Is Complement Factor H a shared risk factor for age-related macular degeneration and cardiovascular disease?

Thesis presented for the Degree of PhD in the Division of Medicine of University College London

Reecha Sofat

Supervisor Professor Aroon Hingorani

Declaration of work

I have performed all experimental work and statistical analysis in the thesis presented here, apart from what is listed below:

Chapter 2: SNP genotyping for both MRC AMD study and the EUREYE was outsourced and carried out either in other institutions (MRC AMD study) or by a commercial provider (the EUREYE study).

Chapter 4: Generation of pure human fH and plasma deplete of fH that was used to re-calibrate the Binding Site standard was carried out by Dr Svetlana Hakobyan in Professor Paul Morgan's laboratory (University of Cardiff). Re-calibration of the Binding Site standard was carried out by Dr Patrizia Mangione in Professor Mark Pepys laboratory (University College London).

Chapter 6: Genome wide analysis was performed by Mr Daniel Barnes at the MRC Epidemiology Unit (University of Cambridge)

Reecha Sofat, April 2011

Acknowledgements

I am extremely privileged to have worked with Professor Aroon Hingorani who is an inspirational scientist, doctor and teacher. I thank him whole heartedly for his supervision of this thesis. I thank him for his unwavering enthusiasm, support, guidance and expert knowledge throughout.

I am indebted to the British Heart Foundation and the Medical Research Council for the funding provided which supported the work presented here.

I would like to thank the many collaborators for their invaluable contribution to this work, in particular Dr. Juan Pablo Casas for informative discussion and statistical advice. I am particularly grateful to collaborators who have provided the data sets used in this thesis, without which none of this work would have been possible, in particular I would like to thank Professor John Yates, Professor Tony Moore, Mr Andrew Webster, Professor Astrid Fletcher, Professor Steve Humphries, Professor Philippa Talmud, Dr Meena Kumari, Dr Helen Ireland, Professor Nick Wareham and Dr Claudia Langenberg. I would also like to thank Professor Mark Pepys, Dr Patrizia Mangione and Professor Paul Morgan for their central role in the development of the assay described in this thesis. For statistical advice and discussion I would also like to thank Professor Liam Smeeth.

I am especially grateful to colleagues and friends for their invaluable support, constructive criticism, wisdom and humour, in particular Professor Raymond MacAllister, Professor Gordon Stewart, Dr Michael Okorie, Dr Lila Mayahi, Dr Manasi Nandi and Dr Ines Pinéda-Torra.

Finally I would like to thank my family for their patience and unfailing support. In particular Mum and Nidhi who have always taught me that all is achievable, and I dedicate this thesis to them.

Abstract

Background and Aims: Inflammation is implicated in common disorders of ageing including atherosclerosis and age-related macular degeneration (AMD), although the link between inflammation and cardiovascular disease (CVD) is the more studied. The recent finding that susceptibility to AMD is increased substantially by common single nucleotide polymorphisms (SNPs) in the gene that encodes complement factor H (*CFH*; a circulating inhibitor of complement activation) provides evidence that inflammation in general, and complement in particular, maybe causally involved in AMD. Since AMD and atherosclerosis share similar pathological features and risk factors, including a link with inflammation, an important question arises: is complement factor H (fH) a shared risk factor for both AMD and CHD? One SNP in particular, which has the most replicated association in AMD, rs1061170, which encodes a putative functional tyrosine to histidine change (Y402H), and has been studied in both AMD and coronary heart disease (CHD). I hypothesised that genetic variants in *CFH*, in particular rs1061170 is associated with risk of both AMD and CHD and that this association may be mediated through changes in circulating fH concentration. I addressed this hypothesis by: (i) precisely defining the effect of the association of the rs1061170 SNP encoding Y402H in *CFH* on AMD risk; (ii) precisely defining the association of rs1061170 on risk of CHD events; and (iii) developing and validating a high throughput assay of circulating fH, to enable further evaluation of the nature of the association between *CFH* genotype and fH concentration, and fH concentration and disease; (iv) measuring fH in a population based sample to determine its non-genetic correlates and genetic determinants (v) measuring fH in case control studies of AMD and genotyping of SNPs in the *CFH* and related genes to determine the concordance of the genetic associations of fH and AMD.

Methods: To address aims (i) and (ii), I conducted a systematic review of published studies investigating the effect of variants in *CFH* on AMD and CHD risk respectively, supplementing data with results from newly genotyped studies in both AMD and CHD. To address aim (iii) I developed and validated a high-throughput assay measuring circulating fH, which I used to undertake studies in aim (iv) in which I measured fH in a population based sample with an existing range of blood and lifestyle measures as well as anthropometric, cardiovascular,

glycaemic, lipid, liver, renal, and inflammation markers. In addition to this genome wide information was also available on ~500,000 SNPs across the genome with additional imputation of un-typed SNPs, giving coverage of ~ 2 million markers across the genome. In order to achieve aim (v) I measured fH in case control studies of AMD, with additional genotyping of SNPs in the *CFH* and *CFH* related gene in order to attain a more high resolution signal of association in this genomic region for both fH concentration and AMD risk.

Results: Data synthesis from published literature and newly genotyped studies, confirmed the strong association of the rs1061170 SNP with risk of AMD (per-allele odds ratio (OR) of 2.30, 99% CI 1.93, 2.73; p<0.001), in individuals of European descent, although the association was less clear in individuals of Chinese or Japanese descent. However, there was no association of rs1061170 with CHD (per-allele OR 1.01 95% CI 0.98, 1.04), or established risk factors for CHD. Adaptation of an existing commercial, low through-put assay allowed the development and validation of a high throughput assay to measure circulating fH concentrations. With an operating range of 7-1000 mg/L, this assay was reliable, repeatable and robust, enabling assay of fH in stored samples. In a large population study, novel associations of fH with lipids, apo-lipoproteins and indices of adiposity were identified and genetic determinants localised to the *CFH/CFHR* gene cluster on Chromosome 1. In case-control analysis, there was no association of fH concentration with AMD risk.

Conclusions: Genetic variation in *CFH*, and in particular the effect of the most replicated rs1061170 SNP is robustly associated with AMD with little attenuation in the effect size as data has accrued. However the effect of the same SNP is not associated with CHD. Circulating fH is associated with a range of cardio-metabolic biomarkers and regulated by common genetic variants in the vicinity of the encoding gene on chromosome 1. However fH itself is not associated with risk of AMD, suggesting the genetic association of *CFH* with AMD is mediated through altered fH function or perhaps through an fH-related protein encoded by an adjacent gene.

List of Abbreviations

ABI	Applied Biosystems Incorporated
aHUS	Atypical Haemolytic Uraemic Syndrome
ALP	Alkaline Phophatase
ALT	Alanine Transaminase
AMD	Age-related macular degeneration
ANOVA	Analysis of Variance
APOE	Apolipoprotien E
AREDS	Age-Related Eye Disease study
ARIC	Atherosclerosis Risk In Communities
ARM	Age-related Maculopathy
BMI	Body Mass Index
BNII	Behring Nephelometer II
BP	Base Pair
BS	Binding Site
BWHHS	British Women's Health and Heart Study
BRLMM	Bayesian Robust Linear Model with Mahalanobis distance classifier
CABG	Coronary Artery Bypass Graft
CAC	Coronary Artery Calcification
CEPH	Centre d'Étude du Polymorphisme
CETP	Cholesteryl ester transfer protein gene
CEU	Centre d'Étude du Polymorphisme
CFH	Complement factor H
CFHR	Complement factor H related gene
CFI	Complement Factor I
CHD	Coronary Heart Disease
CHD	Cardiovascular Health Study
CI	Confidence Interval
CIMT	Carotid Intimal Media Thickness

CNV	Choroidal Neovascularisation
COBBLL1	Cordon-bleu protein-like 1
CRB1	Crumbs Homolog 1
CRP	C-reactive protein
CR1	Complement receptor 1
CT	Computed Tomography
CV	Co-efficient of Variation
CVD	Cardiovascular disease
DAF	Decay accelerating factor
DBP	Diastolic Blood Pressure
ddNTP	Dideoxy nucleoside triphosphates
DHPLC	Denaturing high-performance liquid chromatography
DLC1	Deleted in Liver Cancer 1
DNA	Deoxyribonucleic acid
ECG	Electro-cardiogram
ECTIM	Étude cas-temoin de l'infarctus myocarde
EDS	Ealing Diabetes Study of Coagulation
EDTA	Ethylenediaminetetraacetic acid
ELAVL2	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 2 (Hu antigen B)
ELISA	Enzyme-linked immunosorbent assay
ELSA	English Longitudinal Study of Ageing
EMBASE	Excerpta Medica Database
ERC2	ELKS/RAB6-interacting/CAST family member 2
EUREYE	European Eye Study
FB	Factor B
FHL	Factor H Like
FPLC	Fast Protein liquid Chromatography
FRET	Fluorescence Resonance Energy Transfer
GA	Geographic Atrophy
GAP	GTPase activating protein

GGT	Gamma-glutamyltransferase
GTP	Guanosine triphosphate
GWAS	Genome wide association study
HDL-C	High Density Lipoprotein Cholesterol
HIFMECH	Hypercoagubility and Impaired Fibrinolytic function Mechanisms predisposing to MI Study
HR	Hazard Ratios
HRP	Horseradish Peroxidase
hs CRP	High Sensitivity CRP
HWE	Hardy Weinberg Equilibrium
ICARMS	International Classification System for Age-related Macular Degeneration
IL-6	Interleukin 6
IVUS	Intravascular ultrasound
LD	Linkage Disequilibrium
LDL	Low Density Lipoprotein receptor
LDLR	Low Density Lipoprotein
LED	Light Emitting Diode
LIPC	Hepatic lipase
LOD	Logarithm (base 10) of odds
Lp(a)	Lipoprotein a
LPL	Lipoprotein lipase
LRP6	Low density lipoprotein receptor-related protein 6
MAC	Membrane attack complex
MASP	Mannose associated serine proteases
MAF	Minor Allele Frequency
MALDITOF	Matrix-assisted laser desorption-ionization/ time-of-flight mass spectrometer
MBL	Mannose binding lectins
MBL	Megabase
MCP	Membrane cofactor protein
MEDLINE	Medical Literature Analysis and Retrieval System Online

MHC	Major Hisctocompatibility Complex
MI	Myocardial Infarction
MPGN II	Membranoproliferative Glomerulonephritis type II
MRC	Medical Research Council
NCBI	National Center for Biotechnology Information
NEFA	Non-esterified fatty acid
NPHS II	Northwick Park Heart Study II
OR	Odds Ratio
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PHACTR1	Phosphatase and actin regulator 1
PNG	Portable network graphics
PTCA	Percutaneous transluminal coronary angioplasty
PVD	Peripheral Vascular Disease
QC	Quality Control
QQ	Quantile-Quantile
QTL	Quantitative trait locus
RCA	Regulation of complement activation
RCT	Randomised Control Trial
RFLP	Restriction fragment length Polymorphisms
RID	Radial immune diffusion
RNA	Ribonucleic acid
RPE	Retinal Pigment Epithelium
rs-ID	Reference SNP Identification number
SAP	Serum amyloid P
SBP	Systolic Blood Pressure
SCR	Short consensus repeat
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEM	Standard error of the mean

SERPING1	Serpin peptidase inhibitor, clade G (C1 inhibitor), member 1
SNP	Single Nucleotide Polymorphism
SQRT	Square Root
TC	Total Cholesterol
TNF	Tumour Necrosis Factor
TRG	Triglycerides
TUSC1	Tumour suppressor candidate 1
UCL	University College London
UCLH	University College London Hospital
UDACS	University College Diabetes and Cardiovascular Study
UK	United Kingdom
USA	United States of America
UV	Ultraviolet
VEGF	Vascular Endothelial Growth Factor
VLDLR	Very low density lipoprotein receptor
WHO	World Health Organisation
WII	Whitehall II
WNT5A	Wingless-type MMTV integration site family, member 5A
WTCCC	Wellcome Trust Case Control Consortium
ZBTB41	Zinc finger and BTB domain containing 41

Table of Contents

List of Figures

List of Tables

1 Introduction

Diseases of aging are of mounting public health importance worldwide. Associated with significant mortality and morbidity, they impact not only on current but projected health care resources. One way to reduce this burden is early identification and treatment of those at risk of disease. Cardiovascular disease (CVD) and age-related macular degeneration (AMD) are two such common, age-related, complex diseases which will be the focus of discussion herein.

There are a number of parallels between these two conditions, not only because they are both age-related, but because they share common pathological features, risk factors and evolution. Moreover inflammation is thought to contribute to both CVD and AMD. In this introductory chapter I aim to discuss, both these diseases, highlighting the parallels that exist between them, before moving on to examine the evidence implicating inflammatory pathways in both conditions and in particular the complement pathway. I finally propose an approach to address the limitations of the evidence, in the context of emerging data from genome wide association scans (GWAS), thus setting out the aims of this thesis.

1.1 Atherosclerosis and age-related macular degeneration: parallels and differences

Comparisons have been drawn between AMD and CVD that support the view of a shared common pathological basis. Recent findings in the field of AMD have found compelling links with inflammation, and in particular with the complement cascade, a key part of the innate immune system. Atherosclerosis, the pathological lesion that underlies CVD, is also believed to represent a chronic inflammatory process, in which both innate and adaptive immune responses have been implicated, including the complement cascade.

1.1.1 Coronary heart disease

1.1.1.1 Disease burden

Cardiovascular disease, the most common forms of which are coronary heart disease (CHD) and stroke remain a major cause of mortality and morbidity worldwide. It is the most common

cause of premature death in industrialised nations, accounting for 4.35 million deaths per year in Europe and 35% of deaths in the United Kingdom [1, 2]. CVD is also associated with considerable morbidity, accounting for 20% of disability-adjusted life years [3]. CHD alone is the most common cause of death in Europe with 1.95 million deaths per annum, of these an estimated 94 000 deaths per annum are in the UK [1, 2].

1.1.1.2 Risk factors and risk prediction

Epidemiological evidence has been central in establishing that certain exposures increase risk of CVD [4]. Those “risk factors” that can be modified include smoking, blood pressure, high cholesterol and diabetes. Non-modifiable risk factors include age, family history and male sex. Although measurement of these risk factors has formed the cornerstone of targeting those at increased absolute risk of CVD, it is well known that most events occur in individuals with average levels of most of the modifiable risk factors [5], including blood pressure and cholesterol even though they are causal [6, 7]. This has been the motivation to search for other biomarkers that may influence CVD, in order to improve risk prediction, provide new therapeutic intervention and or preventative strategies and enable identification of novel drug targets, with the aim of lowering the burden of CVD.

1.1.1.3 Patho-physiology of CHD

A key element of the search for biomarkers has been on better understanding the pathological processes of atherosclerosis that underlie CVD. Atherosclerosis is thought to arise from damage to the single cell layer of vascular endothelium that lines all arteries [8]. Damage can occur through one of a number of proposed pathways including high levels of (oxidised) low density lipoprotein cholesterol (LDL-C), local free radical generation, elevated plasma homocysteine, infectious micro-organisms such as *Chlamydia pnuemoniae*, shear stress in areas of turbulent flow, and local release of cytokines (all of which are reviewed in [9]). Endothelial function itself is thought to become altered as a consequence. Under normal physiological conditions the endothelial surface is anti-adhesive, anti-coagulant and impermeable, but in response to injury becomes pro-adhesive, attracting in the early stages blood derived monocytes, which gain entry to the underlying medial layer, as endothelial cells

become permeable, and induction of pro-coagulant pathways as a response to damage. Injury to the endothelium thus triggers an inflammatory response and like injury in other any other tissue, the immune system is mobilised. One of the first features in the development of atherosclerosis is the entry of lipids and monocytes into the media, where monocytes differentiate into macrophages, and take up oxidised LDL-C forming characteristic lipid laden macrophages commonly referred to as foam cells. A collection of foam cells appear as small raised yellow lesions on the intimal surface of the arterial wall. Known as 'fatty streaks', these are hallmarks of very early atherosclerosis and can be found in otherwise healthy adolescents. Not all fatty streaks will progress to mature atherosclerotic plaques. However with continued damage, the immune process continues unabated, becoming chronic, promoting the formation of a lipid laden necrotic core, mediated by elements of the innate immune system migrating into the lesion (T cells and macrophages), activation of which amplifies inflammation further with the release of cytokines, chemokines, hydrolytic enzymes and growth factors. The response to necrosis by any tissue is an attempt at healing by fibrosis, and atherosclerosis is no exception, with the formation of a fibrous cap eventually protruding into the arterial lumen causing flow limiting or clinically catastrophic occlusive symptoms. The interface between the vascular endothelial cell lining of the artery and the outer bounds of the arterial media has been the focus of much investigation and this is where the pathological changes that define atherosclerosis occur.

Clinically, atherosclerosis in the context of CHD can present either as symptoms of angina, form blood flow limitations, or as an acute occlusive event (myocardial infarction), due to the rupture of a plaque leading to thrombosis in an affected coronary vessel. Medical therapies are targeted at dissolving acutely formed thrombus, and reducing further platelet aggregation. Mechanical interventional procedures may be used to re-establish and maintain patency of coronary vessels. In the longer term, therapy is targeted at reducing platelet activation, which plays a crucial role in formation of thrombosis, and control of afore mentioned modifiable risk factors.

1.1.1.4 The complement system

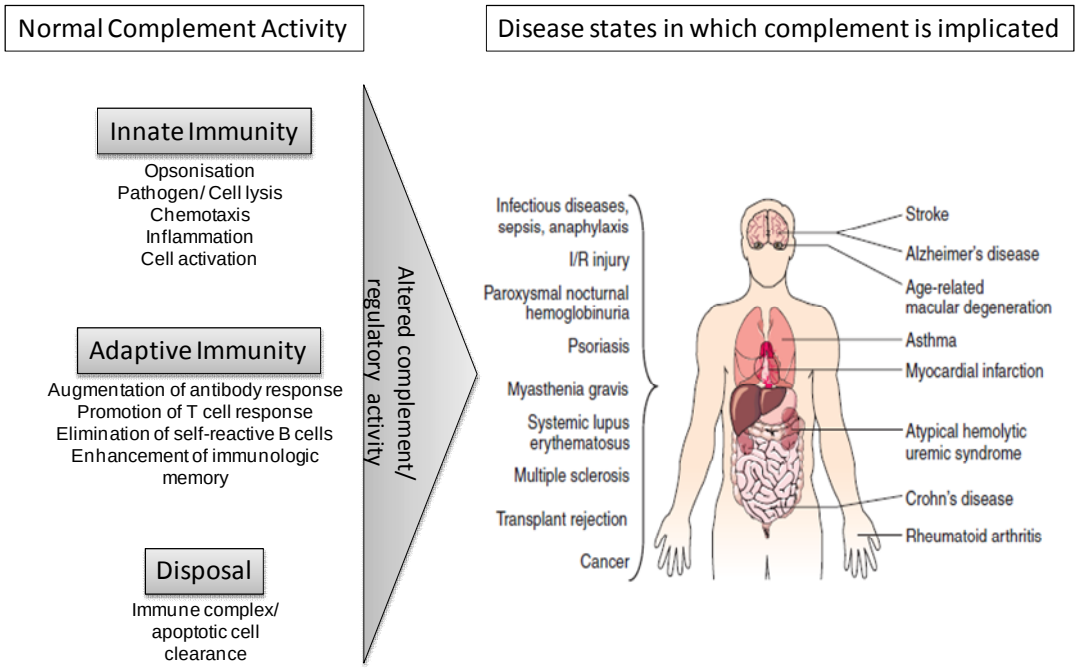
Complement is one component of the immune process implicated in atherosclerosis. Findings from recent GWAS indicate that variants in genes encoding complement components are associated with a number of diseases of aging, including AMD [10] and Alzheimer's disease [11], discussed in further detail in section 1.1.2, as well as susceptibility to infections [12]. This provides evidence of a common mechanism that may be advantageous for protection from early-life infection but which underlies these diseases of aging. A brief overview of the complement system and its function is presented here, in order to be able to place subsequent descriptions of the role of specific complement proteins in atherosclerosis and AMD into context.

An essential part of the immune system, complement is composed of over 30 proteins distributed in plasma and on cell membranes. It provides a generalised non-specific response which identifies targets (via C3b) and destroys them (via direct lysis) while dispatching messengers to seek help (via anaphylotoxins C3a and C5a) thus mediating a more specific antibody-mediated response (reviewed in [13, 14]). The importance of individual constituent proteins has been demonstrated by the diseases that arise from mutations leading to complete deficiency of a component of the complement system, or those that severely affect function of the whole system. In general, individuals with mutations affecting complement components have predisposition to recurrent bacterial infections or immune related disorders. However, complement has also been shown to play a role in a number of diseases which, at first glance, may not be thought of in either of these categories (**Figure 1.1**).

Down-regulation of the complement system is as important as its activation. Although complement is essential for clearance of pathogenic or altered host cells, its activity must also be regulated in order to avoid damage to host cells, which become tagged with C3b at the same time as pathogenic material. Protection and control is achieved through a number of complement regulators, all which have a similar structure and are collectively known as regulators of complement activity (RCA), most being encoded by genes that form the RCA cluster on chromosome 1. Regulatory proteins can either be membrane bound, soluble or both, and protect against complement activation by down-regulating the central proteolytic

activity of the amplification and opsonisation steps (centred on C3b). In theory, deficiencies and mutations in these regulators may lead to excessive complement activation and tissue injury.

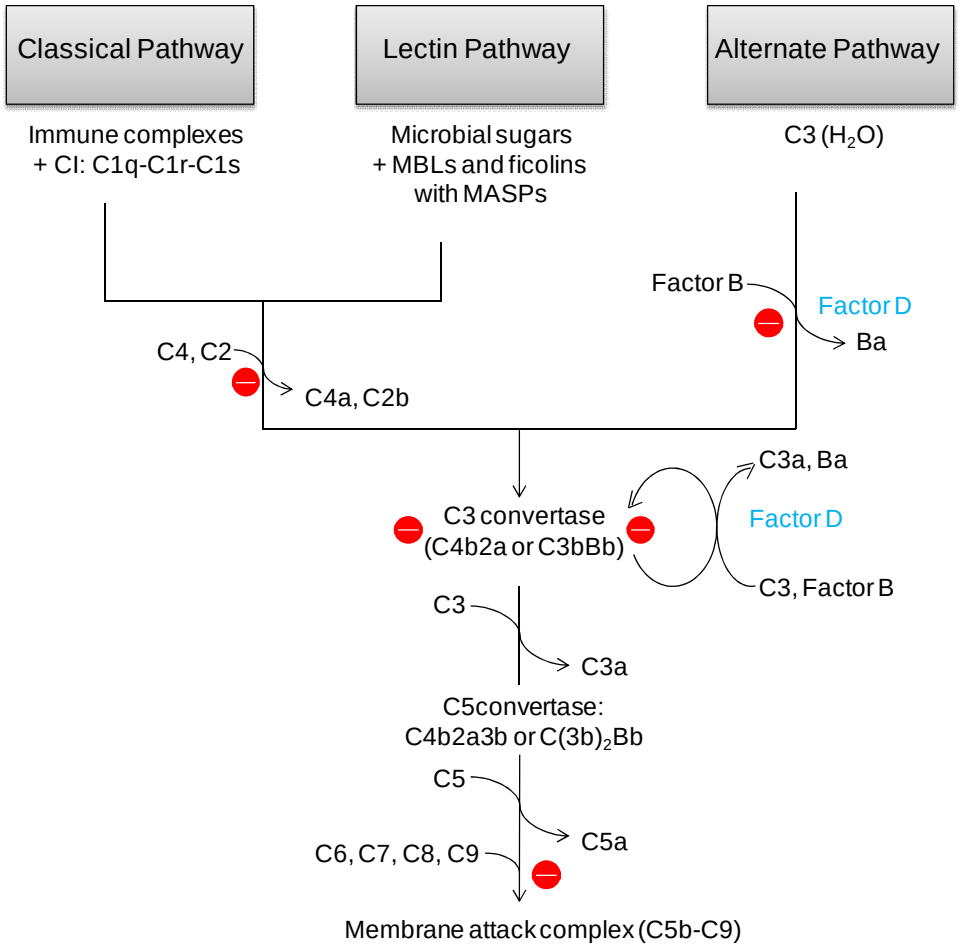
Figure 1.1 The role of complement in health and disease. The functions of the complement cascade in recognition of pathogens, mediating adaptive immunity and effector functions through the disposal of pathogens, immune complexes and apoptotic cells. Dysfunction or uncontrolled regulation of any of these functions has been implicated in a number of organ specific and systemic diseases. Adapted from Ricklin D and Lambris JD, Nature Biotechnology 2007 [15].



Complement is activated by one of three pathways, classical, alternate or lectin (**Figure 1.2**). Of the three only the alternate pathway can be continuously activated in the absence of foreign material, through the action of the continuous, although slow, spontaneous hydrolysis of C3, commonly referred to as the 'tick over' pathway [16]. Importantly, all pathways converge on C3 and the formation of C3 convertase, a proteolytic enzyme complex, without which full activation of complement cannot be achieved. This is also the focal point of the amplification of the activated complement cascade. The C3 convertase formed by the classical and lectin

pathway, although assembled using different components of complement (**Figure 1.2**), is functionally homologous to that which is formed through the alternate pathway.

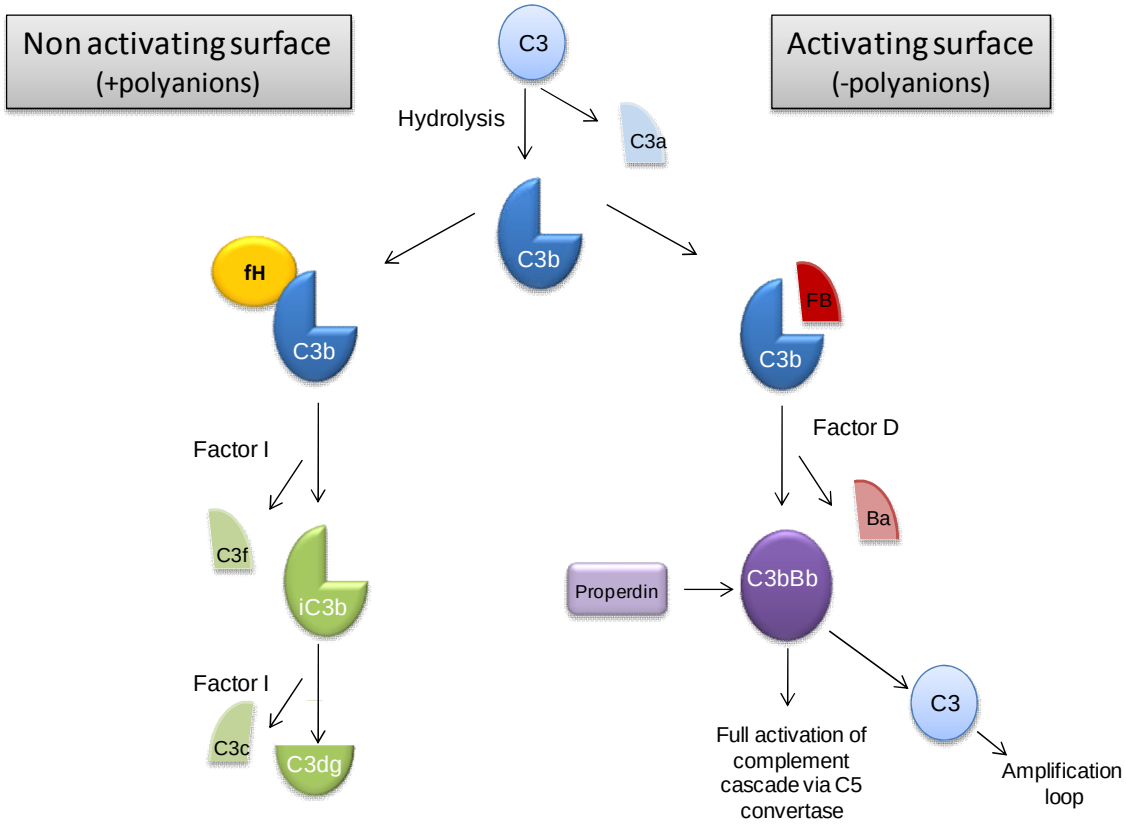
Figure 1.2 Activation of the complement cascade. Activation occurs via three distinct pathways. The alternate pathway is the only one that does not require exposure to foreign material, C3b being generated through continuous hydrolysis of circulating C3. This action exposes an internal thioester bond which is required for the subsequent attachment of C3b to target surfaces, and can also bind Factor B (homologous to C2 of classical/lectin pathway). Bound factor B can be activated by factor D, the resulting Bb combines with C3b to form C3 convertase. The classical pathway is initiated via recognition of immune complexes on target surfaces by C1q , binding which induces a conformational change in C1q which in turn activates serine proteases C1r and C1s, which goes on to proteolytically activate C4, homologous to the alternate pathways C3. The lectin pathway is initiated through recognition of sugars (e.g. mannose) on the microbial cell surface, recognised through mannose binding lectins (MBL) or ficolins which form a complex with mannose associated serine proteases (MASPs). It is MASPs that are activated as a result of recognition, and go on to activate C4 like the classical pathway, and subsequently C3 convertase.



The focus of this project is on the alternative complement pathway, in particular one of its regulatory proteins, and therefore further descriptions are limited to this pathway. The alternative pathway constitutes six major serum proteins: C3, factor B, factor D, properdin, factor I and factor H, the latter three which act to regulate the pathway. Initiation of the pathway can either be by foreign materials (such as bacterial or viral cells walls) or through the aforementioned tick over pathway by slow, continuous hydrolysis of circulating inactive C3 yielding C3a and C3b. The C3b component binds to cells, both foreign and host. In the presence of polyanions (sialic acid and heparin), commonly found on mammalian cell membranes but not on bacterial cell walls, C3b preferentially binds factor H (fH), a regulatory protein described in more detail below, inhibiting further assembly of C3 convertase. In the absence of polyanions, C3b preferentially binds factor B, and this complex serves as the substrate for enzymatically active serum protease factor D, which cleaves C3b releasing Ba, and in the process generating C3bBb, or C3 convertase (**Figure 1.3**).

Under normal physiological conditions, this C3 convertase has a short half life although this is extended by the binding of an additional protein properdin, which acts to stabilise alternative pathway C3 convertase. The presence of properdin is thus a marker of complement activation via the alternative pathway. C3 convertase can cleave C3, thus seeding an amplification loop of complement activation, a key strength and feature of this cascade. Full activation of the complement cascade then proceeds as C3 convertase binds C5 to generate C5b (and in the process releasing the potent anaphylotoxin C5a) which then binds C6 and initiates the formation of the membrane attack complex (MAC). Formed by the sequentially binding components C5a, C6, C7, C8 and C9, the MAC displaces membrane phospholipids forming a large pore in cellular membranes enabling free diffusion of ions and small molecules, which disrupt the function of the target cell, often leading to lysis. Steps subsequent to the formation of C3 convertase are often referred to as the final common pathway, as these steps are common to the complement pathway regardless of how it has been activated.

Figure 1.3 Activation and regulation of the alternate complement pathway Circulating (inactive) C3 is hydrolysed to C3a (an anaphylotoxin which mediates downstream adaptive responses) and C3b. C3b possesses an active thioester bond, which enables C3b to bind covalently to hydroxyl groups on nearby host cell surfaces. Under such conditions fH preferentially binds C3b. This complex is recognised by factor I, which cleaves C3b into inactive products. Alternatively binding of C3b to bacterial cell walls, in the absence of surface heparin and sialic acid leads to C3b preferentially binding to factor B (FB), forming a complex which is recognised and cleaved by the protease factor D, leading to the formation of C3 convertase (C3bBb). C3 convertase of the alternate pathway is stabilised by properidin, increasing its half life. C3 convertase additionally seeds the amplification loop by cleavage of more C3 leading to the deposition of more C3b on host and foreign cell surfaces. In addition to the regulation mediated by fH and factor I, the alternative pathway can also be regulated through action of DAF, which accelerates decay of C3b, and CR1 which is also a cofactor for factor I mediated neutralisation of C3b



Regulation of the alternative pathway is key, given that C3 convertase can be produced under normal physiological conditions without exposure to any foreign material. Prevention of the assembly of C3 convertase is mediated by a number of the RCA proteins, including complement receptor 1 (CR1), membrane cofactor protein (MCP), decay accelerating factor (DAF) and fH, (the focus of this thesis, and the only one of these regulators specific for the alternate pathway). CR1, DAF and MCP also regulate the classical and lectin pathways. In general, regulatory proteins act either by promoting irreversible dissociation (decay

acceleration) of the convertase enzymes, or by acting as co-factors for enzymatic cleavage. CR1, DAF and fH, all act to dissociate C3b from Bb, as well as acting as co-factors for factor I in the processing of C3b (**Figure 1.3**). Regulation of the terminal pathway (subsequent to the formation of C5b) occurs through regulators clusterin and vitronectin and CD59. Lack of any of these regulatory proteins may therefore lead to the unchecked activation of the complement cascade, mediating injury to host cells. Specific lack of fH or loss of its function could be particularly problematic given the continuous, albeit slow, production of C3b.

1.1.1.5 Complement and atherosclerosis

Complement components and their regulators have both been implicated in atherosclerosis. Normal arterial intima lacks signs of complement activation, therefore findings from a number of studies demonstrating complement is associated with atherosclerotic lesions suggests that complement activators may be generated or deposited during atherogenesis. Complement is known to be activated by a number of factors that are present in the atherosclerotic plaque over its life course including modified LDL-C deposits, apoptotic cells and C-reactive protein (CRP).

The location of complement components in post mortem human atherosclerotic plaques, detected by immunohisochemistry, is such that the MAC (indicating full activation of complement), and regulatory proteins associated with MAC (clusterin and vitronectin), are found in the deeper musculo-elastic layer of the intima, whereas regulatory proteins which operate at the C3 convertase level, including fH and C4bp are limited to superficial proteoglycan-rich areas, consistent with the localisation of LDL-C in deeper layers [17]. Moreover, in these experiments MAC was found to co-localise with properedin suggesting complement activation via the alternate pathway. Evidence of full activation of complement, has been linked with severity of atherosclerosis [18, 19], leading to the suggestion that complement activation could be the key initiating event in atherosclerosis, a notion that is supported by findings in animals that suggest complement activation occurs temporally with deposition of lipids in arteries, and more importantly, that this precedes the accumulation of components of cell mediated immunity [20].

A number of other lines of investigation in humans have linked complement to atherosclerosis. These include the expression of membrane bound complement regulators on cells resident in the atherosclerotic plaque (DAF on vascular smooth muscle cells and macrophages, CD59 on T cells and macrophages, CR1 on macrophages) which implicate local activation of complement [21-23] Complement has been shown to be activated in plaques prone to rupture, moreover high circulating levels of anaphylotoxins and terminal complex regulators (CD59) have been reported in individuals presenting with acute coronary syndromes [24-26]. Elevated levels of C5a have been shown in one study to independently predict future major cardiovascular events in patients with documented atherosclerosis [27], and circulating C3 concentration in men has been shown to be associated with established markers of complement activation [28]. C4 concentrations have also been shown to be higher in individuals with acute coronary syndrome and stable angina [29].

Animal studies have supported a role of complement in both acute MI, where complement depletion in a model of myocardial injury in non-human primates [30] and in the mouse [31] limits infarct size; as well as in more long term development of atherosclerosis, for example where the expression of DAF decreased the burden of atherosclerosis, possibly through regulation of C3 convertase in apolipoprotein E (ApoE) and low-density lipoprotein receptor (LDL-R) double knockout mice [32, 33].

More recently, high-density lipoprotein cholesterol (HDL-C), which exhibits an inverse association with CHD in prospective observational studies [34], has been shown to be bound by a number of components of the complement system using proteomic analysis of human plasma [35]. These include C3, C1 inhibitor, fH, C4a, C4b, C9, clusterin and vitronectin, indicating that HDL-C may modify complement activation in some way. Moreover genes that play a role in HDL-C regulation including hepatic lipase (*LIPC*), cholesteryl-ester transfer protein (*CETP*), lipoprotein lipase (*LPL*), and ATP-binding cassette transporter (*ABCA1*), have recently been implicated in AMD [36, 37], supporting shared mechanisms contributing to both atherosclerosis and AMD.

Taken together these data suggest that complement activation might promote development of the atherosclerotic plaque, while the regulatory components might inhibit atherogenesis.

1.1.1.6 Limitations of observational evidence linking inflammation to CHD

Inflammation and atherosclerosis have been studied *in vitro*, in animal models, and in man. Numerous components of the inflammatory response have emerged as potential candidates for use as biomarkers of cardiovascular risk with potential clinical application as predictors or as therapeutic targets. These include for example CRP, fibrinogen, interleukins and leucocyte cell count, [38, 39]. Links between inflammation and CHD have also been made, for example individuals with an acute infection are at a transiently increased risk of MI or stroke [40], and individuals with chronic inflammatory disorders such as rheumatoid arthritis and systemic lupus erythematosus are at increased risk of cardiovascular events later in life [41, 42].

However, most of the evidence linking inflammation and markers of inflammation with CHD comes from observational epidemiology, and such associations are prone to confounding (where associations between an inflammation marker and CHD are not causal but explained by a common association of the two with a causal factor (e.g. smoking)). They may also be affected by reverse causation bias, where atheroma provides a stimulus for inflammation rather than being the target of it. Thus the causal relevance of inflammation in general and inflammatory mediators specifically in atherosclerosis remains uncertain [39, 43].

One way in which the causal relevance of the associations could be confirmed of refuted is by studying the effect of genes associated with activation or deficiency in inflammatory pathways [44-46]. However, genetic studies of this type have previously however been inconsistent in their findings, largely due to the limited number of genetic variants tested and by the low power of individual studies.

1.1.1.7 Overcoming limitations of observational evidence

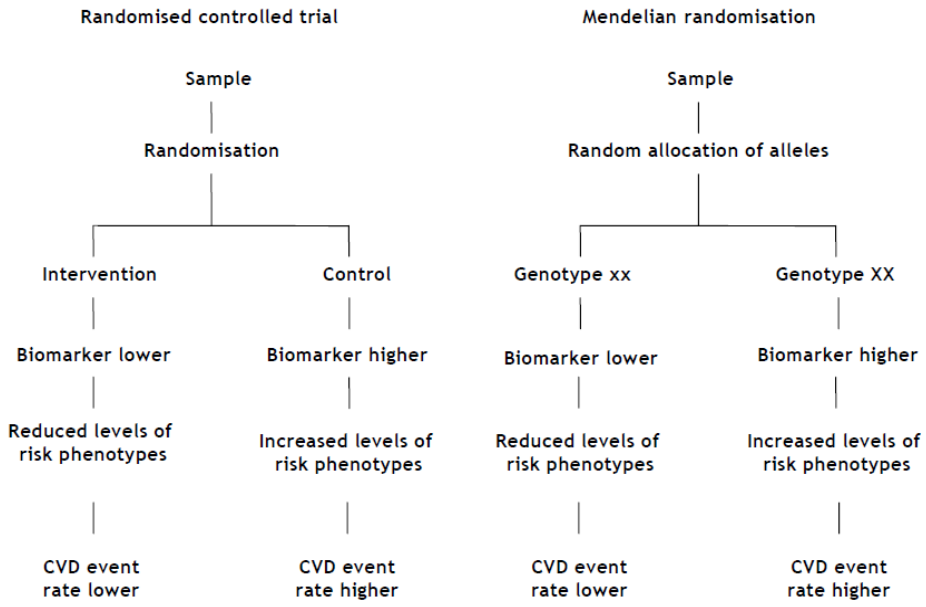
Another way in which to establish whether or not an inflammatory molecule is causally related to disease would be to carry out a randomised controlled trial (RCT) using a therapeutic intervention that modifies the concentration or fraction of each implicated mediator selectively. Randomised intervention studies are free from many of the biases of observational associations. The randomised treatment allocation balances any confounders between treatment groups, and reverse causation is mitigated because the outcome is assessed

following the intervention. Through the use of RCTs, the apparently protective role of anti-oxidants in CHD was shown to be unfounded [47], drugs targeting tumour necrosis factor (TNF), which reduce the inflammatory burden in rheumatoid arthritis, were shown not to confer protection on CHD, and treatment with antibiotics against *Chalmydia spp*. was not found to protect against CHD [48, 49]. More specifically, to the address the causal role of the complement cascade, monoclonal antibodies that target specific components of the complement cascade e.g. pexelizumab, that targets complement component C5, was shown not to be beneficial as an adjunctive therapy to primary-percutaneous coronary intervention in acute MI [50, 51]. These are important negative findings in the CHD arena and perhaps more importantly, highlight the value of defining the precise causal pathways that are of patho-physiological significance in CHD, to allow these to be targeted by appropriate therapies.

RCTs however cannot be used to test every one of the very large number of biomarkers or proposed mechanisms implicated in CHD by observational epidemiology. The life-cycle of any drug from discovery to phase III studies is 15-20 years, with estimated costs of its development running into billions of pounds [52]. One way of exploiting the benefits of randomisation but without resorting to the expensive and lengthy development of a new drug by conducting a RCT, is to harness the random assortment of genotype during gamete formation and fertilisation (Mendel's second law). Natural variation in a gene that alters the concentration or function of its encoded protein is transmitted at random from parent to offspring. This randomisation should balance the distribution of known and unknown confounders between genotype groups, just as confounders are balanced in treatment and placebo arms of a RCT. Genotype is fixed throughout life and non-modifiable and therefore genetic associations are not affected by reverse causality. In essence a genetic study is therefore akin to 'natural' randomised controlled trial (**Figure 1.4**) [45, 53]. Allocation in this case is not to placebo or drug but to a genotype that influences an encoded protein or downstream biomarker. If a biomarker is causally involved in CHD (or any other outcome being investigated) individuals with genotypes that associate with higher levels of the marker, should have a higher risk of CHD, commensurate with the effect of the change in the biomarker [54, 55]. This use of genetic studies as tools to dissect causal association is commonly known as Mendelian randomisation analysis and has already been applied to a

number of putative markers in the cardiovascular arena. For example it is overturning the firmly held views that the inflammatory markers CRP and fibrinogen are likely causally linked to CHD [56-58], whereas the evidence using the same approach indicates that lipoprotein (a) is causally related to MI [59]. If genetic variants were identified that influence the level or function of complement components, the approach could be applied to test the causal nature of complement proteins and their regulators in CHD.

Figure 1.4 Parallels between randomised controlled trials and Mendelian randomisation Comparison of randomised evidence from clinical trials and Nature's randomised trials. A hypothetical drug that lowers a deleterious biomarker would be hypothesised to be protective from CVD according to current observational evidence, although this is unknown. As an alternative or adjunct to inform drug development, study of randomly allocated single nucleotide polymorphisms (SNP) in the gene encoding the deleterious marker may inform its causal nature. Figure taken from Hingorani AD, Humphries SE, Lancet 2005 [45].



1.1.2 Age-related macular degeneration

1.1.2.1 Disease burden

Age-related macular degeneration ranks as the third cause of visual impairment worldwide [60]. It is estimated that approximately 35% of individuals over the age of 75 in the United States of America have some degree of AMD [61]. This will increase with the demographic trend towards an aging population in middle and high income countries. AMD will therefore

become a major economic and public health burden [62], and a greater understanding of its aetiology together with development of improved therapies and preventative strategies is therefore of great importance.

1.1.2.2 Risk factors for AMD, and potential overlap with CHD

Despite the burden of AMD, few factors have been reliably identified as altering the course of disease once early signs are present. The most consistent link is with cigarette smoking [63-69]. Associations also exist with blood pressure [63, 64, 70], lipid levels [63, 64, 70, 71] abdominal obesity [63, 64, 72], physical activity [63, 64, 70, 72] and dietary factors including consumption of fat [73]. There are inverse associations with consumption of caretenoids and antioxidants [74, 75]. All of these have also been proposed as risk or protective factors for CHD, suggesting the aetiology of the two disorders may overlap. Compared to the evidence base in CHD, the epidemiological foundations of these associations in AMD are still maturing. Early associations were reported from small cross sectional and case control studies, and thus prone to chance findings and bias. Only more recently have large prospective cohorts been established whose primary focus is on dissecting the precise epidemiology, disease progression and molecular epidemiology of AMD [76-79]. However, even in the early phase, findings from these studies need to be interpreted cautiously. The low incidence of late stage AMD, which poses the greatest threat to vision, means the power to detect and quantify associations of exposure and disease outcomes in such studies is low. Only with a large evidence base can the determinants of AMD be reliably identified. Thus some additional risk factors, for which the evidence is currently inconclusive, may still, emerge as being common to both AMD and CHD.

The ability to precisely delineate the risk factors for AMD may have also been somewhat limited by the more heterogeneous nature of AMD given its distinct subtypes and clinically identifiable early and late stages. Early AMD may never proceed to the late AMD, which is associated with greatest visual loss. This division into early and late stages is similar to descriptions of lesions that define CHD. Fatty streaks could be analogous to early AMD changes, not all progressing to late stage atherosclerotic plaques. Once plaques have developed, there may be further stratification into those more likely to rupture or cause flow

limiting symptoms. However given that coronary arteries are not amenable to non-invasive visualisation unlike the retina, the focus in CHD has remained on clinical events, which might be analogous to visual loss in AMD. However, the identification of early atherosclerotic lesions is now of interest in CHD, with development of techniques enabling measurement of intimal media thickness in vessels such as the carotid artery (CIMT), coronary artery calcification (CAC) by use of computed tomography (CT) scanning, and detection of plaque by the more invasive intravascular ultrasound (IVUS) in the coronary artery. These are techniques currently reserved for research and thought of as surrogate measures for 'hard' cardiovascular events, and their clinical utility as predictors of disease risk, or as outcome measures in clinical trials is the subject of much debate [80].

As in CHD, RCTs have also played some role in deciphering the causal nature of some of the proposed risk factors for AMD. The Age-Related Eye Disease study (AREDS), a randomised study found support for prior observations that dietary supplementation with high doses of vitamin C, vitamin E, beta-carotene and zinc are effective at lowering overall risk of late AMD once it is established, but are not effective in preventing the development of new lesions [79]. However these effects were seen in a post-hoc analysis of small sub-groups defined according to disease severity in which there were few events. In contrast a role for anti-oxidant and vitamin C supplementation in reducing the risk of CHD events was not supported by RCTs, despite high expectations of success based on prior observational studies [81].

Another common pathway of interest in both AMD and CHD is the biology of vascular endothelial growth factor (VEGF). Blockade of VEGF signalling using monoclonal antibodies [82] or antisense RNA to VEGF-A receptor [83] has been found be beneficial in RCT's of late AMD, probably by inhibiting the formation of new and rupture prone vessels thus reducing the progression of visual disability. The gene encoding VEGF (*VEGF)* has been investigated as a candidate gene for AMD, and in small studies polymorphisms in *VEGF* have been associated with an increase in AMD risk [84, 85], although these finding are not always replicated in larger studies [86]. Similarly *VEGF* polymorphisms have been associated with CHD (see [87] for a summary of all studies), although the exact role of therapeutic manipulation of VEGF in atherosclerosis is controversial given that it can both promote atherosclerosis through pro-

inflammatory pathways but may also have beneficial effects through re-canalisation of occlusive lesions (see [88] for a critical appraisal in this field).

In summary, in terms of the understanding of aetiology and identification of risk factors for disease prediction the proposal that there may be shared risk factors for both AMD and CHD is plausible but there are several potential sources of error and bias in the epidemiological studies of AMD so far, though common associations with several risk factors lends support to the partially shared pathological basis for these two disorders. In the subsequent sections, I detail the pathological features of AMD, drawing further comparisons with CHD.

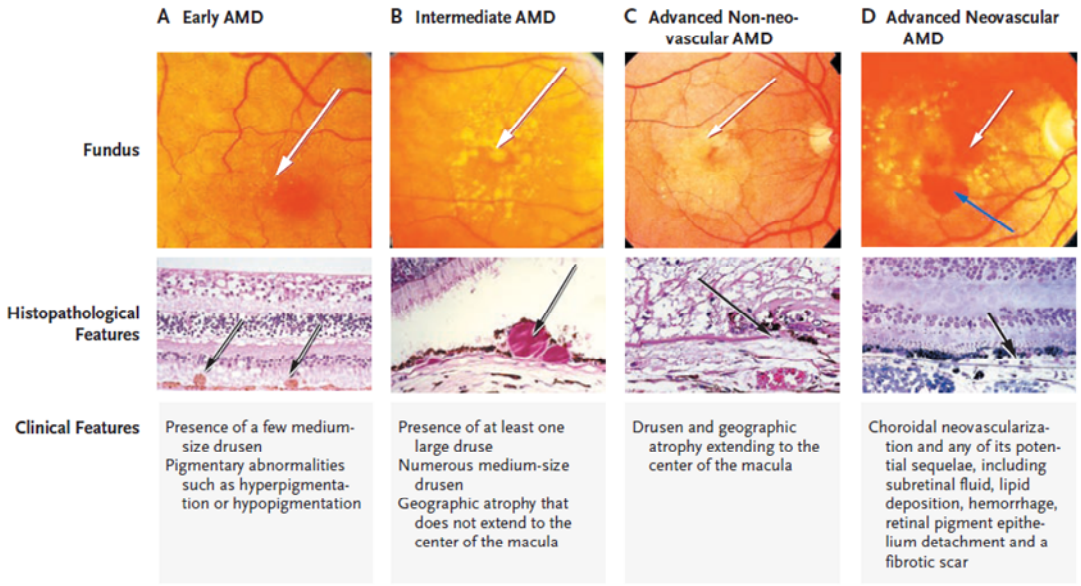
1.1.2.3 Patho-physiology of AMD

The pathological lesion of AMD centres on the area of the retina known as the macula, a small (~6mm diameter), densely packed area of photoreceptor cells (rods and cones) located at the centre of the fovea. This area of the retina is critically responsible for the integrity of the central visual field. The retina can be thought of as two functional units, one comprising photoreceptor cells and their neuronal connections, the other providing support and vascular supply to these cells. The layer underlying photoreceptor cells comprises epithelial cells, known as the retinal pigment epithelium (RPE) that nourish the retinal cells and below lies a basement membrane, called Bruch's membrane. These structures maintain a barrier between the retinal cells and the underlying vascular supply from the choroid. This interface between the photoreceptor cells and vascular supply is where the pathological changes that define AMD are thought to occur [89].

Clinically, AMD has been sub-divided into early and late forms. Early disease, also known as age-related maculopathy (ARM), is diagnosed by the presence of pathological lesions known as drusen, which are amorphous extracellular deposits, and/ or changes in the RPE, including hyper-pigmentation and detachment (**Figure 1.5**). These changes can be asymptomatic (similar to the early lesions of atherosclerosis), although they begin to develop much later in life than fatty streaks in arteries that can be seen from as early as the second decade of life. Drusen appear as punctuate yellow lesions, however these lesions are not specific for AMD, and can be found as a normal feature of ageing. Lipids are a common constituent of both fatty

streaks and drusen, with drusen being first described on the basis of their appearance as “Colloidkuglen” or colloid spheres, rich in “fettreichen” or fat [90]. More recent immunohistochemical and chromatographic techniques have confirmed the presence of lipid, including cerebrosides, gangliosides, esterified and unesterified cholesterol. There is also evidence of local transcription of genes for lipoprotein components (e.g. apolipoprotein E). However there is uncertainty around the exact origin of lipid components within drusen, in particular whether these are derived from the circulation, or synthesised locally (all these findings are reviewed in [91]).

Figure 1.5 Clinical and histological features of sub-classes of AMD. Taken from Jager RD et al NEJM 2008 [92]



Prospective clinical studies have catalogued features of drusen that are more likely to be associated with progression to late stage AMD. These include size, number, confluence and appearance (with a distinction being drawn between hard and soft drusen), and the presence of associated changes in the RPE [93]. These finding contribute to the classification systems for clinical AMD [94].

Individuals who do progress to late disease can present with one or a combination of three clinical types: either ‘dry’, ‘wet’ or mixed AMD. Dry AMD or geographic atrophy (GA) is thought to result from cell death within the RPE, progressing to atrophy of photoreceptor cells. Wet AMD or choroidal neovascularisation (CNV) is thought to be due to the formation of new

vessels sprouting from existing choriocapilliaries, which extend into the areas below the RPE and photoreceptor cells. These new vessels are prone to damage causing local haemorrhage and exudative changes, which can result in sudden visual loss. Advanced stages of disease are characterised by a fibrovasular or atrophic macular scars and are associated with permanent damage to central vision (reviewed in [95]). Although CNV accounts for 10% of overall AMD cases, it importantly contributes to 90% of visual loss [96]. Strategies to target late stage disease overall would potentially save the site of many individuals.

