University of Dundee

DOCTOR OF PHILOSOPHY

Clinical and genetic determinants of diabetic retinopathy

Liu, Yiyuan

Award date:
2014

Awarding institution:
University of Dundee

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 17. Feb. 2017

DOCTOR OF PHILOSOPHY

Clinical and genetic determinants of diabetic retinopathy

Yiyuan Liu

2014

University of Dundee

Conditions for Use and Duplication
Copyright of this work belongs to the author unless otherwise identified in the body of the thesis. It is permitted to use and duplicate this work only for personal and non-commercial research, study or criticism/review. You must obtain prior written consent from the author for any other use. Any quotation from this thesis must be acknowledged using the normal academic conventions. It is not permitted to supply the whole or part of this thesis to any other person or to post the same on any website or other online location without the prior written consent of the author. Contact the Discovery team (discovery@dundee.ac.uk) with any queries about the use or acknowledgement of this work.

DOCTOR OF PHILOSOPHY

Clinical and genetic determinants of diabetic retinopathy

Yiyuan Liu

2014

University of Dundee

Declaration of the Candidate

I declare that this thesis is my own work and include nothing which is the outcome of work done in collaboration except where specifically indicated in the text. No part of this thesis has previously been submitted for an academic degree or any other qualification.

Yiyuan Liu

April 2014

Declaration of the Supervisor

I certify that Yiyuan Liu has completed the equivalent of twelve terms of experimental research and that she has fulfilled the conditions of the University of Dundee so that she is qualified to submit this thesis in application for the degree of Doctor of Philosophy.

Professor Colin Palmer

Professor Ewan Pearson

Acknowledgements

I am immensely grateful to my supervisor Prof. Colin Palmer for his mentorship generously offered to me to grow scientifically and personally. His invaluable advice and guidance have been pivotal in my Ph.D. progression, and indispensable to the birth of this thesis. I am also profoundly thankful to my co-supervisor Prof. Ewan Pearson whose thought-provoking suggestions and advice propelled me to develop with an inquisitive scientific nature, and have cultivated my interest in the pursuit of biomedical research. Additionally, I would like to thank Dr. Alex Doney whose warm advice and encouragements have directly accelerated my Ph.D. progress. Finally, every step of my Ph.D. study has benefited from the unreserved support of Prof. Andrew Morris, Prof. Graham Leese, Prof. Helen Colhoun, Dr. Kaixin Zhou, Dr. Louise Donnelly, Dr. Natalie Van Zuydam, Dr. Harshal Deshmukh, Dr. Roger Tavendale and Ms. Fiona Carr, and I wish to thank you all.

This thesis could not been possible without the wonderful support from families and friends. I can hardly convey in full my thankfulness my mum Haiyan Wei, dad Zhi Liu, my life-long companion and husband Dr. Minghui Wang who all for years have made my life full of growth and joy. Moreover, I would like to thank my close friend Mr David Martin, who

added paramount cheerfulness to my Ph.D. study in Dundee, and kindly offered to proof read this thesis. Finally, I am extremely lucky to have my personal growth companied by my friends in Dundee: Mr. Alan Palmer, Mr. Colin Anderson, Mr. Eugene Clark, Dr. Robert Close, Prof. Sam Eljamel, Dr. Shane Sellarajah; I would like to thank you all from the bottom of my heart.

I have deep gratitude for the College of Medicine Dentistry and Nursing, the University of Dundee, owing to the first-class Ph.D. programme and the generous scholarship that has made this thesis and my scientific development possible. Additionally, I am thankful to the Health Informatics Centre, University of Dundee and NHS Tayside for collecting, managing and supplying the anonymised clinical data. Lastly, I am grateful to the GoDARTS study patients and the staff members involved in the GoDARTS project.

Abstract

Diabetic retinopathy is a microvascular complication of type 1 and type 2 diabetes, affecting the retinal vasculature of the eye. In the Scottish diabetic retinopathy grading scheme (version 1.1), the severity of diabetic retinopathy is classified as no retinopathy, mild background, observable background, severe non-proliferative and proliferative retinopathy. In the GoDARTS (Genetics of Diabetes Audit and Research Tayside) cohort, we have longitudinal data of retinopathy in diabetic patients since 1990. 3,734 and 3,673 GoDARTS patients were genotyped in the Affymetrix Genome-wide Human SNP Array 6.0 and Illumina HumanOmniExpress BeadChip, respectively. As the pathophysiology of diabetic retinopathy remains elusive, the aim of this thesis is to use the GoDARTS phenotype and genotype data to study clinical and genetic determinants for diabetic retinopathy.

The first part detailed the longitudinal analysis of diabetic retinopathy data, using the multi-state Markov model. This methodology permitted us to infer the rates of transitions across retinopathy states under the influence of common population risk factors. We observed a consistent risk effect of HbA1c on the progression (no-retinopathy to mild background diabetic retinopathy, BDR, hazard ratio per standard deviation of HbA1c (HR): 1.42, 95%

confidence interval (CI): 1.32, 1.52; mild BDR to observable BDR HR: 1.32, 95% CI: 1.08, 1.60; observable BDR to severe non-proliferative/proliferative DR (non-PDR/PDR) HR: 2.23, 95% CI: 1.16, 4.29). Similarly, systolic blood pressure (SBP) and diastolic blood pressure (DBP) increased the risk for the transition from the asymptomatic phase to mild BDR (HR: 1.20; 95% CI: 1.11, 1.30) and the mild BDR to observable BDR (HR: 1.87; 95% CI: 1.46, 2.40), respectively. Regression from mild BDR to no DR was associated with lower SBP (HR: 0.79; 95% CI: 0.64, 0.97) and lower HbA1c (HR: 0.76; 95% CI: 0.64, 0.89). Our results provided the evidence that glycaemic exposure and blood pressure are strongly associated with progression and remission of diabetic retinopathy.

The second part studied narrow-sense heritability and genetic correlations of retinopathy and related risk factors in type 1 and type 2 diabetic patients explained by common genome-wide SNPs using a novel variance components methodology. We found that up to 34% of phenotypic variation of diabetic retinopathy was attributable to total additive genetic effects in the GoDARTS 1000 Genomes imputed genetic data. The narrow-sense heritability explained by the study data was 49% for BMI (body mass index), 20% for cholesterol, 23% for serum creatinine, 21% for HbA1c (glycaemic exposure), 40% for HDL (high-density liproprotien), 18% for SBP (systolic blood pressure), 31% for triglycerides. Due to the wide credible intervals, the inference of genetic and residual correlations between retinopathy and the clinical risk factors was limited.

The last part of the thesis described genome-wide meta-analysis of retinopathy in type 1 and type 2 diabetic patients in the SUMMIT collaborative cohorts. This study analysed 9,508,089 genetic variants in 5,422 diabetic retinopathy cases and 4,302 controls. We

identified the most significant signal SNP rs10746970 (*P* value of 2.22×10^{-7}). In this study, we found diabetic retinopathy risk loci 1q43, 2q32.1, 6q16.1 13q32 identified in this study are in close proximity to regions of 1q13-42, 2q31-47, 6q22-27 and 13q14-32 previously reported in smaller scaled genome-wide association studies. We discovered previously unreported retinopathy risk loci of 1p34-p32, 2p12-p11.1, 2q11.2, 3p24.3, 3q24, 4q28-q32, 5p14, 5q31.3, 6p21, q13-q21.2, 11q21, 12p11.22, 17q25.1 and 19q13.11. The identification of extensive susceptibility loci is suggestive of polygeneic effects contributing to the development of diabetic retinopathy. The closest genes to these loci have been implicated in multiple physiological processes including carbohydrate/lipid metabolism, functional activity in the neuronal and glial cells, and transcription regulation.

This series of studies reported novel insights into the clinical and genetic susceptibility of diabetic retinopathy, and provided the scientific basis of informed clinical decisions on diabetic retinopathy prognosis and treatment. The findings are valuable to further studies of diabetic retinopathy.

CONTENTS

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Diabetic retinopathy

1.1.1 Clinical features and classification

The development of type 1 and 2 diabetes mellitus increase patients' risk for developing a wide range of chronic complications. Macrovascular complications are identified by severe atherosclerosis in large blood vessels, and diabetic macroangiopathy affects the heart (coronary artery disease), the brain (cerebrovascular disease) and the lower extremities (peripheral vascular disease). In contrast, microvascular complications are characterised by weakening of capillary walls, leading to bleeding and a reduced blood flow. Diabetic microangiopathy affects the kidney (nephropathy), peripheral nerves (neuropathy) and the retina (retinopathy).

Diabetic retinopathy (DR) is a range of microangiopathies that can affect the retina of both eyes (Figure 1-1 and Figure 1-2), influenced by prolonged hyperglycaemia. Progression of diabetic retinopathy is sight threatening. Diabetic retinopathy is diagnosed by the methods of digital imaging photography or ophthalmoscopy (Figure 1-3), and in Scotland, clinical classification is based on the latest Scottish grading schemes for diabetic retinopathy and maculopathy (Table 1-2 and Table 1-3). The initial phase of diabetic retinopathy can be asymptomatic. The development of diabetic retinopathy is broadly classified into non-proliferative and proliferative stages [1]. The non-proliferative or background stage is characterised by damages to retinal endothelium and the resultant capillary occlusion leading to retinal ischaemia, and related clinical features include microaneurysms, haemorrhages and cotton wool spots in the retinal periphery (Table 1-1, Table 1-2 and Figure 1-4) [2]. The advanced proliferative stage develops with the proliferation of new blood vessels (neovascularisation) on the interface between the retina and the vitreous cavity [1,2]. Rupture of fragile blood vessels causes a large retinal haemorrahage [2]. Patients with a vitreous haemorrhage may experience obscured vision, and patients suffering from a preretinal haemorrhage may notice a moving visual blockage influenced by gravity [3]. When retinopathy affects the macula at the posterior pole of the retina, diabetic maculopathy develops (Table 1-3), and the loss of central vision becomes imminent [4]. Diabetic maculopathy is prevalent in type 2 diabetic patients [5]. Macular oedema and macular ischaemia are two sub-classes of diabetic maculopathy, and these two conditions can be concurrent [4]. Macular oedema is characterised by intraretinal fluid accumulation and swelling at the macular area [4], and the hallmark of macular ischaemia is macular vascular occlusion leading to cell death [4].

Figure 1-1 Human eye anatomy adapted from [6]. The wall of eyeball consists of three layers: the outermost connective tissue layer formed by cornea and sclera, the vascular tissue layer, the uvea composed of the iris, ciliary body and choroid; and the innermost neuronal tissue, the retina. The retina receives 65% to 85% of the blood supply from the choroid for nourishing photoreceptor cells [7].

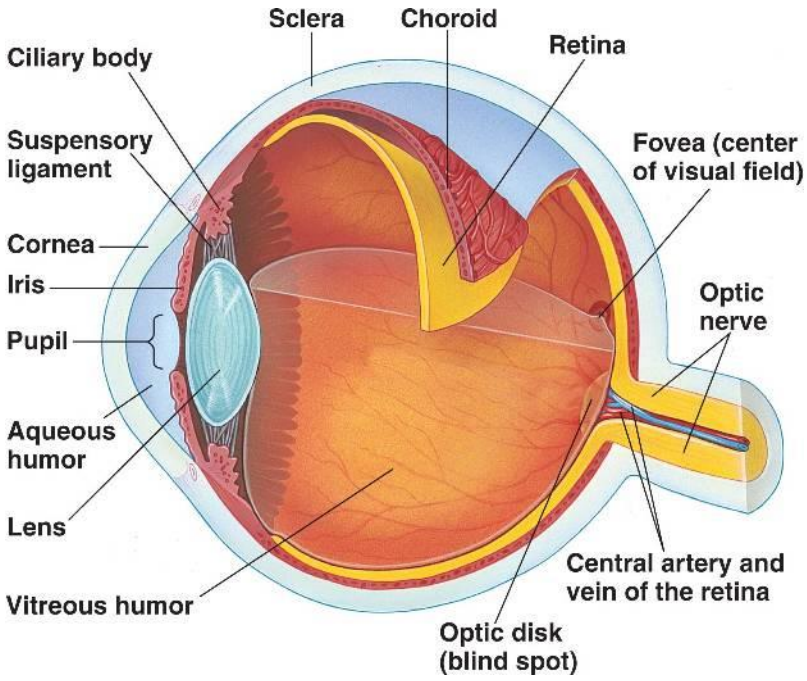


Figure 1-2 The microstructure of human retina adapted from [8]. Light is transmitted through several transparent cellular layers before reaching photoreceptor cells (rods and cones) at the posterior of the retina.

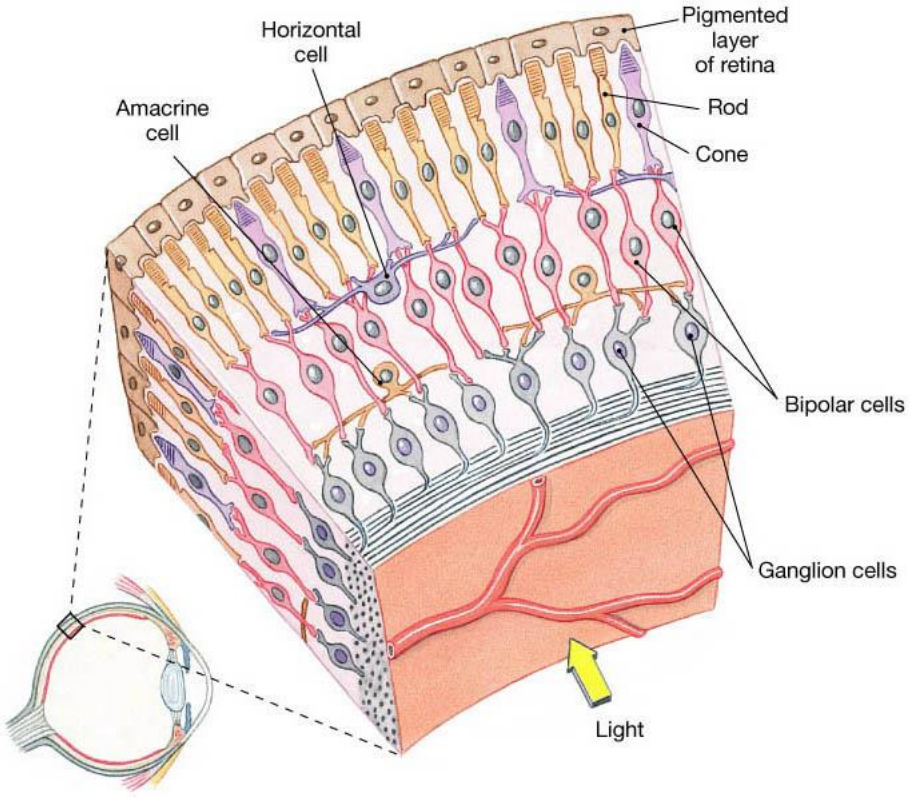


Figure 1-3 A fundus photograph showing central retinal blood vessels radiating from the optical nerve in a healthy eye adapted from [4]. The fovea is seen as a red spot at the centre of the macular area free of blood vessels.

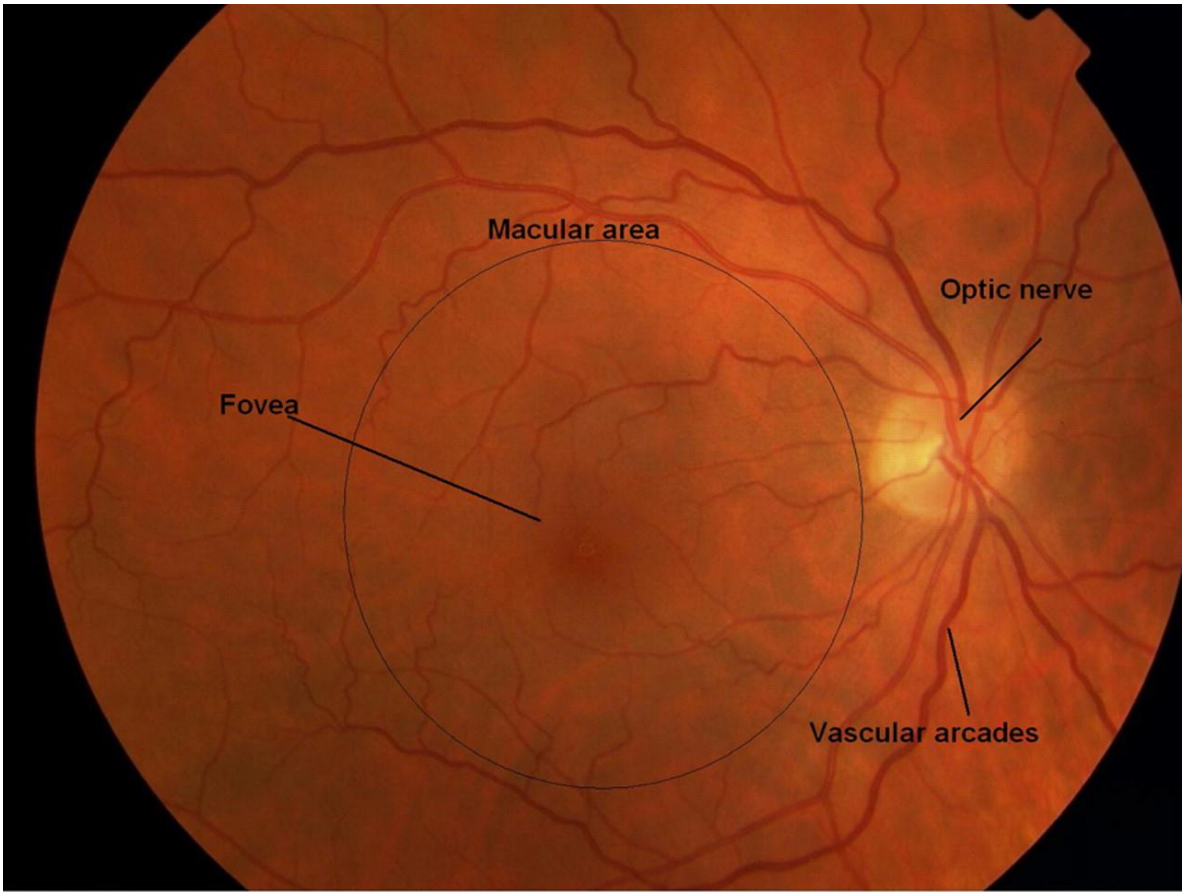


Table 1-1 Ophthalmic features associated with diabetic retinopathy summarised from [3].

Microaneurysms: • Red dots, caused by protrusion of retinal capillaries.
Haemorrhages: • Extravasation of retinal blood vessels. • Flame-shaped haemorrhages: haemorrhages in the nerve fiber layer of the retina. • Dot or Blot-shaped haemorrhages: haemorrhages deep in the connecting neuron layers of the retina.
Exudates: • Yellow-white lesions, caused by deposits of plasma leaked from retinal capillaries.
Cotton-wool spots: • Pale spots, caused by swellings of ischaemic nerve fibers.
Intraretinal microvascular abnormalities (IRMA): • Burgundy-colored area, caused by dilatation of retinal capillaries.

Figure 1-4 Ophthalmic features of background diabetic retinopathy adapted from [9].

FH: flame-shaped haemorrhages; CWS: cotton wool spots; HE: hard exudates.

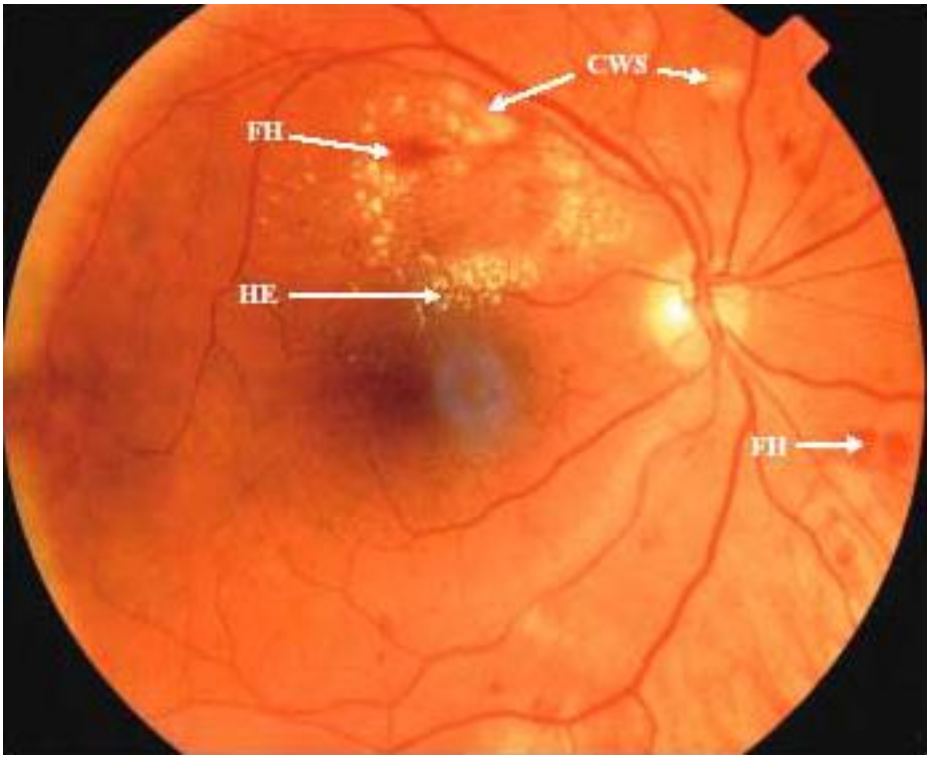


Table 1-2 Clinical diagnosis of diabetic retinopathy according to the Scottish diabetic retinopathy grading scheme 2007 v1.1. IRMA: intraretinal microvascular abnormalities.

Retinopathy	Description
No visible retinopathy	No diabetic retinopathy anywhere
Mild background	The presence of at least one of any of the following features anywhere • dot haemorrages • microaneurysms • hard exudates • cotton wool spots • blot haemorrages • superficial/flame shaped haemorrhages
Observable background	Four or more blot haemorrhages in one hemi-field only (Inferior and superior hemi-fields delineated by a line passing through the centre of the fovea and optic disc)
Severe non-proliferative retinopathy	Any of the following features: • Four or more blot hameorrages in both inferior and superior hemi-fields • Venous beading • IRMA
Proliferative retinopathy	Any of the following features: • Active new vessels • Vitreous haemorrhage

Table 1-3 Clinical diagnosis of diabetic maculopathy based on the centre of the fovea in macula according to the Scottish diabetic maculopathy grading scheme 2007 v1.1.

Maculopathy	Description
No maculopathy	No features ≤disc diameters from the centre of the fovea sufficient to qualify for observable or referable maculopathy as defined below.
Observable maculopathy	Lesions as specified below within a radius of > 1 but ≤ 2 disc diameters the centre of the fovea. • Any hard exudates
Referable maculopathy	Lesions as specified below within a radius of ≤ 1 disc diameter of the centre of the fovea. • Any blot haemorrhages • Any hard exudates

1.1.2 Treatment

Patients with early-stage diabetic retinopathy are regularly monitored for signs of severe non-proliferative/proliferative retinopathy or maculopathy, but effective therapeutic interventions for early-stage diabetic retinopathy remain limited [4]. Patients with proliferative retinopathy and diabetic macular oedema are referred for laser photocoagulation [4]. In the course of the treatment, an ophthalmologist administers flashes of laser beams to peripheral retina or fovea, aiming to reduce the size of ischaemic retina or to impair macular thickening respectively for proliferative retinopathy or macular oedema [3]. Multiple sessions of laser treatment are required for patients with proliferative retinopathy [3]. A study of laser photocoagulation in patients with proliferative retinopathy showed a 95% success rate in preventing further neovascularisation and improving visual acuity [3]. In contrary, a 60% visual improvement was reported for laser treatment in patients with diabetic macular oedema [3].

If an extensive haemorrhage occurs in or behind the vitreous cavity or a fibrous mass threatens to detach from the retina, patients with proliferative retinopathy may undergo vitrectomy surgery [4]. This technique removes abnormal tissues and replaces the vitreous by clear fluid [4]. Despite of this surgical intervention, vitreous haemorrhages are likely to recur [4].

New treatment including intravitreal administration of anti-VEGF (vascular endothelial growth factor) is recommended for patients with proliferative retinopathy [3]. However, effectiveness of this intervention remains unclear.

1.1.3 Diabetic retinopathy screening

Long-term follow-up studies showed that retinal screening is crucial for detecting onset or progression of diabetic retinopathy [4]. In Tayside, patients with diabetes were screened in a retinal exam using ophthalmoscopy or slit lamps, in a hospital by ophthalmologists and diabetologists or by hospital and community based optometrists [10,11]. Since 1986, non-mydriatic (without dilation of the pupil) Polaroid photography has been shown as an effective method for identifying retinopathy in the diabetic population [12–14]. In 1992, a mobile unit mounted with a retinal camera was set up for systematic diabetic retinopathy screening in the Tayside diabetic population [15]. From 2003, digital retinal photography has been integrated into diabetic retinopathy screening, replacing Polaroid photography for a higher image quality and the ease of data storage [16]. In the same year, a Scotland-wide diabetic retinopathy screening program was established with an independent retinal grading scheme recommended by the Health Technology Board Scotland (HTBS) [17]. Currently, all type 1 and type 2 diabetic patients aged 12 years or older in Scotland are offered an annual retinal examination through the National Health Service (NHS) diabetic retinopathy screening program, or more frequently if required [4].

1.1.4 Epidemiology

Diabetic retinopathy is a prevalent microvascular complication in patients with diabetes. Nearly all type 1 diabetic patients develop diabetic retinopathy within 20 years of diagnosis [1]. Other previous studies have reported the prevalence of retinopathy in type 1 diabetic patients varies from 10% to 50%, stratified by the diagnostic procedure, ethnicity, population, the patient age, and the duration of diabetes [18].

At the diagnosis of diabetes, up to 21% of type 2 diabetes sufferers have clinical features for retinopathy [19], and exceeding 60% of type 2 diabetic patients develop retinopathy within 15-20 years [1]. The prevalence of diabetic retinopathy in European ancestry patients with type 2 diabetes ranges from 5% to 52% [18]. Leese *et al* [20] found in Tayside, Scotland, the prevalence of diabetic retinopathy in type 2 diabetic patients of an urban area is 7%, whereas in a rural area, the prevalence rises to 13%.

Population risk factors for diabetic retinopathy includes diabetes duration [19], glycaemic exposure [21], blood pressure [22], serum total cholesterol [23], triglyceride [23], body mass index (BMI) [23], pregnancy [24] and cigarette smoking [25]. Management of diabetes by achieving glycaemic, blood pressure and lipid targets (glycated haemoglobin, HbA1c, 6.5% or 48.0 mmol/mol [26]; diastolic/systolic blood pressure, DBP/SBP, 140/80 mm Hg; total cholesterol, 4.0 mmol/l, or low density lipoprotein cholesterol, LDL, 2.0 mmol/l [27]) is the recommended primary prevention for retinopathy.

Despite the optimal control of diabetic risk factors, patients with over 20 years' diabetes are at risk for developing retinopathy. These population risk factors explain limited variations in the risk of diabetic retinopathy. The Diabetes Control and Complications Trial (DCCT) discovered 11% of variation in diabetic retinopathy susceptibility was accounted for by glycaemic exposure and diabetes duration [28–30]. Similarly, in the Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR) study, the effect of glycaemic exposure, blood pressure and serum total cholesterol contributed 9-11% of variance

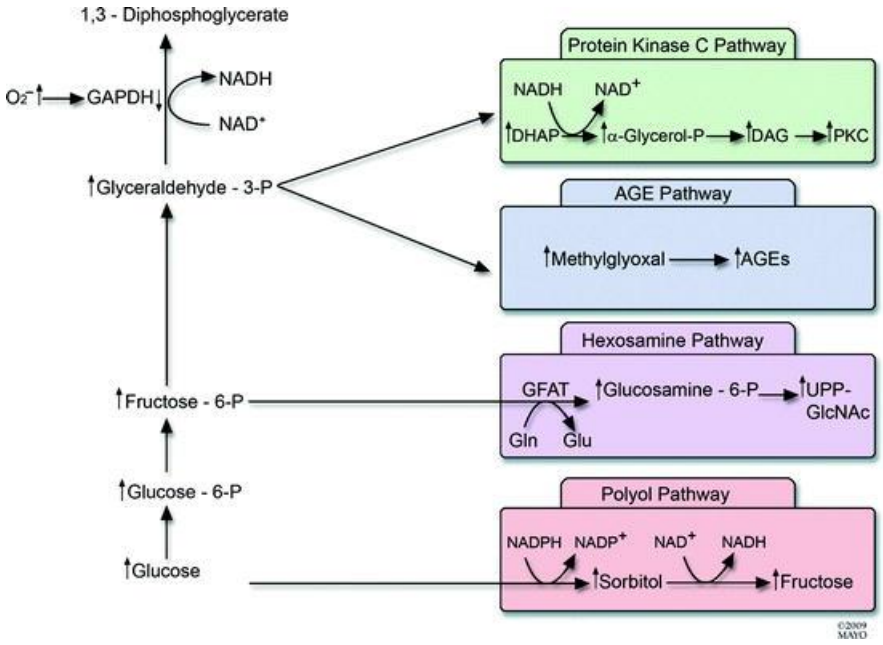
towards diabetic retinopathy risk [31]. These findings suggest a role for non-environmental risk factors for diabetic retinopathy.

Increasing evidence suggests genetic variation can influence the susceptibility to diabetic retinopathy. For example, diabetic retinopathy shows patterns of familial aggregation, and higher prevalence of diabetic retinopathy was observed in siblings than in genetically unrelated individuals [32].

1.1.5 Pathogenesis

Although the pathogenic mechanism of diabetic retinopathy has not been fully elucidated, diabetic retinopathy studies have identified retinal microvascular abnormalities [33]. Sustained hyperglycaemia increases vascular permeability, accompanied by decreased synthesis of vasodilators (nitric oxide), elevated production of vasoconstrictors (angiotensin II, endothelin-1), with the release of vasopermeability-inducing cytokines (vascular endothelial growth factor, VEGF), resulting in programmed cell death (apoptosis) of retinal endothelial cells [33]. These retinal vascular abnormalities are linked to biochemical changes underlying multiple metabolic pathways (Figure 1-5 and Figure 1-6), and hormonal alterations in the endocrine system (Figure 1-7).

Figure 1-5 Multiple interconnecting pathways in glucose metabolism, leading to vascular tissue damage adapted from [33]**. P: phosphate; NAD+/NADH: nicotinamide adenine dinucleotide; NADP+/NADPH: nicotinamide adenine dinucleotide phosphate; GAPDH: glyceraldehyde 3-phosphate dehydrogenase; DHAP: dihydroxyacetone phosphate; DAG: diacylglycerol; PKC: protein kinase C; AGE: advanced glycation end product; GFAT: glutamine fructose-6-phosphate amidotransferase; Gln: glutamine; Glu: glucose; UPP-GlcNAc: uridine diphosphate N-acetylglucosamine.**



1.1.5.1 Polyol pathway

The sorbitol-aldose reductase or the polyol pathway has been implicated in several diabetic complications including retinopathy, nephropathy (microangiopathy in kidney) and neuropathy (microangiopathy in neuronal tissues). The energy source molecule glucose enters cells through glucose transporter proteins (GLUT) by diffusion [34], and five subtypes differ in enzyme kinetics and insulin dependency, enabling a differential use of glucose in a diverse range of tissues [34]. GLUT1 primarily expressed in retinal, neuronal and renal tissues is insulin independent [34]. In diabetes, excessive intracellular glucose saturates glycolysis, the ubiquitous glucose metabolic pathway, and the build-up of glucose is metabolised in the polyol pathway (Figure 1-5) by the enzyme aldose reductase reducing glucose into sorbitol and oxidising the co-factor nicotinamide adenine dinucleotide phosphate (NADPH) [35]. In hyperglycaemia, the burden for aldose reductase for glucose is increased, leading to the accumulation of sorbitol and deprivation of NADPH for other cellular processes including nitric oxide and glutathione synthesis [33]. Acquired glutathione deficiency promotes oxidative stress. Imbalance of these metabolites causes cellular damages [33].

1.1.5.2 Advanced glycation end product (AGE) pathway

In the normal metabolic state, nonenzymatic protein glycation occurs at a low and constant rate [35], and in the diabetic state, high availability of intracellular glucose markedly accelerates the formation of AGEs from sugars reacting with free amino groups of proteins, lipids, and nucleic acids (Figure 1-5) [35]. A feature of AGEs is the formation of inappropriate protein cross links, leading to impaired protein structure and functions [35].

The main influence of AGEs on diabetic retinopathy susceptibility is mediated by AGE-binding receptors including the receptor for advanced glycation endproducts (RAGE) [36], activating proinflammatory signals [35]. AGEs were detected in retinal tissues, and blockade of AGE pathway retarded the development of retinopathy in diabetic patients [33].

1.1.5.3 Protein kinase C (PKC) pathway

Hyperglycaemia triggers a profusion of glucose reflux through glycolysis, and in turn an increased synthesis of an intermediate metabolite diacylglycerol (DAG), a second messenger for activating nine of eleven identified isoforms of PKC (Figure 1-5) [35]. Hyperglycaemia also stimulates the PKC pathway via AGE [37] and polyol [38] mediated pathways. PKC-β isoforms are expressed in retinal and renal vasculature, and the PKC-β activity inducesactivation of endothelin-1 and inhibition of nitric oxide synthesis, causing blood circulation aberration [39]. PKC also activates expression of VEGF in vascular smooth muscle cells [40], inciting angiogenesis and endothelial permeability.

PKC-β inhibitor (ruboxistaurin mesylate) was reported as an effective therapeutic drug in prevention of progressive vision loss amongst diabetic retinopathy patients [35]. Nevertheless, owing to versatile regulatory roles of PKC-β in cell signalling, inhibition of PKC-β has adverse effects [35], and hence, PKC-β has not been recommended as a therapeutic target for diabetic retinopathy.

1.1.5.4 Hexosamine pathway

In diabetes, the build-up of intracellular glucose causes overproduction of fructose-6-phosphate, a glycolytic metabolite, which is also a substrate for synthesis of N-acetylglucosamine (GlcNAc) in the hexosamine pathway (Figure 1-5). GlcNAc glycosylation of the intracellular transcription factor Specificity Protein 1 (Sp1) reduces competitive phosphorylation at the same regulatory domain [33]. Dephosphorylation of Sp1 is associated with increased transcription of the plasminogen activator inhibitor-1 gene (*PAI-1*), leading to vascular endothelial dysfunction and related alterations, consistent with retinal vascular damage observed in patients with diabetic retinopathy[33].

1.1.5.5 Oxidative stress

In hyperglycaemia, when glucose is oxidised, superfluous electron donors (NADH and $FADH_2$) generated in the tricarboxylic acid or citric acid (TCA) cycle oversupply the downstream electron transport chain (Figure 1-6). Consequently, an overactive electron transport chain generates excess free radical by-products, exhausting cellular antioxidants and subsequently, causing oxidative stress [41]. Reactive oxygen radicals are known to destroy cellular functions, and have been implicated in diabetes [42]. Studies demonstrated the link between oxidative stress and diabetic retinopathy severity [43]. Attenuation of reactive oxygen species production suppresses polyol, AGE, PKC and hexosamine pathways, providing strong evidence for a unifying theory of oxidative stress-mediated pathological pathways in hyperglycaemia [44].

Figure 1-6 Biochemical pathways underlying glucose metabolism in humans adapted from [45]**, highlighting intra-mitochondrial reactions in the citric acid cycle and electron transport chain. NAD+/NADH: nicotinamide adenine dinucleotide; ATP: adenosine-5'-triphosphate; ADP: adenosine diphosphate; AMP: adenosine monophosphate ; HSCoA/CoA: coenzyme A; PPi: pyrophosphate; Pi: orthophosphate; CO2: carbon dioxide; FAD+/FADH2: flavin adenine dinucleotide; GTP: guanosine-5'-triphosphate; e-: electron; H+: proton; O2: oxygen; H2O: water, OH-: hydroxide.**

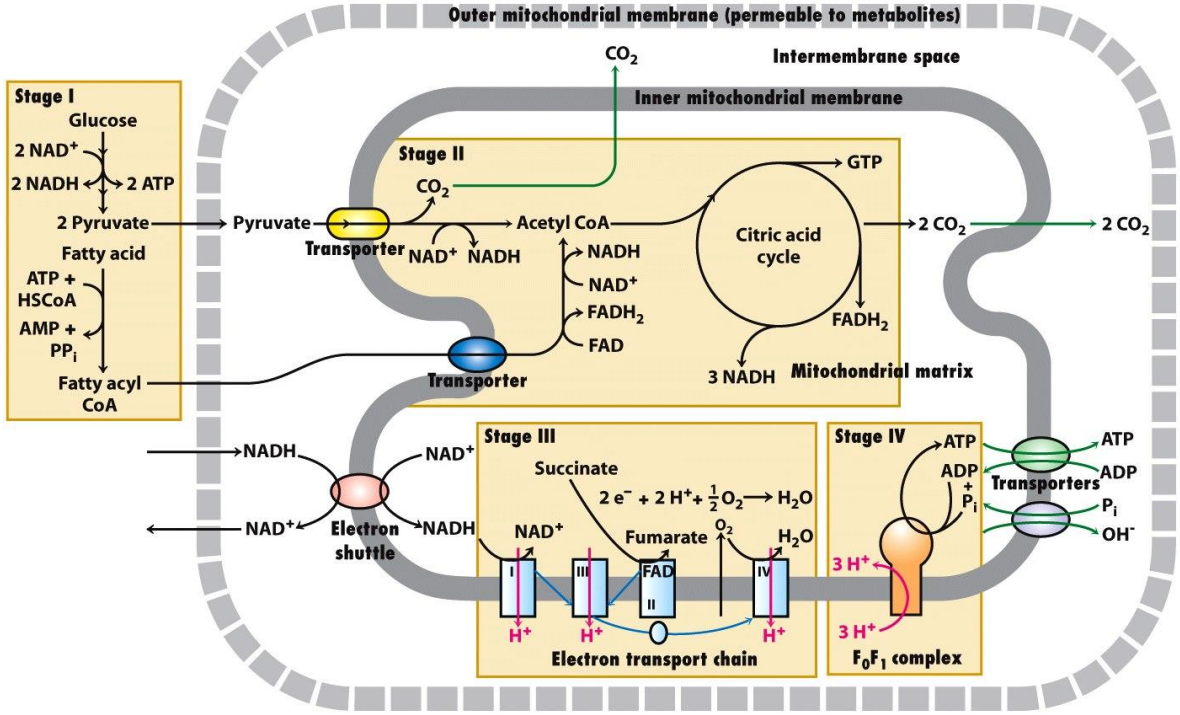
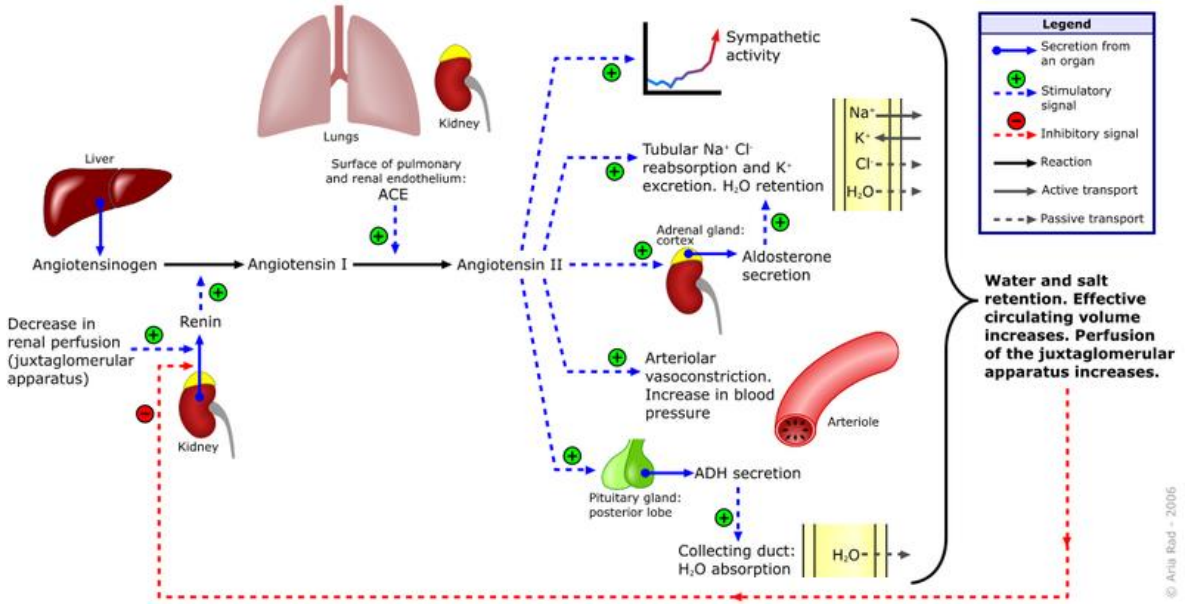


Figure 12-8
Molecular Cell Biology, Sixth Edition
© 2008 W. H. Freeman and Company

1.1.5.6 Renin-angiotensin system

The renin-angiotensin system (RAS, Figure 1-7) is an endocrine system regulating blood pressure and volume of body fluid (blood plasma, lymph and tissue fluid), and has an implication for proliferative diabetic retinopathy. In the classical pathway, granular cells in the kidney, a group of specialised smooth muscle cells located in glomerular arterioles, synthesize and release the enzyme renin into blood circulation, in response to a drop in blood pressure. Plasma renin catalyses the cleavage of the oligopeptide angiotensinogen synthesized in liver, forming angiotensin I. Angiotensin-converting enzyme (ACE) are glycoproteins present in the lung, endothelial and renal cells, where both the membrane bound and secretory forms of ACE were found [46]. ACE mediates cleavage of angiotensin I. The resulting active hormone angiotensin II influences vasoconstriction and electrolyte homeostasis, restoring blood pressure balance.

Figure 1-7 The renin-angiotensin system influences an array of physiological functions in the human body adapted from [47]**.**



However, further studies extended RAS to include additional focal angiotensin II-releasing circuits in neuronal [48], cardiovascular [49], renal [50,51] and adrenal [52] tissues. The precise mechanism for exacerbation of diabetic retinopathy by angiotensin II under a hyperglycaemic state is still enigmatic. Compelling evidence suggests that angiotensin II receptors in retina mediate the release of VEGF, inciting angiogenesis [53], and the inhibition of ACE suppresses neovascularisation [54].

1.1.5.7 Hypertension-induced mechanical stress

Hypertension is a known modifiable risk factor for diabetic retinopathy, and frequently concomitant with hyperglycaemia [55]. A series of studies applied mechanical stress to an *in vitro* cellular environment to investigate retinal responses to hypertension *in vivo*. The

results indicated that mechanic stretch up-regulated the expression of VEGF, a biomarker for proliferative retinopathy, in retinal pigment epithelial [56], endothelial and pericyte cells [57] through singling pathways involving phosphatidylinositol 3-kinase and protein kinase C [58]. Additionally, there was proof that mechanical stretch increased oxidative stress and apoptosis in porcine [59] and bovine [60] retinal pericytes, which was further exacerbated by hyperglycaemia [60]. The loss of retinal pericytes undermines the blood–brain barrier in the retina, characteristic of diabetic retinopathy [55].

1.1.6 Heritability

1.1.6.1 Estimation methods

The premise that many complex traits are heritable is crucial to many genetic epidemiology studies. Determining the degree to which a trait is heritable is a critical first stride towards deciphering the underlying genetic architecture and pertinent to designing appropriate gene-mapping strategies. Heritability is a measure that quantifies the fraction of phenotypic variations due to genetic effects. Phenotypic variations (V_p) in a population are attributable to the variability in environment effects (V_e) and genetic components (V_g) including allelic interactions within a locus (dominance, V_d), across loci (epistasis, V_{ep}) and additive allelic effects within and/or across loci (V_a). Heritability in the broad sense (H^2) is the ratio of the total genetic variance to the phenotypic variance, i.e. $H^2 = V_g / V_p$. Although phenotypic variations are directly measureable from a population sample, the total genetic variations are frequently unobserved and challenging to gauge. In the past, animal and plant breeders discovered that close resemblance between relatives was due to the additive

genetic effects [61], and by considering phenotypic values within genetic related samples, it became feasible to attain approximations for the variance of the total additive genetic effects and the narrow-sense heritability (h^2), which is the ratio of the total additive genetic variance to the phenotypic variance, i.e. $h^2 = V_a / V_p$. Owing to its vital role in breeding programs and quantitative genetics, hereinafter, the discussion is dedicated to narrow-sense heritability, unless otherwise stated.

A number of heritability estimation techniques are commonly utilised, and the adoption of such an approach is on the basis of practical limitations such as the genetic relationship of samples, the sample size and/or the nature of genetic data. Historically, parent-offspring regression outcompeted the rivals by multiple counts. First, the parent-offspring relationship can be identified with ease. For example, in farm fields, the origination of seeds from the corresponding maternal crops can be established with certainty, whereas other types of relations such as full-sibship, half-sibship or paternity may not be accurately identified in a natural population. Second, the regression was statistically simple to perform. Third, the covariance of phenotypic values across a familial generation is considered as influenced by additive genetic effects at the absence of dominance or epistasis effects [62], and thus, the narrow-sense heritability can be estimated from the coefficient of the regression [62]. However, the parent-offspring information may not always be available. For instance, generations may not overlap for some species, and therefore the lineage becomes impossible to establish. Sibship-based estimation methods were considered as a compelling alternative. In this analysis, analysis of variance (ANOVA) is applied to attain phenotypic covariations within- and between-family, which heritability can be estimated from the intraclass correlation [62]. In addition, the familial relationship of twins is

common in human populations [62], and as in Falconer's formula [63], the broad-sense heritability is estimated by $H^2 = 2\left(r_{mz} - r_{dz}\right)$, where r_{mz} is the concordance of monozygotic twins (derived from the same embryo); r_{dz} is the concordance of dizygotic twins (derived from two embryos). Finally, large-scaled human population cohorts frequently consist of largely genetically unrelated individuals and a small number of relatives that may be within a generation or across-generations. With the availability of genome-wide variant data, genetic similarities between samples can be estimated to facilitate the estimation of narrow-sense heritability using a linear mixed model [64].

Owing to the paucity of densely genotyped data for quantitative genetics studies, conventional methods for estimating heritability in relatives were based on inherent assumptions in respect to the expected genetic and environmental resemblance of relatives. For example, the phenotypic resemblance across monozygotic and dizygotic twins was assumed to be attributable to common environmental effects [64], which from the current perspective is questionable. Furthermore, heritability estimates for a trait are likely to be variable, reflective of temporal and spatial fluctuations in the underlying genetic and environmental effects. For instance, heritability for first lactation milk yield in dairy cattle was 25% in the 1970s, and 40% at the present time [64]. Environmental variance may differ between non-identical study populations or in an identical study population but sampled at different times [64]. Genetic variance can be altered by allele frequencies, the emergence or the extinction of variants [64]. Thus, heritability of a trait is specific to a population at a study time.

