (Fragment) OS=Homo sapiens GN=DKFZp686K06216 PE=2 SV=1 - [Q6MZF3_HUMAN]	DKFZp686K06216	0.538	4.7	2
Putative uncharacterized protein MCM6 (Fragment) OS=Homo sapiens GN=MCM6 PE=2 SV=1 - [Q4ZG57_HUMAN]	MCM6	0.652	7.6	25
RBM16 protein (Fragment) OS=Homo sapiens GN=RBM16 PE=2 SV=1 - [Q05BU5_HUMAN]	RBM16	0.010	0.0	3
Ribonucleoside-diphosphate reductase subunit M2 (Fragment) OS=Homo sapiens GN=RRM2 PE=4 SV=1 - [C9JXC1_HUMAN]	RRM2	0.363	2.0	2
RNA 3'-terminal phosphate cyclase-like protein OS=Homo sapiens GN=RCL1 PE=1 SV=3 - [RCL1_HUMAN]	RCL1	0.249	42.5	4
Serine/threonine-protein phosphatase (Fragment) OS=Homo sapiens GN=PPP4C PE=3 SV=1 - [H3BTA2_HUMAN]	PPP4C	0.664	16.0	3
Sister chromatid cohesion protein PDS5 homolog A OS=Homo sapiens GN=PDS5A PE=1 SV=1 - [PDS5A_HUMAN]	PDS5A	0.552	13.9	3
Small nuclear ribonucleoprotein F OS=Homo sapiens GN=SNRPF PE=1 SV=1 - [RUXF_HUMAN]	SNRPF	0.647	7.1	3
Solute carrier family 12 member 2 OS=Homo sapiens GN=SLC12A2 PE=1 SV=1 - [S12A2_HUMAN]	SLC12A2	0.565	13.8	29
Solute carrier family 22 member 18 OS=Homo sapiens GN=SLC22A18 PE=4 SV=1 - [E9PRM7_HUMAN]	SLC22A18	0.667	4.4	2
Stathmin OS=Homo sapiens GN=STMN1 PE=2 SV=1 - [Q96CE4_HUMAN]	STMN1	0.614	8.1	2
Sterol 26-hydroxylase, mitochondrial OS=Homo sapiens GN=CYP27A1 PE=1 SV=1 - [CP27A_HUMAN]	CYP27A1	0.491	6.9	2
SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 2 OS=Homo sapiens GN=SMARCD2 PE=4 SV=1 - [J3KMX2_HUMAN]	SMARCD2	0.636	25.6	2
Syndecan-1 OS=Homo sapiens GN=SDC1 PE=1 SV=3 - [SDC1_HUMAN]	SDC1	0.366	42.5	3
Thioredoxin domain-containing protein 9 OS=Homo sapiens GN=TXNDC9 PE=4 SV=1 - [F8WBV5_HUMAN]	TXNDC9	0.642	12.3	2
Thymidine kinase OS=Homo sapiens GN=TK1 PE=2 SV=1 - [B5BU32_HUMAN]	TK1	0.463	9.7	4
Trans-golgi network protein 2, isoform CRA_a OS=Homo sapiens GN=TGOLN2 PE=2 SV=1 - [B2R686_HUMAN]	TGOLN2	0.550	32.3	2
Transporter (Fragment) OS=Homo sapiens GN=SLC6A6 PE=3 SV=1 - [C9JPV1_HUMAN]	SLC6A6	0.503	7.4	2
U2 small nuclear ribonucleoprotein B'' OS=Homo sapiens GN=SNRPB2 PE=1 SV=1 - [RU2B_HUMAN]	SNRPB2	0.681	13.3	2
U5 small nuclear ribonucleoprotein 40 kDa protein OS=Homo sapiens GN=SNRNP40 PE=1 SV=1 - [SNR40_HUMAN]	SNRNP40	0.596	13.6	3
Ubiquitin-conjugating enzyme E2 R2 OS=Homo sapiens GN=UBE2R2 PE=1 SV=1 - [UB2R2_HUMAN]	UBE2R2	0.454	38.1	2

UBX domain-containing protein 1 (Fragment) OS=Homo sapiens GN=UBXN1 PE=4 SV=1 - [E9PJ81_HUMAN]	UBXN1	0.698	50.4	3
Uncharacterized protein (Fragment) OS=Homo sapiens PE=3 SV=1 - [H7C469_HUMAN]	-	0.681	12.8	10
Uncharacterized protein C19orf21 OS=Homo sapiens GN=C19orf21 PE=1 SV=1 - [CS021_HUMAN]	C19orf21	2.170	4.5	2
Uncharacterized protein C7orf50 (Fragment) OS=Homo sapiens GN=C7orf50 PE=4 SV=1 - [C9JQV0_HUMAN]	C7orf50	0.508	29.5	4
UPF0554 protein C2orf43 OS=Homo sapiens GN=C2orf43 PE=1 SV=1 - [CB043_HUMAN]	C2orf43	0.713	24.9	2
Uridine-cytidine kinase 2 OS=Homo sapiens GN=UCK2 PE=2 SV=1 - [B4DGD3_HUMAN]	UCK2	0.432	4.4	2
UV excision repair protein RAD23 homolog A OS=Homo sapiens GN=RAD23A PE=4 SV=1 - [K7ESE3_HUMAN]	RAD23A	0.632	15.2	3