1.1.2.4 Inflammation and AMD

Like CHD, many of the pathological changes in AMD have been attributed to inflammation, with identification of the contents and mechanisms which lead to the formation of drusen being a focus of current research [91]. The formation of drusen has been thought of as an inflammatory response to damage to the RPE, analogous to the formation of fatty streaks following endothelial cell damage in atherosclerotic arteries. Evidence accrued from human and animal studies has implicated local inflammatory and immune-mediated events in the development of drusen and, by extension, in the development of AMD. Components of the inflammatory response identified in drusen include: ubiqutin, intergrins, tissue inhibitor of metalloproteinaises, fibronectin and components of the complement system including regulators such as components of the MAC, and its regulator vitronectin; complement component C3 and fragments associated with its cleavage as well as other regulators complement including MCP, CR1, fH and clusterin (reviewed in [91]). In addition proteins that are associated with other chronic diseases of ageing including beta-amyloid (associated with Alzheimer's disease), ApoE (associated with both Alzheimer's disease and CVD), and generalised inflammation including MHC class II antigens, have all been shown to be present in drusen [97]. Given that access to circulating complement components may be limited in the brain and neural retina it has been hypothesised that local synthesis may occur [97]. Quantitative methods to assess gene expression of elements of the complement cascade have been carried out with the finding that components that comprise alternate and classical pathway are expressed by the human RPE-choroid complex, but with little evidence to support local generation of lectin and terminal complement pathways [98].

Like atherosclerosis higher concentrations of circulating markers of sub-clinical inflammation including CRP and IL-6 [99, 100] have been found to be associated with a higher risk of AMD, and other circulating biomarkers are also associated with both diseases including homocysteine. Candidate gene studies have also implicated certain inflammatory genes in AMD, as well as other genes which are also implicated in atherosclerosis (e.g. *VEGF, LRP6, APOE*, *VLDLR*, reviewed in [101])

1.1.2.5 Genetic association studies in AMD

The unique attributes of genetic studies for causal inference as well as the limitations of genetic association studies have been discussed previously in the context of CVD. Sequencing of the human genome and subsequent cataloguing of common variation in the genome through the HapMap project [102] (hapmap.ncbi.nlm.nih.gov), has lead to development of technologies that allow the efficient typing of hundreds of thousands of SNPs allowing the conduct of GWAS. These studies overcome the some of the biases inherent in candidate gene studies. In GWAS, several hundreds of thousands of variants (up to 1 million at the time of writing) can be typed in several thousand cases and controls to identify genetic variants that differ between cases and controls providing valuable insights into the genetic loci that may underlie susceptibility to complex diseases [103].

One of the first of this new generation of genetic studies has provided the most compelling evidence of a link between inflammation and AMD and, in particular, a link between complement and risk of AMD [10, 104-106] . In 2005 a GWAS of ~90 late AMD cases and 90 controls found a highly significant association of a SNP in the complement factor H (*CFH*) gene with AMD [10]. The gene product is essential in regulation of the alternate complement pathway. This association has been independently replicated in numerous studies [107-110]. The particular SNP, encoding a tyrosine to a histidine change at residue 402 (Y402H), has been thought of as a putatively a causal variant, because it lies within a functional domain of the protein implicated in the binding of fH with heparin and CRP. Although many SNPs in this and adjacent genes have been associated with risk of AMD, many are in strong linkage disequilibrium; it is the association with this lead SNP in particular that has been the most studied. Other complement related genes have also been found to be strongly associated with

altered risk of AMD based on candidate gene studies. These include the complement regulators factor B (*CFB*) [111], C3 [112] and *SERPING1* also known as C1 inhibitor [113]. In addition, the genes neighbouring *CFH*, the *CFHR*s (*CFH* related genes) which share homology with *CFH*, having arisen through gene duplication, have also shown association with AMD. For example deletions in the *CFHR1* and *CFHR3* genes confer protection from AMD [114].

The 2005 AMD GWAS used a 100,000 SNP array, but a more recent GWAS using a denser 500,000 SNP array validated the association of SNPs in the *CFH* gene cluster with AMD [36, 37]. In addition SNPs in other complement related genes, which previously emerged from candidate gene studies (including *C3 C2/ CFB* and *CFI*) were confirmed to be associated with AMD. Intriguingly genes involved in the regulation of HDL-C (for example *CETP, LIPC, LPL* and *ABCA1*) some of which are also been associated with CHD risk [115] were also found to be associated with AMD [36, 37]. The preliminary finding of genes that appear common to both AMD and CHD provides additional evidence that both disorders may share a common pathological basis. The rs1061170 *CFH* SNP encoding Y402H was also subsequently found to be associated with MI in two candidate gene studies. However, the magnitude of effect on CHD was much smaller than that for AMD [116, 117]. These findings led me to hypothesise that complement factor H could therefore be a shared risk factor for both AMD and atherosclerosis.

1.2 Complement factor H: genetic variation and functional effects

Despite the consistent genetic association of SNPs in *CFH* with AMD, the biological pathway linking variation in the *CFH* gene with the development of the retinal lesions is as yet uncertain. Regulatory SNPs in the vicinity of the *CFH* gene could alter *CFH* gene transcription or mRNA stability leading to an alteration in the concentration of circulating fH. Alternatively, or additionally non-synonymous coding SNPs in the gene itself (for example the variant encoding Y402H) could alter the function of fH. There is substantial precedent for circulating protein biomarkers to be regulated by common genetic variants located in the vicinity of the encoding gene (i.e. *cis*-acting variants). The evidence for this comes from identification of *cis*-acting regulatory variants: (i) through candidate gene studies (e.g. for CRP [56], and fibrinogen [118]

(ii) GWAS of multiple blood proteins [119, 120] and (iii) GWAS of messenger RNA expression profiles [121]. However at the time this work was initiated there were no studies of the association between *CFH* variants and fH level, of the non-genetic determinants of fH in the population, or the relationship between fH concentration and the risk of AMD or CHD. If the association between genetic variation in *CFH* and risk of AMD were mediated through differences in the amount of fH produced, fH itself would emerge as a potential biomarker and therapeutic target for treatment of prevention of AMD, that may be relevant not only to individuals who may carry *CFH* gene variants, but more generally to all those with AMD.

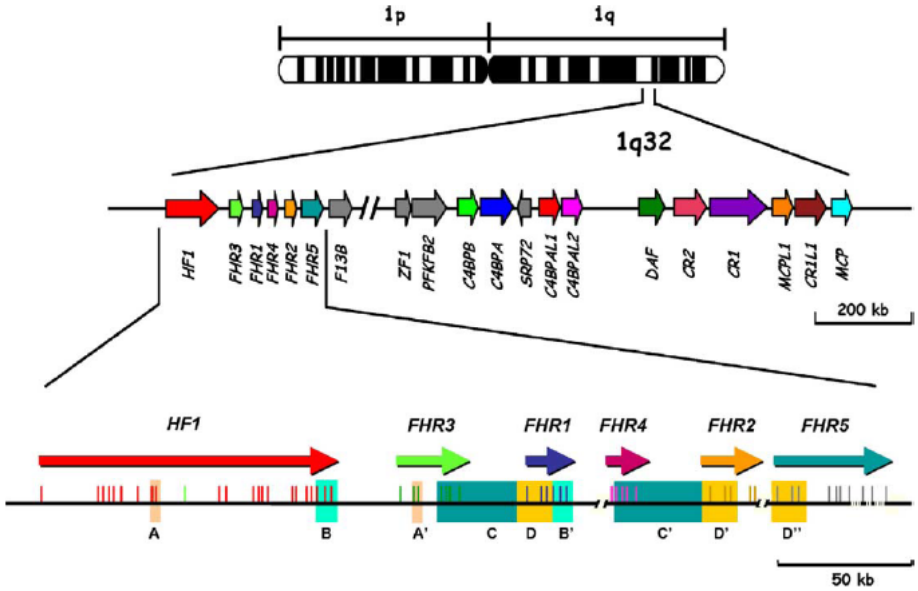
1.2.1 Structure and function of fH

First identified in 1965, complement factor H (fH) is a ~150 kDa protein. A single polypeptide chain plasma glycoprotein [122], fH is composed of 1213 amino acid residues, arranged in 20 repetitive units termed short consensus repeats (SCRs), each of ~60 amino acid residues in length. Functional domains of fH include three C3b binding sites in SCRs 1-4, SCR 12-14 and SCRs 19-20; and three binding sites for heparin/ sialic acid in SCR7, SCR 13 and SCRs 19-20. Thus SCRs 19-20 houses overlapping sites for both polyanion and C3b binding (**Figure 1.7**). A binding site for the acute phase protein CRP has also been identified on SCR7, and it has been proposed that binding to CRP may arrest the alternative pathway. The N-terminal SCRs 1-4 appear to be essential for factor I cofactor activity [123] and C-terminal SCRs 19-20 appear to be the most important for preventing alternate pathway activation on host cells [124]. Recombinant fH lacking SCRs 16-20 cannot function adequately, although removal of the C3b binding site in SCR 7 or SCRs 12-14 only reduces function modestly [125]. The Y402H SNP, implicated in AMD lies in SCR 7, and prior to GWAS had already been highlighted as a potentially important mutation with functional consequences, because of the substitution of a positively charged histidine for a non-charged tyrosine. This substitution event could have significant functional implications, as this amino acid change occurs in a cluster of positively charged amino acids, disruption of which could potentially alter binding properties to cell surface components which determines the further activity of fH [126].

Factor H is encoded by a single gene on chromosome 1q32 in the RCA gene cluster. The cluster contains separate genes for the structurally similar complement regulatory proteins

including MCP, DAF and CR1. In addition, genes whose proteins show marked homology with *CFH*, *FHR1-5* are located immediately downstream of *CFH* (**Figure 1.6**) *CFHRs* are thought to have arisen by genomic rearrangement, and although their precise function is not known, some studies suggest that like fH they may have anti-inflammatory actions. Naturally occurring haplotypes incorporating deletions in *CFHR1* and *CFHR3* have been shown to have a protective effect in AMD. *CFH* itself comprises 23 exons spanning over 94 kb genomic DNA. Each SCR is encoded by a single exon in *CFH*, apart from SCR2 which is encoded by exons 3 and 4 together. Exon 10 is only transcribed as part of the smaller (43 kDa) and less abundant alternatively spliced product of *CFH*, termed factor H like-1 protein (FHL1).

Figure 1.6 The Regulation of Complement activation gene cluster. The RCA cluster includes more than 60 genes, 15 of which are complement related, arranged in tandem forming 2 groups separated by a number of genes un-related to complement. Within the group that houses *CFH* and *CFHR*s there are large genomic duplications (indicated as A-D), which include different exons of the *CFH and CFHR 1-5* genes. Schematic taken from Rodriguez de Cordoba et al., Molecular Immunology 2004 [127]



A key regulator of the alternate complement pathway at the level of C3 convertase, fH functions in accelerating decay of C3b and acts as a co-factor for factor I accelerating enzymatic cleavage of activated C3 (**Figure 1.3**). Soluble fH acts to dissociate circulating C3b

(decay acceleration function), and membrane bound fH binds C3b, forms a complex with factor B, that is recognised by factor I, which cleaves C3 into inactive components iC3b (which remains surface bound) and can longer participate in the formation of C3 convertase. Factor H preferentially binds C3b on host cell surfaces, recognising C3b in conjunction with anionic clusters on host cell surfaces (heparin and sialic acid). Conversely C3b binding to microbial cell surfaces lacking these structures increases the affinity of C3b to bind factor B which leads to formation of C3 convertase, via the action of factor D (**Figure 1.3**). Factor H is therefore considered to protect host cells from complement mediated damage while leaving pathogens open to complement mediated attack. Micro-organisms have however evolved surface molecules which bind fH, similar to polyanions on host cells, thus evading complement mediated destruction. Pathogens that have developed such surface molecules include *Streptococcus pyogenes* [128]*, Neisseria meningitides* [129]*, Borrelia burgdorferi* [130].

Null mutations in fH have been identified that play a role in renal diseases including atypical haemolytic uraemic syndrome (aHUS) where mutations alter the membrane binding capability of fH, and reduce protection to complement mediated damage, and membranoproliferative glomerulonephritis type II (MPGN II), characterised by uncontrolled activation of C3. The *CFH* knockout mice provide a disease model for human MPGN II [131]. Aged *CFH* knockout mice also develop a phenotype akin to AMD, although this is a difficult disease to model in mice given the absence of a macula in this species [132].

Synthesis of both fH and fHL1 is largely in the liver [133], and one study to date has measured fH in a moderately sized population, indicating that circulating concentrations are in the range 110-615 mg/L [134]. The alternatively spliced fHL-1 circulates at a much lower concentration of 10-50 mg/L. Extra-hepatic synthesis has also been shown and, in particular, local synthesis in the RPE, choroid and neural retina, although expression of the alternatively spliced variant has not been studied in such depth. fHL-1 is thought to have complement regulatory activity with an N terminal region homologous to fH (**Figure 1.7**). Five additional proteins share homology with fH although these are the products of different but related genes, *CFHR 1-5*. Functions of the fHR proteins are as yet un-certain. Although they share homology with fH, they lack N- terminal regions that define the regulatory activity of fH of decay accelerating and

co-factor functions (**Figure 1.7**). However complement modulatory activity has been ascribed to some of the fHR proteins [135, 136], and these functions may well extend to others in this family of proteins. It has been suggested that CFHRs have functions which might be co-operative or competitive with fH, but these are yet to be fully defined. More recent evidence suggests that novel a mutation in *CFHR5* predisposes to renal glomerular disease (termed CFHR5 nephropathy) in individuals of Cypriot origin, suggesting there is an overlapping role for factor H and related proteins [137].

1.2.2 Genetic regulation of circulating fH

One study to date has investigated the effects of genetic variation on fH level [134], reporting substantial inter-individual variation in fH concentration (100-600 mg/L), a heritability of 0.6 and using microsatellite markers to conduct linkage analysis that fH is likely to be regulated by a quantitative trait locus (QTL) mapping to chromosome 1q within a region that corresponds to the RCA cluster. Independent mapping of a QTL for fH and a susceptibility locus for AMD to the RCA cluster provides some evidence that variation in the same region, perhaps in the vicinity of *CFH* itself influences fH concentration, and perhaps even that SNPs influencing fH level and AMD risk might be concordant. If this were to be the case, it would provide strong evidence that alterations in fH concentration play a pathogenic role in susceptibility to AMD. However, the exact causal gene was not identified because of the low resolution of linkage studies. Given the precedent for circulating biomarkers to be regulated by common genetic polymorphisms in the vicinity of the encoding genes, I hypothesised that genetic variants may exist in the vicinity of the *CFH* gene and that this might explain the observed association of SNPs in this region with AMD.

Figure 1.7 Homology between factor H, fHL-1, and members of the CFHR protein family., a) Schematic view of complement factor H protein, delineating regions involved in regulation (shaded green) and cell surface binding (shaded blue). Regions involved in binding to C3b and heparin are indicated. The Y402H variant lies in a region that is implicated in C3 binding and in an interaction with CRP. (b) Shared homology with alternative splice variant FHL1 is shown. Critically, FHL1 lacks key surface binding domains although it is postulated to have some regulatory functions. (c) CFHR 1-5 are shown with homology to fH protein indicated by the overlap of SCRs sharing similar activities that are shaded in the same colour. Although the precise function of CFHRs is yet to be determined, structural homology suggests that they may have similar, competitive or even independent actions to fH. CFHR4 has two alternatively spliced forms.

1.3 Hypothesis

Based on the preceding background, I developed the following specific hypotheses:

- Common variants in *CFH* in particular the Y402 polymorphism is associated with risk of both CHD and AMD
- The association of *CFH* variants with AMD, and perhaps CHD, are mediated through changes in circulating fH
- fH may be a therapeutic target and/ or a predictive factor for AMD (and perhaps CHD)

1.4 Aims

My initial aim was to carry out a systematic review and meta-analysis of genetic association studies using both published literature and newly genotyped studies, to more precisely estimate the association of the *CFH* rs1061170 SNP encoding Y402H with AMD and CHD, to ascertain whether this variant was a risk factor for both disorders.

A further aim was to develop and validate an assay for circulating fH in order to (i) study the genetic and non-genetic determinants of fH using stored blood samples, (ii) to examine the association between circulating fH and the risk of AMD and CHD.

Completion of these studies would allow assessment of the consistency of the associations of genetic variation at the *CFH* locus with fH level, and fH level itself with disease risk (Mendelian randomisation), to test if fH is causally involved in both CHD and AMD, and thus if it is a shared risk factor for both disorders.

2 Genetic variation in complement factor H and age-related macular disease: Assessment of effect size, modifiers, and relationship to disease subtype using meta-analysis

Late stage AMD is an important cause of visual loss in the elderly. Early identification of individuals at greater risk of disease and, in particular, the most sight threatening forms of AMD might help target current preventative interventions to those at highest risk. Moreover, better understanding of the causes of AMD could lead to the development of improved treatments. A susceptibility locus on the gene that encodes complement factor H has been identified in those with AMD [10, 104-106]. The most replicated association has been with the SNP: rs1061170 encoding a tyrosine to histidine change at position 402 (Y402H).

rs1061170 is a common variant, the risk allele (C), encoding fH-402H, and has a frequency of 38%. Given that this variant is associated with a large effect on AMD, it has been proposed that genotyping of this variant may be helpful in predicting future risk of AMD [138]. However a number of existing uncertainties require resolution before a genetic test of this type could be recommended for use in clinical practice. First, the precise magnitude of the effect size requires better delineation as large effect sizes in initial 'discovery' studies can become attenuated as the literature matures [139]. Second, it is unclear if the risk conferred by carriage of this variant is the same for all grades/ severity of AMD. Third, although it has been suggested that smoking (a recognised risk factor for late AMD) could modify the effect of the rs1061170 variant, individual studies have not been large enough to address this question reliably. Fourth, it is uncertain whether rs1061170 is itself a causal variant or simply marks another causal site (because of linkage disequilibrium (LD)), either within or outside of *CFH*. Since LD patterns vary among populations of differing ancestry, the magnitude of association and in turn the predictive performance of rs1061170 could be very different in populations of differing ancestry. Moreover, if rs1061170 marks a causal site in an adjacent gene, of which several exist in the region of complement activation (RCA) cluster on chromosome 1, targeting fH or its downstream effects may not provide a useful means of treating or preventing AMD.

To address these uncertainties the association of rs1061170 with AMD was tested in two new genetic association studies that had not previously contributed to GWAS of AMD, together

totalling 2,222 cases and 795 controls. These studies were further incorporated into meta-analysis of 22 prior studies (14,174 cases/ 26,494 individuals) of which 18 reported information on the late sight-threatening forms of AMD and 9 studies (~6500 cases) provided either participant or aggregate level information.

The aims of this chapter were to more precisely quantify genetic effects of rs1061170 polymorphism on AMD risk in individuals of European and non-European descent, assess the effect of genotype on different grades and severity of AMD, and the potential interactive effect of smoking on AMD. Collectively this substantially updates and extends the prior meta-analysis [140] in this area.

2.1 Methods

2.1.1 Newly genotyped Studies

2.1.1.1 MRC AMD Case Control Study

A prospective case control study was conducted in 1,469 unrelated individuals of European ancestry (1234 cases and 194 controls) recruited from Moorfields Eye Hospital. All individuals were examined by an ophthalmologist and completed a health and lifestyle questionnaire to gather information on family history, smoking and other medical history. Clinical examination included colour stereoscopic fundus photography of the macular region and photographs were graded according to the International Classification System for age-related maculopathy [94] (ICARMS) classification by *2* ophthalmologists independently at two stages (preliminary and final), with any discrepancies being resolved by consensus. Cases or controls were excluded if they had evidence of diabetic retinopathy, inflammatory or retinovascular disease. Blood was drawn at the time of examination from an ante cubital vein into EDTA, and samples stored at -20º C for DNA extraction. The rs1061170 polymorphism was typed blind to the clinical status of the participants using Taqman (ABI),which was carried out commercially.

2.1.1.2 EUREYE Study

rs1061170 was genotyped in a nested case control subset of the EUREYE study, a cross sectional study investigating the prevalence and risk factors for AMD across Europe. The

study randomly sampled individuals ≥ 65 years of age who were invited to an eye examination at one of 7 participating centres across Europe (Norway, Estonia, United Kingdom, France, Italy, Greece and Spain). [141]. Fundal images were taken of each eye and were graded masked to clinical status, according to ICARMS classification at a single reading centre. Drusen were categorised according to their size, homogeneity of surface features and outlines, pigmentory irregularities were classified into either hypopigmentation or hyperpigmentation. Early AMD was defined as the presence of soft indistinct drusen (≥ 125 μm) or reticular drusen only or soft distinct drusen (≥ 63 μm) with pigmentary abnormalities or soft indistinct drusen (≥ 125 μm) or reticular drusen with pigmentary abnormalities. Late AMD subtypes were geographic atrophy (GA) or choroidal neovascularisation (CNV). GA was defined as any sharply demarcated round or oval areal of apparent absence of the RPE, larger than 175μm, with visible choroidal vessels, and no CNV. CNV was defined as the presence of a serous or hemorrhagic detachment of the RPE and/or a sub retinal neovascular membrane and/or sub retinal hemorrhage, and/or peri-retinal fibrous scarring, even with patches of GA. When geographic atrophy and choroidal neovascularisation co-existed in the same eye this was categorised as CNV. When in doubt, eyes with other disorders resembling AMD were excluded. Interviews were also conducted with individuals in order to ascertain lifestyle risk factors including smoking history. Blood was drawn at the time of interview into EDTA and DNA was extracted and stored at -80º C. Genotyping was carried out at KBiosciences using KASPar chemistry, a competitive allele specific polymerase chain reaction (PCR) SNP genotyping system using FRET quencher cassette oligos. (http://www.kbioscience.co.uk/genotyping/genotyping-chemistry.htm).

2.1.2 Meta-analysis

2.1.2.1 Search Strategy

MEDLINE, (using Pubmed up to August 2008) and EMBASE (1981-2008) were searched for studies evaluating the association between rs1061170 polymorphism and AMD. Free text and MeSH terms used were ‘age related macular degeneration’, ‘macular degeneration’, ‘complement factor H’ and ‘Y402H’. Searches were limited to “human”. Additional studies were

identified through reference lists of publications and searching through 'related articles' link on Pubmed.

2.1.2.2 Inclusion criteria

To be included, studies had to be cohort or case control (nested/ prospective/ retrospective) studies studying unrelated individuals. Where studies contained duplicated results; the study with the smaller population was excluded. If there were clearly two populations (e.g. discovery and replication populations), both these were included, and notation used for these were as used in the primary study. Where studies contained data sets from related and un-related cases, information extracted was limited to unrelated dataset.

2.1.2.3 Data Collection and Management

Data were extracted on AMD outcome from studies which reported any subtype of AMD, these were categorised either as early AMD or late AMD, and if reported separately data was collected on the discrete sub-categories of ARM, GA, CNV, or mixed GA and CNV. Study quality was assessed according to HuGENET guidelines for genetic association studies (http://www.hugenet.ca). The following information was abstracted from published reports: grading scales used for AMD diagnosis, clinical categories of AMD (early AMD, GA, CNV, late AMD), if genotyping was carried out blinded to outcome, genotyping platform used, deviation from Hardy Weinberg Equilibrium (HWE) and ancestry of population under study. Corresponding authors from the studies that met inclusion criteria and which included a total of ≥500 individuals were contacted on at least three occasions to request further aggregate or participant level data. Information was requested on age, gender, smoking status, rs1061170 genotype, AMD status, and AMD sub-type if affected and grading scale used for diagnosis. Participant level data was reconstructed from limited tabular data where necessary.

2.1.2.4 Study Quality

The Venice criteria were applied to the studies included in the meta-analysis to assess the strength of the cumulative evidence on genetic associations. These criteria are a semi-quantitative index which assigns three levels to the association under study [142]. The levels

assigned include first the amount of evidence, second the extent of replication, and third the degree of protection from bias. Within each category, three scores from A to C were assigned, with A signifying the highest level indicating the strength of data in each category, and C the lowest in each category, indicating paucity of data or poor data quality in each category. A composite score of assessment for an association is generated which can then be graded as strong, moderate or weak.

2.1.3 Statistical analysis

2.1.3.1 Newly genotyped studies

To evaluate differences between cases and controls, unpaired Student's t or $\chi 2$ tests were used as appropriate. Departure from Hardy-Weinberg equilibrium (HWE) was evaluated using a $\chi 2$ analysis in controls. The principal *a priori* hypothesis was that the association between rs1061170 and late AMD follows an additive model according to the number of C alleles, however recessive and dominant models and pair-wise comparisons (that is CC homozygotes *vs* TT homozygotes and CT heterozygotes *vs* TT homozygotes) were also evaluated. For the additive model the OR was compared using logistic regression between cases and controls by assigning scores (0,1,2) for different genotype groups and calculating the OR's and 95% confidence intervals (CI's). For the EUREYE data set, standard errors were adjusted to account for data clustered by country. Concomitant subsidiary analyses assessed the association rs1061170 in the different subtypes of AMD. The joint contribution of rs1061170 genotype and smoking on the risk of AMD was assessed to evaluate any gene- smoking interaction. Smoking was coded as ever and never or as current, ex-smoker and never smoked. Interaction was assessed overall on the multiplicative scale and separately by strata of age and gender. All data were analysed using Stata (Version 11, StataCorp LP College Station, TX, USA)

2.1.3.2 Meta analysis

2.1.3.2.1 Participant level meta-analysis

Nine studies provided either participant level or limited tabular information and these data were analysed as follows:

2.1.3.2.2 Main effect of rs1061170 on late AMD

As for the case control and cross sectional study, the model specified *a priori* was an additive (per-allele) model for late AMD. Secondary analyses explored other genetic models of inheritance. Mixed logistic models treating 'study' as the random factor were used. Interaction by age, gender was assessed on the multiplicative scale using studies able to provide these data.

2.1.3.2.3 AMD subtype

The effect of rs1061170 genotype on different clinical grades of AMD was assessed in using mixed models as above with study as the random factor and using 99% confidence intervals.

2.1.3.2.4 Smoking

A number of approaches were used to explore effect modification by smoking on the effect of rs1061170 genotype on any AMD, late AMD (GA and CNV together) and subtypes of AMD (GA and CNV separately). The first approach was to use all participant level data merged with limited tabular data (9 studies 5,804 cases) to assess the effect of smoking on the rs1061170 effect on all AMD; second using only participant level data (5 studies) allowed me to assess the effect of smoking on rs1061170 genotype effect on all late AMD (CNV and GA together) as well as individual subtypes of late AMD (GA and CNV separately). Thirdly a case only design [143] was used in a subset of studies providing genotype and smoking data to further assess an effect modification with smoking. This analysis is based on the premise that in the absence of any effect modification, smoking distribution among cases will not be related to genotype, provided smoking is independent of *CFH* genotype in the general population. Advantages of the case-only approach are that it excludes potential biases arising from control selection and overcomes problems arising from combining different study designs while maintaining statistical power. Lastly using a further subset of data (3 studies) where pack years of smoking could be determined, association between intensity of smoking and AMD risk stratified by rs1061170 genotype was also calculated. Given the exploratory nature of the analyses 99% confidence intervals were used.

2.1.3.2.5 Aggregate level meta-analysis

Aggregate information was available from 16 studies and these were included in the analysis of the main effect of rs1061170 on AMD and investigation of heterogeneity of effect size based on pre-specified subgroups.

2.1.3.2.6 Biological and technical variables

Heterogeneity can arise because of errors and biases as well as true biological variation. Therefore the effect of study design, genotyping platform, and grading scale for the diagnosis of late AMD (as sources of error and bias) and the effect of smoking and disease subtype (as potential explanations for true biological variation) were evaluated. Using aggregate level meta-analysis summary measures for all variables were calculated using a random effects model with inverse variance weights. Heterogeneity between estimates was assessed using the Der Simonian and Laird Q test and I^2 was used as a measure to describe the percentage of variability in point estimates.

2.1.3.2.7 Non European Populations

Studies investigating the association of rs1061170 in non European populations (mainly Chinese and Japanese) were analysed separately by collecting aggregate level data from published studies. Summary estimates were generated using random effects models as above. Where there were “zero” cells, that is there were no cases for a given genotype, a value of 0.5 was added to all cells and meta-analysis carried out [144]. If case/control counts by genotype were not readily extractable from published studies, but allele frequencies were given, genotype frequencies were calculated. If no data was given from which genotype data could be calculated, summary odds ratios were extracted and included in the meta-analysis. All data analysis was carried out using Stata (Version 11, StataCorp LP College Station, TX, USA) all studies included reported findings of late AMD cases.

2.1.3.2.8 Small study bias and cumulative meta-analysis

Small study bias was assessed using funnel plots and the Egger test, and cumulative meta-analysis was performed to test the cumulative evidence at the time each study was published, to examine attrition in the effect size after initial studies reporting a large effect size.

2.1.3.2.9 Data extraction and database reconstruction

Submitted databases were cleaned and checked and any uncertainties resolved by contact with the authors. Aggregate level data from studies investigating European individuals was converted to “pseudo” participant level data by expanding the aggregate database. A unique “studyid” variable was created and a unique number was assigned to individuals in the aggregate database. Similarly the number of cases and their genotypes were also assigned in the same manner. Details of other covariates: age, gender and smoking were added to the baseline information of genotype and outcome although in separate databases, as not enough information was available to merge all covariates into one database. The individual databases were merged with participant level databases thus creating four databases: one for the main effect of late AMD, one for age, one for gender and one database of pooled studies where all fields at participant level were available. The latter was used to explore interaction between covariates, as this was not possible with databases limited to one covariate

2.2 Results

2.2.1 Newly genotyped studies

Clinical, demographic data and distribution of rs1061170 genotype from the MRC case control study and EUREYE are shown in **Table 2.1**. Allele frequencies in cases and controls were comparable to other published studies, with an increased frequency of the C allele in cases. There was no deviation from Hardy Weinberg Equilibrium in controls. An additive age adjusted model (per *C* allele) demonstrated a significant association between rs1061170 SNP and late AMD risk in both MRC case control study (OR 2.04 (95% CI 1.63, 2.56 p<0.001)) and EUREYE studies (OR 2.43 (95% CI 2.04, 2.90 p<0.001)) (**Table 2.2**). The effect was found to be similar in analysis for each clinical subtype of AMD and, in ever-smokers compared to

never smokers (**Table 2.2**).There was no evidence for a gene-smoking interaction (p=0.42 for the MRC case control study and p=0.78 for the EUREYE study).

Table 2.1 Demographic characteristics of AMD cases and controls with differing rs1061170 genotype from the MRC case-control study and the EUREYE study (* indicates ever smoked, GA=Geographic Atrophy, CNV=Choroidal neovascularisation)

Characteristic	**MRC Case Control Study**		**EUREYE**	
	AMD	**Controls**	**AMD**	**Controls**
Total number	1,234	235	686	604
Age (mean, sd)	77.36 (8.21)	74.88 (8.11)	75.4 (6.43)	74.6 (6.02)
% males	35.17	41.28	45.9	45.6
% Smokers*	62.80	59.57	50.4	44.4
TT genotype	201	71	206	251
CT genotype	591	105	333	285
CC genotype	397	38	147	68
Deviation from HWE for controls (p value for χ^2 test)		0.94		0.33
AMD Grade				
% early	18.56		79.5	
% GA	17.34		6.3	
% CNV	60.00		14.3	

Table 2.2 Effect of rs1061170 polymorphism on AMD risk in MRC case control and EUREYE studies

	Additive OR (95% CI), unadjusted	Additive models, age adjusted OR (95% CI)	Genetic Models, age adjusted TC vs TT OR (95% CI)	Genetic Models, age adjusted CC vs TT OR (95% CI)
		MRC Case Control Study		
Main effect: **Late AMD**	1.90 (1.53, 2.36)	2.04 (1.63, 2.56)	2.24(1.56, 3.21)	4.08 (2.61, 6.39)
ARM/ Early AMD	2.07 (1.56, 2.72)	2.07 (1.56, 2.73)	1.96 (1.20, 3.20)	4.26 (2.43, 7.45)
GA	2.12 (1.60, 2.83)	2.23 (1.66, 3.00)	2.27 (1.36, 3.82)	4.99 (2.77, 9.00)
CNV	1.77 (1.41, 2.22)	1.90 (1.51, 2.41)	2.10 (1.44, 3.06)	3.57 (2.24, 5.70)
Ever Smoked	2.06(1.55, 2.74)	2.25 (1.67, 3.03)	1.67 (1.07, 2.61)	4.58 (2.66, 8.51)
Never Smoked	1.71 (1.22, 2.41)	1.81 (1.28, 2.56)	1.59 (0.89, 2.87)	3.31 (1.63, 6.69)
		EUREYE*		
Main effect: **Late AMD**	2.30 (1.93,2.73)	2.43 (2.04,2.90)	2.59 (2.02,3.32)	5.90 (4.13,8.44)
ARM/ Early AMD	1.43 (1.21, 1.70)	1.44 (1.21, 1.71)	1.27 (1.09, 1.48)	2.22 (1.47, 3.37)
GA	2.41 (1.39, 4.19)	2.74 (1.60, 4.69)	1.73 (0.82, 3.65)	7.16 (2.93, 17.51)
CNV	2.26 (1.64, 3.11)	2.34 (1.61, 3.40)	3.12 (1.69, 5.74)	5.47 (2.42, 12.36)
Ever Smoked	1.99 (1.32,2.99)	2.14 (1.42, 3.21)	2.39 (1.26, 4.54)	4.74 (1.95, 11.52)
Never smoked	2.72 (1.81, 4.08)	2.90 (1.90, 4.43)	3.17 (2.14, 4.71)	8.15 (3.65, 18.18)

* Clustering of data by country has been accounted for in analysis

2.2.2 Meta-analysis

The search identified 40 studies from which relevant information could be abstracted, 24 of which were in European populations [10, 104, 105, 107-110, 145-158], six were in Japanese populations [159-164], six in Chinese [165-170] (including Hong Kong Chinese and Taiwanese), one Korean [171], one in Indian/ South Asians [172], one in Latin Americans

[173] and one in Black South Africans [174] **(Figure 2.1, Table 2.3).** The main focus of the analysis was on the relationship of the rs1061170 SNP and late AMD in individuals of European ancestry, although secondary analyses on association in other ethnic groups were also performed. Nine studies, comprising 5,185 cases of AMD identified from the search, or by direct contact, were able to provide more detailed information, either aggregate or individual level, in addition to the newly genotyped studies [107-109, 154, 158].

Figure 2.1 Flow diagram illustrating identification of studies included in the meta-analysis of the rs1061170 variant in *CFH* and AMD.

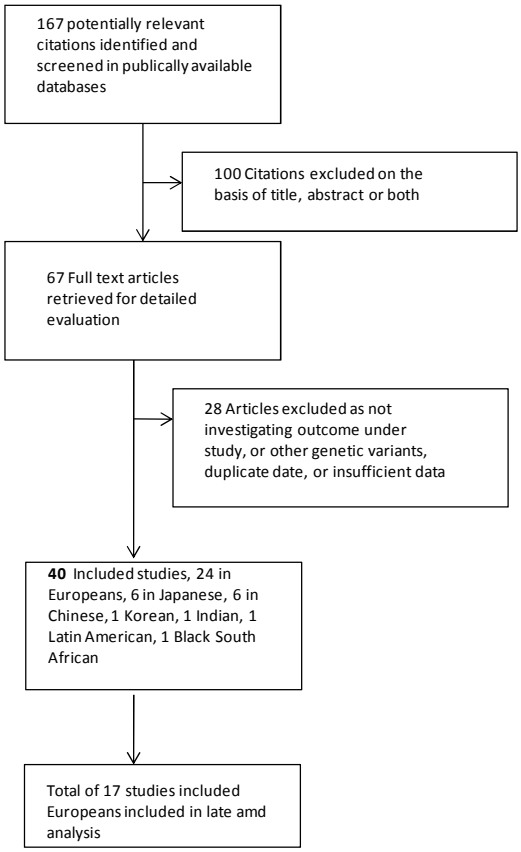


Table 2.3 Studies included in the meta-analysis of the association of *CFH* rs1061170 variant and AMD risk. The table is divided by the type of data that were gathered and according to the ancestry of the participants. The first section details characteristics of studies where information was abstracted from the publication, which studied individuals of European ancestry. Studies indicated with a *, are those for which the outcome of late AMD alone was not reported, rather all AMD (including early disease) reported as one- and were therefore NOT included in the subsequent meta-analysis of late AMD. The second section details the studies that were contacted, and who were able to provide either limited tabular data or participant level data. These studies are also of individuals of European ancestry. The final part of the table includes information of data abstracted from published literature in non-European ethnicities. HWE is not given as all studies report control population in HWE.

Published studies with participants of European Ancestry

Study	Ethnicity	Study Design	Grading Scales used for AMD	Source of population: Cases	Source of population: controls	Genotyping platform
Edwards AO et al., Discovery Set*	European	Case control	Wisconsin	Clinic Based	Clinic/ Spouse controls	TaqMan (ABI); MassARRAY (Sequenom)-analysed by MALDITOF
Edwards AO et al., Replication set*	European	Case control	Wisconsin	Clinic Based	Population	TaqMan (ABI); MassARRAY (Sequenom)-analysed by MALDITOF
Hageman G., et al Columbia population*	European	Case control	ICARMS	Clinic Based	Clinic	TaqMan
Hageman G., et al Iowa population*	European	Case control	ICARMS	Clinic Based	Clinic	TaqMan
Souied EH et al	European	Case control	ICARMS	Clinic Based	Clinic	PCR directed sequencing
Physicians Health Study*	European	Nested Case control	Not specified	Nested in RCT	Nested in RCT	TaqMan
Simonelli F et al	European	Case control	ICARMS	Clinic Based	Clinic	Taqman
Seitsonen S et al*	European	Case control	Not specified	Clinic Based	Clinic and anonymous blood donors	PCR directed sequencing
Baird et al	European	Case control	Wisconsin	Clinic Based	Population	MassArray
Fisher et al*	European	Case control	ICARMS	Clinic Based	Spouse	PCR Directed Sequencing
Nurses Health/ Health Professionals study	European	Nested case control	ICARMS	Population	Population	Taqman

Table 2.3 continued

Published studies with participants of European Ancestry

Study	Ethnicity	Study Design	Grading Scales used for AMD	Source of population: Cases	Source of population: controls	Genotyping platform
Weger et al.	European	Case control	AREDS	Clinic Based	Clinic	RFLP
Pulido et al.	European	Case control	Not specified	Clinic Based	Disease controls (cystic fibrosis)	RFLP
Droz et al.	European	Case control	ICARMS	Clinic Based	Clinic and Spouse	DHPLC
Blue Mountain Eye Study	European	Prospective cohort	Wisconsin	Population	Population	TaqMan
Rotterdam	European	Prospective cohort	Rotterdam	Population	Population	TaqMan

Table 2.3 continued. Limited tabular and participant level data from studies of individuals of European ancestry.

Limited Tabular data, European Ancestry

Study	Ethnicity	Study Design	Grading Scales used for AMD	Source of population: Cases	Source of population: controls	Genotyping platform
Rivera et al	European	Case control	ICARMS	Clinic Based	Clinic	MassArray
Conley et al	European	Case control	In House	Clinic Based	Clinic	RFLP
CHS	European	Nested Case control	In House	Population	Population	RFLP
AREDS	European	Nested Case control	AREDS	Nested in RCT	Nested in RCT	RFLP

Participant level data, European Ancestry

Study	Ethnicity	Study Design	Grading Scales used for AMD	Source of population: Cases	Source of population: controls	Genotyping platform
MRC Case Control Study	European	Case control	ICARMS	Clinic Based	Clinic	Taqman
Zhang et al.	European	Case control	ICARMS	Clinic Based	Clinic	Direct DNA Sequencing
Zareparsi et al.	European	Case control	ICARMS	Clinic Based	Clinic	Not specified
Sepp T et al.	European	Case control	ICARMS	Clinic Based	Clinic	SNaPshot
Southampton	European	Case control	AREDS	Clinic Based	Clinic	GenomeLab SNPstream
EUREYE	European	Nested case control in cross sectional study	ICARMS	Population based	Population Based	KASPar
Amsterdam	European	Population based	Rotterdam	Population based	Population Based	Taqman

Table 2.3 continued. Aggregate level data abstracted from publications of non-European populations.

Study	Ethnicity	Study Design	Grading Scales used for AMD	Source of population: Cases	Source of population: controls	Genotyping platform
Japanese Studies						
Okomoto et al.	Japanese	Case control	Not specified	Clinic Based	Clinic Based	Direct sequencing
Gotoh et al.	Japanese	Case control	Not specified	Clinic Based	Clinic Based	Direct sequencing
Uka et al.	Japanese	Case control	Not specified	Clinic Based	Clinic Based	Taqman
Fuse et al.	Japanese	Case control	Not specified	Clinic Based	Clinic Based	PCR amplification restriction digest
Tanimoto et al.	Japanese	Case control	Not specified	Clinic Based	Clinic Based	Taqman
Mori et al.	Japanese	Case control	AREDS	Clinic Based	Clinic Based	Taqman
Chinese Studies						
Lau et al.	Chinese	Case control	ICARMS	Clinic Based	Clinic Based	PCR amplification restriction digest
Chen et al.	Chinese	Case control	ICARMS	Clinic Based	Clinic Based	Taqman
Lin et al (wet AMD)	Chinese	Case control	ICARMS	Clinic Based	Clinic Based	Melting curve mutation analysis
Ng et al.	Chinese	Case control	ICARMS	Clinic Based	Clinic Based	Direct sequencing
Xu et al.	Chinese	Case control	Not specified	Clinic Based	Clinic Based	PCR amplification restriction digest
Lin et al (early AMD)	Chinese	Case control		Clinic Based	Clinic Based	Melting curve mutation analysis

Table 2.3 continued. Aggregate level data abstracted from publications of non-European populations.

Study	Ethnicity	Study Design	Grading Scales used for AMD	Source of population: Cases	Source of population: controls	Genotyping platform
Korean Study						
Kim et al	Korean	Case control	Not specified	Clinic Based	Clinic Based	Direct Sequencing
Indian/ South Asian study						
Kaur et al.	Indian/ South Asian	Case control	AREDS	Clinic Based	Clinic Based	PCR amplification restriction digest
Latin American/ US based study						
Tedeschi-Blok et al.	Latin American	case control	Wisconsin	Clinic Based	Clinic Based	PCR amplification restriction digest
South African Study						
Ziskind et al.	South African	case control	Rotterdam	Clinic Based	Clinic Based	RFLP

2.2.2.1 Main effect of rs1061170 polymorphism on late AMD

Of a total of 26 studies, 18 reported information on late AMD and included 6,322 cases and 10,467 controls. The per-C allele OR for late AMD was 2.27 (95% CI 2.10, 2.45) (**Figure 2.2**) and for any AMD the OR was 1.86 (95% CI 1.77, 1.97, for 26,494 individuals including 14, 174 cases). These estimates are more precise than the previously published meta-analysis which included 4,856 cases from 8 studies [140]. Genotyping platform (χ^2 p =0.06 on 6 degrees of freedom, I^2=49.6%), clinical grading scale (χ^2 p =0.33 on 5 degrees of freedom, I^2=13.3%), and study design (χ^2p =0.0.80 on 2 degrees of freedom, I^2=0.0%) contributed to some heterogeneity in effect size (**Figure 2.3**)

Figure 2.2 Overall association of the rs1061170 SNP on the risk of late AMD and the effects of this polymorphism on clinical subtypes of AMD. For the outcome of "late AMD" GA, CNV and mixed GA and CNV were combined into a single category.

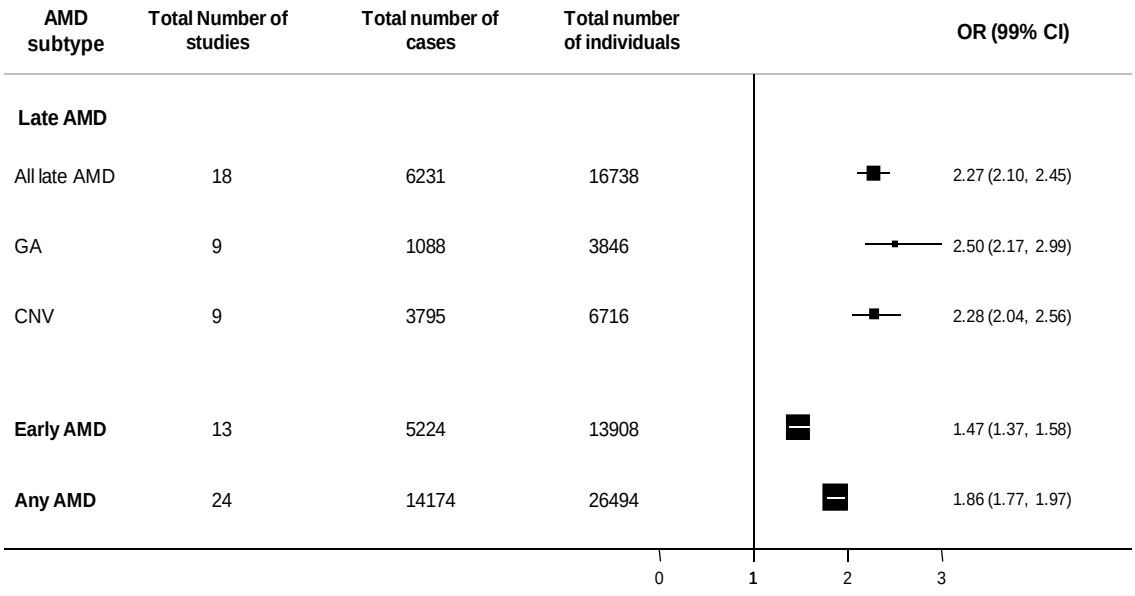
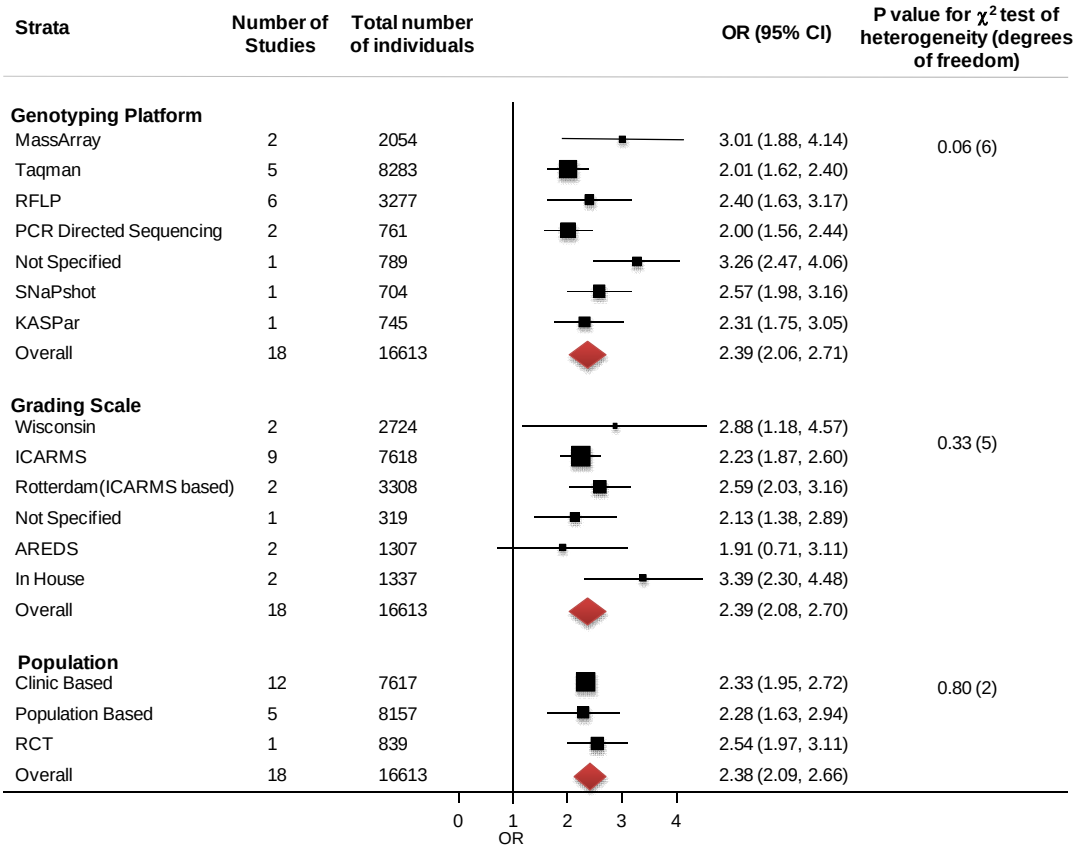


Figure 2.3 Association of rs1061170 variant on late AMD stratified by diagnostic grading scale, population type and genotyping method



Where studies genotyped multiple variants in the *CFH* gene, and did not specify genotyping platform used for rs1061170 encoding Y402H they were coded as "not specified". For AREDS, limited tabular data was available by request as appears in Conley et al.[107] , which stated genotyping was carried out using RFLP rather than Affymetrix 100K Chip as per original genome wide scan.

ICARMS: International classification and grading system for age-related macular degeneration [94]; AREDS: Age-Related Eye Disease Study System for classifying age-related macular degeneration [143], Wisconsin age related maculopathy grading system[175], Rotterdam age-related macular degeneration scale[176]

2.2.2.2 Effect of rs1061170 on grade of AMD, assessed using participant level data

The OR per-*C* allele for the association with early AMD was 1.47 (99% CI 1.37, 1.58; information from 13 studies, 5,224 cases); for GA was 2.50 (99% CI 2.17, 2.99; 9 studies, 1,088 cases), and for CNV was 2.28 (99% CI 2.04, 2.56; 9 studies, 3,795 cases) (**Figure 2.2**).

2.2.2.3 Effect modification of the rs1061170- AMD association by smoking

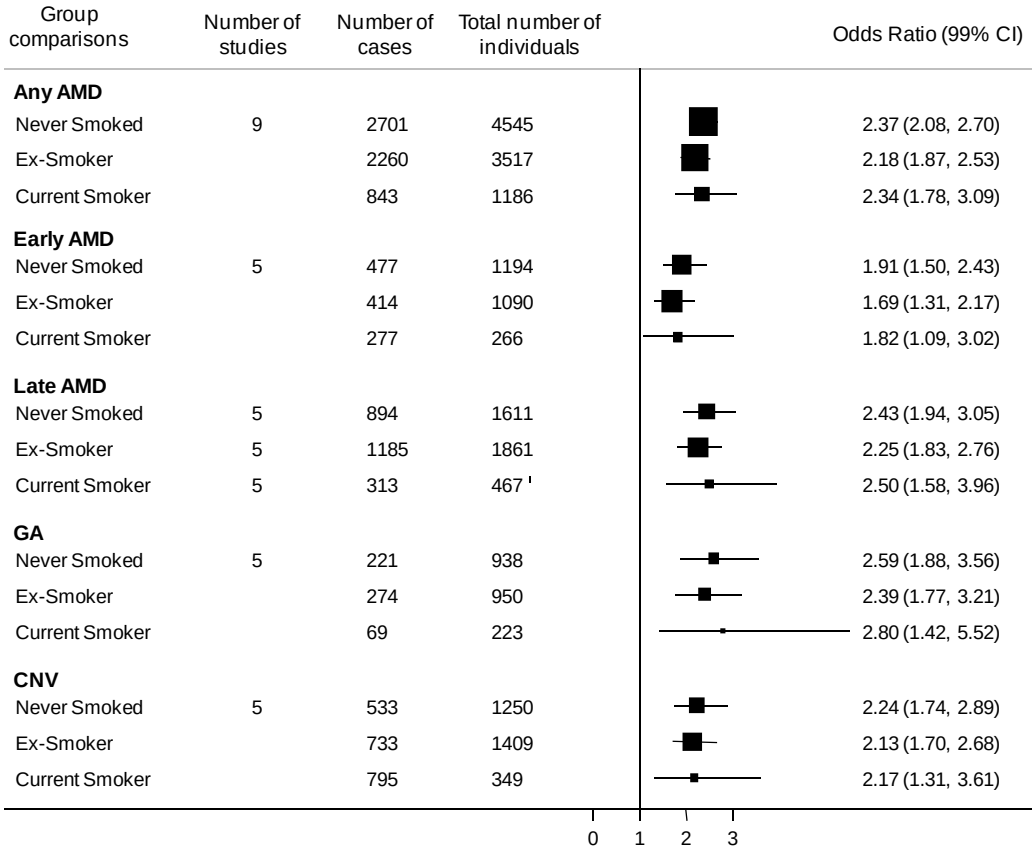
First I assessed the effect of smoking on AMD risk without taking genotype into account. Using participant level data from five studies (2,403 cases of late AMD) the odds ratio for late AMD in

those who had smoked at any time in their life (ever smoked) compared to those who had never smoked was 1.12 (99% CI 0.88, 1.43) .

The smoking effect stratified by rs1061170 genotype was such that the odds for late AMD in smokers versus non-smokers in individuals who were homozygous for the T allele was 1.66 (99% CI 1.00, 2.75), for TC heterozygotes the odds for late AMD was 0.82 (99% CI 0.56, 1.17) and for CC individuals the odds for late AMD in smokers versus never smoked is 1.67 (99% CI 0.90, 3.08).