1.1.6.2 Heritability of diabetes and diabetic retinopathy

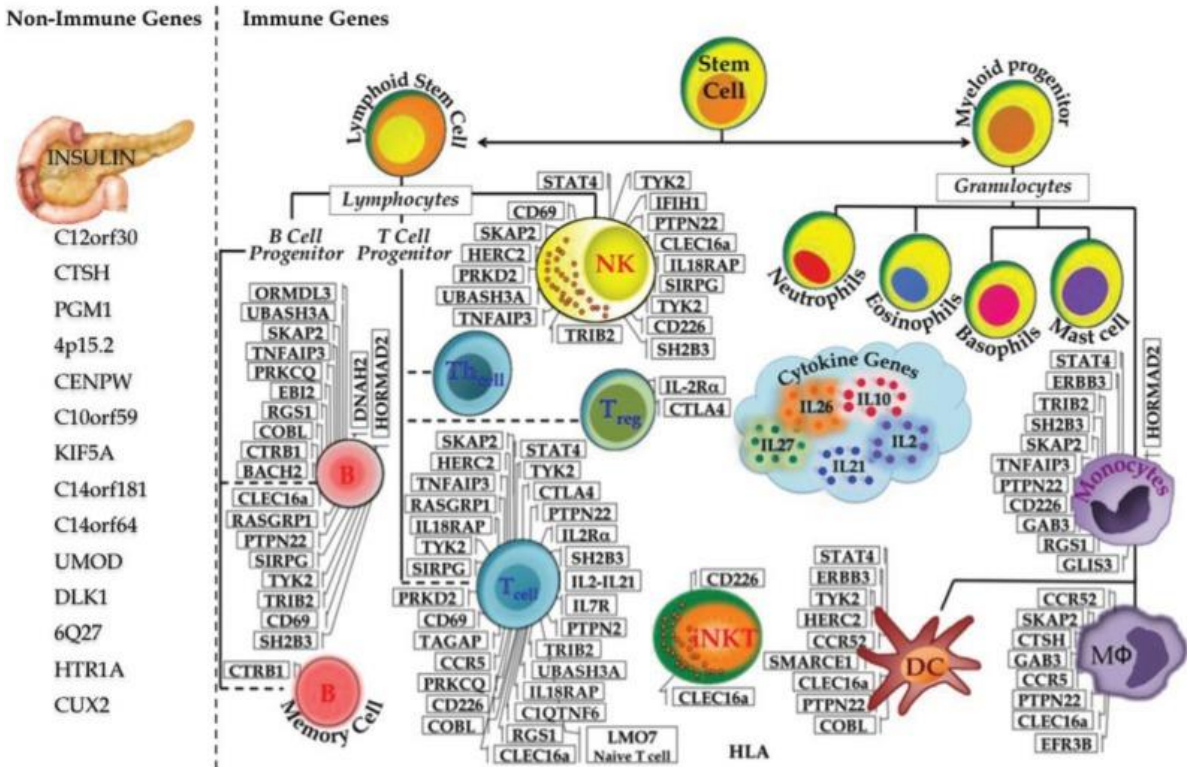
The clinical phenotypes of diabetes and diabetic retinopathy are both categorical in nature, and to estimate heritability, these categorical phenotypes are transformed to underlying unobserved liability on a continuous scale [65]. The heritability estimate was 88% for type 1 diabetes, in the Finnish Twin Cohort, which included 44 monozygotic and 183 dizygotic twin pairs with type 1 diabetes in a 22,650 population-based twin pairs [66]. In the Danish Twin Register where 62 twin pairs out of 606 were type 2 diabetic, the heritability estimate was 26% for type 2 diabetes [67]. It was suggested that the heritability estimate of type 2 diabetes was reflective of heritable components of obesity, a major risk factor for type 2 diabetes [68]. The broad-sense heritability was 27% for background and proliferative diabetic retinopathy in the FIND-Eye study, where most patients were diagnosed with type 2 diabetes [69]. Heritability estimates were 25% and 52% for proliferative retinopathy respectively in the FIND-Eye cohort [69] and the FinnDiane study (a type 1 diabetes cohort) [70].

1.1.7 Genetics of diabetes

Type 1 diabetes is characterised by autoimmune-mediated chronic destruction of insulin-secreting beta cells in the pancreas. Through studies of candidate genes and a limited set of genome-wide nonsynonymous variants, 6 susceptibility genes for type 1 diabetes were successfully identified between 1974 and 2006, including *HLA* (major histocompatibility complex, class I), *INS* (insulin), *CTLA4* (cytotoxic T-lymphocyte-associated protein 4), *PTPN22* (protein tyrosine phosphatase, non-receptor type 22), *IL2RA* (interleukin 2 receptor, alpha), and *IFIH1* (interferon induced with helicase C domain 1) [71]. Since 2007,

several genome-wide association studies (GWAS) [72–74] reported a number of association genes for type 1 diabetes, including *ERBB3* (v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 3), *RAB5B* (RAB5B, member RAS oncogene family), SUOX (sulfite oxidase), *RPS26* (ribosomal protein S26), *CDK2* (cyclin-dependent kinase 2) and *UBASH3A* (ubiquitin associated and SH3 domain containing A). An independent study confirmed the association of UBASH3A with chronic autoimmune destruction and type 1 diabetes [75]. To date, 59 susceptibility loci were identified with more than 40 associated genes implicated for type 1 diabetes [71] (Figure 1-8).

Figure 1-8 Suscpetiability genes for type 1 diabetes, adapted from [71].



The pathogenesis of type 2 diabetes is underscored by the interplay between impaired beta-cell function and insulin sensitivity [76]. Despite of the extensive efforts for years, linkage

and candidate-gene studies only identified *PPARG* (peroxisome proliferator-activated receptor gamma), *KCNJ11* (potassium inwardly-rectifying channel, subfamily J, member 11) and *TCF7L2* (transcription factor 7-like 2, T-cell specific, HMG-box) as susceptibility genes for the polygenic form of type 2 diabetes [77]. The advent of GWAS accelerated the detection of susceptibility loci for type 2 diabetes. It was suggested that published GWAS more frequently mapped susceptibility genes for beta-cell dysfunction compared to genes for insulin resistance [78]. For example, *ADCY5* (adenylate cyclase 5), *PROX1* (prospero homeobox 1), GCK (glucokinase, hexokinase 4), *GCKR* (glucokinase/hexokinase 4 regulator), and *DGKB/TMEM195* (diacylglycerol kinase, beta/ transmembrane protein 195) were implicated for fasting glucose/beta-cell dysfunction , whereas GCKR was the only gene significantly associated with insulin resistance in the MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium) study [79]. In the DIAGRAM+ (DIAbetes Genetics Replication and Meta-analysis plus additional cohorts) study, 10 genes including *MTNR1B* (melatonin receptor 1B), *SLC30A8* (solute carrier family 30/zinc transporter, member 8), *THADA* (thyroid adenoma associated), *TCF7L2*, *KCNQ1* (potassium voltage-gated channel, KQT-like subfamily, member 1), *CAMK1D* (calcium/calmodulin-dependent protein kinase ID), *CDKAL1* (CDK5 regulatory subunit associated protein 1-like 1), *IGF2BP2* (insulin-like growth factor 2 mRNA binding protein 2), *HNF1B* (HNF1 homeobox B), and *CENTD2* (ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 1) were associated with beta-cell function, whereas only *PRARG*, *FTO* (fat mass and obesity associated), and *KLF14* (Kruppel-like factor 14) showed association with insulin sensitivity [80].

1.2 Aim of this thesis

As aforementioned, despite the extensive research efforts, the pathophysiology of diabetic retinopathy remains elusive, as a multitude of questions in respect of the retinopathy development were unexplored. One of the salient features of the chronic development of retinopathy is manifested in the clinical diagnosis as the classification of retinopathy into multiple severity stages. However, so far there have been few epidemiological studies aiming to uncover the progression and remission of retinopathy during the duration of diabetes, and the influence of population risk factors on this development. Additionally, in genetics studies, the heritable components underlying the severity of retinopathy with the consideration of the multi-state development were rarely addressed. Finally, studies to date that assessed genetic determinants for diabetic retinopathy were weakly powered in the study sample size and variants ([28,81–83]), with a large fraction of genetic susceptibility for diabetic retinopathy remaining unexplained.

Thus, the corollary to these neglected questions was the birth of this thesis, which , focused on answering three clinically important questions: (1) given the longitudinal nature of the GoDARTS (Genetics of Diabetes Audit and Research in Tayside Scotland) study cohort and the categorical feature of retinopathy data, can we decipher determinants that affect changes in retinopathy severity over the duration of diabetes? (2) With the abundance of genomic data for the GoDARTS cohort, can we estimate heritability and genetic correlations explained by common SNPs for diabetic retinopathy and related risk factors? (3) With the additional independent retinopathy data available from IMI-SUMMIT (SUrrogate markers for Micro- and Macro-vascular hard endpoints for Innovative diabetes Tools; a pan-European research consortium funded by the Innovative Medicines Initiative;

www.imi-summit.eu), can our study with an increased sample size compared to any of the published GWAS [28,81–83], be more successful in detecting and replicating risk loci for diabetic retinopathy?

Three studies were included in this thesis to answer these questions.

(1) The influence of population risk factors on diabetic retinopathy severity over the duration of diabetes. The project plan was to capture the longitudinal data of retinopathy including intermediate retinopathy events during the follow-up in GoDARTS diabetic patients, and model the patient data in a multi-state Markov model for the inference of covariate effects.

(2) Heritability and genetic correlations explained by common SNPs for diabetic retinopathy and related risk factors. The project plan was to utilise the Gibbs Sampling algorithm for the Bayesian inference of variance components in the analysis of narrow-sense heritability and genetic correlation captured by common genome-wide SNPs for diabetic retinopathy and population risk factors.

(3) The SUMMIT genome-wide meta-analysis of diabetic retinopathy. The project plan was to identify diabetic retinopathy risk loci in human genome by association mapping in the GoDARTS cohort and additionally with independent diabetic retinopathy samples from SUMMIT collaborators.

Together, these three projects form a coherent and comprehensive investigation of clinical and genetic basis for the development of diabetic retinopathy, and should provide an

increased understanding of diabetic retinopathy pathogenesis and prognosis for scientists and clinicians.

Chapter 2

Methods

2.1 Description of clinical data

The Health Informatics Centre (HIC) of the University of Dundee manages and supplies the anonymised patient data owned by the National Health Service (NHS) Tayside, in accordance with the Standard Operating Procedures approved by the Caldicott Guardians, for medical research purposes. Patients registered with a general practitioner (GP) in Scotland are allocated a unique 10-digit identification number, a Community Health Index, which is used for identifying patients' information including the address, the GP registration status and the date of death for a deceased patient. Community Health Indices for Tayside patients are centralised in the Community Health Master Patient Index database held by the Tayside Health Board. The Community Health Index as a unique patient

identifier is in use of all heath care activities, and is useful in medical research for patients' record linkage.

2.1.1 DARTS database

The diabetes audit and research in Tayside Scotland (DARTS) database is a comprehensive and up-to-date register of diabetes in Tayside, and it includes electronic records assimilated from multiple independent sources including the regional biochemistry and diabetes eye screening data [84]. The DARTS register has been shown as an enriched source \ in identifying diabetes in the Tayside population with high sensitivity and positive predictive value (0.96 and 0.95 respectively) [84].

2.1.2 GoDARTS database

Since October 1997, registered DARTS patients have been invited to give informed consent to DNA in the Genetics of DARTS (GoDARTS) study, supported by the Wellcome Trust United Kingdom Type2 Diabetes case control collaborative study and approved by the Tayside Committee for Medical Research Ethics. Between December 1998 and May 2009, 17,602 patients (including 9,829 type 2 diabetic patients) of European ancestry were recruited in Tayside. Patients' blood samples were collected for DNA (deoxyribonucleic acid) extraction and genotyping. Participants were allocated a unique anonymised system identifier. Clinical characteristics of participants were recorded at recruitment.

2.1.3 Diabetes eye screening data

As of June 2011, the diabetes eye screening data originated from the SCI-DC (the Scottish Care Information – Diabetes Collaboration; that is, the Scottish national diabetes disease database) system includes the diagnosis since 1990 of diabetes-related eye conditions including cataract, glaucoma, laser photocoagulation treatment, diabetic retinopathy and maculopathy for 8,910 Tayside patients. The diabetes eye screening database is a constellation of ophthalmology data collated from multiple sources including national and regional retinopathy screening programs, diabetes clinics and regional hospitals, since the diagnosis of diabetes in patients. The diabetes eye screening data documented techniques and specialists for ophthalmic diagnosis. Diabetic retinopathy and maculopathy severity stages (Table 1-2 and Table 1-3) are determined from grading of single-field 45 degree retinal photographs for both eyes where staged mydriasis is given.

2.1.4 Clinical phenotype data

The Tayside laboratory system documents the outcome from clinical pathology tests performed in regional surgeries, clinics and hospitals. The clinical pathology data is dated from 1992 and available in biochemistry, haematology, microbiology, virology and serology sections. Measures of total cholesterol, serum creatinine, glycatedhaemoglobin (HbA1c), high-density lipoprotein cholesterol (HDL-c) and triglycerides that are relevant population risk factors in this study were ascertained from the Tayside laboratory biochemistry database. As of June 2011, biochemistry data for 17,575 GoDARTS patients has been available through the Tayside laboratory database.

Additional Tayside-based longitudinal patient data for blood pressure (BP including diastolic and systolic BP, or DBP and SBP), body mass index (BMI) and HbA1c variables has been available through the SCI-DC network. As of June 2011, the SCI-DC system included data from 9,498 GoDARTS individuals for BP, 9,509 individuals for BMI and 9,510 individuals for HbA1c.

In addition to the routinely collected clinical data, common clinical variables including BMI, total cholesterol, serum creatinine, BP, HbA1c, HDL-c, triglycerides and smoking status were measured in the participants at the GoDARTS recruitment time. As of June 2011, 16,131 patients' recruitment data is available. This number is lower than the total GoDARTS cohort as baseline phenotypic data was not collected between 1997 and 2004.

2.1.5 Demographic data

As of June 2011, 17,602 Tayside patients' demographic data including the gender, the date of birth, the recruitment date of GoDARTS and the date of death if a patient is deceased, has been available through HIC.

2.2 Description of genetic data

In the GoDARTS study, Genomic DNA was extracted from blood specimens following the Promega Wizard or Qiagen procedures at Professor Colin Palmer's laboratory at Ninewells Hospital and Medical School. DNA samples were normalised and tiled to 96 and 384 well plates for genotyping.

2.2.1 Affymetrix Genome-wide Human SNP Array 6.0

The Affymetrix Genome-wide Human SNP Array 6.0 provides the genomic coverage for 906,600 single-nucleotide polymorphisms (SNPs) of the human genome build 36, constituting of 482,000 historical SNPs an unbiased selection from Affymetrix Genome-wide Human SNP Array 5.0 and 424,000 additional SNPs that are tag, mitochondrial, or sex chromosomal SNPs, or SNPs newly added to the dbSNP database or in recombination hotspots [85]. 4,000 GoDARTS diabetic patients that had received statin treatment since the recruitment were genotyped on the Affymetrix Genome-wide Human SNP Array 6.0 at the Affymetrix Service Laboratory, California. Prior to genotyping, DNA samples were quality assessed in order to eliminate experimental errors. Genotyping data intensities were normalised, and genotypes were called using the CHIAMO algorithm [72]. The Affymetrix Service Laboratory also provided preliminary quality control (QC) assessment on the genotype data. Further QC on this data was performed by the Wellcome Trust Case Control Consortium 2 (WTCCC2) study group as described in [86,87].

2.2.2 Illumina HumanOmniExpress BeadChip

The Illumina HumanOmniExpress BeadChip provides the coverage for over 700,000 tag SNPs of the human genome build 37, spanning up to 90% genomic regions for European and Asian ancestry populations of the International HapMap Project [79]. Supported by the SUMMIT consortium, additional 4,000 GoDARTS diabetic patients that mostly had not been genotyped on the Affymetrix Genome-wide Human SNP Array 6.0, and with serious retinal, renal and cardiovascular complications were genotyped on the Illumina

HumanOmniExpress BeadChip at the Diabetes Centre of the University of Lund, Sweden. Data normalization and genotype calling was implemented in the GenomeStudio Software (Illumina, Inc.) environment.

2.2.3 1000 Genomes reference data

The 1000 Genomes Project was initiated in 2008 with the aim to catalogue most of the genome variants with frequencies of 1% or higher in the populations studied. Given the high cost of deep sequencing for whole genomes and the limited number of haplotypes in any specific genomic region, the plan of the 1000 Genomes Project was to sequence 2,500 samples from diverse ethnic populations at low-coverage (4X; that is, the amount of DNA sequence equivalent to 4 times across the genome), and the remaining unidentified variants for each individual can be frequently inferred from the sequence data for the overall sample [89,90].

To inform whether the strategy of light sequencing is adequate in meeting the goal of the project, three pilot studies were carried out: low-coverage (2-4X) sequencing of 179 individuals from four demographic regions; high-coverage (20-60X) sequencing of a pair of mother–father–child trios; and sequencing of exomes (50X) in 697 individuals from seven populations [89]. The pilot studies published in October 2010 characterised 15 million SNPs, 1 million short insertions and deletions, and 20,000 structural variants [89]. The pilot data and methods supported the design of the full-scale project [89].

In October 2012, the 1000 Genomes Project Consortium published low-coverage genome-wide and exome sequence data for 1,092 individuals from 14 populations [90]. The study identified 38 million SNPs, 1.4 million short insertions and deletions, and over 14,000 larger deletions, with the promise for sequencing additional 1,500 individuals from 12 different populations in the last phase of the project [90].

We utilised 1000 Genomes phase 1 integrated haplotype data version 3 (released in March 2012) for genotype imputation and the assessment of population structure in the genotyped GoDARTS sample.

2.3 The influence of population risk factors on diabetic retinopathy severity over the duration of diabetes

2.3.1 Study sample

We performed a prospective cohort study of diabetic retinopathy in the GoDARTS population sample. The ophthalmology data used in this study were from the complete calendar years 1990 to 2011. GoDARTS is a study of patients with a diagnosis of type 2 diabetes, but we further reduced the chance of including misclassified type 1 diabetes patients by only considering subjects who were diagnosed with diabetes at 35 years of age or older. The cohort included patients who had at least two longitudinal retinal records. We observed that the numbers of diabetic retinopathy events and the distributions of follow-up time collected for both eyes were comparable, and to preclude the artefacts reflected as observed remission and recurrence of the proliferative phase of diabetic retinopathy,

produced from compounding longitudinal data from both eyes, we collated and analysed retinopathy data from the same eye.

The primary start point for this study obtained from this data set was the first retinal record indicative of no retinopathy within one year from the date of diabetes diagnosis. Patients were followed until either the onset of severe non-proliferative/proliferative diabetic retinopathy, their date of death, or 16 years duration of diabetes. Intermediate retinopathy observations were included in this study.

Additional independent data sets (e.g., demography and regional biochemistry database) were integrated through electronic record linkage. Population risk factors extracted were sex, smoking status (ever smoked against never smoked) and longitudinal records of age, BMI, total cholesterol, serum creatinine, DBP, HbA1c, HDL-c, SBP and triglycerides. Non-high density lipoprotein cholesterol (non-HDL-c) was estimated from total cholesterol and HDL-c measurements recorded on an identical date. As low-density lipoprotein cholesterol (LDL-c) measurements were often missing, throughout this study, non-HDL-c was considered as a valid surrogate for LDL-c estimation based on the Friedewald formula (LDL-c ≈ total cholesterol – HDL-c – k*triglycerides, where k is 0.20 if the measurements are in mg/dl and 0.45 if in mmol/l). The concordance was shown in this study as the Pearson's correlation coefficient (0.987) we attained using weighted means of longitudinal non-HDL-c and LDL-c measures in the overall GoDARTS sample (16,928 patients). Time-variant covariates were matched to a retinal event that occurred at the closest time point. Covariates measured on a quantitative scale were standardised by sample mean and

standard deviation (SD). In this study, we only included patients with the complete set of covariate data.

2.3.2 Statistical analysis

2.3.2.1 Multi-state Markov model

A multi-state Markov model was fitted to the panel data of diabetic retinopathy in this study. The multi-state Markov model depicts the movement across a series of categorical states over the observational time. The detailed description of this statistical framework for the model was published elsewhere [91,92]. In brief, suppose at an arbitrary time *t*, an individual is observed in the discrete state *i*. The movement into a state *j* ($i \neq j$) at a later observational time $t + \Delta t$ is dependent on the instantaneous propensity of moving from state *i* to *j* (the instantaneous incidence rate, or formally the transition intensity), time elapsed and additionally, time-dependent explanatory variables. Formally, the transition intensity is described by:

$$q_{ij}(t) = \lim_{\Delta t \to 0} \frac{P\{S(t+\Delta t) = j \mid S(t) = i\}}{\Delta t} \qquad 0.1$$

where $S(t)$ is the state observed at time *t*. Considering all possible state transitions, the corresponding transition intensities form the entries in a transition intensity matrix, where by convention rows sum to 0 and the transition intensities for no state change is conventionally defined as

$$q_{ii}(t) = -\sum_{j \neq i} q_{ij}(t) \qquad 0.1$$

In the model specification, a non-feasible or non-permitted instantaneous state transition is pre-set to 0.

In contrast to a transition intensity measure, which expresses an instantaneous propensity, given an elapsed time period, the probability of a state transition observed at two time points can be estimated from the matrix exponential operation of the scaled transition intensity. Formally, the transition probability of being in state j at time $t + u$ given the state at time t is state i is represented as $p_{ij}(t, t+u)$. The probability of a state transition does not suggest the number of transitions occurred for an observed state change during a time period, and the process may have passed through other states between time t and $t + u$.

In a multi-state hidden Markov model, true underlying states are considered as unobserved and observed states are assumed to be reflective of hidden true states. For a patient at an observational time, the observed state is generally conditioned on the true state according to a categorical distribution.

Covariates are introduced as proportional to the baseline transition rate expressed as

$$q_{ij}(x) = q_{ij} \exp(x^T \beta_{ij}) \qquad 0.1$$

where β_{ij} is the vector of regression coefficients associated with the vector of covariables x for the transition between the states i and j; T denotes the transposition of a matrix.

In the process of parameter estimation, the estimates are iteratively updated until the likelihood for observing the data given a set of parameter is maximised. The “msm” package (version 1.1.1) in the R (version 2.14.2) programming environment includes maximum likelihood estimation (MLE) methods for a number of observational schemes

including intermittent observation time, observations from a hidden Markov process, censor states, or a mixture of these schemes.

2.3.2.2 Likelihood ratio test and Akaike's information criterion (AIC)

To identify the better fitted model from the null and the alternative models in the study sample, the likelihood ratio test and the AIC statistic were considered. The likelihood ratio test statistic (D) is formally defined as

$$D = -2\log\left(\frac{Likelihood_{Null}}{Likelihood_{Alternative}}\right) \qquad 0.1$$

The probability distribution of the test statistic D is approximated by the chi-squared distribution with ($df_2 - df_1$) degrees of freedom, where df_1 and df_2 are numbers of parameters in the null and alternative models, respectively. Thus, the likelihood ratio test provides a p-value based statistical framework for hypothesis testing. However, the likelihood ratio test is applicable to nested models only, where the complex alternative model can be simplified into the null model by imposing constraints on the model parameters [93].

With more parameters in the fitting of a complex model, it is possible to inflate the likelihood of the parameter values given the observed data, with the consequence of overfitting (that is, when the model fitting is unduly complex, and the statistical inference is inaccurately based on the noise instead of the true relationship between variables). The AIC statistic penalises a model's likelihood by the number of parameters involved, and is commonly used in model selection. The AIC is defined as

$$AIC = 2k - 2\log(L) \qquad 0.1$$

where *k* is the number of parameters, and *L* is the likelihood of a model. With multiple candidate models fitted to the same data, the most suitable model is selected from the lowestAIC value. The AIC measure is not restricted to the comparison of nested models [94].

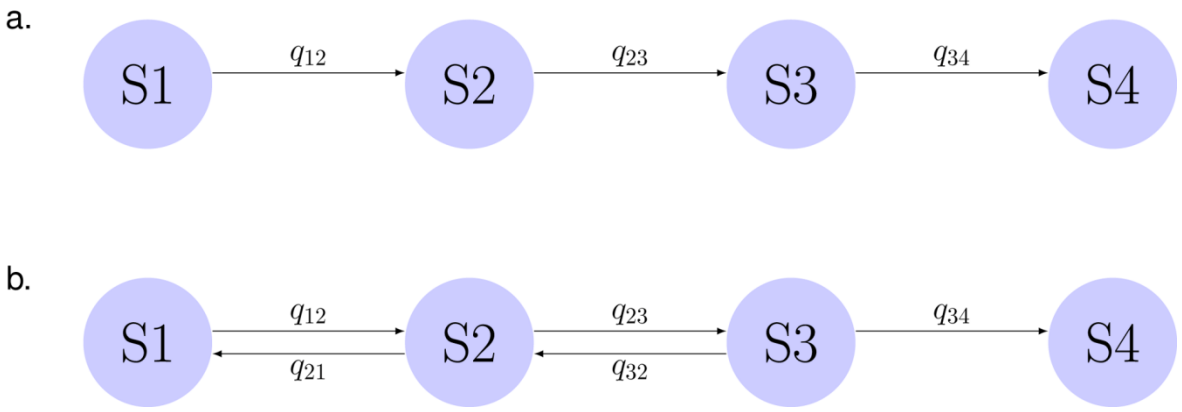
2.3.2.3 Model fitting and comparison

The discrete, non-overlapping stages of diabetic retinopathy were translated into distinctive states in the multi-state model. The effect of explanatory variables on diabetic retinopathy development is modelled in an adapted form of proportional hazard model (Equation 0.1) [92]. Patients' diabetic duration at retinal examination was considered in the model. Between follow-up visits, patients' diabetic retinopathy development is usually unmonitored, and the exact time of transition from one state to the other is unknown. Thus, we specified a relevant sampling scheme to accommodate an intermittently-observed disease process.

We postulated two baseline multi-state models, which together aimed to decipher the process underlying the development of diabetic retinopathy from the observed retinal event data by modelling distinct putative transition paths between states. In the first model, diabetic retinopathy development is modelled as one-way progression (Figure 2-1a; Equation 0.6), and misclassification was allowed to occur between adjacent states except for the absorbing state (Equation 0.7). The second model is specified by a two-way transition intensity matrix and an identical mis-classification probability matrix (Figure 2-1b; Equation 0.7 and 0.8). The best-fitted model was selected from the likelihood ratio

tests and AIC statistic. The selected model was then used for assessing covariate effects. Covariate model selection procedures also utilised LR and AIC measures.

Figure 2-1 (a) A base multi-state Markov model describes one-way transition of four states of diabetic retinopathy. (b) A second base model describes two-way transition of four states of diabetic retinopathy. The process of entering the final absorbing state is irreversible. Rates of transition (or, transition intensities) are specified as q_{ij}, where transition occurs from the current state *i* to the future state *j*.



2.3.2.1 One-way transition with mis-classification

The matrix of transition intensities is specified by

$$Q_1 = \begin{pmatrix} q_{11} & q_{12} & 0 & 0 \\ 0 & q_{22} & q_{23} & 0 \\ 0 & 0 & q_{33} & q_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad 0.1$$

where each entry q_{ij} denotes the instantaneous risk of transition from state i to state j, and $q_{ii} = -\sum_{j \neq i} q_{ij}$ for i = 1,2,3,4. The matrix of classification probabilities is of the form

$$E_1 = \begin{pmatrix} e_{11} & e_{12} & e_{13} & e_{14} \\ e_{21} & e_{22} & e_{23} & e_{24} \\ e_{31} & e_{32} & e_{33} & e_{34} \\ e_{41} & e_{42} & e_{43} & e_{44} \end{pmatrix} \qquad 0.1$$

where each entry e_{ij} gives the probability of observing state j given that the true state is i, and $e_{ii} = 1 - \sum_{j \neq i} e_{ij}$ for i = 1,2,3,4.

2.3.2.2 Two-way transition with mis-classification

The matrix of transition intensities is defined by

$$Q_2 = \begin{pmatrix} q_{11} & q_{12} & 0 & 0 \\ q_{21} & q_{22} & q_{23} & 0 \\ 0 & q_{32} & q_{33} & q_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad 0.1$$

Entries of matrix Q_2 are similarly defined as in matrix Q_1. The classification probability matrix is specified identically as in E_1.

2.4 Heritability and genetic correlations explained by common SNPs for diabetic retinopathy and related risk factors

2.4.1 Study sample

The study sample was ascertained from the GoDARTS cohort, which includes genetically unrelated individuals and some families with diabetes. Diabetic patients underwent a complete ophthalmologic examination in diabetes eye screening, as previously described. The worst eye grade of diabetic retinopathy for each patient up to the GoDARTS recruitment time was collated. Severity of diabetic retinopathy was classified as none, mild background retinopathy, observable background and severe non-proliferative/proliferative retinopathy based on the Scottish diabetic retinopathy grading scheme (Table 1-2). The diagnosis of severe non-proliferative/proliferative retinopathy was also identified from the evidence for laser photocoagulation treatment. Retinopathy samples were pooled from both type 1 and type 2 diabetic populations.

We additionally extracted demographic data and biochemistry measurements taken immediately prior to or on the GoDARTS recruitment date. We log-transformed and subsequently standardised BMI, total cholesterol, serum creatinine, HbA1c, HDL-c, SBP and triglycerides.

2.4.2 Genotype data

2.4.2.1 Typed SNP data quality control

SNP data for 3,734 GoDARTS patients typed on the Affymetrix Genome-wide Human SNP Array 6.0 was included in this analysis. We included biallelic autosomal SNP markers based on the quality control (QC) criteria: missingness < 0.03, Hardy-Weinberg equilibrium (HWE; $P > 10^{-3}$) and minor allele frequency (MAF) > 0.005. SNP alleles were aligned to the forward/+ strand, and the rs numbers and genome positions of SNPs were uplifted from the human genome build 36 to build 37. This retained 732,651 SNPs for the Affymetrix genotype data.

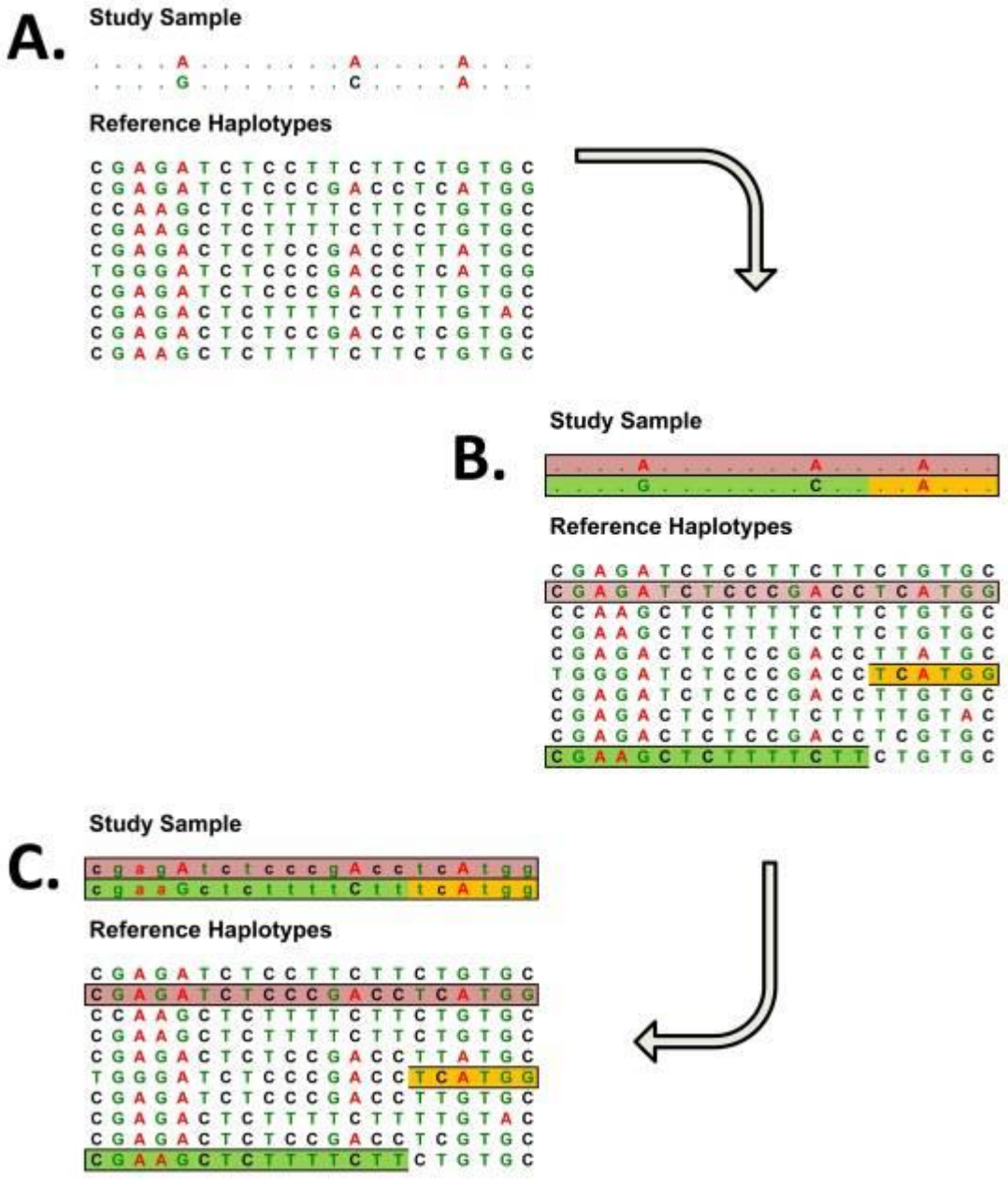
Study subjects were checked for the match between clinical reported and genotype genders, and additionally were filtered based on the missingness threshold of 0.035, heterozygosity (the false discovery rate, or FDR, at level 0.01, following the Benjamini-Hochberge controlling procedure [95], or BH95). Population stratification was assessed in the multidimensional scaling analysis by integrating with SNP data of 1000 Genomes reference populations.

2.4.2.2 Imputation

We used the imputation technique to augment genome-wide SNPs in this study. Imputation is a statistical method for predicting unobserved genotypes in the study sample based on the correlation of observed study haplotypes guided by the reference haplotypes of a similar genetic population [96] (Figure 2-2). To reduce the computational burden, prior to imputation, insertions and deletions, structural variants, singletons (these are, SNPs that

have only one copy of the minor allele in the sample genotype data) and monomorphic SNPs (these are, SNPs with a single form of allele in the population studied) were removed from the 1000 Genomes Project European CEU data (phase 1 version 3; 379 individuals), resulting in over 12.3 million SNPs retained in the reference panel. To reliably impute our genotype data that includes genetic related individuals, we utilised the MaCH-Admix program (version beta 2.0.185), which has been developed with the capabilities of imputing individuals independently based on preceding calibration of model parameters [97]. More than 12.3 million SNPs were successfully imputed for the Affymetrix genotyped sample.

Figure 2-2 The method of genotype imputation applied to a genetically unrelated sample [98]. The study sample consists a set of genotypes that are present in a more densely genotyped reference panel (panel A). The haplotypes of the study samples are compared to that in the reference panel for a match (panel B). Missing genotypes are imputed to the study sample based on the genotypes in the matched haplotype (panel C).



2.4.2.3 Post-imputation genetic and sample data quality control

We included imputed SNPs based on the MACH $\hat{r}^2$ measure (that is, the estimate for the squared correlation between imputed and true genotypes; [99]) ≥ 0.5 and MAF ≥ 0.01. We assessed minor allele counts for each imputed SNP, and none was below 10. We excluded SNPs that have triple alleles reported in the dbSNP database, and common SNPs (MAF ≥ 0.05) with a HWE *P* value < 10^{-3} or less common SNPs (MAF < 0.05) with a HWE *P* value < 10^{-6} in diabetic retinopathy controls (these are defined as patients with four years' duration of diabetes but without retinopathy, maculopathy, or evidence for laser photocoagulation treatment).

We filtered out imputed SNPs with the MACH $\hat{r}^2$ measure < 0.98, and used the most probable genotypes of these SNPs only for the estimation of identity-by-descent (IBD). We identified individuals marked by the same identifier but with the IBD estimate < 0.80, far less than the expected IBD estimate of 1.00. Additionally, we identified individuals with non-identical birth dates but strongly closely related (IBD > 0.80). We removed these mis-identified individuals from the sample, resulting in a final sample of 3,701 individuals.

2.4.3 Statistical analysis

Owing to the intractability of the published software for estimating heritability of a polychotomous trait (for example, retinopathy) and genetic (or residual) correlation between polychotomous and quantitative traits (for example, retinopathy with one of the described risk factors) from genome-wide SNP data, we used Gibbs sampling for

estimating variance components. The strategy for Bayesian inference in multivariate models using Gibbs sampling was previously published [100,101], and thanks to the work of Dr. Minghui Wang, Birmingham University, the efficient implementation of these models was made feasible by reducing the computational time for Henderson's Mixed Model Equations in Gibbs sampling. We leveraged his software for the analysis of heritability and genetic correlations in common genome-wide SNPs for diabetic retinopathy and related risk factors. We briefly describe the basis for these models.

2.4.3.1 Phenotype correlations

Phenotype correlations between a pair of quantitative traits, a single ordinal trait with a continuous outcome, and two ordinal traits, namely, Pearson, polyserial and polychoric correlations respectively, were estimated using the "polycor" package (version 0.7.8) in the R programming environment (version 3.0.0). Confidence intervals were estimated from Fisher transformation as follows [102]

$$z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \qquad 0.1$$

where r is the sample correlation coefficient. The z statistic is approximately normally distributed with mean and variance [102]

$$E(z) = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$
$$Var(z) = \frac{1}{N-3} \qquad 0.1$$

where ρ is the population correlation coefficient; N is the sample size.

2.4.3.2 Genetic similarity matrix estimation

The genetic similarity matrix is estimated from the sum of products of SNP correlation coefficients for pairs of individuals in Fortran 90 programming environment, which enables efficient computation with millions of SNPs. We used whole genome imputed SNP dosage data to construct the relatedness matrix ($\hat{G}$) from standardised genotypes [103] as follows:

$$\hat{G}_{ij} = \frac{1}{L}\sum_{l=1}^{L}\frac{(x_{il}-2p_l)(x_{jl}-2p_l)}{2p_lq_l} \qquad 0.1$$

where individuals *i* and *j* can be the same or different individuals, and x_{il} (x_{jl}) is the genotype dosage based on the minor allele in individual *i* (*j*) at locus *l* with allele frequency p_l and $q_l = 1 - p_l$, where there are L loci in total under consideration. This pairwise relatedness measure can be either positive or negative in value, and thus, the term of "similarity" is prevailing in literature instead of "relationship" in describing this matrix.

To ensure allele frequencies used in computing the genetic similarity matrix was unbiased by sample structure, we kept 6,830,657 imputed autosomal SNPs with minor allele frequencies more than 0.01 in the genetically unrelated sample (IBD less than 0.05) for the estimation of the genetic similarity matrix.

Several related estimation approaches with slight modifications for the genetic similarity matrix have been published, including a relatedness matrix constructed from cantered genotypes $\hat{G}_{ij} = \frac{1}{L}\sum_{l=1}^{L}(x_{il}-2p_l)(x_{jl}-2p_l)$ [103] and an adjustment in estimating the standardised genotype relatedness matrix for the same individuals using

$\hat{G}_{ii} = 1 + \frac{1}{L}\sum_{l=1}^{L}\frac{x_{il}^2 - (1+2p_l)x_{il} + 2p_l^2}{2p_l q_l}$, in order to minimise sampling noise [104]. These matrices performed similarly in modelling of narrow-sense heritability [103]. However, occasionally one or more of these matrices may not be invertible, rendering the modelling process intractable, and thus any matrix that is positive definitive [105] is preferred.

2.4.3.3 Analysis of a single trait

With the estimated genetic similarity matrix, we used univariate mixed models to partition phenotypic variations of single quantitative or polychotomous traits into genetic and residual variations, and the theoretical strategies were previously presented [100,104,106]. For a quantitative trait, the linear mixed model relating phenotypes to genetic relatedness is described by

$$\begin{aligned} y &= X\beta + g + \varepsilon \\ g \mid G_n, \sigma_g^2 &\sim MVN_n(0, G_n\sigma_g^2) \\ \varepsilon \mid \sigma_\varepsilon^2 &\sim MVN_n(0, I_n\sigma_\varepsilon^2) \end{aligned} \qquad 0.1$$

where y is a vector of single-trait quantitative measures for n individuals; β is a vector of regression coefficients of fixed effects for the overall mean and covariates; g is an n-vector of total additive genetic effects, and is treated as random effects normally distributed with variance σ_g^2 and correlation structure represented by additive genetic relatedness matrix G_n; X is a design matrix relating fixed effects to individuals; ε is an n-vector of residual terms normally distributed with variance σ_ε^2 and correlation structure represented by identity matrix I_n. MVN_n represents multivariate normal distribution of order n. Therein, the variance structure of the response variable (V) is

$$V = G_n\sigma_g^2 + I_n\sigma_\varepsilon^2 \qquad 0.1$$

The coefficient of heritability h_g^2 is estimated by $\sigma_g^2 / \left(\sigma_g^2 + \sigma_\varepsilon^2\right)$.

An ordered categorical trait in the studies of quantitative genetics is considered as discrete observed classes derived from an underlying risk gradient that is often unobserved, through fixed thresholds [107]. This latent quantitative variable is named liability (*l*) in genetics studies [107]. For example, considering hypertension versus non-hypertension, the observation would be in the hypertensive category, if the blood pressure measurement (SBP/DBP) at the rest state exceeds the threshold of 140/90 mmHg. The liability model is defined as [107]

$$\begin{aligned} l &= X\beta + g + \varepsilon \\ l \mid \beta, g, \sigma_\varepsilon^2 &\sim MVN_n(X\beta + g, I_n\sigma_\varepsilon^2) \end{aligned} \qquad 0.1$$

where *l* is an *n*-vector of liabilities which are multivariate normally distributed with mean $X\beta + g$ and variance-covariance matrix $I_n\sigma_\varepsilon^2$; by convention, residual variance for liabilities is parameterised as $\sigma_\varepsilon^2 = 1$ [107]; other terms are identically defined as in the linear mixed model, but are quantified on a liability scale.

In the liability model, individuals with observed ordinal categories ($y = \{y_i\}$ where *i* indexes the *i*th individual) have liabilities bounded by a pair of threshold values in the set ($t_{\min} < t_1 < t_2 \cdots < t_{c-1} < t_{\max}$ where there are $c + 1$ hypothetical thresholds for *c* mutually exclusive but exhaustive categories; the extreme thresholds $t_{\min} = t_0$ and $t_{\max} = t_c$) [107]. For example, if a liability value is between t_0 and t_1, the assignment is into the first category . Formally, the categorical assignment is described by

$$P(y_i = j \mid \beta, g, t) = P(t_{j-1} < l_i < t_j \mid \beta, g, t) \qquad 0.1$$

where j is the categorical response (j = 1, 2, …, c) for individual i, and $t = (t_0, t_1, ..., t_c)$ is a vector of thresholds; other terms were described before. It is the convention to fix $t_0 = -\infty$, $t_1 = 0$ and $t_c = +\infty$ so as to centre the liability distribution [107]. In the threshold model for categorical traits, an unselected sample that reflects the prevalence of categories in the population is required for unbiased inference of heritability.

2.4.3.4 Analysis of correlated traits

With the estimated genetic similarity matrix, we also applied multivariate mixed models to partition phenotypic variations for a pair of related traits to genetic and residual variations and covariance for genetic and residual effects. In this analysis, each individual has measurements for two different traits such that in the matrix form,

$$\begin{pmatrix} \eta_{(1)} \\ \eta_{(2)} \end{pmatrix} = \begin{pmatrix} X_{(1)} & 0 \\ 0 & X_{(2)} \end{pmatrix} \begin{pmatrix} \beta_{(1)} \\ \beta_{(2)} \end{pmatrix} + \begin{pmatrix} g_{(1)} \\ g_{(2)} \end{pmatrix} + \begin{pmatrix} \varepsilon_{(1)} \\ \varepsilon_{(2)} \end{pmatrix} \qquad 0.1$$

where the subscripts are used to differentiate the two traits; each of $\eta_{(1)}$ and $\eta_{(2)}$ represents the observed quantity if the trait is continuous, or the measure of liability if the phenotype is polychotomous; other terms have been previously described, and are expressed on the same scale as the variable η. In this form, the variance-covariance matrices are

$$G = Var(g) = Var\begin{pmatrix} g_{(1)} \\ g_{(2)} \end{pmatrix} = \begin{pmatrix} \sigma^2_{g_{(1)}} & \sigma_{g_{(1)}g_{(2)}} \\ \sigma_{g_{(1)}g_{(2)}} & \sigma^2_{g_{(2)}} \end{pmatrix} \otimes G_n \qquad 0.1$$

$$R = Var(\varepsilon) = Var\begin{pmatrix} \varepsilon_{(1)} \\ \varepsilon_{(2)} \end{pmatrix} = \begin{pmatrix} \sigma^2_{\varepsilon_{(1)}} & \sigma_{\varepsilon_{(1)}\varepsilon_{(2)}} \\ \sigma_{\varepsilon_{(1)}\varepsilon_{(2)}} & \sigma^2_{\varepsilon_{(2)}} \end{pmatrix} \otimes I_n \qquad 0.1$$

where $\otimes$ is the kronecker product operator; G is the variance-covariance matrix of additive genetic effects across two traits; R is the variance-covariance matrix of residual

terms of two traits; $\sigma^2_{g_{(i)}}$ ($\sigma^2_{\varepsilon_{(i)}}$) is the genetic (residual) variance for the *i*th trait, and $\sigma_{g_{(1)}g_{(2)}}$ ($\sigma_{\varepsilon_{(1)}\varepsilon_{(2)}}$) is the genetic (residual) covariance between the two traits; other term were previously described. The genetic (residual) correlation coefficient r_g (r_e) is defined as $r_g = \frac{\sigma_{g_{(1)}g_{(2)}}}{\sigma_{g_{(1)}}\sigma_{g_{(2)}}}$ ($r_\varepsilon = \frac{\sigma_{\varepsilon_{(1)}\varepsilon_{(2)}}}{\sigma_{\varepsilon_{(1)}}\sigma_{\varepsilon_{(2)}}}$).