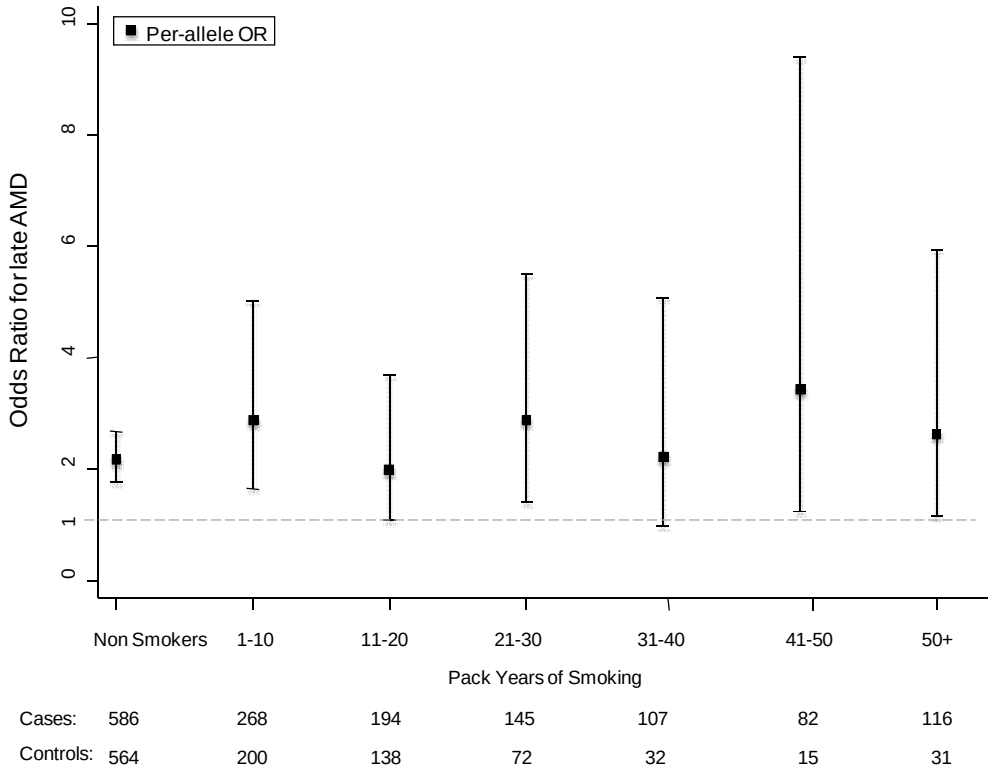
Adding genotype and using an additive (per allele model), the odds ratio for late AMD per C allele in current *versus* never smoked was 2.50 (99% CI 1.58, 3.96) and ex-smokers *versus* never smoker 2.25 (99% CI 1.83, 2.76), with no evidence for interaction using either categorisation (p value for interaction 0.63 and p=0.75 respectively). The findings were similar when using merged participant level and tabular data from 11 studies (6,422cases of any AMD) and are summarised in **Figure 2.4**.

Figure 2.4 Association of rs1061170 genotype on AMD stratified by smoking The boxes in the plot below represent the additive (per-allele) estimate for rs1061170 genotype on AMD and their 99% confidence intervals. These estimates are stratified by smoking, and the category of smoking is displayed on the left hand side. For the analyses shown, smoking was coded as never and ex-smoker and current smoker. Using the additive model the reference group is that coded '0' in the genotype model, which here are individuals homozygous for the rs1061170 'TT' genotype.



Lastly using the case only approach, limited to late AMD cases (n=2,403) from studies providing participant level data, the odd ratio was 0.81 (99% CI 0.60, 1.10), indicating no evidence for a gene-environment interaction. To test the assumption of independence between the genotype and smoking in the general population the control group (n=1556) was also assessed, and the odds ratio for smoking in this group was 1.04 (99% CI 0.79, 1.35). Further assessment of the effect of intensity of smoking on AMD risk by genotype also did not show any evidence of interaction, for both pair-wise and per allele comparisons (**Figure 2.5).**

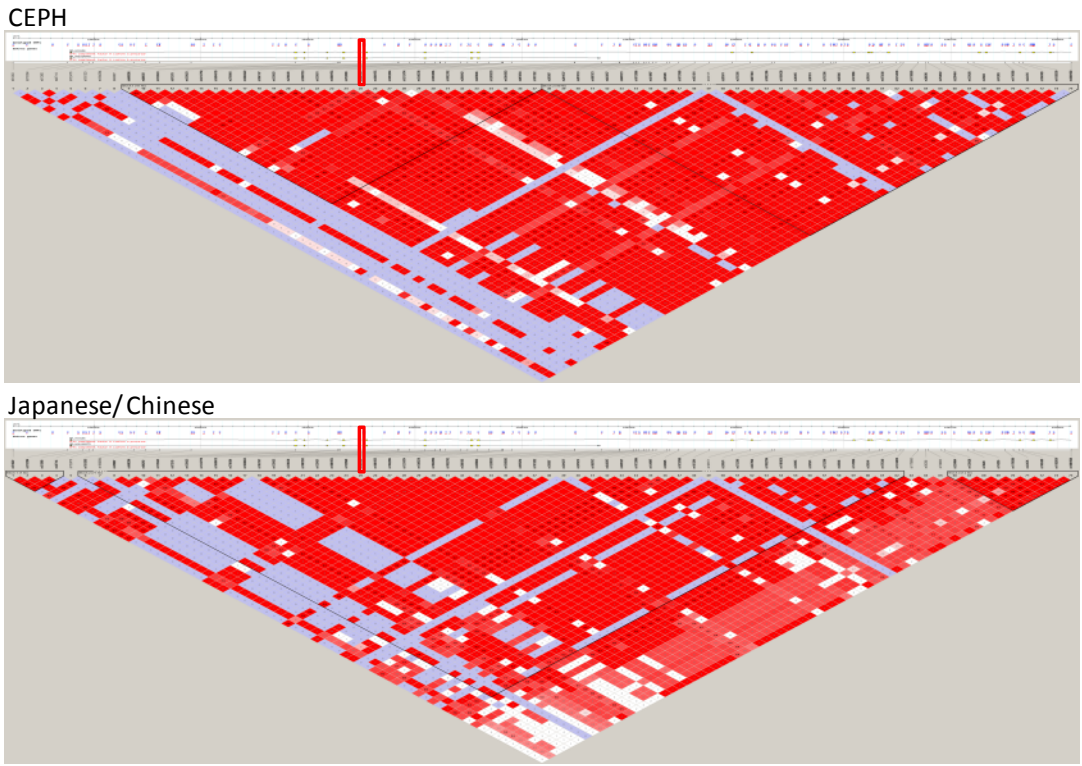
Figure 2.5 Association of rs1061170 and late AMD by stratum of smoking intensity. Additive, per-allele estimates are given above. Data is taken from three studies in which pack years of smoking was recorded.



2.2.2.4 Influence of ethnicity on the association of rs1061170 and late AMD

HapMap (www.hapmap.org) LD patterns for common SNPs in the vicinity of *CFH* differ among ancestral groups, and the frequency of the C allele for rs1061170 is lower in Japanese (MAF=0.06) and Chinese (MAF=0.07), compared to European subjects (MAF=0.36) (**Figure 2.6**).

Figure 2.6 Linkage disequilibrium maps of *CFH* in populations of European ancestry (CEPH) and Japanese and Han Chinese ancestry. LD plots were constructed using Build 22 Hapmap. The rs1061170 SNP is indicated by a red bar in both LD panels to serve as a marker for comparison of the surrounding LD patterns. LD metric displayed is r^2, darker red indicating higher degree of LD.



A systematic review and meta-analysis of published studies on the association of rs1061170 and AMD risk in non-European populations was therefore performed. The search identified 16 studies carried out in non-European populations, with six being in Japanese, six being in Chinese, one in south Asians, one in Black South Africans, and one in Latin Americans. Data was reported in a form amenable to meta-analysis from 14 out of 15 studies, with genotype counts being calculated from allele frequencies in one study [160] and summary odds ratio being used from the Indian/ South Asian study [172] . In ten studies there were no cases or controls with the putative risk allele, and an estimate for odds could only be obtained by adding 0.5 to each cell [144]. Meta-analysis was carried out using data from 2 studies (206 cases) with full data and showed no evidence of an effect of rs1061170 on AMD risk in

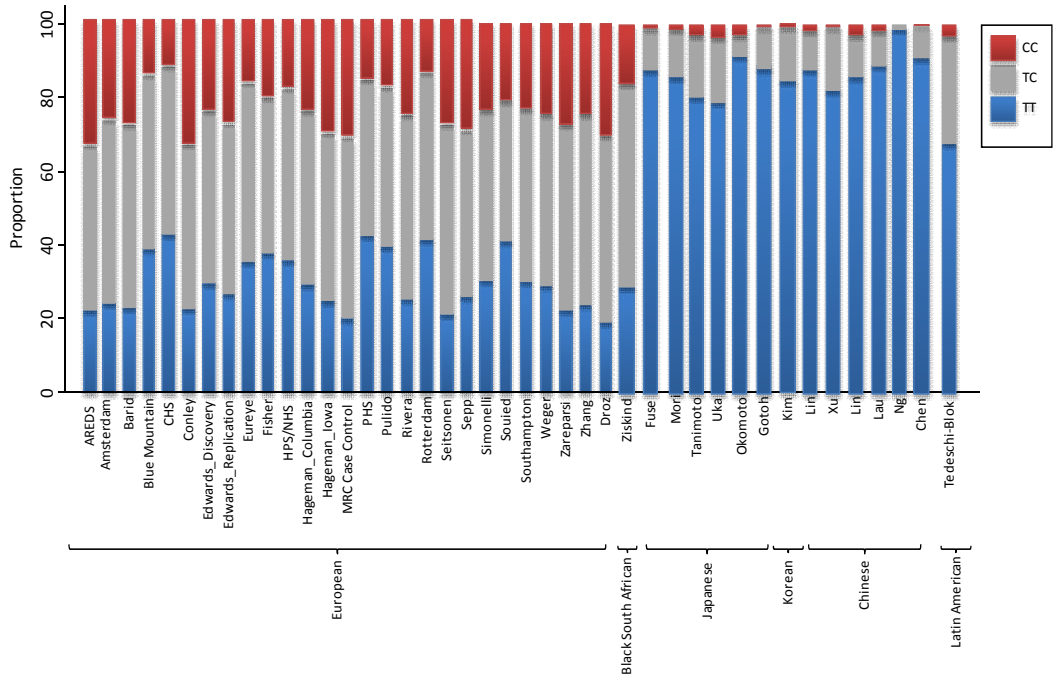
individuals of Japanese ancestry, (OR 1.72 95% CI 0.51, 5.79 for CC compared to TT; and 1.49 95% CI 0.86, 2.58 in CT vs TT). Estimates from analysis using the complete subset of the data from Japanese and Chinese studies adding a value of 0.5 to cells with zero counts (6 Japanese studies,672 cases) demonstrated an odds of 1.71 (95%CI 0.65,4.49) for CC homozygotes compared to TT individuals, and for individuals of Chinese ancestry (6 studies, 839 cases) the odds of disease in CC individuals compared to TT individuals was 5.34 (95% CI 0.81, 23.12), Risk in heterozygous subjects and data from other ancestries is given in **Table 2.4**. The observed differences in disease risk may in part be due to the differences in allele frequencies in the different populations (**Figure 2.7).**

Table 2.4 Results of meta-analysis from non-European populations, on the effect of rs1061170 on AMD. All summary statistics are random effects estimates, 0.5 has been added to cells with zero counts to allow pooling of the individual studies. For populations represented by a single study, summary measures were not included.

Ethnicity	No of studies	Summary OR for CC vs TT (95% CI)	Summary OR for TC vs TT (95%CI)
Japanese	6	1.79 (0.69, 4.66)	1.03 (0.71, 1.50)
Chinese	4	5.34 (1.05, 27.07)	1.73 (1.08, 2.77)
Taiwan* Chinese	2	22.26 (2.85, 173.83)	1.83 (1.08, 3.08)

*Two studies combined, one wet AMD , one early AMD

Figure 2.7 Frequency of rs1061170 genotypes in studies included in the analysis of the main effect on AMD risk.



2.2.2.5 Venice Criteria applied to the association of rs1061170 and AMD in European and non-European populations

Venice criteria were applied to evidence from populations of all ancestries, the largest amount of information coming from genetic association studies in those of white European ancestry. The findings are summarised in **Table 2.5.** The evidence for the association was greatest in European subjects (graded AAA) and least in Indian, Latin American and Black African subjects (CCC). The strength of the evidence from Japanese studies (BCB) and Chinese subjects (BCC) was intermediate.

Table 2.5 Strength of evidence for the association of rs1061170 genotype with AMD risk in individuals of differing ancestry assessed using Venice Criteria. A score of A indicates strong epidemiological credibility, B indicates moderate and C indicates weak, therefore a score of AAA in all categories indicates a large amount of evidence with extensive replication and protection from bias. Studies amongst European descent populations present strong epidemiological credibility although the relative paucity of data amongst other ancestries makes this association less so, although this may reflect inadequate power to detect an effect given the MAF particularly in Japanese and Chinese populations.

Population	Amount of Evidence	Replication	Protection from Bias	Venice Score
White European	A	A	A	AAA
Japanese	B	C	B	BCB
Chinese	B	C	C	BCC
Indian (South Asian)	C	C	C	CCC
Latin American	C	C	C	CCC
Black African	C	C	C	CCC

2.3 Discussion

2.3.1.1 Summary of main findings

By conducting a meta-analysis using a combination of new studies and published data, what I have been able to present here is a more precise and reliable estimate of late AMD conferred by carriage of the rs1061170 variant in the *CFH* gene. This analysis indicates that the association has remained robust, with the best estimate of the effect size being largely undiminished despite the continuing accrual of information on this association since the original GWAS. Of the 813 GWAS of common diseases and disease related risk factors conducted to date (www.genome.gov/26525384), the association of *CFH* (in particular the rs1061170 variant) with AMD, is amongst the largest for a common allele conferring susceptibility for a complex disease with a per-allele odds ratio of 2.27 (95% CI 2.10, 2.45) in European populations.

Exploiting the large size of the assembled data set, I also found no evidence for effect modification in Europeans by smoking habit, which has been reported previously in other, smaller (and therefore lower power) studies. However, there was evidence of a modest difference of genetic effect by disease category, with a slightly higher risk of GA compared to other forms of AMD among individuals carrying risk alleles for rs1061170. However given the overlap of the confidence intervals for the association with GA and CNV, it remains somewhat inconclusive whether the variant confers greater risk of GA compared to other disease sub-types. Importantly, the association of rs1061170 was not replicated in individuals of non-European (mainly Chinese and Japanese) ancestry, though published information from all non-European populations is limited and precludes firm conclusions in these groups.

These findings from a meta-analysis of association studies that followed the initial discovery GWAS are of specific relevance to AMD, but the approach to collation of information are likely to be informative for other complex disorders where GWAS have provided new leads on causative genes that are subsequently assessed by an additional studies in independent data sets.

2.3.1.2 Implication of main effect estimate for predictive genetic testing

Irrespective of whether it is directly causal or not, the strong association of rs1061170 and AMD risk, and the high frequency of the variant in the population has led to an interest in its use as a potential screening or predictive tool. This variant has been included, for example, in a genetic risk profile offered by DeCode genetics (www.decodeme.com) and 23 and me (www.23andme.com), two commercial providers of genetic screening tests for complex diseases. However, initial effect estimates in genetic association studies, whether candidate based or GWAS, can be prone to the winner's curse [139] – inflation of the reported effect estimate over its true value. This can be a particular problem where the initial discovery study is small (and low in power) and declaration of association is based on its crossing a pre-specified p-value threshold. The initial GWAS in AMD was based on a collection of only 96 cases and 50 controls [10] which would be considered small for a GWAS of common disease by current standards, where a minimum of 2000 cases and 2000 controls are considered necessary. Thus while the reported association of rs1061170 can be considered robust, the initial effect estimate of a per-allele odds ratio of 7.4 (95% CI 2.9.19: for homozygotes) is likely to exceed the true effect. In our analysis of 6,231 late AMD cases including the initial discovery study, there is some attenuation of the effect size as data on this association has expanded (**Figure 2.8**), and the summary estimate identified by the meta-analysis presented here is likely to be closer to the true effect.

A smaller effect estimate in prospective compared to case-control studies was also noted. Case-control studies which are the preferred design for gene discovery and tend to be based on highly-selected samples chosen to maximise the differences between cases and controls by choosing patients with younger onset or more advanced disease, and excluding controls whose fundoscopic examinations are not entirely normal.

Genetic association studies that follow a GWAS tend to focus on a subset of SNPs and often only the lead SNP(s) with the most extreme p-value that may not itself be causal. The literature may also become prone to publication bias as positive replications may be more likely to be published than negative ones. However, this particular association does not seem

to be substantially influenced by publication bias as shown by the symmetry of the funnel plot (**Figure 2.9**) though a test for small study bias (Egger test p<0.001), was significant.

Figure 2.8 Cumulative Meta-analysis of the rs1061170-AMD association in studies of European individuals. Effect estimates are per-allele odds rations for the association with any form of AMD

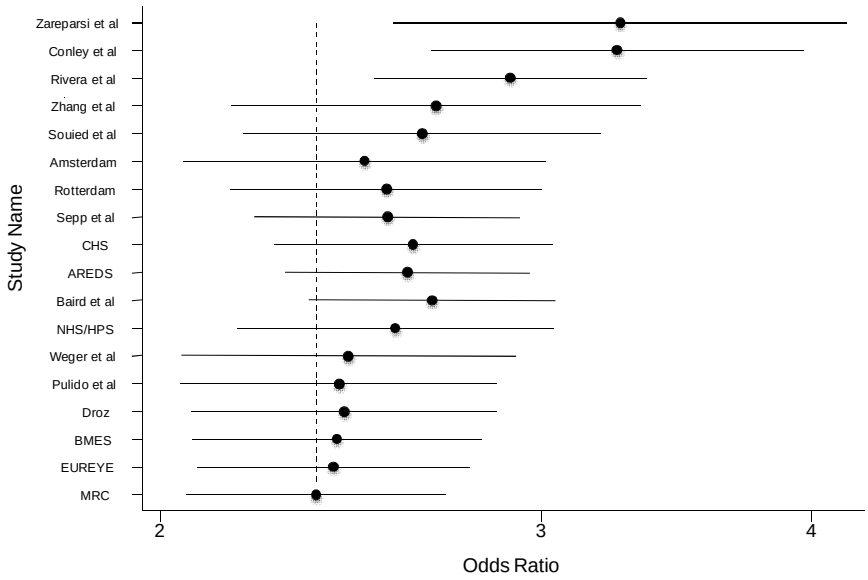
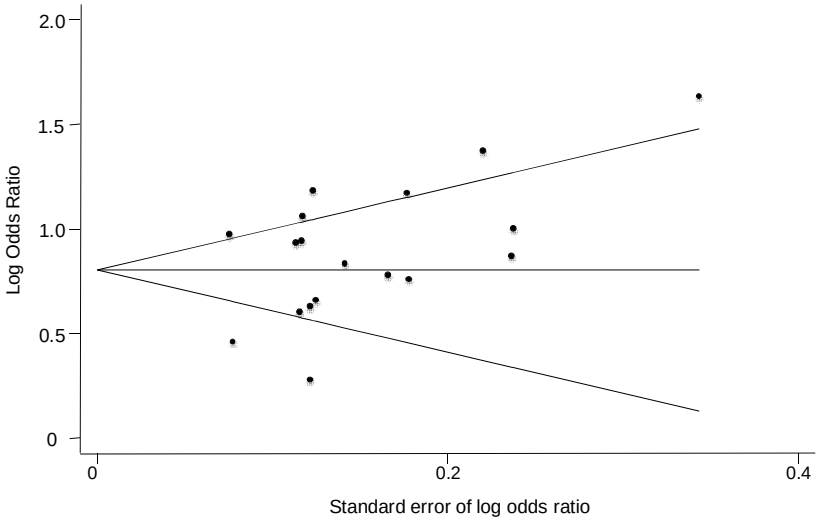


Figure 2.9 Begg's funnel plot with pseudo 95% confidence limits, illustrating the relationship between an index of study size (the SE of the log odds ratio) and the effect estimate for studies in Figure 2.8



2.3.1.3 Heterogeneity of effect estimates

One important attribute of a meta-analysis of genetic association is that it allows exploration of the existence and reasons for variation of effect size across data sets. In the analysis of the main effect of *CFH* genotype on late AMD risk there was no evidence for heterogeneity and explored reasons for this. Potential modifiers of the effect size were further tested with particular focus on investigating the effect of modification by smoking, which has aroused considerable interest. Study design, genotyping and grading scale all contributed to the observed heterogeneity but the influence was not large suggesting that these sources of variation do little to alter the summary effect estimate.

The absence of a smoking interaction in this meta-analysis is at odds with the findings and interpretation from some individual studies. The analysis of effect modification by smoking was conducted in a subset of the database in which participant level or new tabular information on effect size by smoking habit was available through direct contact with the study authors. Despite being limited to part of the dataset, the smoking analysis still incorporated data on 2,392 cases AMD cases of which 1,498 individuals were classified as current or prior smokers. Quantitative data on smoking exposure (in terms of pack years) was also available from 3,279 individuals. In none of these analyses was there any positive evidence for a gene smoking interaction. Further case only analyses indicated an absence of gene-environment interaction, highlighting a consistent lack of evidence of interaction with smoking given the number of methods used to explore its presence. Even with the large size of the dataset, this may still be underpowered to detect a modest interaction, but the analysis presented here suggests that the detection of smoking interaction in individual studies in this field is just as likely to be due to chance as a real finding.

The patho-physiological relationship between the different subtypes of AMD is poorly understood and the mechanisms may be different. For this reason a separate analysis of the association of rs1061170 with the different subtypes of AMD was performed, an analysis that no individual study to date has been sufficiently powered to do. There was some evidence for a larger effect of this variant on risk of GA compared to other forms of the disease. However this finding should be interpreted with caution because it is based on a subgroup analysis

(though attempts were made to reduce the potential for type 1 error by pre-specifying the analysis and using 99% rather than 95% confidence intervals). Furthermore, the apparently more modest effect of rs1061170 on early AMD may simply reflect the fact that prospective observational studies with a surveillance approach to assessment of AMD contributed the majority of cases of this end-point. Although the difference is probably insufficiently large for this information to be useful for disease prediction, it may shed light on mechanistic differences in the disease subtypes that could be enhanced by meta-analyses of the effect of other variants in *CFH* and variants in other risk genes on disease subtype.

2.3.1.4 Rationale for the specific interest in rs1061170

Although, multiple SNPs in and around the *CFH* gene demonstrated association with AMD in the initial discovery studies, including a copy number variant and SNPs in downstream genes [114] apparently independently of rs1061170, it is this SNP that has been most widely studied based on the significance level of the initial association and the fact that it encodes a rs1061170 substitution in the encoded protein, which may alter function. However, evidence on the functionality of the variant is by no means conclusive [177, 178], and this SNP may simply mark another causal site in the gene or region. The absence of an association of rs1061170 with AMD risk in patients of non-European ancestry in this meta-analysis provides some evidence for that, although this interpretation must be treated with caution because of the much more modest size of the data set in non-European subjects. However, the lack of, or attenuated association in non-European individuals could occur if rs1061170 is not the causal SNP but a marker in non-European populations. Because there are some differences in the LD structure of *CFH*, rs1061170 may mark the causal variants less well. Nevertheless, this finding suggests that other variants in this region are worthy of further investigation as identification of these would have important implications for use of genetic information in risk prediction of AMD, and importantly the development of therapeutic interventions that target factor H related pathways. Given that the initial GWAS in AMD was modest in size by current standards (96 cases and 50 controls) and a SNP array only of ~100,000 variants (compared to those currently in use by GWAS, i.e. up to 1 million markers), larger GWAS analyses in AMD, involving more dense next generation whole genome SNP arrays together with imputation of

un-typed SNPs using standard methods are likely to be fruitful in refining the association signal in this region.

2.4 Conclusions

The rs1061170 variant of the *CFH* gene is associated with the risk of AMD, with slightly larger effect on GA than other forms of the disease. As the evidence base on this association has matured, the association remains robust and in keeping with findings from the initial discovery study. There was no evidence for a true interaction of this variant with smoking though both exposures independently increase risk of AMD. The attenuated association of this variant with AMD in non-European subjects provides some evidence that rs1061170 could be a marker rather than a causal variant in these populations, or that sample sizes need to be much greater in order to detect an adequate signal. Since the signals of association in the region span more than one gene, further genetic analyses, as well as studies that measure the complement factor H protein itself are required to assess whether complement factor H itself or the product of another complement related gene in the near vicinity mediates the association with AMD risk. This information will be essential to convert the exciting genetic findings into new therapies for AMD. Further prospective studies in general populations with records of incident AMD will be required to precisely define the absolute risk of AMD associated with the carriage of this and other SNPs which have been linked to AMD before a genotype based predictive test could be adopted in clinical practice.

3 Genetic variation in complement factor H and risk of coronary heart disease: systematic review and meta-analysis of published, unpublished and newly genotyped studies

There is evidence for an epidemiological association between both AMD and CHD [179-181] and that the complement pathway may contribute to both (as discussed in chapter 1). Following evidence of the robust association of *CFH* genetic variants with AMD risk, and the previous association of circulating complement components with CHD risk, investigations have evaluated the role of the rs1061170 variant in the *CFH* gene with the risk of CHD. Due the nature of randomised allocation of genotypes at conception (which balances the distribution of potential confounding factors among genotypic classes) and the fact that genotype is invariant and unaffected by the presence of disease, genetic studies are less affected by biases that may affect non-genetic observational associations [45]. If the association of *CFH* SNPs with CHD were verified, it would provide strong evidence for a partially overlapping pathological basis and common routes to the prevention of the AMD and CHD.

Individual genetic association studies of the rs1061170 variant in CHD have thus far been small, and the findings inconsistent. Moreover, few studies have evaluated the effect of *CFH* SNPs on blood lipids or inflammation markers that have been linked to CHD, and which might provide a plausible mechanism for the reported association between *CFH* SNPs and CHD. There has been some interest in the association of SNPs in *CFH* and C-reactive protein (CRP), as fH is thought to harbour a binding site for CRP and CRP itself been implicated in CHD. More recent evidence has also suggested that combing both *CFH* and *CRP* genetic profiles could be useful in the prediction of incident MI [182].

In order to address some of the uncertainties, in this chapter I conducted a systematic review of published genetic association studies and supplemented these data with information from several studies in which genotyping of the rs1061170 variant had already been completed as well as a further four studies in which genotyping was conducted *de novo* for the purposes of addressing the research question. In addition, where possible, I also evaluated the association

of the rs1061170 variant with established CHD risk factors as well as biomarkers of inflammation.

3.1 Methods

3.1.1 Published Studies- Systematic review

I searched two electronic databases (MEDLINE using PubMed and EMBASE up to January 2009). Using free text and MeSH terms “complement factor H”, “rs1061170”, “Y402H”, “myocardial infarction”, “cardiovascular disease”, “coronary heart disease” with the search being limited to “human” studies. Further studies were identified through reference lists of the identified publications and by using the ‘related articles’ link in PubMed. Electronic searches of published data were supplemented by personal contact with a number of studies known to have repositories of genetic data and information on cardiovascular outcomes.

3.1.1.1 Inclusion Criteria

To be included, studies had to be cohort or case control studies evaluating unrelated individuals, and had typed the rs1061170 variant or one in strong linkage disequilibrium (LD; r^2>0.7) with it. Outcomes recorded were fatal and non-fatal myocardial infarction, angiographic evidence of atherosclerosis, and coronary artery bypass grafting. Both incident and prevalent CHD outcomes were recorded and analyses were limited to studies investigating individuals of European ancestry.

3.1.1.2 Data extraction

I extracted data and it was checked for consistency by others. Any disagreements were resolved by consensus. Following HuGENET Guidelines (http://www.hugenet.ca), I extracted information on whether genotyping was carried out masked to outcome, deviation from HWE and the method used for genotyping. Where directly available, information on coronary events in each genotype group was used. Where this was unavailable the number of cases by genotype category was estimated using published information on allele frequencies and odds ratios for coronary events.

3.1.2 Studies in which genotyping was performed *de novo*

Genotyping of the rs1061170 variant was performed in six new studies.

The Second Northwick Park Heart Study (NPHSII) is a prospective study of 3,012 healthy Caucasian men aged 50 to 64, recruited from 9 general practices across the UK in 1986 [183]. At the time of recruitment individuals were healthy with no history of CHD. Baseline characteristics were ascertained using a questionnaire completed at the beginning of the study. Individuals were followed up annually at which time events were ascertained and blood samples taken. End points recorded included fatal and non-fatal CHD, coronary artery by-pass surgery (CABG) and silent MI (ascertained using ECG). In total 227 CHD events were recorded, with a median length of follow-up of 10.6 years (9 days to 13.7 years).

University College Diabetes and Cardiovascular Study (UDACS) comprises 1020 consecutive subjects recruited from the diabetes clinic at University College London Hospitals NHS Trust (UCLH) between the years 2001–2 [184]. All patients had diabetes according to WHO criteria. No subjects requiring renal dialysis were recruited. Subjects were categorised by the presence/absence of clinically manifest CVD. CVD was recorded if a patient had one or more of CHD, peripheral vascular disease (PVD) or cerebrovascular disease (CVD). The presence of CHD was recorded if any patient had positive coronary angiography or angioplasty, coronary artery bypass, a positive cardiac thallium scan or exercise tolerance test, documented evidence of myocardial infarction or symptomatic/treated angina.

Ealing Diabetes Study of Coagulation (EDS) is a cross sectional study of 927 patients recruited consecutively from the diabetic clinic at Ealing Hospital over a two year period of time [185]. 927 patients remained in the study. Three ethnic groups were collected, although for the purpose of the analyses presented here, investigation was limited to those of European ancestry. Patients were investigated for CHD if they had symptoms (chest pain). If neither the patient nor the clinic records suggested clinical symptoms of CHD, these patients were recorded as negative for CHD. If either the patient or clinic records suggested they had had an event or had angina the full medical records were searched to identify the criteria used to determine the event. Patients with chest pain but with normal ECG, Holter exercise test,

thallium scan or angiogram, were classified negative for CHD. Patients positive for CHD included patients with severe atherosclerosis, shown by angiogram, and/or requiring surgical intervention (per-cutaneous intervention or CABG), those who had been admitted with an acute MI determined by ECG/increase in cardiac enzymes, patients who had a silent MI but showed unequivocal ECG changes suggesting previous MI, and those with a positive exercise test or thallium scan for ischaemia.

The Hypercoagulability and Impaired Fibrinolytic function MECHanisms predisposing to MI (HIFMECH) study is a European multicentre case-control study of Caucasian men under the age of 60 who have suffered a MI [186].The dataset excluded individuals with familial hypercholesterolaemia and insulin-dependent diabetes mellitus. Age-matched controls from the same regions as cases were recruited. Recruitment centres included Stockholm, London, Marseilles and San Giovanni Rotondo. Blood samples from both affected individuals and controls were taken following an overnight fast. In total 491 individuals with MI and 517 controls were available for genotyping.

Whitehall II (WII) study recruited 10,308 individual (70% men) between 1985 and 1989 from 20 London based Civil service departments. Genetic samples were collected in 2004 from over 6,000 participants and are included in the analyses presented here [187]. The study is highly phenotyped for cardiovascular outcomes with 9 phases of follow up. In total 5,039 individuals were available for genotyping and are included here, of which 88 are cases of CHD. Data were used as a nested case control study rather than prospective.

English Longitudinal Study of Ageing (ELSA) is a national cohort of participants (48% men) aged over 50 years recruited from the Health Surveys for England in 1998, 1999 and 2001. Genetic data were collected at wave 2 of the study (2004/5) [188]. A wide range of phenotypic measures relevant to ageing are available. These measures were made at Wave 0 of the study (1998, 1999 and 2001) and at follow up (2004/5).

3.1.3 Previously unreported studies

The British Women's Health and Heart Study (BWHHS) [189], were known to have a genetic repository and prevalent and incident measures of CHD. Principal investigators of the study

were contacted to ascertain if rs1061170 or SNPs in the *CFH* gene had been typed, and second the Wellcome Case Control Consortium 1 (WTCCC 1) in which a genome wide association scan for CHD had previously been completed [103]. Information on a SNP in LD with the rs1061170 variant was extracted as part of a previous data application by Professor Hingorani's group.

BWHHS is a prospective population based study which recruited 4286 women between the ages of 60-79. Women were randomly selected from 23 general practices across the United Kingdom. Follow up for disease outcomes is by annual medical review (with validation checks) and obtaining copies of death certificates. More detailed follow-up is collected on coronary heart disease, stroke and type 2 diabetes and cancer events.

The WTCCC1 CAD collection includes 1988 individuals with either a history of MI or coronary revascularisation before the age of 66 years, as well as a strong family history of CAD with controls from two independent control groups the National Blood Service and the 1958 Birth Cohort- recruited as part of the WTCCC project. These studies provided information on genotype for the rs1061170 variant, or one in LD with it based on prior genotyping.

3.2 Statistical analysis

3.2.1 Newly genotyped studies

Continuous variables were logarithmically transformed where appropriate and differences tested by unpaired t-test or ANOVA. Differences in the distribution of categorical variables were assessed by the χ^2 test. Allele frequencies between groups were compared using χ^2tests, and tested for departure from HWE in controls. Hazard ratios (HR) were calculated in the prospective study (NPHSII), using Cox regression models with corresponding 95% intervals. Odds ratios (OR) were calculated from case-control and cross sectional studies using logistic regression models. Genetic associations were tested using a pair wise comparison, individuals homozygote for the T allele being the reference group. Dominant and recessive models were also tested. All gene-disease associations were adjusted for age as a continuous outcome and, in NPHS II, further adjustment was made for recruitment site. To evaluate the association between genotype and continuous risk markers the mean difference

in each variable between genotype groups for each quantitative trait within each study was calculated, and then the overall weighted mean difference using random effect models was derived.

Thus, three sources of information contributed to the meta-analysis: (i) newly genotyped studies: NPHSII, HIFMECH, UDACS and EDS. (ii) studies with unpublished data of *CFH* SNP and CHD status identified by direct contact with study authors and (iii) published studies identified by literature search;

3.2.2 Meta-analysis

For studies that typed the *CFH* SNP rs1061170 directly, pair-wise comparisons using the TT homozygote group as the reference group were made. Where information on rs1061170 was unavailable, information on any SNP in linkage disequilibrium (r^2>0.7) with rs1061170 was incorporated. A random effects model meta-analysis was used weighting effect size in individual studies by the inverse of the variance. Heterogeneity was tested using the Der Simonian and Laird Q test and quantified using the I^2. Evidence for publication bias was assessed by visual inspection of Begg's funnel plots.

3.3 Genotyping

Genotyping for NPHSII, UDACS, EDS and HIFMECH was carried out using ABI Taqman assay by design (primers, probes and PCR conditions described below). Variants in ELSA were genotyped commercially using SNPLex, and *CFH* SNPs were included as part of a custom SNP array designed by the Institute of Translational Medicine and Therapeutics, the Broad institute and the National Lung and Blood Institute supported Candidate gene Association Resource Consortium (Human CVD BeadChip; Illumina) [190], and typed at the Bart's Genome Centre by Jutta Palmen. Methods for the newly genotyped studies are described below, and from all studies included in the meta-analysis are summarised in **Table 3.1.**

Taqman

Primers and probes designed by Applied Biosystems (ABI TaqMan assays by design, Warrington, UK. Forward Primers GGTCCTTAGGAAAATGTTATTTTCCTTATTTGG; reverse primer: GGCAGGCAACGTCTATAGATTTACC; probes- forward labelled with VIC: TTTCTTCCATGATTTTG, reverse labelled with FAM: TTCTTCCATAATTTTG) were used in 384-well Thermo-fast PCR plates (Fisher Scientific). 2.5 ng DNA was used per reaction. Samples were left to dry prior to adding PCR reaction mastermix (containing AmpiTaq Gold DNA polymerase, AMPErase uracil-N-glycosylate (UNG), dNTPs with dUTP and dichlororhodamine acceptor dye, passive reference to which samples are normalised), the SNP assay containing primers and VIC and FAM probes. Cycling conditions were 50°C for 2 min, 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec, and 60°C for 1 min. Genotypes were read on a 7900HT sequence detector with laser fluorescence detection (Applied Biosystems) using Sequence Detection Systems (SDSv2.1) software. Genotypes were automatically called and grouped into three distinct clusters according to genotype (allele 1 homozygote, allele 1/2 heterozygote, or allele 2 homozygote). Negative controls were present in each plate to test for DNA contamination. The rs1061170 SNP is common with a minor allele frequency (MAF) of 0.38, therefore positive controls were not required. Call rates and Hardy Weinberg frequency were determined following genotyping.

3.4 Results

3.4.1 New Genotyped Studies NPHSII, HIFMECH and UDACS

The distribution of clinical and demographic variables by rs1061170 genotypes in all newly genotyped studies is shown in **Table 3.1**. There was no deviation of genotype frequency from Hardy-Weinberg proportions in controls for all studies. There were no differences in genotype frequency for rs1061170 by case-control status in any of the studies regardless of study design and the inclusion of incident or prevalent CHD cases (**Table 3.2**). When finding from all studies were pooled using random effects meta-analysis. there was no association between the rs1061170 SNP and the established risk factors for CHD including cholesterol (total LDL and HDL cholesterol), smoking, age, or systolic blood pressure either within study or after

pooled analysis (**Table 3.1** for individual study results and **Table 3.3** for pooled results). There was a borderline difference in triglycerides between CC homozygotes and the reference TT group (p=0.06) although no difference between the heterozygote (TC) and reference (TT) group. There was also no evidence of association between rs1061170 and other circulating inflammation markers, including CRP. (**Table 3.1, and Figure 3.1**).

Table 3.1 Association of established cardiovascular risk factors with rs1061170 in 4 newly genotyped studies *indicates where variables have been log transformed

	rs1061170									
	TT			TC			CC			
NPHS	**Total**	**Mean/ %**	**SD**	**Total**	**Mean/ %**	**SD**	**Total**	**Mean/ %**	**SD**	**P Value (Anova/ chi^2)**
Age	1030	56.22	3.39	1268	55.99	3.46	401	56.21	3.45	0.23
Systolic Blood Pressure*	1030	136.91	18.43	1267	137.27	18.73	400	138.02	19.53	0.60
Diastolic Blood Pressure	1030	84.92	11.17	1267	84.69	11.25	400	83.82	11.42	0.24
Total Cholesterol	1022	5.73	1.02	1259	5.72	1.02	397	5.84	0.98	0.11
LDL-Cholesterol	668	3.98	0.97	798	4.00	0.96	229	4.06	0.93	0.57
HDL-Cholesterol	698	0.84	0.26	850	0.83	0.26	245	0.85	0.26	0.45
Triglycerides*	1023	1.77	0.90	1260	1.80	0.97	397	1.88	1.00	0.20
BMI*	1029	26.20	3.30	1268	26.28	3.48	399	26.23	3.19	0.85
Smoking Status (% Current)	279	27.10		364	28.70		120	29.90		0.50
C-Reactive Protein*	938	3.07	3.20	1159	3.09	3.21	355	2.97	3.09	0.83

Table 3.1 continued

	rs1061170									
	TT			TC			CC			
Ealing Diabetes	**Total**	**Mean/ %**	**SD**	**Total**	**Mean/ %**	**SD**	**Total**	**Mean/ %**	**SD**	**P Value (Anova/ chi^2)**
Age	123	64.11	12.78	158	63.46	14.00	41	61.51	15.69	0.58
Systolic Blood Pressure*	121	134.97	19.28	155	138.48	23.05	41	4.89	132.49	0.17
Diastolic Blood Pressure	121	71.65	11.57	155	73.86	11.71	40	71.90	8.27	0.24
Total Cholesterol	117	4.92	1.13	151	5.02	1.12	40	4.61	1.09	0.13
LDL-Cholesterol	104	2.82	0.83	128	2.90	0.93	34	2.65	0.95	0.34
HDL-Cholesterol*	111	1.20	0.36	139	1.13	0.35	36	1.17	0.33	0.33
Triglycerides*	111	1.68	1.04	139	1.98	1.15	36	2.03	1.20	0.06
BMI*	113	28.60	5.36	145	29.84	5.79	40	29.12	5.23	0.20
Smoking Status (% Current)	23	18.70%		32	0.20		5	0.12		0.80
C-Reactive Protein*	28	2.29	3.13	42	2.63	2.60	12	1.68	1.11	0.46

Table 3.1 continued

	rs1061170									
	TT			TC			CC			
UCL Diabetics	**Total**	**Mean/ %**	**SD**	**Total**	**Mean/ %**	**SD**	**Total**	**Mean/ %**	**SD**	**P Value (Anova/ chi^2)**
Age	47	57.96	10.03	45	60.49	12.00	12	60.17	9.20	0.52
Systolic Blood Pressure*	47	138.68	17.12	45	138.75	16.94	12	139.03	18.46	1.00
Diastolic Blood Pressure	47	81.96	11.53	45	78.73	7.84	12	82.25	9.63	0.25
Total Cholesterol	47	4.95	1.05	45	4.84	0.98	12	4.73	0.89	0.75
LDL-Cholesterol	45	2.82	1.14	40	2.66	0.94	12	2.52	0.88	0.59
HDL-Cholesterol*	47	1.14	0.29	45	1.13	0.34	12	1.14	0.14	0.99
Triglycerides*	47	1.97	0.96	45	2.16	1.40	12	2.13	0.95	0.72
BMI*	46	27.50	3.65	45	25.72	4.28	11	26.98	3.56	0.10
Smoking Status (% Current)	11	0.24		8	0.19		3	0.25		
C-Reactive Protein*	46	1.71	1.23	45	1.75	1.49	12	1.63	1.48	0.96

Table 3.1 continued

	rs1061170									
	TT			TC			CC			
HIFMECH	**Total**	**Mean/ %**	**SD**	**Total**	**Mean/ %**	**SD**	**Total**	**Mean/ %**	**SD**	**P Value (Anova/ chi^2)**
Age	396	51.40	5.60	541	51.70	5.28	139	51.70	5.71	0.75
Systolic Blood Pressure*	395	128.34	15.03	534	127.62	15.62	136	127.06	17.16	0.65
Diastolic Blood Pressure	394	83.58	9.50	534	82.50	9.36	136	82.93	9.80	0.23
Total Cholesterol	389	5.46	1.10	511	5.48	1.09	129	5.43	0.97	0.88
Triglycerides*	389	1.65	0.73	511	1.62	0.70	129	1.65	0.74	0.71
BMI*	396	26.54	3.32	539	26.53	3.19	138	26.54	3.53	1.00
Smoking Status (% Current)	285	0.72		386	0.71		102	0.73		0.89
C-Reactive Protein*	388	1.80	1.98	511	1.59	1.92	134	1.36	1.73	0.04

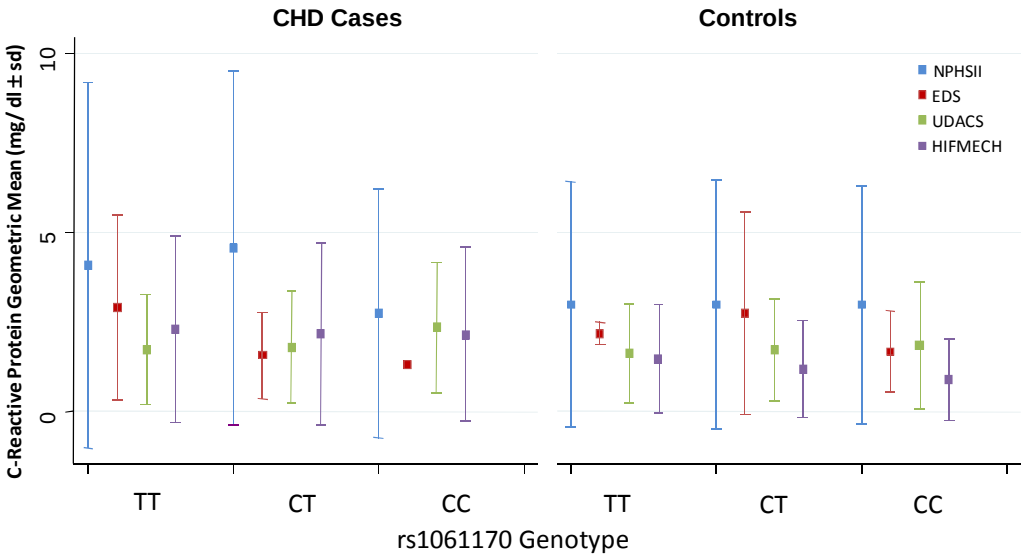
Table 3.2 Association of the *CFH* variant rs1061170 with CHD in four newly genotyped studies *Hazard ratios are given for NPHS II rather than odds ratios, calculated using Cox proportional hazards, the remainder of the studies were analysed using logistic regression.

	Case Control Status according to genotype group						OR (95% CI) for association with CHD	
	TT		TC		CC			
Study	Case	Control	Case	Control	Case	Control	TC vs TT	CC vs TT
NPHS*	110	920	120	1,148	38	363	0.87 (0.67, 1.12)	0.90 (0.63, 1.13)
HIFMECH	183	215	271	271	64	75	1.18 (0.91, 1.52)	1.0 (0.68, 1.47)
Ealing Diabetes	22	95	24	126	5	35	0.93 (0.65, 1.31)	1.0 (0.57, 1.76)
UCL Diabetics	54	180	76	195	16	57	1.27 (0.88, 1.81)	0.89 (0.50, 1.58)

Table 3.3 Weighted mean difference in the level of CHD risk factors between genotype groups of rs1061170, data pooled from 4 studies. Pairwise comparisons are shown of individuals homozygous or heterozygous for the risk 'C' allele compared to those homozygous for the common 'T' allele, which is used as the reference group.

Trait	No. of Individuals	No. of Studies	CC vs TT	CT vs TT
			Weighted mean difference (95% CI)	Weighted mean difference (95% CI)
Age (Years)	2189	4	0.02 (-0.35, 0.39)	-0.11 (-0.46, 0.25)
DBP (mmHg)	2180	4	-0.81 (-1.82, 0.20)	-0.41 (-1.70, 0.88)
TC (mmol/L)	2153	4	-0.02 (-0.19 ,0.15)	0.00(-0.07, 0.07)
HDL-C (mmol/L)	1149	3	-0.05 (-0.11, 0.02)	-0.04 (-0.09, 0.01)
LDL-C (mmol/L)	1092	3	-0.02 (-0.23, 0.19)	0.02 (-0.07, 0.11)
TRG (mmol/L)	2144	4	0.08 (-0.01, 0.17)	0.04 (-0.06, 0.14)
CRP (mg/L)	1913	4	-0.28 (-0.52, -0.04)	-0.08 (-0.26, 0.10)
BMI	2172	4	0.03 (-0.29, 0.35)	0.02 (-0.49, 0.54)

Figure 3.1 CRP concentrations stratified by the *CFH* rs1061170 variant in CHD cases and controls



3.4.2 Meta-analysis

Eleven eligible studies were identified among individuals of European descent [116, 117, 191-198]. Nine of these (4,510 cases, 18,381 controls) included unrelated individuals [116, 117, 192-196, 198], and had information on genotype frequencies according to case or control status that could be abstracted from the published report (**Figure 3.2**). All studied the association of SNP rs1061170. Two further studies were excluded on the basis that they examined association within families [191, 197]. The Whitehall II prospective study, ELSA, BWHHS, together with the WTCCC 1 genome wide association scan of coronary disease with controls from the National Blood Service and the 1958 Birth Cohort [103] contributed a further 2,417 cases and 16,526 controls. SNPs in LD (r^2>0.7) were used as a proxies for rs1061170 in ELSA, Whitehall II and WTCCC1 (**Figure 3.3**). Thus, in total, 7,909 cases and 38,576 controls contributed to the meta-analysis. Details of all studies, the genotyping platform used, and outcomes measured are given in (**Table 3.1)**

Figure 3.2 Flow diagram indicating identification of studies included in the meta-analysis of rs1061170 and CHD

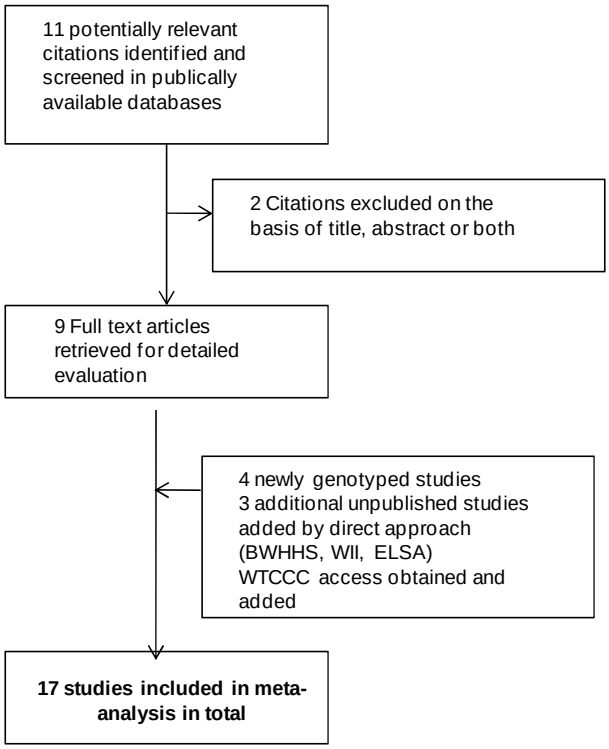


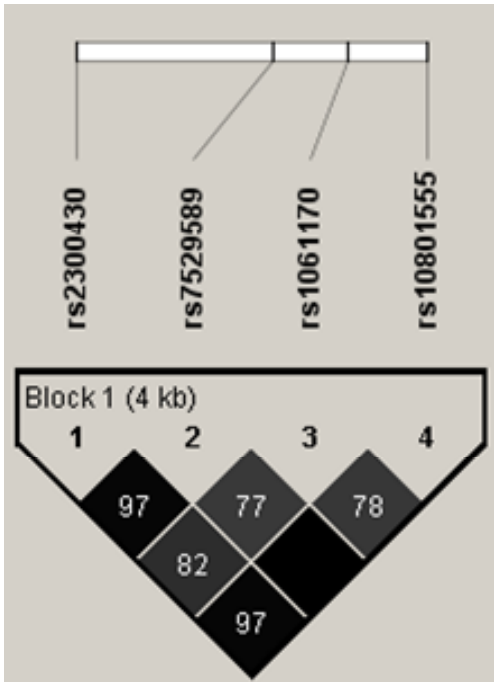
Table 3.4 Studies contributing to the meta-analysis of the association of the *CFH* variant rs1061170 and CHD. The first part of the table indicates the newly genotyped studies, and previously unreported studies and the second is a summary of the studies identified using the search strategy.

Study	Study Design	Sample Size	Outcome reported	SNP Typed	Genotyping Method
Newly genotyped and previously unreported studies					
NPHS II	Prospective Cohort study	3012	Fatal and non Fatal CHD events	rs1061170	Taqman (Applied Biosystems)
HIFMECH	Case Control	1074	MI	rs1061170	Taqman (Applied Biosystems)
UDACS	Cross Sectional	571	MI/ Positive exercise tolerance test / CABG or PTCA for CHD	rs1061170	Taqman (Applied Biosystems)
Ealing Diabetes	Cross Sectional	307	MI/ Positive exercise tolerance test / CABG or PTCA for CHD	rs1061170	Taqman (Applied Biosystems)
BWHHS	Nested Case Control (Cohort)	3437	Fatal and non fatal CHD events	rs1061170	Taqman (Applied Biosystems)
ELSA	Nested Case Control (Cohort)	5475	MI	rs2300430	SNPLex
Whitehall II	Nested Case Control (Cohort)	3445	MI	rs7529589	Illumina IBD 50K Cardiochip
WTCCC1	Case Control	4992	MI	rs10801555	Affymetrix 500K

Table 3.1 continued, Studies included in meta-analysis

Studies from Literature based Search					
Physicians Health Study	Nested Case Control (RCT)	670	MI	rs1061170	Taqman (Applied Biosystems)
Rotterdam	Nested Case Control (Cohort)	5529	MI	rs1061170	Taqman (Applied Biosystems)
Southampton	Case Control	448	Angiographically Confirmed CHD	rs1061170	Taqman (Applied Biosystems)
Atherogene	Case Control	1292	MI	rs1061170	Taqman (Applied Biosystems)
ECTIM	Case Control	2073	MI	rs1061170	Taqman (Applied Biosystems)
Health Professionals Study	Nested Case Control (Cohort)	2359	Any CHD	rs1061170	Taqman (Applied Biosystems)
Nurses Health Study	Nested Case Control (Cohort)	712	Non fatal MI or fatal CHD	rs1061170	Taqman (Applied Biosystems)
Mayo Clinic	Case Control	384	Angiographically Confirmed CHD	rs1061170	RFLP
ARIC	Nested Case Control (Cohort)	9424	CHD	rs1061170	Taqman (Applied Biosystems)

Figure 3.3 Linkage Disequlibrium plot of SNPs used for meta-analysis of rs1061170 and CHD. The analysis was based and plotted using HapMap data (NCBI Build 36). The LD values in the squares are the pairwise r^2 values between rs1061170 and other SNPs typed as a proxy for this SNP is shown.



The odds ratio for CHD in individuals homozygous for the C allele at rs1061170 encoding the 402H variant compared to those homozygous for the common T allele was 1.02 (95% CI 0.91, 1.15) and for heterozygous individuals the odds ratio was 1.05 (95% CI 0.99, 1.11), with little evidence of heterogeneity (I^2 37.3% for homozygous comparison and 0% for heterozygote comparison) (**Figure 3.4 a and b**). The summary point estimates were similar for previously published studies, previously unpublished studies and the newly-genotyped studies with no evidence for publication bias (**Figure 3.5**)

Figure 3.4 Meta-analysis of the association of the rs1061170 SNP encoding Y402H and CHD: (a) and individuals homozygous for the C allele for the C allele; **(b)** and heterozygous for the C allele, both with reference to TT homozygous subjects.

(a)

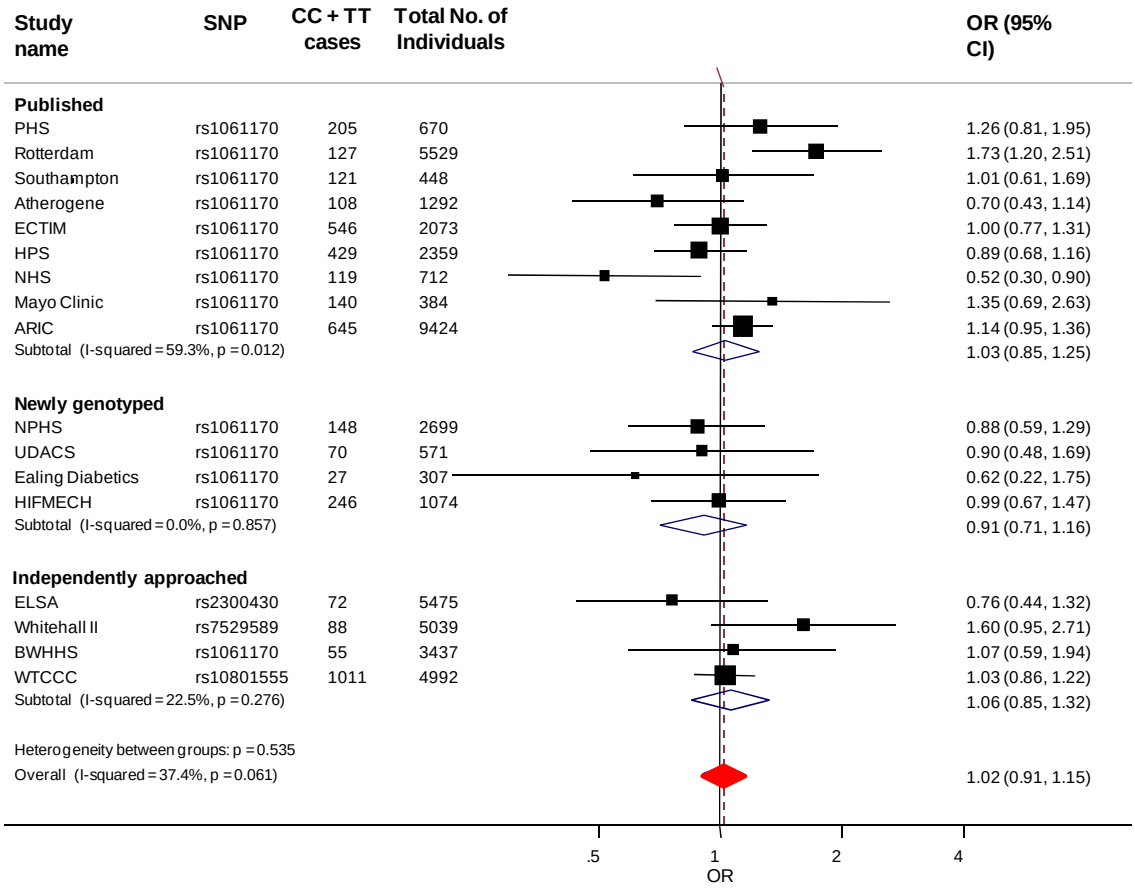


Figure 3.4 (b)

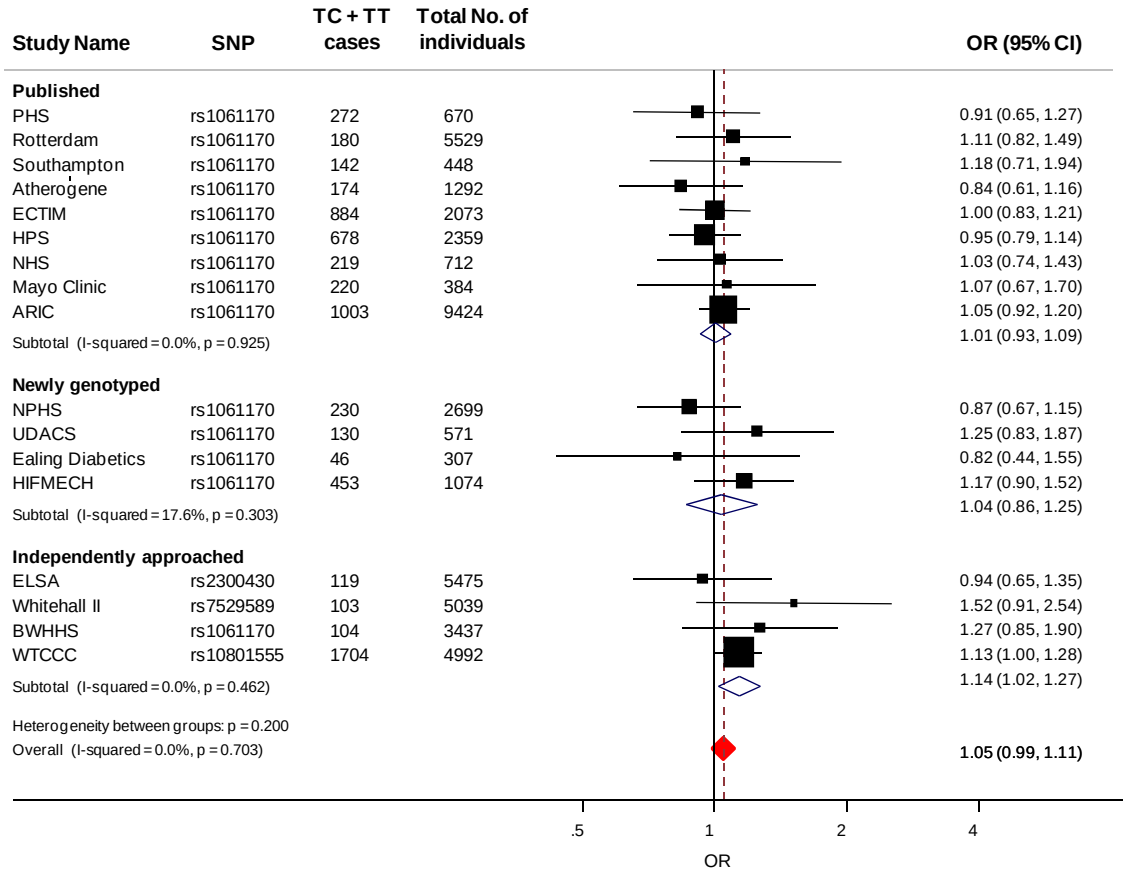
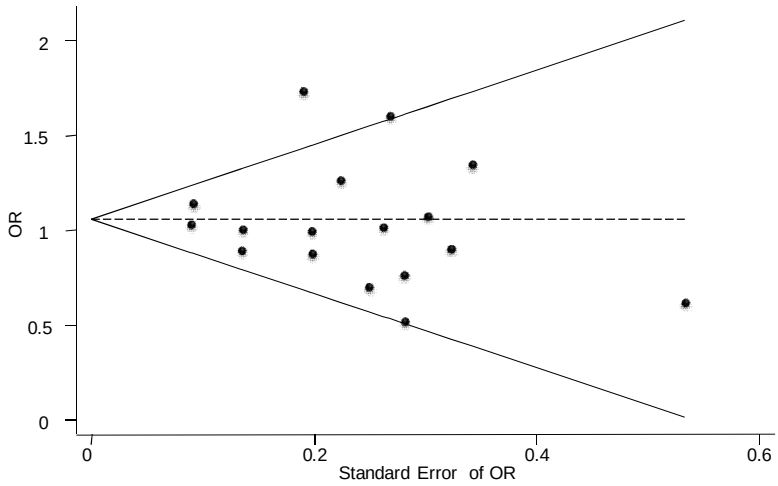


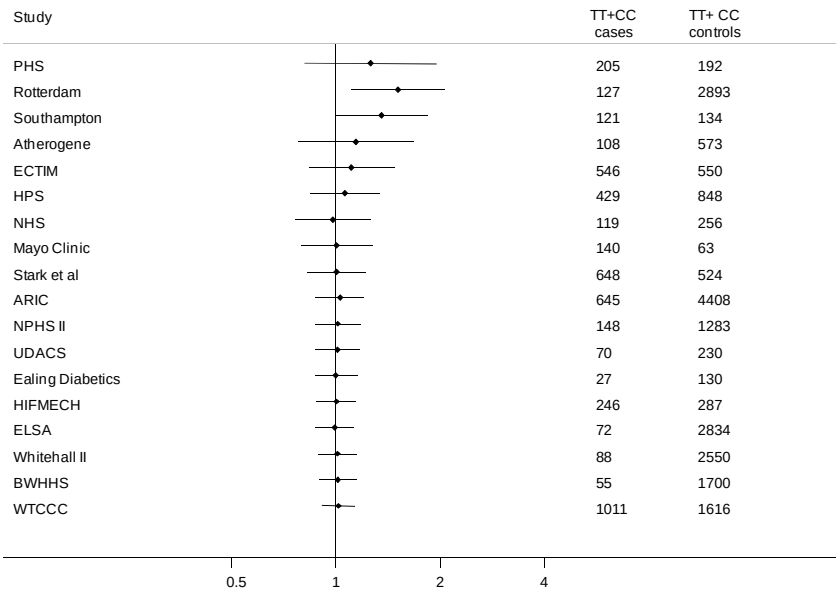
Figure 3.5 Begg's Funnel plot with pseudo 95% confidence intervals for the studies included in Figure 3.4



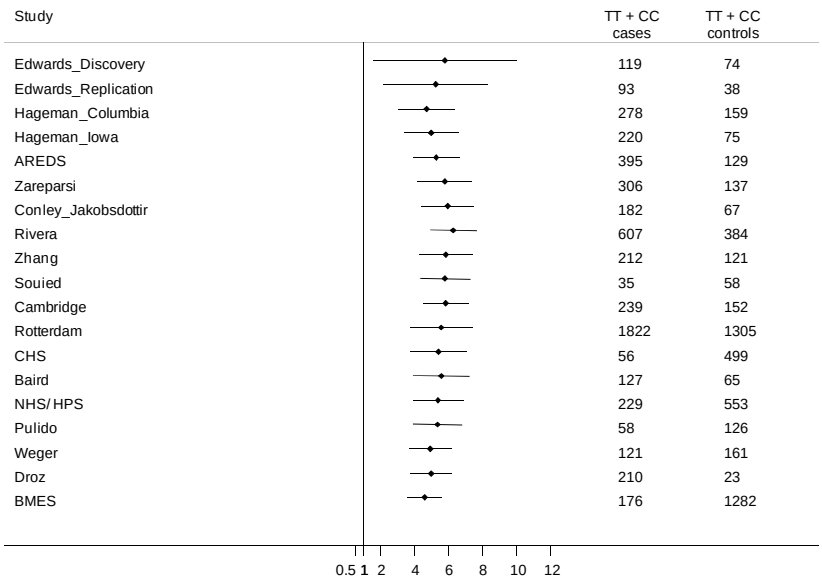
In a comparative cumulative meta-analysis of the published association studies of rs1061170 in CHD and AMD, the summary effect estimate exhibited only a minor attenuation over time for AMD while the summary effect estimate in the field of CHD attenuated to the null. **(Figure 3.6 a and b**)

Figure 3.6 (a) Cumulative meta-analysis of the association between rs1061170 encoding rs1061170 in *CFH* and CHD; and (b) cumulative meta-analysis of association between rs1061170 and late AMD. Both plots show comparisons for CC homozygous individuals compared to the TT reference group

(a)



(b)



3.5 Discussion

Using 7,909 cases and 38,576 controls, the largest accumulated data set on this question to have been studied to date, I identified a null association of the rs1061170 SNP in *CFH* with CHD events. The meta-analysis had >99% power to detect a per-allele odds ratio 1.2, at a 0.05 level of significance.

With the exception of a possible association of the minor (C) allele of rs10611760 SNP with higher circulating triglyceride concentration (which should be regarded as hypothesis generating but not definitive) this variant was not associated with any of the other risk factors or biomarkers evaluated. However, confirmation of the triglyceride association will be important and is of potential interest because a physical interaction of fH with triglyceride-rich proteins has been reported, which may have bearing on the mechanism of the fH regulation of complement function in health and disease [35]. Of the inflammatory markers studied there has previously been particular interest in the association of *CFH* genotype with CRP, given fH is thought to bind CRP. However recent evidence suggests that binding of CRP to fH does not occur under physiological conditions [177] and, that moreover, the association of CRP with CHD events is non-causal [56-58, 199].

Previous studies, reported an interaction of BMI and the rs1061170 SNP in *CFH* on the risk of AMD [152]. However there was no evidence for such an interaction in relation to CHD, nor was there any evidence that the effect of the SNP on CHD risk was different in those with hypertension which had also been suggested previously [198] (**Table 3.4**). The reliable demonstration of interactions in genetic studies requires very much larger sample sizes than those studied in prior reports or even this large meta-analysis. However, the null association of with CHD events identified here indicates that investigation of any possible interaction is unlikely to be a high priority for future research.

Table 3.5 Association of rs1061170 in *CFH* with CHD limited to individuals with a blood pressure recording of >140/90, irrespective of blood pressure lowering medication.

Study	OR (95% CI) for association with CHD in individuals with Hypertension only	
	TC *vs* TT	CC *vs* TT
NPHS	0.76 (0.53, 1.10)	0.79 (0.38, 1.55)
Ealing Diabetes	0.81 (0.49, 1.34)	0.41 (0.15, 1.13)
UCL Diabetics	1.85 (1.07, 3.18)	1.71 (0.75, 3.93)
HIFMECH	0.93 (0.59, 1.47)	0.77 (0.38, 1.55)

The rs1061170 SNP was first highlighted in a genome wide analysis for its association with AMD rather than CHD. AMD shares many risk factors with CHD, and inflammation has been implicated in both. Although the association of this SNP with AMD has proved robust in numerous subsequent studies, the initially positive association with CHD (which was smaller in effect size than that for AMD) has progressively declined to the null as more data on this association have accrued added. By contrast, the association with AMD has been reinforced and the confidence limits around the effect estimate have narrowed. (**Figure 3.6 a and b**).

Inflammation has been implicated in a number of common diseases of ageing, of which CHD is perhaps the most studied and complement activation is one (but not the only) way in which the inflammatory response can be elicited. However, much of the evidence for the link between inflammation and CHD is observational and may therefore be prone to confounding since many inflammation markers are also associated with established risk factors for CHD. SNPs in genes involved with inflammatory pathways have been associated recently in several common late onset disorders including diabetes, Alzheimer's and certain forms of cancer, providing strong evidence of the importance of inflammation in the pathogenesis of these disorders. However, none of the genetic variants associated with CHD to date relates directly to inflammation although it is likely that only a very small proportion of SNPs contributing to

CHD risk has been uncovered so far. More loci have recently been identified by meta-analysis of genome wide and association studies of CHD [200], which implicate loci associated with lipids and lipid handling, primary protein structure, tissue specific gene expression as well as established risk factors for CHD. Some of the loci were notable in that they associate with other diseases and disease traits, supporting the view of common pathologies existing between diseases.

Although *CFH* does not appear to provide a point of overlap in the pathogenesis of AMD and CHD, recent evidence suggests that overlaps do exist in other pathways. A recent genome wide-association study in AMD found several new susceptibility loci known for their association with circulating HDL-C and triglycerides [36], pointing to possible common mechanisms and perhaps common preventative treatments for both AMD and CHD. More broadly, it may prove informative to continue to study SNPs identified for an index association with AMD for their effect on CHD risk and *vice versa.*

There are however some limitations to the findings that have been described here. Although the rs1061170 variant in *CFH* is a non-synonymous coding SNP, it remains uncertain whether the encoded variant is itself functional and therefore whether its association with AMD is explained by LD with another causal allele. The null association of this variant with CHD therefore does not complete exclude the possibility of association of other SNPs in this gene with CHD, and more fine mapping approaches may be helpful in answering this question. These findings however do not completely rule out a potential role for complement activation in the development of CHD either, because variants encoding other components of the complement cascade, which have been associated with AMD, were not studied here. Moreover, this analysis does not exclude a potential causal role of inflammation mediated by pathways other than complement activation in the pathogenesis of CHD.

3.6 Conclusion

In summary, despite a large sample size and high degree of power to identify modest effects, these findings were unable to confirm the previously reported association of the AMD-associated SNP rs1061170 with CHD risk. With the possible exception of triglycerides, this

SNP was also not associated with any of a wide range of risk factors and biomarkers of CHD studied. However, bearing in mind the possibility of partially overlapping pathogenesis, studies of AMD-associated SNPs in CHD and CHD-associated SNPs in AMD may prove mutually informative for the detection and prevention of both disorders.

4 Development and validation of a high-throughput assay for circulating complement factor H.

If alterations in fH concentration contribute to AMD risk, its measurement could aid risk stratification and lead to development of interventions that modify fH concentration or function which may help prevent or treat AMD. At the time this work started, no studies comprehensively examined the association of genetic variation in the *CFH* and *CFHR* region with fH concentration, or fH concentration with AMD risk. Studies reporting measurement of fH in patients with AMD and controls after the commencement of this work used different assays and have studied relatively few individuals (for a summary see **Table 4.7**).

A key limitation of research in this area to date has been the lack of a high-throughput assay for fH that can be used to measure fH in stored samples on a large scale. In order to overcome this major limitation, I helped develop and validate a new high throughout assay for circulating fH, using an existing commercially available anti-fH antibody used in the Human Factor H 'NL' BINDARID™ radial immune diffusion assay (RID) marketed by The Binding Site (referred to as BS from here on). The RID is a low-throughput method of assaying fH, measuring 14 samples per plate, requiring up to 72 hours for completion of the reaction, which is based on the formation of antigen-antibody precipitin rings. I helped adapt reagents sold as part of this commercial assay (antibody and calibrator) for a high throughput nephelometric assay on the Siemens BNII nephelometer. This is a standard assay methodology used for the quantification of a variety of circulating antigens. In this assay, antigen-antibody complexes in the sample solution scatter a beam of light with the degree of scatter being used to quantify antigen concentration. The same method is widely used to measure CRP, Lp (a) and SAP, on automated clinical chemistry analysers and thus the method could readily be adopted in hospital clinical practice, given that the instrument used for these measurements is standard piece of equipment in most hospital clinical biochemistry laboratories.

In this chapter I aim to describe, in detail, the adaptation of the commercially available reagents for use on the BN II nephelometer. In brief, I re-quantified the fH protein supplied as a 'calibrator' in the commercial BS assay with reference to a gold-standard purified fH preparation from Professor Paul Morgan's laboratory (Cardiff). Second, using the recalibrated

BS fH, I measured fH concentration in a large volume of pooled post-cryoprecipitate material (supplied as a gift from the National Blood Transfusion Service), which was used as a control sample for all the subsequent assay work. To confirm the validity and appraise the performance of the newly developed assay, I assessed agreement between the existing RID assay and the new nephelometric assay, compared the observed and predicted concentration of fH in test samples prepared by mixing varying proportions of a plasma sample of known fH concentration and a plasma sample completely depleted of fH, and assessed the reproducibility of measurement of fH by determining the intra- and inter-assay coefficient of variation. In the second part of this series of experiments I evaluated the properties of fH as an analyte. This was done by determining stability of fH to repeated freeze thaw cycles and to extremes of temperature. Finally the new nephelometric assay was used to assess the distribution of fH concentration in a small population of healthy individuals, to assess its distribution and range of measurement as a pilot study prior to the determination of these properties in a larger population based sample, the MRC-Fenland study (Chapter 5).

4.1 Materials and Methods

4.1.1 Adaptation of The Binding Site reagents for use on the Siemens BNII nephelometer, and assessment of validity of the nephelometric method

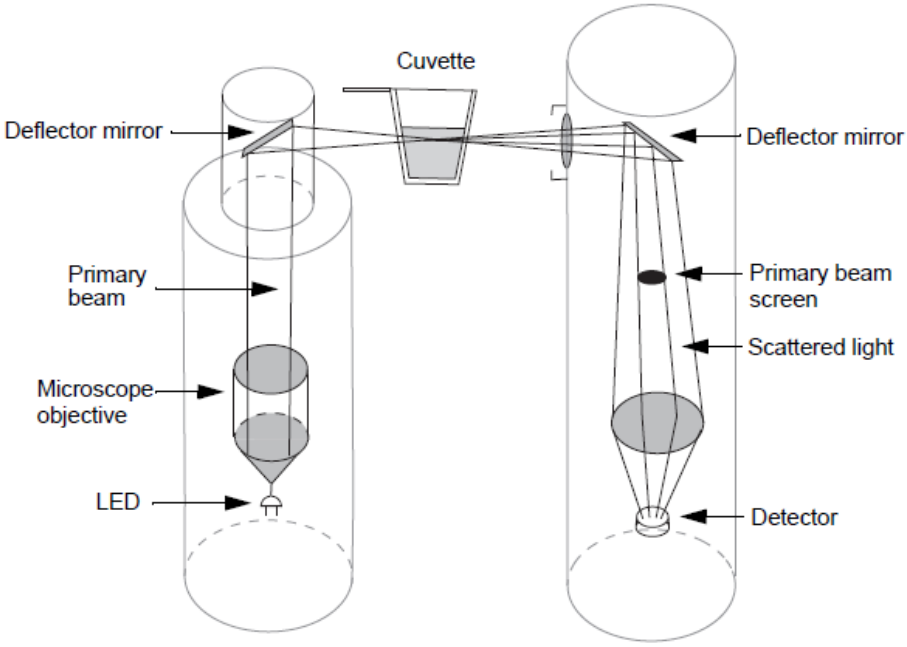
Immunonephelometry is one of the most frequently used methods to determine protein concentration in serum, urine and other body fluids. It is based on the property of antigen-antibody complexes to scatter a light beam. The amount of light scattered by the antigen-antibody complex is proportional to the amount of antigen in solution, with increased protein concentrations resulting in more light scatter. Operationally the BNII places the test sample into a cuvette robotically, into which is added the antibody specific to the analyte whose measurement is sought. Light emitted from a light emitting diode (LED) at a wavelength of 840 nm is transmitted through the cuvette, and light scattered thereafter is focussed onto a photodiode (i.e. a photodetector that is capable of converting light to either current or voltage) through a number of lenses and mirrors (**Figure 4.1**), which allows the system to be placed and function in a relatively small space. Once detected at the photodiode the signal is converted into a concentration of protein by measurement against a calibration curve

constructed against a standard protein of known concentration. Measurements are taken at two time points, the first when antibody has been added to the test sample when no immune complexes have formed, and second when all immune complexes have formed. The final concentration of protein is calculated by subtracting the initial measure from the final measure.

Each assay therefore requires the following reagents:

1. A specific antibody.
2. A standard preparation of the antigen of interest, necessary for generation of a calibration curve, against which subsequent measures can be read.
3. Control samples of known antigen concentration to check stability of measurement over the long term use of the assay.

Figure 4.1 Operation of the BNII nephelometer, the primary light beam is emitted from the LED, which is deflected through a mirror and transmitted through the cuvette containing a sample with unknown concentration of analyte and antibody. The light scattered through the cuvette is directly proportional to the concentration of the protein under measurement. Light scatter is detected by a photo detector able to convert this information into a concentration of protein. Schematic taken from the BNII Operator's Manual



Reagents from the commercial RID assay were identified that could be adapted for use with this method. The polyclonal sheep anti-human fH antibody used sold as part of the commercial assay was purchased in bulk (Human Factor H NL calibrator code RP030, Batch 247673), and stored in 10 ml aliquots at -80°C, and used as the primary antibody. The BS fH calibrator, assigned with a nominal value by the company of 700 mg/ L, when re-suspended in 655 uL of distilled water was used as the fH standard for nephelometric assay, following re-calculation of its true concentration with reference to a highly purified fH preparation of known quantity. The concentration of this highly purified fH was determined by measurement of its extinction coefficient, as described below. Finally pooled post-cryoprecipitate plasma (that is plasma material that remains after production of cryoprecipitate, rich in clotting factors VIII, XIII, von Willebrand factor and fibrinogen used for clinical treatment of clotting disorders has

been removed) was used as a control. We received 2L of pooled post-cryoprecipitate material as a gift from the National Blood Service, and this was stored in 5 ml aliquots at -80°C.

The operational parameters for the nephelometric assay of fH on the BNII analyser were established by The Binding Site (**Appendix 1**), and then field tested by Ruth Gallimore (Professor Mark Pepys's laboratory, Centre for Amyloid and Acute Phase Proteins, UCL), to determine an initial operating range on the BNII.

4.1.1.1 Preparation of a highly purified preparation of factor H from human plasma

Complement factor H was purified from plasma of individuals homozygous for either the H402 or Y402 variants by a sequential three-step FPLC method at 4°C on ÄKTAprime (GE Healthcare) in Professor Paul Morgan's laboratory, Cardiff, by Dr Svetlana Hakobyan. One hundred mL of filtered (EDTA) plasma was applied to a 5 ml HiTrap column (GE Healthcare) to which 10mg of mouse anti-human fH monoclonal antibody 35H9 (generated in the Morgan laboratory) was coupled. Bound protein was eluted at low pH and fractions containing fH (identified by ELISA) were pooled, dialysed against phosphate buffered saline (PBS: 137 mM NaCl, 10 mM Phosphate, 2.7 mM KCl, pH 7.4) and applied to a HiTrap-Heparin column (GE Healthcare) equilibrated with PBS. Bound proteins were eluted with 1M NaCl in PBS. Fractions containing fH were pooled and polished by gel filtration on Superdex-200 preparative grade matrix in a XK16/70 column (GE Healthcare). Purity was confirmed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) using plasma from identified homozygous donors.

4.1.1.2 Preparation of fH depleted plasma

Plasma depleted of fH during the experiments described above, was also used in subsequent validation experiments. In order to ensure that these samples were completely depleted of fH, they were run twice over the column. Plasma samples, taken before and after each run were analysed by dot blotting and Western blotting. For dot-blots, serial dilutions of plasma samples deplete of fH were spotted on nitrocellulose using a Bio-Dot SF microfiltration apparatus (Bio-Rad, UK). Blocked membranes were incubated with mouse monoclonal anti-fH antibody MRC OX-24 for detection of fH [201]. This antibody is specific for the C3b binding domain in the N-

terminal SCR 5 domain of fH. Bound monoclonal antibody was detected using horseradish peroxidise (HRP) conjugated rabbit anti-mouse immunoglobulin (Dako, UK) and visualised by ECL chemiluminescence (GE Healthcare). For Western blotting, purified fH, normal plasma and depleted plasma were blotted onto a nitrocellulose membrane which was subsequently blocked and probed with MRC OX-24 as for dot blots, with bound monoclonal antibody being detected using HRP-conjugated rabbit anti-mouse immunoglobulin as above.

4.1.1.3 Determination of extinction coefficient of purified factor H

The extinction coefficient is a method of determining the concentration of a purified protein based on the ability of the protein to absorb light at a given wave length, measured at a given distance of path length. The extinction coefficient of fH was determined directly by the dried weight analysis of the protein, carried out by Dr Patrizia Mangione in Professor Mark Pepys laboratory. A 20 ml solution of the purified fH in PBS was extensively dialysed against MilliQ water for 72 h at 4ºC. Before dialysis, absorbance at 280 nm, corrected for light scattering at 320 nm of 1/10 diluted samples of fH in PBS, was measured in a standard 1 cm light path cell using a DU650 UV-vis scanning spectrophotometer (Beckman-Coulter UK Ltd.) A final value of 0.108225 ± 0.0014 (mean ± SD, n= 4) was obtained.

After dialysis the solution was lyophilised, dissolved in MilliQ water (~1.4-1.5 ml) and placed in a 20 ml straight wall nickel crucible (Sigma) previously cleaned with ethyl acetate and dried at high temperature before its weight could be determined. The sample was lyophilized for 18 hr and then cryogenically dried over activated charcoal at -196º C (5×10^{-6} bar) for 48 hr.

4.1.1.4 Quantification of the concentration of Factor H of The Binding Site Calibrator

The fH calibrator supplied by The Binding Site (Human Factor H NL calibrator code RP030 Batch 247673, expiry date: 2010/07) had been assigned a nominal value of 700 mg/mL but this had not been previously calibrated against a purified fH preparation. Therefore experiments were conducted to recalibrate The Binding Site calibrator fH with reference to the purified fH protein obtained from the experiments described in the previous section.

In order to determine the concentration of the BS calibrator, purified fH (prepared according to procedure above) was diluted to concentrations of 500, 400, 200 and 100 mg/ L and loaded onto the RID plate in keeping with manufacturer's instructions (5 µl loaded per well). After loading, the plate lid was closed tightly, and placed flat in a moist box incubated at room temperature until precipitin rings developed to completeness. Samples were loaded in duplicate, and measures taken from these were used to establish the standard curve for fH. The BS calibrator was loaded on the same plate, having been reconstituted with 655 µl distilled water as recommended by the manufacturer, with further samples diluted at 1:2 dilution and run alongside the newly purified fH. On reaching completion, the diameter of the precipitin rings was measured by capturing an image of the RID plate using BIO-RAD GS-800 densitometer and measuring ring diameters using Adobe Photoshop software. A standard curve for fH was constructed by plotting the square of the diameters of fH precipitin rings against their specific protein concentrations, generating a line of best linear fit, using measures from newly purified fH plated at 4 different concentrations. The concentration of the BS calibrator could therefore be measured against this standard curve as determined by newly purified fH, and assigned a new value for use on the BNII nephelometer. In addition, plasma deplete of fH was loaded onto RID plates in order to ensure these samples generated no precipitin rings.

The newly assigned value for the Binding Site Calibrator was used as the standard for all subsequent analyses using the BNII nephelometer. Further calibrations on the BNII were carried out every month, or earlier if blank values of the cuvettes were high.

4.1.2 Validation of assay on BNII Nephelometer

4.1.2.1 Agreement between radial immune-diffusion assay and immune-nephelometric assay

Fifty plasma (EDTA) samples taken from healthy volunteers (kindly provided by Professor Tim Goodship, University of Newcastle) were used to measure fH in both the commercially available RID (The Binding Site) assay, and the newly developed nephelometric assay. For the assay 5 ul of sample was applied to wells in duplicate and allowed to diffuse to completion. Each plate was scanned using standard scanning software, images saved at original

dimensions as PNG files and ring diameter was measured using Image J software. Factor H concentration was determined by measuring ring diameter and calculating the concentration relative to the ring diameter produced by the BS Calibrator (Batch 247673), using its newly assigned concentration. Agreement between measurements was evaluated using methods described by Bland and Altman for agreement between two tests [202].

4.1.2.2 Recovery of known concentrations of fH from plasma deplete of fH

To determine the ability of the newly developed assay to detect fH, a recovery experiment was carried out. Venous blood was drawn into EDTA plasma from one healthy volunteer and the mean concentration of fH measured using the BNII from 4 replicates. This value was designated 100%. A fH depleted EDTA plasma sample (from the Morgan laboratory, as described above) was designated 0%. The two samples were mixed in varying proportions to generate samples with expected concentrations of 10, 20, 30, 40, 50, 60, 70, 80, or 90% of the value of the healthy donor. Each dilution was assayed using nephelometry on the BNII in quadruplicate. Expected concentrations were compared to observed concentrations in order to determine the degree of recovery.

4.1.2.3 Reproducibility of the nephelometric assay of fH

Non-fasting venous whole blood was drawn from the ante-cubital vein from 5 healthy volunteers, into EDTA, citrate, heparin and no anti-coagulant. The plasma samples were separated immediately after centrifugation for 15 minutes at 3000g, and frozen at -80 °C. Serum samples were allowed to clot at room temperature for 30 minutes, the clots rimmed and left for a further 90 minutes after which serum was separated and stored in aliquots -80 °C. All subjects gave informed consent and collection procedures and approval for the assay as a service development was given by the Chair of the local ethics committee.

Five replicate measurements of fH were assayed by nephelometry using fresh plasma and serum from five different individuals to determine average levels of fresh fH. Subsequently 5 measurements were taken over 5 days on the same 5 individuals to determine the inter- and intra- assay coefficient of variation (CV). Intra-assay CV was calculated for each sample on

each day (using freshly thawed material on the day of analysis), with inter-assay CV being calculated on measurements made on separate days.

4.1.3 Evaluation of analyte properties

4.1.3.1 Stability to freeze thaw.

Samples from the same 5 individuals, described above, were subjected to 5 repeated freeze-thaw cycles, by freezing at -80 °C, and thawing at 37° C, with re-measurement of fH using nephelometry at each thaw, to determine the stability of fH to repeated freeze-thaw.

4.1.3.2 Stability to temperature extremes

The stability of fH at extreme temperatures was tested by selecting samples from 2 of the above 5 to storage at -80°C; -20°C; 4°C; 24°C; 37°C; and 45°C over a period of 4. Measurements of fH were made using nephelometry at a weekly interval.

4.1.3.3 Measurement of fH in one hundred healthy individuals

Factor H was measured in plasma (EDTA) from 100 healthy individuals (samples courtesy of Professor Tim Goodship, Newcastle University), to establish an initial operating range for the assay and dilutions required to measure fH in a population based sample.

4.2 Results

4.2.1 Adaption of reagents from Binding Site RID Factor H for use on Siemens BNII Nephelometer

4.2.1.1 Isolation of plasma deplete of Factor H

Preparations of fH-H402 and fH-Y402 were obtained without any detectable contaminants or aggregates by this three-step method. The proteins were also free of fH-like-1 (FHL-1) and fH-related proteins. Yield was 25-50% and was similar for the two variants [203].

Dot blots confirmed plasma from donors was deplete of fH (**Figure 4.2**), and Western blotting (**Figure 4.3**) using the same antibodies confirmed the presence of the appropriately sized protein in normal plasma, with a higher signal from purified fH, and its complete absence from fH depleted plasma.

Figure 4.2 Dot blot of fH depleted and normal plasma. Samples of fH-depleted or normal plasma underwent serial dilution before blotting onto a nitrocellulose membrane, incubation with MRC OX-24 and then detection with a secondary HRP-conjugated rabbit anti-mouse antibody.

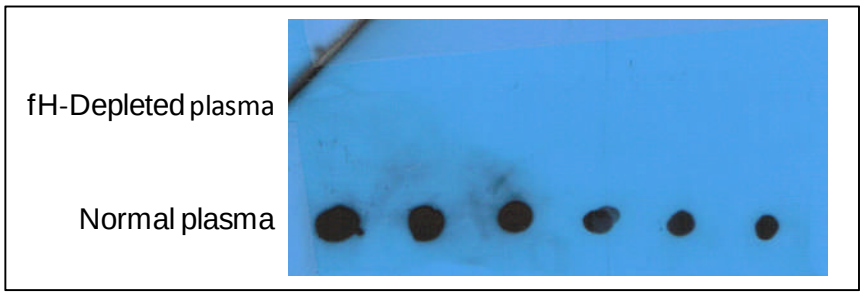
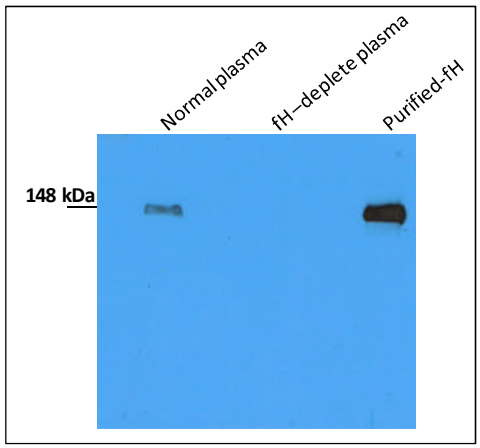


Figure 4.3 Western Blot of normal, fh-deplete plasma and purified fH. Using the MRC OX-24 anti-fH antibody, detection was with a HRP conjugated mouse anti-human monoclonal antibody (see Methods).



4.2.1.2 Determination of extinction coefficient of purified Factor H

The extinction coefficient of factor H was determined directly by the dried weight analysis of the protein. The weight of cryogenic dried protein was 0.01g (**Table 4.1**). This combined with the volume and absorbance measured where a value of 0.108225 ± 0.0014 (mean ± SD, n= 4) was obtained, allowed us to establish the concentration of the original solution of factor H in PBS which was 550 mg/L. Given the linear relationship between absorbance and concentration of the absorbent species (according to the Beer-Lambert law) we could derive the value of factor H extinction coefficient at 1mg/ml which was 1.9677.

Table 4.1 Weight of purified fH protein at various stages of drying. Tare value indicates the vessel with no material in situ.

	Weight (g)
Tare	**37.49**
After 18 hr lyophilization	**37.50**
After cryogenic drying	**37.50**
Cryogenic dried protein	**0.01**

4.2.1.3 Quantification of the concentration of Factor H of The Binding Site Calibrator

Using the BS RID assay, a standard curve was constructed using the concentration of the newly purified fH, previously determined as 550 mg/L. This is shown in **Figure 4.4**. The concentration of the BS calibrator was then determined using this standard curve, and found to be 389 mg/ L, and much lower than the nominal value of 700 mg/L reported by the company. The same batch of BS calibrator was used for all subsequent assays described in this thesis.

Figure 4.4 Standard curve and re-calibration of the Binding Site fH control with reference to purified fH using the commercial RID assay.

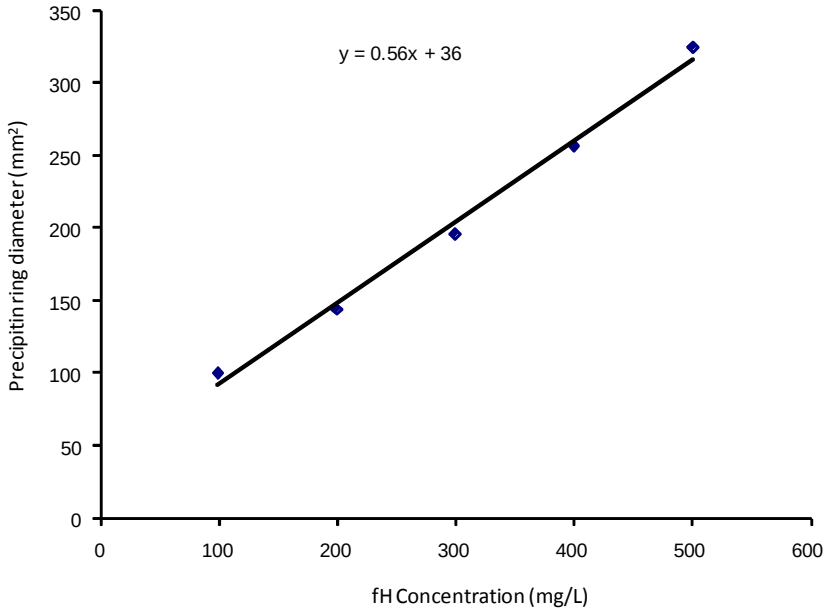
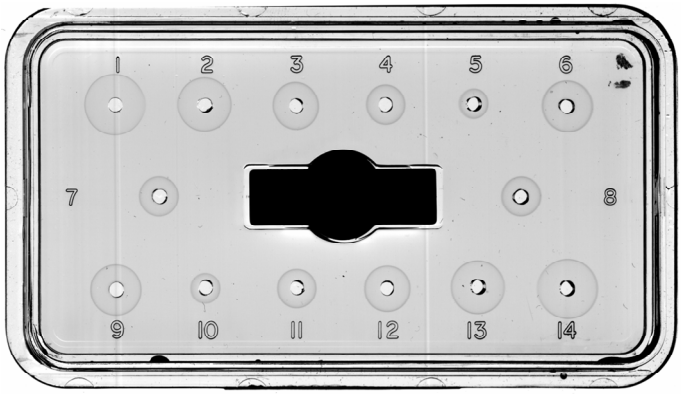


Figure 4.5 RID re-calibration plate. An image of the RID plate that was used to re-calibrate the BS calibrator using pure fH and plasma deplete fH. Neat BS calibrator is shown in wells 7 and 8, plasma depleted fH spiked with BS calibrator is shown in wells 6 and 9, and pure fH loaded at concentrations of 500 mg/L (wells 1 and 14), 400 mg/ L (wells 2 and 13), 300 mg/L (wells 3 and 12), 200 mg/L (wells 4 and 11) and 100 mg/ L (wells 5 and 10) are shown.



4.2.1.4 Quantification of factor H concentration in post cryo-precipitate material

Factor H concentration was measured in 10 replicates of pooled post cryo-precipitate material using the BS calibrator at the newly assigned concentration. A mean value of 288 mg/L was obtained, which was used as the control value on the BNII nephelometer.

4.2.1.5 Agreement between radial immune-diffusion assay and immune-nephelometric assay

Complement factor H was measured in plasma from 50 healthy individuals using both the commercial RID kit and the newly established BNII assay (**Table 4.2**), to test for agreement between the two methods using methods described by Bland and Altman [202]. Initial analyses, plotting the average (of BNII and RID measures) against the difference (BNII minus RID measures) demonstrated a larger difference in values between the two assays in samples either at both the higher and lower end of the range of fH concentrations, giving non uniform limits of agreement. It would be expected that the nephelometric assay would be more precise than the RID assay.

Given this is not an unusual situation in the evaluation for the measurement of biological analytes, Bland and Altman have, in addition to the classical Bland-Altman test, developed an

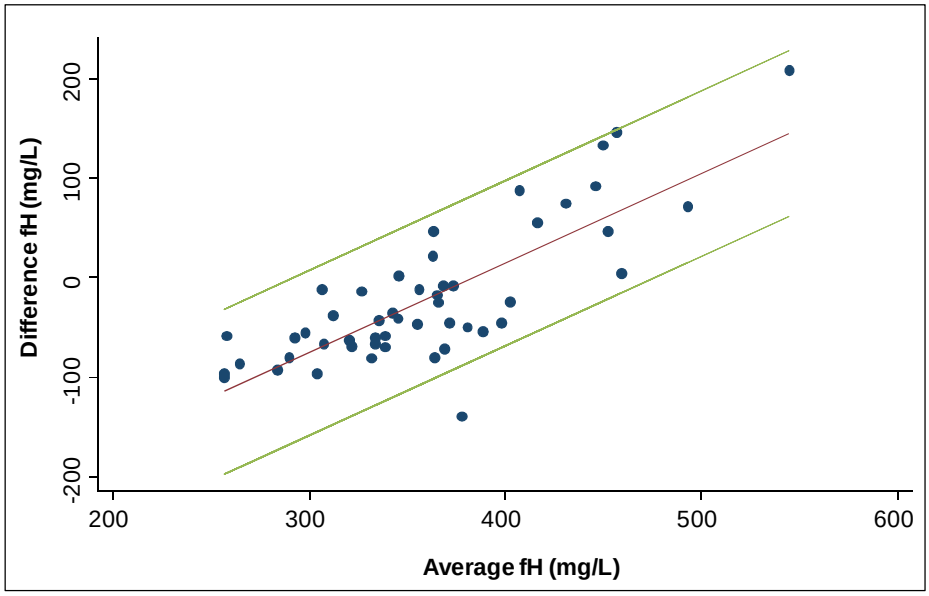
approach to assess agreement between two methods for the same marker, where the differences in assay values vary as a function of the absolute measure of the marker [204]. The approach uses a two step regression method. For the first step, the difference between the marker values obtained from the two assays is regressed on the average of the marker values as measured by the two assays, and in the second step, the limits of agreement are calculated and the residuals from the first regression are modelled on the average measure of the two tests.

Using this method, the nephelometric measures were more sensitive than the RID method (**Figure 4.5**). For example, at the higher than average reading of 493.2 mg/L the nephelometric assay reads 71.2 mg/L higher than the RID assay, with 95% limits of agreement at 15.4 mg/ L and 181.4 mg/L, and similarly at the lower average reading of 256.4 mg/L the nephelometric assay reads 96.9 mg/L lower that the RID assay with 95% limits of agreement at -197.8 mg/l and -31.8 mg/l.

Table 4.2 Comparison of the distribution of fH measured by the commercial RID assay and the newly developed nephelometric assay.

Summary statistics	fH Nephelometric measurement	fh RID measurement
Number of individuals	50	50
Mean	348.49	370.36
Standard deviation	92.93	40.28
25th Centile	291.50	341.04
Median	328.50	367.68
75th Centile	375.50	394.40
Range	443.00	169.58

Figure 4.6 Regression based Bland-Altman comparison of fH concentration in 50 samples analysed using the RID assay and the newly developed nephelometric assay. The difference refers to the value of fH estimated using nephelometry minus the value using RID.



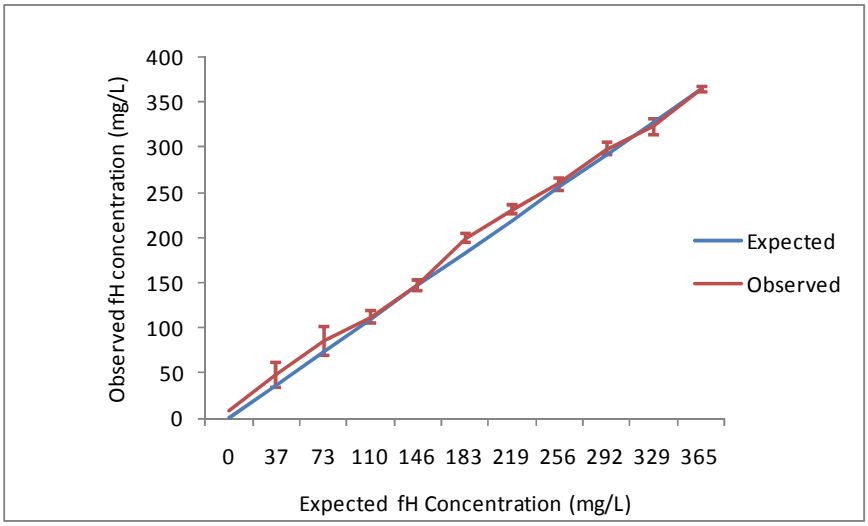
4.2.1.6 Recovery

The mean fH concentration calculated from 4 repeated measures of EDTA plasma taken from one healthy volunteer was 365 mg/L. Samples of varying fH concentration were then prepared by mixing this sample with fH depleted EDTA plasma to generate samples at concentrations that varied between 0-100% of the volunteer sample. After measurement on the BNII nephelometer, the percentage recovery was calculated by comparison of observed to expected values, and these are shown in **Table 4.3** and **Figure 4.6.** Recovery was close to 100% at all plasma concentrations. For the sample depleted of fH the observed fH concentration was 7.69 mg/L (SD 0.24 mg/L).

Table 4.3 Percentage recovery of fH from solutions of plasma containing a range of known fH concentrations

Concentration of fH in mixed sample, expressed as percentage of the undiluted control	Expected fH concentrations (mg/L)	Observed fH concentration (SD) (mg/L)	% Recovery (observed-expected/observed x100)
0%	0.00	7.69 (0.24)	.
10%	36.50	48.20 (13.79)	132%
20%	73.00	85.725 (15.80)	117%
30%	109.50	111.75 (6.50)	102%
40%	146.00	147.25 (5.44)	101%
50%	182.50	199.25 (4.57)	109%
60%	219.00	231.00 (5.60)	105%
70%	255.50	259.25 (6.99)	101%
80%	292.00	298.75 (7.54)	102%
90%	328.50	322.75 (9.54)	98%
100%	365.00	364.50 (3.32)	100%

Figure 4.7 Observed and expected concentration of fH in samples containing 0-100% of the fH concentration of an individual control plasma sample. Each point represents the mean ± standard deviation of the 4 replicate findings



4.2.1.7 Stability of Factor H using newly developed assay on BNII

4.2.1.8 Reproducibility and difference of fH measures in plasma and serum

To determine the influence of anti-coagulant, I assayed venous blood collected from 5 healthy volunteers in plasma, using anticoagulants EDTA, citrate and lithium heparin, as well as no anticoagulant (serum) (**Table 4.4**). Measures were broadly consistent between sample types.

Table 4.4 Mean (±SEM) of 5 measurements of fH from each of 5 healthy individuals

Fresh sample	EDTA	Citrate	Lithium-Heparin	Serum
Individual 1	313.00 (7.12)	306.30 (6.04)	328.20 (7.28)	319.40 (15.8)
Individual 2	267.00 (4.93)	272.00 (4.49)	287.40 (8.38)	287.20 (7.94)
Individual 3	271.00 (6.12)	270.50 (5.50)	286.00 (6.38)	288.40 (6.79)
Individual 4	241.60 (5.80)	236.30 (3.88)	253.80 (4.66)	267.00 (5.50)
Individual 5	229.80 (5.04)	227.00 (3.94)	235.00 (9.10)	247.00 (4.44)

Mean values of fH in serum and plasma were in the range 227 to 328 mg/L (**Table 4.4**), with intra-assay CV's ranging from 4 to 11% (**Table 4.5),** and inter-assay CV's ranging from 5-15% (**Table 4.6**). The assay has an overall intra-assay and inter-assay precision and accuracy within 7% and 10% respectively.

Table 4.5 Intra-assay coefficient of variation, calculated separately for plasma and serum using 5 replicate freshly thawed samples each day from 5 separate individuals.

Intra-assay coefficient of variation				
	EDTA	Citrate	Lithium-Heparin	Serum
Individual 1	5%	4%	5%	11%
Individual 2	4%	4%	7%	6%
Individual 3	5%	5%	5%	5%
Individual 4	5%	4%	4%	5%
Individual 5	5%	4%	9%	4%

Table 4.6 Inter-assay coefficient of variation, calculated separately for plasma and serum using 5 replicate freshly thawed samples each day from 5 separate individuals.

Inter-assay coefficient of variation				
	EDTA	**Citrate**	**Lithium Heparin**	**Serum**
Individual 1	10%	8%	7%	13%
Individual 2	10%	6%	5%	13%
Individual 3	10%	7%	7%	14%
Individual 4	10%	7%	8%	15%
Individual 5	14%	9%	13%	15%

4.2.2 Stability of factor H using newly developed assay on BNII

4.2.2.1 Stability to freeze thaw and extreme temperatures of measured factor H on the Siemens BNII

Complement factor H was stable to at least 4 freeze thaw cycles, after which some decrease in concentration was observed (**Figure 4.7**) On subjecting samples from two healthy volunteers to a range of temperatures, fH was found to be is stable at all temperatures up to 37°C. However no fH was detectable after incubation at 45°C for one week (**Figure 4.8**).