2.4.3.5 Bayesian Markov chain Monte Carlo (MCMC) inference

Analysis of single and joint quantitative traits to estimate model parameters including variance and covariance for genetic and residual components can be readily implemented using the restricted maximum likelihood (REML) method [104,108]. However, the analysis of single polychotomous trait or the joint analysis of ordinal and quantitative traits becomes intractable with the REML estimation procedure, and thus at the current time, no REML-based software is available for this analysis. The Bayesian MCMC approach provides an alternative strategy for estimating variance components in the analysis of variations of single or joint traits of quantitative and/or polychotomous nature. In this study, ten independent Markov chains were run for each trait or combination of traits based on random initials for genetic and residual variance, with starting values for fixed effects and random effects set to zero. To ensure stochastic sampling, random seeds were set for these independent chains. For the univariate mixed model, a single MCMC chain was run for 3 million iterations, with first 0.5 million samples discarded (burn-in), and thereafter samples were saved every 100 iterations. For a bivariate mixed model where both phenotypes were quantitative, a pair of chains was run according to the same specification. Either stationary chain was used in this study. If an MCMC diagnostic indicates more MCMC samples

required for either chain and both chains passed the stationarity test, samples from both chains were combined. If only one chain reached stationarity but more MCMC samples required, ten short MCMC chains were run in parallel, using the last sample of the stationary chain as the initial parameter values. Each of these short chains was run with random seeds for 0.1 million iterations without burn-in, and samples were saved every 100 iterations. Short chains were merged with the long chain. These strategies aimed to augment MCMC samples without re-implement MCMC sampling from the start, which was computationally intensive. As we anticipated intense compute time for a bivariate mixed model where at least one phenotype was ordinal, four chains were run for 1 million iterations, with first 0.5 million samples burnt in, and thereafter sampling at the interval of 100 iterations. These MCMC samples were assessed for stationarity, and were combined.

In contrast to the REML method, which evaluates the probability of observed data given a set of parameter values (that is, the likelihood), the Bayesian approach estimates the probability of a set of parameter values given observed data (that is, the posterior probability). In Bayes' theorem, the posterior probability is proportional to the product of the likelihood and the prior probabilities for the set of parameters. In combination with the Bayesian approach, MCMC methods construct a Markov chain for each model parameter and sample from the target distribution [107]. Gibbs sampling algorithm is a special case of MCMC methods, in which the joint posterior probability distribution is unknown, and each model parameter is sampled from the conditional posterior probability distribution [107].

In the analysis of single quantitative trait, the joint posterior probability based on model 0.1 is written as

$$p(\beta, g, \sigma_g^2, \sigma_\varepsilon^2 \mid y, X, G_n, v_g, S_g^2, v_\varepsilon, S_\varepsilon^2) \propto p(y \mid X, \beta, g, \sigma_\varepsilon^2) p(\beta) p(g \mid G_n, \sigma_g^2) p(\sigma_g^2 \mid v_g, S_g^2) p(\sigma_\varepsilon^2 \mid v_\varepsilon, S_\varepsilon^2) \quad 0.1$$

where $p(y \mid X, \beta, g, \sigma_\varepsilon^2)$ is the likelihood function; $p(\beta)$, $p(g \mid G_n, \sigma_g^2)$, $p(\sigma_g^2 \mid v_g, S_g^2)$ and $p(\sigma_\varepsilon^2 \mid v_\varepsilon, S_\varepsilon^2)$ are prior distributions for model parameters β (regression coefficients for the fixed effect), g (total additive genetic effects), σ_g^2 (additive genetic variance) and σ_ε^2 (residual variance), respectively; other terms are defined identically as in Model 0.1. The prior distribution for regression coefficients β is a uniform distribution [100], and total additive genetic effects g are sampled from multivariate normal distribution as specified in Model 0.1. The prior distribution for genetic and residual variance is inverse-gamma distribution with probability density function [100]

$$p(\sigma_g^2 \mid v_g, S_g^2) \propto (\sigma_g^2)^{-((v_g/2)+1)} \exp(-\frac{v_g S_g^2}{2\sigma_g^2})$$
$$p(\sigma_\varepsilon^2 \mid v_\varepsilon, S_\varepsilon^2) \propto (\sigma_\varepsilon^2)^{-((v_\varepsilon/2)+1)} \exp(-\frac{v_\varepsilon S_\varepsilon^2}{2\sigma_\varepsilon^2}) \quad 0.1$$

where v_g and S_g^2 (v_ε and S_ε^2) are hyper-parameters for genetic (residual) variance. We specified slightly informative prior parameters with $v_g = v_\varepsilon = 3$ and $S_g^2 = S_\varepsilon^2 = 1$. The Gibbs sampler updates conditional posterior probabilities for each of the model parameters (β, g, σ_g^2 and σ_ε^2) sequentially through

$$p(\beta, g \mid y, X, \sigma_\varepsilon^2) \propto p(y \mid X, \beta, g, \sigma_\varepsilon^2) p(g \mid G_n, \sigma_g^2) \quad 0.1$$

$$p(\sigma_g^2 \mid g, v_g, S_g^2) \propto p(g \mid \sigma_g^2, G_n) p(\sigma_g^2 \mid v_g, S_g^2) \quad 0.1$$

$$p(\sigma_\varepsilon^2 \mid \beta, g, \sigma_\varepsilon^2, y, X, v_\varepsilon, S_\varepsilon^2) \propto p(y \mid X, \beta, g, \sigma_\varepsilon^2) p(\sigma_\varepsilon^2 \mid v_\varepsilon, S_\varepsilon^2) \quad 0.1$$

In the analysis of a single polychotomous trait, joint posterior probability distribution for Model 0.1 is similar to that for based on a single quantitative trait (Model 0.1)

$$\begin{aligned} & p(\beta, g, \sigma_g^2, \sigma_\varepsilon^2, l, t \mid y, X, G_n, v_g, S_g^2, v_\varepsilon, S_\varepsilon^2) \\ & \propto p(y \mid l, t)\, p(l \mid X, \beta, g, \sigma_\varepsilon^2)\, p(t)\, p(\beta)\, p(g \mid G_n, \sigma_g^2)\, p(\sigma_g^2 \mid v_g, S_g^2)\, p(\sigma_\varepsilon^2 \mid v_\varepsilon, S_\varepsilon^2) \end{aligned} \qquad 0.1$$

where l is a vector of liabilities; t is a set of thresholds previously defined for Model 0.1; $p(y \mid l, t)$ is the likelihood function; $p(l \mid X, \beta, g, \sigma_\varepsilon^2)$, $p(t)$ are prior distributions for parameters l and t, respectively; other terms have been described previously. $p(l \mid X, \beta, g, \sigma_\varepsilon^2)$ is a multivariate normal density function, and $p(t)$ has a uniform distribution [100]. Parameter values were iteratively updated through a conditional posterior probability density function derived from (Model 0.1) by the Gibbs sampler, and these formulae were omitted here.

In the analysis of a pair of quantitative traits, based on Model 0.1, if we let $\eta = (\eta_{(1)}^T, \eta_{(2)}^T)^T$ and $\theta = (\beta_{(1)}^T, \beta_{(2)}^T, g_{(1)}^T, g_{(2)}^T)^T$, where T denotes transposition and η represents quantitative outcome y for the two traits, the joint posterior distribution is described by

$$p(\theta, G, R \mid y) \propto p(y \mid \theta, R)\, p(\theta \mid G)\, p(G)\, p(R) \qquad 0.1$$

where $p(y|\theta,R)$ is the likelihood function, and $p(\theta|G)$, $p(G)$ and $p(R)$ are prior probabilities for model parameters θ, G and R, respectively; for simplicity, genetic similarity matrix (G_n), the design matrix for fixed effects (X) are omitted; other terms have been previously described in Model 0.1. In the analysis of a pair of quantitative and ordinal traits jointly, η simultaneously represents quantitative outcome for the quantitative trait and liability for the ordinal trait, and joint posterior distribution is written as

$$
\begin{aligned}
& p(\theta,\eta,G,R,t \mid y) \\
& \propto p(y \mid \theta,\eta,G,R,t)\, p(\eta \mid \theta,R)\, p(\theta \mid G)\, p(G)\, p(R)\, p(t) \\
& = p(y_c \mid \eta_c,t)\, p(\eta \mid \theta,R)\, p(\theta \mid G)\, p(G)\, p(R)\, p(t)
\end{aligned}
\qquad 0.1
$$

where y_c and η_c are observed categorical outcome and liability respectively for the ordinal traits in the bivariate model; other terms were previously described. Prior uncertainty about genetic and residual covariance matrix *G* and *R* is an inverse Wishart distribution of order 3 (that is, genetic variance for the two traits respectively and covariance between these two) specified by [101]

$$
\begin{aligned}
G \mid v_g, S_g^2 &\propto \mid G \mid^{-\frac{1}{2}(v_g+3+1)} \exp\left\{ tr\left(-\frac{v_g-3-1}{2} S_g^2 G^{-1} \right) \right\} \\
R \mid v_\varepsilon, S_\varepsilon^2 &\propto \mid R \mid^{-\frac{1}{2}(v_\varepsilon+3+1)} \exp\left\{ tr\left(-\frac{v_\varepsilon-3-1}{2} S_\varepsilon^2 R^{-1} \right) \right\}
\end{aligned}
\qquad 0.1
$$

where $\mid G \mid$ ($\mid R \mid$) is the determinant of the genetic (residual) covariance matrix; *tr* denotes the operation of trace for a square matrix; v_g and S_g^2 (v_ε and S_ε^2) are parameters for the genetic (residual) variance matrix, and were set as slightly informative prior parameters with $v_g = v_\varepsilon = 5$ and $S_g^2 = S_\varepsilon^2 = I_2$; *tr* denotes the trace operation.

2.4.3.6 MCMC stationarity diagnostics

The "coda" package (version 0.16.1) in the R programming environment (version 3.0.0) provides utilities for diagnosing convergence of MCMC chains. In this study, quality control (QC) for MCMC chains was based on the Geweke and Heidelberg-Welch diagnostics, and additionally, the Gelman-Rubin multi-sequence diagnostic was applied to assess the stationarity of multiple MCMC chains. We only used MCMC data that passed through the stationarity diagnostics.

The Geweke test estimates the difference of means for a model parameter sampled from two non-overlapping parts (by default, the initial 0.1 and final 0.5 portions) in the same Markov chain [109]. When the Markov chain is convergent, no difference in the means is observed (the null hypothesis) [109]. If MCMC chain reaches stationarity, the Geweke's standard Z-score should be between -1.96 and 1.96 in keeping with alpha level (the false positive rate) of 0.05.

Additionally, we used the Heidelberg-Welch diagnostic for assessing whether a sequence of parameter values or a proportion of this sequence converged to a stationary distribution (that is, the target probability distribution, described fully in [110]). The Heidelberg-Welch diagnostic also provides the assessment whether an MCMC chain should be run for longer to have narrow halfwidth (half the width of the credible interval around the mean). The halfwidth test statistic is the halfwidth normalised by the posterior mean, and to pass the halfwidth test, the test statistic should be smaller than a target value (0.1 by default). If the posterior mean of MCMC samples is close to 0, the halfwidth test statistic is substantially large, resulting in the failure of the halfwidth test. In this case or the case where multiple parallel MCMC chains were run, the halfwidth test was not necessary and therefore not applied.

The Gelman-Rubin multi-sequence diagnostic estimates the within-chain and between-chain variances [111]. If the chains have reached stationarity, these are unbiased estimates of the variance of the stationary distribution. Otherwise, the within-chain variance is an underestimate for the variance of the stationary distribution, and the between-chain

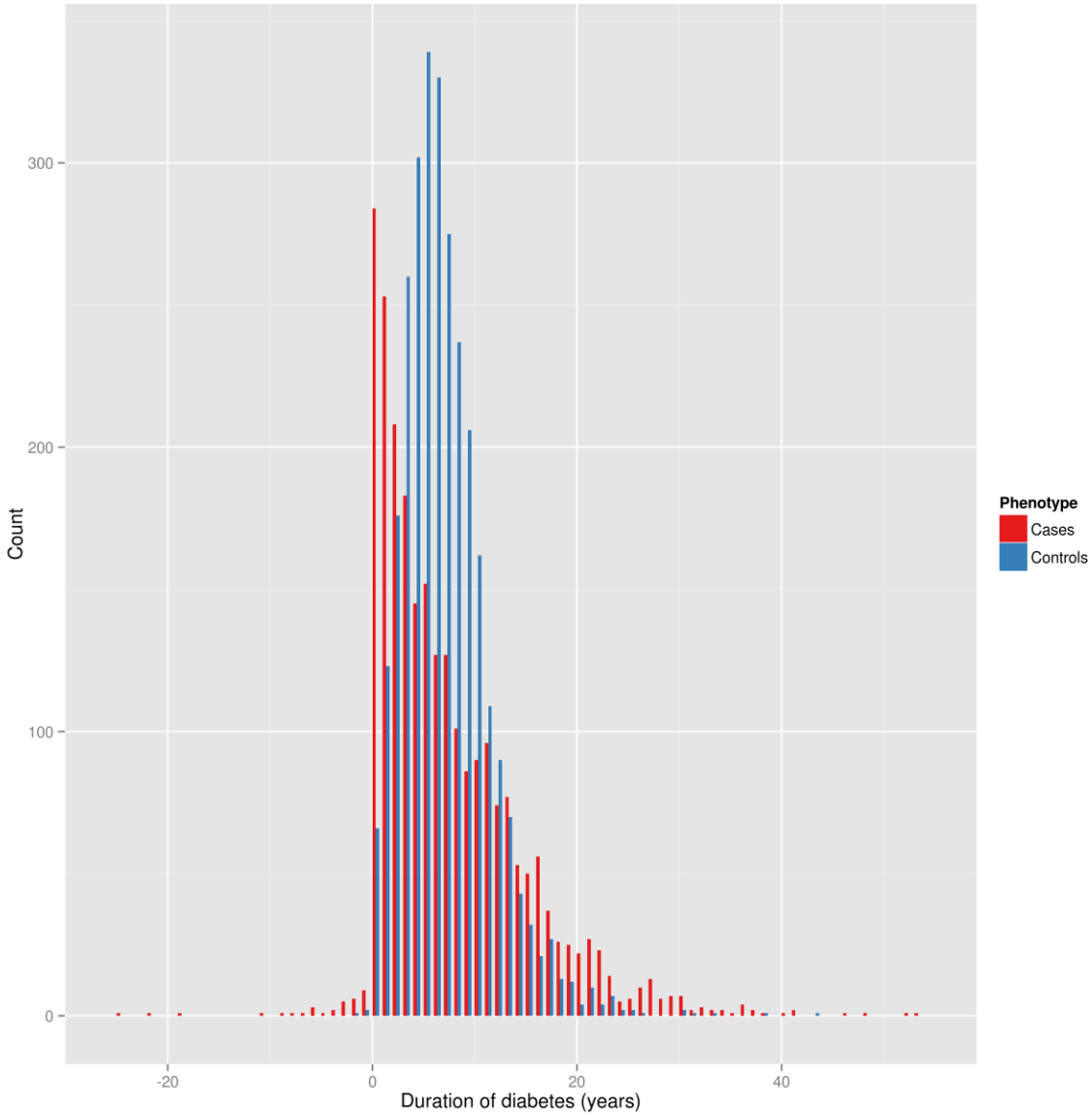
variance is an overestimate for the same parameter. The potential scale reduction factor is the reflection of the overdispersion of the stationary distribution variance compared to the within-chain variance. If this test statistic is larger than 1.2, we consider these parallel MCMC chains have not reached stationarity.

2.5 The SUMMIT genome-wide meta-analysis of diabetic retinopathy

2.5.1 Study sample

In the GoDARTS cohort, the data has been based on diabetic retinopathy, maculopathy data and the evidence for laser photocoagulation treatment ascertained from the June 2011 release of the SCI-DC diabetes eye screening data set. Observable background, severe non-proliferative and proliferative retinal events (Table 1-2) are registered in the national and regional retinal screening service or the validation database, and these programs have vigorous adherence to the national clinical guidelines. Evidence for laser photocoagulation treatment, diabetic maculopathy and data of no-retinopathy were included from all retinal examinations documented in the diabetes eye screening data. Diabetic retinopathy was defined as the diagnosis of observable background, severe non-proliferative and proliferative retinopathy, or the evidence for laser photocoagulation treatment based on the worst affected eye. Individuals with diabetes duration of at least 4 years (Figure 2-3), free of any retinopathy or maculopathy (Table 1-2 and Table 1-3) and if alive, with available retinal records within between January 2006 and June 2011 were included as controls. This definition aimed to ensure the case-control selection was robust.

Figure 2-3 Duration of diabetes for genotyped retinopathy cases (n = 2447) and controls* (n = 2968) in the GoDARTS cohort prior to the SUMMIT whole-genome meta-analysis. Negative duration of diabetes is likely to be noise associated with primary retinopathy data. *No restriction on patients' diabetes duration or the last date of retinopathy data.



In addition, we had the support from the SUMMIT cohorts for genome-wide meta-analysis. The SUMMIT study of retinopathy meta-analysis in type 1 and type 2 diabetic cohorts included the EURODIAB (Epidemiology and Prevention of Diabetes) IDDM (Insulin-Dependent Diabetes Mellitus) complications study [112], the FinnDiane (Finnish Diabetic Nephropathy) study [70] and the SDR (Scania Diabetes Registry) cohort [113].

The EURODIAB cohort is a cross-sectional study of microangiopathy and acute complications in 3,250 type 1 diabetic patients recruited from 16 nations in Europe [112]. Patients aged 15-60 stratified by age, diabetes duration and gender were invited for enrolment [112,114]. Two-field retinal photographs were taken for both eyes, and graded with reference to standard photographs [114].

The FinnDiane cohort is a cross-sectional study of genetic and environmental influences for nephropathy in 4,800 adults with type 1 diabetes recruited across Finland [70,115]. Cohort subjects were diagnosed before age 35 and offered insulin treatment from the first year of diagnosis [70,115].

The SDR cohort is a longitudinal study of diabetes and chronic complications in southern Sweden since 2000 [113,116]. At the time of the study, the SDR cohort recruited 1,264 type 1 and 5,123 type 2 diabetic patients [113].

In the SUMMIT meta-analysis, diagnosis of diabetic retinopathy was based on information on fundus photography, ophthalmoscopy or laser treatment for diabetic retinopathy. To be informative for the definition of proliferative retinopathy, the specification for the fundus

photograph was at least 45-degree coverage of the retina or 30-degree coverage of the macula. Diabetic retinopathy was defined as type 1 or type 2 diabetic patients with at least moderate background retinopathy (ETDRS>40, NSC, SDRGS=R2), mild background retinopathy (ESDRS=30-39, NSC) with duration of diabetes at retinopathy assessment < 10 years, or treated with panretinal laser therapy. Controls were defined as individuals with no recorded background retinopathy, maculopathy or panretinal laser therapy. These individuals should have at least 4 years duration of diabetes. If retinal photographs were unavailable but positive evidence of no laser treatment was available, individuals with duration of diabetes > 20 years were also included in the study as controls. Individuals with maculopathy were excluded from the control sample.

Table 2-1 Equivalence of diabetic retinopathy coding in differeent clinical classification systems, tabluated by Professor Colin Palmer and Dr Alex Doney, University of Dundee. ETDRS: the Early Treatment Diabetic Retinopathy Study; NSC: the National Screening Committee; SDRGS: the Scottish Diabetic Retinopathy Grading Scheme; BDR: background diabetic retinopathy; NPDR: non-proliferative diabetic retinopathy; PDR: proliferative diabetic retinopathy.

ETDRS	NSC	SDRGS
10 None	R0 None	R0 None
20 Microaneurysms only	R1 Background	R1 Mild BDR
35 Mild NPDR		
43 Moderate NPDR	R2 Pre-proliferative	R2 Moderate BDR
47 Moderately Severe NPDR		
53 A-D Severe NPDR		R3 Severe BDR
53 E Very severe NPDR		
61 Mild PDR	R3 Proliferative	R4 PDR
65 Moderate PDR		
71, 74 High risk PDR		
81, 85 Advanced PDR		

2.5.2 Genotype data

This study constitutes of Work Package 1 that focussed on genetic markers for type 1 and 2 diabetes complications in SUMMIT, and Natalie Van Zuydam, Oxford University, kindly offered to perform pre-imputation quality control and imputation in GoDARTS samples typed on Affymetrix Genome-wide Human SNP Array 6.0 and Illumina HumanOmniExpress BeadChip for SUMMIT project use within GoDARTS research groups. Harshal Deshmukh (Dundee University), Niina Sandholm (Helsinki University)

and Emma Ahlqvist (Lund University) performed genome-wide association scans of diabetic retinopathy for the EURODIAB, FinnDiane and SDR cohorts, respectively.

Genome-wide genotyping information is provided in Table 2-2. Standard quality control filters were applied (Table 2-3). In this study, bilallelic autosomal SNPs were included, and the coding of SNPs was aligned to the forward/+ strand of the human genome build 37.

Study subjects were reviewed for the match between clinical reported and genotype genders, and additionally were filtered for heterozygosity. Population stratification was assessed in the multidimensional scaling analysis (plotted with 1000 Genomes reference populations).

Genome-wide SNP markers were imputed to 1000 Genomes reference data (phase 1 version 3; 379 individuals). SHAPEIT (version v1.ESHG, [117]) was used for inferring haplotypes in the study samples, and IMPUTE (version 2, [96]) was employed for genotype imputation in haplotype data. The resulting imputed data is in genotype probabilities. In the FinnDiane and SDR studies, MACH (version 1.0, [118]) was also used for genotype imputation, and the resulting genotypes are expressed in dosages, which was suited for association analysis, where genotype dosages was required. Genotype probabilities can also be transformed to dosages in custom scripts following the formula as in [119]: $\sum_{x=0}^{2} P(\mathrm{G} = \mathrm{x}) * \mathrm{x}$ where x is the additively coded genotype value; G is the variable for genotype; $P(\mathrm{G} = \mathrm{x})$ is the genotype probabilities. Imputed variant data that failed the quality control procedures (Table 2-3) are not reported.

Table 2-2 Genotyping information for the meta-analysis study cohorts.

Cohort	Genotyping chip(s)
EURODIAB	Illumina HumanOmniExpress BeadChip
FinnDiane	Illumina 610Quad chip
GoDARTS	Affymetrix Human SNP Array 6.0 and Illumina HumanOmniExpress BeadChip
SDR	Illumina HumanOmniExpress BeadChip

Table 2-3 Quality control (QC) filters applied to genome-wide meta-analysis study cohorts.HWE-P: Hardy–Weinberg equilibrium P value; INFO: the IMPUTE software quality control measure [99]**; SCR: sample call rate; MAC: minor allele count; MACH r2: the MACH software quality control measure** [99]**; MAF: minor allele frequency; VCR: variant call rate.**

Cohort	Pre-imputation QC	Post-imputation QC
EURODIAB	HWE-P>10^{-6}; MAF>0.01; VCR>0.95	HWE-P>10^{-6}; INFO>0.40; MAF>0.01
FinnDiane	HWE-P>$10^{-7;}$ MAF> 0.01; SCR>0.95, VCR>0.90	HWE-P>10^{-4} for MAF<0.05; HWE-P>5.7×10^{-7} for MAF>0.05; INFO>0.40 or MACH r^2>0.30; MAF/MAC>0.01/10
GoDARTS	HWE-P>10^{-6}; MAF>0.01; VCR>0.99	HWE-P>10^{-6}; MACH r^2>0.50; MAF>0.01
SDR	HWE-P>10^{-7}; SCR>0.98; VCR>0.95	HWE-P>10^{-4} for MAF<0.05; HWE-P>5.7×10^{-7} for MAF>0.05; INFO>0.40 or MACH r^2>0.30; MAF>0.01

Table 2-4 Imputation software employed for the whole-genome meta-analysis study cohorts. IMPUTE2 (version 2.3.0); MACH (version 1.0); SHAPEIT (version 2).

Cohort	Imputation software
EURODIAB	SHAPEIT with IMPUTE2
FinnDiane	SHAPEIT with IMPUTE2; MACH
GoDARTS	SHAPEIT with IMPUTE2
SDR	SHAPEIT with IMPUTE2; MACH

2.5.3 Statistical analyses

The genetically unrelated sample was defined by an IBD estimate <0.05 between any pair of individuals, and the genetically mixed sample was inclusive of related and unrelated samples. Genome-wide association analyses were performed in each study cohort, and the whole genome meta-analysis of diabetic retinopathy was analysed by Dr. Niina Sandholm, Helsinki University. Subsequently, I annotated and summarised the data of genome-wide meta-analysis was annotated and summarised for this study. Genome-wide association analysis methodologies in each study are provided in Table 2-5. We used SNPTEST (version 2.4.1) for the genome-wide association analysis of the unrelated sample, and the genotype uncertainty was accounted for in the likelihood score test. In the analysis of the genetically mixed sample, EMMAX (version beta, [120]) or GEMMA (version 0.93, [121]) was used for efficient implementation of linear mixed models for genome-wide variants, accounting for population structure in the genetically mixed sample.

As EMMAX or GEMMA provides *P* values as the test statistic as opposed to the log odds ratio as the effect estimate in SNPTEST, genome-wide meta-analysis for the genetically mixed sample was implemented in METAL (released on 2011-03-25, [122]) based on the Z score approach. In this approach, *P* values were transformed into Z scores according to the standard normal cumulative distribution function, and the sign of the *Z* score is reflective an increased or decreased risk associated with the reference allele [122]. *Z* scores were combined across studies in a weighted sum, and the weights are proportional to the square root of study sample sizes [122]. In the genetically unrelated sample, fixed effect meta-analysis of genome-wide variants was implemented in GWAMA (version 2.1, [123]). This

methodology is based on the inverse variance weighting, in which the effect estimates weighted by the inverse of squared standard error were combined across studies [123].

Heterogeneity of effect sizes across studies were assessed by Cochran's Q test and I^2 index [124]. Cochran's Q statistic is the weighted sum of squared difference of study effects and the pooled effect in the meta-analysis. The weights are identically defined as in the Z score method or the fixed effect inverse variance model [124]. Cochran's Q statistic follows the chi-square distribution with the degree of freedom being the number of studies subtracted by one. I^2 index depicts the variation across studies attributable to heterogeneity rather than by chance, and the estimation equation is described in [124].

With the use of the CaTS (version 0.01) [125], we estimated power for each additive association test of single markers in the genome-wide meta-analysis of diabetic retinopathy, assuming the prevalence rate of 9.59% as in GoDARTS with a variability of relative risks and SNP minor allele frequency. Relative risk and odds ratio is interconvertible by [126]

$$RR = \frac{OR}{\left(1 - P_{ref}\right) + \left(P_{ref} * OR\right)} \qquad 0.1$$

where OR is the odds ratio; P_{ref} is the baseline prevalence in the non-exposed group; RR is the relative risk.

Table 2-5 Genome-wide association analysis information of each cohort. EMMAX (version beta); GEMMA (version 0.93); SNPTEST (version 2.4.1). PC1: the first principal component; PC2: the second principal component.

Cohort	Analysis software	Covariates
EURODIAB	SNPTEST and EMMAX	Age, diabetes duration, PC1, PC2, sex
FinnDiane	SNPTEST and EMMAX	Age, diabetes duration, HbA1c, PC1, PC2, sex
GoDARTS	SNPTEST and GEMMA	Age, HbA1c , sex
SDR	SNPTEST and EMMAX	Age, diabetes duration, HbA1c, sex

Chapter 3

The influence of population risk factors on diabetic retinopathy severity over the duration of diabetes

3.1 Introduction

The multi-stage clinical classification of diabetic retinopathy development has prompted the wide use of categorical data analysis strategy in clinical studies. Commonly, cross-sectional studies use diabetic retinopathy case and control samples in logistic regression analysis and/or contingency tables for modelling population risk factor effects [127–130]. Other studies have utilised the longitudinal nature of diabetic retinopathy progression in proportional hazard models [131,132]. To date, however, only one study [92] has included

intermediate states from longitudinal, multi-state diabetic retinopathy data in the analysis, an approach which provides an increased ability to decipher the stage-wise development of retinopathy compared with a simple survival analysis. In the Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS) database, we have ongoing, longitudinal collection of diabetic retinopathy clinical outcome from 1990 for Tayside patients with diabetes, and additionally we have access to all biochemistry measurements for these patients. These rich data resources enable us to investigate changes in patients' retinal status over the duration of their diabetes. A multi-state Markov model was developed to analyse panel data of a complex, multi-staged disease process in continuous time [133]. This longitudinal analysis approach has recently been applied in a wide range of medical fields, including hepatic cancer [134], diabetic complications [92,135], breast cancer screening [136] and liver cirrhosis [137]. The early study [92] on diabetic retinopathy using the multi-state Markov approach was not able to assess the clinical effects of relevant risk factors on diabetic retinopathy state transitions, possibly owing to insufficient computational power back in the mid-1990s. In this study we have used the GoDARTS database to incorporate longitudinal measures of multiple risk factors and assess their role in the specific developmental stages of diabetic retinopathy.

3.2 Results

3.2.1 Characteristics of the study sample

Overall 49,959 retinal measurements were studied in 4,758 diabetes patients who were retinopathy free at diagnosis of diabetes. At the end of this study, 100 patients developed

severe non-proliferative/proliferative diabetic retinopathy. The raw data on clinical characteristics, observed state transitions, numbers of retinal events over duration of diabetes, and the prevalence of diabetic retinopathy with and without stratification by first-year DBP, SBP and HbA1c is shown in Table 3-1, Table 3-2, Table 3-3, Figure 3-1 and Figure 3-2.

Table 3-1 Death, diabetes treatment, cardiovascular (a constellation of coronary artery disease, ischaemic stroke and lower extremity arterial disease) and chronic kidney disease (estimated glomerular filtration rate < 60) profiles for diabetic retinopathy patients included in the longitudinal study.

Diabetic retinopathy patients n (%)	4758 (100%)
Death n (%)	711 (14.9%)
Diabetes treatment (n, %)	Diet only (1285, 27.0 %), Insulin (428, 9.0 %), Oral agents (2444, 51.4 %), Oral agents and Insulin (148, 3.1 %)
Cardiovascular disease n (%)	1427 (30.0 %)
Chronic Kidney Disease n (%)	1171 (24.6 %)

Table 3-2 Frequency counts of the transition across states in successive clinical visits and the longitudinal diabetic retinopathy events (in parentheses). State 1: no-retinopathy; State 2: mild background retinopathy; State 3: observable background retinopathy; State 4: severe non-proliferative/proliferative retinopathy. Interval censored retinal states 21, 31 and 32 were defined to represent multiple observed states on the same record date, which were either state 1 or 2, either of state 1 and 3, and one of the set of state 2 and 3, respectively. A right censor was considered to be state-unknown (either state 1, 2, 3 or 4).

From	To State 1	State 2	State 3	State 4	State 2 or 1	State 3 or 1	State 3 or 2	Right censor
	(39282)	(4478)	(1011)	(100)	(561)	(64)	(143)	(4320)
State 1	32198	2273	549	34	338	13	63	3619
State 2	1640	1777	190	30	91	3	21	611
State 3	442	187	243	30	3	19	32	49
State 2 or 1	198	170	2	4	126	2	1	38
State 3 or 1	18	6	3	0	3	26	8	0
State 3 or 2	28	65	24	2	0	1	18	3

Table 3-3 The numbers of individuals occupying each observed state (non-censored state) at one year interval in the GODARTS panel data of retinopathy. A prior retinopathy event recorded the latest in relation to the time point specified was used in the estimation, even if more than one retinal event was recorded per year. The last observed retinal event for each patient were summed at the closest, later time point, and with a non-absorbing retinal event, the last observed retinal event was not accumulated into the estimation at a later time interval. The diagnosis of severe non-proliferative/proliferative retinopathy, an absorbing retinal event, was accumulated into the later estimations till 16 years.

Year	No-retinopathy	Mild background retinopathy	Observable background retinopathy	Severe non-proliferative/proliferative retinopathy	Total
1	4662	44	50	2	4758
2	4485	175	81	10	4751
3	4256	260	89	15	4620
4	3984	318	89	21	4412
5	3630	348	88	24	4090
6	3202	356	91	30	3679
7	2796	329	84	39	3248
8	2344	344	67	50	2805
9	1906	349	58	58	2371
10	1526	335	58	64	1983
11	1215	303	39	81	1638
12	894	261	28	89	1272
13	675	232	20	94	1021
14	480	198	14	96	788
15	323	148	13	99	583
16	196	117	6	100	419

Figure 3-1 Prevalence of diabetic retinopathy in the GODARTS panel data by duration of diabetes. This shows the retinopathy state as a percentage of the sample, recorded at each year of duration of diabetes from 1 to 16 years of diabetes duration.

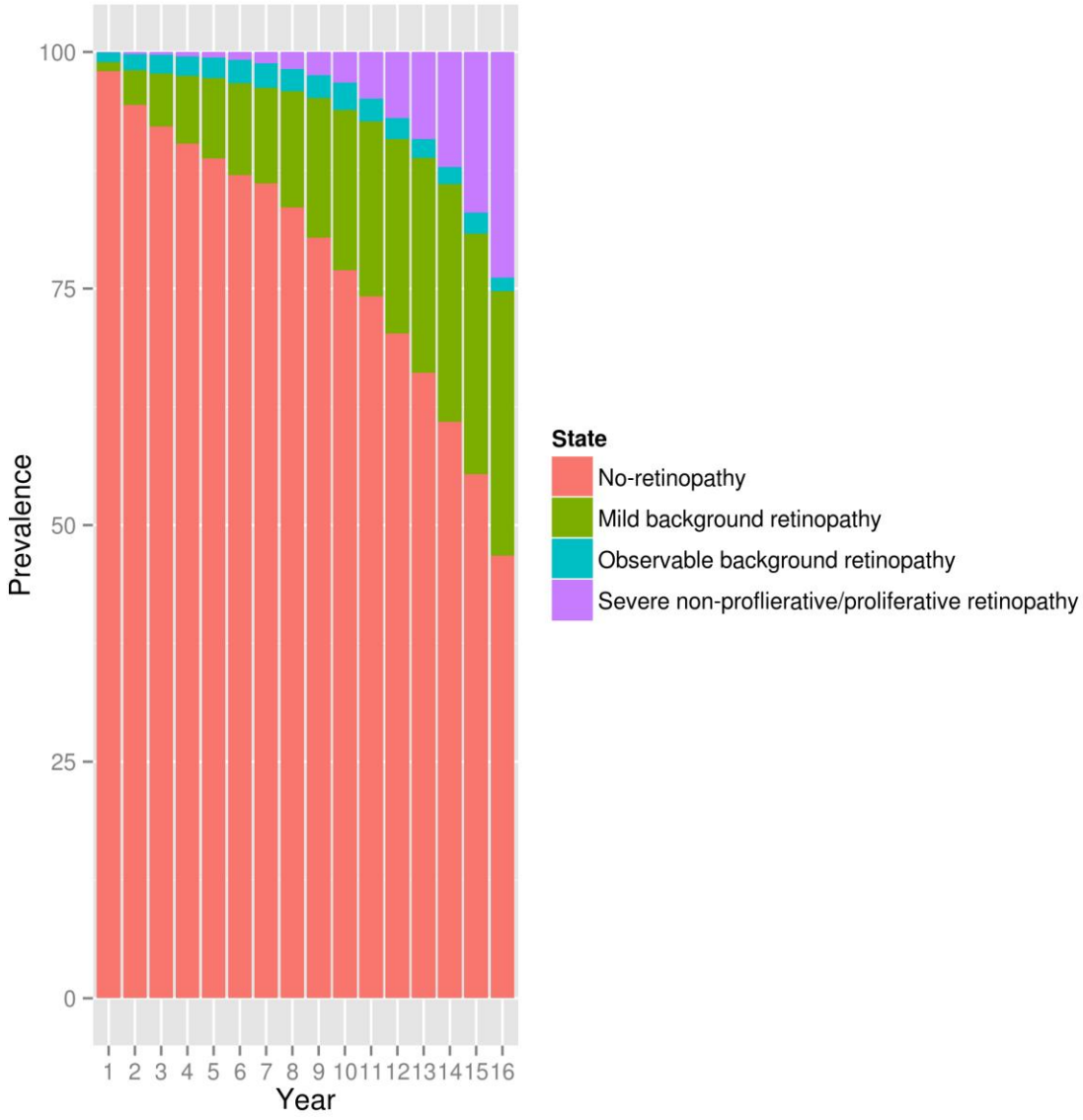
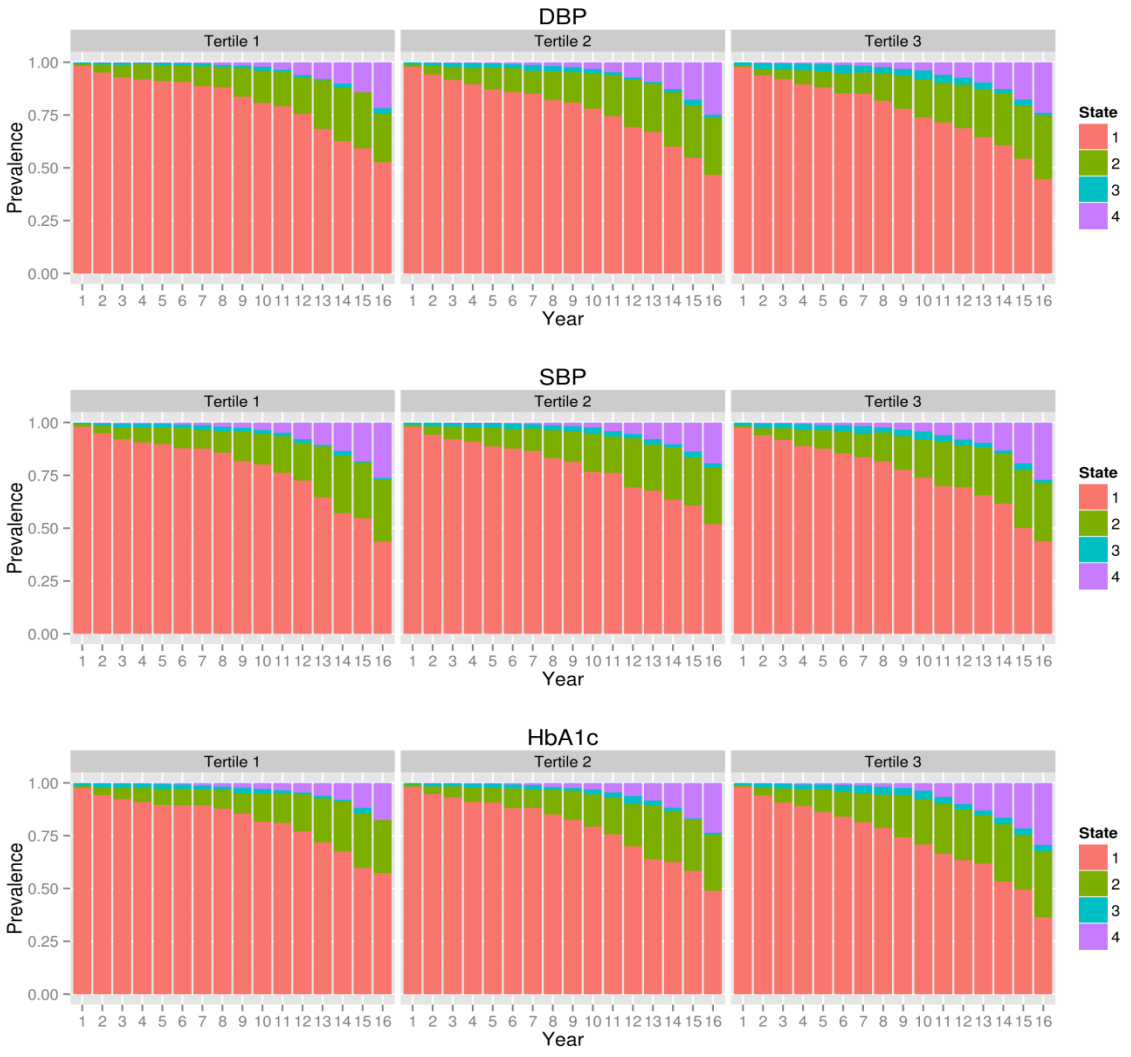


Figure 3-2 Prevalence of diabetic retinopathy states in the GoDARTS panel data, stratified by first-year DBP, SBP and HbA1c respectively. Sample sizes for DBP: tertile 1 (n = 1573), tertile 2 (n = 1690), and tertile 3 (n = 1495). Sample sizes for SBP: tertile 1 (n = 1609), tertile 2 (n = 1853), and tertile 3 (n = 1296). Sample sizes for HbA1c: tertile 1 (n = 1668), tertile 2 (n = 1514), and tertile 3 (n = 1576). DBP tertile cutoffs: 77.0 mmHg and 86.0 mmHg; SBP tertile cutoffs: 132.0 mmHg and 150.0 mmHg; HbA1c tertile cuffoffs: 45.0 mmol/mol (6.3%) and 57.0 mmol/mol (7.4 %). State 1: no-retinopathy; State 2: mild background retinopathy; State 3: observable background retinopathy; State 4: severe non-proliferative/proliferative retinopathy.



3.2.2 Baseline model without risk factor adjustment

Initial unadjusted modeling demonstrated the better fit of the two way transition model (one-way transition model AIC = 32042.2; two-way transition model AIC = 31574.2, $p < 0.0001$). A comparison between the observed and model-predicted prevalence indicated a close fit of the model to the DR data and so supported the internal validity of the model (Figure 3-3). This model indicated that the rates of remission from mild background diabetic retinopathy to diabetic retinopathy free state and from observable background diabetic retinopathy to mild background diabetic retinopathy were significantly faster than

the rates of progression (2.0 times faster and 4.2 times faster respectively), with the remission from observable to mild being almost double the rate of that observed for mild to no retinopathy (Table 3-4). The expected total length of time for diabetic retinopathy free, mild background retinopathy, observable background and severe non-proliferative/proliferative retinopathy states were 12.6 (95% confidence interval or CI: 12.41, 12.83) years, 2.91 (95% CI: 2.70, 3.11) years, 0.37 (95% CI: 0.27, 0.48) years and 0.11 (95% CI: 0.06, 0.19) years, respectively. For the maximum follow-up time (16 years), the estimated transition matrix showed 26%, 4.3% and 2% probabilities that a patient free of diabetic retinopathy will progress to mild background, observable and severe non-proliferative/proliferative retinopathy, respectively (Table 3-5). In the two-way transition model, the probabilities for correctly classifying retinopathy-free, mild background, observable, pre-proliferative/proliferative states were 0.971 (95% CI: 0.941, 0.986), 0.656 (95% CI: 0.503, 0.781), 0.613 (95% CI: 6.16×10^{-2}, 0.974) and 0.438 (1.53×10^{-4}, 1.000), respectively (Table 3-6). To investigate whether the maximum likelihood estimates of the two-way transition model converged to reliable values, we ran the base model using multiple sets of initial values randomly generated from the standard uniform distribution. For eight out of nine models used random initial values, the point estimates of transition intensities and classification probabilities were close to the estimates derived from the default crude initial values generated by the “msm” package (Table 3-7 and Table 3-8).

Table 3-4 Estimated ratios of transition intensities, with standard errors (SE) and lower (L) and upper (U) bound of 95% CI estimated from the delta method, in the two-way transition base model. State 1: no-retinopathy; State 2: mild background retinopathy; State 3: observable background retinopathy; State 4: severe non-proliferative/proliferative retinopathy.

Transitions intensities	Estimates	SE	L	U
State 2 → State 1 vs. State 1 → State 2	2.04	0.17	1.74	2.40
State 3 → State 2 vs. State 2 → State 3	4.20	0.70	3.04	5.82
State 1 → State 2 vs. State 2 → State 3	1.05	0.15	0.80	1.38
State 2 → State 3 vs. State 3 → State 4	0.84	0.35	0.37	1.92
State 1 → State 2 vs. State 3 → State 4	0.88	0.34	0.42	1.86
State 2 → State 1 vs. State 3 → State 2	0.51	0.09	0.36	0.73

Table 3-5 The estimated transition probability matrix for 16 years' time interval and 95% CI (in parentheses) for the two-way transition base model. State 1: no-retinopathy; State 2: mild background retinopathy; State 3: observable background retinopathy; State 4: severe non-proliferative/proliferative retinopathy.

	To			
From	State 1	State 2	State 3	State 4
State 1	0.677 (0.654,0.700)	0.260 (0.237,0.281)	0.043 (0.031,0.057)	0.020 (0.011,0.035)
State 2	0.531 (0.481,0.581)	0.329 (0.286,0.370)	0.071 (0.047,0.099)	0.069 (0.037,0.116)
State 3	0.368 (0.288,0.442)	0.296 (0.224,0.356)	0.079 (0.041,0.127)	0.257 (0.139,0.424)
State 4	0 (0,0)	0 (0,0)	0 (0,0)	1 (1,1)

Table 3-6 The maximum likelihood estimates of the probability classification matrix and 95% CI (in parentheses) for the two-way transition base model. State 1: no-retinopathy; State 2: mild background retinopathy; State 3: observable background retinopathy; State 4: severe non-proliferative/proliferative retinopathy.

	Observed State			
True State	State 1	State 2	State 3	State 4
State 1	0.971 (0.941,0.986)	0.019 (0.017,0.022)	9.62×10^{-3} $(8.56\times10^{-3},0.011)$	4.4×10^{-4} $(2.14\times10^{-4},9.07\times10^{-4})$
State 2	0.311 (0.267,0.360)	0.656 (0.503,0.781)	0.026 (0.018,0.037)	7.51×10^{-3} $(4.66\times10^{-3},0.023)$
State 3	0.145 (0.100,0.205)	0.232 (0.136,0.369)	0.613 (0.063,0.974)	0.010 $(4.28\times10^{-4},0.196)$
State 4	3.12×10^{-3} $(3.39\times10^{-6},0.743)$	9.26×10^{-3} $(6.44\times10^{-6},0.576)$	0.550 (0.032,0.978)	0.438 $(1.53\times10^{-4},1.000)$

Table 3-7 Point estimates of transition intensity modelled in the diabetic retinopathy two-way transition base model using initial parameter estimates calculated from the "crudeinits.msm" function and random deviates generated from standard uniform distribution. State 1: no-retinopathy; State 2: mild background retinopathy; State 3: observable background retinopathy; State 4: severe non-proliferative/proliferative retinopathy.

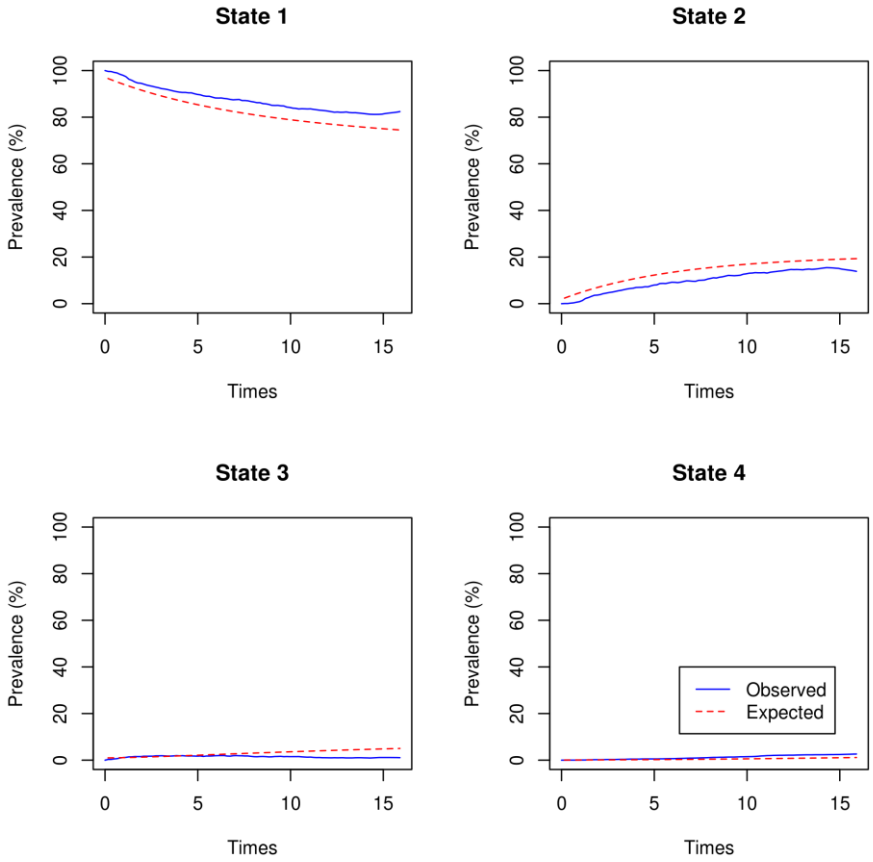
	Transition intensities estimates					Log-likelihood
	S1=>S2	S2=>S1	S2=>S3	S3=>S2	S3=>S4	
Crude initial	0.048	0.099	0.046	0.194	0.055	-15770.1
Random initial 1	0.048	0.100	0.047	0.173	0.032	-15773.5
Random initial 2	0.048	0.099	0.046	0.193	0.055	-15770.1
Random initial 3	0.048	0.099	0.046	0.176	0.035	-15773.9
Random initial 4	0.048	0.099	0.046	0.175	0.027	-15774.2
Random initial 5	0.048	0.099	0.046	0.194	0.053	-15770.3
Random initial 6	0.053	0.113	0.056	0.137	0.102	-15731.5
Random initial 7	0.048	0.099	0.046	0.192	0.053	-15770.2
Random initial 8	0.048	0.099	0.046	0.193	0.054	-15770.2
Random initial 9	0.048	0.099	0.046	0.193	0.053	-15770.2

Table 3-8 Point estimates of classification probability modelled in the diabetic retinopathy two-way transition base model using initial parameter estimates calculated from the "crudeinits.msm" function and random deviates generated from standard uniform distribution. State 1: no-retinopathy; State 2: mild background retinopathy; State 3: observable background retinopathy; State 4: severe non-proliferative/proliferative retinopathy.

	Classification probability estimates							
	S1=>S1	S1=>S2	S1=>S3	S1=>S4	S2=>S1	S2=>S2	S2=>S3	S2=>S4
Crude initial	0.971	0.019	0.010	0.000	0.311	0.656	0.025	0.008
Random initial 1	0.971	0.019	0.010	0.001	0.313	0.657	0.026	0.005
Random initial 2	0.971	0.019	0.010	0.000	0.312	0.656	0.025	0.007
Random initial 3	0.971	0.019	0.010	0.000	0.311	0.657	0.026	0.006
Random initial 4	0.971	0.019	0.010	0.000	0.310	0.658	0.027	0.006
Random initial 5	0.971	0.019	0.010	0.000	0.312	0.656	0.025	0.007
Random initial 6	0.974	0.016	0.010	0.000	0.365	0.613	0.018	0.004
Random initial 7	0.971	0.019	0.010	0.000	0.311	0.655	0.026	0.008
Random initial 8	0.971	0.019	0.010	0.000	0.311	0.656	0.025	0.007
Random initial 9	0.971	0.019	0.010	0.000	0.311	0.656	0.025	0.008
	S3=>S1	S3=>S2	S3=>S3	S3=>S4	S4=>S1	S4=>S2	S4=>S3	S4=>S4

	Classification probability estimates							
Crude initial	0.145	0.232	0.613	0.010	0.003	0.009	0.550	0.438
Random initial 1	0.132	0.207	0.621	0.039	0.008	0.360	0.001	0.630
Random initial 2	0.145	0.233	0.612	0.010	0.003	0.010	0.552	0.435
Random initial 3	0.133	0.208	0.626	0.033	0.011	0.109	0.109	0.772
Random initial 4	0.133	0.204	0.626	0.037	0.012	0.016	0.028	0.944
Random initial 5	0.144	0.233	0.611	0.012	0.007	0.014	0.555	0.424
Random initial 6	0.163	0.188	0.597	0.052	0.000	0.861	0.111	0.028
Random initial 7	0.145	0.231	0.613	0.012	0.007	0.007	0.553	0.433
Random initial 8	0.145	0.232	0.612	0.011	0.003	0.011	0.557	0.428
Random initial 9	0.145	0.233	0.611	0.012	0.008	0.004	0.565	0.423

Figure 3-3 Expected prevalence modelled by the two-way transition base model, compared to the observed prevalence. State 1: no-retinopathy; State 2: mild background retinopathy; State 3: observable background retinopathy; State 4: severe non-proliferative/proliferative retinopathy.



3.2.3 Assessment of tradition risk factors

We standardised values of BMI, cholesterol, creatinine, DBP, HbA1c, HDL-c, SBP, triglycerides and non-HDL-c (Table 3-9). As we have the full longitudinal medical record of each patient we adjusted each specific retinal event using risk factor data that was measured as close to that event as available. We found that BMI, DBP, HbA1c, SBP provided generally very close measures for each retinal assessment, probably due to their measure by diabetes specialists (Table 3-9). However, measures of vascular risk such as cholesterol, creatinine, HDL-c, triglycerides and non-HDL-c were measured more distally to the retinal screening events (Table 3-9).

In a univariate analyses there was a significant effect on progression rates for age of diagnosis, age, cholesterol, DBP, HbA1c, SBP, triglycerides, and non-HDL-c, even after Bonferroni correction (threshold 0.0038) (Table 3-10 and Table 3-11). In contrast, there was no significant effect of BMI, serum creatinine, HDL-c, sex and smoking status (Table 3-10 and Table 3-12). We then examined the effects of the risk factors on the individual transitions between disease states. An increase in HbA1c level by one standard deviation (SD) (15.83 mmol/mol, 1.4%) had a 42% increased risk of progression from the no retinopathy state to mild background retinopathy, a 32% increased risk of progression from mild background retinopathy to observable background retinopathy, and a 123% increased risk in progression from observable background retinopathy to severe non-proliferative/proliferative retinopathy (Table 3-11). Conversely, a reduction in the HbA1c level by one SD was associated with a 24% increased possibility of recovering from mild background retinopathy to the retinopathy free state (Table 3-11), but the HbA1c level was

unrelated to the regression from observable background retinopathy to mild background retinopathy in this cohort. A raised level of DBP by one SD (10.41 mmHg) elevated the risk for developing observable background retinopathy from the mild background retinopathy by 87% (Table 3-11). SBP was also significant risk factor for progression to mild background retinopathy from the initial retinopathy-free state (Table 3-11), and the reduction in SBP by one SD (17.28 mmHg) was associated a 20% increased chance of regression back to the retinopathy-free state (Table 3-11). The risk effect of cholesterol and non-HDL-c on the progression from the mild background retinopathy to observable background retinopathy reached statistical significance in the univariate models (Table 3-11), but was insignificant after adjustment in the multivariate model (Table 3-11). In the multivariate analysis, at the 5% significance level, triglyceride values influenced the transition from the retinopathy-free state to mild background retinopathy (Table 3-11), which however was statistically insignificant in the univariate assessment (Table 3-11).

Table 3-9 Mean and standard deviation (SD) of longitudinally measured covariates, and time difference between a covariate measurement date and the closest date of retinal events.

* The lower quartile, median and upper quartile of the time difference (in days). A negative value in time difference indicates a covariate measured prior to a retinal event.

Longitudinal covariate	Mean	SD	Record time*
BMI (kg/m^2)	30.77	5.43	0,0,0
Cholesterol (mmol/l)	4.57	1.05	-50,0,31
Serum creatinine (μmol/l)	89.81	24.31	-57,0,39
DBP (mmHg)	77.01	10.41	0,0,0
HbA1c % (mmol/mol)	7.2% (54.94)	1.4% (15.83)	0,0,0
HDL-c (mmol/l)	1.27	0.35	-60,0,29
SBP (mmHg)	139.60	17.28	0,0,0
Triglycerides (mmol/l)	2.14	1.08	-182,0,90
Non-HDL-c (mmol/l)	3.26	1.04	-78,0,29

Table 3-10 Likelihood ratio tests of single-covariable model against the two-way transition model (base model AIC: 31574.2).

Covariate	LR Statistic D	DF	*P*	AIC
Diabetes diagnosis age	61.786	5	5.19×10^{-12}	31522.4
Age	60.317	5	1.05×10^{-11}	31523.9
Sex	3.451	5	6.31×10^{-1}	31580.8
Smoking	7.333	5	1.97×10^{-1}	31576.9
BMI	3.119	5	6.82×10^{-1}	31581.1
Cholesterol	71.054	5	6.18×10^{-14}	31513.1
Serum creatinine	6.981	5	2.22×10^{-1}	31577.2
DBP	84.247	5	1.11×10^{-16}	31500.0
HbA1c	196.379	5	0.00	31387.8
HDL-c	14.135	5	1.48×10^{-2}	31570.1
SBP	51.344	5	7.35×10^{-10}	31532.9
Triglycerides	25.141	5	1.31×10^{-4}	31559.1
Non-HDL-c	80.744	5	5.55×10^{-16}	31503.5

Table 3-11 The estimates and 95% CI (in parentheses) of hazard ratios (HR) for diabetic retinopathy progression (state 1-2, 2-3, 3-4) and regression (state 2-1, 3-2). State 1: no-retinopathy; State 2: mild background retinopathy; State 3: observable background retinopathy; State 4: severe non-proliferative/proliferative retinopathy. Covariates were z-transformed in relation to the mean and standard deviation presented in Table 3-9, and all hazard ratios refer to per standard deviation of covariate. Risk factors statistically significant in the single-covariate test after the Bonferroni adjustment (the statistical significance level: 0.0038) are shown.

	Cholesterol	DBP	HbA1c	SBP	Triglycerides	Non-HDL-c
	Single-covariate analysis (HR,95%CI)					
State 1 - State 2	0.99 (0.92,1.06)	1.09 (1.01,1.17)	1.41 (1.32,1.51)	1.17 (1.09,1.26)	0.93 (0.87,1.00)	0.98 (0.92,1.05)
State 2 - State 1	0.95 (0.78,1.15)	0.85 (0.70,1.03)	0.75 (0.64,0.88)	0.76 (0.63,0.92)	0.85 (0.70,1.03)	0.93 (0.77,1.12)
State 2 - State 3	1.99 (1.63,2.44)	2.20 (1.77,2.74)	1.60 (1.32,1.94)	1.35 (1.07,1.70)	1.30 (1.11,1.52)	2.02 (1.67,2.45)
State 3 - State 2	0.95 (0.69,1.31)	0.91 (0.63,1.31)	0.86 (0.64,1.14)	1.00 (0.71,1.40)	0.84 (0.59,1.20)	0.98 (0.71,1.35)
State 3 - State 4	1.33 (0.87,2.03)	1.12 (0.64,1.96)	2.46 (1.57,3.84)	0.76 (0.43,1.33)	1.48 (1.06,2.07)	1.28 (0.94,1.74)

	Cholesterol	DBP	HbA1c	SBP	Triglycerides	Non-HDL-c
	Multi-covariate analysis (HR,95%CI)					
State 1 - State 2	0.99 (0.82,1.19)	0.98 (0.90,1.07)	1.42 (1.32,1.52)	1.20 (1.11,1.30)	0.92 (0.85,0.99)	0.98 (0.81,1.19)
State 2 - State 1	1.11 (0.70,1.78)	1.01 (0.82,1.25)	0.76 (0.64,0.89)	0.79 (0.64,0.97)	0.95 (0.76,1.18)	0.91 (0.55,1.50)
State 2 - State 3	0.86 (0.40,1.86)	1.87 (1.46,2.40)	1.32 (1.08,1.60)	0.96 (0.76,1.22)	0.99 (0.78,1.24)	1.94 (0.84,4.45)
State 3 - State 2	0.84 (0.44,1.58)	0.86 (0.58,1.29)	0.88 (0.66,1.16)	1.14 (0.78,1.67)	0.91 (0.62,1.33)	1.19 (0.58,2.45)
State 3 - State 4	1.69 (0.62,4.55)	0.92 (0.53,1.58)	2.23 (1.16,4.29)	0.87 (0.52,1.46)	1.27 (0.89,1.81)	0.72 (0.26,1.99)

Table 3-12 The estimates and 95% CI (in parentheses) of hazard ratios (HR) for statistically insignificant (Bonferroni threshold: 0.0038) covariates modelled in the two-way transition process in a single-covariate analysis. Quantitative covariates were z-transformed in relation to the mean and standard deviation presented in Table 3-9, and all hazard ratios refer to per standard deviation of covariate. State 1: no-retinopathy; State 2: mild background retinopathy; State 3: observable background retinopathy; State 4: severe non-proliferative/proliferative retinopathy.

	Sex (Male)	Smoking	BMI	Creatinine	HDL-c
State 1 - State 2	1.05 (0.91,1.20)	0.91 (0.78,1.07)	0.97 (0.91,1.04)	1.05 (0.97,1.13)	1.03 (0.96,1.10)
State 2 - State 1	0.95 (0.68,1.32)	1.22 (0.81,1.84)	0.89 (0.74,1.07)	1.10 (0.92,1.30)	1.07 (0.91,1.25)
State 2 - State 3	0.76 (0.50,1.14)	1.53 (0.93,2.52)	0.91 (0.75,1.11)	1.16 (1.01,1.33)	0.69 (0.54,0.89)
State 3 - State 2	0.93 (0.51,1.67)	1.14 (0.58,2.25)	0.88 (0.65,1.18)	1.08 (0.83,1.40)	1.05 (0.76,1.43)
State 3 - State 4	1.51 (0.60,3.80)	1.34 (0.40,4.43)	1.14 (0.72,1.79)	0.72 (0.39,1.34)	1.15 (0.68,1.93)

3.3 Conclusions

Our analysis has demonstrated that in the development of diabetic retinopathy, the initial, asymptomatic phase was stable, followed by transient mid-stages, and that substantial rates of disease regression could be observed. The risk of diabetic retinopathy progression from the retinopathy-free state to mild background retinopathy, from mild background retinopathy to observable background retinopathy and from observable background retinopathy to severe non-proliferative/proliferative retinopathy was strongly positively associated with glycaemic exposure. We also found a significant risk effect of DBP on the progression of mild background retinopathy to observable background retinopathy and of SBP on the state transition from the retinopathy-free state to mild background retinopathy. There was no evidence in this study that the risk effects for diabetic retinopathy state transitions were influenced by sex, smoking status, BMI, serum creatinine or HDL-c. We did not find the evidence for independent risk effects of cholesterol, triglycerides and non-HDL-c on diabetic retinopathy state transitions with the adjustment for blood pressure and glycaemic control. This study has provided the first evidence to show that better HbA1c and SBP are strongly correlated with the regression from mild background retinopathy back to the retinopathy-free state.

One of the strengths of this study is the 15-fold greater overall sample size compared to an earlier study on diabetic retinopathy using an identical approach and a substantially extended follow-up time. A potential limitation in this longitudinal study of historical events remains the paucity of follow-up data on the study subjects that were recruited more

recently. Also, half way through the follow up screening was switched from Polaroid films to digital images, although a similar grading category was followed.

In this study, our data yielded an important novel estimation about the time spent in each state in this cohort. To date, most longitudinal studies on diabetic retinopathy development have been directed at estimating incidence and/or progression rate in a study sample, and few have examined the average length of time spent in each stage of diabetic retinopathy. In the late 1980s, the Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR) reported 0.4% of patients with diabetes diagnosed at 30 years of age or above and without retinopathy at the first retinal examination progressed to proliferative diabetic retinopathy within four years [138]. The United Kingdom Prospective Diabetes Study (UKPDS) identified 0.2% of 2,316 type 2 diabetic patients with no retinopathy at baseline required photocoagulation treatments at 3 years, 1.1% at 6 years and 2.6% at 9 years [139,140]. A recent study on 16,444 patients with type 2 diabetes without retinopathy at the first retinal examination found the cumulative incidence of non-proliferative retinopathy, severe non-proliferative retinopathy and proliferative retinopathy was 36%, 4% and 0.68%, respectively after 5 years follow up, and after 10 years follow up, these estimates rose to 66%, 16% and 1.5% respectively [141]. These findings broadly support the estimated total length of time in the retinopathy-free state reported in this study. .

Extensive evidence from published randomised clinical trials [21,142,143], prospective [144–150] and retrospective [151] studies support our findings on HbA1c as an important risk factor on diabetic retinopathy progression (Table 3-11). The Diabetes Control and Complications Trial (DCCT) [21] reported a hazard rate of 1.63 ($P<0.001$) for the risk

effect of one SD of HbA1C in type 1 diabetic patients. UKPDS [143] has found per one SD increase in the HbA1c variable a hazard rate of 1.48 (95% CI: 1.40, 1.61) and 1.96 (95% CI: 1.79, 2.16) respectively for microvascular complications in patients with type 2 diabetes. These results broadly support the hazard ratios we found in this study for HbA1c on diabetic retinopathy. Additionally, we demonstrate that lower HbA1c is associated with regression of retinopathy from mild background retinopathy to no retinopathy. However, once a more severe retinopathy state e.g. observable BDR is reached, the protective effect associated with lowering HbA1c is not observed, suggesting that good glycaemic control only facilitates retinopathy remission at an early stage.

Previous studies have shown mixed results on the association between blood pressure and diabetic retinopathy. The UKPDS [140] demonstrated that the incidence of retinopathy was associated with SBP values in top vs. bottom tertiles and lowering blood pressure resulted in a marked reduction in development or progression of diabetic retinopathy. In one of the WESDR reports [145], in which a prospective cohort of type 1 diabetic patients was followed up for 14 years, the baseline DBP variable was a significant predictor of progression to PDR. A study in the late-1980s [152] showed no association of SBP and DBP variables in the highest and lowest quartiles with the incidence or the progression of retinopathy in type 2 diabetic patients. In contrast, it was shown in the same study that in type 1 diabetic patients, SBP and DBP were correlated with the progression of retinopathy. Our study firmly supports a role for blood pressure in diabetic retinopathy progression in individuals with type 2 diabetes.

In this study, we have applied an innovative approach for the analysis of population-based longitudinal retinopathy cohort data. Our findings delineated state-by-state transitions underlying diabetic retinopathy development, and our assessment of population risk factors influencing progressive and regressive state transitions yielded the evidence for the role of blood pressure and glycaemic control in diabetic retinopathy development. Furthermore, the analytical approach utilised in this study holds the potential to be extended for investigating the additional independent effect from anti-diabetic oral agents on the course of diabetic retinopathy, or the interaction between anti-diabetic medications with HbA1c on the development diabetic retinopathy. These lines of interest on the front of pharmacoepidemiology may deserve a separate, thorough investigation, with additional input from population prescribing data sets. However, we have the confidence that the strategy we applied here will become the cornerstone for increasingly more clinical studies.

Chapter 4

Heritability and genetic correlations explained by common SNPs for diabetic retinopathy and related risk factors

4.1 Introduction

Diabetic retinopathy and the related metabolic traits are highly heritable [69,70,143,153–156]. Uncovering the genetic influences underlying these traits and the genetic correlation between these related traits is the area of clinical interest. Recently, high-throughput array scans including genome-wide association studies (GWAS) identified putative genetic signals for diabetic retinopathy . However, significant genetic variants discovered in GWAS for a number of clinical traits explain only a limited fraction of the phenotypic

variance, substantially lower than the expected narrow-sense heritability, a genetic measure for the ratio of additive genetic variance over phenotypic variance. Yang *et al* [104] discovered that the missing proportion of heritability for GWAS may be linked to incomplete linkage disequilibrium (LD) between causal variants and genetic markers, and small effects associated with genetic variants against a stringent genome-wide significance *P* value threshold. Simultaneously, common GWAS SNPs explains a considerable portion of "missing" heritability [104,157].

The restricted maximum likelihood (REML) analytical approach for decomposing phenotypic variation has been developed to study narrow-sense heritability for quantitative or dichotomous traits only based on common genome-wide SNP data, and additionally genetic and residual correlations between two quantitative traits or two binary traits [104,108,157]. However, clinically relevant phenotypes are frequently measured as an ordinal trait, as with the diabetic retinopathy phenotype. In studying narrow-sense heritability and genetic/residual correlations, a feasible analytical strategy is also needed for ordinal traits. Here, we explored the Bayesian Markov chain Monte Carlo approach, and uncovered the heritable proportion of the phenotypic variance for diabetic retinopathy and related risk factors in the GoDARTS cohort, and for the first-time, estimated the pairwise genetic and residual correlations for diabetic retinopathy and related risk factors.

4.2 Results

4.2.1 Characteristics of the study data

In this study, our base population consists of the GoDARTS sample genotyped on the Affymetrix 6 chip, which had random selections of retinopathy patients from the diabetic population. This study consisted of separate samples for genetically related individuals with identical by descent (IBD) estimates more than 0.05, and genetically unrelated individuals with IBD estimates less than 0.05. The clinical characteristics of the study population are provided in Table 4-1, and these characteristics were similar between genetically related and unrelated samples. Phenotypic correlations of body mass index (BMI), cholesterol (CHL), serum creatinine (CRE), glycosylated haemoglobin/HbA1c (HBA), high-density lipoprotein cholesterol/HDL-c (HDL), systolic blood pressure (SBP), triglycerides (TG) and retinopathy (RET) estimated in both samples are provided in Table 4-2 and Table 4-3. Here, we report consistent correlations observed in genetically related and unrelated samples of diabetes. We found positive phenotypic correlations of BMI-HBA, CHL-HBA, HBA-TG, BMI-TG, CHL-TG, CHL-RET, CRE-RET, HBA-RET and negative phenotypic correlations of BMI-HDL, HBA-HDL and HDL-TG (Table 4-2 and Table 4-3). The point estimates of phenotypic correlation coefficients of CRE-HDL were almost the same (Table 4-2), but with the adjustment for multiple testing, this correlation in the genetically related sample was insignificant with $P>0.05$, owing to the smaller sample size.

Table 4-1 Characteristics of the related sample (the IBD estimate between at least one pair of individuals > 0.05) and the unrelated sample (the IBD estimate between any pair of individuals ≤ 0.05). Retinopathy category 0: no-retinopathy; category 1: mild background retinopathy; category 2: observable background retinopathy; category 3: severe non-proliferative/proliferative retinopathy.

(a) Quantitative traits.

	Related sample			Unrelated sample		
Trait	Mean	SD	n	Mean	SD	n
Age (years)	64.2	10.8	1255	65.5	10.7	3039
Diabetes duration (years)	7.0	6.6	1255	7.8	7.1	3039
BMI (kg/m^2)	31.0	5.6	1193	30.8	5.6	2970
Cholesterol (mmol/l)	4.7	1.1	1193	4.6	1.1	2970
Serum creatinine (μmol/l)	95.0	25.7	1193	96.5	25.4	2970
HbA1c % (mmol/mol)	7.5 (58.9)	1.4% (15.4)	1193	7.5% (58.3)	1.4% (15.2)	2970
HDL-c (mmol/l)	1.3	0.4	1193	1.4	0.4	2970
SBP (mmHg)	141.6	18.5	1193	141.6	18.3	2970
Triglycerides (mmol/l)	2.4	1.4	1193	2.2	1.3	2970

(b) Dichotomous traits.

	Related sample		Unrelated sample	
Trait	Counts (n, %)		Counts (n, %)	
Sex	F (601, 48%)	M (654, 52%)	F (1401, 46%)	M (1638, 54%)

(c) Polychotomous traits. Retinopathy categories: 0 (without retinopathy), 1 (mild background retinopathy), 2 (observable background retinopathy), 3 (severe non-proliferative/proliferative retinopathy).

	Related sample				Unrelated sample			
Trait	Counts (n, %)				Counts (n, %)			
Retinopathy	0 (372, 30%)	1 (442, 36%)	2 (267, 22%)	3 (158, 13%)	0 (806, 27%)	1 (1091, 37%)	2 (657, 22%)	3 (432, 15%)

Table 4-2 Pearson correlation coefficients and 95% confidence intervals (separate lines below) for quantitative risk factors of diabetic retinopathy in the GoDARTS related (n = 1193) and unrelated (n = 2970) sample. BMI: body mass index; CHL: cholesterol; CRE: serum creatinine; HBA (HbA1c): glycosylated haemoglobin; HDL (HDL-c): high-density lipoprotein cholesterol; SBP: systolic blood pressure; TG: triglycerides. *Significance at the 0.05 level adjusted for 21 hypotheses using Fisher-transformed correlation coefficients.

	Related sample						Unrelated sample					
	BMI	CHL	CRE	HBA	HDL	SBP	BMI	CHL	CRE	HBA	HDL	SBP
CHL	0.019						0.007					
	-0.037						-0.029					
	0.076						0.043					
CRE	-0.007	-0.046					-0.071*	-0.108*				
	-0.064	-0.103					-0.106	-0.143				
	0.050	0.010					-0.035	-0.072				
HBA	0.113*	0.099*	-0.027				0.110*	0.131*	-0.046			
	0.057	0.042	-0.084				0.074	0.096	-0.082			
	0.169	0.155	0.029				0.145	0.166	-0.010			
HDL	-0.116*	0.093*	-0.070	-0.112*			-0.181*	0.153*	-0.071*	-0.067*		
	-0.172	0.037	-0.127	-0.167			-0.216	0.117	-0.107	-0.102		

	Related sample						Unrelated sample					
	BMI	CHL	CRE	HBA	HDL	SBP	BMI	CHL	CRE	HBA	HDL	SBP
	-0.060	0.149	-0.014	-0.055			-0.146	0.188	-0.036	-0.031		
SBP	0.054	0.001	-0.011	-0.017	0.097*		0.025	0.100*	-0.010	0.021	0.096*	
	-0.003	-0.056	-0.067	-0.074	0.040		-0.011	0.065	-0.046	-0.015	0.060	
	0.110	0.058	0.046	0.040	0.153		0.060	0.136	0.026	0.057	0.132	
TG	0.113*	0.391*	0.004	0.214*	-0.295*	0.012	0.196*	0.292*	-0.009	0.185*	-0.372*	0.052
	0.057	0.342	-0.052	0.159	-0.346	-0.045	0.161	0.259	-0.045	0.150	-0.403	0.016
	0.169	0.438	0.061	0.267	-0.242	0.069	0.230	0.325	0.027	0.219	-0.341	0.088

Table 4-3 Phenotypic correlation coefficients and 95% confidence intervals (separate lines below) between diabetic retinopathy and related quantitative risk factors in the GoDARTS related and unrelated sample. CHL: cholesterol; CRE: serum creatinine; HBA: HbA1c; RET: retinopathy; TG: triglycerides. *Significance at the 0.05 level adjusted for 7 hypotheses using Fisher-transformed correlation coefficients.

RET	BMI	CHL	CRE	HBA	HDL	SBP	TG
Related sample	-0.080*	0.058*	0.239*	0.189*	-0.052	0.022	-0.013
n = 1178	-0.137	0.001	0.184	0.133	-0.109	-0.035	-0.070
	-0.023	0.115	0.292	0.243	0.005	0.079	0.045
Unrelated sample	-0.036	0.110*	0.123*	0.201*	0.031	0.060*	-0.060*
n = 2903	-0.072	0.074	0.087	0.166	-0.005	0.024	-0.096
	0.001	0.146	0.159	0.236	0.068	0.097	-0.023

4.2.2 Heritability explained by common genome-wide SNPs for diabetic retinopathy and related risk factors

We estimated heritability explained by common genome-wide SNPs with the adjustment for age, sex and duration of diabetes in the genetically related and unrelated samples using the Markov chain Monte Carlo (MCMC) sampling strategy. The heritability estimate is the total additive genetic variance normalised by total phenotypic variance. MCMC data is valid (mixed) or formally described as having reached stationarity if it approximates the unknown target distribution of the parameter of interest. In this study, MCMC data for each trait passed through the Geweke and Heidelberg-Welch stationarity tests. The trace plots provide a visual tool for inspecting parameter values sampled. Any systematic upward or downward trends indicate poor mixing of MCMC data. The traces of MCMC data in this study showed small fluctuations, suggesting of good mixing of MCMC chains (Appendix Figure A-1). In the density plots, genetic variance parameter estimates were positively skewed (Appendix Figure A-1), because values for variance are always non-negative. As a convention, the residual variance was pre-defined to be 1 for any categorical trait in the liability threshold model. Autocorrelations of MCMC data is an assessment for the efficiency of mixing. The autocorrelation plot suggests higher autocorrelations of MCMC samples in the genetically unrelated sample compared to the related sample (Appendix Figure A-1), owing to a larger sample size and more total additive genetic effect parameters sampled in the MCMC algorithm. We obtained 25,000 MCMC samples for the heritability estimation for each trait in the related and unrelated samples.

Long-range linkage disequilibrium blocks that span known SNP markers and unknown genomic variations were anticipated in the genetically related sample, and thus it was probable that estimates of heritability and genetic correlation may be higher in the genetically related sample, even with a smaller sample size compared to that in the unrelated sample. The heritability estimates were close in both samples for CHL, CRE, HBA and SBP (Table 4-4). BMI and TG showed higher heritability in the genetically related and unrelated samples, respectively (Table 4-4). 22% and 34% of phenotypic variations of diabetic retinopathy were heritable in the genetically related and unrelated samples, respectively (Table 4-4).

As aforementioned, densities of variance parameter estimates were positively skewed, and thus, the posterior median was used as the posterior estimator. The other alternative posterior estimator, the posterior mode, required the mathematical integration of joint posterior probabilities, which became an intractable mathematical problem with this model. Since the posterior distribution were non-symmetric, we used 95% highest posterior density (HPD) interval, which is the credible interval in parameter space that has 95% posterior probabilities, with the minimal density in the interval greater than or equal to the density of any point outside of the interval. In simple terms, the HPD interval can be interpreted as the range of parameter values that occur 95% of time with repeated sampling of probable parameter values for a given data, where the true parameter values are considered as dispersed following a distribution. In contrast, the frequentist confidence interval estimation involves no sampling, and it is interpreted as the range of parameter values that cover the true, fixed parameter value for 95% of time, if the study data is hypothetically repeatedly sampled from the population. *P* values are associated with the frequenstist hypothesis

testing, whereas in Bayesian inference, the Bayes factor is used in comparing competing models for the same data for the better-fitted model. The Bayes factor is analogous to the frequentist likelihood ratio, and it provides a means for assessing the strength of evidence. However, the Bayes factor is difficult to compute for mixed models we described here, and thus, it was not feasible to estimate *P* values adjusted for multiple testing in this study. A comprehensive review on MCMC Bayesian inference is provided in [107].

Table 4-4 Genetic (V_g) and residual (V_e) variance that have been used for estimating heritability (h_g^2) explained by common genome-wide SNPs in the GoDARTS sample. 1,193 (2,970) genetically related (unrelated) subjects' data were used in the heritability estimation of BMI, CHL, CRE, HBA, HDL, SBP and TG; 1,236 (2,985) genetically related (unrelated) subjects' data were used in the heritability estimation of RET. BMI: body mass index; CHL: cholesterol; CRE: serum creatinine; HBA (HbA1c): glycosylated haemoglobin; HDL (HDL-c): high-density lipoprotein cholesterol; RET: retinopathy; SBP: systolic blood pressure; TG: triglycerides. **Residual variance is a pre-specified parameter for categorical traits.

	Related sample			Unrelated sample		
Trait	V_g	V_e	h_g^2	V_g	V_e	h_g^2
BMI	0.47 (0.22, 0.72)	0.50 (0.27, 0.72)	**0.49 (0.24, 0.72)**	0.30 (0.12, 0.49)	0.65 (0.47, 0.82)	**0.31 (0.13, 0.51)**
CHL	0.17 (0.07, 0.32)	0.77 (0.62, 0.89)	**0.18 (0.07, 0.33)**	0.19 (0.08, 0.33)	0.75 (0.62, 0.87)	**0.20 (0.08, 0.34)**
CRE	0.18 (0.08, 0.33)	0.62 (0.49, 0.74)	**0.23 (0.09, 0.40)**	0.15 (0.07, 0.26)	0.65 (0.54, 0.74)	**0.19 (0.08, 0.32)**
HBA	0.18 (0.07, 0.33)	0.76 (0.61, 0.89)	**0.19 (0.07, 0.34)**	0.20 (0.08, 0.36)	0.74 (0.59, 0.86)	**0.21 (0.08, 0.37)**
HDL	0.38 (0.17, 0.61)	0.57 (0.37, 0.77)	**0.40 (0.19, 0.62)**	0.21 (0.08, 0.37)	0.73 (0.59, 0.86)	**0.23 (0.09, 0.39)**
SBP	0.17 (0.06, 0.32)	0.81 (0.66, 0.94)	**0.18 (0.07, 0.32)**	0.13 (0.06, 0.24)	0.85 (0.74, 0.94)	**0.14 (0.06, 0.24)**
TG	0.23 (0.08, 0.43)	0.76 (0.58, 0.92)	**0.24 (0.09, 0.42)**	0.30 (0.13, 0.49)	0.68 (0.50, 0.84)	**0.31 (0.14, 0.50)**
RET	0.29 (0.06, 0.78)	1.00 (1.00, 1.00)**	**0.22 (0.07, 0.45)**	0.51 (0.08, 1.35)	1.00 (1.00, 1.00) **	**0.34 (0.10, 0.59)**

4.2.3 Genetic correlation explained by common genome-wide SNPs for diabetic retinopathy and related risk factors

We estimated pairwise genetic correlations for diabetic retinopathy and related risk factors in the bivariate mixed model with the adjustment for age, sex and duration of diabetes, using MCMC sampling (Appendix Figure B-1 and Appendix Figure B-2). The genetic (residual) correlation coefficient is the additive genetic (residual) covariance of a pair of traits normalized by genetic (residual) standard deviation of each trait. The number of MCMC samples collected for each pair of traits is provided in Appendix Table B-1.

In the related sample, we identified negative genetic correlation of HDL-TG (Table 4-5), and significant residual correlations of HDL-TG, BMI-TG, CHL-TG and HBA-TG (Table 4-5). Genetic and residual correlations of diabetic retinopathy with related risk factors were insignificant at the 0.05 significance level in the related sample (Table 4-7). In the unrelated sample, we identified significant residual correlations at the 0.05 significance level of BMI-HBA, BMI-HDL, BMI-SBP, CHL-HDL, CHL-SBP, CHL-TG, CRE-HDL, HBA-TG and HDL-TG (Table 4-6). Diabetic retinopathy showed positive residual correlation with cholesterol in the unrelated sample (Table 4-7). Out of all pairs of traits studied, significant residual correlations of CHL-TG, HBA-TG and HDL-TG were consistently observed in the related and unrelated samples. However, we could not preclude the existence of genetic and residual correlations for pairs of traits that were insignificant in this study due to the wide credible intervals.

Table 4-5 Posterior median and 95% highest posterior density intervals of the genetic (upper triangle) and residual (lower triangle) correlation coefficient in the genetically related sample for retinopathy-related population risk factors. Sample size: 1,193 patients. BMI: body mass index; CHL: cholesterol; CRE: serum creatinine; HBA (HbA1c): glycosylated haemoglobin; HDL (HDL-c): high-density lipoprotein cholesterol; RET: retinopathy; SBP: systolic blood pressure; TG: triglycerides. * Significant at the 0.05 level.

	Related sample						
	BMI	CHL	CRE	HBA	HDL	SBP	TG
BMI		-0.36 (-0.74, 0.12)	0.39 (-0.06, 0.74)	0.20 (-0.31, 0.64)	-0.29 (-0.68, 0.13)	0.16 (-0.35, 0.64)	-0.19 (-0.64, 0.29)
CHL	0.07 (-0.15, 0.32)		0.15 (-0.36, 0.60)	0.05 (-0.46, 0.54)	0.34 (-0.16, 0.74)	0.18 (-0.33, 0.65)	0.07 (-0.46, 0.56)
CRE	-0.02 (-0.28, 0.22)	0.03 (-0.12, 0.16)		0.20 (-0.30, 0.64)	0.27 (-0.18, 0.68)	0.10 (-0.40, 0.57)	0.07 (-0.44, 0.55)
HBA	0.04 (-0.20, 0.26)	0.04 (-0.10, 0.17)	-0.06 (-0.22, 0.08)		-0.34 (-0.73, 0.14)	-0.05 (-0.54, 0.46)	0.03 (-0.49, 0.52)
HDL	0.00 (-0.30, 0.31)	0.05 (-0.19, 0.23)	-0.21 (-0.46, 0.00)	-0.04 (-0.23, 0.17)		0.10 (-0.42, 0.57)	-0.48* (-0.80, -0.02)
SBP	0.09 (-0.16, 0.30)	0.01 (-0.13, 0.14)	-0.10 (-0.24, 0.04)	0.03 (-0.09, 0.17)	0.07 (-0.12, 0.26)		0.14 (-0.38, 0.63)
TG	0.28* (0.04, 0.53)	0.42* (0.28, 0.56)	0.08 (-0.08, 0.24)	0.27* (0.12, 0.41)	-0.39* (-0.56, -0.20)	0.02 (-0.13, 0.17)	

Table 4-6 Posterior median and 95% highest posterior density intervals of the genetic (upper triangle) and residual (lower triangle) correlation coefficient in the genetically unrelated sample for retinopathy-related population risk factors. BMI: body mass index; CHL: cholesterol; CRE: serum creatinine; HBA (HbA1c): glycosylated haemoglobin; HDL (HDL-c): high-density lipoprotein cholesterol; RET: retinopathy; SBP: systolic blood pressure; TG: triglycerides. * Significant at the 0.05 level.

	Unrelated sample						
	BMI	CHL	CRE	HBA	HDL	SBP	TG
BMI		0.18 (-0.31, 0.62)	0.17 (-0.31, 0.61)	-0.29 (-0.69, 0.18)	-0.26 (-0.68, 0.23)	-0.35 (-0.72, 0.11)	0.36 (-0.09, 0.73)
CHL	-0.13 (-0.31, 0.03)		0.18 (-0.29, 0.62)	-0.06 (-0.54, 0.43)	0.22 (-0.29, 0.65)	0.13 (-0.36, 0.59)	0.38 (-0.08, 0.75)
CRE	0.02 (-0.15, 0.16)	-0.05 (-0.18, 0.07)		0.08 (-0.42, 0.52)	0.23 (-0.24, 0.64)	0.13 (-0.35, 0.56)	0.15 (-0.33, 0.59)
HBA	0.23* (0.06, 0.43)	0.08 (-0.04, 0.22)	-0.06 (-0.18, 0.07)		-0.12 (-0.61, 0.37)	-0.05 (-0.53, 0.43)	0.16 (-0.34, 0.62)
HDL	-0.18* (-0.35, -0.01)	0.17* (0.03, 0.30)	-0.16* (-0.31, -0.03)	-0.09 (-0.23, 0.06)		0.04 (-0.45, 0.51)	-0.43 (-0.76, 0.04)
SBP	0.20* (0.04, 0.37)	0.13* (0.02, 0.23)	-0.09 (-0.20, 0.01)	0.06 (-0.05, 0.18)	0.08 (-0.04, 0.19)		0.19 (-0.29, 0.64)
TG	0.17 (-0.04, 0.35)	0.28* (0.12, 0.41)	0.00 (-0.16, 0.15)	0.21* (0.04, 0.36)	-0.44* (-0.57, -0.30)	0.05 (-0.10, 0.17)	

Table 4-7 Posterior median and 95% highest posterior density intervals of genetic (residual) correlation coefficient, r_g (r_e), between retinopathy and related population risk factors. BMI: body mass index; CHL: cholesterol; CRE: serum creatinine; HBA (HbA1c): glycosylated haemoglobin; HDL (HDL-c): high-density lipoprotein cholesterol; RET: retinopathy; SBP: systolic blood pressure; TG: triglycerides. * Significant at the 0.05 level.

	Related sample		Unrelated sample	
Trait	r_g	r_e	r_g	r_e
RET-BMI	0.33 (-0.30, 0.81)	-0.25 (-0.68, 0.12)	0.19 (-0.46, 0.71)	0.00 (-0.29, 0.27)
RET-CHL	0.10 (-0.51, 0.65)	0.17 (-0.05, 0.39)	-0.46 (-0.81, 0.04)	0.37 (0.12, 0.64) *
RET-CRE	0.32 (-0.27, 0.77)	0.16 (-0.10, 0.36)	0.16 (-0.40, 0.67)	0.09 (-0.12, 0.28)
RET-HBA	0.13 (-0.47, 0.65)	0.14 (-0.07, 0.34)	0.34 (-0.25, 0.78)	0.05 (-0.22, 0.27)
RET-HDL	0.09 (-0.55, 0.68)	-0.20 (-0.55, 0.10)	-0.15 (-0.67, 0.47)	-0.03 (-0.27, 0.21)
RET-SBP	-0.11 (-0.66, 0.50)	0.08 (-0.12, 0.31)	0.27 (-0.31, 0.72)	0.04 (-0.16, 0.20)
RET-TG	-0.06 (-0.66, 0.55)	0.13 (-0.11, 0.40)	-0.10 (-0.67, 0.49)	0.10 (-0.18, 0.38)

4.3 Conclusions

In this study, we analysed the narrow-sense heritability and genetic correlations captured by common whole-genome SNPs for the severity of retinopathy and related population risk factors in the GoDARTS diabetic population. Whilst the same methodology for analysing heritability and the genetic/residual correlation was also applicable to a genetically mixed samples, the partition of the study cohort into genetically related and unrelated samples permitted separate considerations for similar environment effects and stronger genetic similarities amongst related individuals compared to those unrelated. This study was based on MCMC Bayesian inference, which made the analysis of multi-categorical clinical data of diabetic retinopathy in univariate and bivariate mixed models feasible. We found statistically significant heritability estimates for BMI (49%), CHL (20%), CRE (23%), HBA (21%), HDL (40%), SBP (18%) and TG (31%) in the GoDARTS study data, and the estimates for BMI, HDL, SBP and TG were similar to the reported values using a comparable approach in a non-diabetic population [108]. We found that 34% of retinopathy phenotypic variations wereinherited additive genetic effects. In both related and unrelated samples, we identified positive residual (environmental) correlations for TG with CHL and HBA, respectively, and additionally confirmed negative genetic and residual correlation for TG with HDL, which was previously reported [108]. Residual correlation of diabetic retinopathy with CHL was observed in the unrelated sample. These residual correlations suggest shared environmental aetiology for the observed physiological relation between lipids and lipids with glycaemic exposure. We found no significant genetic correlation between any of the metabolic traits with retinopathy; however, the sample sizes we had in this study may be the limiting factor for the accurate determination of plausible genetic

correlations between retinopathy and related risk factors, using the MCMC estimation approach.

We estimated the genetic similarity matrix based on 6,830,657 imputed autosomal SNPs, which was substantially a larger set of genetic variants compared to genotyped SNPs used in previous studies [104,108,157]. In one study [108], it was reported that extra SNPs had little additional effects on the heritability estimates, once analysed SNPs reached a certain size. The Gibbs sampling approach we applied was computationally intensive and required days and sometimes months of computation for bivariate mixed models, which may be an opportunity for further algorithm development to accelerating mixing of MCMC samples and reducing compute time.

This study estimated narrow-sense heritability (34%) captured by genome-wide SNPs for the severity of diabetic retinopathy, which was analysed as an ordinal variable including all stages of retinopathy defined by the Scottish diabetic retinopathy grading scheme 2007 v.1.1. This estimate was supported by previous evidence that the narrow-sense heritability of proliferative diabetic retinopathy was 0.52 ± 0.31 [70], and the broad-sense heritability estimate was 27% for the severity of diabetic retinopathy [69]. Few studies reported the heritability estimate for total cholesterol (CHL). Previous studies identified that the heritability estimates in the non-diabetic population for BMI, CRE, HBA, HDL, SBP and TG can be as high as 58% [158], 64.1% [159], 59% [160], 46% [161], 57% [159] and 36% [162], respectively. These results provided strong evidence for the heritability estimates in this study.

In this study, we estimated genetic and residual correlation for diabetic retinopathy and related risk factors. Negative genetic correlation (posterior median: -0.48; 95% highest posterior density/HPD interval: -0.80, -0.02) was observed for HDL-TG in the related sample. Although this result was not replicated in the unrelated sample of this study (posterior median: -0.43; 95% HPD interval: -0.76, 0.04) owing to a large credible interval, genetic correlations of HDL-TG were consistently observed in the non-diabetic related (unrelated) sample of Atherosclerosis Risk in Communities (ARIC) [108], with the maximum likelihood estimate/MLE: -0.59, 95% confidence interval/CI: -0.84, -0.34 (MLE: -0.57, 95% CI: -0.94, -0.20). The ARIC study included 530 related (the relatedness coefficient between 0.35 and 0.65) individuals and 5,647 unrelated (the relatedness coefficient less than 0.025) individuals. This large genetic correlation between HDL-TG is directly reflective of a known physiological relationship [163]. In the ARIC study [108], no significant genetic correlation was detected for BMI-HDL, BMI-SBP, BMI-TG, HDL-SBP, SBP-TG, which is consistent with findings from our study. Compared to the ARIC study [108], we had approximately half of the size in the unrelated sample, and additionally, we had similarly ranged credible intervals. In this study, it is probable that the current sample size is not adequate in power to estimate genetic and residual correlations with modest effect sizes.