Figure 4.8 Stability of fH to 5 freeze thaw cycles. Triplicate samples from 5 healthy volunteers and the post-cryoprecipitate control were subjected to 5 freeze thaw cycles from -80°C to 37°C. with the exception of cycle 5 (where only duplicate assays were possible because of technical difficulties with the nephelometer). Cycle 0-fresh samples; cycle 1 - 7 days after blood draw, cycle 2 and 3 – 10days after blood draw; cycles 4 and 5 – 15 days after blood draw. *(NB unable to complete all cycles with individual 3 lithium heparin plasma due to lack of sample)*

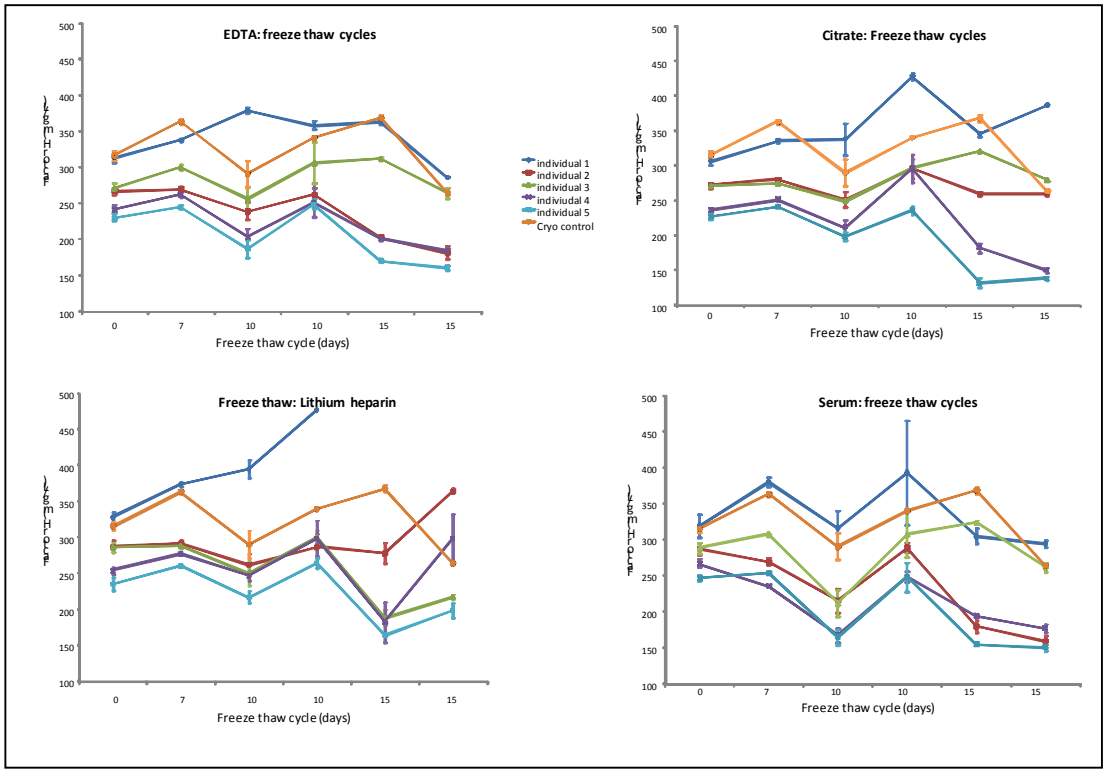
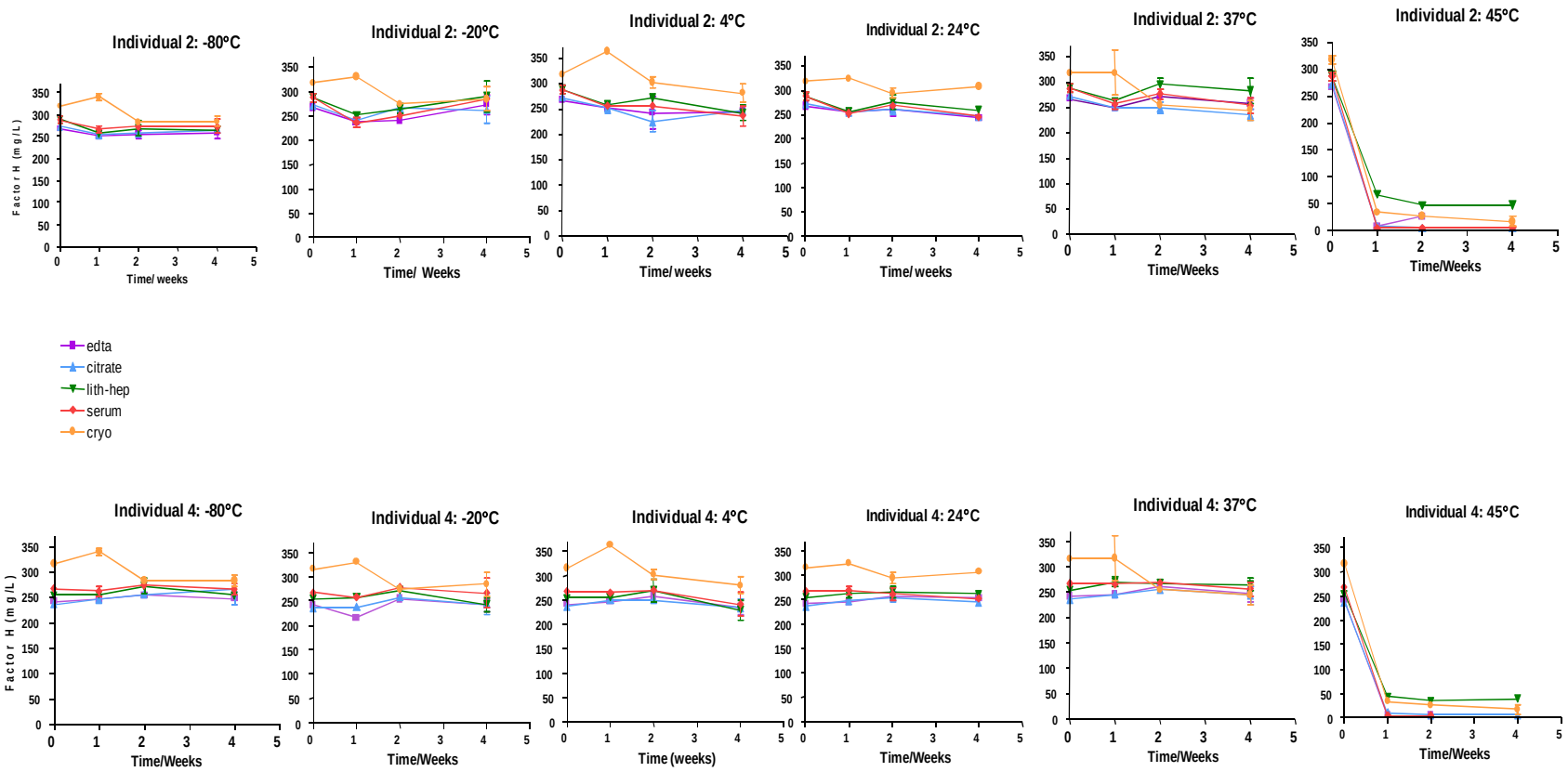


Figure 4.9 Stability of fH to temperatures. Samples from 2 individuals from around the 25th and 75th centiles of the distribution, and the post-cryo material were stored at -80°C, -20°C, 4°C, 24°C, 37°C and 45°C. Aliquots were assayed fresh and at weekly intervals for 4 weeks.



4.2.2.2 Variation and distribution of fH in healthy volunteers

Circulating fH followed a skewed distribution in 100 healthy volunteer samples, with a median of 348 mg/L and a range of 440 mg/L. Summary statistics of fH measures as carried out by immune-nephelometry are shown in **Table 4.3**. A similar distribution is seen for fH measured using RID (median 370.36 mg/ L) although the range measured is narrower (169.58 mg/L).

4.3 Discussion

By using reagents from a pre-existing commercial low throughput assay I have described the development of a newer high throughput assay for fH which is robust, reliable and simple to perform, and which uses standard equipment present in any healthcare associated biochemistry laboratory.

The new assay is more sensitive than the current RID assay, for values across the whole range of fH found in populations and, in particular, at the extremes of the distribution of fH values. Although degraded quickly at temperatures > 37°C, fH is remarkably stable over a number of weeks to ambient and low temperatures, and is resistant to multiple freeze-thaw cycles. Importantly there are no major differences in measurement of fH across the range of plasma (EDTA, citrate and lithium heparin) and serum. Collectively these properties of the analyte render it suitable for measurement in stored samples from population based and control studies in AMD and CHD. Moreover these properties mean that a test for blood fH could be transferred quickly to a clinical setting should this marker be found to be useful in predicting risk of AMD or CHD or the response to treatment.

Samples of fH depleted plasma were found to have a nominal concentration of fH, ~7 mg/L. This may be because the polyclonal sheep anti-fH antibody used in the assay may have some cross reactivity with other homologous circulating proteins such as FLH1 or one or more of the CFHRs. An alternative explanation is that the sample of fH depleted plasma still contains a small concentration of fH that was detected by the BNII assay. However, given the concentration range of fH in serum and plasma even if there were some cross reactivity this is likely to be small and constant, as determined by these experiments.

Other methods for measuring fH have been reported since this work was initiated (**Table 4.7**). However none have been shown to be as robust as the assay described here given the validity of the assays have not been as comprehensively determined, and stability of fH not been shown with other assays, as has been demonstrated here. In addition the assay described here is high throughput, compared to the ELISA based methods of other assays, and given the platform on which it has been established it can be rapidly moved to a clinical setting. Moreover the current assay offers an advantage to more traditional assays of fH using an erythrocyte lysis assay, which would be unable to detect subtle differences in fH concentration on a population scale. However one advantage that the latter does offer that it tests function of fH which is lacking in the current assay. Nevertheless, the previously described QTL mapping to the *CFH* gene has been associated with concentrations rather than activity. Another possible disadvantage maybe that the current assay is not able to distinguish between Y402 and 402H forms of the fH protein, although this information could be obtained by correlation of genetic and phenotypic data, should the polymorphism underlying this protein change be shown to be functional.

This method therefore allows fH concentration to be measured on a population scale, thereby allowing description of its molecular epidemiology, the genetic and non genetic determinants to test the association with both CHD and AMD, and to delineate wither the association of genetic variation in the CFH gene with AMD is mediated in whole or in part by differences in blood fH concentration.

Table 4.7 Tabular summary of the results of fH assay in previously reported studies Blank cells indicate where values have not been reported.

Source	Total number of individuals	Mean fH	Standard deviation	Median	Maximum	Minimum	Range
Esparza-Gordillo J et al., [134]	385 (21 pedigrees)	319.9	71.4		562	116	446
Scholl HP et al., AMD CONTROLS [205]	67			515			
Hakobyan S et al., AMD CONTROLS (Cardiff) [203]	63	233.24	56.65		349.27	135.54	213.73
Hakobyan S et al., AMD CONTROLS (Spanish) [203]	75	268.69	55.52		398.38	83.05	315.33

5 Molecular Epidemiology of circulating complement factor H

Given common variants in *CFH* are associated with AMD, and that a QTL was previously mapped to chromosome 1 in the RCA cluster, there is strong biological plausibility that fH concentrations may mediate the link between SNPs in this region and AMD risk. If so fH might be a potentially useful biomarker with utility for disease prediction and as a potential drug target. However very little is known about non-genetic correlates of circulating fH at a population level. Having established a high-throughput method for measuring fH applicable at a population level, I aim to describe in this Chapter, measurement of fH in a population based study which has a wide range of demographic, lifestyle, anthropometric measures and blood biomarkers already measured. This has allowed for the first time, the determination of the distribution of circulating fH in an adult white European population, and the cross-sectional correlations and associations of fH.

5.1 Materials and Methods

5.1.1 Population

The Fenland study is a MRC funded prospective, population based study, which recruited individuals born between 1950 and 1975 registered at general practices in the east Cambridgeshire area. It is an ongoing study which aims to recruit 10,000 individuals. A wide range of blood, metabolic, anthropometric and lifestyle data have been collected in individuals recruited thus far. I studied a subset of 1514 subjects where fasting serum was available to measure fH. The samples used for estimation of fH were stored at -80°C, and had not been through a previous freeze-thaw cycle. All participants of this study gave informed consent and protocols and collection procedures were approved by the relevant ethics committees.

5.1.2 fH measurement

Factor H was measured once in each sample using the newly developed nephelometric assay described in Chapter 4. Samples were thawed at room temperature, and centrifuged at 3000g for 5 minutes prior to loading on the BNII. A volume of 40 uL was used by the analyser to perform a measurement and concentration was estimated as described in Chapter 4 and

Appendix 1. A subset of 200 samples, chosen at random were re-measured to ensure repeatability of the assay within the sample set.

5.1.3 Statistical analysis

For normally distributed continuous variables, means and standard deviations were calculated. For non-normally distributed variables natural logarithms (log_e) were used to transform the variables where appropriate, and geometric means and approximate standard deviations were reported. If no normalising transformation was possible, medians and inter-quartile ranges were calculated. For categorical variables, proportions were calculated. Results were reported overall and stratified by gender.

The shape of the fH distribution was determined, and any appropriate transformations were carried out. The central tendency for both untransformed and transformed fH was determined, as well as appropriate metrics for degree of variability and dispersion. Univariate associations for cross-sectional associations were performed using Pearson correlation coefficient, correcting reported p values for multiple testing using the Bonferroni method. In order to further delineate, and visualise the shape of associations of fH with individual measures, the geometric mean of log_e fH, and its 95% confidence intervals, for each decile of the related marker were calculated and plotted, for males and females separately. All data-analysis described in this chapter was carried out using the statistical package Stata (Version 11, StataCorp LP College Station, TX, USA)

5.2 Results

Baseline characteristics overall and by gender are shown in **Table 5.1**

5.2.1 Distribution of fH

The distribution of fH concentration in the MRC Fenland study was found to be right skewed, and normalised by a natural logarithm transformation (log_e) (**Figure 5.1**). The median and the inter-quartile range for the untransformed distribution were 226.6 mg/L (111.3 mg/L), and the

transformed means and approximate standard deviation were 232.66 mg/L (74.52). These measures are summarised in **Table 5.2.**

Table 5.1 Baseline characteristics of the MRC Fenland study participants. Means or proportions are given for each variable, for the overall group as well as gender specific values. *Log transformed, ** 1/SQRT transformation

	Male			Female			Total sample		
	N	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
Age (at study entry)	882	44.85	7.27	1092	45.34	7.08	1974	45.12	7.17
Lifestyle measures									
		%			**%**			**%**	
Never smoked		49.94			52.89			51.57	
Ex-Smokers		29.51			30.30			29.95	
Current Smokers		12.37			10.84			11.52	
Never consumed alcohol		5.57			7.07			6.40	
Ex-drinkers		0.11			1.10			0.66	
Ever drinkers		0.57			0.73			0.66	
Current		81.93			75.94			78.62	
Alcohol units/ week	709	11.98	13.79	886	5.05	6.20	1595	8.13	10.85
Anthropometric measures									
Dexa-Total Bone Mass	660	3254.83	457.57	848	2550.85	362.76	1508	2837.98	528.17
Dexa-Total fat Mass	660	23056.57	9013.40	848	25531.06	9842.40	1508	24416.95	9548.13
Dexa-Total lean mass	660	58227.59	6829.63	848	40827.52	5187.0			
Body fat Percent (impedance measure)									
Waist hip ratio*	878	0.94	0.06	1090	0.83	0.06	1968	0.88	0.08
BMI*	881	27.48	3.89	1091	26.30	5.13	1972	26.82	4.68

Table 5.1 continued

	Male			Female			Total sample		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Cardiovascular Measures									
Mean SBP	874	128.58	14.11	1076	118.56	15.48	1950	123.05	15.69
Mean DBP	877	78.35	10.40	1082	73.87	9.86	1959	75.88	10.34
Resting heart rate	878	66.78	8.744	1087	63.83	8.49	1965	63.83	8.49
Indices of Glycaemic Control									
Fasting Glucose*	882	5.04	0.59	1078	4.73	0.51	1960	4.87	0.57
Glucose 120mins*	870	5.14	1.51	1061	5.04	1.42	1931	5.08	1.46
HbA1c*	880	5.45	0.44	1085	5.36	0.37	1965	5.40	0.41
NEFA*	666	286.71	131.81	848	340.18	160.03	1514	315.53	149.33
Leptin*	667	5.32	4.74	848	16.91	14.62	1515	10.16	10.64
Adiponectin*	667	3.89	1.68	848	6.80	3.18	1515	5.32	2.82
Insulin (pmol/L)*	666	44.08	25.71	848	35.46	20.79	1514	39.02	23.20
Major lipids and apolipoproteins									
Total Cholesterol*	881	5.38	1.04	1087	5.24	0.99	1968	5.30	1.01
Triglyceride*	881	1.28	0.74	1087	0.89	0.47	1968	1.05	0.61
HDL-C*	881	1.23	0.30	1087	1.56	0.40	1968	1.40	0.38
LDL-C	867	3.56	0.91	1082	3.30	0.88	1949	3.41	0.90
ApoE*	667	0.03	0.01	846	0.03	0.01	1513	0.03	0.01
Indices of renal function									
Urea*	881	5.21	1.21	1085	4.62	1.13	1966	4.87	1.20
Creatinine	879	84.68	13.63	1086	66.05	11.87	1965	74.38	15.71
Indices of liver function and bone metabolism									
Serum albumin	881	43.31	2.61	1086	41.53	2.77	1967	42.33	2.84
Bilirubin*	881	11.14	4.61	1085	8.84	3.76	1966	9.80	4.27
Alkaline Phosphatase*	880	76.34	20.04	1085	69.06	20.92	1965	72.24	20.93
Alanine Transaminase*	881	35.44	15.65	1085	22.06	8.99	1966	27.28	13.22

Table 5.1 continued

	Male			Female			Total sample		
	N	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
Gamma GT**	881	0.17	0.04	1085	0.21	0.04	1966	0.19	0.05
Calcium	880	2.29	0.08	1085	2.26	0.09	1965	2.27	0.09
Corrected Calcium	880	2.20	0.08	1085	2.20	0.08	1965	2.20	0.08
Markers of inflammation									
hs CRP*	660	1.40	1.67	842	1.46	1.87	1502	1.43	1.78
fH*	668	258.44	80.65	846	214.13	64.68	1514	232.66	74.52

Figure 5.1 Distribution of circulating fH in the MRC Fenland study and the effect of a number of transformations

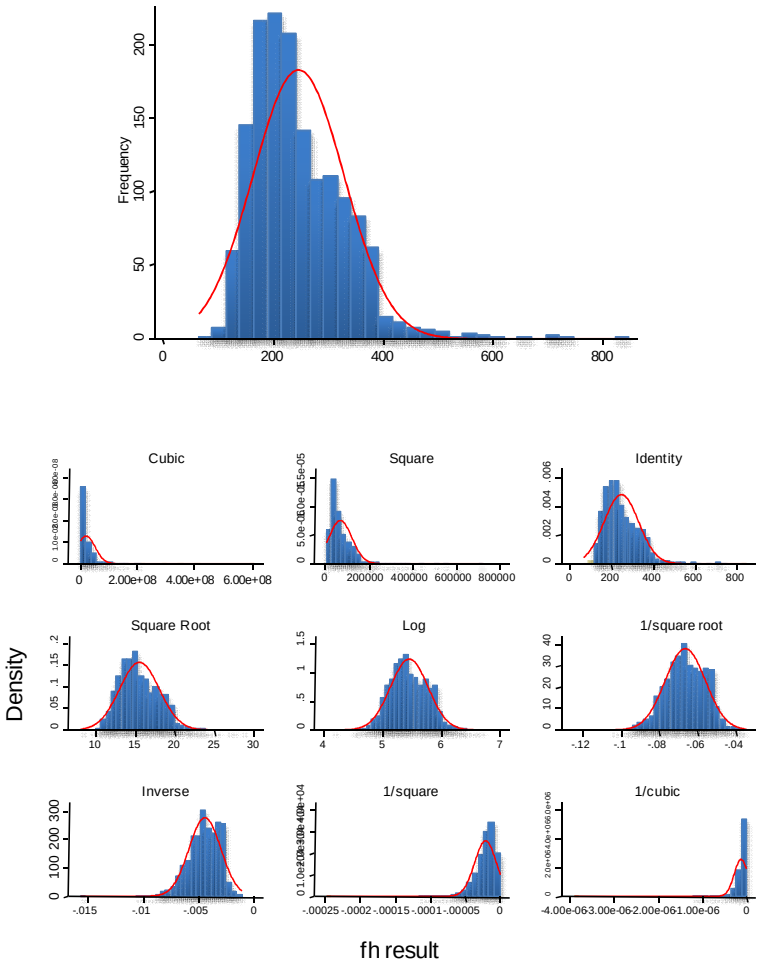


Table 5.2 Distribution of circulating fH in the MRC Fenland study

	Total Number	Minimun	Maximum	25th centile	50th centile	75th centile	Interquartile range	Mean	SD
fh	1514	63.48	847.55	184.77	226.56	296.06	111.29	245.15	83.25
Log fH	1514	63.48	847.55	184.77	226.56	296.06	111.29	232.66	74.52

Univariate associations of log transformed fH were observed with a number of the demographic, anthropometric measures, all of which are shown in **Table 5.3**. No clear trend was observed with age (correlation coefficient 0.01) or smoking (correlation coefficients=0.11) as has been

previously reported in a smaller (n~300) family based linkage study [134]. However there were a number of significant positive correlations with major blood lipids, including triglycerides (correlation coefficient=0.67) total cholesterol (correlation coefficient=0.33) and LDL-C (correlation coefficient=0.29). fH was inversely correlated with HDL-C (correlation coefficient=-0.4). Circulating fH was also correlated with indices of adiposity including BMI, (correlation coefficient=0.43), waist hip ratio (correlation coefficient=0.4). Significant associations were also observed with liver markers (including ALT, ALP and gamma GT), although no association of fH with indices of renal function. Log fH was in addition also significantly associated with the sensitive inflammation marker and acute phase reactant CRP (correlation coefficient =0.4).

Table 5.3 Unadjusted correlation coefficients of log transformed fH with a range of anthropometric measures and other biological measures in the MRC Fenland study. P values reports are adjusted for multiple testing using the Bonferroni method

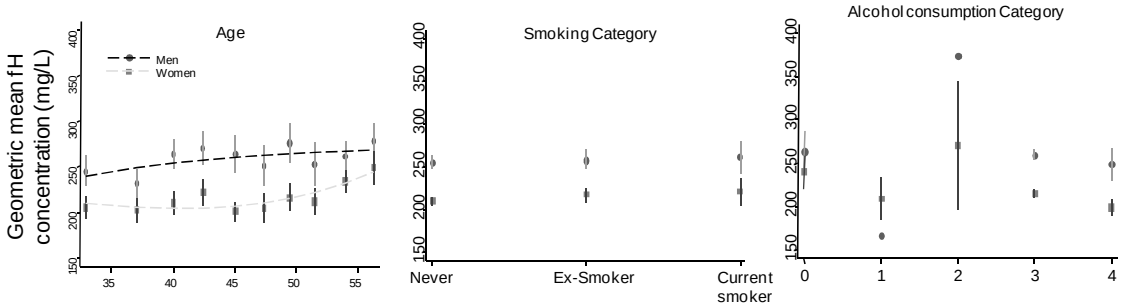
Variable	Correlation Coefficient	p value
Lifestyle measures		
Age	0.01	<0.001
Alcohol (units per day)	0.09	0.37
Smoking (cigarettes per day)	0.11	1
Anthropometric measures		
Log Dexa: total bone mass	0.24	<0.001
Log Dexa: total fat mass	0.4	<0.001
Dexa: total lean mass	0.31	<0.001
Body Fat Percent (Impedance)	0.13	0
Log Waist Hip Ratio	0.4	<0.001
Log BMI	0.43	<0.001
Cardiovascular Measures		
Mean SBP	0.27	<0.001
Mean DBP	0.32	<0.001
Resting Heart Rate	0.13	0
Indices of Glycaemic Control		
LogFasting Glucose	0.25	<0.001
Glucose 120 minutes	0.3	<0.001
HbA1C	0.3	<0.001
Log NEFA	0	1
Log Leptin	0.21	<0.001
Log Adiponectin	-0.41	<0.001
Insulin	0.51	<0.001
Major lipids and apolipoproteins		
Total cholesterol	0.33	<0.001
LDL cholesterol	0.29	<0.001
Log HDL Cholesterol	-0.4	<0.001
Log Triglycerides	0.67	<0.001
Log ApoE	0.53	<0.001
Indices of renal function		
Log Urea	0.09	0.06
Creatinine	0.23	<0.001
Indices of liver function and bone metabolism		
Albumin	0.01	1
Log Bilirubin	-0.05	1
Log Alkaline Phophatase	0.31	<0.001
Log Alananine Transaminase	0.34	<0.001
SQRT Gamma GT	-0.42	<0.001
Calcium	0.14	<0.001
Corrected Calcium	0.15	<0.001
Markers of inflammation		
Log hsCRP	0.4	<0.001

5.2.2 Cross-sectional associations of circulating fH with other variables measured in the MRC Fenland Study

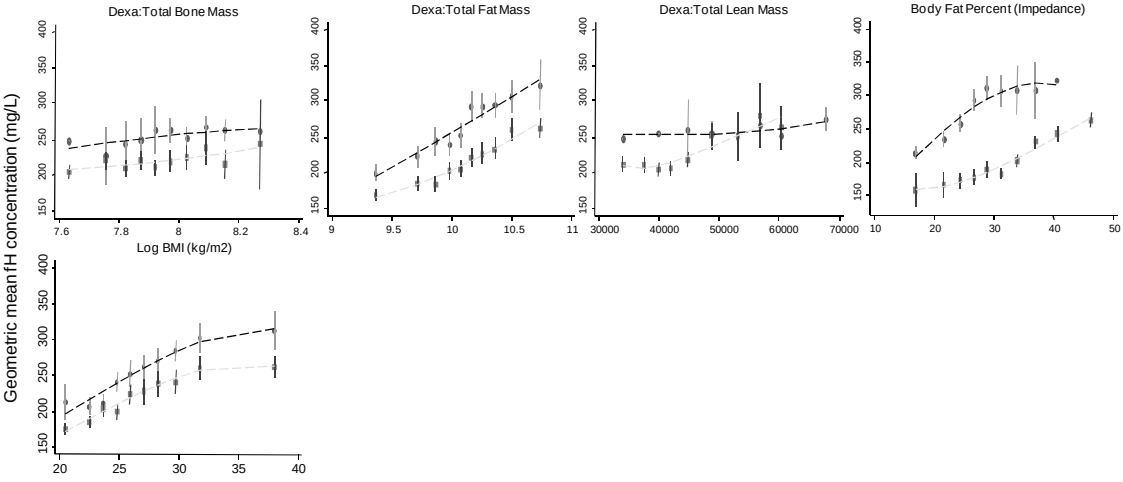
To further explore the relationship between fH and indices measured broadly in the MRC Fenland study, the geometric mean of log transformed fH (95%CI) was plotted against the mean value of each decile of the associated co-variate (**Figure 5.2**). Associations of fH with anthropometric measures reflecting fat mass (including Dexa total fat mass, impedance based measures of body fat percent and log BMI), lipid biomarkers including total cholesterol, triglyceride and apo-lipoprotein E were positive and approximately linear, while associations with adiponectin log HDL-C were negative and approximately linear. These were the strongest associations. Circulating fH was also associated with liver markers of ALT, ALP (positively) and bilirbun and albumin (negatively), although there was interestingly no association with renal markers of urea and creatinine. The difference in fH concentration between opposite deciles was greatest for triglycerides, BMI and HDL-C being as high as 150 mg/L which is approximately equal to 2 standard deviations of the fH distribution.

Figure 5.2 Cross-sectional associations of circulating fH in 1500 individuals from the MRC Fenland study. Mean log fH and its 95% CI was calculated for each decile of the associated variable and plotted to visualise the shape of the association. Plots are divided into panels with respect to which system they best reflect.

Panel A: demographic and lifestyle measures. For the alcohol variable 0= never drinkers,1= not current, 2=ex-drinker, 3=ever drinker 4=current drinker



Panel B: Anthropometric measures



Panel C: Cardiovascular Measures

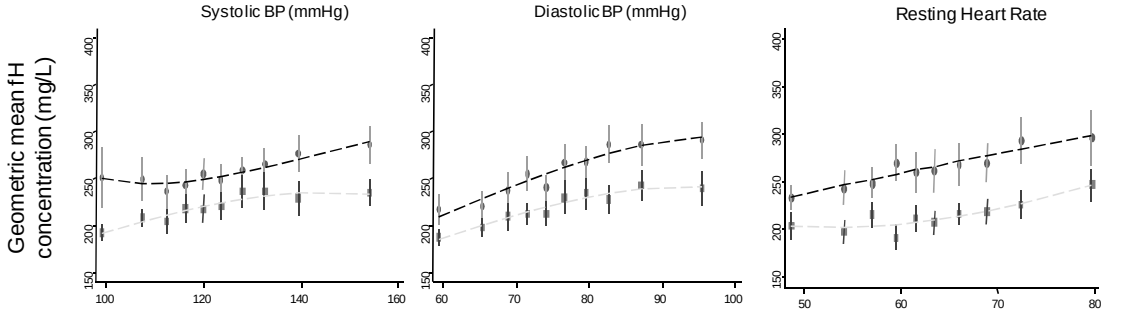
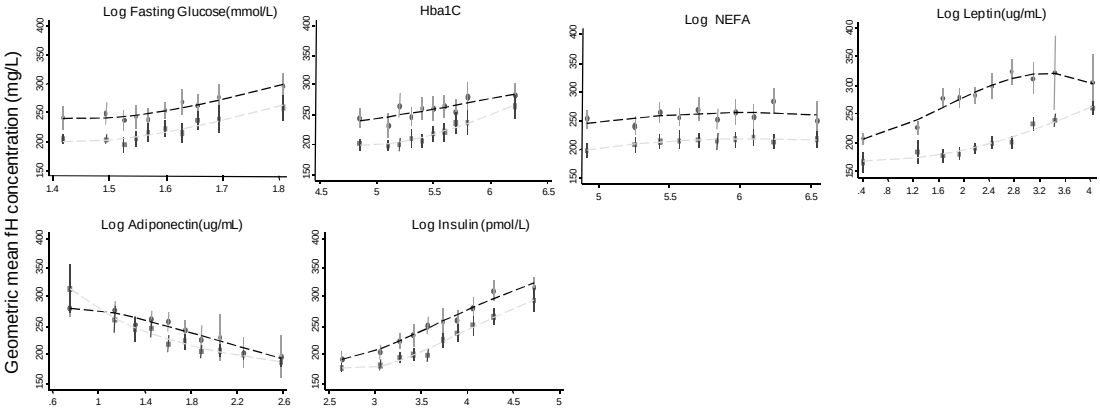
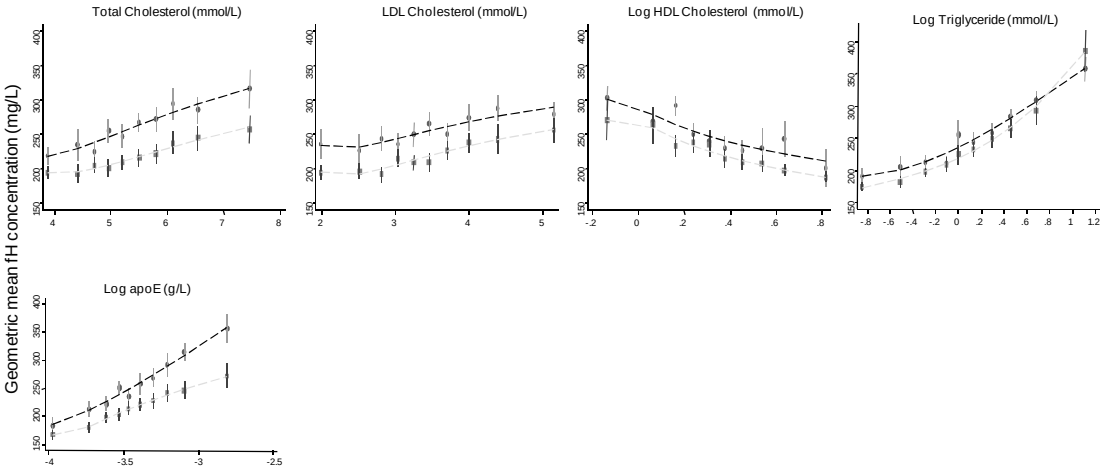


Figure 5.2 continued

Panel D: Indices of glycaemic control



Panel E: Major lipids and lipoproteins



Panel F: Indices of renal function

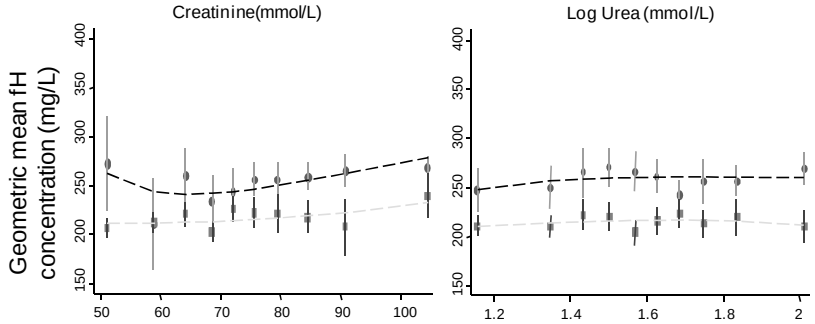
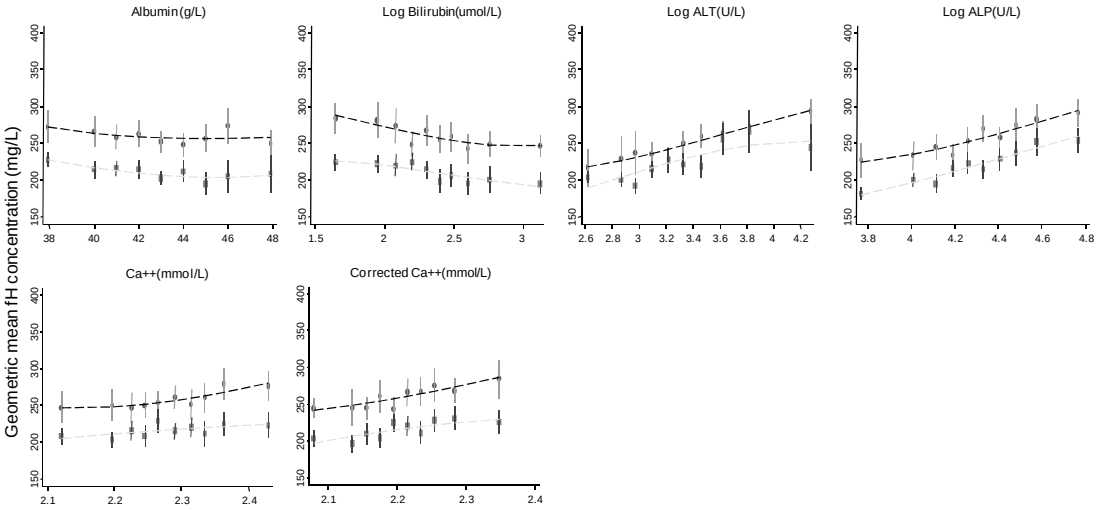
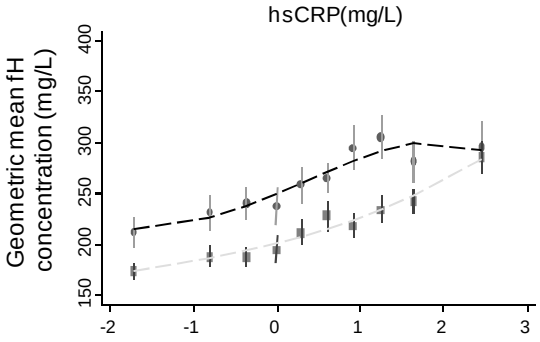


Figure 5.2 continued

Panel G: Indices of liver and bone metabolism



Panel H: Indices of inflammation



5.3 Discussion

In this Chapter I have described the first use of a newly developed high-throughput nephelometric assay for circulating fH on a population scale. I have also described, for the first time, the correlates and associations of circulating fH with age, gender, and a wide range of lifestyle, anthropometric markers and blood biomarkers related to the risk of late-onset disorders. Although a smaller scale family based study has been carried out as part of a linkage scan to identify the QTL for fH [134], which estimated that 60% of variation in circulating plasma

fH is determined genetically, but also suggested that there were important environmental determinants of fH such as age and smoking. However the small size of that study (n=358) and the number of covariates, limit conclusions about the associations of fH. More recently a further small study (n=398) evaluated adipose *CFH* gene expression in a subset of individuals as well as measuring both circulating fH and fB by in-house ELISA methods [206]. They demonstrated a positive association with BMI, waist circumference, triglycerides and inflammatory markers, and a negative association with insulin sensitivity and HDL-C, in keeping with our observations on a larger scale. Isolated adipocytes were demonstrated to have a low basal production of fH, and concentrations were higher in individuals with altered glucose tolerance and insulin resistance, with a demonstrable reduction of fH and fB levels with weight loss and treatment with the anti-diabetic drug rosiglitazone (a member of the thiazolidinediones class of drugs). However given the larger scale of the data described in this chapter and the number of co-variates that have been measured, the current description of the molecular epidemiology of fH the largest and most comprehensive to date.

The association of fH with measured blood lipid fractions is of particular interest given that recent proteomic analyses have demonstrated that circulating complement components complex with HDL-C [35]. However, fH was found to be inversely associated with HDL-C in the MRC Fenland study, which cannot simply be explained by binding complexes between HDL-C and fH which would be expected to produce a positive association. These associations may have implications for disease mechanism. There is observational evidence linking HDL-C and triglyceride to CHD risk, however the effect of these markers is less clear in AMD. However more direct evidence of a link between HDL-C with AMD in particular has emerged from denser coverage GWAS where variants in genes that underlie variation in circulating HDL-C are also associated with AMD [36, 37]. Further support for a relationship between lipid and lipid related markers with fH is described in Chapter 3, where 402H homozygote individuals were observed to have a significant association with circulating triglyceride.

Beyond the association of fH with particular subsets of markers, the results of measurement of fH on this scale are important in terms of relating what is known about the biology and pathological function of fH to its measurement in a healthy population based sample. fH is made

in the liver [133] and in MRC-Fenland, there is a correlation with other measures of synthetic function of the liver (albumin) and liver function in general (measures of ALP, ALT and bilirubin). The initial observation that fH was produced by the liver [133], also indicated that fH mRNA expression was independent of the acute phase marker serum amyloid A (SAA), indicating that fH is not an acute phase reactant, but rather produced at a steady state. Complete absence of the protein is associated with uncontrolled activation of the complement system, supporting its constant production. Half life measurements of factor H have come from studies where fH is infused into an individual deficient of fH secondary to a premature stop codon preventing adequate section of the protein. Infusion of purified fH into this individual thus allowed an estimation of a half life of about 6 days [207].

In addition to production by the liver, as already discussed, there is evidence of production of fH at local sites outside of the liver which may contribute to the overall plasma concentration or may represent local concentrations particular if fH at these sites in surface bound, which serves to inhibit complement activation in specific micro-environments. This may be one explanation of the lack of evidence of an association with renal function (creatinine and urea) in this study, despite the well known associations of fH deficiency in certain rare forms of renal disease.

There are some limitations to this analysis which are worthy of discussion. The analyses presented here are restricted largely to white Europeans and therefore may not be reflective of fH distributions in samples of other ethnicities. I have already demonstrated in Chapter 2, that there is a difference in the frequency of the rs1061170 SNP encoding Y402H, the putative functional SNP in AMD, between white Europeans and east Asian populations. As I have hypothesised that variation in the gene underlies differences in circulating fH, it would be important to determine distributions of fH in ethnically diverse populations where dense genotyping is also available, to allow for a comparison of how genotype and fH levels are related. Moreover, it is also important to replicate these findings in other European samples in order to confirm the distributions and cross-sectional observations. Other aspects that would be of interest are (i) the biological variability in fH concentration measured in serial samples from the same individual and (ii) to confirm or refute earlier findings that fH is not an acute phase reactant given its central role in regulation of complement activation.

The importance of the new assay for fH will now be in investigating differences in levels of fH in individuals with and without AMD (the finding of no effect of the genetic variants in *CFH* and CHD makes this association less likely), and to assess if it is indeed a putative biomarker for one of both of these diseases. This knowledge could impact not only on prediction of both disorders and early targeting of expensive therapies for AMD but may also inform on whether targeting fH or its downstream actions will be useful therapeutically.

6 Genetic determinants of circulating complement factor H using a genome wide association scan

In this chapter I aim to describe the experiments used to identify the genetic loci influencing the concentration of circulating fH using a genome wide association scan. Genome wide scans by their design provide a hypothesis free approach to identify possible genetic loci associated with a disease or trait. They are therefore a useful tool in determining genetic contributions of seemingly unrelated loci to a trait or disease, and thus have a clear advantage over previous candidate gene studies.

A previous genome wide, family-based linkage study has mapped a number of regions as being associated with circulating fH [134], although none of these reached the pre-specified threshold for linkage (LOD >3). However a biologically plausible QTL was mapped to the RCA cluster of chromosome 1 for fH (LOD score 2.03), which not only includes the *CFH* gene, but also the *CFHR* genes and other genes with complement regulatory activities (see introduction for details of genes in this region). Importantly the linkage peak also extended to other complement associated genes outside the *CFH/CFHR* gene family, including *CR1* (complement receptor 1) which has most recently been associated with Alzheimer's disease using a genome wide approach [11]. Other regions of the genome putatively involved with fH regulation were additionally identified including a locus on chromosome 2p21-24.1 (LOD score 2.15) and chromosome 15q22.1-24.1 (LOD score 2.00). In comparison to the standards of current genome wide scans, this was a relatively small study of 358 individuals belonging to 21 three- to five-generation extended pedigrees. Further evidence for the role of the RCA cluster in influencing circulating fH comes from a study in which variants in the *CFH* promoter have been associated with alteration of fH levels in a cohort of paediatric cases of meningitis [12]. Using newer technologies, with higher density coverage of variants across the genome, and with the added advantage of being able to infer non-genotyped variants against a densely genotyped reference panels (HapMap 3 (www. hapmap.ncbi.nlm.nih.gov) and the 1000 genomes project (www.1000genomes.org)), allowing for an even higher resolution of signal, allows confirmation and extension of these findings to further pin-point genetic regulation of circulating fH. Finding loci for fH in this way may provide an additional route for locating regions of the genome in other

diseases in which we know complement has been implicated, including AMD, aHUS, MPGNII, Alzheimer's disease and meningitis in which complement regulation has been implicated [11, 12, 131].

6.1 Materials and Methods

6.1.1 Population

The MRC Fenland study, which was described in more detail in Chapter 5, was used for this experiment. The 1514 individuals who had measures for serum fH in whom there was SNP data available which allowed me to carry out a genome wide scan for circulating fH. Of these individuals 1500 Caucasian individuals were used for this study in order to eliminate possible confounding due to population stratification. Identification of individuals of Caucasian decent were identified using the self-reported ethnicity data in the Fenland study.

6.1.2 Genotyping

Genotyping in MRC Fenland was carried out in the MRC Epidemiology Unit in Cambridge, using the Affymetrix 500k Human GeneChip®. Five hundred thousand single nucleotide polymorphisms were genotyped in each individual using this commercial chip. Genotyping by this method is based on the use of single oligonucleotide primers for amplification coupled to allelic discrimination on the DNA mirco-arrays. Markers on this chip have been chosen using a tagging SNP approach from the reference HapMap panel, in order to capture variation across the whole genome for SNPs with a minor allele frequency of greater than 1%. Genotyping was carried out in all individuals using one chip per person, with an automated genotyping algorithm used for calling BRLMM (Bayesian Robust Linear Model with Mahalanobis distance classifier http://media.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf). Where the algorithm was unable to call the genotype, allelic discrimination plots were examined visually, and manually called if appropriate. For genotypes that could not be called, SNPs were excluded from the subsequent analysis.

6.1.3 Measurement of fH

Measurement of fH was carried out using the nephelometric assay described in chapter 4, and values attained for the Fenland data set have already been extensively described in chapter 5. For purposes of analysis, values of fH were log transformed given its skewed distribution.

6.1.4 Statistical analysis

All statistical analyses presented in this chapter were carried out by Daniel Barnes (statistician at the MRC Epidemiology Unit, Strangeways Research Laboratory, Cambridge), using methods developed by Jian'an Luan (statistician at the MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge). Methods are summarised below and are also available in reference [208].

Quality control statistics for genotypes were carried out on genotypes successfully typed or called. Departure from HWE was calculated using a χ^2 test, which also serves as a measure of genotyping quality. A threshold of significance was set at a p value of less than $2x10^{-8}$, assuming a 5% significance level Bonferroni adjusted for ~2.5 million SNPs. Analyses were limited to self-reported Caucasians in the Fenland dataset, and the resulting quantile-quantile (QQ) plot was inspected visually to determine if there was evidence of significant population stratification, and a genomic control calculated at that stage. Missing autosomal genotypes were imputed that were present in HapMap Phase II/ NCBI build 36, but were not present in the genome-wide chip or did not pass direct genotyping quality control. For each sample, genotypes were imputed using data from the GWAS chip and HapMap phased genotype data from the 60 CEU HapMap founders. SNPs that were retained included those with a MAF>1%. Missing genotypes were imputed using the statistical software IMPUTE version 0.4(mathgen.stats.ox.ac.uk/impute/), which determines the probability distribution of missing genotypes conditioned on a set of known haplotypes and an estimated fine-scale recombination map. Imputation was based on 362,000 autosomal SNPs with a MAF of >1% that appear on HapMap. Imputed SNPs were excluded if they had an imputation information score of 0.4 a MAF <1% or were not in HWE equilibrium (p value $<1x10^{-6}$). Both directly typed and imputed SNPs were then tested for association with log transformed fH using an additive genetic model adjusting for age and sex. Log transformed fH was used as fH distribution is skewed and

normalised by log to base e transformation, as described in Chapter 5. Association analyses were carried out using SNPTEST version 1 [209]. The threshold p value was set at $2x10^{-8}$, to account for multiple testing.

6.2 Results

6.2.1 Quality Control

362,059 directly genotyped SNPs passed initial quality control, with call rates of >95% and that did not deviate from HWE using a threshold of $\geq 1x10^{-7}$. These included SNPs that were not initially called by automated calling algorithm and were called using visual inspection of allelic discrimination plots. Using information from these SNPs a further 2,080,823 SNPs were imputed, of which 1,891,264 were included in the final analysis after quality control. Therefore a total of 2,253,323 SNPs in total were used for the genome wide association analysis. Genomic control was not applied as was calculated to be 1.0057

6.2.2 Association results

No SNP, directly genotyped or imputed crossed a pre-specified genome wide statistical significance (p value of $2x10^{-8}$). However 40 SNPs achieved a p value of $\leq 1x10^{-5}$, of these, 19 were located in a 1 MB region including the *CFH-CFHR* gene region on chromosome 1 (**Figure 6.1, Figure 6.2, Table 6.1**). The strongest signal for log fH was rs644598, an inter-genic SNP whose closest reference gene is *CFHR3*, although SNPs in *CFH*, *CFHR1* and *CFHR3* share p values at the 10^{-6} level (**Table 6.1**). Of the other loci with a p value of $\leq 1x10^{-5}$, 15 SNPs were in or close to the *ELAVL2* gene, 2 in the *CRB1* gene, and 1 each in the *WNT5A*, *DCL1 GRB14* and *PHACTR1* genes. The Q-Q plot (**Figure 6.2**), confirms the likelihood of these associations and confirms that these are unlikely to be due to population sub-structure, with the genomic inflation factor being calculated as 1.0057.

Figure 6.1 Graphical representation (Manhattan plot) for genome wide scan of circulating log fH concentration in the MRC Fenland study. In this scan no markers reached genome wide significance although there are a number of peaks which are worthy of replication.

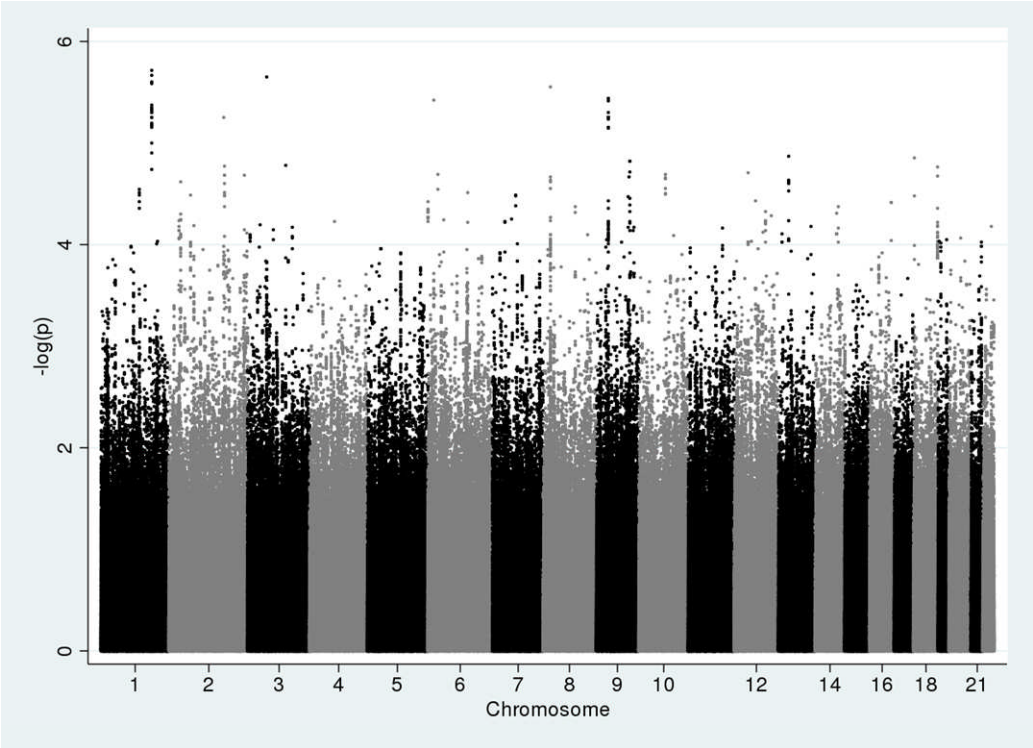


Figure 6.2 Quantile-quanitle plot of genome wide association scan of circulating log fH concentration in the MRC Fenland study. Genomic inflation factor λ=1.0057

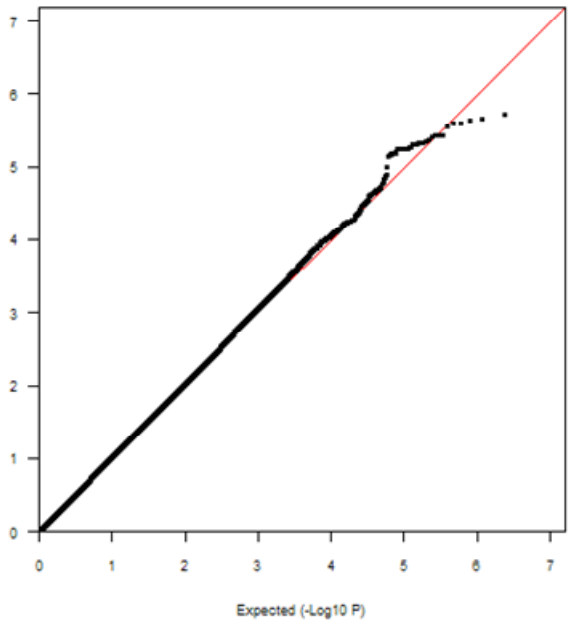


Table 6.1 Top association signals from a GWAS for circulating fH in the MRC Fenland population. Hits are organised from the smallest p value to largest, and the table has been restricted to just show signals below a p value of <0.000001) SNPs may occur within genes themselves or within intergenic regions which are not covered by the Affymetrix chip used or typed in the HapMap reference panel, and thus the closest and second closest reference genes are also given, as markers may highlight areas of association in genes which are in LD with these markers.