We found large residual correlations for CHL-TG, HBA-TG, HDL-TG in the related and unrelated samples, and RET-CHL in the unrelated sample. Observed residual correlations for lipids are likely attributed to a known environmental aetiology [163]. The residual correlation of HBA-CHL shown in this study suggests shared environmental aetiology to hyperglycaemia and hyperlipidaemia, supported by studies of heart diseases and diabetes

[164], although the full explanation for this link remains elusive. We detected a strong residual correlation for RET-CHL, which suggests the increased susceptibility of retinopathy may be associated with hyperlipidaemia due to a common environmental exposure including diet and lifestyle. The physiological relation between HBA, CHL and RET may be further characterised in studies for common genetic and environmental aetiology underlying hyperglycaemia, hyperlipidaemia and retinopathy, in a single analysis using multi-variate mixed models.

A drawback of this study lies within the use of clinical data unadjusted for the treatment effects of hyperglycaemia, hyperlipidaemia and hypertension. However, given the successful detection of known genetic and residual correlations between lipids, we consider the analytical approach applied here is robust for studying treatment-unadjusted clinical response data.

In summary, this study found a large portion of phenotypic variation in retinopathy and related risk factors in the diabetic population is heritable, as captured by common genome-wide SNPs. These findings are complementary to discoveries made in the population-based association studies for diabetic retinopathy and related risk factors using common genome-wide SNPs [28,71–73]. This study provided the evidence for genetic and environmental correlations between metabolic traits in the diabetic populations, which has been previously observed in non-diabetic population [108]. This is consistent with the premise of the analysis of genetic correlated phenotypes in genome-wide association scans [165]. We detected positive environmental correlation of retinopathy with total cholesterol, and cholesterol with HbA1c, which prompts joint correlation analysis of retinopathy with lipids

and glycaemic exposure for improving clinical management of diabetic retinopathy. In agreement with a previous study using a similar approach [108], we found the precision for heritability and correlation estimates from genome-wide SNPs could be improved with an increased sample size in future studies.

Chapter 5

The SUMMIT genome-wide meta-analysis of diabetic retinopathy

5.1 Introduction

As demonstrated in the previous chapter, mounting evidence suggests there is a strong heritable genetic contribution to the development of retinopathy in the GoDARTS population sample. To date, genome-wide association studies (GWAS) have been shown to be a fruitful strategy in identifying natural genetic variations and causal genes associated with type 1 and type 2 diabetes [72,80,166,167], pharmacoresponses in the type 2 diabetic population [86] and additional diabetes-related complex traits [168]. However, the handful of diabetic retinopathy GWAS that are published have not identified any loci with genome wide significance [28,81–83]. These analyses are based upon limited sample sizes of

diabetic retinopathy cases and controls, and for example, the whole-genome meta-analysis of severe diabetic retinopathy, the largest retinopathy data analysis of all published retinopathy GWAS, included 937 cases and 1,856 controls. These sample sizes are underpowered to identify retinopathy risk loci with moderate effects.

Compared to these previous analyses, in this study, the GoDARTS cohort contributed 1,942 diabetic retinopathy cases and 2,014 controls in the discovery phase of a whole-genome association scan, and overall with collaborators in SUMMIT, the meta-analysis included 5,422 cases and 4,302 controls in the GWAS analysis, which becomes the largest GWAS of diabetic retinopathy to date. In addition, this study analysed 1000 Genomes imputed genetic variants, which is the largest set of variants being studied for the association with diabetic retinopathy. Through these initiatives, we aimed at achieving adequate power to detect diabetic retinopathy susceptibility loci from genome-wide genetic variants, and robustly identify causal genes.

5.2 Results

5.2.1 Characteristics of the GoDARTS study sample

Clinical characteristics of the GoDARTS study sample are provided in Table 5-1. Controls have minimally higher mean of age and lower mean of serum creatinine and HbA1c, but other clinical profiles are comparable.

Table 5-1 Clinical characteristics of retinopathy cases (n = 1942) and controls (n = 2014) of the GoDARTS cohort.

	Cases		Controls	
Trait	Mean	SD	Mean	SD
Age (years)	60.1	11.5	68.9	10.8
BMI (kg/m^2)	30.3	5.5	31.4	6.1
Cholesterol (mmol/l)	4.7	1.1	4.5	1.0
Serum creatinine (μmol/l)	100.3	28.9	89.7	27.2
DBP (mmHg)	76.0	10.3	76.9	10.6
HbA1c % (mmol/mol)	7.8% (61.4)	1.5% (16.4)	7.1% (54.5)	1.2% (13.4)
HDL-c (mmol/l)	1.3	0.4	1.3	0.4
SBP (mmHg)	142.8	18.8	139.2	18.0
Triglycerides (mmol/l)	2.1	1.2	2.3	1.3
Non-HDL-c (mmol/l)	3.4	1.1	3.2	1.0
Sex (%)	F (44.4%)	M (55.6%)	F (45.3%)	M (54.7%)

5.2.2 Meta-analysis study data

The meta-analysis study included 9,724 individuals from EURODIAB, FinnDiane, GoDARTS and SDR cohorts in 1000 Genomes imputed variants (Table 5-2 and Table 5-3). GoDARTS and SDR population samples comprised of genetically unrelated individuals and related relatives. Genome-wide meta-analyses were performed independently in genetically unrelated samples and mixed samples (section 2.8.3).

Table 5-2 Sample sizes for meta-analysis study cohorts. Sizes for genetically unrelated samples are shown in the brackets.

Cohort	Cases	Controls	Total
EURODIAB (T1D)	201	249	450
SDR (T1D)	490 (472)	142 (140)	632 (612)
SDR (T2D)	1151 (1066)	865 (815)	2016 (1881)
GoDARTS (T1,2D)	1142 (1022)	882 (778)	2024 (1800)
GoDARTS-SUMMIT (T1,2D)	800 (725)	1132 (1020)	1932 (1745)
FinnDiane (T1D)	1638	1032	2670
Total	5422 (5124)	4302 (4034)	9724 (9158)

Table 5-3 1000 Genomes imputed genetic variants included in the meta-analyses of genetically mixed and genetically unrelated (numbers enclosed in the brackets) study samples. QC: quality control.

Cohort	Pre-QC genetic variants	Post-QC genetic variants
EURODIAB (T1D)	37,524,910 (37,524,910)	9,091,275 (9,091,275)
SDR (T1D)	15,077,368 (15,091,253)	7,044,378 (7,051,662)
SDR (T2D)	15,078,384 (15,092,269)	9,258,178 (9,267,780)
GoDARTS (T1,2D)	13,658,156 (13,658,156)	8,364,025 (8,364,025)
GoDARTS-SUMMIT (T1,2D)	15,005,754 (15,005,754)	9,037,594 (9,037,594)
FinnDiane (T1D)	14,324,304 (14,337,518)	9,330,927 (9,344,876)
Meta-analysis	---	9,508,089 (10,444,012)

5.2.3 Loci associated with diabetic retinopathy

The resulting Manhattan plots of meta-analysis association results are shown in Figure 5-1, which showed top association signals clustered on chromosome 1 (1p34-p32 and 1q43), chromosome 2 (2p12-p11.1, 2q11.2 and 2q32.1), chromosome 3 (3p24.3 and 3q24), chromosome 4 (4q28-q32), chromosome 5 (5p14 and 5q31.3), chromosome 6 (6p21 and 6q16.1), chromosome 9 (q13-q21.2), chromosome 11 (11q21), chromosome 12 (12p11.22), chromosome 13 (13q32), chromosome 17 (17q25.1), chromosome 19 (19q13.11) (Table 5-4 and Table 5-5).

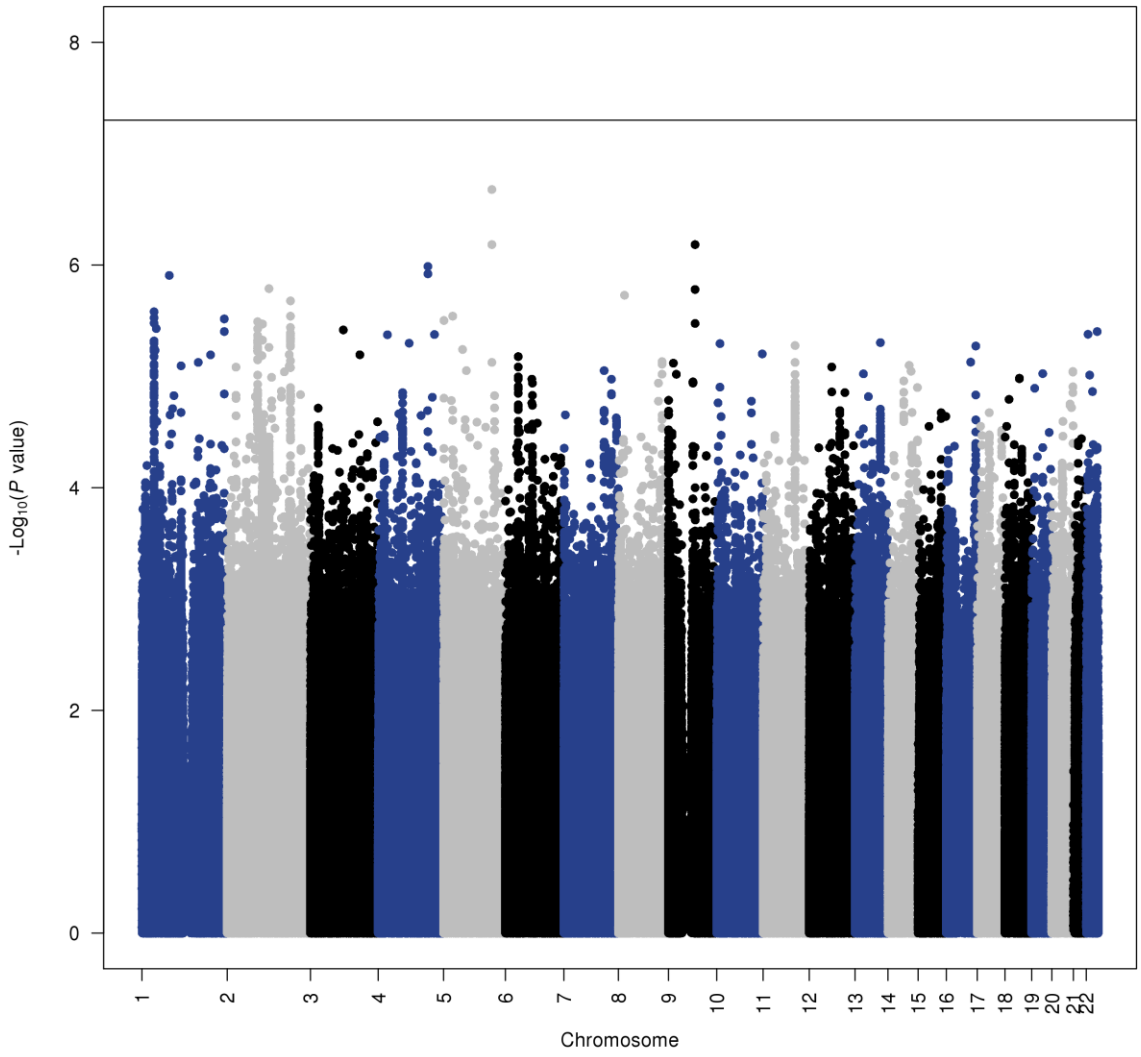
The resulting quantile-quantile plots are shown in Figure 5-2, and the genomic inflation factor estimates are 1.00 and 1.01, respectively. The quantile-quantile plot of meta-analysis in the mixed genetic relatedness samples showed more associations with significant *P* values observed than expected (Figure 5-2b), although P values of both analyses were strongly correlated (Figure 5-3). *P* values of the top independent signal SNPs were close to the genome-wide significance (that is, 5.00×10^{-8}) (Table 5-4). Top signal variants in independent regions of associations with *P* value < 10^{-5} are provided in Appendix Table C-1, and the functional annotation of the closest gene is shown in Appendix Table C-2. Four independent signal SNPs had *P* values below 10^{-6}, with the highest significant locus being rs10746970 (*P* value of 2.22×10^{-7}). Heterogeneity of genetic effects was minimally small for top ranked SNPs in the meta-analysis (Table 5-4). In the regional association plots (Figure 5-4), the four top ranked SNPs each showed strong linkage disequilibrium with adjacent SNPs, suggestingthe association signal originated from a single region. We found that top signal SNPs are close to genes that are implicated in carbohydrate/lipid metabolic processes, function, activity of neuronal and glial cells, and transcription regulation (Table

5-5). As of 18 October 2014, none of these top independent association SNPs was previously reported in the database of published GWAS (http://www.genome.gov/gwastudies/) [169].

With the estimated prevalence (9.59%) of diabetic retinopathy in GoDARTS at the recruitment time, statistical power for achieving genome-wide significance in each single-marker test of the genome-wide meta-analysis rose with increased relative risk and minor allele frequencies (MAFs) (Table 5-6 and Figure 5-5). With the current sample size, maximum achievable power at the relative risk of 1.1 or less was 0.025, and when the relative risk was 1.5 or more and MAF was 0.05 or greater, power was 0.973 or greater (Table 5-6).

Figure 5-1 Manhattan plots of meta-analysis association results in the genetically unrelated samples (a) and mixed samples (b). The genome-wide significance threshold (*P* value of 5×10^{-8}) is shown as the horizontal line.

(a)



(b)

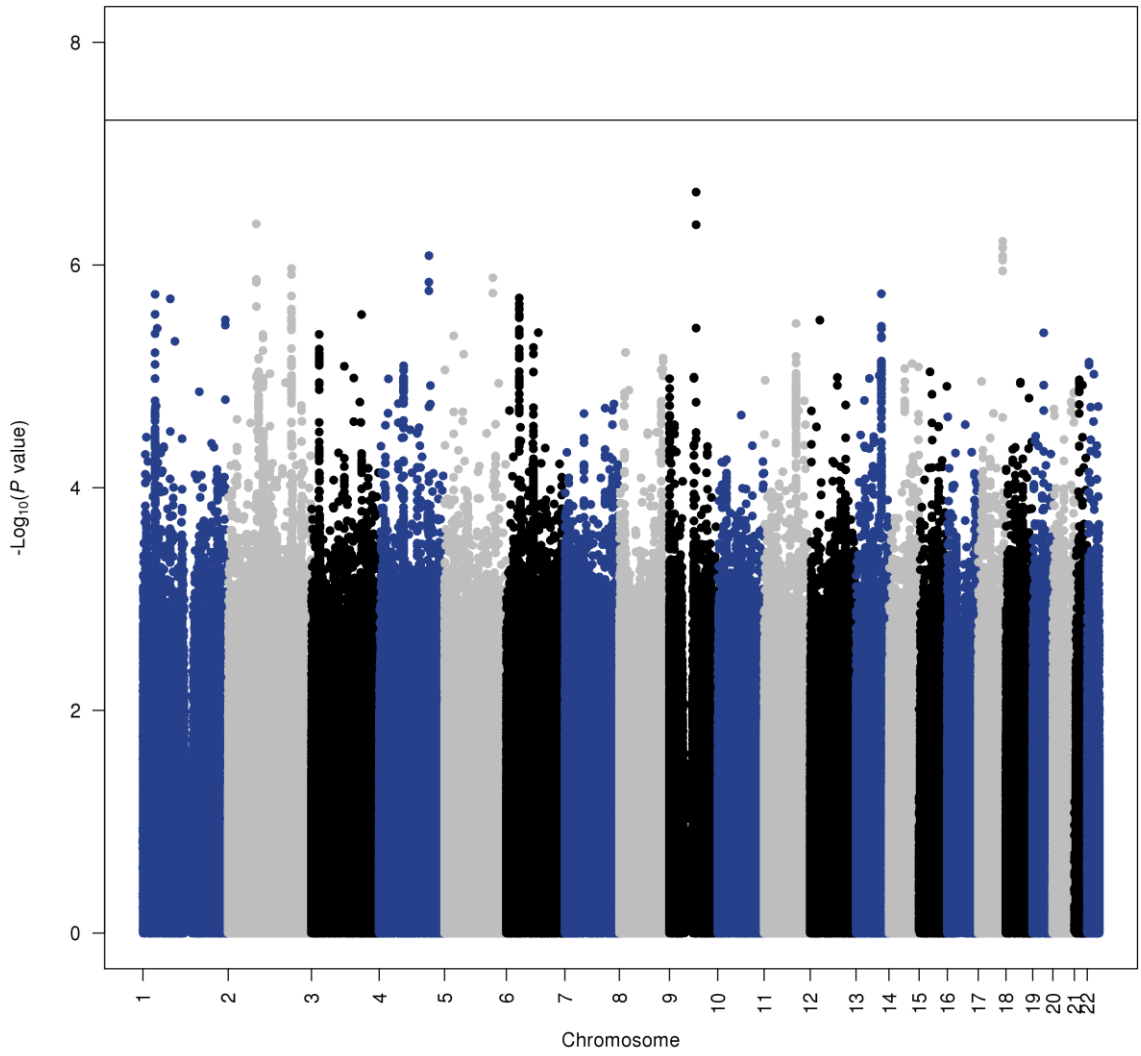
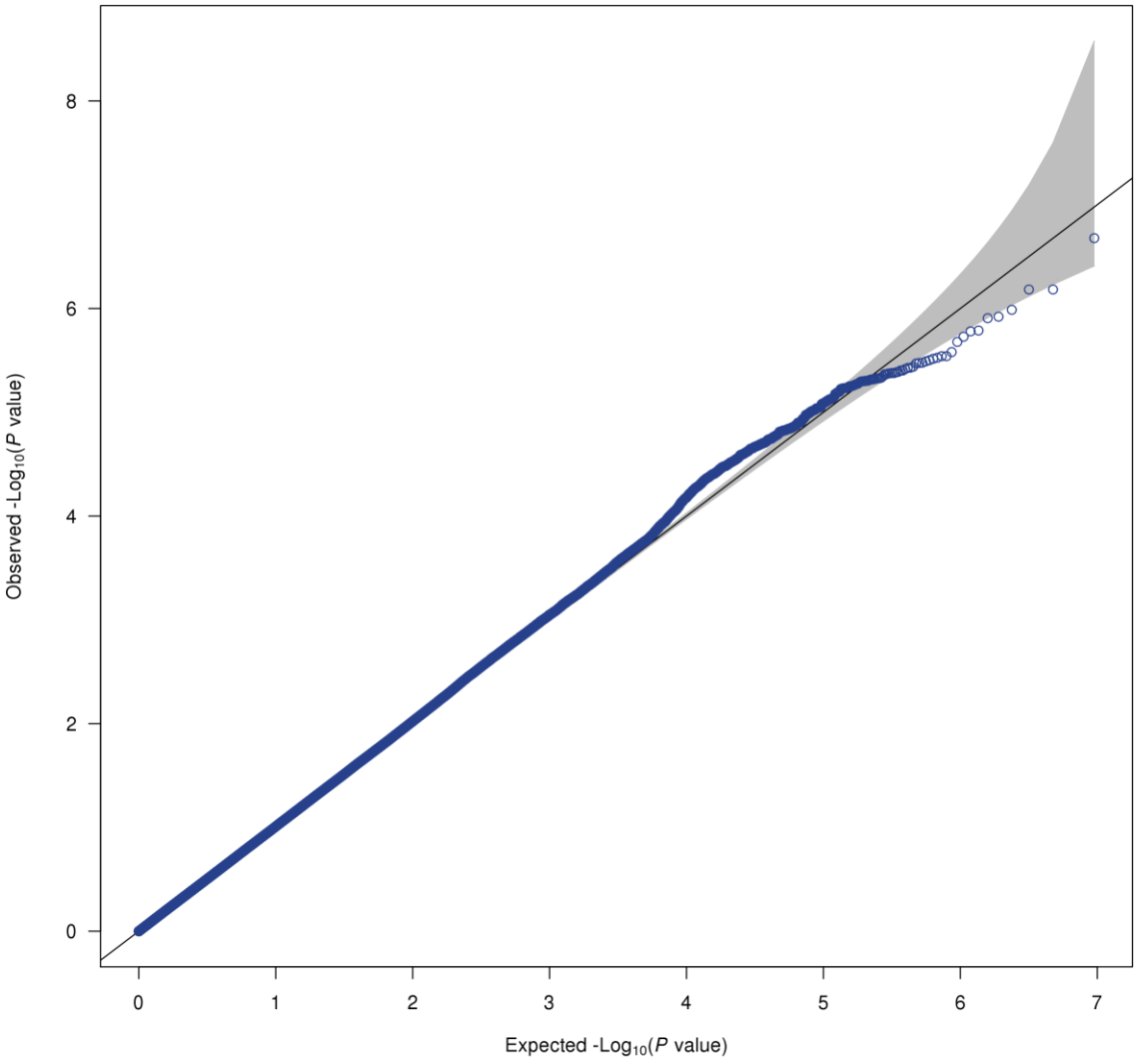

8
6
4
2
0
-Log10(P value)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

Figure 5-2 Quantile-quantile plots of meta-analysis association results in the genetically unrelated samples (a) and mixed samples (b). 95% confidence intervals are shown as the shaded area.

(a)



(b)

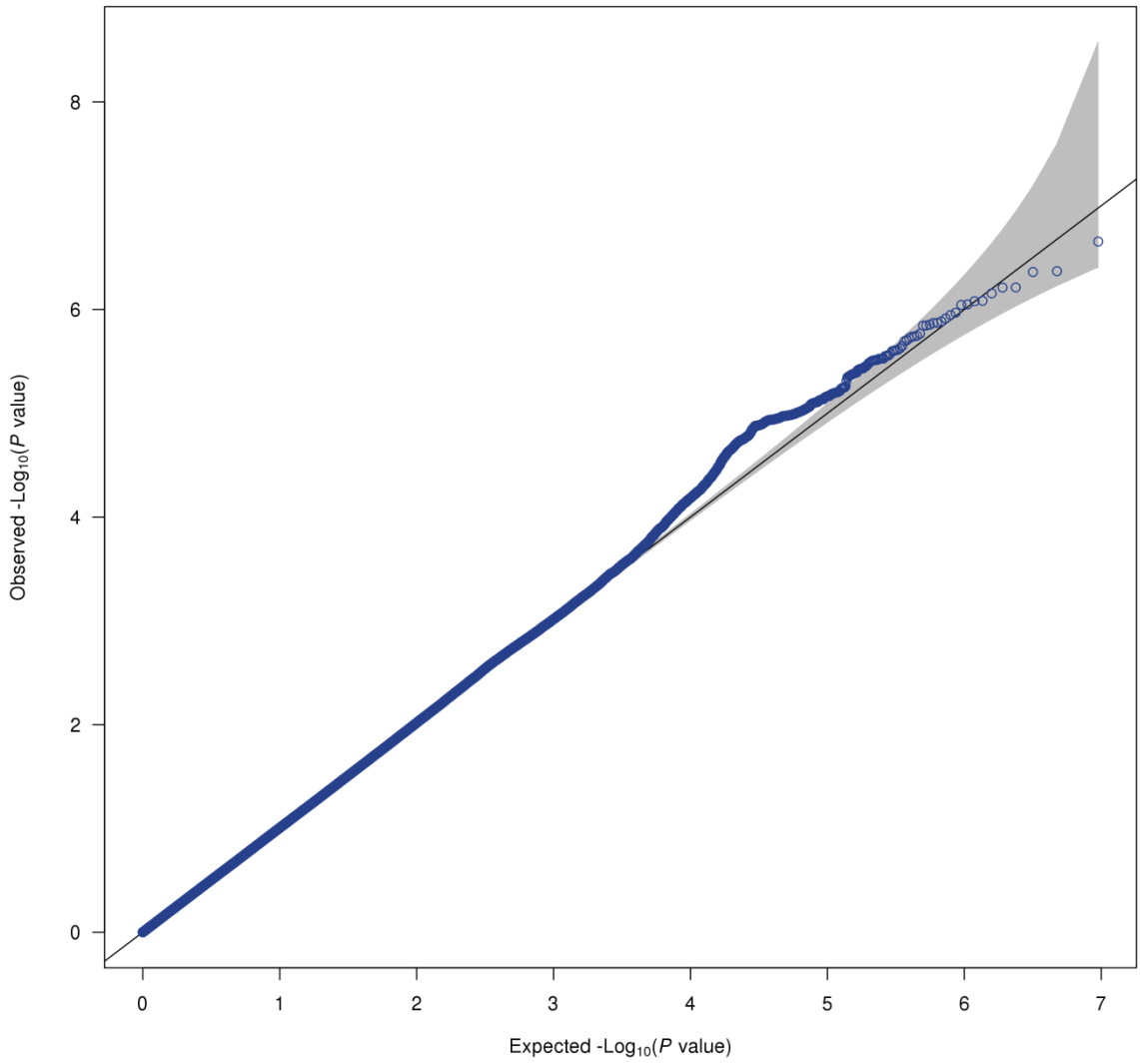

Observed -Log10(P value)
Expected -Log10(P value)
0
2
4
6
8
0
1
2
3
4
5
6
7

Figure 5-3 Correlations of meta-analysis *P* values, provided by Dr William Rayner, Oxford University, based on the genetically unrelated sample analysed by the SNPTEST software (version 2.4.1) and the genetically mixed samples analysed by EMMAX beta or GEMMA (version 0.93).

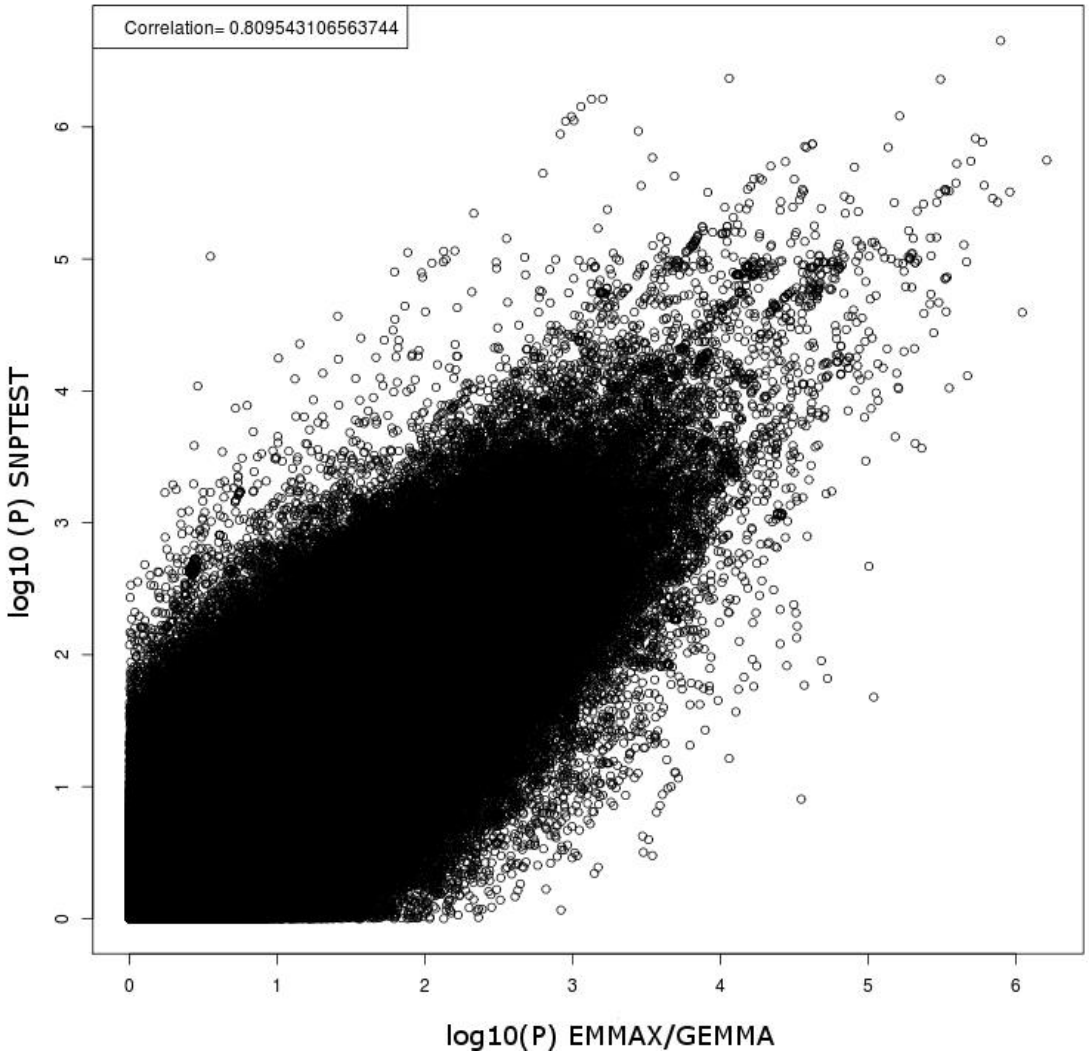
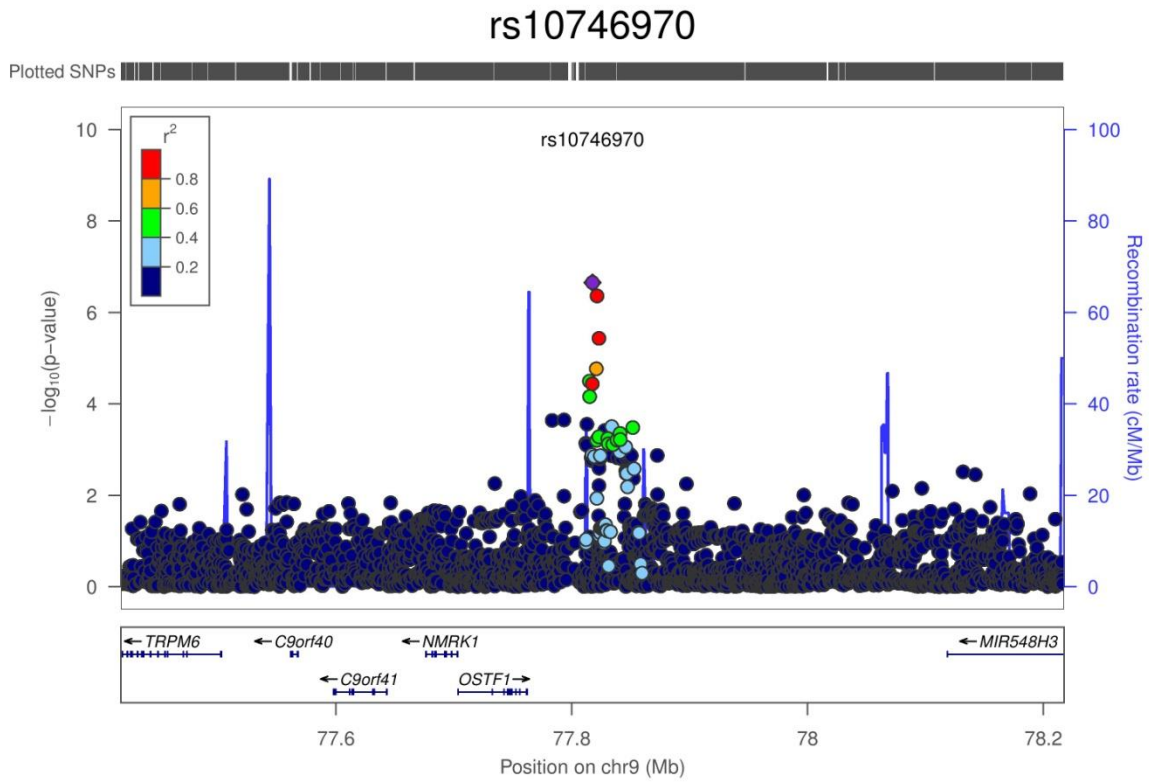
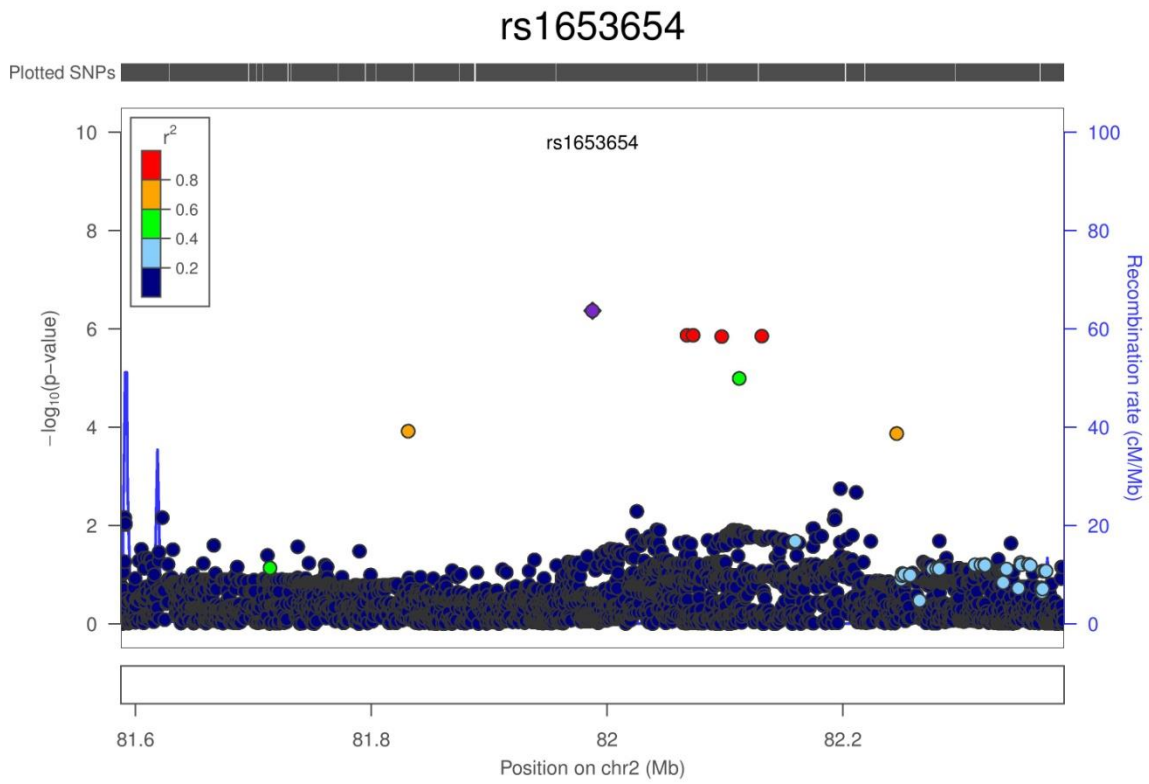


Figure 5-4 Regional association plots for top independent signal SNPs of (a) rs10746970, (b) rs1653654, (c) rs75125621 and (d) rs2657795 in the SUMMIT meta-analysis of diabetic retinopathy. In the association region, neighbouring SNPs were coloured according to the linkage disequilibrium (r^2) with the index SNP (purple). Blue spikes indicate loci of high recombination rates within the 400 K base pair (bp) either side of the index SNP.

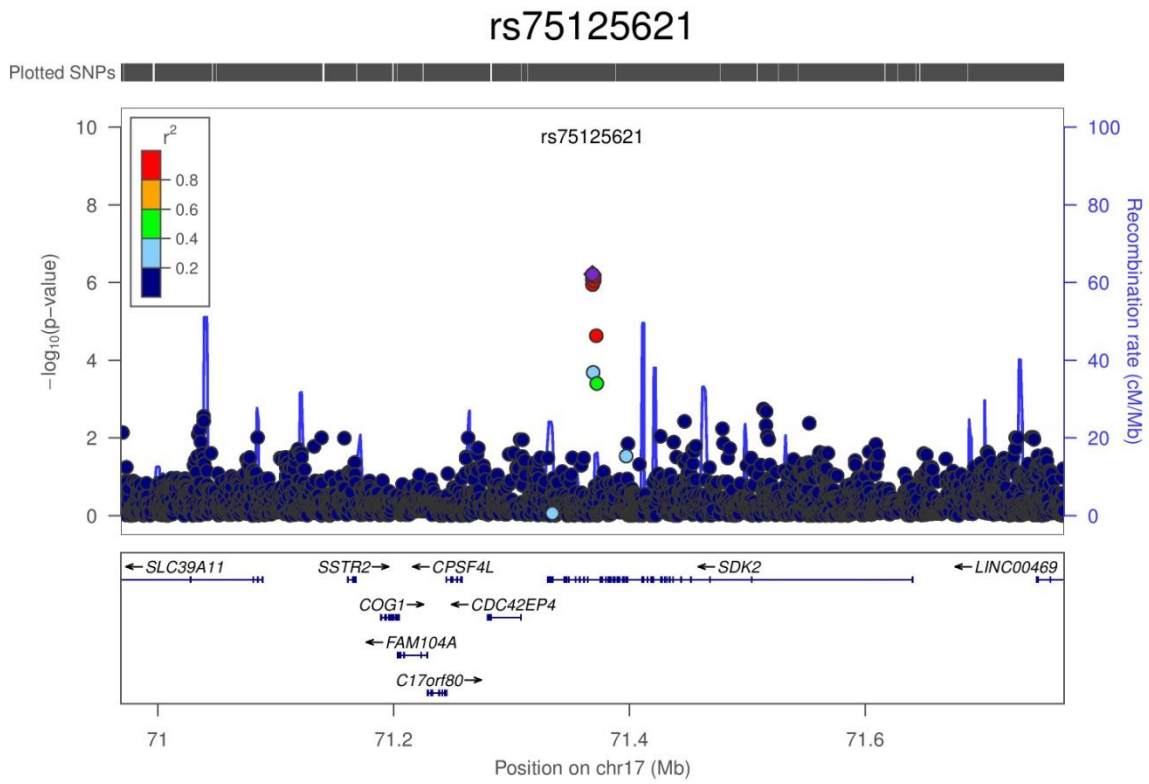
(a)



(b)


rs1653654
Plotted SNPs
rs1653654
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
10
8
6
4
2
0
100
80
60
40
20
0
81.6
81.8
82
82.2
Position on chr2 (Mb)

(c)


rs75125621
Plotted SNPs
rs75125621
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
SLC39A11
SSTR2
CPSF4L
SDK2
LINC00469
COG1
CDC42EP4
FAM104A
C17orf80
71
71.2
71.4
71.6
Position on chr17 (Mb)

(d)

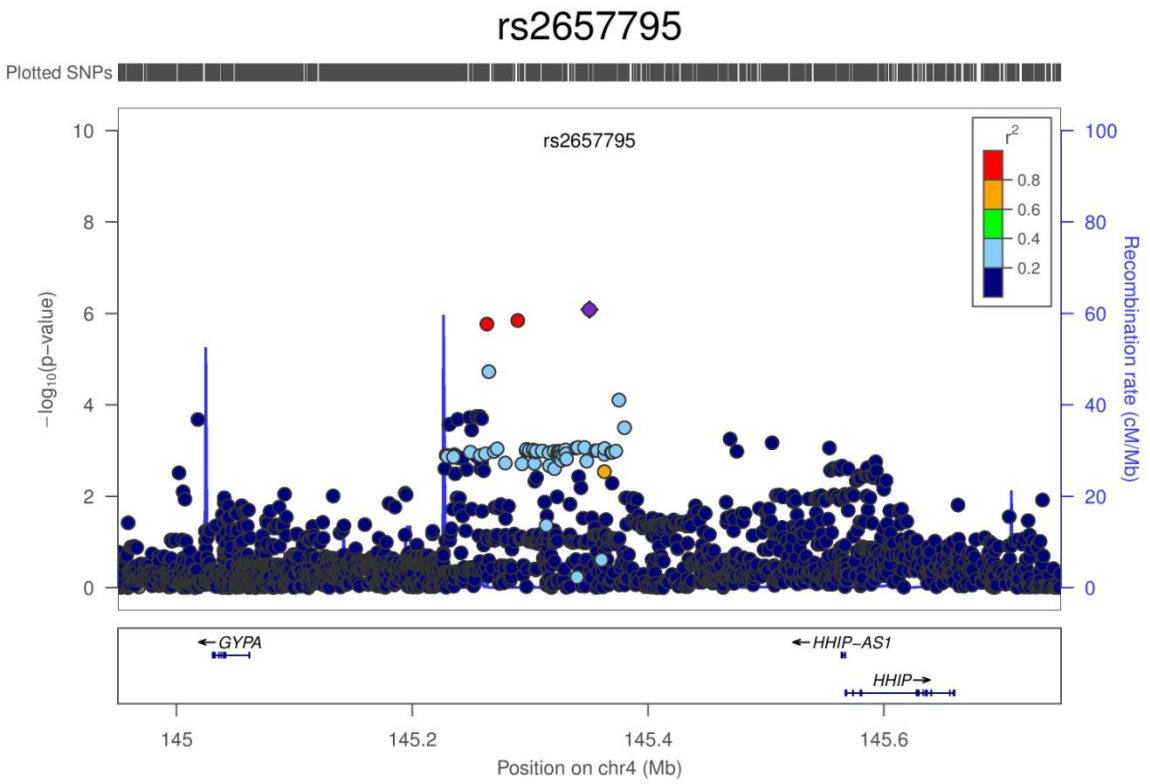

rs2657795
Plotted SNPs
rs2657795
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
GYPA
HHIP−AS1
HHIP
145
145.2
145.4
145.6
Position on chr4 (Mb)

Table 5-4 Top independent (defined by >100 Kbp) signal SNPs in the whole-genome meta-analysis of diabetic retinopathy. Chr: chromosome.

Chr.	Position[a]	SNP	Effective allele frequency[b]	Effective/ alternative allele[c]	OR (95% CI)[d]	Meta *P* value[e]	I^2 (%), *P* value[f]	Effect directions[g]
9	77817536	rs10746970	0.68	A/T	1.19 (1.11, 1.27)	2.22×10^{-7}	0, 6.46×10^{-1}	++++++
2	81987573	rs1653654	0.03	A/G	0.63 (0.51, 0.78)	4.27×10^{-7}	0, 8.32×10^{-1}	---?--
17	71368680	rs75125621	0.03	G/A	0.61 (0.49, 0.77)	6.13×10^{-7}	60, 4.03×10^{-2}	-+-?--
4	145350264	rs2657795	0.21	A/C	0.82 (0.76, 0.89)	8.25×10^{-7}	2, 3.98×10^{-1}	------
2	184738091	rs1682430	0.89	C/T	1.24 (1.10, 1.39)	1.07×10^{-6}	21, 2.71×10^{-1}	+++??+
5	141158686	rs10072382	0.02	A/G	1.99 (1.52, 2.62)	1.30×10^{-6}	28, 2.41×10^{-1}	+?+?++
13	93033916	rs9516067	0.56	A/G	1.16 (1.09, 1.24)	1.81×10^{-6}	51, 6.69×10^{-2}	++++++
1	35627132	rs6425936	0.50	A/G	0.76 (0.68, 0.85)	1.83×10^{-6}	16, 3.08×10^{-1}	-??---
6	37583911	rs810855	0.21	C/T	0.83 (0.77, 0.90)	1.98×10^{-6}	0, 7.97×10^{-1}	------
1	80022362	rs55939932	0.02	T/C	0.48 (0.36, 0.65)	2.01×10^{-6}	0, 8.98×10^{-1}	---?--
3	146324544	rs163757	0.36	A/T	0.86 (0.80, 0.93)	2.79×10^{-6}	0, 9.99×10^{-1}	------
1	240674842	rs35428289	0.21	A/G	1.22 (1.12, 1.32)	3.11×10^{-6}	0, 4.73×10^{-1}	++++++
12	27969088	rs10771375	0.44	G/A	0.87 (0.81, 0.93)	3.13×10^{-6}	0, 8.48×10^{-1}	------
11	93564393	rs601711	0.64	T/C	1.17 (1.09, 1.25)	3.35×10^{-6}	0, 4.55×10^{-1}	++++++
1	41835871	rs4660191	0.97	G/A	1.75 (1.38, 2.21)	3.71×10^{-6}	6, 3.76×10^{-1}	-+++++

Chr.	Position[a]	SNP	Effective allele frequency[b]	Effective/ alternative allele[c]	OR (95% CI)[d]	Meta *P* value[e]	I^2 (%), *P* value[f]	Effect directions[g]
6	93361564	rs7750013	0.94	C/A	1.45 (1.22, 1.73)	4.05×10^{-6}	0, 7.35×10^{-1}	++++++
19	32473222	rs8113622	0.33	T/G	0.86 (0.80, 0.92)	4.06×10^{-6}	0, 6.13×10^{-1}	------
3	22624716	rs73033654	0.08	T/C	1.30 (1.15, 1.47)	4.20×10^{-6}	34, 1.78×10^{-1}	+++-++
2	101346696	rs7579862	0.19	G/A	1.19 (1.08, 1.30)	4.22×10^{-6}	0, 5.36×10^{-1}	+++??+
5	26698737	rs72758936	0.03	A/G	0.63 (0.52, 0.77)	4.33×10^{-6}	39, 1.61×10^{-1}	+--?--

[a] Human genome build 37, assembly hg19.

[b, c & d] Effective allele frequency, effective/alternative allele, odd ratio and 95% CI were extracted from the meta-analysis of the genetically unrelated samples, whereas in the genetically mixed samples, estimates of effective allele frequency and odds ratio may be biased.

[e & f] Meta and heterogeneity *P* values were extracted from the meta-analysis of the genetically mixed samples, which included more individuals, but this modelling strategy does not provide the estimation of odds ratios.

[g] Directions of effect (that is the sign of log odd ratio) were extracted from the meta-analysis of the genetically unrelated samples in the order of EURODIAB, GoDARTS Affymetrix and Illumina genotyped samples, SDR type 1 and 2 diabetes samples and FinnDiane.

Table 5-5 The closest gene and the gene functional annotation for top independent (defined by >100 Kbp) signal SNPs in the whole-genome meta-analysis of diabetic retinopathy. The regulatory region of a gene is defined by <250 Kbp upstream from the transcription start site (TSS), and >250 Kbp upstream from TSS is characterised as intergenic.

Locus	SNP	Region	Closest Gene	Gene Function
9q13-q21.2	rs10746970	Regulatory	*OSTF1*	Ossification
2p12-p11.1	rs1653654	Intergenic	*CTNNA2*	Structural constituent of cytoskeleton
17q25.1	rs75125621	Intronic	*SDK2*	Cell adhesion
4q28-q32	rs2657795	Regulatory	*HHIP*	Carbohydrate metabolic process
2q32.1	rs1682430	Intergenic	*ZNF804A*	Transcriptional control
5q31.3	rs10072382	Regulatory	*PCDH1*	Nervous system development
13q32	rs9516067	Intronic	*GPC5*	Carbohydrate metabolic process
1p34.3	rs6425936	Regulatory	*SFPQ*	DNA recombination and repair, RNA splicing
6p21	rs810855	Regulatory	*MDGA1*	Neuron migration
1p33-p32	rs55939932	Regulatory	*ELTD1*	Neuropeptide signaling pathway
3q24	rs163757	Exonic	*PLSCR5*	Unknown
1q43	rs35428289	Intronic	*GREM2*	Regulation of cytokine activity
12p11.22	rs10771375	Regulatory	*KLHL42*	Regulation of microtubule-based process
11q21	rs601711	Intronic	*VSTM5*	Transmembrane protein
1p34.2	rs4660191	Intronic	*FOXO6*	Sequence-specific DNA binding transcription factor activity

Locus	SNP	Region	Closest Gene	Gene Function
6q16.1	rs7750013	Intergenic	*EPHA7*	Retinal ganglion cell axon guidance
19q13.11	rs8113622	Intergenic	*ZNF507*	Regulation of transcription
3p24.3	rs73033654	Regulatory	*ZNF385D*	Regulation of transcription
2q11.2	rs7579862	Regulatory	*NPAS2*	Cellular lipid metabolic process, central nervous system development
5p14	rs72758936	Regulatory	*CDH9*	Cell-cell adhesion

Figure 5-5 Power estimates for a additive single-marker test in a genome-wide meta-analysis/association studty of diabetic retinopathy, with retinopathy prevalence of 9.59% estimated in the GoDARTS diabetic populaiton at the recruitment time. The data were stratified by minor allele frequencies (MAF) and the effect sizes (relative risk) displayed in panels from left to right. Power and the case/control size (assuming equally sized cases and controls) were shown on the y- and x-axes respectively.