SNP	Chromosome	In reference gene	Reference gene 60 kb away from SNP	Closest reference gene	Distance from closest reference gene (kb)	Second closest reference gene	Distance from second reference gene (kb)	Beta coefficient (log fH)	Standard error	P value
rs644598	1		*CFH;CFHR3;CFHR1*	*CFHR3*	3705	*CFHR1*	22283	0.08	0.02	$1.92x10^{-06}$
rs446713	1		*CFHR3;CFHR1*	*CFHR1*	6984	*CFHR3*	19004	0.08	0.02	2.18x10-06
rs6445740	3		*WNT5A*	*WNT5A*	37573	*ERC2*	80165	0.06	0.01	2.26x10-06
rs377298	1	*CFHR3*	*CFH;CFHR3;CFHR1*	*CFHR3*	4331	*CFHR1*	30319	-0.07	0.02	2.49x10-06
rs449847	1		*CFH;CFHR3;CFHR1*	*CFHR3*	12781	*CFH*	14517	-0.08	0.02	2.60x10-06
rs11985642	8	*DLC1*	*DLC1*	*DLC1*	132255	*C8orf79*	352856	0.07	0.02	2.79x10-06
rs4131556	9			*ELAVL2*	253702	*TUSC1*	1596628	-0.06	0.01	3.65x10-06
rs10966201	9			*ELAVL2*	253125	*TUSC1*	1597205	-0.06	0.01	3.67x10-06
rs1223552	6	*PHACTR1*	*PHACTR1*	*PHACTR1*	113688	*TBC1D7*	131345	0.06	0.01	3.77x10-06
rs12379069	9			*ELAVL2*	249305	*TUSC1*	1601025	0.06	0.01	3.84x10-06
rs371075	1		*CFHR3;CFHR1*	*CFHR1*	12024	*CFHR3*	13964	0.07	0.02	4.28x10-06
rs426736	1	*CFHR3*	*CFH;CFHR3;CFHR1*	*CFHR3*	2455	*CFHR1*	28443	-0.07	0.02	4.46x10-06
rs385390	1		*CFH;CFHR3;CFHR1*	*CFHR3*	2	*CFH*	27296	-0.07	0.02	4.66x10-06

Table 6.1 continued

SNP	Chromosome	In reference gene	Reference gene 60 kb away from SNP	Closest reference gene	Distance from closest reference gene (kb)	Second closest reference gene	Distance from second reference gene (kb)	Beta coefficient (log fH)	Standard error	P value
rs1065489	1	*CFH*	*CFH;CFHR3*	*CFH*	6857	*CFHR3*	34155	-0.07	0.02	$4.71x10^{-06}$
rs11582939	1	*CFH*	*CFH;CFHR3*	*CFH*	6474	*CFHR3*	33772	-0.07	0.02	$4.71x10^{-06}$
rs11799380	1	*CFH*	*CFH;CFHR3*	*CFH*	8176	*CFHR3*	35474	-0.07	0.02	$5.00x10^{-06}$
rs3828022	1	*CRB1*	*CRB1*	*CRB1*	70674	*ZBTB41*	138409	-0.12	0.03	$5.02x10^{-06}$
rs10966195	9			*ELAVL2*	249019	*TUSC1*	1601311	-0.06	0.01	$5.04x10^{-06}$
rs13027837	2	*GRB14*	*GRB14*	*GRB14*	40638	*COBLL1*	103537	0.08	0.02	$5.57x10^{-06}$
rs4487866	9			*ELAVL2*	261273	*TUSC1*	1589057	-0.06	0.01	$5.57x10^{-06}$
rs10511739	9			*ELAVL2*	265279	*TUSC1*	1585051	0.06	0.01	$5.60x10^{-06}$
rs7033446	9			*ELAVL2*	283688	*TUSC1*	1566642	0.06	0.01	$5.60x10^{-06}$
rs10489456	1	*CFH*	*CFH;CFHR3*	*CFH*	16822	*CFHR3*	56414	0.07	0.02	$5.62x10^{-06}$
rs7032051	9			*ELAVL2*	279432	*TUSC1*	1570898	-0.06	0.01	$5.62x10^{-06}$
rs3753396	1	*CFH*	*CFH;CFHR3*	*CFH*	20889	*CFHR3*	48187	-0.07	0.02	$5.63x10^{-06}$
rs10645026	9			*ELAVL2*	272629	*TUSC1*	1577701	-0.06	0.01	$5.63x10^{-06}$
rs10966216	9			*ELAVL2*	279907	*TUSC1*	1570423	-0.06	0.01	$5.64x10^{-06}$
rs4445308	9			*ELAVL2*	269299	*TUSC1*	1581031	-0.06	0.01	$5.67x10^{-06}$
rs4072605	9			*ELAVL2*	261787	*TUSC1*	1588543	0.06	0.01	$5.82x10^{-06}$
rs70620	1	*CFH*	*CFH;CFHR3*	*CFH*	11634	*CFHR3*	38932	0.07	0.02	$6.42x10^{-06}$

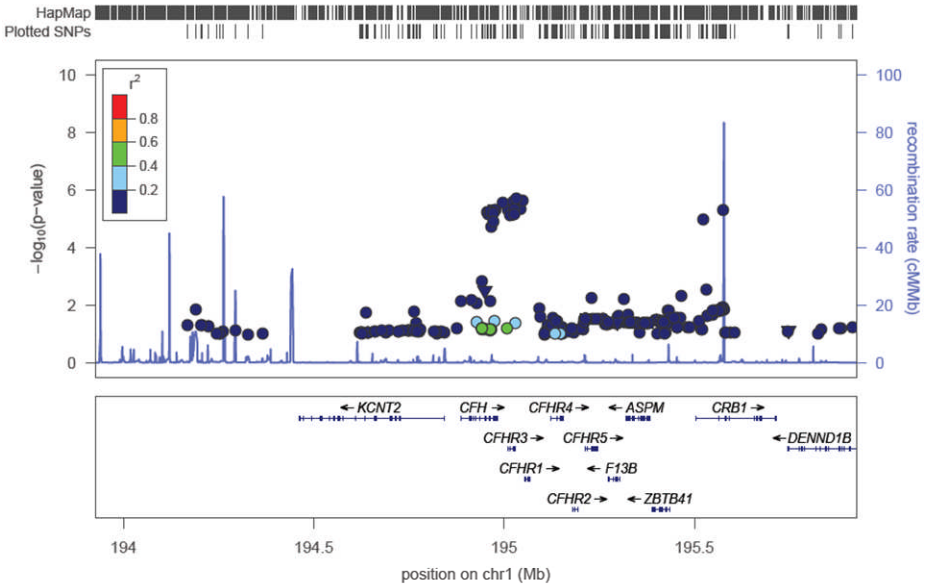
Table 6.1 continued

SNP	Chromosome	In reference gene	Reference gene 60 kb away from SNP	Closest reference gene	Distance from closest reference gene (kb)	Second closest reference gene	Distance from second reference gene (kb)	Beta coefficient (log fH)	Standard error	P value
rs378283	1	*CFHR3*	*CFH;CFHR3;CFHR1*	*CFHR3*	4586	*CFHR1*	30574	0.07	0.02	$6.47x10^{-06}$
rs11801630	1	*CFH*	*CFH;CFHR3*	*CFH*	21455	*CFHR3*	51781	-0.07	0.02	$6.48x10^{-06}$
rs389897	1	*CFHR3*	*CFH;CFHR3;CFHR1*	*CFHR3*	4442	*CFHR1*	30430	-0.07	0.02	$6.50x10^{-06}$
rs643781	1	*CFHR3*	*CFH;CFHR3;CFHR1*	*CFHR3*	2692	*CFHR1*	28680	-0.07	0.02	$6.72x10^{-06}$
rs370789	1	*CFHR3*	*CFH;CFHR3;CFHR1*	*CFHR3*	6632	*CFHR1*	32620	-0.07	0.02	$6.73x10^{-06}$
rs445207	1	*CFHR3*	*CFH;CFHR3;CFHR1*	*CFHR3*	5816	*CFH*	33114	0.07	0.02	$7.03x10^{-06}$
rs4246854	9			*ELAVL2*	256335	*TUSC1*	1593995	0.06	0.01	$7.05x10^{-06}$
rs4977910	9			*ELAVL2*	256513	*TUSC1*	1593817	0.06	0.01	$7.05x10^{-06}$
rs12380144	9			*ELAVL2*	257376	*TUSC1*	1592954	-0.06	0.01	$7.18x10^{-06}$
rs1009187	1	*CRB1*	*CRB1*	*CRB1*	19512	*ZBTB41*	87247	-0.13	0.03	$9.98x10^{-06}$

6.2.2.1 Chromosome 1

Most SNPs identified in *CFH/CFHR* region cluster around *CFH, CFHR1* and *CFHR3*, rather than extending to any of the downstream *CFHR* related genes (**Table 6.1**). This represents the locus that incorporates the deletion that has been shown to be protective for AMD [114], and it is feasible that a deletion may alter circulating levels of fH. Although centred around these 3 genes, there is an additional signal 1 Mb further downstream where 2 SNPs (rs3828022 and rs1009187; ($p= 5.02 \times 10^{-06}$ and 9.98×10^{-06} respectively) in the *CRB1* (Crumbs homologue 1) gene (**Figure 6.3**). *CRB1* encodes a homologue of a gene first identified in *Drosophila* encoding the crumbs protein which is one part of the molecular scaffold that controls the development of the eye, particularly in determination of polarity of photoreceptor cells in the fruit fly. In humans the homolog protein localises to the inner segment of photoreceptor cells. Mutations in this gene are associated with a number of inherited retinal dystrophies, although the most severe being Leber's congenital amaurosis, which represents a clinically severe form of retinitis pigmentosa. This region is not in LD with the *CFH/ CFHR* region, and has not been identified, as yet, as being associated with AMD, even using the more recently published denser genome wide platforms [36, 37]. However unlike these GWAS which do identify other complement related genes (namely *CFI, CFB* and *C3*) as being associated with AMD, no other complement related genes outside of the *CFH/ CFHR* locus were identified by the recent GWAS as being associated with circulating fH levels. Moreover no other inflammatory pathway genes were associated with fH levels.

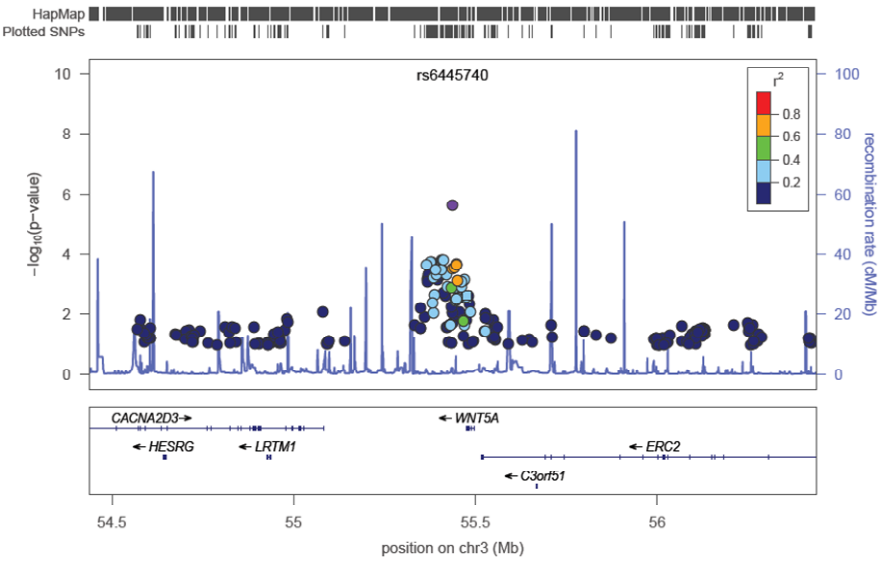
Figure 6.3 Regional association plot for SNPs in the *CFH* gene cluster for circulating fH concentration. Most significant association is shown in purple and the colour of the remaining SNP associations is based on the LD between these SNPs and the most associated SNP. The number of SNPs typed in HapMap are shown in the topmost track, with the number of SNPs directly genotyped/ imputed displayed in the track labelled plotted SNPs.



6.2.2.2 Chromosome 3

Of the loci that are outside of the *CFH/ CFHR* family and chromosome 1, the top signals of association in order, included *WNT5A* (wingless-type MMTV integration site family, member 5A, **Figure 6.4**), which belongs to the WNT gene family of structurally related genes, encoding secreted signalling proteins. *WNT5A* has been identified in *Xenopus* to bind to the human frizzled-5(hFz5) receptor, which mediates induction of cell polarity and migration, which may also be the mechanism that mediates its role in tumour-genesis, although this is less well understood (reviewed in [210]). The second closest gene to this marker is *ERC2* (ELKS/RAB6 interacting/CAST family member) which is thought to be involved in the organisation of the cellular cyto-matrix at active zones of nerve terminals which are involved in the regulation of neurotransmitter release [211]. It has most recently been found to be associated with behaviour associated with smoking cessation, in genome wide studies [212].

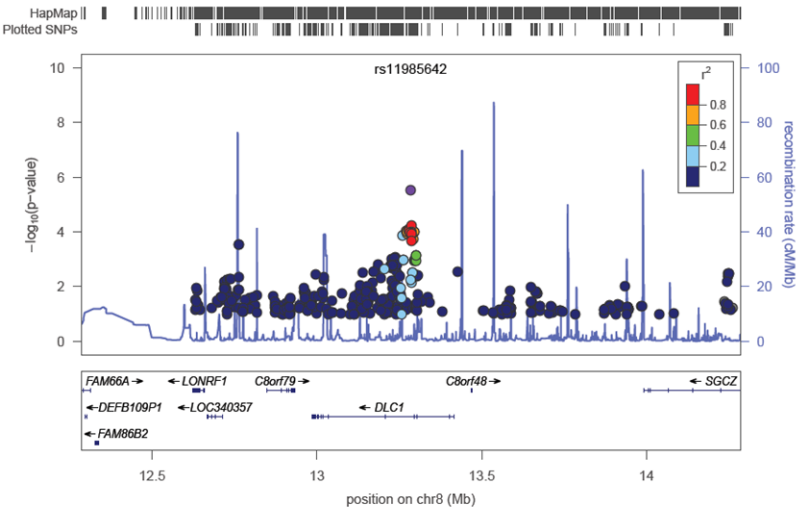
Figure 6.4 Regional association plot of signal in and surrounding WNT5a for circulating fH concentration.



6.2.2.3 Chromosome 8

The next most significant marker was in the gene *DLC1* (deleted in liver cancer) on chromosome 8, (**Figure 6.5**), encoding a GTPase-activating protein, and a member of the rho-GAP family, which is important in regulation of small GTP binding proteins, essential in cellular signalling pathways and regulating in particular processes which are involved in cytoskeletal changes[213]. In humans the most common disease association for *DLC1* is cancer, having been identified as a tumour suppressor gene, and is implicated in a number of common cancers (reviewed in [214] being thought to mediate cancer growth by alteration in cytoskeletal function in cell migration.

Figure 6.5 Regional association plot for *DLC1* and circulating fH concentration.



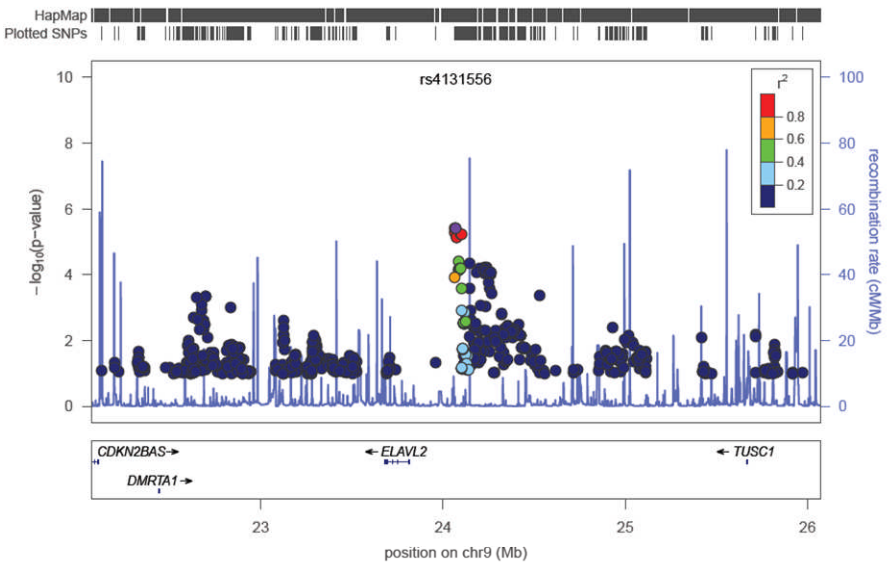
Although both signals- *WNT5A* and *DLC1* were single SNPs at this p value (10^{-6}), they are unlikely to be false positive findings given the signal of association extends to other SNPs with higher p values (**Figure 6.1**).

6.2.2.4 Chromosome 9

Similarly markers in an intergenic region on chromosome 9p were not only well represented at the p<0.00001 level but also extend to higher p values. (**Figure 6.1**). Chromosome 9p has, like *DLC1*, been implicated in the development of a number of tumours, housing a number of putative tumour suppressor genes. The region identified in this genome scan is downstream of the inter-genic region on chromosome 9p that has previously been associated with CHD [215, 216] and T2DM [217, 218], the signals for which themselves are in inter-genic regions, although closest to cell cycle regulator genes *CDKN2A and CDKN2B* . In the current genome scan, the closest gene to the most significant markers in the inter-genic region was *ELAVL2* ((embryonic lethal, abnormal vision, *Drosophila)*-like 2), (**Figure 6.6**). The encoded protein is essential to regulation of transcription through RNA binding, playing a pivotal role in development of the fruit fly nervous system. The human homolog *Hel-N1* [219] is a member of family of proteins with similar functions ascribed to it as in the fruit fly, that is mainly in post-transcriptional processing. Increased levels *ELAVL2* mRNA has been found in individuals with small cell lung cancer [220], although it has been also implicated in other cancer types. The next closest gene to this region is *TUSC 1*

(tumour suppressor candidate 1) which has also been implicated in lung tumourogenesis [221] (**Figure 6.6**), although from the regional association plot, **Figure 6.6**, it is clear that the signal is largely due to SNPs closest to *ELAVL2*, with *TUSC1* being further downstream.

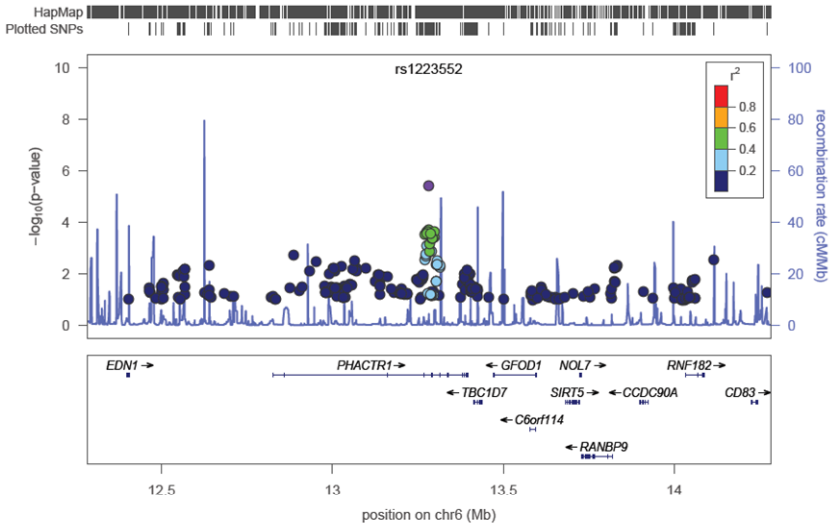
Figure 6.6 Regional association plot for locus on Chromosome 9 and circulating fH concentration. Closest reference genes being *ELAVL2* and further downstream *TUSC1*.



6.2.2.5 Chromosome 6

Of the other signals that only contain one SNP at this level of significance, although other SNPs at higher p values follow, is *PHACTR1* (phosphate and actin regulator 1), on chromosome 6, which belongs to a family of 4 structurally related proteins thought to function as protein phosphate 1 and actin regulators [222] with functionally diverse roles within the brain. Of note markers in this locus have also been identified in a GWAS of early onset MI [223]. The closest next gene to *PHACTR1* as well as markers in intergenic regions with p values of <0.0001/ <0.0001, is *TBC1D7* (TBC domain family member 7) (**Figure 6.7**). This gene belongs to family of proteins that are thought to have a role in regulating cell growth and differentiation, and although have not come up as hits in GWAS, proteins encoded by this gene are found to be expressed in lung cancers using cDNA microarray experiments, where they are thought to confer growth promoting activity of the tumour [224].

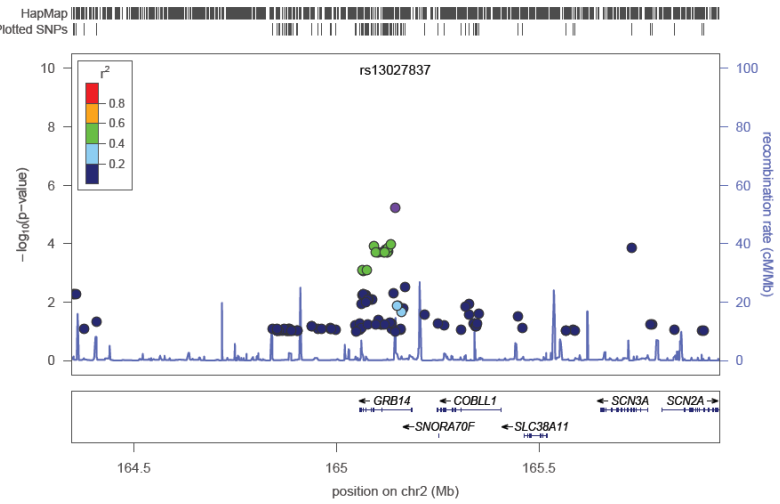
Figure 6.7 Regional association scan for PHACTR1 and signal on chromosome 6 and circulating fH concentration.



6.2.2.6 Chromosome 2

The final signal at this p-value comes from *Grb14* (growth factor receptor-bound protein 14), on chromosome 2 (**Figure 6.8**). This protein is a member of a family of structurally related multi domain adapter proteins, which, as a family, have been shown to have diverse cellular functions, and *Grb14*, in particular has been implicated in regulation of insulin receptor signalling [225]. The next closest gene to *Grb14*, is *COBLL1*, which is a protein encoding gene thus far implicated in cancer and thought to be of use in predicting survival of patients undergoing surgery for the tumour of the pleural lining: mesothelioma [226].

Figure 6.8 Regional association plot for signal in chromosome 2 and circulating fH concentration.



6.3 Discussion

The findings described in this chapter are consistent with the hypotheses described at the outset in that SNPs in the vicinity of *CFH* gene and extending the RCA cluster on chromosome 1q encompassing *CFHR1-5* acting in *cis* contribute to the variation in circulating fH. The associations also provide further confirmation of the validity of the newly developed nephelometric assay for fH described in chapter 4 whose development has allowed the conduct of a GWAS of the requisite scale.

An important question which arises is whether the SNPs in this region associated with fH overlap with those associated with the risk of AMD. An overlap of this type would provide strong evidence that the associations of SNPs in this region with AMD risk are explained in whole or part by alterations in circulating fH. This question is discussed in greater detail in the next chapter.

SNPs at a number other loci also showed associations with fH in this GWAS. These are also of interest as they may provide insight into pathways involved in the determination of fH concentration in the steady state. The association signals are all in regions which appear to play some role in the regulation of cellular cytoskeletal structures and signalling. In general, proteins encoded by the genes in which these signals are from, or the gene closest to these signals appear to play a role in human cancers. The exception to this is the *CRB1* locus, which is known

to be associated with an inherited eye disease, although this signal has not been reported in either the first generation of GWAS for AMD, or indeed the more recent GWAS which have employed denser platforms with additional imputation to further refine signals. In these recent AMD studies the notable successes were other regions of the complement cascade, including the previously identified SNP in *C3*, SNPs in *CFI* and *CFB.* Apart from these, these two reports largely focussed on SNPs in *CETP*, which contributes to the determination of circulating HDL-C and *LIPC* which is associated with circulating triglycerides. In Chapter 5, determination of the molecular epidemiology, and non-genetic correlates of fH highlighted numerous lipid markers as being associated with circulating fH levels, moreover in Chapter 3, the rs1061170 genotype was also associated with circulating triglyceride levels. Findings from this scan however do not provide evidence of an association with any of the 90 or so loci known to be associated with the major circulating lipids [227], which we may take to interpret that although fH is correlated with circulating lipid associated particles, which may themselves be a risk for AMD, none of the lipid associated genes contribute to determination of circulating concentrations of fH. The mechanism of their association is therefore more likely to be at the protein level, which has already been suggested by recent shot-gun proteomic studies, discussed previously [35].

The genome scan approach is useful in elucidating how circulating fH may be controlled. However, it does not tell us what the mechanisms that govern the action of fH, which could be important for disease mechanisms, not only in AMD but for other diseases with which fH protein function and or concentration is thought to be altered. These pathways must still be elucidated by mechanistic studies, using GWAS as a guiding tool. In this scan, the genetic determinants of circulating fH were from genes that would not obviously be associated with inflammation, and perhaps the interpretation is more challenging. This is in contrast to the recent finding of variants in *CFH* being associated in a genome scan with the risk of meningitis [12]. This observation is explained with greater ease as bacteria are known to use binding to fH as a way of avoiding detection by the immune system, and thereby causing disease.

Other limitations of the genome wide approach are the absence of rare variants, which are not well represented by whole genome arrays. Most variation will be common, and therefore expected to be associated with a modest effect size (the exception of course being GWAS in

AMD). However the contribution of rare variation has not really been elucidated to common traits and disease, and may be worthy to capture as it is these variants which may offer potential therapeutic targets, which will not only be useful for individuals with a given trait/ disease, but may be useful in more general terms. The obvious example of this strategy is the use of lipid lowering drugs which were targeted at those with familial hypercholesterolaemia or those at high absolute CVD risk, but now that they are off patent, thus cheaper, and the benefits of lowering cholesterol in the general population are known, statins are used more widely. The detection of rare variants determining traits and disease will begin to emerge with sequencing projects, both small and large scale- the 1000 genomes project being an example of such a resource. With sequence information becoming available across a wide range of ethnic groups worldwide will give the area more of a sense of the genetic regulation of traits and disease, the value of rare variants and the contribution of private mutations.

7 High resolution association analysis of the SNPs in the RCA cluster, fH concentration and AMD risk

SNPs in the RCA cluster are associated with AMD risk. Although much of the focus has been on rs1061170 encoding Y402H, many other SNPs in this region are associated with AMD. The region of association encompasses SNPs spanning ~1Mb between *KCNT2* and *ZBTB41* (position 194844122 to 195389437) including complement factor H and 5 other complement factor H related genes [36, 37]. In chapters 6, I identified SNPs within the same region as being associated with circulating fH. A recent GWAS also identified SNPs in the vicinity of *CFH* that were associated with risk of invasive meningococcal disease [12]. However there are several unanswered questions regarding these associations which are addressed by work in this chapter:

1. Which SNPs in the RCA gene cluster are associated with a higher risk of AMD, and is the association explained by rs1061170 encoding Y402H alone?

2. To what extent do SNPs associated with fH concentration (identified by GWAS in chapter 6) overlap with the SNPs associated with AMD risk

3. Is fH concentration itself associated with risk of AMD

Work in this chapter had the aim of 1) applying tagging SNPs as an efficient means of capturing genetic variation in the CFH/ CFHR gene cluster; 2) imputing unobserved genotypes across the whole 1Mb region in order to provide a more high resolution map of genotypes in this region using a fine scale recombination map and densely genotyped references panels. Using these two strategies I was able to identify if variants other than rs1061170 were associated with AMD risk and circulating fH levels. A third and separate aim of this chapter was to establish the relationship between circulating fH levels and AMD risk.

7.1 Methods

7.1.1 Studies

The analyses for single SNP associations were based on two studies of AMD (in which measurement of fH concentration had also been made). The first was the EUREYE study which was described previously in Chapter 2. Briefly this was a cross-sectional study investigating the

prevalence and risk factors for AMD in 7 European countries. AMD was diagnosed based on fundal photographs which were graded centrally using standardised protocols based on the ICARMS grading scheme [94]. The second study was the Cambridge Eye Study, a case-control study of unrelated individuals, over the age of 50, ascertained through hospital based ophthalmic clinics, general practices, optometrists and charitable societies for people with visual impairments in the East Anglian region. Where willing, spouses of identified cases were examined, and if unaffected were used as controls and, if affected included as cases. Cases and controls were examined by an ophthalmologist and all individuals completed a health and lifestyle questionnaire modified from that used for the European Prospective Investigation of Cancer (EPIC). All individuals had colour stereoscopic fundus photography of the macular region with grading according to the ICARMS classification [94]. Grading took place at Moorfields Eye Hospital by two independent ophthalmologists, with any discrepancy in grading being referred to a third ophthalmologist for arbitration using additional information from medical records and any previous fundal photographs and flourescin angiography where available. Individuals were included as cases if they were confirmed to have GA or CNV in one or both eyes. Individuals with early AMD were not included in the Cambridge study. Controls were similarly examined and included if they had a normal macula or non-extensive small or intermediate drusen and minimal hypo- or hyper-pigmentation. Individuals with inflammatory or other retino-vasuclar diseases were excluded.

7.1.2 SNP selection

For the EUREYE study SNPs were identified in the region encompassing *CFH* to *CFHR5* that maximised the variation across this region in European individuals, also known as a tagging SNP approach (carried out by Dr Tina Shah, UCL). The approach utilises prior knowledge of the linkage disequilibrium between SNPs catalogued by the Human HapMap. Two statistics can be used to determine the association between SNPs, the first is D' and the second is r^2, the latter of which is a correlation co-efficient, and calculated using D', and is the more stringent of the two, and therefore used in preference. Thresholds used in the selection of tagging SNPs were an r^2 of 0.8 and a minor allele frequency (MAF) of >5%. Where possible the directly typed SNP panel was enriched for non-synonymous SNPs as these may encode functional variation in the protein. Tagging SNPs were further supplemented by SNPs that had been reported in the literature as

being associated with AMD, where it was felt that these SNPs should be typed directly in order to preserve consistency with previously reported findings. These included the rs1061170 variant as well as the variant identified to tag the CNV in this region [228]. Tagging SNPs identified using this approach were typed in the EUREYE case-control subset.

The Cambridge AMD study had previously genotyped a number of SNPs in this region, which had been chosen and typed *a priori* by Professor Alan Wright (University of Edinburgh). SNPs were chosen based on evidence in Professor Wright's lab of a potential association with fH level. The rs1061170 SNP in the Cambridge AMD study was typed and reported independently [154].

7.1.3 Measurement of fH

Circulating fH was measured in plasma samples from EUREYE (drawn in EDTA) and the Cambridge study (drawn in Lithium Heparin) using the newly established nephelometric method as described in Chapter 4. Previous analysis has indicated that the values of fH are equivalent in plasma samples obtained with different anti-coagulants (Chapter 4).

7.1.4 Genotyping

Genotyping in the EUREYE study was carried out as described in Chapter 2, using KASPar technology by KBiosciences. The rs1061170 SNP in the Cambridge AMD study was typed by Dr Tiina Sepp using the ABI PRISM SNaPshot ddNTP Primer Extension Kit. Genotypes were called using an automated allelic discrimination method using a 3100 Genetic Analyses (Applied Biosystems). The remainder of the SNPs in the Cambridge AMD study were typed in Professor Wright's laboratory Taqman.

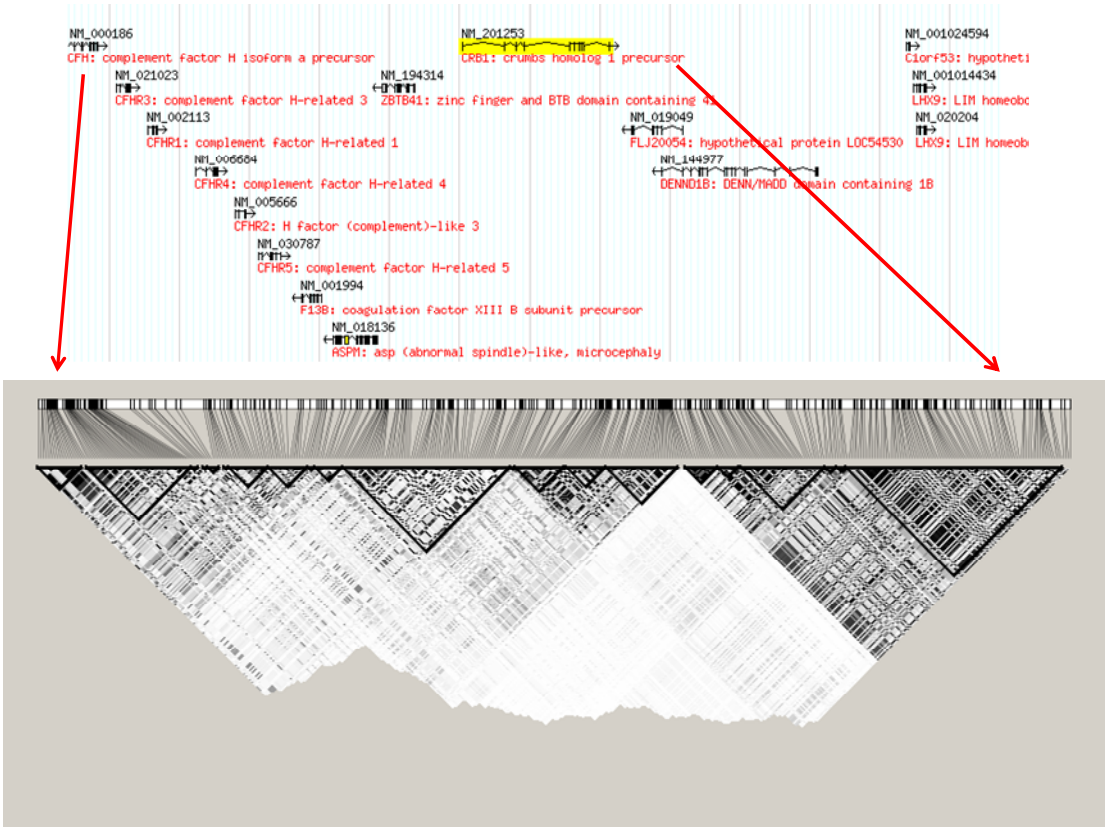
7.1.5 Imputation of untyped SNPs

Missing genotypes were imputed in the EUREYE study alone where there was a wider range of typed SNPs across which imputation could be performed. Missing genotypes were imputed against a European reference panel taken from the pilot release of the 1000 Genomes Project and the HapMap project phase 3. Both reference panels which contain two different sets of SNPs were incorporated into one imputation analysis using IMPUTE2 [229]. The 1000 Genome

reference panel provides a wide coverage of SNPs, in that it contains many more SNPs than available on the HapMap catalogue, although the latter provides deeper coverage of typed SNPs as it contains a greater sampling of chromosomes from each represented population on the catalogue. Combining these panels in a single imputation analysis therefore results in an extensive coverage of the genome (using 1000 Genomes) and increased accuracy at a subset of SNPs (using HapMap 3). Reference panels from chromosome 1 from HapMap and 1000 genomes were used, together with the recombination map from HapMap as provided on the IMPUTE2 website (www.mathgen.stats.ox.ac.uk/impute/impute_v2). Genotypes were imputed for a 1Mb region extending from *CFH* to *CRB1,* which encompassed signals from both the fH GWAS described in Chapter 6 and the recently published AMD GWAS (**Figure 7.1**).

Quality control filters were applied at the stage of association testing of the imputed genotypes, as software used for this stage (SNPTest version 2) takes into account genotype uncertainty. Filters that are commonly applied are first exclusion of SNPs with MAF <1% and second filtering is done on the basis of a metric which measures the degree of uncertainty around the imputed genotype. In IMPUTE2 and SNPTEST v2 this metric is commonly referred to as “info”, which usually takes a value from 0 to 1, and is a ratio of the observed to the complete information available. There is no consensus on what value of a cut off should be used [230], although a commonly used cut-off that is applied is 0.3, and this was used here. No cut off for MAF was used, given imputing against 1000 Genomes may reveal some important hits which are rare (rare variants are often thought to account for some of the dark matter of the genome [231]), although an important caveat to this is that identification of any rare variants by such a method would need to be replicated in another data set so as to ensure any finding is true rather than a false positive association.

Figure 7.1 Imputation of SNPs in the *CFH* gene cluster. SNPs across the CFH/ CFHR locus extending down to the CRB locus were imputed, based on the genome wide findings described in Chapter 5, where the signal for fH extends to CRB, the LD plot (based on r^2 values) is also shown, and were taken from HapMap.



7.1.6 Statistical analysis

7.1.6.1 Single SNP analysis

For both studies, genotyping call rates were calculated, and a threshold of 90% was set for inclusion of SNPs to be analysed. Departure from HWE was tested in controls only using a χ^2 test. SNPs that did not meet these quality control criteria in either study were excluded from further analyses. Association analyses were carried out in two stages: the first was based on the genotyped SNPs in both studies, and the second used imputed genotypes in the EUREYE dataset alone. For the first stage of analysis, using both data sets, association analyses were carried out in three stages:

(i) Genotype-AMD association

(ii) Genotype-circulating-fH association

(iii) Association of fH with AMD

Analyses were carried out separately for Cambridge and EUREYE, and consistency was checked across both studies.

7.1.6.1.1 Genotype AMD association

For association analyses each SNP from both EUREYE and Cambridge AMD case-control data sets that satisfied the QC criteria were tested for association with AMD. Cases in the Cambridge AMD study were all late AMD, although information on both early and late AMD cases was available in EUREYE, however results presented in this section are limited to late AMD, given this is the clinically more significant form of the disease and to ensure comparability of results across the two studies. SNPs were re-coded as 0, representing homozygote individuals for the common allele, 1 representing heterozygotes and 2 representing homozygotes for the minor allele. Odds ratios with 95% CI's for association of SNPs with late AMD were calculated using logistic regression on an additive (per allele) scale where the reference is the common rather than the non-risk allele. Adjustments were made for multiple testing using the Bonferroni method.

7.1.6.1.2 Genotype-fH association

SNPs were then tested for association with circulating fH level on an additive scale, using linear regression, reporting beta coefficients and 95% CI's. Coding of SNPs was consistent with that described above, again correcting for multiple testing using the Bonferroni correction.

For SNPs where an association with AMD has been previously reported the genotype and fH association was probed in a little more detail. Box plots indicating median, 25^{th} and 75^{th} centile values of fH levels were plotted for each SNP according to genotype group and association tested using ANOVA.

7.1.6.1.3 Association of fH concentration with AMD risk

Association of fH levels and AMD was performed using logistic regression models, using standardised log fH as a predictor variable, so as to express the odds ratio per one standard deviation of a change in log fH. Consistent with the analyses described above unadjusted analyses are presented.

SNPs that were significant in the above analyses were then entered into a forward fitting logistic (for late AMD as an outcome) or linear (for fH) stepwise model, with a p value of 0.05 set as the cut off for entry into the model, to test the statistical independence of association signals.

7.1.6.2 Imputation analysis

For the second stage of analysis, EUREYE alone was used to infer missing genotypes across a 1 Mb region (**Figure 7.1**) using reference panels from HapMap and 1000 Genomes, as described. Associations for this region were then tested using the SNPTEST Version 2 software (www.stats.ox.ac.uk/~marchini/software/gwas/snptest.html), using a frequentist, additive model, where alleles are automatically coded 0 and 1, and thus and additive test estimates the increase in log-odds that can be attributed to the allele coded 1. Where a model cannot be fitted to the data the p-value is set to -1.Genotype uncertainty was accounted for by using the “method – expected” option, using expected genotype counts (also known as genotype dosages). Filtering for uncertainly around genotypes was carried out using a threshold value of 0.3, and no cut off for MAF was used.

The EUREYE study has information on both early and late disease and therefore imputation analyses was carried out with (i) all AMD- encompassing early and late as one outcome (ii) early AMD only (iii) late AMD only and (iv) log fH. Further conditional analyses on the rs1061170 SNP were performed for all above outcomes.

7.2 Results

7.2.1 Baseline Characteristics of studies

Baseline characteristics for the EUREYE study are given in **Table 2.2** and for the Cambridge Eye study are given in the publication pertaining to the study [154] indicating no differences between characteristics of cases and controls apart from disease status.

7.2.2 Identification of tag SNPs and genotyping quality

SNPs that were identified using the tagging SNP approach which were subsequently directly typed in both studies are indicated in **Figure 7.2** and **Figure 7.3** show the LD plot for both the EUREYE and Cambridge AMD datasets, indicating that there is a low level of LD between the SNPs typed, which is what would be expected using such an approach to capture haplotype diversity across the region. Quality control statistics, including call rates, HWE and MAF of SNPs typed in these studies is given in **Table 7.1**

Figure 7.2 SNPs directly typed in EUREYE and Cambridge AMD study

SNPs typed in EUREYE *(Ensemble annotations)*

SNPs typed in Cambridge AMD study *(Ensemble annotations)*

Genes in Region/ Chromosome 1 ideogram

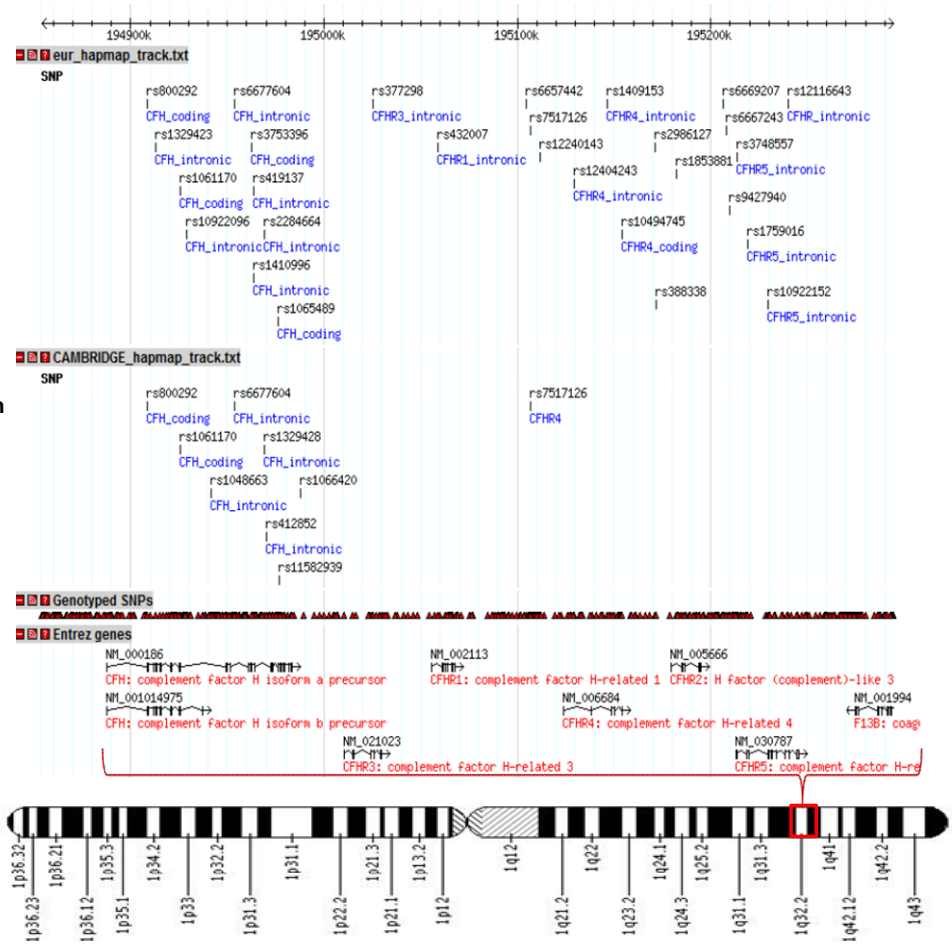
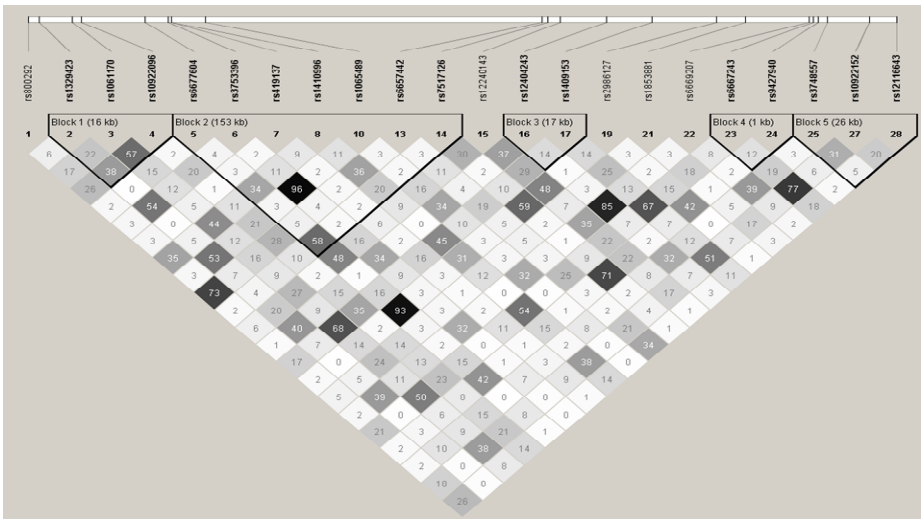


Figure 7.3 LD (r2) between SNPs typed in Cambridge AMD and EUREYE studies. These show the success of the tagging approach in that the LD is low between chosen SNPs, which enables variation across the haplotype to be captured. LD plots were created using Haploview [232].

SNPs typed in EUREYE



SNPs typed in Cambridge

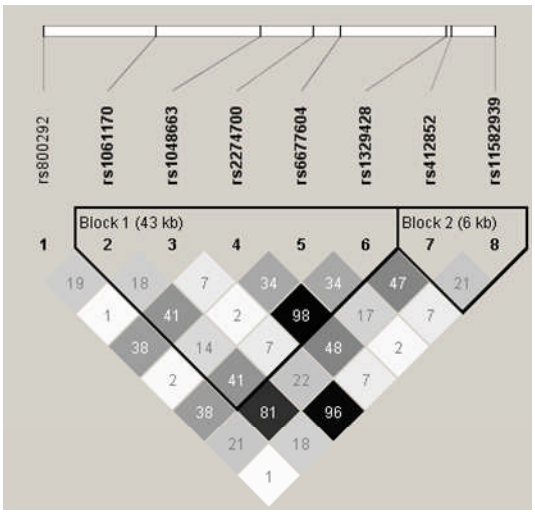


Table 7.1 Quality Control statistics for SNPs directly typed in EUREYE and the Cambridge AMD studies

Marker	Chromosome 1 position	EUREYE	EUREYE/ controls		Cambridge	Cambridge/ controls	
		Call rate	HWE chi^2	HWE p	Call rate	HWE chi^2	HWE p
rs800292	194908856	88%	4.24	0.40	94%	0.03	0.86
rs1329423	194913010	90%	0.59	0.44			
rs1061170	194925860	90%	0.93	0.33	99%	4.55	0.03
rs10922096	194929082	91%	0.93	0.33			
rs1048663	194941605				95%	0.03	0.85
rs2274700	194949570				95%	2.29	0.13
rs6677604	194953541	91%	1.55	0.21	95%	0.02	0.89
rs3753396	194962365	90%	0.09	0.76			
rs419137	194963498	91%	5.49	0.02			
rs1410996	194963556	91%	0.05	0.83			
rs2284664	194969148	91%	0.06	0.81			
rs1329428	194969433					1.85	0.17
rs412852	194970330					1.51	0.22
rs1065489	194976397	91%	0.00	0.97			
rs11582939	194976780				95%	0.09	0.77
rs1066420	194987678				96%	9.02	0.00
rs377298	195025164	87%	71.73	0.00			
rs432007	195059087	90%	119.16	0.00			
rs6657442	195104683	91%	0.16	0.69			
rs7517126	195106895	92%	0.86	0.35	93%	0.33	0.56
rs12240143	195111640	90%	0.01	0.93			
rs12404243	195129192	91%	3.43	0.06			
rs1409153	195146628	88%	2.10	0.15			
rs10494745	195154080	91%	0.01	0.93			
rs2986127	195171294	83%	0.71	0.40			
rs388338	195171556	22%	58.23	0.00			
rs1853881	195182368	91%	0.06	0.81			
rs6669207	195206465	91%	0.18	0.67			
rs6667243	195208116	91%	0.90	0.34			
rs9427940	195209976	91%	0.55	0.46			
rs3748557	195213492	91%	1.49	0.22			
rs1759016	195219121	100%	51.87	0.00			
rs10922152	195229629	91%	0.22	0.64			
rs12116643	195239806	90%	0.90	0.34			

Blue: SNPs typed in EUREYE and Cambridge; Red: SNPs typed in Cambridge only and **Black**: SNPs typed in EUREYE only

Based on the statistics presented in **Table 7.1**, SNPs with a call rate of below 90% and a HWE p value <0.05 were excluded from subsequent single SNP analyses. These SNPs included

rs800292, rs377298, rs1409153, rs2986127, rs388338, rs1065489, rs12240143 and rs10494745 in the EUREYE study. All SNPs in the Cambridge AMD study had call rates above 90% although rs1066420 was excluded on the basis of deviation from HWE. The remaining SNPs were carried forward to the subsequent analyses.

7.2.3 Analysis using directly typed SNPs

7.2.3.1 Genotype-AMD association

Single marker analyses from EUREYE and Cambridge AMD study re-affirmed the strong association of late AMD with the rs1061170 SNP encoding Y402H, with an unadjusted OR of 2.30 95% CI (1.94, 2.73) in EUREYE and similarly an odds of 2.55 95% CI (2.44, 2.66) in the Cambridge data set. In addition a number of other SNPs in this region are significant for their association with late AMD. In EUREYE these include SNPs in *CFH* rs10922096; a coding variant, rs6677604; and rs1410996, rs2284664; all intronic variants, and of note rs1410996 has previously been shown to confer AMD risk that is independent of rs1061170 [233]; SNPs in *CFHR1* included rs432007, and intergenic SNPs rs6657442, rs7517126, rs6667243, rs9427940 (between *CFHR4* and *CFHR5*) and SNPs in *CFHR2* included rs1853881 (**Table 7.2**). Of these SNPs, rs6677604 was also present and significant for its association with late AMD in the Cambridge AMD dataset (OR 0.37 95%CI 0.34, 0.76), although rs7517126 was only marginally significant in the Cambridge AMD study. An additional SNP rs800292 which was excluded from the EUREYE analyses because of QC, was significant in Cambridge AMD (OR 0.38 95% CI (0.37, 0.39)), and has been previously reported for its association with AMD, of note tagging a protective haplotype [233].

SNPs which were significant at this stage for each study were then taken forward into a regression model to test the statistical independence of these signals. For EUREYE, out of the 11 significant SNPs only 2 were retained within the model with p values of <0.05. These were rs1061170 (the Y402H SNP) and rs1410996, an intronic *CFH* SNP. When adjusting for age, SNPs rs6677604, rs7517126 and rs1853881 were no longer significantly associated with late AMD and taking the remaining 7 SNPs forward into an age adjusted stepwise analysis (using p<0.05 as a cut off), again only rs1061170 and rs1410996 were retained.

Similarly for the Cambridge AMD data set rs800292, rs1061170, rs1048663, rs1329428, rs412852 were all significant in the initial analysis, after correcting for multiple comparisons. Taking these forward into a stepwise regression model, again the rs1061170 SNP was retained and in addition, the rs1329428 an intronic SNP in *CFH* was retained, suggesting that the signal for late AMD is narrowed down to the *CFH* gene itself.

A further exploratory analysis of early AMD in EUREYE only show that only rs1061170 and rs1410996 were significant at the first stage of SNP by SNP analyses, prior to proceeding to a step-wise regression analysis.

Table 7.2 Association of SNPs in the RCA cluster with late AMD and log fH in the Cambridge AMD and EUREYE case-control datasets. Both corrected and uncorrected p values are shown.

SNP	EUREYE- AMD	AMD association P value	Corrected P value	β Coefficient for association with fH (95% CI)	Uncorrected lfH P value	Corrected P value	Cambridge AMD	AMD association P value	Corrected P value	β Coefficient for association with lfH (95% CI)	Uncorrected lfH P value	Corrected P value
rs800292	Excluded						0.38(0.37,0.39)	8.53×10^{-17}	6.83×10^{-16}	-0.02(-0.05,0.01)	0.19	1.0
rs1329423	0.89(0.7,1.13)	0.34	0.68	-0.01(-0.05,0.03)	0.66	1.0						
rs1061170	2.3(1.94,2.73)	3.10×10^{-21}	6.20×10^{-20}	-0.01(-0.06,0.04)	0.64	1.0	2.55(2.44,2.66)	1.17×10^{-23}	9.38×10^{-23}	0.02(0,0.05)	4.80×10^{-2}	0.38
rs10922096	0.58(0.47,0.71)	3.00×10^{-7}	6.00×10^{-6}	-0.03(-0.07,0.02)	0.19	1.0						
rs1048663							0.37(0.28,0.49)	8.00×10^{-12}	6.40×10^{-11}	0.01(-0.02,0.04)	0.45	1.0
rs2274700							Excluded					
rs6677604	0.62(0.5,0.77)	1.20×10^{-5}	2.40×10^{-4}	0.07(0.03,0.12)	7.50×10^{-3}	0.15	0.51(0.34,0.76)	1.00×10^{-3}	8.00×10^{-3}	0.04(0,0.07)	5.30×10^{-2}	0.42
rs3753396	0.87(0.6,1.26)	4.50×10^{-1}	0.90	-0.04(-0.11,0.02)	0.16	1.0						
rs419137	1.32(0.97,1.8)	8.20×10^{-2}	0.16	0.04(-0.01,0.08)	0.09	1.0						
rs1410996	0.43(0.33,0.57)	8.60×10^{-10}	1.72×10^{-8}	0.02(-0.03,0.07)	0.33	1.0						
rs2284664	0.46(0.34,0.62)	2.80×10^{-7}	5.60×10^{-6}	-0.04(-0.07,0)	0.05	1.0						
rs1329428							0.37(0.27,0.5)	8.00×10^{-11}	6.40×10^{-10}	0.01(-0.01,0.04)	0.31	1.0
rs412852							2.47(2.19,2.79)	8.60×10^{-49}	6.88×10^{-48}	0.03(0.01,0.06)	7.50×10^{-3}	6.00×10^{-2}
rs1065489	Excluded											
rs11582939							0.91(0.74,1.12)	0.39	1.0	-0.08(-0.11,-0.05)	4.00×10^{-6}	3.20×10^{-5}
rs1066420							Excluded					
rs377298	Excluded											
rs432007	0.54(0.45,0.65)	7.80×10^{-10}	1.56×10^{-9}	-0.02(-0.06,0.02)	0.31	1.0						
rs6657442	0.49(0.38,0.65)	2.60×10^{-7}	5.20E-06	-0.04(-0.07,0)	0.03	0.6						
rs7517126	0.56(0.48,0.66)	6.80×10^{-12}	1.36×10^{-10}	0.07(0.01,0.12)	2.10×10^{-2}	0.42	0.61(0.35,1.06)	7.90×10^{-2}	0.63	0.04(0,0.08)	6.10×10^{-2}	0.49
rs12240143	Excluded											
rs12404243	0.82(0.64,1.05)	0.01	1.0	0(-0.06,0.05)	0.094	0.19						

SNP	EUREYE- AMD	AMD association P value	Corrected P value	β Coefficient for association with fH (95% CI)	Uncorrected IfH P value	Corrected P value	Cambridge AMD	AMD association P value	Corrected P value	β Coefficient for association with IfH (95% CI)	Uncorrected IfH P value	Corrected P value
rs1409153	Excluded											
rs10494745	Excluded											
rs2986127	Excluded											
rs388338	Excluded											
rs1853881	1.59(1.3,1.93)	5.40×10^{-6}	1.08×10^{-4}	-0.01(-0.07,0.05)	0.69	1.0						
rs6669207	0.79(0.59,1.06)	0.012	1.0	0(-0.05,0.05)	0.91	1.0						
rs6667243	0.46(0.39,0.54)	4.40×10^{-20}	8.80×10^{-19}	-0.02(-0.06,0.03)	0.39	1.0						
rs9427940	0.59(0.48,0.72)	4.80×10^{-7}	9.60×10^{-6}	0.06(0.02,0.1)	0.011	0.2						
rs3748557	0.75(0.52,1.1)	0.014	1.0	-0.01(-0.08,0.05)	0.61	1.0						
rs1759016	0.72(0.48,1.07)	0.01	1.0	0.01(-0.02,0.04)	0.57	1.0						
rs10922152	0.55(0.46,0.65)	3.30×10^{-11}	6.60×10^{-10}	-0.01(-0.05,0.03)	0.46	1.0						
rs12116643	0.44(0.31,0.64)	1.70×10^{-5}	3.40×10^{-4}	-0.04(-0.09,0)	0.062	1.0						

Blue: SNPs typed in EUREYE and Cambridge, Red: SNPs typed in Cambridge only and **Black**: SNPs typed in EUREYE only. SNPs that have been excluded in these analyses because they have not passed QC criteria are indicated

7.2.3.2 Genotype-fH association

The same SNPs that were tested for association with AMD were then tested for their association with circulating fH levels which had been measured in both studies. Log transformed values of fH were used for analyses, given skewed distributions of fH in both studies which were normalised with this transformation. After adjusting the p values for multiple testing, none of the SNPs associated with late AMD were significantly associated with fH in the EUREYE dataset **Table 7.2**. In the Cambridge AMD case-control study however 2 SNPs were associated with fH, rs10922096 (β coefficient of log fH -0.09 95% CI (-0.13, -0.06) p value 1.04×10^{-6}) and rs11582939 (β coefficient of log fH -0.08 95% CI (-0.11, -0.05) p value 4×10^{-6}), and one marginally associated with fH levels, rs412852 (β coefficient of log fH 0.03 95% CI (0.01, 0.06) p value 7.5×10^{-3}). Of these, only rs412852 was the only SNP also associated with AMD (OR 2.47, 95% CI (2.19, 2.79) and p value 8.6×10^{-49}). The results for all SNPs analysed for both EUREYE and Cambridge AMD study are given in **Table 7.2**. In order to compare the SNP effects on AMD and fH together, SNPs were re-coded to compare the risk versus non risk genotype, and OR's were plotted against β coefficients for effect on log fH (**Figure 7.4)**

SNPs that had previously been reported as having an association with AMD were probed a little further in terms of their fH associations per genotype group. Even though the above analyses show no or marginal associations between SNPs and fH, interrogating individual SNP associations with fH level may give an indication of a trend in change of levels, which may in itself be important and borne out by replication of any findings in other studies with both genotypes and fH measures. SNPs selected, based on previous publications [114, 233, 234] were rs800292, rs1061170 (Y402H), rs3753396 and rs1410996, which were all genotyped in EUREYE, although out of these, genotypes for the rs1061170 and rs800292 were only available in the Cambridge AMD study.