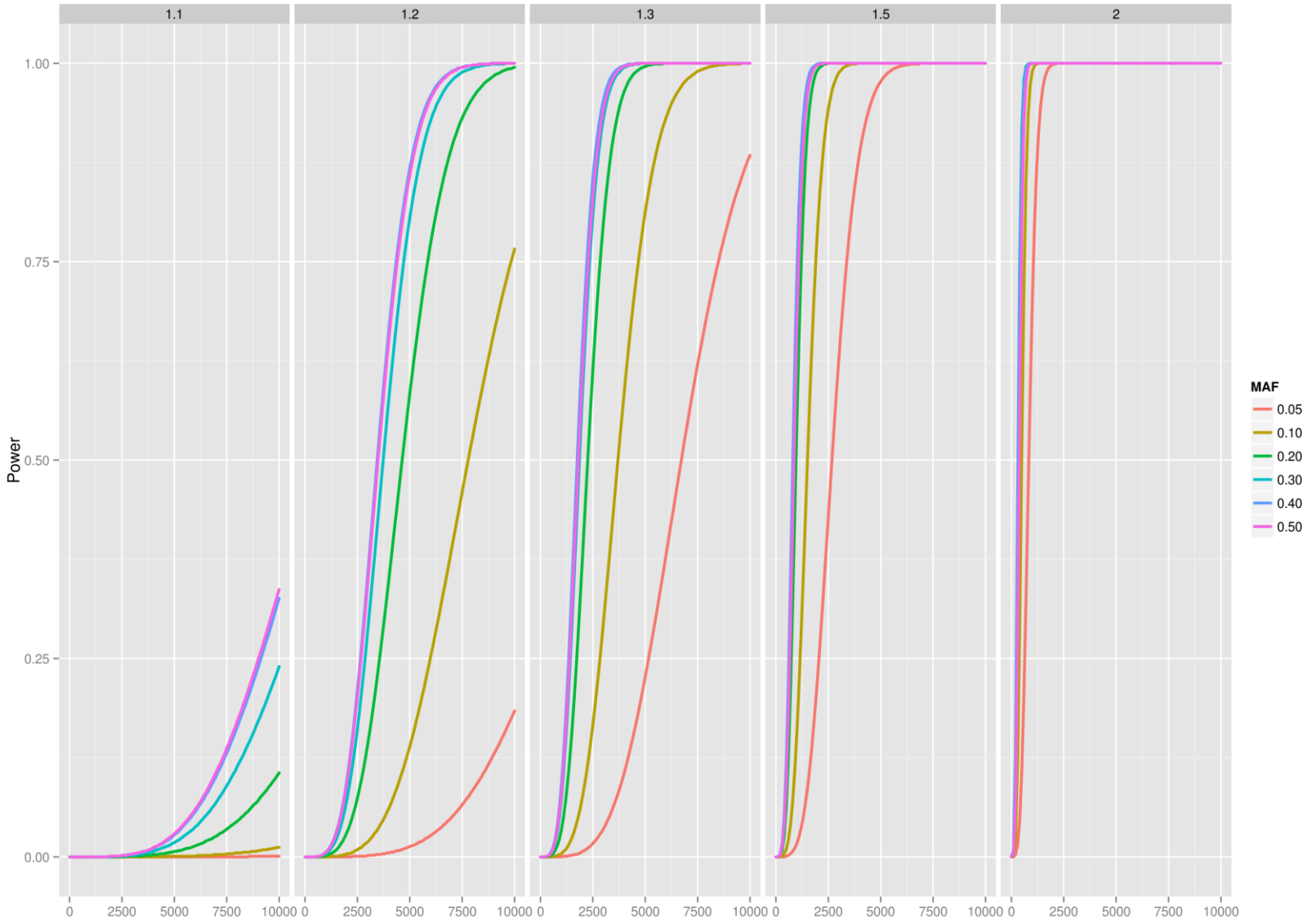

1.1
1.2
1.3
1.5
2
Power
1.00
0.75
0.50
0.25
0.00
0
2500
5000
7500
10000
MAF
0.05
0.10
0.20
0.30
0.40
0.50

Table 5-6 Power estimates for each addititive single-marker test of genome-wide meta-analysis of diabetic retinopathy (5,422 cases and 4,302 controls). MAF: minor allele frequency. The assumed prevalence rate was 9.59% as in GoDARTS at the recruitment time.

Relative risk	MAF	Power
1.1	0.05	0.000
1.1	0.10	0.001
1.1	0.20	0.006
1.1	0.30	0.015
1.1	0.40	0.023
1.1	0.50	0.025
1.2	0.05	0.011
1.2	0.10	0.124
1.2	0.20	0.545
1.2	0.30	0.771
1.2	0.40	0.838
1.2	0.50	0.827
1.3	0.05	0.209
1.3	0.10	0.787
1.3	0.20	0.994
1.3	0.30	1.000
1.3	0.40	1.000
1.3	0.50	1.000
1.5	0.05	0.973
1.5	0.10	1.000
1.5	0.20	1.000
1.5	0.30	1.000
1.5	0.40	1.000
1.5	0.50	1.000
2.0	0.05	1.000

2.0	0.10	1.000
2.0	0.20	1.000
2.0	0.30	1.000
2.0	0.40	1.000
2.0	0.50	1.000

5.3 Conclusions

In this study, we performed the largest genome-wide meta-analysis to date of retinopathy from patients with type 1 and type 2 diabetes. This study considered genetic unrelated and mixed relatedness samples separately, as the unbiased estimation of odds ratio requires the study sample reflective the underlying population that consists of largely genetically unrelated individuals. Besides, so as to increase study power, linear mixed model was applied to the analysis of the genetically mixed sample, which in contrary to the logistic regression approach employed separately for the unrelated sample, does not provide an estimate for odds ratio. Thus, we included the statistics from these analyses. In the current study, we identified the most significant signal SNP rs10746970 (*P* value of 2.22×10^{-7}). In this study, we found diabetic retinopathy risk loci 1q43, 2q32.1, 6q16.1 13q32 identified in this study are in close proximity to regions of 1q13-42, 2q31-47, 6q22-27 and 13q14-32 that were previously reported in smaller sized genome-wide association studies of diabetic retinopathy [28,81–83]. Additionally, we discovered previously unreported retinopathy risk loci of 1p34-p32, 2p12-p11.1, 2q11.2, 3p24.3, 3q24, 4q28-q32, 5p14, 5q31.3, 6p21, q13-q21.2, 11q21, 12p11.22, 17q25.1 and 19q13.11 (Table 5-9). The identification of extensive susceptibility loci is suggestive of polygeneic effects contributing to the development of diabetic retinopathy. The closest genes to these loci have been implicated in multiple physiological processes including carbohydrate/lipid metabolism, functional activity in the neuronal and glial cells, and transcription regulation.

The strongest signal SNP rs10746970 is located in the intergenic region of *OSTF1* (osteoclast stimulating factor 1) at 9q13-21.2. Encoded OSTF1 is an intracellular protein, and its known function include the indirect stimulation of the formation of osteoclasts and

bone resorption, a continual bone renewal process involving bone breaking down by osteoclasts, releasing of minerals in the medium of bone fluid into the circulation [170]. The OSTF1 protein is expressed in brain, platelets and white blood cells including monocytes, neutrophils, and lymphocytes (based on the protein expression databases of MOPED [171], PaxDb [172] and MaxQB [173]). OSTF1 was studied in a SNP array of candidate genes for osteoporosis, an imbalance between the resorption and the formation of bone [174]. However, as of 18 October 2014, neither this SNP nor the closest gene OSTF1 was reported in the GWAS Catalog. . Patients with the genetic disorder of 9q21 deletions showed mental retardation with speech delay, epilepsy and facial dysmorphy [175]. The physiological role of OSTF1 in the aetiology of diabetic retinopathy is currently unknown.

The second strongest association signal rs1653654 is in an intergenic region of *CTNNA2,* catenin (cadherin-associated protein) alpha 2 at 2p12-p11.1. Catenin alpha-2 is a scaffold protein in the cytoskeleton, and its function is implicated in cell-cell adhesion, cell morphogenesis and differentiation of the nervous system. Catenin alpha-2 is highly expressed in the brain (based on the protein expression databases of MOPED, PaxDb and MaxQB). No reported association was found for rs1653654 in any published GWAS. However, the *CTNNA2* gene was reported in the genome-wide association analyses of eye colours [176], bipolar disorder [177], Alzheimer's disease [178], orthostatic hypotension [179] and excitement-seeking [180]. The connection of CTNNA2 with diabetic retinopathy awaits further investigations.

These support suggestive functional associations of top ranked SNPs in the meta-analysis with diabetic retinopathy. However, statistically significant SNPs identified in this study

were novel (Appendix Table A-1). These variants and the nearest genes have not been reported in candidate polymorphism, genome-wide linkage or previous genome-wide association studies (Table 5-7, Table 5-8 and Table 5-9). None of the previous GWAS signals shown in Table 5-9 achieved the genome-wide significance in our meta-analysis after correction for multiple testing (data not shown).

A major strength of this study is the deeply phenotyped data of retinopathy in the large, homogenous European diabetic cohorts. This should help control population stratification and reduce differential environmental influences. This study attained cohort samples overall close to 10,000 subjects, which increased the opportunity of discovering retinopathy susceptibility loci. Another advantage of this study is the utility of 1000 Genomes imputation which increased the size of genetic variants to several fold greater compared to Hapmap Phase II imputation. This should help increasing the probability of identifying association variants across the allele frequency spectrum, including relatively rare variants. In addition, this study analysed diabetic retinopathy associations in both genetically unrelated and mixed samples. Association data from the genetically unrelated sample provided an unbiased estimation of effect size for the genetic effect. The analysis of the genetically mixed sample enlarged the study sample sizes, resulting in a modest increase in the association results.

In this study, we made every effort to standardise different clinical classification and grading systems for diabetic retinopathy between cohorts in SUMMIT. However, the complete compatibility of diabetic retinopathy phenotype definitions across studies may be difficult to attain. Additionally, the use of imputation to augment the number of genetic

variants in the study may introduce inaccurately predicted genotypes, as this is reflected by the mean correlation between imputed and true genotypes of 0.82 to 0.91 [181]. When this error rate is applied to the scale of several million variants, the size of inaccurately imputed genotypes may be substantial. Thirdly, we did not adjust retinopathy for the duration of diabetes, as negative durations were observed (Figure 2-3), despite of cross-validation of diabetes diagnosis dates with laboratory biochemistry records. Thus, the fidelity of dates of retinopathy screening in the database may be open to questioning, which however, was recognised as a drawback of an observational study in contrast to a prospective cohort, as the retinal screening was initiated in the early 1990s when the recording, entry, and annotation of the vast amount of retinopathy data was completely manual. In our data, we observed retinopathy cases declining over the durations of diabetes since the time of diabetes diagnosis, whilst the number of retinopathy controls peaked at approximately 5 years of duration of diabetes. This was likely reflective of possible imprecision of diabetes diagnosis dates in the GoDARTS database, and a more probable justification is that patients with diabetes may be undiagnosed for years till the manifestation of secondary complications, such as retinopathy. Finally, population outliers were identified from multidimensional scaling of identity by descent (IBD) and subsequently removed. Thus, within the genetically unrelated sample, there was no further population stratification such as genetic relatedness to be accounted for, whereas in the genetically mixed sample, linear mixed model used in the analysis captured genetic relatedness simultaneously [121]. Thus, further adjustment by principal components was not imperative as exemplified by previous GWAS publications [182].

We estimated statistical power of each additive test of markers in the genome-wide meta-analysis of diabetic retinopathy. This data was suggestive of a rapid rise in power for detecting genome-wide significant SNPs with an increase of the relative risk from 1.1 to 1.3. This trend was supported by a previous study (Figure 5-6) [87]. Additionally, the study power was 0.973 or greater for SNPs with MAF of 0.05 or greater, which was supportive of the proposition that diabetic retinopathy is a polygenic phenotype with each genetic variant conferring a small risk. These rare variants may remain undetected in this meta-analysis. Nevertheless, power estimation assumed the same rate of diabetic retinopathy prevalence across study cohorts, and the actual power of this meta-analysis may be influenced by the prevalence rate within each cohort.

In summary, this study utilised the whole-genome meta-analysis approach to investigate genetic susceptibility of diabetic retinopathy. We observed strong genetic associations with diabetic retinopathy across the genome, but none have achieved the stringent genome wide significance. A number of these association regions were confirmed by smaller-sized association studies of diabetic retinopathy. While the pathophysiology of diabetic retinopathy remains elusive, genes in the pathways of carbohydrate/lipid metabolism, the functional activity in the neuronal and glial cells, and transcription regulation are implicated in the development of diabetic retinopathy. We expect the addition of further cohorts to the genome-wide meta-analysis from the CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology) and East Asia Eye consortiums where we hope to achieve over 10,000 cases versus 10,000 controls. This is likely to improve the statistical power for detecting variants with moderate effect sizes, and thus *P* values of diabetic retinopathy association results are more likely to attain the genome-wide significance.

Table 5-7 Genetic variants of candidate genes for diabetic retinopathy (DR) in the European heritage population reported in a systematic meta-analysis [183]**. INDEL: insertion-deletion polymorphisms; PDR: proliferative DR.**

Gene	Variant	Risk allele	DR definition	Type of diabetes	*P*	Reference
ACE	INDEL at intron 16	287 bp deletion	No DR vs. any DR	Type 1	0.54	[184–189]
				Type 2	0.92	[186,190–192]
				Total	0.72	
			No DR vs. PDR	Type 1	0.44	[184,185,187,188]
				Type 2	0.56	[192]
				Total	0.30	
			NPDR vs. PDR	Type 1	3.0×10^{-4}	[187]
				Type 2	0.19	[192,193]
				Total	0.05	
NOS3	rs3138808	393 bp insertion	No DR vs. any DR	Type 1	3.6×10^{-3}	[194]
				Type 2	0.62	[195]
				Total	0.45	
VEGF	rs2010963	G	No DR vs. any DR	Type 2	0.16	[196,197]
AKR1B1	rs759853	T	No DR vs. any DR	Type 1	1.0×10^{-4}	[198,199]
				Type 2	0.68	[200]
				Total	0.21	
AKR1B1	microsatellite	z	No DR vs. any DR	Type 1	0.38	[198,201,202]
				Type 2	0.05	[203]
				Total	0.81	

	z	No DR vs. PDR	Type 1	0.20	[202]
			Type 2	0.05	[203]
			Total	0.79	
	z-2	No DR vs. any DR	Type 1	0.12	[198,201,202]
			Type 2	9.1×10^{-3}	[203]
			Total	0.03	
	z-2	No DR vs. PDR	Type 1	0.80	[202]
			Type 2	9.1×10^{-3}	[203]
			Total	0.42	
	z+2	No DR vs. any DR	Type 1	0.07	[202]
			Type 2	0.30	[203]
			Total	0.04	
	z+2	No DR vs. PDR	Type 1	0.88	[198,201,202]
			Type 2	0.30	[203]
			Total	0.46	

Table 5-8 Human chromosome regions showing evidence of linkage to diabetic retinopathy (DR) (logarithm of odds, LOD > 1) [204]**. NPDR/PDR: non-proliferative/proliferative DR.**

Locus	Closest microsatellite marker	Diabetic retinopathy definition	Population	LOD score	Reference
1p36	D1S3669	Retinopathy score in worst eye	Pima Indians	3.10	[153]
1p36	GGAT2A07	Any DR	Mexican Americans	1.24	[205]
2q37	AFM112yd4	Severe NPDR/PDR	Mexican Americans	1.11	[205]
3p26	GATA22G12	Severe NPDR/PDR	Mexican Americans	1.29	[205]
3q12	GATA68D03	Severe NPDR/PDR	Mexican Americans	1.40	[205]
3q12	GATA68D03	Any DR	Mexican Americans	2.41	[205]
3q26	D3S3053, D3S2427	Presence of at least one microaneurysm, haemorrhage or PDR	Pima Indians	1.36	[206]
7p15	GATA41G07	Any DR	Mexican Americans	1.02	[205]
9q21	D9S1120, D9S910	Presence of at least one microaneurysm, haemorrhage or PDR	Pima Indians	1.46	[206]
12p13	GATA49D12	Any DR	Mexican Americans	2.47	[205]
12q23	GATA85A04	Severe NPDR/PDR	Mexican Americans	1.03	[205]
15q25	ATA28G05	Any DR	Mexican Americans	1.07	[205]
15q26	GATA22F01	Any DR	Mexican Americans	1.16	[205]

Table 5-9 Diabetic retinopathy (DR) associated SNPs reported in GWAS to date. DR characterisation in references [81]**,** [82]**,** [83] **and [28] is moderate-to-severe non-proliferative and proliferative DR (PDR) in type 2 diabetic patients, non-proliferative and PDR in type 2 diabetic patients, diabetic macular edema and PDR in type 1 diabetic patients, and PDR in type 2 diabetic patients, respectively.**

Locus	SNP	Population	Reference allele	OR (95% CI)	*P*	Closest Gene	Function	Reference
6p12.1	rs6909083	Mexican Americans	--	--	1.80×10^{-5}	*TINAG*	Nephrogenesis	[81]
6q22.31	rs17083119	Mexican Americans	--	--	2.76×10^{-5}	*C6orf179*	Unknown	[81]
1q23	rs1033465	Mexican Americans	--	--	4.50×10^{-5}	*TNFSF18*	T lymphocyte survival in peripheral tissues	[81]
1q13	rs11583330	Mexican Americans	--	--	5.35×10^{-5}	*GNAI3*	Transmembrane signaling pathways	[81]
1q42.11	rs3014267	Mexican Americans	--	--	6.58×10^{-5}	*CDC42BPA*	Peripheral actin formation and cytoskeletal reorganization	[81]
15q13.3	rs11635920	Mexican Americans	--	--	7.18×10^{-5}	*FMN1*	Unknown	[81]
2q34	rs6726798	Mexican Americans	--	--	8.66×10^{-5}	*FN1*	Cell adhesion and migration processes	[81]
10q21.1	rs11812882	Mexican	--	--	8.85×10^{-5}	*CISD1*	Regulation of oxidation	[81]

Locus	SNP	Population	Reference allele	OR (95% CI)	*P*	Closest Gene	Function	Reference
		Americans						
2q35	rs1106412	Mexican Americans	--	--	8.9×10^{-5}	*USP37*	Ubiquitin specific hydrolysis	[81]
3p24.2	rs11927173	Mexican Americans	--	--	9.4×10^{-5}	*UBE2E2*	Ubiquitin conjugation	[81]
8q22.3	rs3098241	Mexican Americans	--	--	9.7×10^{-5}	*DCAF13*	Signal transduction, pre-mRNA processing and cytoskeleton assembly	[81]
1p32.1	rs2811893	Taiwanese	T	1.50 (1.03:2.20)	3.1×10^{-7}	*MYSM1*	Transcription regulation	[82]
4q32.1	rs4470583	Taiwanese	A	1.16 (0.70:1..92)	4.3×10^{-7}	*RPS14P7*	Ribosomal protein	[82]
5q15	rs13163610	Taiwanese	A	3.59 (1.36:9.47)	3.2×10^{-16}	*KIAA0825*	Unknown	[82]
10p12.31	rs12219125	Taiwanese	T	1.62 (1.02:2.58)	9.3×10^{-9}	*PLXDC2*	Unknown	[82]
10q11.22	rs4838605	Taiwanese	C	1.58 (1.00:2.52)	1.9×10^{-9}	*ARHGAP22*	Regulation of cell motility	[82]
10q21.1	rs4462262	Taiwanese	C	1.54 (0.79:2.99)	9.2×10^{-9}	*IPMK*	Nuclear mRNA export	[82]
13q32.1	rs2038823	Taiwanese	C	2.33 (1.13:4.77)	4.7×10^{-11}	*HS6ST3*	Proliferation, differentiation, adhesion, migration, inflammation and blood coagulation	[82]

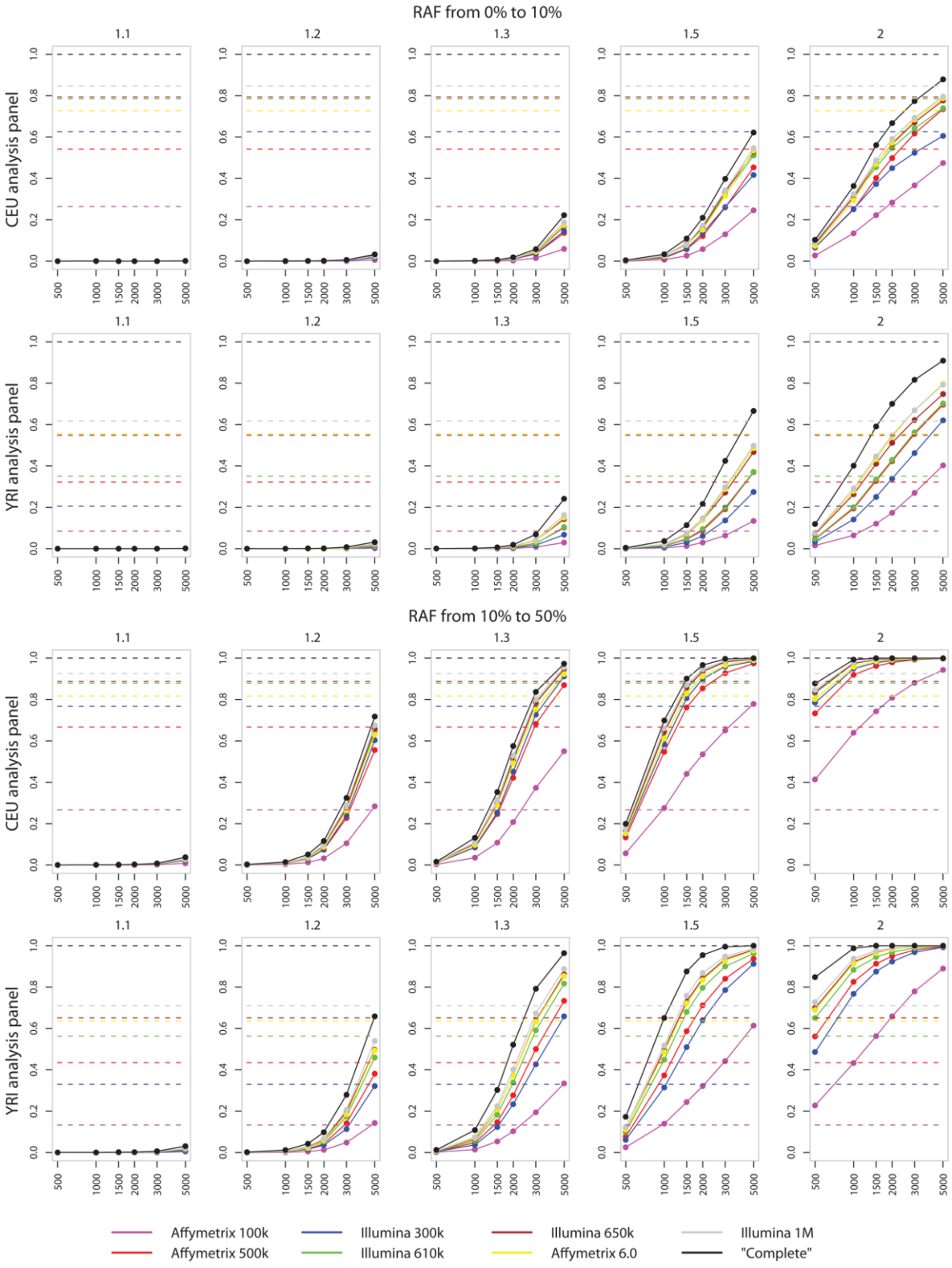
Locus	SNP	Population	Reference allele	OR (95% CI)	*P*	Closest Gene	Function	Reference
6q27	rs227455	Caucasian	C	0.53	1.6×10^{-7}	*C6orf118*	Unknown	[83]
16p11.2	rs151320	Caucasian	A	0.58	3.1×10^{-6}	*CCDC101*	Regulation of transcription	[83]
17p11.2	rs11871508	Caucasian	A	0.27	4.8×10^{-6}	*TVP23B*	Unknown	[83]
13q14.11	rs238252	Caucasian	C	1.41	6.3×10^{-6}	*AKAP11*	Cell cycle control	[83]
16p12.1	rs11074904	Caucasian	C	0.58	7.8×10^{-6}	*SULT1A1*	Sulfate conjugation of hormones, neurotransmitters, drugs, and xenobiotic compounds	[83]
10p12.31	rs17670074	Caucasian	A	1.58	8.1×10^{-6}	*PLXDC2*	Unknown	[83]
13q22.2	rs9565164	Taiwanese	C	1.7	4.4×10^{-7}	*TBC1D4*	Unknown	[28]
2q31.1	rs1399634	Taiwanese	A	1.5	4.2×10^{-6}	*LRP2*	Cell signaling	[28]
2q37.1	rs2380261	Taiwanese	T	1.5	4.7×10^{-6}	*ARL4C*	Cholesterol transport	[28]

The study by Fu *et al.* [81]: Reference allele and odds ratio were not represented for these imputed SNPs.

The study by Huang *et al*. [82]: Genotypes were dominantly coded, in contrast to the additive coding of genotypes in other referenced studies.

The studies by Grassi *et al*. [83] and Sheu *et al*. [28]: 95% confidence intervals of odds ratio were not reported.

Figure 5-6 Power (solid lines) and genome coverage (dotted lines) assuming increasing relative risks (above each plot) with a range of sample sizes and frequencies of relative risk allele (RAF) in the CEU and YRI Hapmap population samples, adapted from [87].


RAF from 0% to 10%
RAF from 10% to 50%
1.1
1.2
1.3
1.5
2
CEU analysis panel
YRI analysis panel
Affymetrix 100k
Affymetrix 500k
Illumina 300k
Illumina 610k
Illumina 650k
Affymetrix 6.0
Illumina 1M
"Complete"

Chapter 6

Concluding remarks

Deciphering the pathophysiology of diabetic retinopathy has been a long endeavour for research scientists. During the past two decades, the establishment and the growth of densely phenotyped diabetic retinopathy cohorts exemplified by GoDARTS raised the opportunity for studying common clinical risk factors for retinopathy at a population-wide scale [10,11,15,17,20]. Much of the aetiology of diabetic retinopathy remains elusive. Thus, the main objective of our research was to study clinical and genetic determinants of diabetic retinopathy in the GoDARTS cohort. Retinopathy risk factors adjusted in the current studies for this thesis did not form the exhaustive list, but rather the most common life style factors towards the development of retinopathy in the diabetic population. It is possible that other confounding factors with moderate risk effect may exist, and yet to be further studied. For example, pregnancy was one of the contributing factors towards the exacerbation of diabetic retinopathy in 10% of cases [218].

First, we utilised the multi-state Markov model for studying the influence of common population risk factors on the progression and remission of diabetic retinopathy by capturing longitudinal retinal events in 4,758 GoDARTS patients since the diagnosis of diabetes for up to 16 years' follow up. As this statistical model was computational intensive for each single analysis, the application of this methodology to a genome-wide scale became futile. In the later study, we used one-time measures of diabetic retinopathy to investigate heritability and genetic correlation for diabetic retinopathy and related risk factors explained by common genome-wide SNPs. The estimated narrow-sense heritability for diabetic retinopathy captured by whole-genome SNP data was supportive of common genetic polymorphisms underpinning diabetic retinopathy. This provided a strong basis for the genome-wide meta-analysis of diabetic retinopathy with the SUMMIT collaborators (EURODIAB, FinnDiane and SDR cohorts).

In the genetic study of retinopathy, we pooled samples from type 1 and type 2 diabetic populations, as we aimed to attain greater study power to detect moderate genetic effects. Despite that the clinical profiles of type 1 and type 2 diabetes may differ considerably, retinal vascular dysfunctions, such as haemorrhages, occlusion and retinal neovascularisation was previously shown to be common in both types of diabetes [207,208]. Akin to the previous studies [207,208], the sample pooling approach we adopted here for studying retinopathy was considered unbiased and sensitive in detecting retinal abnormalities. In the following sections, we will discuss the main conclusions for these studies.

6.1 The influence of population risk factors on diabetic retinopathy severity over the duration of diabetes

As described in Chapter 3, in the GoDARTS, the severity of diabetic retinopathy of each patient was captured by a series of follow-up observations for the duration of diabetes. This enabled us to delineate the rates of transitions across retinopathy states dependent upon the influence of common population risk factors. The key finding was that glycaemic exposure and blood pressure were strongly associated with the progression and remission of diabetic retinopathy. This discovery was consistent with previously reported diabetic retinopathy studies in which case-control or time-to-event phenotypes were analysed [21,142–151,155]. In this study, we found that with the adjustment of glycaemic exposure and blood pressure, the evidence for independent associations with transition rates across retinopathy states was insignificant for cholesterol, triglycerides and non-HDL-c. Our data suggested that the evidence for BMI, serum creatinine, HDL-c, sex and smoking influencing diabetic retinopathy progression and regression may be weak in the single covariate analyses after the adjustment of multiple testing.

There are a number of conflicting reports of the risk effect of cholesterol, triglycerides, non-HDL-c, BMI, serum creatinine, HDL-c, sex and smoking status [209–211]. We suggest that these disagreements may have arisen from sources including covariates adjusted in these analyses, the units of these covariates (direct measures or transformed scales), the sample sizes of these studies, the ethnicity of the study sample, and the study definition of retinopathy (case-control, ordinal categorical phenotypes, time-to-retinopathy, or time serial data of retinopathy as in this study). The analytical approach we presented

here is permissible in the future research to study the stage-by-stage development of diabetic retinopathy influenced by therapeutic medicine and genetic variants. Further work is required to optimise the computational performance of the methodology in order for its application to genome-wide association analysis of millions of genetic variants. Nevertheless, our results are valuable for clinical prognosis and treatment of diabetic retinopathy. Indeed the Scottish Diabetes Research Group is now investigating this methodology to optimise retinal screening intervals at the population level. Importantly, the long duration of diabetes for patients in the states of no-retinopathy and mild background retinopathy due to the tight regulation of blood pressure and glycaemic exposure provides the evidence for a stratified approach to personalise screening intervals. This may have a great impact on increasing the efficacy and decreasing the cost of the national screening programme.

6.2 Heritability and genetic correlations explained by common SNPs for diabetic retinopathy and related risk factors

The study presented in Chapter 4 was motivated by the observation that a large proportion of narrow-sense heritability and genetic correlations of metabolic traits in a non-diabetic population sample were captured by common genome-wide SNPs [108]. However, the methodology (the restricted maximum likelihood estimation) was only applicable to traits quantitative in nature, and it is computationally intractable for the analysis of ordinal traits such as diabetic retinopathy. Therefore, a Markov chain Monte Carlo (MCMC) sampling

algorithm was the only feasible method for a study of this kind. Dr Minghui Wang, Birmingham University, developed the software for running an efficient Gibbs sampler (an MCMC algorithm), and we applied it for the estimation of narrow-sense heritability and genetic correlation for diabetic retinopathy and related risk factors with the adjustment for age, sex and duration of diabetes in the GoDARTS Affymetrix 6 genotyped samples.

In this study data, we found that up to 34% phenotypic variations of diabetic retinopathy were accounted for by total additive genetic variations. In the analysis of genetic and residual correlations, the wide credible intervals of our estimates limited our ability to make more meaningful inference. However, at the 5% significance level, retinopathy with cholesterol showed positive residual correlation in the genetically unrelated sample. Our data confirmed the negative genetic and residual correlation between triglycerides and HDL-c [108]. Triglycerides also showed residual correlation with total cholesterol and glycaemic exposure.

We expect to discover more genetically correlated traits with diabetic retinopathy based on genome-wide SNP data with a larger sample. The main constraint with this plan lies within the slow convergence with the MCMC method, which took days and weeks to complete the analytical process at the current sample size. However, the utility of this methodology was vital, as the alternative, time-saving method remains applicable to pairs of numerical or binary traits [108]. In addition, for this methodology, it required an unbiased population sample, which is reflective of retinopathy prevalence in the population. Thus, due to this limitation, we did not include the GoDARTS Illumina genotyped samples that were selected for retinopathy. Future studies may be motivated to investigate an approach for

correcting ascertainment bias for ordinal traits in a non-random sample. The methodology we utilised here is also applicable to family data, where genome-wide variant data is unavailable, but trait phenotypes and familial relationship are known. In this case, the results of heritability and genetic correlations are accounted for by all genetic effects including additive, dominant, epistatic, maternal and paternal effects [63]. Future studies may also be interested in investigating pleiotropic effects of genome-wide variants [165]. Overall, this study may be the cornerstone of future studies of correlated clinical phenotypes.

6.3 The SUMMIT genome-wide meta-analysis of diabetic retinopathy

In Chapter 5, we presented the largest whole-genome meta-analysis of diabetic retinopathy to date, which included 9,724 retinopathy patients with either type 1 or type 2 diabetes. In this study, we identified top ranked association SNPs after the adjustment for age, sex and glycaemic exposure for diabetic retinopathy, and rs10746970 at 9q13-q21.2 was the strongest association signal (*P* value of 2.2×10^{-7}). The association regions of 1q43, 2q32.1, 6q22-27, and 13q14-32 were confirmed by previously published smaller-scaled association studies of diabetic retinopathy [28,81–83]. The closest genes are implicated in carbohydrate/lipid metabolism, functional activity in the neuronal and glial cells, and transcription regulation. These findings support the argument that a large number of genetic variants with moderate effects contribute to the development of diabetic retinopathy [212].

It is possible that SNPs with minor allele frequencies < 1%, un-genotyped loci with weak linkage disequilibrium with the genotyped/imputed SNPs, and copy number variants that have not been investigated in this study are also accountable for the predisposition of diabetic retinopathy [213]. It is increasingly likely that these challenges are solvable through the whole-genome sequencing technologies [214,215]. The plan for the next phase of this project is to seek the locus-based and whole-genome meta-analysis replication of diabetic retinopathy with the CHARGE (Cohorts for Heart and Aging Research in Genetic Epidemiology) [216] and East Asia Eye consortiums. With an increased sample size, it is hopeful that the top ranked association SNPs may be replicated to attain higher statistical significance. As the GoDARTS study is a source of high quality phenotype and genotype data of diabetic retinopathy, there may also an opportunity to integrate metabolomics and gene expression, when available, to investigate the complex system behaviour of diabetic retinopathy pathophysiology [217]. In conclusion, this study is the latest and the largest effort in mapping diabetic retinopathy susceptibility loci in the human genome. The follow-up work may shed additional light on the pathological mechanism of diabetic retinopathy.

6.4 Future studies

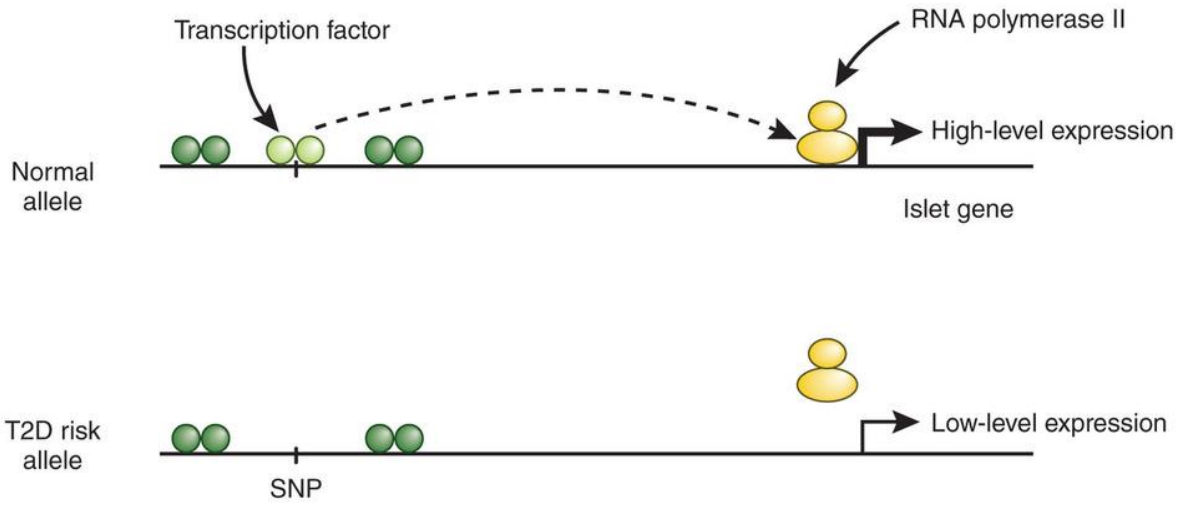
This series of studies provided the stepping stone for future studies. For instance, research interest may arise from the study of the extreme phenotypes that may be retinopathy cases with good glycaemic control or progressed to retinopathy shortly after diabetes diagnosis. Wherever the dates of diabetes diagnosis were accurate, we would encourage the use of diabetes duration in the analysis. Despite of the unparalleled research efforts in the susceptibility of diabetic retinopathy, much of the genetic architecture linking

polymorphisms with cellular functions remains an under-explored area. Newer research motivation lies within the paradigm that genetic polymorphisms mediate functional impacts through differential gene transcription and mRNA translation. For example, the study by Pasquali *et al*. [219] formed the strong evidence that genetic polymorphisms of type 2 diabetes are frequently mapped to non-coding regions of genes, and the observed risk alterations were attributable to the susceptibility loci differentially interacting with transcription factors regulating gene transcription (Figure 6-1). Similarly, the regulatory link between genetic polymorphisms and gene expression has been shown in other clinical phenotypes [220].

In 2011, the study by Bixler *et al*. [221] showed that expression alterations of 5% genes in the retina under the diabetic state were not reversible by the insulin treatment in rats, implicating dysregulated gene expressions underlying the pathophysiology of diabetic retinopathy. In the following year, Kandpal *et al*. [222] identified transcriptomic signatures for diabetic retinopathy in mice and these differentially expressed genes were associated with a diverse range of cellular functions including inflammation, microvasculature and glucose metabolism. These observed expression changes were inhibited by attenuating the activity of RAGE or p38 MAP kinase [222]. In the same year, the study by Brondani *et al*. [223] became the first study to date investigating genetic polymorphisms (within the gene of uncoupling protein 1) associated with diabetic retinopathy through the correlation with differentially expressed genes (uncoupling protein 1 and superoxide dismutase-2) in human retina. However, few genetic polymorphisms and retina-expressed genes were examined in this study [223]. This offers the unprecedented potential for leveraging genome-wide polymorphism data in studying regulated gene expressions underlying the pathophysiology

of diabetic retinopathy. A study of this kind may demonstrate regulatome influenced by genetic polymorphisms associated with diabetic retinopathy, and also highlight possible interaction of genetic variants in the regulatory network. This may provide a detailed functional annotation of genetic polymorphisms underlying the susceptibility of diabetic retinopathy.

Figure 6-1 Alleles haboured at an enhancer cluster differentially regulate gene expression in the pancreatic islet [219]. In the pancreatic islet, the wild-type allele enhances the binding of transcription factor to the enhancer, and subsequently induces high-level transcription (the top panel). The type 2 diabetes risk allele inhibits the binding of transcription factor to the enhancer, leading to reduced gene expression in the pancreatic islet (to bottom panel). This was associated with pancreatic dysfunction.



6.5 Publications

The longitudinal analysis of diabetic retinopathy described in Chapter 3 was published [224]. We also participated in the replication of diabetic retinopathy risk loci for the Candidate gene Association Resource (CARe) study [225].

References

1. Williams R, Airey M, Baxter H, Forrester J, Kennedy-Martin T, et al. (2004) Epidemiology of diabetic retinopathy and macular oedema: a systematic review. Eye 18: 963–983. doi:10.1038/sj.eye.6701476.

2. Fowler MJ (2008) Microvascular and Macrovascular Complications of Diabetes. Clin Diabetes 26: 77–82. doi:10.2337/diaclin.26.2.77.

3. Taylor R, Batey D (2012) Handbook of Retinal Screening in Diabetes: Diagnosis and Management. John Wiley & Sons. 186 p.

4. Ockrim Z, Yorston D (2010) Managing diabetic retinopathy. BMJ 341: c5400–c5400. doi:10.1136/bmj.c5400.

5. Zander E, Herfurth S, Bohl B, Heinke P, Herrmann U, et al. (2000) Maculopathy in patients with diabetes mellitus type 1 and type 2: associations with risk factors. Br J Ophthalmol 84: 871–876. doi:10.1136/bjo.84.8.871.

6. Urry LA, Cain ML, Wasserman SA, Minorsky PV (2011) Campbell Biology. Pearson Benjamin Cummings. 1263 p.

7. Records RE (1979) Physiology of the human eye and visual system. Harper & Row. 716 p.

8. The Organisaiton of Retina (n.d.). Available: http://droualb.faculty.mjc.edu/Lecture%20Notes/Unit%205/retina_2.jpg. Accessed 25 June 2013.

9. Diabetic Retinopathy Screening Services in Scotland: A Training Handbook - July 2003 (n.d.). Available: http://www.ndrs.scot.nhs.uk/Train/Handbook/drh-19.htm. Accessed 30 May 2013.

10. Leese GP, Broadbent DM, Harding SP, Vora JP (1996) Detection of sight-threatening diabetic eye disease. Diabet Med J Br Diabet Assoc 13: 850–853. doi:10.1002/(SICI)1096-9136(199610)13:10<850::AID-DIA167>3.0.CO;2-E.

11. Leese GP, Tesfaye S, Dengler-Harles M, Laws F, Clark DI, et al. (1997) Screening for diabetic eye disease by optometrists using slit lamps. J R Coll Physicians Lond 31: 65–69.

12. Williams R, Nussey S, Humphry R, Thompson G (1986) Assessment of non-mydriatic fundus photography in detection of diabetic retinopathy. Br Med J Clin Res Ed 293: 1140–1142.

13. Taylor R, Lovelock L, Tunbridge WM, Alberti KG, Brackenridge RG, et al. (1990) Comparison of non-mydriatic retinal photography with ophthalmoscopy in 2159 patients: mobile retinal camera study. BMJ 301: 1243–1247.

14. Nathan DM, Fogel HA, Godine JE, Lou PL, D'Amico DJ, et al. (1991) Role of diabetologist in evaluating diabetic retinopathy. Diabetes Care 14: 26–33.

15. Leese GP, Newton RW, Jung RT, Haining W, Ellingford A (1992) Screening for diabetic retinopathy in a widely spaced population using non-mydriatic fundus photography in a mobile unit. Tayside Mobile Eye Screening Unit. Diabet Med J Br Diabet Assoc 9: 459–462.

16. Leese GP, Ellingford A, Morris AD, Ellis JD, Cunningham S (2003) Screening Using Compressed Digital Retinal Images Successfully Identifies Retinopathy. Diabetes Care 26: 247–247. doi:10.2337/diacare.26.1.247.

17. Leese G, Morris A, Olson J (2003) A national retinal screening programme for diabetes in Scotland. Diabet Med 20: 962–964.

18. Sivaprasad S, Gupta B, Crosby-Nwaobi R, Evans J (2012) Prevalence of Diabetic Retinopathy in Various Ethnic Groups: A Worldwide Perspective. Surv Ophthalmol 57: 347–370. doi:10.1016/j.survophthal.2012.01.004.

19. Fong DS, Aiello L, Gardner TW, King GL, Blankenship G, et al. (2004) Retinopathy in Diabetes. Diabetes Care 27: s84–s87. doi:10.2337/diacare.27.2007.S84.

20. Leese GP, Ahmed S, Newton RW, Jung RT, Ellingford A, et al. (1993) Use of mobile screening unit for diabetic retinopathy in rural and urban areas. BMJ 306: 187–189.

21. DCCT (1995) The relationship of glycemic exposure (HbA1c) to the risk of development and progression of retinopathy in the diabetes control and complications trial. Diabetes 44: 968–983.

22. Ishihara M, Yukimura Y, Aizawa T, Yamada T, Ohto K, et al. (1987) High Blood Pressure as Risk Factor in Diabetic Retinopathy Development in NIDDM Patients. Diabetes Care 10: 20–25. doi:10.2337/diacare.10.1.20.

23. Van Leiden HA, Dekker JM, Moll AC, Nijpels G, Heine RJ, et al. (2002) Blood pressure, lipids, and obesity are associated with retinopathy: the hoorn study. Diabetes Care 25: 1320–1325.

24. Best RM, Chakravarthy U (1997) Diabetic retinopathy in pregnancy. Br J Ophthalmol 81: 249–251. doi:10.1136/bjo.81.3.249.

25. Moss SE, Klein R, Klein BE (1991) Association of cigarette smoking with diabetic retinopathy. Diabetes Care 14: 119–126.

26. Home P (2008) Safety of very tight blood glucose control in type 2 diabetes. BMJ 336: 458–459. doi:10.1136/bmj.39499.514410.80.

27. Conditions NCGC for A and C, NICE (n.d.) Type 2 diabetes (partially updated by CG87). NICE. Available: http://www.nice.org.uk/. Accessed 28 June 2013.

28. Sheu WH-H, Kuo JZ, Lee I-T, Hung Y-J, Lee W-J, et al. (2013) Genome-wide association study in a Chinese population with diabetic retinopathy. Hum Mol Genet. doi:10.1093/hmg/ddt161.

29. Hirsch IB, Brownlee M (2010) Beyond hemoglobin A1c--need for additional markers of risk for diabetic microvascular complications. JAMA J Am Med Assoc 303: 2291–2292. doi:10.1001/jama.2010.785.

30. Lachin JM, Genuth S, Nathan DM, Zinman B, Rutledge BN (2008) Effect of Glycemic Exposure on the Risk of Microvascular Complications in the Diabetes Control and Complications Trial—Revisited. Diabetes 57: 995–1001. doi:10.2337/db07-1618.

31. Antonetti DA, Klein R, Gardner TW (2012) Diabetic retinopathy. N Engl J Med 366: 1227–1239. doi:10.1056/NEJMra1005073.

32. Leslie RD, Pyke DA (1982) Diabetic retinopathy in identical twins. Diabetes 31: 19–21.

33. Pathophysiology of Diabetic Retinopathy - Springer (n.d.). Available: http://link.springer.com.libproxy.dundee.ac.uk/chapter/10.1007/978-0-387-85900-2_1/fulltext.html. Accessed 29 May 2013.

34. Wilcox G (2005) Insulin and Insulin Resistance. Clin Biochem Rev 26: 19–39.

35. Tarr JM, Kaul K, Chopra M, Kohner EM, Chibber R (2013) Pathophysiology of Diabetic Retinopathy. ISRN Ophthalmol 2013: 1–13. doi:10.1155/2013/343560.

36. Hammes HP, Martin S, Federlin K, Geisen K, Brownlee M (1991) Aminoguanidine treatment inhibits the development of experimental diabetic retinopathy. Proc Natl Acad Sci U S A 88: 11555–11558.

37. Portilla D, Dai G, Peters JM, Gonzalez FJ, Crew MD, et al. (2000) Etomoxir-induced PPARalpha-modulated enzymes protect during acute renal failure. Am J Physiol Renal Physiol 278: F667–F675.

38. Keogh RJ, Dunlop ME, Larkins RG (1997) Effect of inhibition of aldose reductase on glucose flux, diacylglycerol formation, protein kinase C, and phospholipase A2 activation. Metabolism 46: 41–47.

39. Ishii H, Jirousek MR, Koya D, Takagi C, Xia P, et al. (1996) Amelioration of vascular dysfunctions in diabetic rats by an oral PKC beta inhibitor. Science 272: 728–731.

40. Williams B, Gallacher B, Patel H, Orme C (1997) Glucose-induced protein kinase C activation regulates vascular permeability factor mRNA expression and peptide production by human vascular smooth muscle cells in vitro. Diabetes 46: 1497–1503.

41. Kagan VE, Quinn PJ (2010) Coenzyme Q: Molecular Mechanisms in Health and Disease. CRC Press. 408 p.

42. Baynes JW (1991) Role of oxidative stress in development of complications in diabetes. Diabetes 40: 405–412.

43. Hartnett ME, Stratton RD, Browne RW, Rosner BA, Lanham RJ, et al. (2000) Serum markers of oxidative stress and severity of diabetic retinopathy. Diabetes Care 23: 234–240.

44. Brownlee M (2005) The Pathobiology of Diabetic Complications A Unifying Mechanism. Diabetes 54: 1615–1625. doi:10.2337/diabetes.54.6.1615.

45. Lodish H (2008) Molecular Cell Biology. W. H. Freeman. 1272 p.

46. Naim HY (1996) Secretion of human intestinal angiotensin-converting enzyme and its association with the differentiation state of intestinal cells. Biochem J 316: 259–264.

47. File:Renin-angiotensin-aldosterone system.png (n.d.). Wikipedia Free Encycl. Available: http://en.wikipedia.org/wiki/File:Renin-angiotensin-aldosterone_system.png. Accessed 5 June 2013.

48. Kumar A, Rassoli A, Raizada MK (1988) Angiotensinogen gene expression in neuronal and glial cells in primary cultures of rat brain. J Neurosci Res 19: 287–290. doi:10.1002/jnr.490190302.

49. Balcells E, Meng QC, Hageman GR, Palmer RW, Durand JN, et al. (1996) Angiotensin II formation in dog heart is mediated by different pathways in vivo and in vitro. Am J Physiol 271: H417–H421.

50. Navar LG, Imig JD, Zou L, Wang CT (1997) Intrarenal production of angiotensin II. Semin Nephrol 17: 412–422.

51. Siragy HM (2004) AT1 and AT2 receptor in the kidney: role in health and disease. Semin Nephrol 24: 93–100.

52. Rong P, Wilkinson-Berka JL, Skinner SL (2001) Control of renin secretion from adrenal gland in transgenic Ren-2 and normal rats. Mol Cell Endocrinol 173: 203–212.

53. Otani A, Takagi H, Suzuma K, Honda Y (1998) Angiotensin II potentiates vascular endothelial growth factor-induced angiogenic activity in retinal microcapillary endothelial cells. Circ Res 82: 619–628.

54. Nagai N, Noda K, Urano T, Kubota Y, Shinoda H, et al. (2005) Selective suppression of pathologic, but not physiologic, retinal neovascularization by blocking the angiotensin II type 1 receptor. Invest Ophthalmol Vis Sci 46: 1078–1084. doi:10.1167/iovs.04-1101.

55. Lopes de Faria JB, Silva KC, Lopes de Faria JM (2011) The contribution of hypertension to diabetic nephropathy and retinopathy: the role of inflammation and oxidative stress. Hypertens Res 34: 413–422. doi:10.1038/hr.2010.263.

56. Seko Y, Seko Y, Fujikura H, Pang J, Tokoro T, et al. (1999) Induction of vascular endothelial growth factor after application of mechanical stress to retinal pigment epithelium of the rat in vitro. Invest Ophthalmol Vis Sci 40: 3287–3291.

57. Suzuma I, Hata Y, Clermont A, Pokras F, Rook SL, et al. (2001) Cyclic stretch and hypertension induce retinal expression of vascular endothelial growth factor and vascular endothelial growth factor receptor-2: potential mechanisms for exacerbation of diabetic retinopathy by hypertension. Diabetes 50: 444–454.

58. Suzuma I, Suzuma K, Ueki K, Hata Y, Feener EP, et al. (2002) Stretch-induced retinal vascular endothelial growth factor expression is mediated by phosphatidylinositol 3-kinase and protein kinase C (PKC)-zeta but not by stretch-induced ERK1/2, Akt, Ras, or classical/novel PKC pathways. J Biol Chem 277: 1047–1057. doi:10.1074/jbc.M105336200.

59. Suzuma I, Murakami T, Suzuma K, Kaneto H, Watanabe D, et al. (2007) Cyclic stretch-induced reactive oxygen species generation enhances apoptosis in retinal pericytes through c-jun NH2-terminal kinase activation. Hypertension 49: 347–354. doi:10.1161/01.HYP.0000253535.26659.2f.

60. Beltramo E, Berrone E, Giunti S, Gruden G, Perin PC, et al. (2006) Effects of mechanical stress and high glucose on pericyte proliferation, apoptosis and contractile phenotype. Exp Eye Res 83: 989–994. doi:10.1016/j.exer.2006.05.008.

61. Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet 4: e1000008. doi:10.1371/journal.pgen.1000008.

62. Lynch M, Walsh B (1998) Genetics and Analysis of Quantitative Traits. 1 edition. Sunderland, Mass: Sinauer Associates. 980 p.

63. Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. Essex, England: Longman.

64. Visscher PM, Hill WG, Wray NR (2008) Heritability in the genomics era — concepts and misconceptions. Nat Rev Genet 9: 255–266. doi:10.1038/nrg2322.

65. Agarwala V, Flannick J, Sunyaev S, Altshuler D (2013) Evaluating empirical bounds on complex disease genetic architecture. Nat Genet 45: 1418–1427. doi:10.1038/ng.2804.

66. Hyttinen V, Kaprio J, Kinnunen L, Koskenvuo M, Tuomilehto J (2003) Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs: a nationwide follow-up study. Diabetes 52: 1052–1055.

67. Poulsen P, Kyvik KO, Vaag A, Beck-Nielsen H (1999) Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. Diabetologia 42: 139–145.

68. Ali O (2013) Genetics of type 2 diabetes. World J Diabetes 4: 114–123. doi:10.4239/wjd.v4.i4.114.

69. Arar NH, Freedman BI, Adler SG, Iyengar SK, Chew EY, et al. (2008) Heritability of the severity of diabetic retinopathy: the FIND-Eye study. Invest Ophthalmol Vis Sci 49: 3839–3845. doi:10.1167/iovs.07-1633.

70. Hietala K, Forsblom C, Summanen P, Groop P-H, FinnDiane Study Group (2008) Heritability of proliferative diabetic retinopathy. Diabetes 57: 2176–2180. doi:10.2337/db07-1495.

71. Bakay M, Pandey R, Hakonarson H (2013) Genes involved in type 1 diabetes: an update. Genes 4: 499–521. doi:10.3390/genes4030499.

72. Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, et al. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661–678. doi:10.1038/nature05911.

73. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, et al. (2007) Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet 39: 857–864. doi:10.1038/ng2068.

74. Hakonarson H, Qu H-Q, Bradfield JP, Marchand L, Kim CE, et al. (2008) A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study. Diabetes 57: 1143–1146. doi:10.2337/db07-1305.

75. Johnson K, Wong R, Barriga KJ, Klingensmith G, Ziegler A-G, et al. (2012) rs11203203 is associated with type 1 diabetes risk in population pre-screened for high-risk HLA-DR,DQ genotypes. Pediatr Diabetes 13: 611–615. doi:10.1111/j.1399-5448.2012.00888.x.

76. DeFronzo RA (2004) Pathogenesis of type 2 diabetes mellitus. Med Clin North Am 88: 787–835, ix. doi:10.1016/j.mcna.2004.04.013.

77. Zeggini E (2007) A new era for Type 2 diabetes genetics. Diabet Med 24: 1181–1186. doi:10.1111/j.1464-5491.2007.02274.x.

78. Imamura M, Maeda S (2011) Genetics of type 2 diabetes: the GWAS era and future perspectives [Review]. Endocr J 58: 723–739.

79. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, et al. (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 42: 105–116. doi:10.1038/ng.520.

80. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, et al. (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 42: 579–589. doi:10.1038/ng.609.

81. Fu Y-P, Hallman DM, Gonzalez VH, Klein BEK, Klein R, et al. (2010) Identification of Diabetic Retinopathy Genes through a Genome-Wide Association Study among Mexican-Americans from Starr County, Texas. J Ophthalmol 2010. doi:10.1155/2010/861291.

82. Huang Y-C, Lin J-M, Lin H-J, Chen C-C, Chen S-Y, et al. (2011) Genome-wide association study of diabetic retinopathy in a Taiwanese population. Ophthalmology 118: 642–648. doi:10.1016/j.ophtha.2010.07.020.

83. Grassi MA, Tikhomirov A, Ramalingam S, Below JE, Cox NJ, et al. (2011) Genome-wide meta-analysis for severe diabetic retinopathy. Hum Mol Genet 20: 2472–2481. doi:10.1093/hmg/ddr121.

84. Morris AD, Boyle DIR, MacAlpine R, Emslie-Smith A, Jung RT, et al. (1997) The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. BMJ 315: 524–528.

85. Affymetrix Genome-Wide Human SNP Array 6.0 (n.d.). Affymetrix. Available: http://www.affymetrix.com/estore/browse/products.jsp?productId=131533#1_1. Accessed 15 July 2013.

86. Group TG and UDPS, 2 TWTCCC (2011) Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes. Nat Genet 43: 117–120. doi:10.1038/ng.735.

87. Spencer CCA, Su Z, Donnelly P, Marchini J (2009) Designing Genome-Wide Association Studies: Sample Size, Power, Imputation, and the Choice of Genotyping Chip. PLoS Genet 5: e1000477. doi:10.1371/journal.pgen.1000477.

88. Illumina HumanOmniExpress BeadChip (n.d.). Illumina. Available: http://investor.illumina.com/phoenix.zhtml?c=121127&p=irol-newsArticle&id=1373309. Accessed 15 July 2013.

89. Consortium T 1000 GP (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061–1073. doi:10.1038/nature09534.

90. Consortium T 1000 GP (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491: 56–65. doi:10.1038/nature11632.

91. Jackson CH (2011) Multi-state models for panel data: the msm package for R. J Stat Softw 38: 1–29.

92. Marshall G, Jones RH (1995) Multi-state models and diabetic retinopathy. Stat Med 14: 1975–1983.

93. Huelsenbeck JP, Hillis DM, Nielsen R (1996) A Likelihood-Ratio Test of Monophyly. Syst Biol 45: 546–558. doi:10.1093/sysbio/45.4.546.

94. Anderson DR (2008) Model Based Inference in the Life Sciences: A Primer on Evidence. Springer. 203 p.

95. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol 57: 289–300. doi:10.2307/2346101.

96. Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5: e1000529. doi:10.1371/journal.pgen.1000529.

97. Liu EY, Li M, Wang W, Li Y (2013) MaCH-admix: genotype imputation for admixed populations. Genet Epidemiol 37: 25–37. doi:10.1002/gepi.21690.

98. Li Y, Willer C, Sanna S, Abecasis G (2009) Genotype Imputation. Annu Rev Genomics Hum Genet 10: 387–406. doi:10.1146/annurev.genom.9.081307.164242.

99. Marchini J, Howie B (2010) Genotype imputation for genome-wide association studies. Nat Rev Genet 11: 499–511. doi:10.1038/nrg2796.

100. Sorensen DA, Andersen S, Gianola D, Korsgaard I (1995) Bayesian inference in threshold models using Gibbs sampling. Genet Sel Evol 27: 1–21. doi:10.1186/1297-9686-27-3-229.

101. Wang CS, Quaas RL, Pollak EJ (1997) Bayesian analysis of calving ease scores and birth weights. Genet Sel Evol 29: 1–27. doi:10.1186/1297-9686-29-2-117.

102. Fisher transformation (2013). Wikipedia Free Encycl. Available: http://en.wikipedia.org/w/index.php?title=Fisher_transformation&oldid=558763917. Accessed 20 August 2013.

103. Zhou X, Carbonetto P, Stephens M (2013) Polygenic Modeling with Bayesian Sparse Linear Mixed Models. PLoS Genet 9: e1003264. doi:10.1371/journal.pgen.1003264.

104. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42: 565–569. doi:10.1038/ng.608.

105. Weisstein EW (n.d.) Positive Definite Matrix -- from Wolfram MathWorld. Available: http://mathworld.wolfram.com/PositiveDefiniteMatrix.html. Accessed 23 July 2013.

106. Lee SH, DeCandia TR, Ripke S, Yang J, (pgc-Scz) TSPG-WASC, et al. (2012) Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nat Genet 44: 247–250. doi:10.1038/ng.1108.

107. Sorensen D, Gianola D (2002) Likelihood, Bayesian and MCMC Methods in Quantitative Genetics. Springer. 768 p.

108. Vattikuti S, Guo J, Chow CC (2012) Heritability and Genetic Correlations Explained by Common SNPs for Metabolic Syndrome Traits. PLoS Genet 8: e1002637. doi:10.1371/journal.pgen.1002637.

109. Geweke J (1992) Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. IN BAYESIAN STATISTICS. Oxford University Press. pp. 169–193.

110. Markov chain (2013). Wikipedia Free Encycl. Available: https://en.wikipedia.org/w/index.php?title=Markov_chain&oldid=566351706. Accessed 3 August 2013.

111. Brooks SP, Gelman A (1998) General Methods for Monitoring Convergence of Iterative Simulations. J Comput Graph Stat 7: 434. doi:10.2307/1390675.

112. Microvascular and acute complications in IDDM patients: the EURODIAB IDDM Complications Study (1994). Diabetologia 37: 278–285.

113. Ahluwalia TS, Lindholm E, Groop LC (2011) Common variants in CNDP1 and CNDP2, and risk of nephropathy in type 2 diabetes. Diabetologia 54: 2295–2302. doi:10.1007/s00125-011-2178-5.

114. Roglic G, Colhoun H m., Stevens L k., Lemkes H h., Manes C, et al. (1998) Parental history of hypertension and parental history of diabetes and microvascular complications in insulin-dependent diabetes mellitus: the EURODIAB IDDM complications study. Diabet Med 15: 418–426. doi:10.1002/(SICI)1096-9136(199805)15:5<418::AID-DIA604>3.0.CO;2-P.

115. Thorn LM, Forsblom C, Fagerudd J, Thomas MC, Pettersson-Fernholm K, et al. (2005) Metabolic syndrome in type 1 diabetes: association with diabetic nephropathy and glycemic control (the FinnDiane study). Diabetes Care 28: 2019–2024.

116. Lindholm E, Agardh E, Tuomi T, Groop L, Agardh C-D (2001) Classifying diabetes according to the new WHO clinical stages. Eur J Epidemiol 17: 983–989. doi:10.1023/A:1020036805655.

117. Delaneau O, Marchini J, Zagury J-F (2012) A linear complexity phasing method for thousands of genomes. Nat Methods 9: 179–181. doi:10.1038/nmeth.1785.

118. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 34: 816–834. doi:10.1002/gepi.20533.

119. Howie B, Marchini J, Stephens M (2011) Genotype Imputation with Thousands of Genomes. G3 Genes Genomes Genet 1: 457–470. doi:10.1534/g3.111.001198.

120. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S, et al. (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42: 348–354. doi:10.1038/ng.548.

121. Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. Nat Genet 44: 821–824. doi:10.1038/ng.2310.

122. Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26: 2190–2191. doi:10.1093/bioinformatics/btq340.

123. Mägi R, Morris AP (2010) GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics 11: 288. doi:10.1186/1471-2105-11-288.

124. Evangelou E, Ioannidis JPA (2013) Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet 14: 379–389. doi:10.1038/nrg3472.

125. Skol AD, Scott LJ, Abecasis GR, Boehnke M (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet 38: 209–213. doi:10.1038/ng1706.

126. Zhang J, Yu KF (1998) What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. JAMA J Am Med Assoc 280: 1690–1691.

127. He B-B, Wei L, Gu Y-J, Han J-F, Li M, et al. (2012) Factors Associated with Diabetic Retinopathy in Chinese Patients with Type 2 Diabetes Mellitus. Int J Endocrinol 2012: 157940. doi:10.1155/2012/157940.

128. Klein R, Klein BE, Moss SE, Davis MD, DeMets DL (1988) Glycosylated hemoglobin predicts the incidence and progression of diabetic retinopathy. JAMA J Am Med Assoc 260: 2864–2871.

129. Teuscher A, Schnell H, Wilson PW (1988) Incidence of diabetic retinopathy and relationship to baseline plasma glucose and blood pressure. Diabetes Care 11: 246–251.

130. Janka HU, Warram JH, Rand LI, Krolewski AS (1989) Risk factors for progression of background retinopathy in long-standing IDDM. Diabetes 38: 460–464.

131. Semeraro F, Parrinello G, Cancarini A, Pasquini L, Zarra E, et al. (2011) Predicting the risk of diabetic retinopathy in type 2 diabetic patients. J Diabetes Complications 25: 292–297. doi:10.1016/j.jdiacomp.2010.12.002.

132. Ahmed KR, Karim MN, Bhowmik B, Habib SH, Bukht MS, et al. (2012) Incidence of Diabetic Retinopathy in Bangladesh: (a 15 year follow-up study). J Diabetes. Available: http://www.ncbi.nlm.nih.gov/pubmed/22613259. Accessed 20 August 2012.

133. Jackson CH, Sharples LD, Thompson SG, Duffy SW, Couto E (2003) Multistate Markov models for disease progression with classification error. J R Stat Soc Ser Stat 52: 193–209. doi:10.1111/1467-9884.00351.

134. Kay R (1986) A Markov model for analysing cancer markers and disease states in survival studies. Biometrics 42: 855–865.

135. Andersen PK (1988) Multistate models in survival analysis: a study of nephropathy and mortality in diabetes. Stat Med 7: 661–670.

136. Duffy SW, Chen HH, Tabar L, Day NE (1995) Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase. Stat Med 14: 1531–1543.

137. Andersen PK, Hansen LS, Keiding N (1991) Assessing the influence of reversible disease indicators on survival. Stat Med 10: 1061–1067.

138. Klein R, Klein BE, Moss SE, Davis MD, DeMets DL (1989) The Wisconsin Epidemiologic Study of Diabetic Retinopathy. X. Four-year incidence and progression of diabetic retinopathy when age at diagnosis is 30 years or more. Arch Ophthalmol 107: 244–249.

139. Group UPDS (ukpds), Kohner EM, Stratton IM, Aldington SJ, Holman RR, et al. (2001) Relationship between the severity of retinopathy and progression to photocoagulation in patients with Type 2 diabetes mellitus in the UKPDS (UKPDS 52). Diabet Med 18: 178–184. doi:10.1046/j.1464-5491.2001.00458.x.

140. Stratton I, Kohner E, Aldington S, Turner R, Holman R, et al. (2001) UKPDS 50: risk factors for incidence and progression of retinopathy in Type II diabetes over 6 years from diagnosis. Diabetologia 44: 156–163.

141. Jones CD, Greenwood RH, Misra A, Bachmann MO (2012) Incidence and Progression of Diabetic Retinopathy During 17 Years of a Population-Based Screening Program in England. Diabetes Care 35: 592–596. doi:10.2337/dc11-0943.

142. Warram J, Scott L, Hanna L, Wantman M, Cohen S, et al. (2000) Progression of microalbuminuria to proteinuria in type 1 diabetes: nonlinear relationship with hyperglycemia. Diabetes 49: 94–100.

143. Stratton IM, Adler AI, Neil HAW, Matthews DR, Manley SE, et al. (2000) Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. BMJ 321: 405–412.

144. Gerstein H, Pogue J, Mann JFE, Lonn E, Dagenais G, et al. (2005) The relationship between dysglycaemia and cardiovascular and renal risk in diabetic and non-diabetic participants in the HOPE study: a prospective epidemiological analysis. Diabetologia 48: 1749–1755.

145. Klein R, Klein BEK, Moss SE, Cruickshanks KJ (1998) The Wisconsin epidemiologic study of diabetic retinopathy: XVII: The 14-year incidence and progression of diabetic retinopathy and associated risk factors in type 1 diabetes11Proprietary interest: none. Ophthalmology 105: 1801–1815.

146. Klein R, Klein BEK, Moss SE (1996) Relation of glycemic control to diabetic microvascular complications in diabetes mellitus. Ann Intern Med 124: 90–96.

147. Klein R, Klein BEK, Moss SE, Cruickshanks KJ (1994) Relationship of hyperglycemia to the long-term incidence and progression of diabetic retinopathy. Arch Intern Med 154: 2169.

148. Florkowski CM, Scott RS, Coope PA, Graham PJ, Moir CL (2001) Age at diagnosis, glycaemic control and the development of retinopathy in a population-based cohort of Type 1 diabetic subjects in Canterbury, New Zealand. Diabetes Res Clin Pract 52: 125–131.

149. Yoshida Y, Hagura R, Hara Y, Sugasawa G, Akanuma Y (2001) Risk factors for the development of diabetic retinopathy in Japanese type 2 diabetic patients. Diabetes Res Clin Pract 51: 195–203.

150. Nakagami T, Kawahara R, Hori S, Omori Y (1997) Glycemic control and prevention of retinopathy in Japanese NIDDM patients: a 10-year follow-up study. Diabetes Care 20: 621–622.

151. Arun C, Pandit R, Taylor R (2004) Long-term progression of retinopathy after initiation of insulin therapy in Type 2 diabetes: an observational study. Diabetologia 47: 1380–1384.

152. Klein R, Klein BEK, Moss SE, Davis MD, DeMets DL (1989) Is blood pressure a predictor of the incidence or progression of diabetic retinopathy? Arch Intern Med 149: 2427.

153. Looker HC, Nelson RG, Chew E, Klein R, Klein BEK, et al. (2007) Genome-wide linkage analyses to identify Loci for diabetic retinopathy. Diabetes 56: 1160–1166. doi:10.2337/db06-1299.

154. Lim LS, Tai ES, Mitchell P, Wang JJ, Tay WT, et al. (2010) C-reactive protein, body mass index, and diabetic retinopathy. Invest Ophthalmol Vis Sci 51: 4458–4463. doi:10.1167/iovs.09-4939.

155. Stratton I, Kohner E, Aldington S, Turner R, Holman R, et al. (2001) UKPDS 50: risk factors for incidence and progression of retinopathy in Type II diabetes over 6 years from diagnosis. Diabetologia 44: 156–163.

156. Busik JV, Esselman WJ, Reid GE (2012) Examining the role of lipid mediators in diabetic retinopathy. Clin Lipidol 7: 661–675. doi:10.2217/clp.12.68.

157. Lee SH, Wray NR, Goddard ME, Visscher PM (2011) Estimating Missing Heritability for Disease from Genome-wide Association Studies. Am J Hum Genet 88: 294–305. doi:10.1016/j.ajhg.2011.02.002.

158. Magnusson PKE, Rasmussen F (2002) Familial resemblance of body mass index and familial risk of high and low body mass index. A study of young men in Sweden. Int J Obes Relat Metab Disord J Int Assoc Study Obes 26: 1225–1231. doi:10.1038/sj.ijo.0802041.

159. Jermendy G, Horváth T, Littvay L, Steinbach R, Jermendy AL, et al. (2011) Effect of genetic and environmental influences on cardiometabolic risk factors: a twin study. Cardiovasc Diabetol 10: 96. doi:10.1186/1475-2840-10-96.

160. Soranzo N (2011) Genetic determinants of variability in glycated hemoglobin (HbA(1c)) in humans: review of recent progress and prospects for use in diabetes care. Curr Diab Rep 11: 562–569. doi:10.1007/s11892-011-0232-9.

161. Browning SR, Browning BL (2013) Identity-by-descent-based heritability analysis in the Northern Finland Birth Cohort. Hum Genet 132: 129–138. doi:10.1007/s00439-012-1230-y.

162. Zarkesh M, Daneshpour MS, Faam B, Fallah MS, Hosseinzadeh N, et al. (2012) Heritability of the metabolic syndrome and its components in the Tehran Lipid and Glucose Study (TLGS). Genet Res 94: 331–337. doi:10.1017/S001667231200050X.

163. Frayn KN (2010) Metabolic regulation a human perspective. Chichester, U.K.; Malden, MA: Wiley-Blackwell Pub.

164. Fisher M (2012) Heart Disease and Diabetes. Oxford University Press. 174 p.

165. O'Reilly PF, Hoggart CJ, Pomyen Y, Calboli FCF, Elliott P, et al. (2012) MultiPhen: Joint Model of Multiple Phenotypes Can Increase Discovery in GWAS. PLoS ONE 7. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3342314/. Accessed 31 December 2013.

166. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, et al. (2009) Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet 41: 703–707. doi:10.1038/ng.381.

167. Consortium the DiaGRAM (DIAGRAM) (2012) Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 44: 981–990. doi:10.1038/ng.2383.

168. Sandholm N, Salem RM, McKnight AJ, Brennan EP, Forsblom C, et al. (2012) New susceptibility loci associated with kidney disease in type 1 diabetes. PLoS Genet 8: e1002921. doi:10.1371/journal.pgen.1002921.

169. Welter D, MacArthur J, Morales J, Burdett T, Hall P, et al. (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42: D1001–D1006. doi:10.1093/nar/gkt1229.

170. Reddy S, Devlin R, Menaa C, Nishimura R, Choi SJ, et al. (1998) Isolation and characterization of a cDNA clone encoding a novel peptide (OSF) that enhances osteoclast formation and bone resorption. J Cell Physiol 177: 636–645. doi:10.1002/(SICI)1097-4652(199812)177:4<636::AID-JCP14>3.0.CO;2-H.

171. Kolker E, Higdon R, Haynes W, Welch D, Broomall W, et al. (2012) MOPED: Model Organism Protein Expression Database. Nucleic Acids Res 40: D1093–D1099. doi:10.1093/nar/gkr1177.

172. Wang M, Weiss M, Simonovic M, Haertinger G, Schrimpf SP, et al. (2012) PaxDb, a database of protein abundance averages across all three domains of life. Mol Cell Proteomics MCP 11: 492–500. doi:10.1074/mcp.O111.014704.

173. Schaab C, Geiger T, Stoehr G, Cox J, Mann M (2012) Analysis of high accuracy, quantitative proteomics data in the MaxQB database. Mol Cell Proteomics MCP 11: M111.014068. doi:10.1074/mcp.M111.014068.

174. Yerges LM, Klei L, Cauley JA, Roeder K, Kammerer CM, et al. (2009) High-Density Association Study of 383 Candidate Genes for Volumetric BMD at the Femoral Neck and Lumbar Spine Among Older Men. J Bone Miner Res 24: 2039–2049. doi:10.1359/jbmr.090524.

175. Boudry-Labis E, Demeer B, Le Caignec C, Isidor B, Mathieu-Dramard M, et al. (2013) A novel microdeletion syndrome at 9q21.13 characterised by mental retardation, speech delay, epilepsy and characteristic facial features. Eur J Med Genet 56: 163–170. doi:10.1016/j.ejmg.2012.12.006.

176. Ulivi S, Mezzavilla M, Gasparini P (2013) Genetics of eye colours in different rural populations on the Silk Road. Eur J Hum Genet EJHG 21: 1320–1323. doi:10.1038/ejhg.2013.41.

177. Scott LJ, Muglia P, Kong XQ, Guan W, Flickinger M, et al. (2009) Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. Proc Natl Acad Sci U S A 106: 7501–7506. doi:10.1073/pnas.0813386106.

178. Sherva R, Tripodis Y, Bennett DA, Chibnik LB, Crane PK, et al. (2014) Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. Alzheimers Dement J Alzheimers Assoc 10: 45–52. doi:10.1016/j.jalz.2013.01.008.

179. Hong K-W, Kim SS, Kim Y (2013) Genome-wide association study of orthostatic hypotension and supine-standing blood pressure changes in two korean populations. Genomics Inform 11: 129–134. doi:10.5808/GI.2013.11.3.129.

180. Terracciano A, Esko T, Sutin AR, de Moor MHM, Meirelles O, et al. (2011) Meta-analysis of genome-wide association studies identifies common variants in CTNNA2 associated with excitement-seeking. Transl Psychiatry 1: e49. doi:10.1038/tp.2011.42.

181. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 44: 955–959. doi:10.1038/ng.2354.

182. Simon-Sanchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, et al. (2009) Genome-Wide Association Study reveals genetic risk underlying Parkinson's disease. Nat Genet 41: 1308–1312. doi:10.1038/ng.487.

183. Abhary S, Hewitt AW, Burdon KP, Craig JE (2009) A Systematic Meta-Analysis of Genetic Association Studies for Diabetic Retinopathy. Diabetes 58: 2137–2147. doi:10.2337/db09-0059.

184. Agardh E, Gaur LK, Lernmark A, Agardh C-D (2004) HLA-DRB1, -DQA1, and -DQB1 subtypes or ACE gene polymorphisms do not seem to be risk markers for severe retinopathy in younger Type 1 diabetic patients. J Diabetes Complications 18: 32–36. doi:10.1016/S1056-8727(03)00040-0.

185. Marre M, Bernadet P, Gallois Y, Savagner F, Guyene TT, et al. (1994) Relationships between angiotensin I converting enzyme gene polymorphism, plasma levels, and diabetic retinal and renal complications. Diabetes 43: 384–388.

186. Nagi DK, Mansfield MW, Stickland MH, Grant PJ (1995) Angiotensin converting enzyme (ACE) insertion/deletion (I/D) polymorphism, and diabetic retinopathy in subjects with IDDM and NIDDM. Diabet Med J Br Diabet Assoc 12: 997–1001.

187. Rabensteiner D, Abrahamian H, Irsigler K, Hermann KM, Kiener HP, et al. (1999) ACE gene polymorphism and proliferative retinopathy in type 1 diabetes: results of a case-control study. Diabetes Care 22: 1530–1535.

188. Tarnow L, Cambien F, Rossing P, Nielsen FS, Hansen BV, et al. (1995) Lack of relationship between an insertion/deletion polymorphism in the angiotensin I-converting enzyme gene and diabetic nephropathy and proliferative retinopathy in IDDM patients. Diabetes 44: 489–494.

189. Van Ittersum FJ, de Man AM, Thijssen S, de Knijff P, Slagboom E, et al. (2000) Genetic polymorphisms of the renin-angiotensin system and complications of insulin-dependent diabetes mellitus. Nephrol Dial Transplant Off Publ Eur Dial Transpl Assoc - Eur Ren Assoc 15: 1000–1007.

190. Araz M, Yilmaz N, Güngör K, Okan V, Kepekci Y, et al. (2001) Angiotensin-converting enzyme gene polymorphism and microvascular complications in Turkish type 2 diabetic patients. Diabetes Res Clin Pract 54: 95–104.

191. Degirmenci I, Kebapci N, Basaran A, Efe B, Gunes HV, et al. (2005) Frequency of angiotensin-converting enzyme gene polymorphism in Turkish type 2 diabetic patients. Int J Clin Pract 59: 1137–1142. doi:10.1111/j.1368-5031.2005.00586.x.

192. Globocnik-Petrovic M, Hawlina M, Peterlin B, Petrovic D (2003) Insertion/deletion plasminogen activator inhibitor 1 and insertion/deletion angiotensin-converting enzyme gene polymorphisms in diabetic retinopathy in type 2 diabetes. Ophthalmol J Int Ophtalmol Int J Ophthalmol Z Für Augenheilkd 217: 219–224. doi:68975.

193. Kanková K, Muzík J, Karásková J, Beránek M, Hájek D, et al. (2001) Duration of non-Insulin-dependent diabetes mellitus and the TNF-beta NcoI genotype as predictive factors in proliferative diabetic retinopathy. Ophthalmol J Int Ophtalmol Int J Ophthalmol Z Für Augenheilkd 215: 294–298. doi:50877.

194. Taverna MJ, Sola A, Guyot-Argenton C, Pacher N, Bruzzo F, et al. (2002) eNOS4 polymorphism of the endothelial nitric oxide synthase predicts risk for severe diabetic retinopathy. Diabet Med J Br Diabet Assoc 19: 240–245.

195. Petrovic MG, Cilensek I, Petrovic D (2008) Manganese superoxide dismutase gene polymorphism (V16A) is associated with diabetic retinopathy in Slovene (Caucasians) type 2 diabetes patients. Dis Markers 24: 59–64.

196. Petrovic MG, Korosec P, Kosnik M, Osredkar J, Hawlina M, et al. (2008) Local and genetic determinants of vascular endothelial growth factor expression in advanced proliferative diabetic retinopathy. Mol Vis 14: 1382–1387.

197. Szaflik JP, Wysocki T, Kowalski M, Majsterek I, Borucka AI, et al. (2008) An association between vascular endothelial growth factor gene promoter polymorphisms and diabetic retinopathy. Graefes Arch Clin Exp Ophthalmol Albrecht Von Graefes Arch Für Klin Exp Ophthalmol 246: 39–43. doi:10.1007/s00417-007-0674-6.

198. Demaine A, Cross D, Millward A (2000) Polymorphisms of the aldose reductase gene and susceptibility to retinopathy in type 1 diabetes mellitus. Invest Ophthalmol Vis Sci 41: 4064–4068.

199. Kao YL, Donaghue K, Chan A, Knight J, Silink M (1999) A novel polymorphism in the aldose reductase gene promoter region is strongly associated with diabetic retinopathy in adolescents with type 1 diabetes. Diabetes 48: 1338–1340.

200. Dos Santos KG, Canani LH, Gross JL, Tschiedel B, Souto KEP, et al. (2006) The -106CC genotype of the aldose reductase gene is associated with an increased risk of proliferative diabetic retinopathy in Caucasian-Brazilians with type 2 diabetes. Mol Genet Metab 88: 280–284. doi:10.1016/j.ymgme.2006.02.002.

201. Heesom AE, Hibberd ML, Millward A, Demaine AG (1997) Polymorphism in the 5'-end of the aldose reductase gene is strongly associated with the development of diabetic nephropathy in type I diabetes. Diabetes 46: 287–291.

202. Lajer M, Tarnow L, Fleckner J, Hansen BV, Edwards DG, et al. (2004) Association of aldose reductase gene Z+2 polymorphism with reduced susceptibility to diabetic nephropathy in Caucasian Type 1 diabetic patients. Diabet Med J Br Diabet Assoc 21: 867–873. doi:10.1111/j.1464-5491.2004.01259.x.

203. Petrovic MG, Peterlin B, Hawlina M, Petrovic D (2005) Aldose reductase (AC)n gene polymorphism and susceptibility to diabetic retinopathy in Type 2 diabetes in Caucasians. J Diabetes Complications 19: 70–73. doi:10.1016/j.jdiacomp.2004.08.004.

204. Ng DPK (2010) Human Genetics of Diabetic Retinopathy: Current Perspectives. J Ophthalmol 2010. Available: http://www.hindawi.com/journals/jop/2010/172593/abs/. Accessed 31 May 2013.

205. Hallman DM, Boerwinkle E, Gonzalez VH, Klein BEK, Klein R, et al. (2007) A genome-wide linkage scan for diabetic retinopathy susceptibility genes in Mexican

Americans with type 2 diabetes from Starr County, Texas. Diabetes 56: 1167–1173. doi:10.2337/db06-1373.

206. Imperatore G, Hanson RL, Pettitt DJ, Kobes S, Bennett PH, et al. (1998) Sib-pair linkage analysis for susceptibility genes for microvascular complications among Pima Indians with type 2 diabetes. Pima Diabetes Genes Group. Diabetes 47: 821–830.

207. De Benedetto U, Querques G, Lattanzio R, Borrelli E, Triolo G, et al. (2014) MACULAR DYSFUNCTION IS COMMON IN BOTH TYPE 1 AND TYPE 2 DIABETIC PATIENTS WITHOUT MACULAR EDEMA. Retina Phila Pa. doi:10.1097/IAE.0000000000000205.

208. Abcouwer SF, Gardner TW (2014) Diabetic retinopathy: loss of neuroretinal adaptation to the diabetic metabolic environment. Ann N Y Acad Sci 1311: 174–190. doi:10.1111/nyas.12412.

209. Kamoi K, Takeda K, Hashimoto K, Tanaka R, Okuyama S (2013) Identifying risk factors for clinically significant diabetic macula edema in patients with type 2 diabetes mellitus. Curr Diabetes Rev 9: 209–217.

210. Nagaoka T, Yoshida A (2013) Relationship between retinal blood flow and renal function in patients with type 2 diabetes and chronic kidney disease. Diabetes Care 36: 957–961. doi:10.2337/dc12-0864.

211. Agroiya P, Philip R, Saran S, Gutch M, Tyagi R, et al. (2013) Association of serum lipids with diabetic retinopathy in type 2 diabetes. Indian J Endocrinol Metab 17: S335–S337. doi:10.4103/2230-8210.119637.

212. Gibson G (2011) Rare and common variants: twenty arguments. Nat Rev Genet 13: 135–145. doi:10.1038/nrg3118.

213. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. (2009) Finding the missing heritability of complex diseases. Nature 461: 747–753. doi:10.1038/nature08494.

214. Project the SJCRH-WUPCG (2013) Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. Nat Genet 45: 602–612. doi:10.1038/ng.2611.

215. the Cohorts for Heart and Aging Research in Genetic Epidemiology (CHARGE) Consortium (2013) Whole-genome sequence-based analysis of high-density lipoprotein cholesterol. Nat Genet 45: 899–901. doi:10.1038/ng.2671.

216. Jensen RA, Sim X, Li X, Cotch MF, Ikram MK, et al. (2013) Genome-Wide Association Study of Retinopathy in Individuals without Diabetes. PLoS ONE 8: e54232. doi:10.1371/journal.pone.0054232.

217. Zhu J, Sova P, Xu Q, Dombek KM, Xu EY, et al. (2012) Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation. PLoS Biol 10: e1001301. doi:10.1371/journal.pbio.1001301.

218. Pescosolido N, Campagna O, Barbato A (2014) Diabetic retinopathy and pregnancy. Int Ophthalmol 34: 989–997. doi:10.1007/s10792-014-9906-z.

219. Pasquali L, Gaulton KJ, Rodr ́guez-Segu ́ıSA, Mularoni L, Miguel-Escalada I, et al. (2014) Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat Genet 46: 136–143. doi:10.1038/ng.2870.

220. Huang Q, Whitington T, Gao P, Lindberg JF, Yang Y, et al. (2014) A prostate cancer susceptibility allele at 6q22 increases RFX6 expression by modulating HOXB13 chromatin binding. Nat Genet 46: 126–135. doi:10.1038/ng.2862.

221. Bixler GV, VanGuilder HD, Brucklacher RM, Kimball SR, Bronson SK, et al. (2011) Chronic insulin treatment of diabetes does not fully normalize alterations in the retinal transcriptome. BMC Med Genomics 4: 40. doi:10.1186/1755-8794-4-40.

222. Kandpal RP, Rajasimha HK, Brooks MJ, Nellissery J, Wan J, et al. (2012) Transcriptome analysis using next generation sequencing reveals molecular signatures of diabetic retinopathy and efficacy of candidate drugs. Mol Vis 18: 1123–1146.

223. Brondani LA, de Souza BM, Duarte GCK, Kliemann LM, Esteves JF, et al. (2012) The UCP1 -3826A/G polymorphism is associated with diabetic retinopathy and increased UCP1 and MnSOD2 gene expression in human retina. Invest Ophthalmol Vis Sci 53: 7449–7457. doi:10.1167/iovs.12-10660.

224. Liu Y, Wang M, Morris AD, Doney ASF, Leese GP, et al. (2013) Glycemic Exposure and Blood Pressure Influencing Progression and Remission of Diabetic Retinopathy: A longitudinal cohort study in GoDARTS. Diabetes Care. doi:10.2337/dc12-2392.

225. Sobrin L, Green T, Sim X, Jensen RA, Tai ES, et al. (2011) Candidate Gene Association Study for Diabetic Retinopathy in Persons with Type 2 Diabetes: The Candidate Gene Association Resource (CARe). Invest Ophthalmol Vis Sci 52: 7593–7602. doi:10.1167/iovs.11-7510.

226. MCMC convergence assessment via a trace plot (n.d.). Available: https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_introbayes_sect008.htm.

227. Box GEP, Jenkins GM, Reinsel GC (2013) Time Series Analysis: Forecasting and Control. John Wiley & Sons. 781 p.

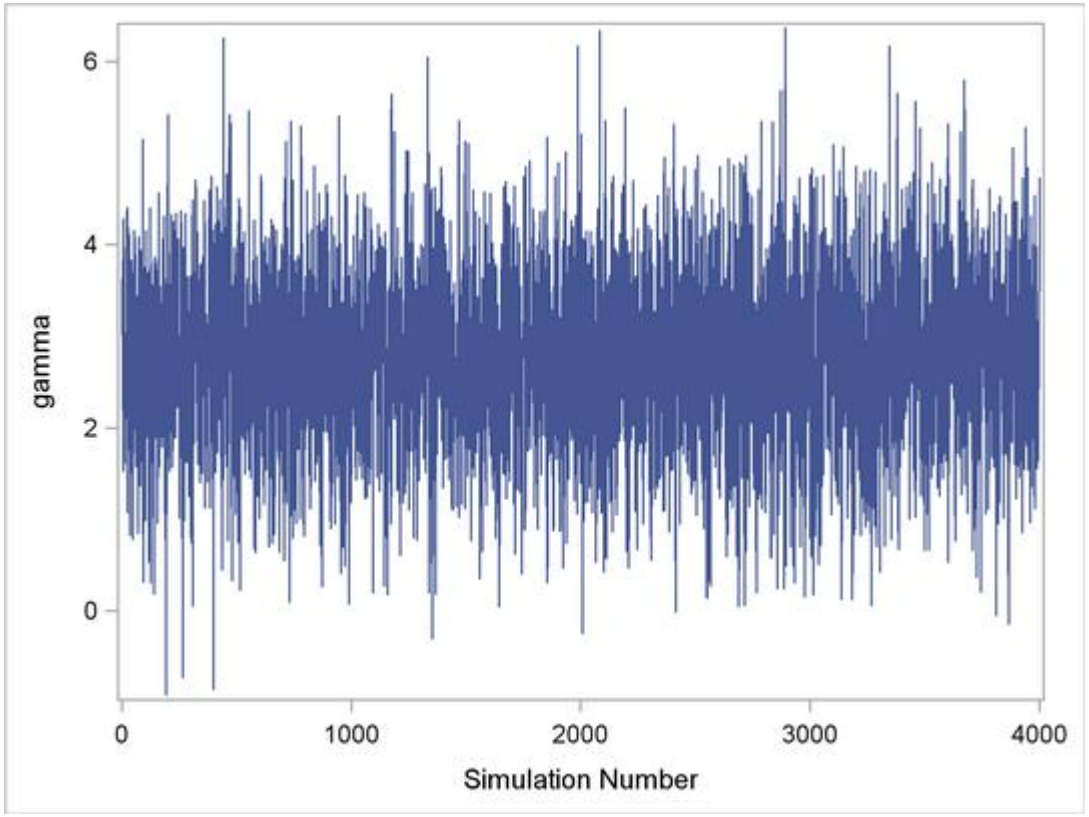
Appendix A MCMC data inspection tools: trace, density and autocorrelation plots

A trace plot provides a visual tool for assessing the convergence of MCMC (here, Gibbs sampling) data. In a trace plot, sampled parameter values are plotted against the iteration number. Frequently, sampling data is collected after a period of software running, to ensure only non-convergent data is filtered out. The MCMC sampling data is also named as a (Markove) chain, as this is sampled from population distributions guided by the Markov chain. The trace is informative of whether the data has converged to its stationary distribution (the unobserved distribution of true parameter values). An MCMC algorithm is efficient (mixed well), if it samples the space of parameter values rapidly. The trace plot is also informational of whether sampling is efficient.

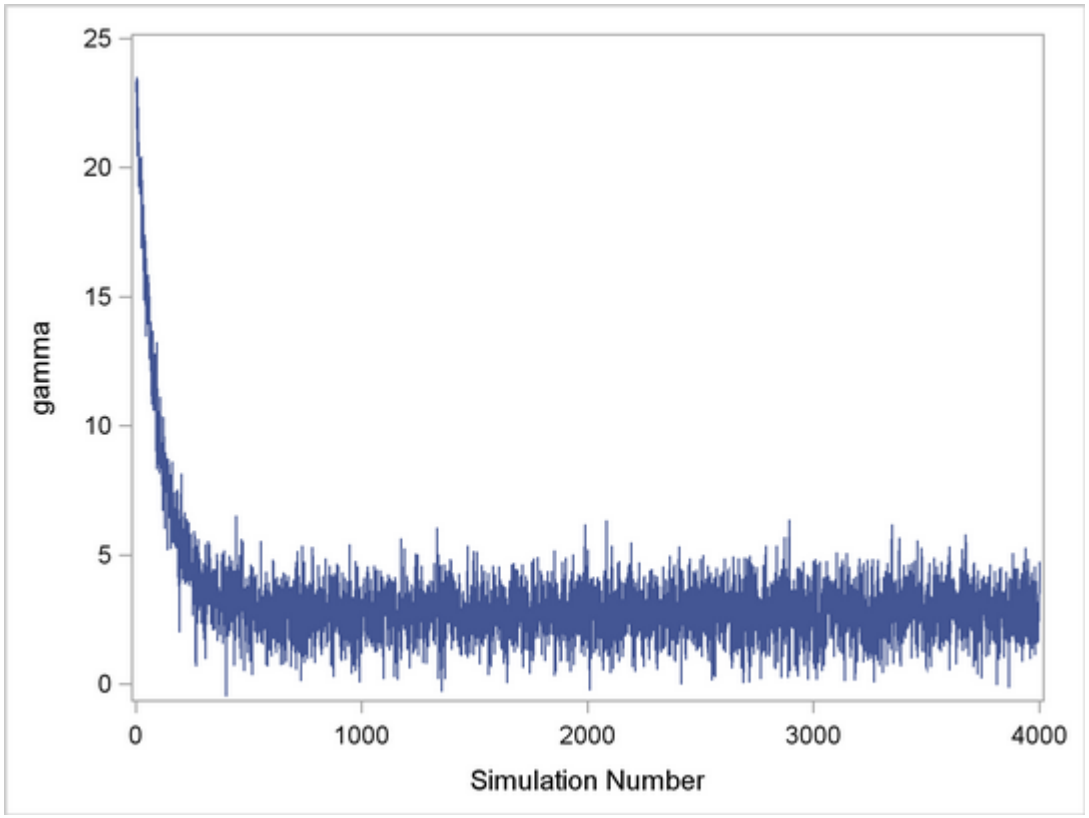
The sampled data may be convergent, if the distribution of sampled values does not show a systematic trend of upward or downward variations, as the iteration number increases. In a trace plot, this is recognised by a constant mean and variance.