Figure 7.4 Scatter plot of OR for late AMD and beta coefficient for the SNP effect on log fH. Analysis indicates per allele estimates, SNPs have been re-coded so to compare the risk verses the non risk genotypes.

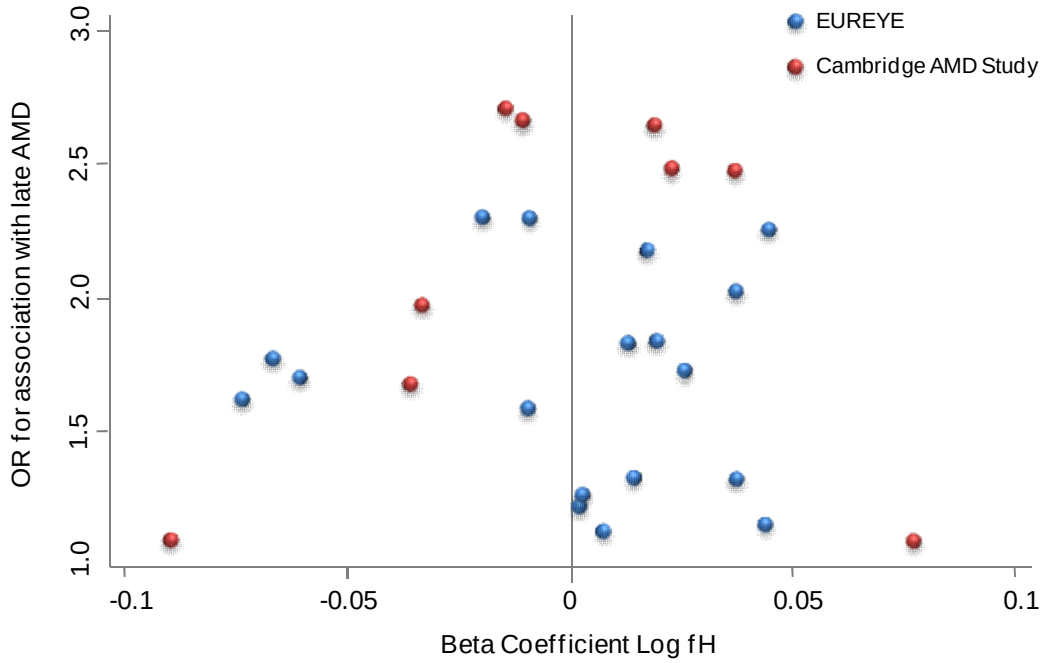
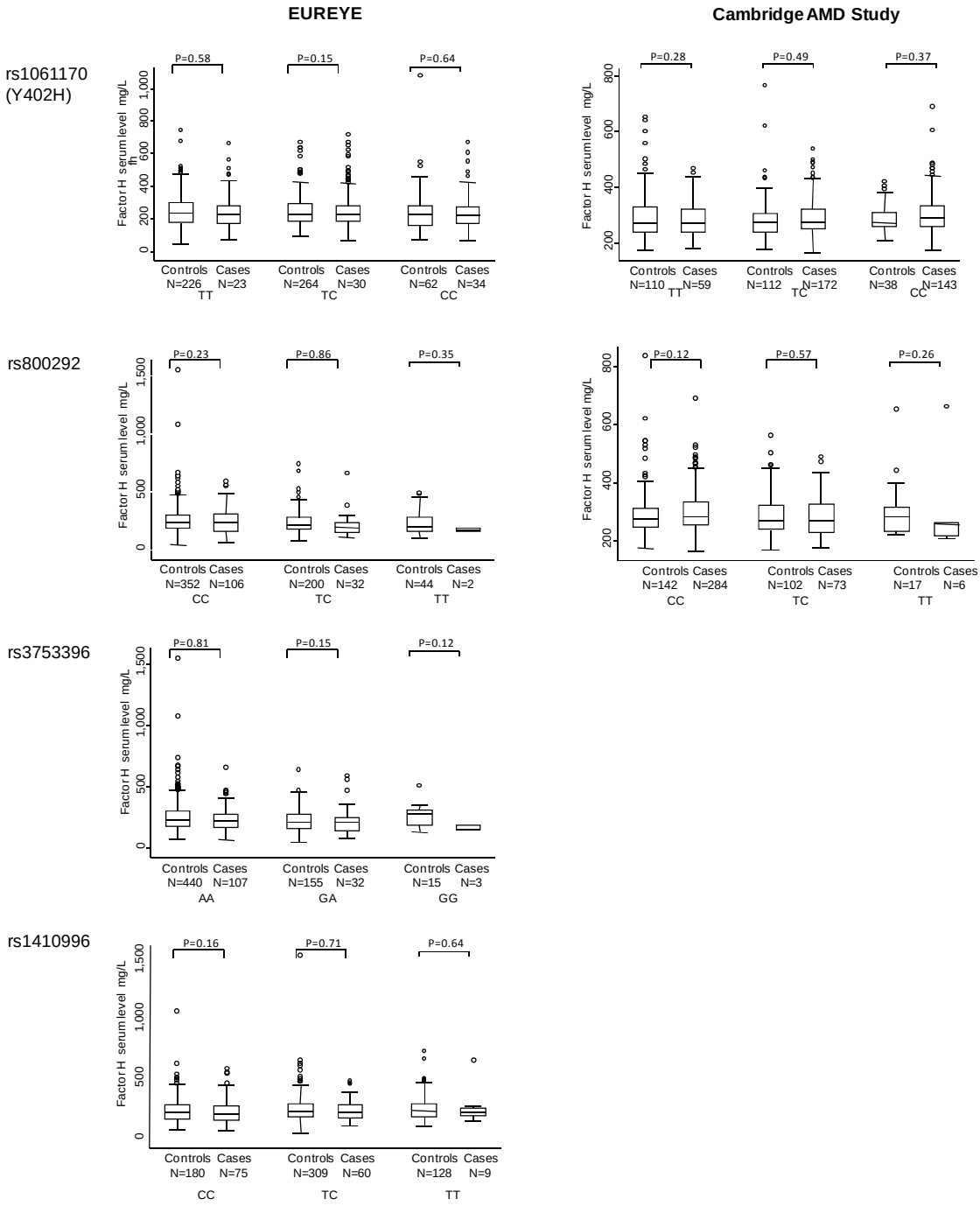


Figure 7.5 SNPs previously reported for association with AMD and their association with fH



7.2.3.3 Circulating fH-AMD association

Circulating levels of fH were not associated with late AMD in either EUREYE or the Cambridge Eye study. There was no significant difference in fH levels between cases of late AMD and controls (**Table 7.3 and 7.4)**. Odds ratios were expressed as one standard deviation increase in

log fH. Unadjusted odds in late AMD in EUREYE were 0.87 95% CI (0.70, 1.08), and in Cambridge were 1.08 95% CI (0.92, 1.26).

Table 7.3 Association of fH with late AMD in EUREYE

EUREYE	Controls N=568	Late AMD Cases N=124	OR* (95% CI)	P value
fH**	231.85 (93.06)	218.10 (95.33)	0.87 (0.70, 1.08)	0.20

Table 7.4 Association of fH with late AMD in the Cambridge AMD study

CAMBRIDGE	Controls N=264	Late AMD Cases N=378	OR* (95% CI)	P value
fH**	284.68 (68.12)	289.65 (65.25)	1.08 (0.92, 1.26)	0.35

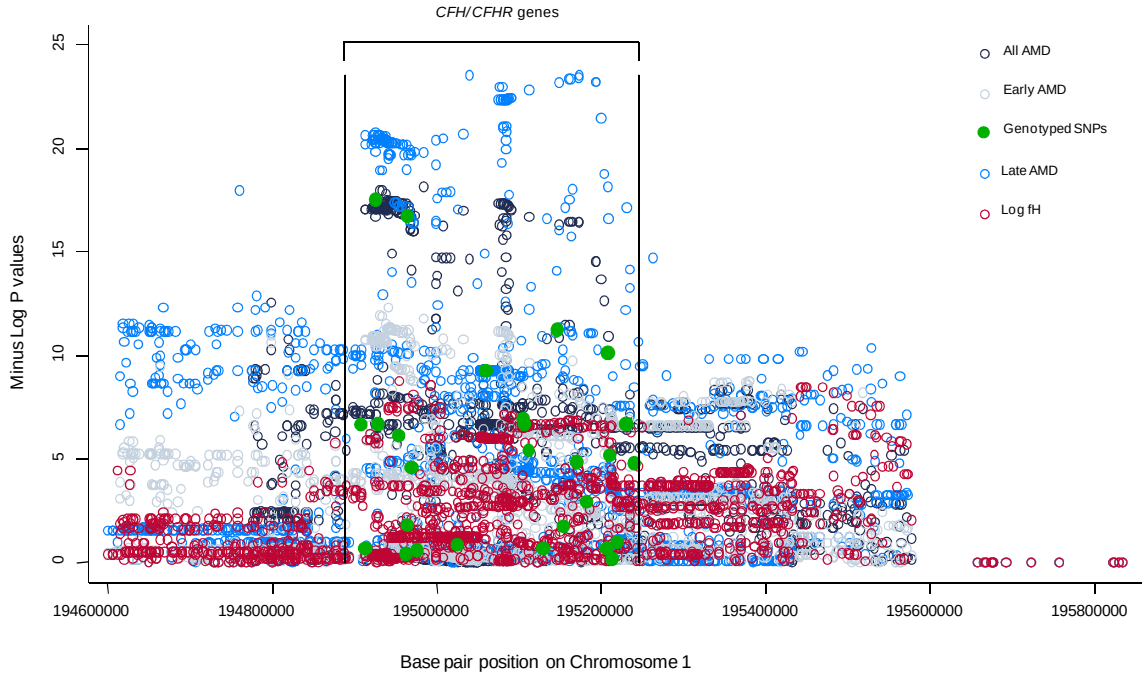
*Odds for a 1 SD increase in log fH, ** log transformed data with approximate SD's.

Exploratory unadjusted analysis in the EUREYE data set showed no association of fH with early AMD either.

7.2.4 Imputation and association of imputed SNPs with fH and AMD in EUREYE

Imputation was specified for a ~1Mb region from *KCNT2* to *DENND1B*, a region which incorporates *CFH* and the *CFHR*s (**Figure 7.1**). This region in addition incorporates the *CRB1* gene which although not a complement related protein, was found to be associated with fH levels in Chapter 6. Samples used from EUREYE are given in **Table 7.1**, which were used to infer genotypes. SNPs with initial low call rates, <80% were excluded (rs388338). Following imputation 2764 SNPs were available, from which a further 867 were excluded by QC criteria, leaving 1897 SNPs across this 1Mb region available to test for association. The imputation results are summarised in **Figure 7.5** and **Tables 7.5-7.8**. In the tables, the top 25 most significant SNPs are given, and if the rs1061170 SNP is not in this list it is indicated as an addition.

Figure 7.6 Association of typed and imputed SNPs in the CFH region with AMD and fH concentration. The region of *CFH* and *CFHR's* is indicated. The –log p values for the association are shown of directly typed SNPs and AMD (closed green circles), and imputed SNPs with any AMD (open dark blue circles), late AMD (open mid blue circles), early AMD open light blue circles and log fH (open red circles).



Although results are shown for any AMD (all, early and late), late AMD encompassing GA and CNV is thought to represent the most unequivocal AMD phenotype and account for the most sight threatening forms of disease. Analysis of the association plot shows that signals for late AMD span the whole *CFH/ CFHR* cluster, and extend further down to include *CRB1* which was identified in the fH GWAS, and has role in the developing retina. However further extension beyond this region is not excluded as imputation was limited to this area. However recent GWAS did not report findings of further extension of this signal extending into other RCA associated genes in addition to the *CFH/ CFHR* cluster. **Tables 7.5-7.8** outline the 25 top ranking signals for these associations. For late AMD, these are largely intergenic SNPs although analysis of signals lower down to this are largely *CFH* SNPs (**Appendix 2**).

Table 7.5 Association of imputed SNPs with any AMD

Rank	rs-ID	BP Position	P value	Gene	Predicted function of SNP
1	rs506317	194983776	$1.40x10^{-08}$	*CFH*	3downstream
2	rs10922097	194931105	$1.60x10^{-08}$	*CFH*	intronic
3	rs1089033	194933416	$1.60x10^{-08}$	*CFH*	intronic
4	1-194938293	194938293	$1.90x10^{-08}$		
5	rs9970075	194940053	$2.50x10^{-08}$	*CFH*	intronic
6	rs1061170	194925860	$2.50x10^{-08}$	*CFH*	coding
7	rs9970784	194940425	$2.60x10^{-08}$	*CFH*	intronic
8	rs402991	194942726	$2.60x10^{-08}$	*CFH*	intronic
9	rs6664705	194944465	$2.60x10^{-08}$	*CFH*	intronic
10	rs1887973	194941802	$2.60x10^{-08}$	*CFH*	intronic
11	rs203685	194944568	$2.60x10^{-08}$	*CFH*	intronic
12	rs399469	194942728	$2.60x10^{-08}$	*CFH*	intronic
13	rs203677	194949989	$2.70x10^{-08}$	*CFH*	intronic
14	rs203680	194947998	$2.70x10^{-08}$	*CFH*	intronic
15	rs203683	194945055	$2.70x10^{-08}$	*CFH*	intronic
16	rs203679	194949141	$2.70x10^{-08}$	*CFH*	intronic
17	rs7522681	194945228	$2.70x10^{-08}$	*CFH*	intronic
18	rs203674	194951248	$2.70x10^{-08}$	*CFH*	intronic
19	rs374896	194959001	$2.90x10^{-08}$	*CFH*	intronic
20	rs403846	194963360	$2.90x10^{-08}$	*CFH*	intronic
21	rs381974	194959295	$2.90x10^{-08}$	*CFH*	intronic
22	rs393955	194959093	$2.90x10^{-08}$	*CFH*	intronic
23	rs2860102	194934942	$2.90x10^{-08}$	*CFH*	intronic
24	rs203672	194955685	$2.90x10^{-08}$	*CFH*	intronic
25	rs456761	195078019	$3.00x10^{-08}$	*Intergenic (CFH and CFR3)*	

Table 7.6 Association of imputed SNPs with early AMD

Rank	rs-ID	BP Position	P value	Gene	Predicted function of SNP
1	rs203687	194940893	$4.70x10^{-06}$	*CFH*	intronic
2	rs10922097	194931105	$6.60x10^{-06}$	*CFH*	intronic
3	rs1089033	194933416	$6.80x10^{-06}$	*CFH*	intronic
4	rs506317	194983776	$7.60x10^{-06}$	*CFH*	3downstream
5	1-194938293	194938293	$1.00x10^{-05}$	*CFH*	
6	rs10733086	194943558	$1.2x10^{-05}$	*CFH*	intronic
7	rs1410997	194943786	$1.2x10^{-05}$	*CFH*	intronic
8	rs2019724	194941540	$1.2x10^{-05}$	*CFH*	intronic
9	rs6428357	194942194	$1.2x10^{-05}$	*CFH*	intronic
10	rs9970075	194940053	$1.2x10^{-05}$	*CFH*	intronic
11	rs1831282	194940616	$1.3x10^{-05}$	*CFH*	intronic
12	rs1887973	194941802	$1.3x10^{-05}$	*CFH*	intronic
13	rs203674	194951248	$1.3x10^{-05}$	*CFH*	intronic
14	rs203677	194949989	$1.3x10^{-05}$	*CFH*	intronic
15	rs203679	194949141	$1.3x10^{-05}$	*CFH*	intronic
16	rs203680	194947998	$1.3x10^{-05}$	*CFH*	intronic
17	rs203683	194945055	$1.3x10^{-05}$	*CFH*	intronic
18	rs203685	194944568	$1.3x10^{-05}$	*CFH*	intronic
19	rs379489	194960074	$1.3x10^{-05}$	*CFH*	intronic
20	rs399469	194942728	$1.3x10^{-05}$	*CFH*	intronic
21	rs402991	194942726	$1.3x10^{-05}$	*CFH*	intronic
22	rs6664705	194944465	$1.3x10^{-05}$	*CFH*	intronic
23	rs7522681	194945228	$1.3x10^{-05}$	*CFH*	intronic
24	rs9970784	194940425	$1.3x10^{-05}$	*CFH*	intronic
25	1-195081157	195081157	$1.4x10^{-05}$	*Intergenic (CFHR1-3)*	
100	rs1061170	194925860	$2.5x10^{-05}$	*CFH*	Coding

Table 7.7 Association of imputed SNPs with late AMD

Rank	rs-ID	BP Position	P value	Gene	Predicted function of SNP
1	rs410895	195039969	6.20×10^{-11}	*Intergenic (CFHR1-3)*	
2	rs6699120	195173543	6.20×10^{-11}	*Intergenic (CFHR4-2)*	
3	rs570965	195172614	7.00×10^{-11}	*Intergenic (CFHR4-2)*	
4	rs2878690	195160176	7.10×10^{-11}	*Intergenic (CFHR4-2)*	
5	rs4428917	195162983	7.10×10^{-11}	*Intergenic (CFHR4-2)*	
6	rs1359594	195194051	8.70×10^{-11}	*CFHR2*	intronic
7	rs2026547	195194414	8.70×10^{-11}	*CFHR2*	intronic
8	rs1853883	195148223	8.80×10^{-11}	*CFHR4*	intronic
9	1-195079152	195079152	1.10×10^{-10}	*Intergenic (CFH-CFH3)*	
10	rs57720283	195076226	1.10×10^{-10}	*Intergenic (CFH-CFH3)*	
11	rs6695525	195112144	1.20×10^{-10}	*LOC100289145 (CFHR1-4)*	
12	rs1963606	195089476	1.80×10^{-10}	*LOC100289145 (CFHR1-4)*	
13	rs1963607	195089582	1.80×10^{-10}	*LOC100289145 (CFHR1-4)*	
14	rs10754206	195087547	1.90×10^{-10}	*LOC100289145 (CFHR1-4)*	
15	rs10922142	195087731	1.90×10^{-10}	*LOC100289145 (CFHR1-4)*	
16	rs6665824	195086857	1.90×10^{-10}	*LOC100289145 (CFHR1-4)*	
17	1-195081157	195081157	2.00×10^{-10}	*LOC100289145 (CFHR1-4)*	
18	1-195081164	195081164	2.00×10^{-10}	*LOC100289145 (CFHR1-4)*	
19	1-195082766	195082766	2.00×10^{-10}	*LOC100289145 (between CFHR1-4)*	
20	rs10922140	195085116	2.00×10^{-10}	*LOC100289145 (CFHR1-4)*	
21	rs1738740	195075263	2.00×10^{-10}	*Intergenic (CHFR1-4)*	
22	rs412967	195080339	2.00×10^{-10}	*Intergenic (CHFR1-4)*	
23	rs455199	195077757	2.00×10^{-10}	*Intergenic (between CFH and CFR3)*	
24	rs456761	195078019	2.00×10^{-10}	*Intergenic (CFH and CFR3)*	
25	rs61822193	195081192	2.00×10^{-10}	*Intergenic (CFH and CFR3)*	
32	rs1061170	194925860	9.8×10^{-10}	*CFH*	coding

Table 7.8 Association of imputed SNPs with log fH

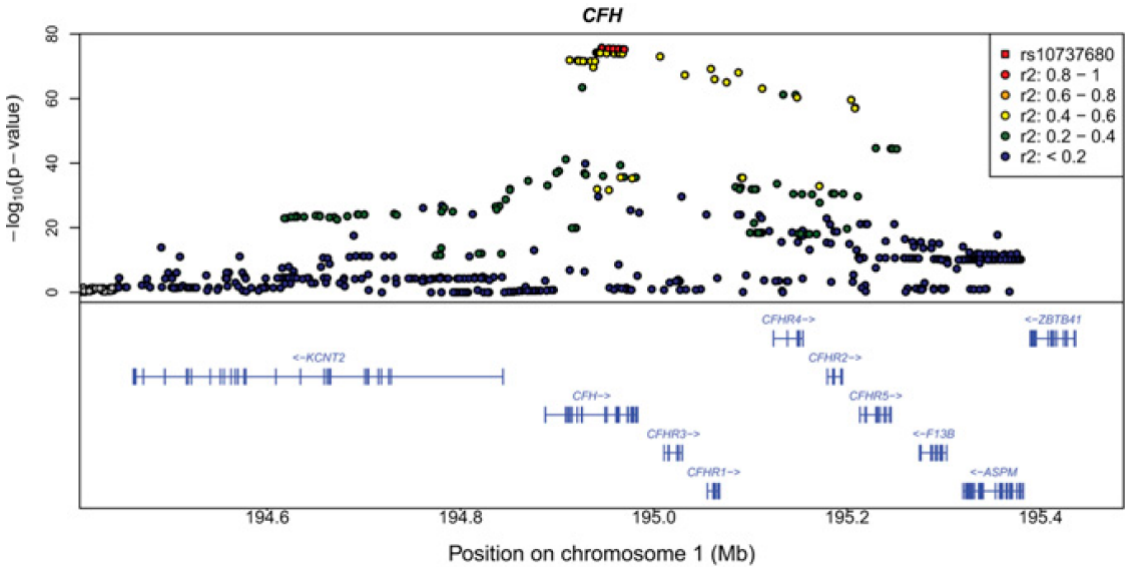
Rank	rs-ID	BP Position	P value	Gene	Predicted function of SNP
1	rs6677604	194953541	$1.65x10^{-04}$	*CFH*	intronic
2	1-194992562	194992562	$1.92x10^{-04}$	*CFH*	
3	1-194993006	194993006	$1.93x10^{-04}$	*CFH*	
4	rs12026252	195447943	$2.06x10^{-04}$	*Intergenic (between ZBTB41-CRB1)*	
5	rs10801598	195450279	$2.06x10^{-04}$	*Intergenic (between ZBTB41-CRB1)*	
6	rs1577802	195468419	$2.07x10^{-04}$	*Intergenic (between ZBTB41-CRB1)*	
7	rs4915393	195444410	$2.06x10^{-04}$	*Intergenic (between ZBTB41-CRB1)*	
8	rs7547972	195442782	$2.08x10^{-04}$	*Intergenic (between ZBTB41-CRB1)*	
9	rs10801600	195483589	$2.79x10^{-04}$	*Intergenic (between ZBTB41-CRB1)*	
10	rs2064456	194971066	$3.05x10^{-04}$	*CFH*	intronic
11	rs4915440	195498434	$3.26x10^{-04}$	*Intergenic (between ZBTB41-CRB1)*	
12	rs10801597	195441837	$3.32x10^{-04}$	*Intergenic (between ZBTB41-CRB1)*	
13	1-194995616	194995616	$4.01x10^{-04}$	*CFH*	
14	rs16840522	194977539	$4.49x10^{-04}$	*CFH*	intronic
15	rs12022402	195504560	$5.28x10^{-04}$	*CRB1*	intronic
16	rs7548481	195514437	$5.28x10^{-04}$	*CRB1*	intronic
17	rs1337170	195535116	$5.31x10^{-04}$	*CRB1*	intronic
18	rs7418217	195538871	$5.33x10^{-04}$	*CRB1*	intronic
19	rs6689009	194965086	$5.51x10^{-04}$	*CFH*	intronic
20	rs6677460	194951198	$5.56x10^{-04}$	*CFH*	intronic
21	rs2300429	194941979	$5.58x10^{-04}$	*CFH*	intronic
22	rs33915960	194943336	$5.58x10^{-04}$	*CFH*	intronic
23	rs35617250	194946305	$5.59x10^{-04}$	*CFH*	intronic
24	rs12134598	194947624	$5.59x10^{-04}$	*CFH*	intronic
25	1-195003848	195003848	$5.64x10^{-04}$	*CFH*	
1405	rs1061170	194925860	0.58	*CFH*	coding

Further conditioning analysis on rs1061170 for late AMD data did not change the top ranking signals, the position of some of these changed in the ranking order although no new signals were found with conditioning analysis.

7.3 Discussion

In this chapter I have re-affirmed the signal from the most replicated rs1061170 SNP with late AMD in the EUREYE and Cambridge eye studies. In addition to this SNP, I have shown that the signal is not confined to rs1061170 but spans *CFH* and also includes *CFHR* genes. Imputation analysis demonstrates that the extension of this signal could be up to 1Mb downstream of the *CFH/ CFHR* cluster. Analysis conditioning on rs1061170 in addition did not yield a signal that was different to the unconditioned analysis. In so doing however I was not able to refine the signal for AMD in itself, but rather shown that the region of association is broad, and extends 1 Mb downstream of the *CFH/ CFHR* cluster. The recent AMD GWAS [36] which has greater power (2157 cases and 1150 controls) which used a denser SNP array in comparison with the first of the AMD GWAS, and in addition imputed against HapMap resulting in the study of ~2.5 million SNPs, also shows that although the region of association is greatest for the *CFH* gene, the signal is broader, extending both upstream and downstream (**Figure 7.6**).

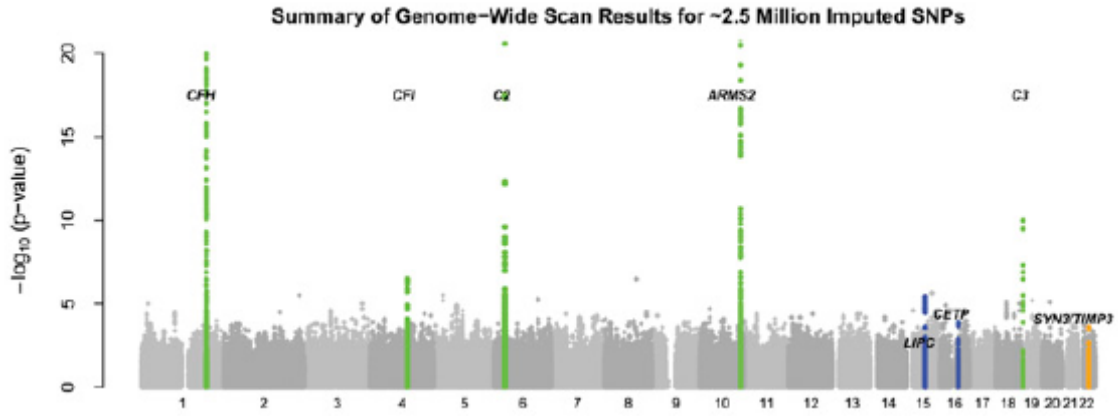
Figure 7.7 Regional association plot for the CFH region from recent AMD GWAS, taken from Chen et al, PNAS, 2010 [36]



The association of other regions of the genome which are related to complement regulation and in particular the alternative pathway (*CFB, CFI, C3*) were also demonstrated in this GWAS, a finding not reported in the initial AMD GWAS (**Figure 7.7**), supporting the view that AMD is likely to result from multiple disease associated alleles at more than one locus. Although this may be the case, the alternative complement pathway is by far the strongest pathway signal arising from the genome scan which poses the important question of whether variation in these genes cause alternations in the cognate protein as the mechanism by which risk is increased.

However there was a lack of overlap between the disease-causing SNPs identified in this chapter and those that were shown to be associated with fH levels in Chapter 6. One reason for this could be that the AMD signal is dependent on fH function rather than on fH concentration. Given the complement pathway is an amplification cascade, it may be that small changes in concentration may have large functional effects which would be missed by just measuring fH levels. However there are currently no scalable functional activity assays of the alternative pathway in general and of fH in particular and therefore this remains a hypothesis that still needs to be tested. Another reason which has already been touched upon in prior chapters is local synthesis of fH may account for local retinal effects and therefore circulating levels of fH may be a poor indicator of local activity or function.

Figure 7.8 Manhattan plot from the updated AMD GWAS. Apart from the SNPs at the ARMS2 locus, whose function is not completely understood, the loci identified contain genes in the complement pathway, and also implicated are genes in the regulation of circulating lipids. This figure is taken from Chen et al, PNAS, 2010 [36]

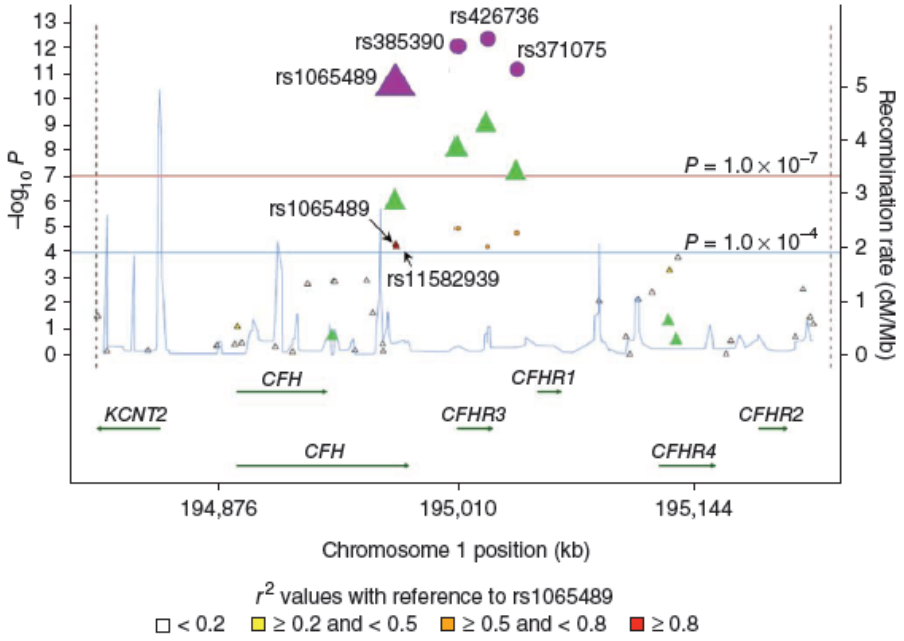


Data emerging from the area of renal disease however appears to suggest that SNPs in *CFH* do contribute to alterations in the circulating pool of fH, in keeping with finding from the original linkage studies which provided evidence for a QTL at the RCA cluster [134]. By developing antibodies that specifically detect either the Y402 or the H402 protein product, Hakobyan et al [203, 235] have shown what the contribution of each allele to the circulating fH pool is by heterozygotes, which has aided in the identification of novel mutations in aHUS [235]. However another important finding demonstrated by the latter study which is pertinent to the discussion here, is that on investigating a pedigree of aHUS, the index case, a rs1061170 heterozygote, had normal levels of Y402 although low levels of H402, that is they had a low expression allele. This individual also carried another common mutation associated with aHUS (R1210C) also in a heterozygote state. However siblings with this mutation did not develop aHUS and others who were also rs1061170 heterozygote were found to have normal levels of H402. Although a genetic basis of this difference and the cause of the low expression of H402 in the index case was not identified, it was shown that this individual had a 5 fold higher than expected level of FHL-1, suggesting that a mutation which disrupts alternative splicing of fH could be responsible. This is an important consideration in the interpretation of the results presented here given the shared epitopes between fH, FHL-1 and the FHR proteins. The commercial antibody used to develop the assay discussed in Chapter 3 is a polyclonal antibody, and may well recognise

shared epitopes between fH and FHL-1 as well as the FHR's. Therefore the association or lack thereof of fH with AMD in the data presented here could be because the signal is noisy, and other proteins are being measured. On the other hand if the epitopes which are recognised by the antibody are limited to fH specific regions, this would indicate that it may be the alternatively spliced form of the protein, FHL-1 or FHR proteins are linked with AMD rather than fH itself. This may be an explanation as to why no overlap is seen between SNPs identified in the fH GWAS (Chapter 6) and the SNPs identified through imputation as well as those directly genotyped for association with AMD in this Chapter. It may also be an explanation of why the signal for the genotype- AMD signal in this and other studies is very compelling, and in this situation the expectation would be that the signal from the encoded protein is just as strong if not stronger, however this is not the case.

The lack of overlap between fH and AMD SNPs is however in contrast to the overlap that is seen for the fH GWAS SNPs (Chapter 6) and findings from the recent GWAS in meningococcal disease [12]. Four of the top ranking SNPs in this study (**Figure 7.8**) have also been identified the GWAS presented in Chapter 6 (**Table 6.1**) with the remaining SNPs also appearing as high ranking signals in the fH GWAS. This and a previous report by the same group of a *CFH* promoter SNP implicated in susceptibility to meningococcal disease [236] and altering levels of circulating fH lends further support to the initial hypothesis that changes in levels of the protein are likely to mediate disease, however as already mentioned bearing in mind the amplification nature of the complement cascade these changes may well be small. It remains to be investigated as to what the changes in fH levels are in these individuals with meningococcal disease compared to controls.

Figure 7.9 Associations of SNPs in the *CFH* region with meningococcal disease, taken from Davila S et al [12]. Amongst the top ranking signals in this study SNPs rs426736, rs385390, rs371075, rs1065489 were associated with fH concentration in the MRC Fenland study (Table 6.1).



Whether related to fH function or concentration, the genetic association of *CFH* for AMD remains compelling, with further evidence of a possible pharmacogenetic effect of the rs1061170 SNP, with HH402 individuals having a poorer response to intravitreal injections of the VEGF monoclonal antibody bevacizumab [237].

Other limitations of the methods of analysis used in this chapter are worthy of discussion. In the imputation process employed to refine the signal for AMD may have excluded by QC filters rarer SNPs which may have greater uncertainty when imputation is attempted. One way to overcome this would be to not use any QC filters, which has been advocated before [238], and it is for this reason that no MAF filters were used, although bearing in mind any finding of a SNP which is rare would need to be replicated. One other way to attempt to identify rarer mutations would be used genotyped data and impute against reference panels (HapMap and 1000 Genomes) of African descent populations who have smaller haplotype blocks and mutations which are found at a reduced frequency in other populations. These results could then be validated through

direct genotyping and comparison across populations, and aid in the identification of possible causal mutations. .

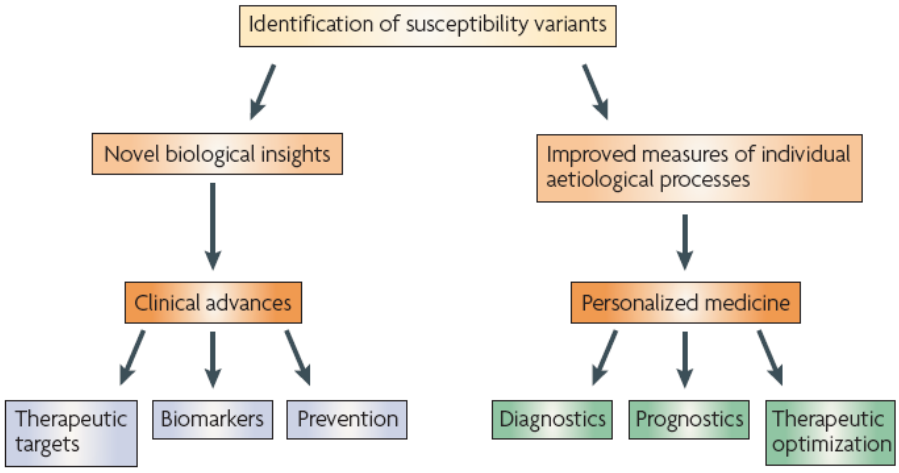
8 Discussion, and future work

Age-related macular degeneration is a devastating condition and emerging as the most common cause of blindness in the ageing population. Unlike cardiovascular disease, risk factors and benefits of risk factor modification in slowing disease progression are only now emerging from prospective studies of AMD. No clinical tests are available which adequately predict those who are most susceptible to disease, or that can aid in risk stratification. Moreover no treatments have been identified that reduce progression of early stages of disease thus avoiding the more sight threatening form of AMD. These factors together have been the impetus for identifying biological pathways that may confer susceptibility to AMD and which may be targetable with therapies at an early stage so as to reduce the public health burden of AMD.

The contribution of genetics in the post-genome era has been central in attempting to address these questions for AMD. Although the RCA cluster and complement was identified by family based linkage scans [134] and observational data of analysis of drusen in affected individuals [105], a GWAS was responsible for refining the signal down to complement factor H [10], and as such led to candidate gene association studies of variants in complement related genes which have subsequently been shown to confer AMD risk (see results of meta-analysis Chapter 2). AMD is thus an exemplar of the integration of different genetic approaches in delineating genetic contributions to a disease. The efforts have combined to place the alternative complement pathway firmly on the AMD map. AMD in this respect is also the success story of GWAS, in that resulting discoveries have the potential to achieve all that is hoped of this approach (**Figure 8.1**). Specifically these aims are to identify novel biological pathways which may result in clinical advances including the identification of biomarkers that may have a role in prediction and stratification of disease, and within these pathways identifying therapeutic targets whose manipulation may be of benefit in preventing or slowing disease progression. All these criteria are satisfied by the example of AMD, identification of the alternative complement pathway has been a major step forward in identifying disease mechanisms, and also potential overlapping mechanisms with other diseases given the existing association of *CFH* mutations in renal diseases such as MPGN II and aHUS (both reviewed in [127]) as well as newly identified associations such as meningococcal disease [12] Biological measurement of proteins encoded

by these genes may offer potential biomarkers which themselves may be open to pharmacological manipulation, and as such are already on the drug development radar [15]. Cheap and high throughput methods of genotyping in themselves offer the potential of a predictive test, and this is an attractive approach given the high risk associated with the rs1061170 SNP. The potential of combining a number of SNP markers has already been evaluated in one study, albeit small, in order to test the predictive utility of SNPs in complement genes for AMD [138]. Furthermore the aim of achieving individual level therapeutic optimisation is also evolving in the AMD field with the finding that the rs1061170 SNP is associated with differential response to VEGF blockade in individuals with late stage AMD [239]. So in terms of GWAS AMD certainly ticks, or appears to tick all the boxes.

Figure 8.1 Potential translational applications of findings from GWAS in common disease
Figure taken from McCarthy et al Nature Genetics 2007 [240]



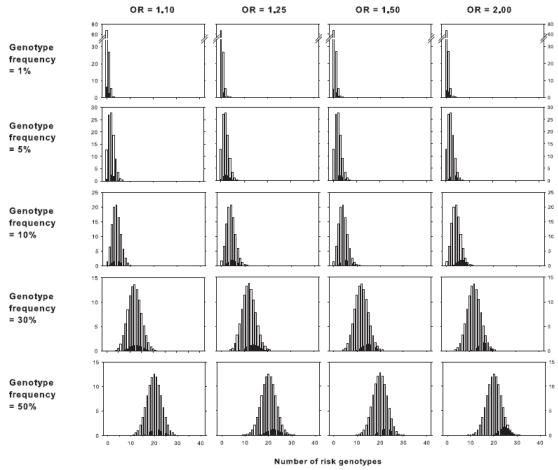
These are however headlines and discussion of the finer print is necessary in understanding where we are in reality in translating these findings into clinical practice, and putting findings from AMD into the broader context of GWAS in other diseases to understand that although apparently successful here, this approach is not a panacea for all common diseases.

AMD is touted as the GWAS success story, the exception to the rule, or rather the hypothesis being that common diseases are likely to arise from the presence of common variants which are predicted to have only a modest influence on disease, commonly referred to the “common-disease, common variant hypothesis” [241]. It was this hypothesis that underpinned the cataloguing of common variants (HapMap) after completion of the draft sequence of the human

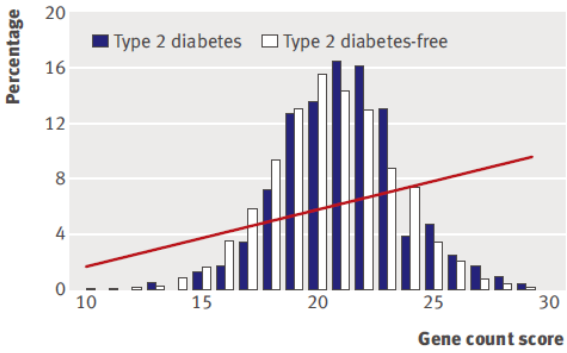
genome, enabling subsequent genome wide scans. Such scans in common diseases for example T2DM [242] and Crohn's Disease [243] have revealed ≥20 different loci which are associated with disease. Like AMD these scans have provided novel insights which are already directing research in these areas, such as the importance of autophagy as a mechanism in Crohn's and the importance of insulin secretion over signalling in T2DM. However it is also evident from these scans that the relative contribution to disease for each SNP across the different loci is small, and as such GWAS in general have fallen short in explaining the heritability of common diseases or traits. The best example of this is probably that of human height where large twin studies have shown that 90% of variation in height is attributable to inherited factors [244]. GWAS have uncovered 180 loci associated with human height [245] although these explain less than 10% of the variation. In effect the observation that Ronald Fisher made almost a century ago that a number of genetic variants, whose individual effect is small, can explain a quantitative trait [246] is supported by the results of genome wide scans. How do these findings then translate, if at all into being clinically useful? Either used singly or as aggregate gene based scores these common variants are poor discriminators of disease both in models and in real examples (**Figure 8.2**), taking for example T2DM. Common SNPs with intermediate risk define individuals where most events are likely to be observed, and could therefore be compared to biomarkers such as blood pressure or cholesterol in prediction of CHD risk for example. However it may be that identification of rarer alleles which are associated with larger effect sizes may be useful as predictive tests. These data will perhaps emerge with the completion of the 1000 genomes project and expansion of genetic studies across the globe, so that population frequencies of variants in themselves may provide important insights into risk and mechanisms of disease, and may importantly lead to the identification of therapies that are applicable not only to those at greatest risk, although at a population based level. In this way genetics can contribute to public health in reducing the risk of populations as opposed to individuals [247].

Figure 8.2 Performance of a hypothetical genotype based risk score (a) and empirical evaluation of a genotype risk score (b) for Type 2 Diabetes. In the simulation study in (a), Janssens et al simulated the effect of 40 SNPs with variable genotype frequencies and a population disease risk of 10%. With low genotype frequencies (1%) no individual carried more than 6/40 of the risk genotypes, with an increasing genotype frequency (30%) all individuals carried at least 6 of the risk genotypes. Because the frequency distribution of common inherited alleles is normal more cases of those with T2DM would be expected to have "bell" of the curve that is intermediate risk alleles, and as shown in the simulation study a genotype based score would not be of clinical utility. Talmud et al (b) found this to be true when using a genotyped based score in a population based cohort that were genotyped for common T2DM SNPs, illustrating that thus far a phenotype score currently has more predictive value than a genotyped based score.

(a)



(b)



So even though AMD stands out as a beacon of success of GWAS, like other common diseases it does also suffer the same problems of clinical utility of its genetic findings as other GWAS mentioned above, and this is perhaps why there currently is no translation of the findings from these studies. Some of these issues are discussed in a discursive paper highlighting the short falls of genetics in AMD prediction [248]. Here it is argued that genetic studies to date in AMD have largely been case-control studies, and although this is a useful design to maximise power of uncovering risk variants, when it comes to applying a genetic test at a population level, it is likely to be over-estimating the effect size, and in reality such tests will be poorly

discriminative in individuals with intermediate findings. One solution to this is longitudinal data and the prospective assessment of variants in predicting incident disease, although using this approach the clinical utility of some of these variants may not be realised for many years to come. Moreover, although useful to investigate, there is no effective intervention that can be offered to those even if they are identified to be at greater risk of AMD, although show few signs of the disease at the point of screening. Although the focus in genetics is moving at a fast pace, what is described here- that is GWAS and generating an inventory of regional variation followed by genotyping efforts to fine map regions with the strongest claims, this is now being overtaken by large scale re-sequencing projects which are not always just tailored to test a hypothesis as the next step in confirming or refuting an association. Therefore focus on the pathways that are highlighted from such studies may be a good first step in attempting to change the clinical course of disease, and that has what has been attempted by this thesis.

8.1 Summary

Evidence emerging from different disciplines in AMD research have converged on the idea that there may be shared factors that contribute to both AMD and CVD. Findings from GWAS in AMD show that complement in particular is likely to play a causal role in AMD. Observational evidence in the cardiovascular arena has previously linked complement with atherosclerosis. In this thesis I therefore set out to establish if complement and in particular complement factor H, the most significant signal from AMD GWAS, was a shared risk factor for AMD and CHD, with a focus first on the most replicated SNP from the AMD studies, and second in establishing if risk is mediated by the encoded protein, circulating fH. In doing so I have for the first time reported the most comprehensive analysis of the rs1061170 SNP encoding Y402H in AMD and CHD showing a consistent link of this putatively functional SNP with AMD, and by collecting a large number of studies have been able to more precisely define the effect size associated with rs1061170. A similar approach in CHD showed no association of rs1061170 with CHD or any conventional risk factors of CHD, supporting the view that *CFH* may not be a shared risk factor for CHD and AMD. However given the strong signal associated with AMD I continued on my line of investigation in trying to establish if the basis of risk in AMD is conferred by alterations in the concentration of circulating fH, which was supported by findings from previous linkage scan of

fH identifying a QTL in the RCA cluster of chromosome 1. In order to address this hypothesis I, with others, developed a high throughput robust assay for measuring circulating fH using an existing commercial assay and scaling up the throughput on a platform that is rapidly translatable into a clinical setting. By establishing this assay I have been able to determine the molecular epidemiology of fH for the first time, taking advantage of a wide number of markers already measured in a population based study. This demonstrated novel associations with circulating lipids and lipoproteins as well as anthropometric measures of body fat. Another advantage in using the Fenland MRC study was the availability of genome wide data (both directly genotyped and imputed against HapMap), again providing for the first time a comprehensive investigation of the genetic determinants of circulating fH. This confirmed the association of SNPs in the *CFH/ CFHR* region as being associated with fH levels, although no signals reached genome wide significance. In addition other loci, largely those that appear to be involved in determination of cell polarity, cyto-skeletal elements and cell growth are also implicated. There is however a lack of signals from other complement related genes. Establishment of the assay for circulating fH allowed me to then address the hypothesis that SNPs associated with AMD may also be associated with alterations of fH and changes in circulating concentrations may predispose to AMD. However no association between SNPs associated with AMD and fH was established, and moreover no association of fH with AMD was found.

8.2 Future work

In order to take the findings forward it would be important to firstly establish the specificity of the commercial polyclonal antibody that has been used in establishing the high throughput assay, refining the signal further down to either fH, FHL-1 or indeed fH related proteins. Secondly it would be important to replicate the findings of the GWAS described in Chapter 6. One approach could be to measure fH in a dataset with genome wide data, so as to jointly analyse two studies thus prioritising signals for further replication. This would be important also in terms of defining the potential overlapping SNPs which regulate fH and are associated with meningococcal disease, which may lead to important insights into biological mechanisms of subtypes of meningococcal disease as well as potential therapeutic pathways. Although a convincing

argument has been presented that fH concentrations are likely to be altered by variation in the gene and related genes, it may be that this is not the whole story, or as discussed already small alterations in concentrations could lead to large functional effects. There is currently no good test to test fH function in population based samples, where one would predict small differences compared to extreme all-or-nothing mutations. It would therefore be of interest to adapt an existing assay of either fH or the alternative complement pathway to a scale and sensitivity that could then be applied to population based cases in order to test this hypothesis. Finally in order to refine even further the genetic signals that have been identified in GWAS and association studies of this region, further studies of ethnically diverse populations with differing LD structures my aid localisation of signals. As a first step however this can be done in silico using the 1000 genome data and imputing current data against panels from other ethnicities where the LD structure is different. Together with targeted re-sequencing this approach may help further refine and identify the causal variants that confer AMD risk.

9 References

1. Allender, S., et al., *Coronary Heart Disease Statistics.* British Heart Foundation, London, 2008.

2. Allender, S., et al., *European Cardiovascular Disease Statisitcs 2008.* British Heart Foundation, London, 2008.

3. World Health Organisation, *The World Health Report 2002: Reducing risks, promoting healthy life. Geneva, WHO.* 2002.

4. Pearson, T.A., et al., *AHA Guidelines for Primary Prevention of Cardiovascular Disease and Stroke: 2002 Update: Consensus Panel Guide to Comprehensive Risk Reduction for Adult Patients Without Coronary or Other Atherosclerotic Vascular Diseases. American Heart Association Science Advisory and Coordinating Committee.* Circulation, 2002. **106**(3): p. 388-91.

5. Law, M.R. and N.J. Wald, *Risk factor thresholds: their existence under scrutiny.* Bmj, 2002. **324**(7353): p. 1570-6.

6. Turnbull, F., *Effects of different blood-pressure-lowering regimens on major cardiovascular events: results of prospectively-designed overviews of randomised trials.* Lancet, 2003. **362**(9395): p. 1527-35.

7. Baigent, C., et al., *Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins.* Lancet, 2005. **366**(9493): p. 1267-78.

8. Ross, R. and J.A. Glomset, *Atherosclerosis and the arterial smooth muscle cell: Proliferation of smooth muscle is a key event in the genesis of the lesions of atherosclerosis.* Science, 1973. **180**(93): p. 1332-9.

9. Libby, P., et al., *Inflammation in Atherosclerosis.* Circ J, 2010.

10. Klein, R.J., et al., *Complement factor H polymorphism in age-related macular degeneration.* Science, 2005. **308**(5720): p. 385-9.

11. Lambert, J.C., et al., *Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease.* Nat Genet, 2009. **41**(10): p. 1094-9.

12. Davila, S., et al., *Genome-wide association study identifies variants in the CFH region associated with host susceptibility to meningococcal disease.* Nat Genet, 2010. **42**(9): p. 772-6.

13. Walport, M.J., *Complement. First of two parts.* N Engl J Med, 2001. **344**(14): p. 1058-66.

14. Walport, M.J., *Complement. Second of two parts.* N Engl J Med, 2001. **344**(15): p. 1140-4.

15. Ricklin, D. and J.D. Lambris, *Complement-targeted therapeutics.* Nat Biotechnol, 2007. **25**(11): p. 1265-75.

16. Pangburn, M.K., R.D. Schreiber, and H.J. Muller-Eberhard, *Formation of the initial C3 convertase of the alternative complement pathway. Acquisition of C3b-like activities by*

spontaneous hydrolysis of the putative thioester in native C3. J Exp Med, 1981. **154**(3): p. 856-67.

17. Oksjoki, R., et al., *Association between complement factor H and proteoglycans in early human coronary atherosclerotic lesions: implications for local regulation of complement activation.* Arterioscler Thromb Vasc Biol, 2003. **23**(4): p. 630-6.

18. Vlaicu, R., et al., *Quantitative determinations of immunoglobulins and complement components in human aortic atherosclerotic wall.* Med Interne, 1985. **23**(1): p. 29-35.

19. Vlaicu, R., et al., *Immunohistochemical localization of the terminal C5b-9 complement complex in human aortic fibrous plaque.* Atherosclerosis, 1985. **57**(2-3): p. 163-77.

20. Seifert, P.S., et al., *Prelesional complement activation in experimental atherosclerosis. Terminal C5b-9 complement deposition coincides with cholesterol accumulation in the aortic intima of hypercholesterolemic rabbits.* Lab Invest, 1989. **60**(6): p. 747-54.

21. Seifert, P.S. and G.K. Hansson, *Decay-accelerating factor is expressed on vascular smooth muscle cells in human atherosclerotic lesions.* J Clin Invest, 1989. **84**(2): p. 597-604.

22. Seifert, P.S., et al., *CD59 (homologous restriction factor 20), a plasma membrane protein that protects against complement C5b-9 attack, in human atherosclerotic lesions.* Atherosclerosis, 1992. **96**(2-3): p. 135-45.

23. Saito, E., et al., *Complement receptors in atherosclerotic lesions.* Artery, 1992. **19**(1): p. 47-62.

24. Laine, P., et al., *Evidence for complement activation in ruptured coronary plaques in acute myocardial infarction.* Am J Cardiol, 2002. **90**(4): p. 404-8.

25. Kostner, K.M., et al., *Inflammation, complement activation and endothelial function in stable and unstable coronary artery disease.* Clin Chim Acta, 2006. **365**(1-2): p. 129-34.