Appendix Figure A-1 Traces [226] for (a) convergent MCMC data, (b) mixed convergence data with non-convergent initial values and (c) poor convergence data with unacceptable mixing.

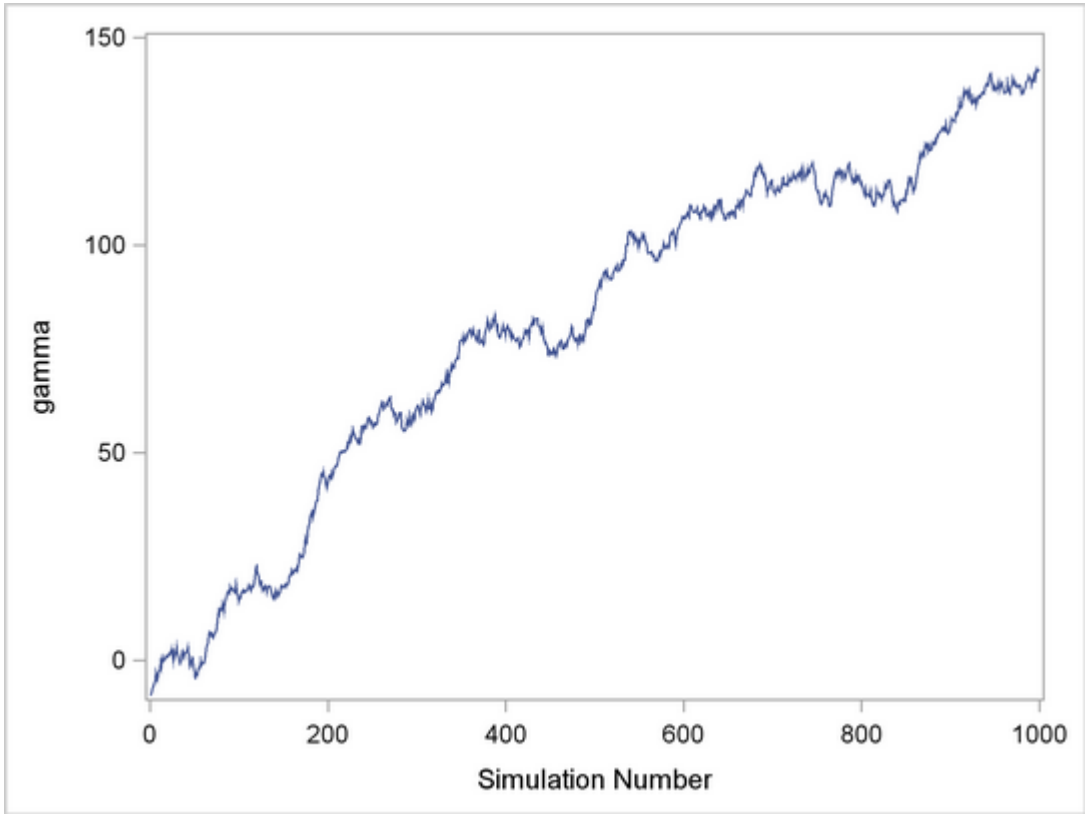
(a) Sampled data appeared centred at 3, with very small fluctuations, suggesting that the MCMC chain may have converged. The example MCMC data traversed regions with low density. Therefore, it can be concluded that data is well mixed and converged.



(b) A chain with initial non-convergent values settled to convergence as sampling proceeded. The initial sampled data should be discarded.



(c) The trace explored very narrow range of parameter values within a period of iterations, indicating of poor mixing. The upward trend indicates the MCMC data is non-convergent.

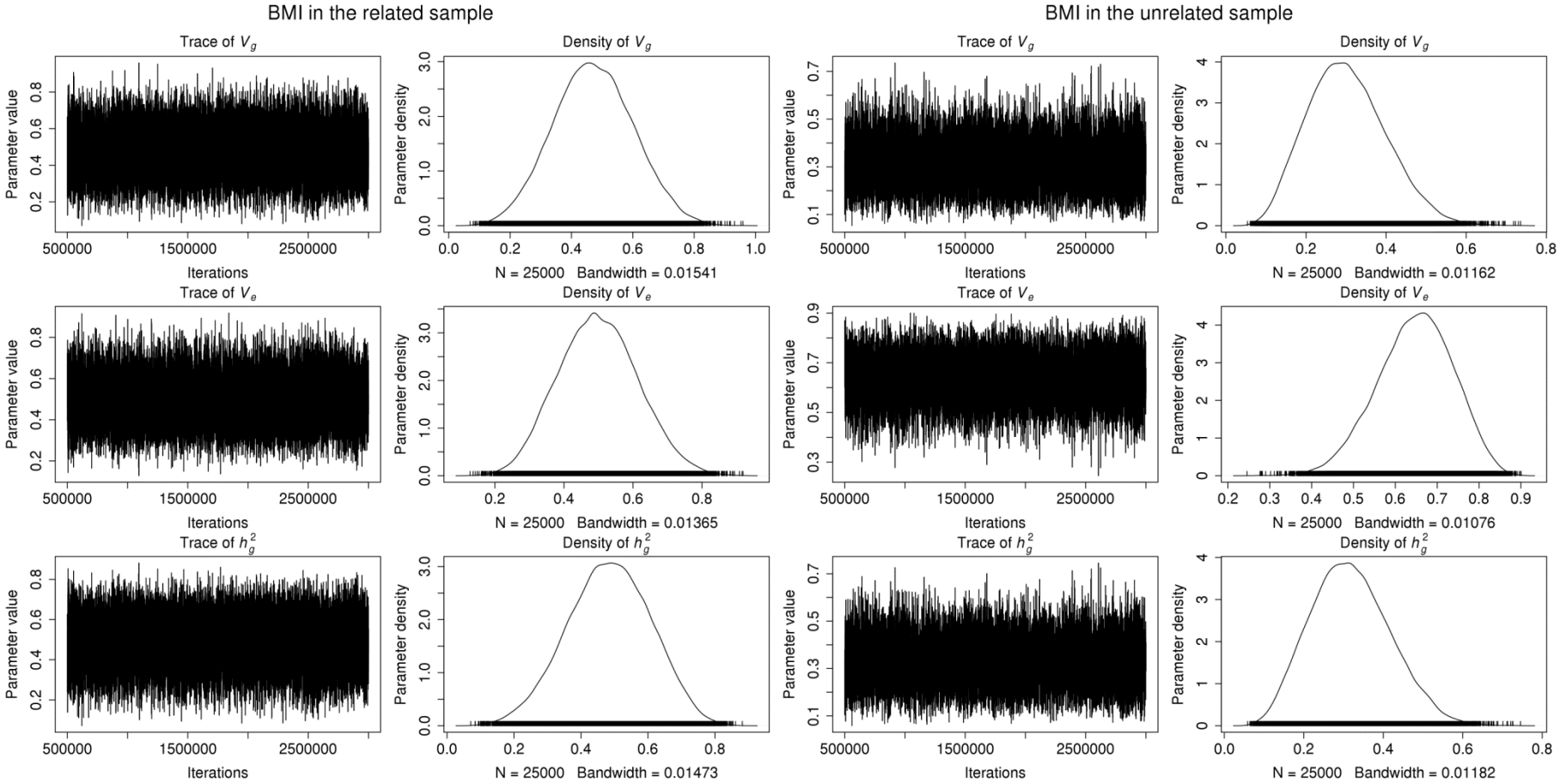


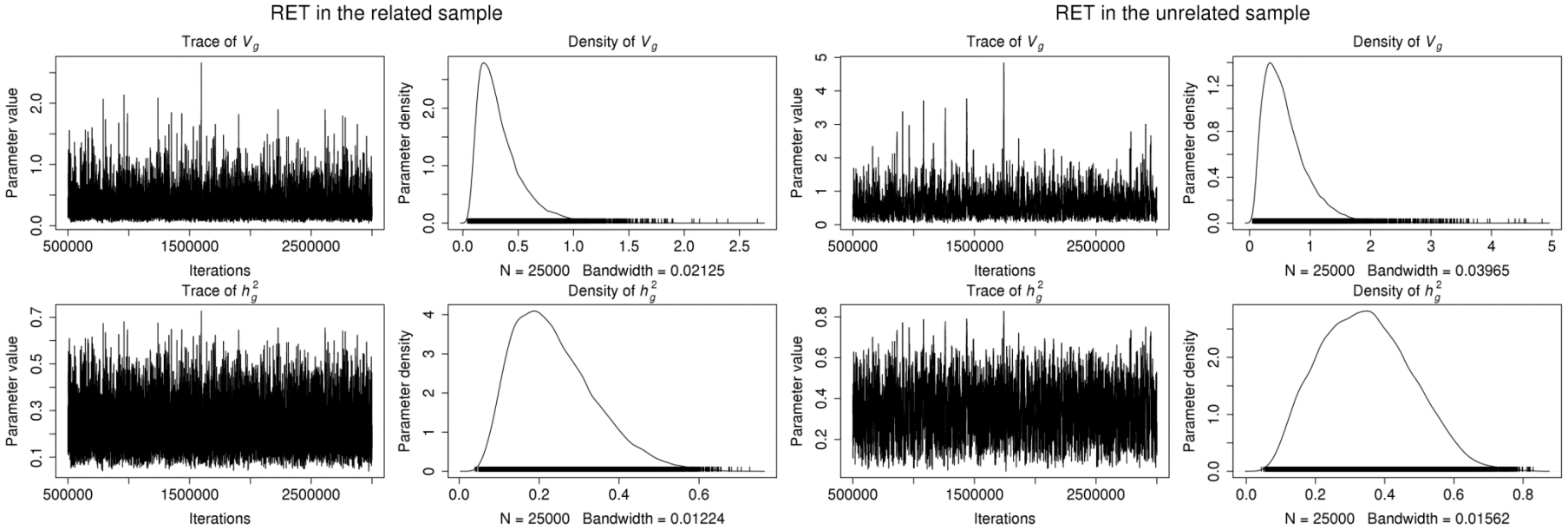
MCMC data is a series of parameter values Y_1, Y_2, ..., Y_N at equally spaced sampling time X_1, X_2, ..., X_N. The lag k autocorrelation is defined by $r_k = \frac{\sum_{i=1}^{N-k}\left(Y_i - \bar{Y}\right)\left(Y_{i+k} - \bar{Y}\right)}{\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2}$ [227]. It is anticipated that in a well-mixed MCMC data, autocorrelation would decrease with an increasing lag k. For example, the autocorrelation between the 1st sampled value of parameters and the 100th value should be smaller than the autocorrelation between the 1st value and 10th value. Consistently high autocorrelations would indicate poor mixing, which is manifested in an autocorrelation plot as similarly heights of bars (values of autocorrelation).

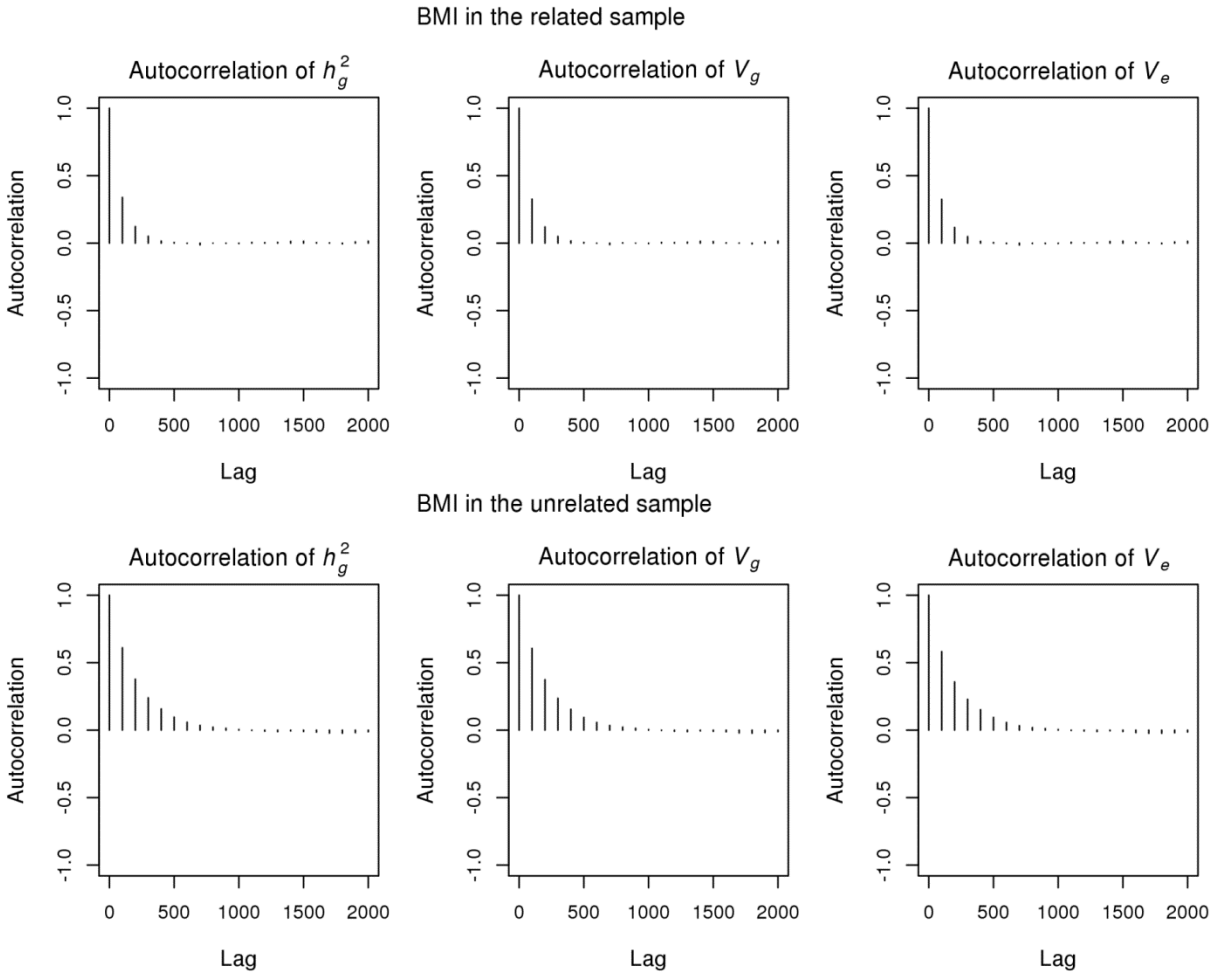
The density plot visually presents the distribution of sampled MCMC data, and can be used in inferring the statistical characteristics of this distribution (for example, the mode and the dispersion of the MCMC data).

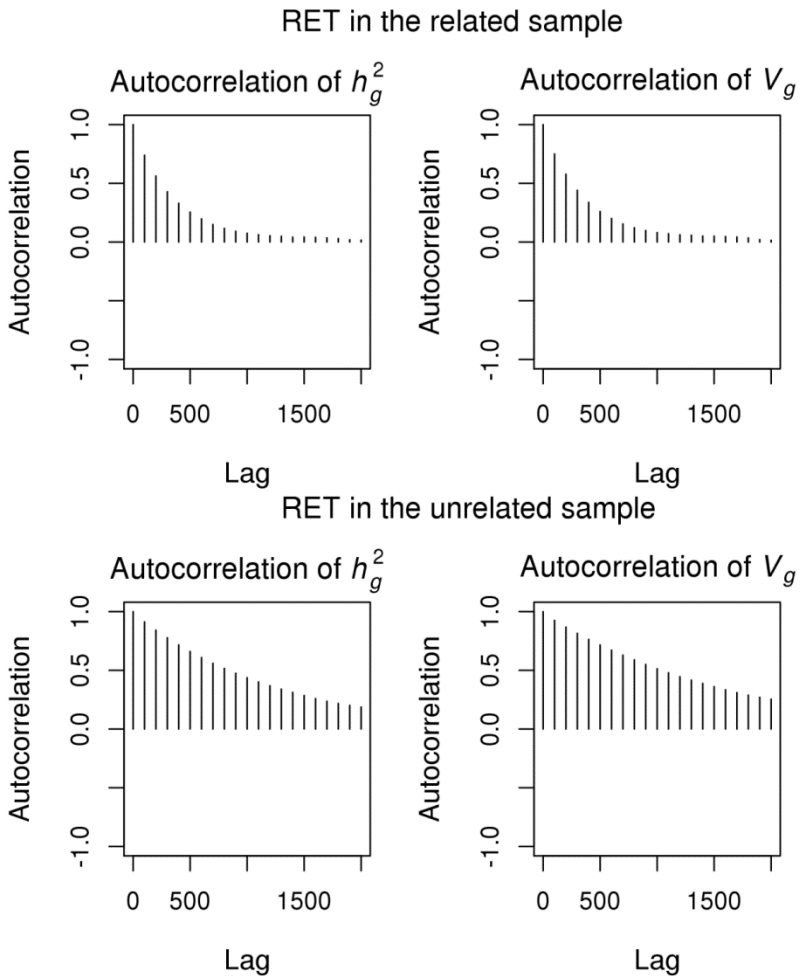
Appendix B MCMC data plots for univariate mixed models

Appendix Figure B-1 Trace, density and autocorrelation plots of genetic (V_g) and/or residual (V_e) variance, heritability (h_g^2) captured by common genome-wide SNPs for BMI and RET (an examplified sample) are shown. BMI: body mass index; RET: retinopathy.


BMI in the related sample
Trace of V_g
Density of V_g
Parameter value
Parameter density
Iterations
N = 25000 Bandwidth = 0.01541
Trace of V_e
Density of V_e
N = 25000 Bandwidth = 0.01365
Trace of h_g^2
Density of h_g^2
N = 25000 Bandwidth = 0.01473
BMI in the unrelated sample
Trace of V_g
Density of V_g
N = 25000 Bandwidth = 0.01162
Trace of V_e
Density of V_e
N = 25000 Bandwidth = 0.01076
Trace of h_g^2
Density of h_g^2
N = 25000 Bandwidth = 0.01182

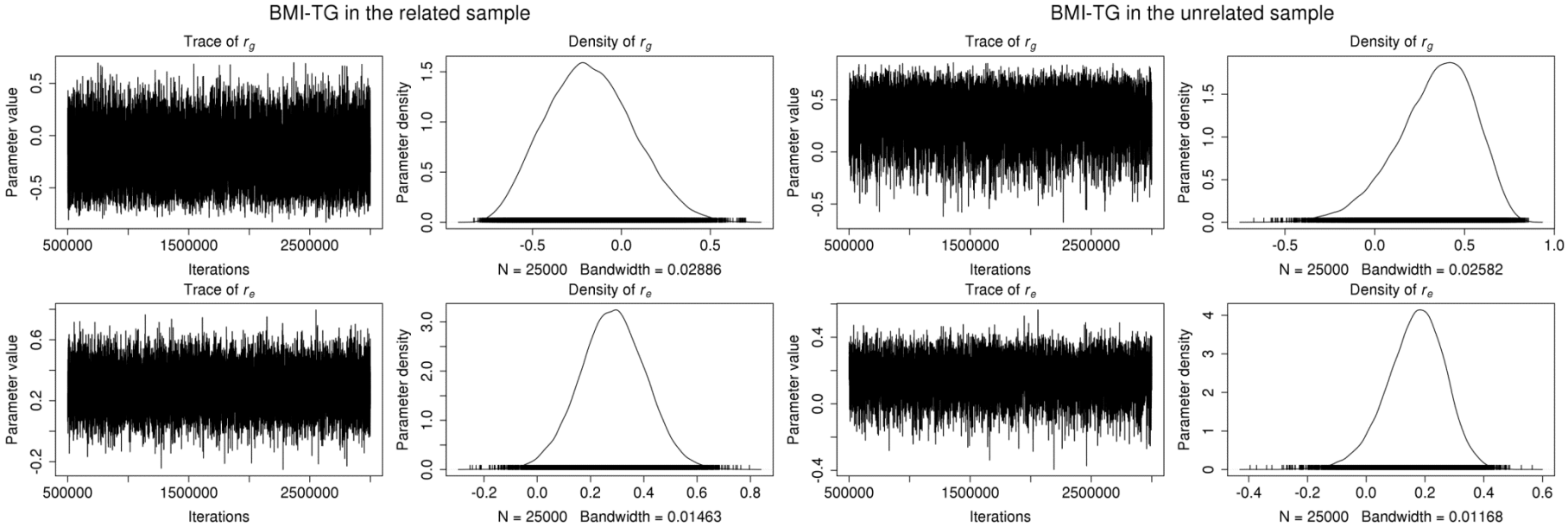

RET in the related sample
Trace of V_g
Parameter value
Iterations
Density of V_g
Parameter density
N = 25000 Bandwidth = 0.02125
Trace of h_g^2
Density of h_g^2
N = 25000 Bandwidth = 0.01224
RET in the unrelated sample
Trace of V_g
Density of V_g
N = 25000 Bandwidth = 0.03965
Trace of h_g^2
Density of h_g^2
N = 25000 Bandwidth = 0.01562

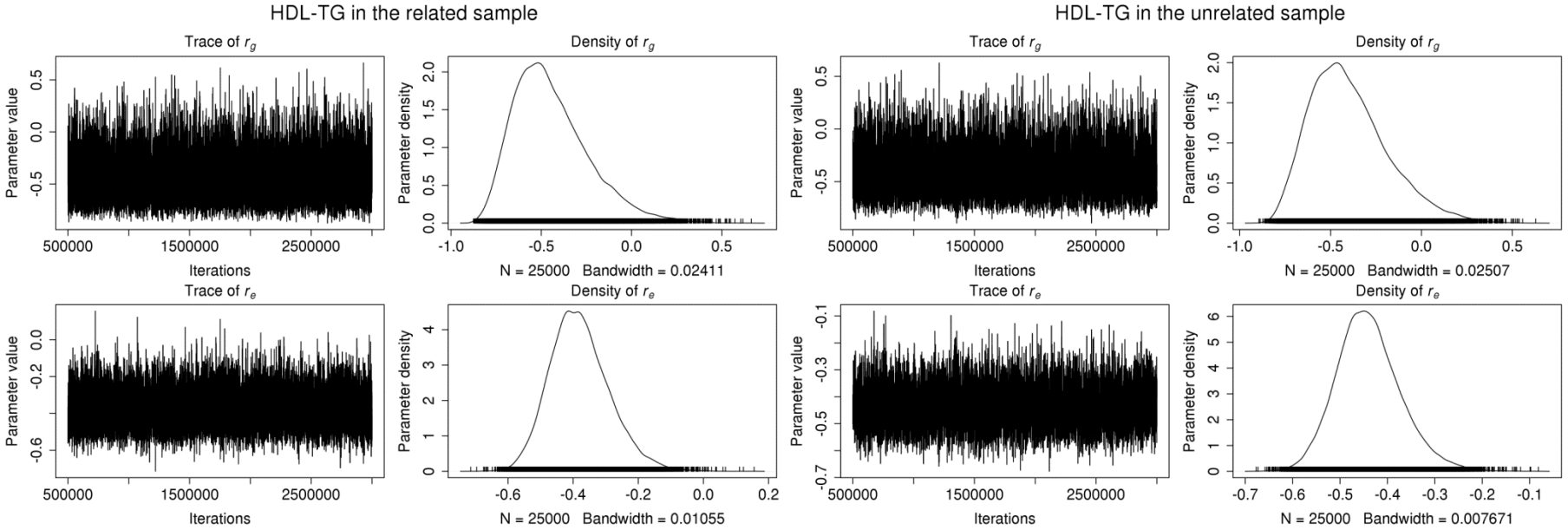

BMI in the related sample
Autocorrelation of h_g^2
Autocorrelation of V_g
Autocorrelation of V_e
Autocorrelation
Lag
BMI in the unrelated sample
Autocorrelation of h_g^2
Autocorrelation of V_g
Autocorrelation of V_e
Autocorrelation
Lag

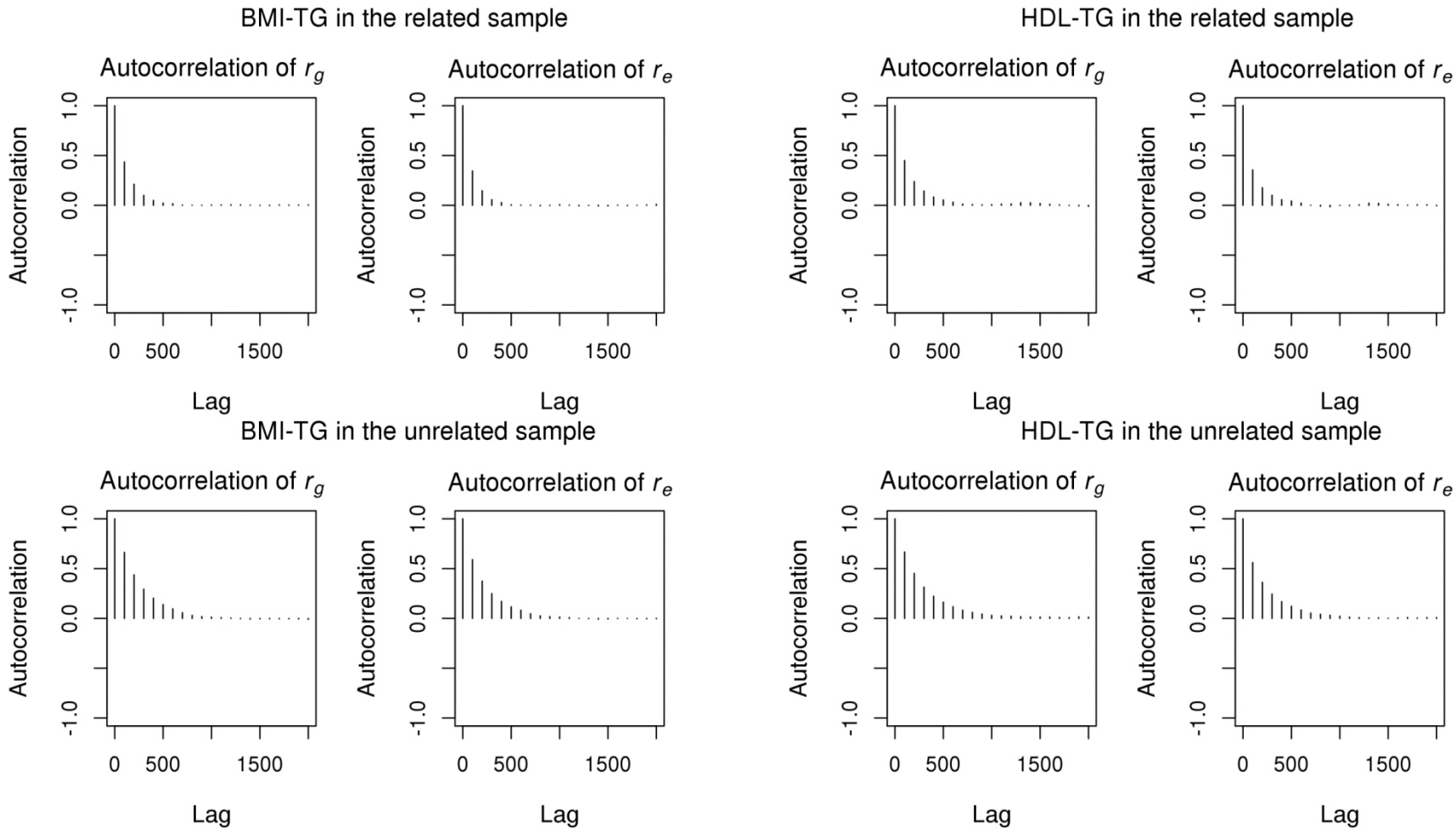

RET in the related sample
Autocorrelation of h_g^2
Autocorrelation of V_g
Autocorrelation
-1.0
0.0
0.5
1.0
0
500
1500
Lag
RET in the unrelated sample
Autocorrelation of h_g^2
Autocorrelation of V_g
Autocorrelation
-1.0
0.0
0.5
1.0
0
500
1500
Lag

Appendix C MCMC data plots for bivariate mixed models

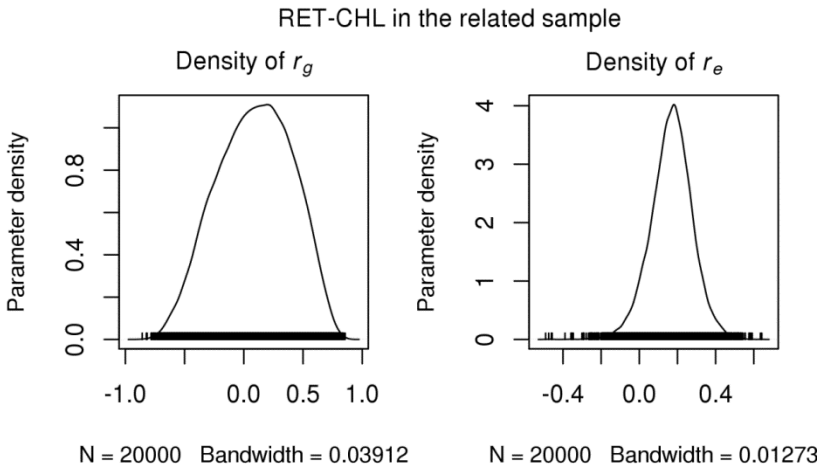
Appendix Figure C-1 Trace, density and autocorrelation plots of genetic (r_g) and residual (r_e) correlation coefficients captured by common genome-wide SNPs for BMI-TG and HDL-TG (an examplified sample)are shown. BMI: body mass index; HDL (HDL-c): high-density lipoprotein cholesterol; TG: triglycerides.

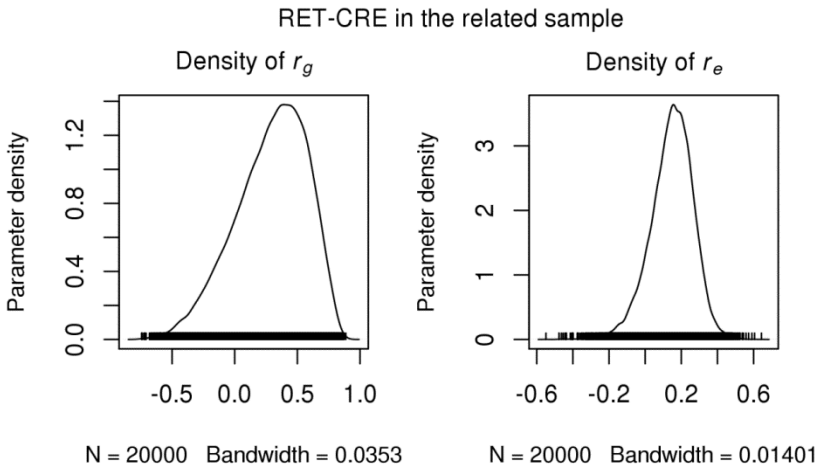

BMI-TG in the related sample
Trace of r_g
Parameter value
Iterations
Density of r_g
Parameter density
N = 25000 Bandwidth = 0.02886
Trace of r_e
Density of r_e
N = 25000 Bandwidth = 0.01463
BMI-TG in the unrelated sample
Trace of r_g
Density of r_g
N = 25000 Bandwidth = 0.02582
Trace of r_e
Density of r_e
N = 25000 Bandwidth = 0.01168

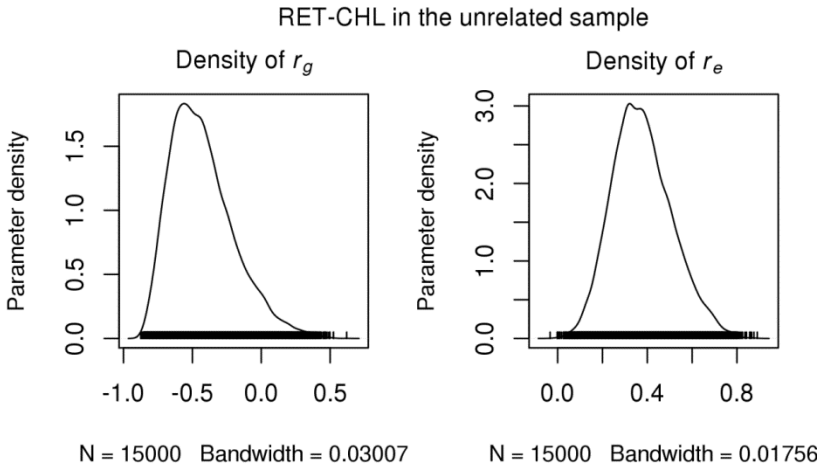

HDL-TG in the related sample
Trace of r_g
Parameter value
Iterations
Density of r_g
Parameter density
N = 25000 Bandwidth = 0.02411
Trace of r_e
Parameter value
Iterations
Density of r_e
Parameter density
N = 25000 Bandwidth = 0.01055
HDL-TG in the unrelated sample
Trace of r_g
Parameter value
Iterations
Density of r_g
Parameter density
N = 25000 Bandwidth = 0.02507
Trace of r_e
Parameter value
Iterations
Density of r_e
Parameter density
N = 25000 Bandwidth = 0.007671

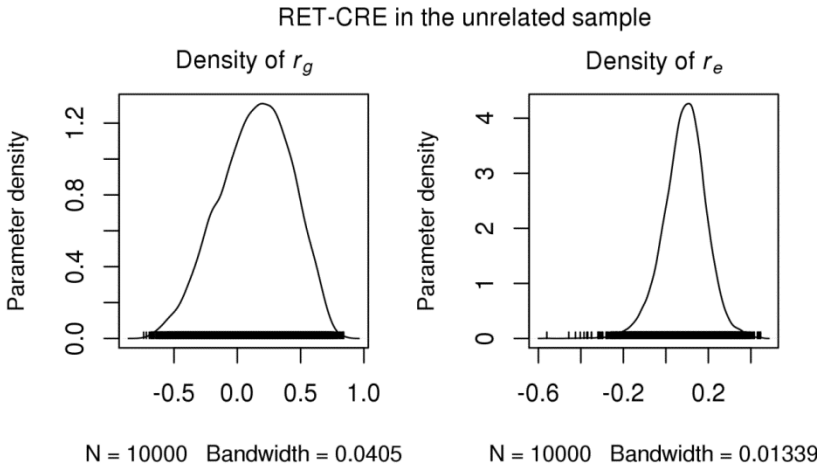

BMI-TG in the related sample
Autocorrelation of r_g
Autocorrelation of r_e
Autocorrelation
-1.0
0.0
0.5
1.0
0
500
1500
Lag
HDL-TG in the related sample
BMI-TG in the unrelated sample
HDL-TG in the unrelated sample

Appendix Figure C-2 Density plots of genetic (r_g) and residual (r_e) correlation coefficients captured by common genome-wide SNPs for RET-CHL and RET-CRE (an exemplified sample) are shown. CHL: cholesterol; CRE: serum creatinine; RET: retinopathy.


RET-CHL in the related sample
Density of r_g
Parameter density
N = 20000 Bandwidth = 0.03912
Density of r_e
Parameter density
N = 20000 Bandwidth = 0.01273


RET-CRE in the related sample
Density of r_g
Parameter density
N = 20000 Bandwidth = 0.0353
Density of r_e
Parameter density
N = 20000 Bandwidth = 0.01401


RET-CHL in the unrelated sample
Density of r_g
Parameter density
N = 15000 Bandwidth = 0.03007
Density of r_e
Parameter density
N = 15000 Bandwidth = 0.01756


RET-CRE in the unrelated sample
Density of r_g
Parameter density
N = 10000 Bandwidth = 0.0405
Density of r_e
Parameter density
N = 10000 Bandwidth = 0.01339

Appendix D The SUMMIT genome-wide meta-anlaysis of diabetic retinopathy

Appendix Table D-1 Top signal SNPs in the indpendent regions (>100 Kbp) with the assocation significance (*P* value < 10^{-5}) based on the SUMMIT genome-wide meta-analysis of diabetic retinopathy.

Chr.	Position[a]	SNP	Effective allele frequency[b]	Effective/ alternative allele[c]	OR (95% CI)[d]	Meta *P* value[e]	I^2 (%), *P* value[f]
9	77817536	rs10746970	0.68	A/T	1.19 (1.11, 1.27)	2.22×10^{-7}	0, 6.46×10^{-1}
2	81987573	rs1653654	0.03	A/G	0.63 (0.51, 0.78)	4.27×10^{-7}	0, 8.32×10^{-1}
17	71368680	rs75125621	0.03	G/A	0.61 (0.49, 0.77)	6.13×10^{-7}	60, 4.03×10^{-2}
4	145350264	rs2657795	0.21	A/C	0.82 (0.76, 0.89)	8.25×10^{-7}	2, 3.98×10^{-1}
2	184738091	rs1682430	0.89	C/T	1.24 (1.10, 1.39)	1.07×10^{-6}	21, 2.71×10^{-1}
5	141158686	rs10072382	0.02	A/G	1.99 (1.52, 2.62)	1.30×10^{-6}	28, 2.41×10^{-1}
13	93033916	rs9516067	0.56	A/G	1.16 (1.09, 1.24)	1.81×10^{-6}	51, 6.69×10^{-2}
1	35627132	rs6425936	0.50	A/G	0.76 (0.68, 0.85)	1.83×10^{-6}	16, 3.08×10^{-1}

Chr.	Position[a]	SNP	Effective allele frequency[b]	Effective/ alternative allele[c]	OR (95% CI)[d]	Meta *P* value[e]	I^2 (%), *P* value[f]
6	37583911	rs810855	0.21	C/T	0.83 (0.77, 0.90)	1.98×10^{-6}	0, 7.97×10^{-1}
1	80022362	rs55939932	0.02	T/C	0.48 (0.36, 0.65)	2.01×10^{-6}	0, 8.98×10^{-1}
3	146324544	rs163757	0.36	A/T	0.86 (0.80, 0.93)	2.79×10^{-6}	0, 9.99×10^{-1}
1	240674842	rs35428289	0.21	A/G	1.22 (1.12, 1.32)	3.11×10^{-6}	0, 4.73×10^{-1}
12	27969088	rs10771375	0.44	G/A	0.87 (0.81, 0.93)	3.13×10^{-6}	0, 8.48×10^{-1}
11	93564393	rs601711	0.64	T/C	1.17 (1.09, 1.25)	3.35×10^{-6}	0, 4.55×10^{-1}
1	41835871	rs4660191	0.97	G/A	1.75 (1.38, 2.21)	3.71×10^{-6}	6, 3.76×10^{-1}
6	93361564	rs7750013	0.94	C/A	1.45 (1.22, 1.73)	4.05×10^{-6}	0, 7.35×10^{-1}
19	32473222	rs8113622	0.33	T/G	0.86 (0.80, 0.92)	4.06×10^{-6}	0, 6.13×10^{-1}
3	22624716	rs73033654	0.08	T/C	1.30 (1.15, 1.47)	4.20×10^{-6}	34, 1.78×10^{-1}
2	101346696	rs7579862	0.19	G/A	1.19 (1.08, 1.30)	4.22×10^{-6}	0, 5.36×10^{-1}
5	26698737	rs72758936	0.03	A/G	0.63 (0.52, 0.77)	4.33×10^{-6}	39, 1.61×10^{-1}
6	79036864	rs1338321	0.63	G/A	1.16 (1.09, 1.25)	5.50×10^{-6}	27, 2.32×10^{-1}
8	17983016	rs2739683	0.02	C/T	1.68 (1.36, 2.08)	6.10×10^{-6}	0, 8.98×10^{-1}
1	35384605	rs6699355	0.87	T/C	0.78 (0.70, 0.87)	6.13×10^{-6}	45, 1.02×10^{-1}
5	55643386	rs184989476	0.04	C/T	0.62 (0.50, 0.76)	6.33×10^{-6}	21, 2.71×10^{-1}
8	127535431	rs10101440	0.87	T/G	0.80 (0.73, 0.88)	6.83×10^{-6}	0, 4.23×10^{-1}
2	88720222	rs72929232	0.05	G/A	0.71 (0.62, 0.82)	6.87×10^{-6}	29, 2.17×10^{-1}
22	21357602	rs396130	0.20	C/T	1.23 (1.13, 1.34)	7.45×10^{-6}	0, 6.22×10^{-1}

Chr.	Position[a]	SNP	Effective allele frequency[b]	Effective/ alternative allele[c]	OR (95% CI)[d]	Meta *P* value[e]	I^2 (%), *P* value[f]
14	86711866	rs186178322	0.02	A/T	1.74 (1.36, 2.22)	7.68×10^{-6}	0, 8.28×10^{-1}
4	71430467	rs62323371	0.06	A/T	0.73 (0.64, 0.84)	8.03×10^{-6}	0, 8.68×10^{-1}
3	95993574	rs1118907	0.14	G/T	0.79 (0.71, 0.87)	8.12×10^{-6}	0, 6.65×10^{-1}
14	105239192	rs2494732	0.43	C/T	0.84 (0.78, 0.91)	8.22×10^{-6}	0, 9.59×10^{-1}
14	65247373	rs8018785	0.18	C/T	0.82 (0.75, 0.90)	8.51×10^{-6}	0, 7.53×10^{-1}
8	122124302	rs2196051	0.33	G/A	0.87 (0.80, 0.94)	8.75×10^{-6}	0, 6.23×10^{-1}
5	1109646	rs139161838	0.02	A/G	0.50 (0.37, 0.67)	8.77×10^{-6}	25, 2.54×10^{-1}
15	51827471	rs77245046	0.03	C/T	0.66 (0.54, 0.80)	9.12×10^{-6}	38, 1.47×10^{-1}
2	121877105	rs11899778	0.23	C/T	1.21 (1.12, 1.30)	9.46×10^{-6}	35, 1.69×10^{-1}
2	84852570	rs145123753	0.02	T/G	2.08 (1.49, 2.91)	9.65×10^{-6}	0, 6.92×10^{-1}
13	86650174	rs17705805	0.20	T/G	1.18 (1.08, 1.29)	9.86×10^{-6}	10, 3.48×10^{-1}

[a] Human genome build 37, assembly hg19.

[b, c & d] Effective allele frequency, effective/alternative allele, odd ratio and 95% CI were extracted from the meta-analysis of the genetically unrelated samples, whereas in the genetically mixed samples, estimates of effective allele frequency and odds ratio may be biased.

[e & f] Meta and heterogeneity *P* values were extracted from the meta-analysis of the genetically mixed samples, which included more individuals, but this modelling strategy does not provide the estimation of odds ratios.

Appendix Table D-2 The closest gene and gene functional annotation for top independent signal SNPs (>100 Kbp) with the assocation significance (*P* value < 10^{-5}) based on the SUMMIT genome-wide meta-analysis of diabetic retinopathy.

Locus	SNP	Closest Gene	Gene Function
9q13-q21.2	rs10746970	*OSTF1*	Ossification
2p12-p11.1	rs1653654	*CTNNA2*	Structural constituent of cytoskeleton
17q25.1	rs75125621	*SDK2*	Cell adhesion, retinal laminar neuron integrity
4q28-q32	rs2657795	*HHIP*	Carbohydrate metabolic process
2q32.1	rs1682430	*ZNF804A*	Transcriptional control
5q31.3	rs10072382	*PCDH1*	Nervous system development
13q32	rs9516067	*GPC5*	Carbohydrate metabolic process
1p34.3	rs6425936	*SFPQ*	DNA recombination and repair, RNA splicing
6p21	rs810855	*MDGA1*	Neuron migration
1p33-p32	rs55939932	*ELTD1*	Neuropeptide signaling pathway
3q24	rs163757	*PLSCR5*	Unknown
1q43	rs35428289	*GREM2*	Regulation of cytokine activity
12p11.22	rs10771375	*KLHL42*	Regulation of microtubule-based process
11q21	rs601711	*VSTM5*	Transmembrane protein
1p34.2	rs4660191	*FOXO6*	Sequence-specific DNA binding transcription factor activity, glucose sensing
6q16.1	rs7750013	*EPHA7*	Retinal ganglion cell axon guidance

Locus	SNP	Closest Gene	Gene Function
19q13.11	rs8113622	*ZNF507*	Regulation of transcription
3p24.3	rs73033654	*ZNF385D*	Regulation of transcription
2q11.2	rs7579862	*NPAS2*	Cellular lipid metabolic process, central nervous system development
5p14	rs72758936	*CDH9*	Cell-cell adhesion
6q13	rs1338321	*HTR1B*	Serotonin receptor activity
8p22	rs2739683	*ASAH1*	Small molecule metabolic process
1p35.3-p34.1	rs6699355	*DLGAP3*	Cell-cell signaling
5q11.2	rs184989476	*ANKRD55*	Unknown
8q24.21	rs10101440	*MYC*	Cell cycle arrest
2p12	rs72929232	*IGK*	Unknown
22q11.21	rs396130	*THAP7-AS1*	Non-coding RNA
14q31	rs186178322	*GALC*	Carbohydrate metabolic process
4q13.3-q21.1	rs62323371	*DCK*	Nucleotide biosynthetic process
3q11.2	rs1118907	*EPHA6*	Ephrin receptor activity
14q32.32	rs2494732	*AKT1*	Activation-induced cell death of T cells
14q23-q24.2	rs8018785	*SPTB*	Structural constituent of cytoskeleton
8q24.1	rs2196051	*ENPP2*	Immune response
5p15	rs139161838	*SLC12A7*	Cell volume homeostasis
15q21.2	rs77245046	*DMXL2*	Cell junction
2q14	rs11899778	*GLI2*	Cell proliferation

Locus	SNP	Closest Gene	Gene Function
2p11.2	rs145123753	*DNAH6*	Unknown
13q31.1	rs17705805	*SPRY2*	Branching morphogenesis of an epithelial tube