26. Hoffmeister, H.M., et al., *Comparison of C-reactive protein and terminal complement complex in patients with unstable angina pectoris versus stable angina pectoris.* Am J Cardiol, 2002. **89**(8): p. 909-12.

27. Speidl, W.S., et al., *Complement component C5a predicts future cardiovascular events in patients with advanced atherosclerosis.* Eur Heart J, 2005. **26**(21): p. 2294-9.

28. Muscari, A., et al., *Relationship between serum C3 levels and traditional risk factors for myocardial infarction.* Acta Cardiol, 1998. **53**(6): p. 345-54.

29. Iltumur, K., et al., *Complement activation in acute coronary syndromes.* APMIS, 2005. **113**(3): p. 167-74.

30. Pinckard, R.N., et al., *Complement localization and mediation of ischemic injury in baboon myocardium.* J Clin Invest, 1980. **66**(5): p. 1050-6.

31. Griselli, M., et al., *C-reactive protein and complement are important mediators of tissue damage in acute myocardial infarction.* J Exp Med, 1999. **190**(12): p. 1733-40.

32. An, G., et al., *CD59 but not DAF deficiency accelerates atherosclerosis in female ApoE knockout mice.* Mol Immunol, 2009. **46**(8-9): p. 1702-9.

33. Leung, V.W., et al., *Decay-accelerating factor suppresses complement C3 activation and retards atherosclerosis in low-density lipoprotein receptor-deficient mice.* Am J Pathol, 2009. **175**(4): p. 1757-67.

34. Lewington, S., et al., *Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55,000 vascular deaths.* Lancet, 2007. **370**(9602): p. 1829-39.

35. Vaisar, T., et al., *Shotgun proteomics implicates protease inhibition and complement activation in the antiinflammatory properties of HDL.* J Clin Invest, 2007. **117**(3): p. 746-56.

36. Chen, W., et al., *Genetic variants near TIMP3 and high-density lipoprotein-associated loci influence susceptibility to age-related macular degeneration.* Proc Natl Acad Sci U S A, 2010. **107**(16): p. 7401-6.

37. Neale, B.M., et al., *Genome-wide association study of advanced age-related macular degeneration identifies a role of the hepatic lipase gene (LIPC).* Proc Natl Acad Sci U S A, 2010. **107**(16): p. 7395-400.

38. Danesh, J., et al., *Long-term interleukin-6 levels and subsequent risk of coronary heart disease: two new prospective studies and a systematic review.* PLoS Med, 2008. **5**(4): p. e78.

39. Danesh, J., et al., *Association of fibrinogen, C-reactive protein, albumin, or leukocyte count with coronary heart disease: meta-analyses of prospective studies.* JAMA, 1998. **279**(18): p. 1477-82.

40. Smeeth, L., et al., *Risk of myocardial infarction and stroke after acute infection or vaccination.* N Engl J Med, 2004. **351**(25): p. 2611-8.

41. Manzi, S., et al., *Age-specific incidence rates of myocardial infarction and angina in women with systemic lupus erythematosus: comparison with the Framingham Study.* Am J Epidemiol, 1997. **145**(5): p. 408-15.

42. Maradit-Kremers, H., et al., *Increased unrecognized coronary heart disease and sudden deaths in rheumatoid arthritis: a population-based cohort study.* Arthritis Rheum, 2005. **52**(2): p. 402-11.

43. Kaptoge, S., et al., *C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: an individual participant meta-analysis.* Lancet, 2010. **375**(9709): p. 132-40.

44. Davey Smith, G. and S. Ebrahim, *'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?* Int. J. Epidemiol., 2003. **32**(1): p. 1-22.

45. Hingorani, A. and S. Humphries, *Nature's randomised trials.* Lancet, 2005. **366**(9501): p. 1906-8.

46. Hingorani, A.D., T. Shah, and J.P. Casas, *Linking observational and genetic approaches to determine the role of C-reactive protein in heart disease risk.* Eur Heart J, 2006. **27**(11): p. 1261-3.

47. Marchioli, R., et al., *Antioxidant vitamins and prevention of cardiovascular disease: epidemiological and clinical trial data.* Lipids, 2001. **36 Suppl**: p. S53-63.

48. Anker, S.D. and A.J. Coats, *How to RECOVER from RENAISSANCE? The significance of the results of RECOVER, RENAISSANCE, RENEWAL and ATTACH.* Int J Cardiol, 2002. **86**(2-3): p. 123-30.

49. Danesh, J., *Antibiotics in the prevention of heart attacks.* Lancet, 2005. **365**(9457): p. 365-7.

50. Granger, C.B., et al., *Pexelizumab, an anti-C5 complement antibody, as adjunctive therapy to primary percutaneous coronary intervention in acute myocardial infarction: the COMplement inhibition in Myocardial infarction treated with Angioplasty (COMMA) trial.* Circulation, 2003. **108**(10): p. 1184-90.

51. Mahaffey, K.W., et al., *Effect of pexelizumab, an anti-C5 complement antibody, as adjunctive therapy to fibrinolysis in acute myocardial infarction: the COMPlement inhibition in myocardial infarction treated with thromboLYtics (COMPLY) trial.* Circulation, 2003. **108**(10): p. 1176-83.

52. Munos, B., *Lessons from 60 years of pharmaceutical innovation.* Nat Rev Drug Discov, 2009. **8**(12): p. 959-68.

53. Davey Smith, G. and S. Ebrahim, *'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?* Int J Epidemiol, 2003. **32**(1): p. 1-22.

54. Drenos, F., et al., *Integrated associations of genotypes with multiple blood biomarkers linked to coronary heart disease risk.* Hum Mol Genet, 2009. **18**(12): p. 2305-16.

55. Smith, G.D., et al., *Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology.* PLoS Med, 2007. **4**(12): p. e352.

56. Casas, J.P., et al., *Insight into the nature of the CRP-coronary event association using Mendelian randomization.* Int J Epidemiol, 2006. **35**(4): p. 922-31.

57. Elliott, P., et al., *Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease.* JAMA, 2009. **302**(1): p. 37-48.

58. Wensley, F., et al., *Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data.* Bmj, 2011. **342**: p. d548.

59. Kamstrup, P.R., et al., *Genetically elevated lipoprotein(a) and increased risk of myocardial infarction.* JAMA, 2009. **301**(22): p. 2331-9.

60. Resnikoff, S., et al., *Global data on visual impairment in the year 2002.* Bull World Health Organ, 2004. **82**(11): p. 844-51.

61. Klein, R., B.E. Klein, and K.L. Linton, *Prevalence of age-related maculopathy. The Beaver Dam Eye Study.* Ophthalmology, 1992. **99**(6): p. 933-43.

62. Schmier, J.K., M.L. Jones, and M.T. Halpern, *The burden of age-related macular degeneration.* Pharmacoeconomics, 2006. **24**(4): p. 319-34.

63. Berger, J.W., S.L. Fine, and M.G. Maguire, *Age-related macular degeneration*. 1999, St. Louis, Mo. ; London: Mosby. xiii, 463 p.

64. Ryan, S.J. and A.P. Schachat, *Retina*. 3rd ed. ed. 2001, St. Louis ; London: Mosby. xxix, p.875-1847, p.II-1-II-53.

65. Christen, W.G., et al., *A prospective study of cigarette smoking and risk of age-related macular degeneration in men.* JAMA, 1996. **276**(14): p. 1147-51.

66. Klein, R., et al., *Ten-year incidence of age-related maculopathy and smoking and drinking: the Beaver Dam Eye Study.* Am J Epidemiol, 2002. **156**(7): p. 589-98.

67. Seddon, J.M., et al., *A prospective study of cigarette smoking and age-related macular degeneration in women.* JAMA, 1996. **276**(14): p. 1141-6.

68. Smith, W., et al., *Risk factors for age-related macular degeneration: Pooled findings from three continents.* Ophthalmology, 2001. **108**(4): p. 697-704.

69. Tan, J.S., et al., *Smoking and the long-term incidence of cataract: the Blue Mountains Eye Study.* Ophthalmic Epidemiol, 2008. **15**(3): p. 155-61.

70. Klein, R., et al., *The association of cardiovascular disease with the long-term incidence of age-related maculopathy: the Beaver Dam Eye Study.* Ophthalmology, 2003. **110**(6): p. 1273-80.

71. van Leeuwen, R., et al., *Blood pressure, atherosclerosis, and the incidence of age-related maculopathy: the Rotterdam Study.* Invest Ophthalmol Vis Sci, 2003. **44**(9): p. 3771-7.

72. Seddon, J.M., et al., *Progression of age-related macular degeneration: association with body mass index, waist circumference, and waist-hip ratio.* Arch Ophthalmol, 2003. **121**(6): p. 785-92.

73. Mares-Perlman, J.A., et al., *Dietary fat and age-related maculopathy.* Arch Ophthalmol, 1995. **113**(6): p. 743-8.

74. Seddon, J.M., et al., *Dietary carotenoids, vitamins A, C, and E, and advanced age-related macular degeneration. Eye Disease Case-Control Study Group.* JAMA, 1994. **272**(18): p. 1413-20.

75. Tan, A.G., et al., *Antioxidant nutrient intake and the long-term incidence of age-related cataract: the Blue Mountains Eye Study.* Am J Clin Nutr, 2008. **87**(6): p. 1899-905.

76. Hofman, A., et al., *The Rotterdam Study: 2010 objectives and design update.* Eur J Epidemiol, 2009. **24**(9): p. 553-72.

77. Wang, J.J., et al., *Ten-year incidence and progression of age-related maculopathy: the blue Mountains Eye Study.* Ophthalmology, 2007. **114**(1): p. 92-8.

78. Klein, R., et al., *Fifteen-year cumulative incidence of age-related macular degeneration: the Beaver Dam Eye Study.* Ophthalmology, 2007. **114**(2): p. 253-62.

79. *A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E, beta carotene, and zinc for age-related macular degeneration and vision loss: AREDS report no. 8.* Arch Ophthalmol, 2001. **119**(10): p. 1417-36.

80. Nissen, S.E., *ENHANCE and ACCORD: controversy over surrogate end points.* Curr Cardiol Rep, 2008. **10**(3): p. 159-61.

81. Kris-Etherton, P.M., et al., *Antioxidant vitamin supplements and cardiovascular disease.* Circulation, 2004. **110**(5): p. 637-41.

82. Rosenfeld, P.J., et al., *Ranibizumab for neovascular age-related macular degeneration.* N Engl J Med, 2006. **355**(14): p. 1419-31.

83. Gragoudas, E.S., et al., *Pegaptanib for neovascular age-related macular degeneration.* N Engl J Med, 2004. **351**(27): p. 2805-16.

84. Galan, A., et al., *Association of age-related macular degeneration with polymorphisms in vascular endothelial growth factor and its receptor.* Ophthalmology, 2010. **117**(9): p. 1769-74.

85. Haines, J.L., et al., *Functional candidate genes in age-related macular degeneration: significant association with VEGF, VLDLR, and LRP6.* Invest Ophthalmol Vis Sci, 2006. **47**(1): p. 329-35.

86. Boekhoorn, S.S., et al., *Polymorphisms in the vascular endothelial growth factor gene and risk of age-related macular degeneration: the Rotterdam Study.* Ophthalmology, 2008. **115**(11): p. 1899-903.

87. Awata, T., *Vascular endothelial growth factor gene polymorphisms in susceptibility to coronary artery disease.* Am J Hypertens, 2010. **23**(9): p. 938-9.

88. Khurana, R., et al., *Role of angiogenesis in cardiovascular disease: a critical appraisal.* Circulation, 2005. **112**(12): p. 1813-24.

89. Franklin, R.M. and New Orleans Academy of Ophthalmology., *Retina and vitreous : proceedings of the Symposium on Retina and Vitreous, New Orleans, LA, USA, March 12-15, 1992*. 1993, Amsterdam ; New York: Kluger Publications. xvi, 322 p.

90. Donders, F.C., *Beitrage zur pathologischen Anatomie des Auges.* Graefe's Archive for Clinical and Experimental Ophthalmology, 1855. **1**(2): p. 106-118.

91. Hageman, G.S., et al., *An integrated hypothesis that considers drusen as biomarkers of immune-mediated processes at the RPE-Bruch's membrane interface in aging and age-related macular degeneration.* Prog Retin Eye Res, 2001. **20**(6): p. 705-32.

92. Jager, R.D., W.F. Mieler, and J.W. Miller, *Age-related macular degeneration.* N Engl J Med, 2008. **358**(24): p. 2606-17.

93. Pauleikhoff, D., et al., *Drusen as risk factors in age-related macular disease.* Am J Ophthalmol, 1990. **109**(1): p. 38-43.

94. Bird, A.C., et al., *An international classification and grading system for age-related maculopathy and age-related macular degeneration. The International ARM Epidemiological Study Group.* Surv Ophthalmol, 1995. **39**(5): p. 367-74.

95. de Jong, P.T., *Age-related macular degeneration.* N Engl J Med, 2006. **355**(14): p. 1474-85.

96. Ferris, F.L., 3rd, S.L. Fine, and L. Hyman, *Age-related macular degeneration and blindness due to neovascular maculopathy.* Arch Ophthalmol, 1984. **102**(11): p. 1640-2.

97. Mullins, R.F., et al., *Drusen associated with aging and age-related macular degeneration contain proteins common to extracellular deposits associated with*

atherosclerosis, elastosis, amyloidosis, and dense deposit disease. FASEB J, 2000. **14**(7): p. 835-46.

98. Anderson, D.H., et al., *The pivotal role of the complement system in aging and age-related macular degeneration: Hypothesis re-visited.* Prog Retin Eye Res, 2009.

99. Vine, A.K., et al., *Biomarkers of cardiovascular disease as risk factors for age-related macular degeneration.* Ophthalmology, 2005. **112**(12): p. 2076-80.

100. Seddon, J.M., et al., *Progression of age-related macular degeneration: prospective assessment of C-reactive protein, interleukin 6, and other cardiovascular biomarkers.* Arch Ophthalmol, 2005. **123**(6): p. 774-82.

101. Haddad, S., et al., *The genetics of age-related macular degeneration: a review of progress to date.* Surv Ophthalmol, 2006. **51**(4): p. 316-63.

102. *A haplotype map of the human genome.* Nature, 2005. **437**(7063): p. 1299-320.

103. *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* Nature, 2007. **447**(7145): p. 661-78.

104. Edwards, A.O., et al., *Complement factor H polymorphism and age-related macular degeneration.* Science, 2005. **308**(5720): p. 421-4.

105. Hageman, G.S., et al., *A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration.* Proc Natl Acad Sci U S A, 2005. **102**(20): p. 7227-32.

106. Haines, J.L., et al., *Complement factor H variant increases the risk of age-related macular degeneration.* Science, 2005. **308**(5720): p. 419-21.

107. Conley, Y.P., et al., *CFH, ELOVL4, PLEKHA1 and LOC387715 genes and susceptibility to age-related maculopathy: AREDS and CHS cohorts and meta-analyses.* Hum Mol Genet, 2006. **15**(21): p. 3206-18.

108. Magnusson, K.P., et al., *CFH Y402H confers similar risk of soft drusen and both forms of advanced AMD.* PLoS Med, 2006. **3**(1): p. e5.

109. Rivera, A., et al., *Hypothetical LOC387715 is a second major susceptibility gene for age-related macular degeneration, contributing independently of complement factor H to disease risk.* Hum Mol Genet, 2005. **14**(21): p. 3227-36.

110. Zareparsi, S., et al., *Strong association of the Y402H variant in complement factor H at 1q32 with susceptibility to age-related macular degeneration.* Am J Hum Genet, 2005. **77**(1): p. 149-53.

111. Gold, B., et al., *Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration.* Nat Genet, 2006. **38**(4): p. 458-62.

112. Yates, J.R., et al., *Complement C3 variant and the risk of age-related macular degeneration.* N Engl J Med, 2007. **357**(6): p. 553-61.

113. Ennis, S., et al., *Association between the SERPING1 gene and age-related macular degeneration: a two-stage case-control study.* Lancet, 2008. **372**(9652): p. 1828-34.

114. Hughes, A.E., et al., *A common CFH haplotype, with deletion of CFHR1 and CFHR3, is associated with lower risk of age-related macular degeneration.* Nat Genet, 2006. **38**(10): p. 1173-7.

115. Thompson, A., et al., *Association of cholesteryl ester transfer protein genotypes with CETP mass and activity, lipid levels, and coronary risk.* JAMA, 2008. **299**(23): p. 2777-88.

116. Goverdhan, S.V., et al., *Complement factor H Y402H gene polymorphism in coronary artery disease and atherosclerosis.* Atherosclerosis, 2006. **188**(1): p. 213-4.

117. Kardys, I., et al., *A common polymorphism in the complement factor H gene is associated with increased risk of myocardial infarction: the Rotterdam Study.* J Am Coll Cardiol, 2006. **47**(8): p. 1568-75.

118. Smith, G.D., et al., *Does elevated plasma fibrinogen increase the risk of coronary heart disease? Evidence from a meta-analysis of genetic association studies.* Arterioscler Thromb Vasc Biol, 2005. **25**(10): p. 2228-33.

119. Melzer, D., et al., *A genome-wide association study identifies protein quantitative trait loci (pQTLs).* PLoS Genet, 2008. **4**(5): p. e1000072.

120. Hindroff, L.A., et al., *A Catalog of Published Genome-Wide Association Studies. .* Available at www.genome.gov/gwastudies, Accessed January 2010.

121. Duan, S., et al., *Genetic architecture of transcript-level variation in humans.* Am J Hum Genet, 2008. **82**(5): p. 1101-13.

122. Nilsson, U.R. and H.J. Mueller-Eberhard, *Isolation of Beta If-Globulin from Human Serum and Its Characterization as the Fifth Component of Complement.* J Exp Med, 1965. **122**: p. 277-98.

123. Sharma, A.K. and M.K. Pangburn, *Identification of three physically and functionally distinct binding sites for C3b in human complement factor H by deletion mutagenesis.* Proc Natl Acad Sci U S A, 1996. **93**(20): p. 10996-1001.

124. Pangburn, M.K., et al., *Molecular mechanisms of target recognition in an innate immune system: interactions among factor H, C3b, and target in the alternative pathway of human complement.* J Immunol, 2000. **164**(9): p. 4742-51.

125. Pangburn, M.K., *Cutting edge: localization of the host recognition functions of complement factor H at the carboxyl-terminal: implications for hemolytic uremic syndrome.* J Immunol, 2002. **169**(9): p. 4702-6.

126. Giannakis, E., et al., *A common site within factor H SCR 7 responsible for binding heparin, C-reactive protein and streptococcal M protein.* Eur J Immunol, 2003. **33**(4): p. 962-9.

127. Rodriguez de Cordoba, S., et al., *The human complement factor H: functional roles, genetic variations and disease associations.* Mol Immunol, 2004. **41**(4): p. 355-67.

128. Horstmann, R.D., et al., *Antiphagocytic activity of streptococcal M protein: selective binding of complement control protein factor H.* Proc Natl Acad Sci U S A, 1988. **85**(5): p. 1657-61.

129. Schneider, M.C., et al., *Functional significance of factor H binding to Neisseria meningitidis.* J Immunol, 2006. **176**(12): p. 7566-75.

130. Hellwage, J., et al., *The complement regulator factor H binds to the surface protein OspE of Borrelia burgdorferi.* J Biol Chem, 2001. **276**(11): p. 8427-35.

131. Pickering, M.C., et al., *Uncontrolled C3 activation causes membranoproliferative glomerulonephritis in mice deficient in complement factor H.* Nat Genet, 2002. **31**(4): p. 424-8.

132. Malek, G., et al., *Apolipoprotein E allele-dependent pathogenesis: a model for age-related retinal degeneration.* Proc Natl Acad Sci U S A, 2005. **102**(33): p. 11900-5.

133. Schwaeble, W., et al., *Human complement factor H: expression of an additional truncated gene product of 43 kDa in human liver.* Eur J Immunol, 1987. **17**(10): p. 1485-9.

134. Esparza-Gordillo, J., et al., *Genetic and environmental factors influencing the human factor H plasma levels.* Immunogenetics, 2004. **56**(2): p. 77-82.

135. Mihlan, M., et al., *Human complement factor H-related protein 4 binds and recruits native pentameric C-reactive protein to necrotic cells.* Mol Immunol, 2009. **46**(3): p. 335-44.

136. Heinen, S., et al., *Factor H-related protein 1 (CFHR-1) inhibits complement C5 convertase activity and terminal complex formation.* Blood, 2009. **114**(12): p. 2439-47.

137. Gale, D.P., et al., *Identification of a mutation in complement factor H-related protein 5 in patients of Cypriot origin with glomerulonephritis.* Lancet, 2010. **376**(9743): p. 794-801.

138. Seddon, J.M., et al., *Prediction model for prevalence and incidence of advanced age-related macular degeneration based on genetic, demographic, and environmental variables.* Invest Ophthalmol Vis Sci, 2009. **50**(5): p. 2044-53.

139. Ioannidis, J.P. and T.A. Trikalinos, *Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials.* J Clin Epidemiol, 2005. **58**(6): p. 543-9.

140. Thakkinstian, A., et al., *Systematic review and meta-analysis of the association between complement factor H Y402H polymorphisms and age-related macular degeneration.* Hum Mol Genet, 2006. **15**(18): p. 2784-90.

141. Augood, C., et al., *Methods for a population-based study of the prevalence of and risk factors for age-related maculopathy and macular degeneration in elderly European populations: the EUREYE study.* Ophthalmic Epidemiol, 2004. **11**(2): p. 117-29.

142. Ioannidis, J.P., et al., *Assessment of cumulative evidence on genetic associations: interim guidelines.* Int J Epidemiol, 2008. **37**(1): p. 120-32.

143. Clayton, D. and P.M. McKeigue, *Epidemiological methods for studying genes and environmental factors in complex diseases.* Lancet, 2001. **358**(9290): p. 1356-60.

144. Cox, D.R. and E.J. Snell, *Analysis of Binary Data*. Second Edition ed. 1989: Taylor & Francis.

145. Baird, P.N., et al., *Analysis of the Y402H variant of the complement factor H gene in age-related macular degeneration.* Invest Ophthalmol Vis Sci, 2006. **47**(10): p. 4194-8.

146. Conley, Y.P., et al., *Candidate gene analysis suggests a role for fatty acid biosynthesis and regulation of the complement system in the etiology of age-related maculopathy.* Hum Mol Genet, 2005. **14**(14): p. 1991-2002.

147. Despriet, D.D., et al., *Complement factor H polymorphism, complement activators, and risk of age-related macular degeneration.* Jama, 2006. **296**(3): p. 301-9.

148. Droz, I., et al., *Genotype-phenotype correlation of age-related macular degeneration: influence of complement factor H polymorphism.* Br J Ophthalmol, 2008. **92**(4): p. 513-7.

149. Fisher, S.A., et al., *Meta-analysis of genome scans of age-related macular degeneration.* Hum Mol Genet, 2005. **14**(15): p. 2257-64.

150. Jakobsdottir, J., et al., *Susceptibility genes for age-related maculopathy on chromosome 10q26.* Am J Hum Genet, 2005. **77**(3): p. 389-407.

151. Schaumberg, D.A., et al., *A Prospective Assessment of the Y402H Variant in Complement Factor H, Genetic Variants in C-Reactive Protein, and Risk of Age-Related Macular Degeneration.* Invest Ophthalmol Vis Sci, 2006. **47**(6): p. 2336-40.

152. Schaumberg, D.A., et al., *A prospective study of 2 major age-related macular degeneration susceptibility alleles and interactions with modifiable risk factors.* Arch Ophthalmol, 2007. **125**(1): p. 55-62.

153. Seitsonen, S., et al., *Analysis of variants in the complement factor H, the elongation of very long chain fatty acids-like 4 and the hemicentin 1 genes of age-related macular degeneration in the Finnish population.* Mol Vis, 2006. **12**: p. 796-801.

154. Sepp, T., et al., *Complement factor H variant Y402H is a major risk determinant for geographic atrophy and choroidal neovascularization in smokers and nonsmokers.* Invest Ophthalmol Vis Sci, 2006. **47**(2): p. 536-40.

155. Simonelli, F., et al., *Polymorphism p.402Y>H in the complement factor H protein is a risk factor for age related macular degeneration in an Italian population.* Br J Ophthalmol, 2006. **90**(9): p. 1142-5.

156. Souied, E.H., et al., *Y402H complement factor H polymorphism associated with exudative age-related macular degeneration in the French population.* Mol Vis, 2005. **11**: p. 1135-40.

157. Weger, M., et al., *Association of the HTRA1 -625G>A promoter gene polymorphism with exudative age-related macular degeneration in a Central European population.* Mol Vis, 2007. **13**: p. 1274-9.

158. Ennis, S., et al., *Fine-scale linkage disequilibrium mapping of age-related macular degeneration in the complement factor H gene region.* Br J Ophthalmol, 2007. **91**(7): p. 966-70.

159. Fuse, N., et al., *Polymorphisms in Complement Factor H and Hemicentin-1 genes in a Japanese population with dry-type age-related macular degeneration.* Am J Ophthalmol, 2006. **142**(6): p. 1074-6.

160. Gotoh, N., et al., *No association between complement factor H gene polymorphism and exudative age-related macular degeneration in Japanese.* Hum Genet, 2006. **120**(1): p. 139-43.

161. Mori, K., et al., *Coding and noncoding variants in the CFH gene and cigarette smoking influence the risk of age-related macular degeneration in a Japanese population.* Invest Ophthalmol Vis Sci, 2007. **48**(11): p. 5315-9.

162. Okamoto, H., et al., *Complement factor H polymorphisms in Japanese population with age-related macular degeneration.* Mol Vis, 2006. **12**: p. 156-8.

163. Tanimoto, S., et al., *A polymorphism of LOC387715 gene is associated with age-related macular degeneration in the Japanese population.* Neurosci Lett, 2007. **414**(1): p. 71-4.

164. Uka, J., et al., *No association of complement factor H gene polymorphism and age-related macular degeneration in the Japanese population.* Retina, 2006. **26**(9): p. 985-7.

165. Chen, L.J., et al., *Association of complement factor H polymorphisms with exudative age-related macular degeneration.* Mol Vis, 2006. **12**: p. 1536-42.

166. Lau, L.I., et al., *Association of the Y402H polymorphism in complement factor H gene and neovascular age-related macular degeneration in Chinese patients.* Invest Ophthalmol Vis Sci, 2006. **47**(8): p. 3242-6.

167. Lin, J.M., et al., *Complement factor H variant increases the risk for early age-related macular degeneration.* Retina, 2008. **28**(10): p. 1416-20.

168. Lin, J.M., et al., *Vascular endothelial growth factor gene polymorphisms in age-related macular degeneration.* Am J Ophthalmol, 2008. **145**(6): p. 1045-1051.

169. Ng, T.K., et al., *Multiple gene polymorphisms in the complement factor h gene are associated with exudative age-related macular degeneration in chinese.* Invest Ophthalmol Vis Sci, 2008. **49**(8): p. 3312-7.

170. Xu, Y., et al., *Association of CFH, LOC387715, and HTRA1 polymorphisms with exudative age-related macular degeneration in a northern Chinese population.* Mol Vis, 2008. **14**: p. 1373-81.

171. Kim, N.R., et al., *Association between complement factor H gene polymorphisms and neovascular age-related macular degeneration in Koreans.* Invest Ophthalmol Vis Sci, 2008. **49**(5): p. 2071-6.

172. Kaur, I., et al., *Analysis of CFH, TLR4, and APOE polymorphism in India suggests the Tyr402His variant of CFH to be a global marker for age-related macular degeneration.* Invest Ophthalmol Vis Sci, 2006. **47**(9): p. 3729-35.

173. Tedeschi-Blok, N., et al., *Population-based study of early age-related macular degeneration: role of the complement factor H Y402H polymorphism in bilateral but not unilateral disease.* Ophthalmology, 2007. **114**(1): p. 99-103.

174. Ziskind, A., et al., *The frequency of the H402 allele of CFH and its involvement with age-related maculopathy in an aged Black African Xhosa population.* Ophthalmic Genet, 2008. **29**(3): p. 117-9.

175. Klein, R., et al., *The Wisconsin age-related maculopathy grading system.* Ophthalmology, 1991. **98**(7): p. 1128-34.

176. van Leeuwen, R., et al., *The risk and natural course of age-related maculopathy: follow-up at 6 1/2 years in the Rotterdam study.* Arch Ophthalmol, 2003. **121**(4): p. 519-26.

177. Hakobyan, S., et al., *Complement factor H binds to denatured rather than to native pentameric C-reactive protein.* J Biol Chem, 2008. **283**(45): p. 30451-60.

178. Johnson, P.T., et al., *Individuals homozygous for the age-related macular degeneration risk-conferring variant of complement factor H have elevated levels of CRP in the choroid.* Proc Natl Acad Sci U S A, 2006. **103**(46): p. 17456-61.

179. Sun, C., R. Klein, and T.Y. Wong, *Age-related macular degeneration and risk of coronary heart disease and stroke: the Cardiovascular Health Study.* Ophthalmology, 2009. **116**(10): p. 1913-9.

180. Tan, J.S., et al., *Cardiovascular risk factors and the long-term incidence of age-related macular degeneration: the Blue Mountains Eye Study.* Ophthalmology, 2007. **114**(6): p. 1143-50.

181. Wong, T.Y., et al., *Age-related macular degeneration and risk of coronary heart disease: the Atherosclerosis Risk in Communities Study.* Ophthalmology, 2007. **114**(1): p. 86-91.

182. Kardys, I., et al., *Usefulness of combining complement factor H and C-reactive protein genetic profiles for predicting myocardial infarction (from the Rotterdam Study).* Am J Cardiol, 2007. **100**(4): p. 646-8.

183. Miller, G.J., et al., *Increased activation of the haemostatic system in men at high risk of fatal coronary heart disease.* Thromb Haemost, 1996. **75**(5): p. 767-71.

184. Stephens, J.W., et al., *An interaction between the interleukin-6 -174G>C gene variant and urinary protein excretion influences plasma oxidative stress in subjects with type 2 diabetes.* Cardiovasc Diabetol, 2004. **3**: p. 2.

185. Ireland, H., et al., *EPCR Ser219Gly: elevated sEPCR, prothrombin F1+2, risk for coronary heart disease, and increased sEPCR shedding in vitro.* Atherosclerosis, 2005. **183**(2): p. 283-92.

186. Juhan-Vague, I., et al., *The plasminogen activator inhibitor-1 -675 4G/5G genotype influences the risk of myocardial infarction associated with elevated plasma proinsulin and insulin concentrations in men from Europe: the HIFMECH study.* J Thromb Haemost, 2003. **1**(11): p. 2322-9.

187. Marmot, M. and E. Brunner, *Cohort Profile: the Whitehall II study.* Int J Epidemiol, 2005. **34**(2): p. 251-6.

188. Marmot, M.G., *Health, wealth and lifestyles of the older population in England : the 2002 English Longitudinal Study of Ageing*. 2003, London: Institute for Fiscal Studies. ix, 374 p.

189. Lawlor, D.A., et al., *The association of the PON1 Q192R polymorphism with coronary heart disease: findings from the British Women's Heart and Health cohort study and a meta-analysis.* BMC Genet, 2004. **5**: p. 17.

190. Keating, B.J., et al., *Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies.* PLoS One, 2008. **3**(10): p. e3583.

191. Mooijaart, S.P., et al., *Complement Factor H polymorphism Y402H associates with inflammation, visual acuity, and cardiovascular mortality in the elderly population at large.* Exp Gerontol, 2007. **42**(11): p. 1116-22.

192. Nicaud, V., et al., *Lack of association between complement factor H polymorphisms and coronary artery disease or myocardial infarction.* J Mol Med, 2007. **85**(7): p. 771-5.

193. Pai, J.K., et al., *Complement factor H (Y402H) polymorphism and risk of coronary heart disease in US men and women.* Eur Heart J, 2007. **28**(11): p. 1297-303.

194. Pulido, J.S., et al., *Relationship between age-related macular degeneration-associated variants of complement factor H and LOC387715 with coronary artery disease.* Mayo Clin Proc, 2007. **82**(3): p. 301-7.

195. Stark, K., et al., *The common Y402H variant in complement factor H gene is not associated with susceptibility to myocardial infarction and its related risk factors.* Clin Sci (Lond), 2007. **113**(4): p. 213-8.

196. Zee, R.Y., K.A. Diehl, and P.M. Ridker, *Complement factor H Y402H gene polymorphism, C-reactive protein, and risk of incident myocardial infarction, ischaemic stroke, and venous thromboembolism: a nested case-control study.* Atherosclerosis, 2006. **187**(2): p. 332-5.

197. Meng, W., et al., *Genetic variants of complement factor H gene are not associated with premature coronary heart disease: a family-based study in the Irish population.* BMC Med Genet, 2007. **8**: p. 62.

198. Volcik, K.A., et al., *Association of the complement factor H Y402H polymorphism with cardiovascular disease is dependent upon hypertension status: The ARIC study.* Am J Hypertens, 2008. **21**(5): p. 533-8.

199. Lawlor, D.A., et al., *The association of C-reactive protein and CRP genotype with coronary heart disease: findings from five studies with 4,610 cases amongst 18,637 participants.* PLoS One, 2008. **3**(8): p. e3011.

200. Schunkert, H., et al., *Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease.* Nat Genet, 2011.

201. Sim, E., et al., *Monoclonal antibodies against the complement control protein factor H (beta 1 H).* Biosci Rep, 1983. **3**(12): p. 1119-31.

202. Bland, J.M. and D.G. Altman, *Statistical methods for assessing agreement between two methods of clinical measurement.* Lancet, 1986. **1**(8476): p. 307-10.

203. Hakobyan, S., et al., *Measurement of factor H variants in plasma using variant-specific monoclonal antibodies: application to assessing risk of age-related macular degeneration.* Invest Ophthalmol Vis Sci, 2008. **49**(5): p. 1983-90.

204. Bland, J.M. and D.G. Altman, *Measuring agreement in method comparison studies.* Stat Methods Med Res, 1999. **8**(2): p. 135-60.

205. Scholl, H.P., et al., *Systemic complement activation in age-related macular degeneration.* PLoS One, 2008. **3**(7): p. e2593.

206. Moreno-Navarrete, J.M., et al., *Complement Factor H is expressed in adipose tissue in association with insulin resistance.* Diabetes, 2009.

207. Licht, C., et al., *Successful plasma therapy for atypical hemolytic uremic syndrome caused by factor H deficiency owing to a novel mutation in the complement cofactor protein domain 15.* Am J Kidney Dis, 2005. **45**(2): p. 415-21.

208. Saxena, R., et al., *Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge.* Nat Genet, 2010. **42**(2): p. 142-8.

209. Marchini, J., et al., *A new multipoint method for genome-wide association studies by imputation of genotypes.* Nat Genet, 2007. **39**(7): p. 906-13.

210. Nishita, M., et al., *Cell/tissue-tropic functions of Wnt5a signaling in normal and cancer cells.* Trends Cell Biol, 2010. **20**(6): p. 346-54.

211. Wang, Y., et al., *A family of RIM-binding proteins regulated by alternative splicing: Implications for the genesis of synaptic active zones.* Proc Natl Acad Sci U S A, 2002. **99**(22): p. 14464-9.

212. Uhl, G.R., et al., *Molecular genetics of successful smoking cessation: convergent genome-wide association study results.* Arch Gen Psychiatry, 2008. **65**(6): p. 683-93.

213. Kim, T.Y., et al., *Role of DLC-1, a tumor suppressor protein with RhoGAP activity, in regulation of the cytoskeleton and cell motility.* Cancer Metastasis Rev, 2009. **28**(1-2): p. 77-83.

214. Liao, Y.C. and S.H. Lo, *Deleted in liver cancer-1 (DLC-1): a tumor suppressor not just for liver.* Int J Biochem Cell Biol, 2008. **40**(5): p. 843-7.

215. Helgadottir, A., et al., *A common variant on chromosome 9p21 affects the risk of myocardial infarction.* Science, 2007. **316**(5830): p. 1491-3.

216. McPherson, R., et al., *A common allele on chromosome 9 associated with coronary heart disease.* Science, 2007. **316**(5830): p. 1488-91.

217. Saxena, R., et al., *Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels.* Science, 2007. **316**(5829): p. 1331-6.

218. Zeggini, E., et al., *Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes.* Science, 2007. **316**(5829): p. 1336-41.

219. Levine, T.D., et al., *Hel-N1: an autoimmune RNA-binding protein with specificity for 3' uridylate-rich untranslated regions of growth factor mRNAs.* Mol Cell Biol, 1993. **13**(6): p. 3494-504.

220. D'Alessandro, V., et al., *Molecular detection of neuron-specific ELAV-like-positive cells in the peripheral blood of patients with small-cell lung cancer.* Cell Oncol, 2008. **30**(4): p. 291-7.

221. Shan, Z., T. Parker, and J.S. Wiest, *Identifying novel homozygous deletions by microsatellite analysis and characterization of tumor suppressor candidate 1 gene, TUSC1, on chromosome 9p in human lung cancer.* Oncogene, 2004. **23**(39): p. 6612-20.

222. Allen, P.B., et al., *Phactrs 1-4: A family of protein phosphatase 1 and actin regulatory proteins.* Proc Natl Acad Sci U S A, 2004. **101**(18): p. 7187-92.

223. Kathiresan, S., et al., *Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants.* Nat Genet, 2009. **41**(3): p. 334-41.

224. Sato, N., et al., *Activation of an oncogenic TBC1D7 (TBC1 domain family, member 7) protein in pulmonary carcinogenesis.* Genes Chromosomes Cancer, 2010. **49**(4): p. 353-67.

225. Holt, L.J. and K. Siddle, *Grb10 and Grb14: enigmatic regulators of insulin action--and more?* Biochem J, 2005. **388**(Pt 2): p. 393-406.

226. Gordon, G.J., et al., *Four-gene expression ratio test for survival in patients undergoing surgery for mesothelioma.* J Natl Cancer Inst, 2009. **101**(9): p. 678-86.

227. Teslovich, T.M., et al., *Biological, clinical and population relevance of 95 loci for blood lipids.* Nature, 2010. **466**(7307): p. 707-13.

228. Redon, R., et al., *Global variation in copy number in the human genome.* Nature, 2006. **444**(7118): p. 444-54.

229. Howie, B.N., P. Donnelly, and J. Marchini, *A flexible and accurate genotype imputation method for the next generation of genome-wide association studies.* PLoS Genet, 2009. **5**(6): p. e1000529.

230. de Bakker, P.I., et al., *Practical aspects of imputation-driven meta-analysis of genome-wide association studies.* Hum Mol Genet, 2008. **17**(R2): p. R122-8.

231. Maher, B., *Personal genomes: The case of the missing heritability.* Nature, 2008. **456**(7218): p. 18-21.

232. Barrett, J.C., et al., *Haploview: analysis and visualization of LD and haplotype maps.* Bioinformatics, 2005. **21**(2): p. 263-5.

233. Maller, J., et al., *Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration.* Nat Genet, 2006. **38**(9): p. 1055-9.

234. Pickering, M.C., et al., *Spontaneous hemolytic uremic syndrome triggered by complement factor H lacking surface recognition domains.* J Exp Med, 2007. **204**(6): p. 1249-56.

235. Hakobyan, S., et al., *Variant-specific quantification of factor H in plasma identifies null alleles associated with atypical hemolytic uremic syndrome.* Kidney Int, 2010. **78**(8): p. 782-8.

236. Haralambous, E., et al., *Factor H, a regulator of complement activity, is a major determinant of meningococcal disease susceptibility in UK Caucasian patients.* Scand J Infect Dis, 2006. **38**(9): p. 764-71.

237. Nischler, C., et al., *Complement factor H Y402H gene polymorphism and response to intravitreal bevacizumab in exudative age-related macular degeneration.* Acta Ophthalmol, 2011.

238. Kutalik, Z., et al., *Methods for testing association between uncertain genotypes and quantitative traits.* Biostatistics, 2011. **12**(1): p. 1-17.

239. Lee, A.Y., et al., *Pharmacogenetics of complement factor H (Y402H) and treatment of exudative age-related macular degeneration with ranibizumab.* Br J Ophthalmol, 2009. **93**(5): p. 610-3.

240. McCarthy, M.I., et al., *Genome-wide association studies for complex traits: consensus, uncertainty and challenges.* Nat Rev Genet, 2008. **9**(5): p. 356-69.

241. Pritchard, J.K. and N.J. Cox, *The allelic architecture of human disease genes: common disease-common variant...or not?* Hum Mol Genet, 2002. **11**(20): p. 2417-23.

242. Voight, B.F., et al., *Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis.* Nat Genet, 2010. **42**(7): p. 579-89.

243. Anderson, C.A., et al., *Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47.* Nat Genet, 2011. **43**(3): p. 246-52.

244. Perola, M., et al., *Combined genome scans for body stature in 6,602 European twins: evidence for common Caucasian loci.* PLoS Genet, 2007. **3**(6): p. e97.

245. Lango Allen, H., et al., *Hundreds of variants clustered in genomic loci and biological pathways affect human height.* Nature, 2010. **467**(7317): p. 832-8.

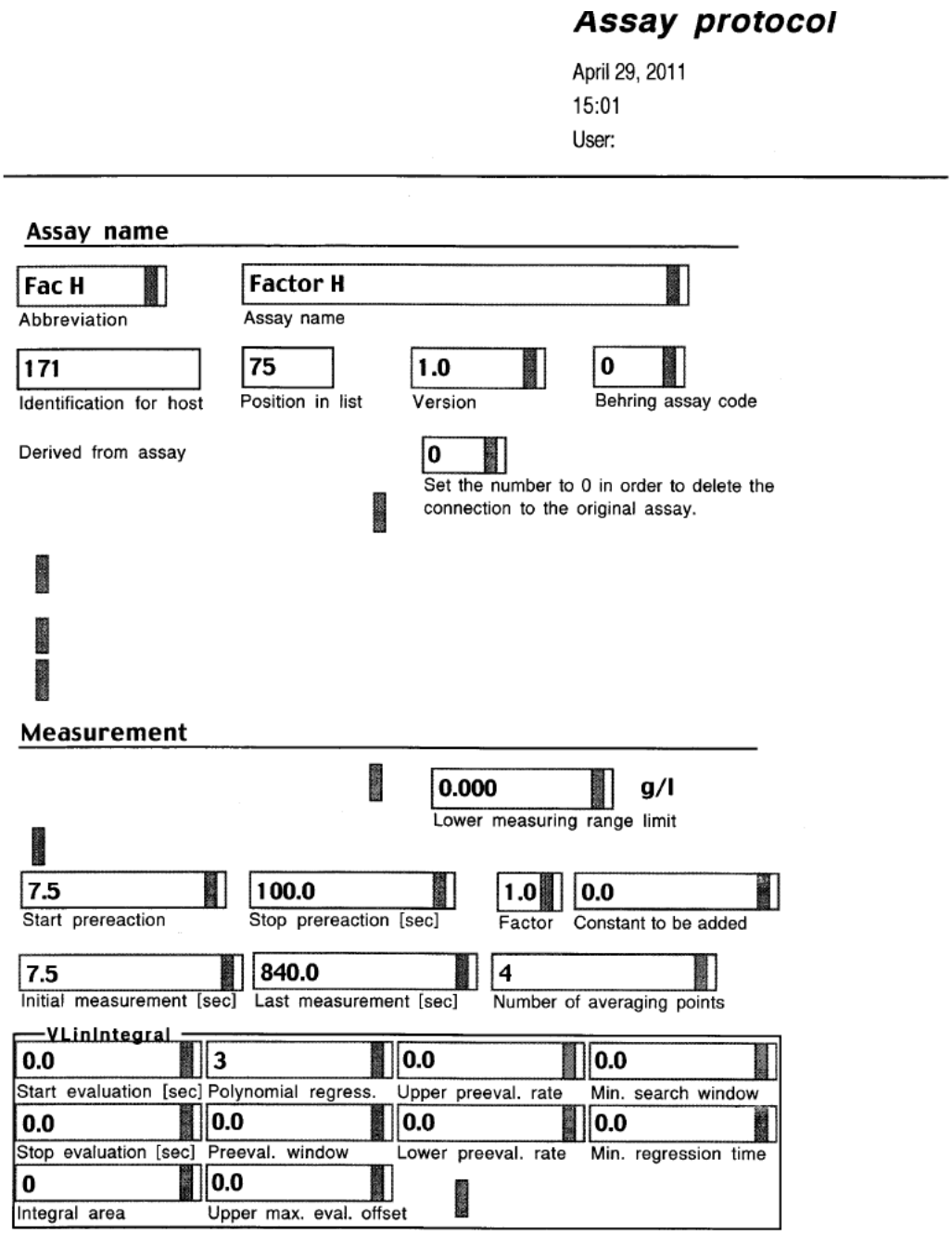
246. Fisher, R.A., *The correlation between relatives on the supposition of Mendelian inheritance.* Transactions from the Royal Society of Edinburgh, 1918. **52**: p. 34.

247. Rose, G., *Sick individuals and sick populations. 1985.* Bull World Health Organ, 2001. **79**(10): p. 990-6.

248. Jakobsdottir, J., et al., *Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers.* PLoS Genet, 2009. **5**(2): p. e1000337.

10 Appendices

10.1 Appendix 1: Assay protocol for Binding Site RID reagents on BNII Nephelometer



Assay protocol

April 29, 2011
15:01
User:

Assay name

Fac H	Factor H
Abbreviation	Assay name

171	75	1.0	0
Identification for host	Position in list	Version	Behring assay code

Derived from assay: 0 — Set the number to 0 in order to delete the connection to the original assay.

Measurement

0.000 g/l — Lower measuring range limit

7.5	100.0	1.0	0.0
Start prereaction	Stop prereaction [sec]	Factor	Constant to be added

7.5	840.0	4
Initial measurement [sec]	Last measurement [sec]	Number of averaging points

VLinIntegral

0.0	3	0.0	0.0
Start evaluation [sec]	Polynomial regress.	Upper preeval. rate	Min. search window
0.0	0.0	0.0	0.0
Stop evaluation [sec]	Preeval. window	Lower preeval. rate	Min. regression time
0	0.0		
Integral area	Upper max. eval. offset		

Remeasurement

Remeasurement in

0.000	0.000	5
Lower remeasurement limit	Upper remeasurement limit	max. no. of remeas.

Result

2	1.000000	1.000000	1.000000
no. of digits after decimal point	Conversion factor from mg/l to IU/l	Conversion factor from mg/l to U/l	Conversion factor from mg/l to mol

Sample dilution

Turbidity check

0 Bit	0 %	0 Bit
Turbidity threshold	Turbidity factor	Upper limit of turbidity check

If you do not wish the turbidity check to be performed, set all values to 0.

Washing

10000	1000000
Cuvette contamination	Dilution probe contamination
500	100
Cuvette washing intensity	Dilution probe rinsing intensity

Assay protocol

April 29, 2011
15:01
User:

Technical Parameters

Transfer: 1, Transfer arm: right
Probe cleaning intensity: 1000000, Dispensing probe rinsing intensity:100
Washing program no.: 1
Mixing time [sec]: 0.00
Buffer
Volume[µl]: 80, Dispensing program no.: 4
Sample
Volume[µl]: 20, Dispensing program no.: 2
Dispense
Dispensing program no.: 7
Transfer: 2, Transfer arm: left
Probe cleaning intensity: 1000000, Dispensing probe rinsing intensity:100
Washing program no.: 1
Mixing time [sec]: 0.60
Start measurement after this transfer step
Buffer
Volume[µl]: 80, Dispensing program no.: 1
Reagent, 171, Influences reference curve
Volume[µl]: 40, Dispensing program no.: 2
Dispense
Dispensing program no.: 1

10.2 Appendix 2: Top 100 signals from imputation analysis for late AMD

Table shows signals that are not shown in Table 7.7, indicating that the top 100 signals are in and around the *CFH* gene.

Rank	rs-ID	BP Position	P value	Gene	Predicted function of SNP
33	rs9427913	195032090	1.00×10^{-09}	*CFHR3*	3downstream
34	1-194938293	194938293	1.10×10^{-09}	*Intergenic*	
35	rs10922092	194924037	1.10×10^{-09}	*CFH*	intronic
36	rs10922099	194931664	1.10×10^{-09}	*CFH*	intronic
37	rs10922100	194931831	1.10×10^{-09}	*CFH*	intronic
38	rs1329422	194921903	1.10×10^{-09}	*CFH*	intronic
39	rs1329424	194912799	1.10×10^{-09}	*CFH*	intronic
40	rs3645	194929639	1.10×10^{-09}	*CFH*	intronic
41	rs528298	194927618	1.10×10^{-09}	*CFH*	intronic
42	rs528298	194927618	1.10×10^{-09}	*CFH*	3downstream
43	rs2860102	194934942	1.20×10^{-09}	*CFH*	intronic
44	rs10922114	195008412	1.30×10^{-09}	*Intergenic*	
45	rs10801553	194922366	1.40×10^{-09}	*CFH*	intronic
46	rs10801554	194924278	1.40×10^{-09}	*CFH*	intronic
47	rs10801557	194936204	1.40×10^{-09}	*CFH*	intronic
48	rs10922094	194928128	1.40×10^{-09}	*CFH*	3downstream
49	rs10922098	194931274	1.40×10^{-09}	*CFH*	intronic
50	rs10922101	194932667	1.40×10^{-09}	*CFH*	intronic
51	rs10922102	194934910	1.40×10^{-09}	*CFH*	intronic
52	rs10922103	194937741	1.40×10^{-09}	*CFH*	3downstream
53	rs12029785	194925528	1.40×10^{-09}	*CFH*	intronic
54	rs12033127	194930878	1.40×10^{-09}	*CFH*	intronic
55	rs12038333	194939077	1.40×10^{-09}	*CFH*	3downstream
56	rs12038674	194932274	1.40×10^{-09}	*CFH*	intronic
57	rs12045503	194939096	1.40×10^{-09}	*CFH*	3downstream
58	rs1329421	194922828	1.40×10^{-09}	*CFH*	intronic
59	rs203688	194939008	1.40×10^{-09}	*CFH*	intronic
60	rs2300430	194922336	1.40×10^{-09}	*CFH*	intronic
61	rs28664709	194937840	1.40×10^{-09}	*CFH*	3downstream
62	rs28853072	194933711	1.40×10^{-09}	*CFH*	intronic
63	rs36137052	194933636	1.40×10^{-09}	*CFH*	intronic
64	rs368465	194938604	1.40×10^{-09}	*CFH*	3downstream
65	rs402056	194938609	1.40×10^{-09}	*CFH*	intronic
66	rs485632	194931416	1.40×10^{-09}	*CFH*	intronic
67	rs488380	194931128	1.40×10^{-09}	*CFH*	intronic
68	rs529899	194934051	1.40×10^{-09}	*CFH*	intronic
69	rs570618	194923687	1.40×10^{-09}	*CFH*	intronic
70	rs579745	194931199	1.40×10^{-09}	*CFH*	intronic

Rank	rs-ID	BP Position	P value	Gene	Predicted function of SNP
71	rs7546015	194930705	1.40x 10^{-09}	*CFH*	intronic
72	rs10922094	194928128	1.40x 10^{-09}	*CFH*	intronic
73	rs10922103	194937741	1.40x 10^{-09}	*CFH*	intronic
74	rs12038333	194939077	1.40x 10^{-09}	*CFH*	intronic
75	rs12045503	194939096	1.40x 10^{-09}	*CFH*	intronic
76	rs203688	194939008	1.40x 10^{-09}	*CFH*	3downstream
77	rs28664709	194937840	1.40x 10^{-09}	*CFH*	intronic
78	rs368465	194938604	1.40x 10^{-09}	*CFH*	intronic
79	rs402056	194938609	1.40x 10^{-09}	*CFH*	3downstream
80	rs10922111	195083436	1.50x 10^{-09}	*Intergenic*	
81	rs10922112	194998766	1.50x 10^{-09}	*Intergenic*	
82	rs4658046	194937380	1.50x 10^{-09}	*CFH*	3downstream
83	rs4658046	194937380	1.50x 10^{-09}	*CFH*	intronic
84	rs10754199	194937462	1.60x 10^{-09}	*CFH*	3downstream
85	rs1887973	194941802	1.60x 10^{-09}	*CFH*	intronic
86	rs203674	194951248	1.60x 10^{-09}	*CFH*	intronic
87	rs203677	194949989	1.60x 10^{-09}	*CFH*	intronic
88	rs203679	194949141	1.60x 10^{-09}	*CFH*	intronic
89	rs203680	194947998	1.60x 10^{-09}	*CFH*	intronic
90	rs203683	194945055	1.60x 10^{-09}	*CFH*	intronic
91	rs203685	194944568	1.60x 10^{-09}	*CFH*	intronic
92	rs399469	194942728	1.60x 10^{-09}	*CFH*	intronic
93	rs402991	194942726	1.60x 10^{-09}	*CFH*	intronic
94	rs6664705	194944465	1.60x 10^{-09}	*CFH*	intronic
95	rs7522681	194945228	1.60x 10^{-09}	*CFH*	intronic
96	rs9970075	194940053	1.60x 10^{-09}	*CFH*	intronic
97	rs9970784	194940425	1.60x 10^{-09}	*CFH*	intronic
98	rs10754199	194937462	1.60x 10^{-09}	*CFH*	intronic
99	rs1061147	194920947	1.70x 10^{-09}	*CFH*	coding
100	rs10801555	194926884	1.70x 10^{-09}	*CFH*	intronic