2020

Quantifying the heterogeneity of the immunoglobulin G N-Glycome in an ageing Australian population: The Busselton Healthy Ageing Study

Alyce Russell
Edith Cowan University

Edith Cowan University

Copyright Warning

Quantifying the Heterogeneity of the Immunoglobulin G N-Glycome in an Ageing Australian Population: The Busselton Healthy Ageing Study

This thesis is presented for the degree of

Doctor of Philosophy

Alyce Christine Russell

BSc(BiomedSc) MSc(HumBiol)

Edith Cowan University

School of Medical and Health Sciences

2019

Principal Supervisor:

Professor Wei Wang, MD PhD FFPH

Associate Supervisors:

Associate Professor Simon Laws, BSc(Hons) PhD

Doctor Ivo Ugrina, MSc PhD

Doctor Michael Hunter, BSc(Hons) PhD

Doctor Jennie Hui, BSc(Hons) PhD

USE OF THESIS

The Use of Thesis statement is not included in this version of the thesis.

Declaration

I certify that this thesis does not, to the best of my knowledge and belief:

(i) incorporate without acknowledgement any material previously submitted for a degree or diploma in any institution of higher education;

(ii) contain any material previously published or written by another person except where due reference is made in the text; or

(iii) contain any defamatory material.

I also grant permission for the Library at Edith Cowan University to make duplicate copies of my thesis as required.

Alyce Christine Russell

4th of December, 2019

Page intentionally left blank

Abstract

The use of immunoglobulin G N-glycomics to study chronic non-communicable disorders and other complex phenotypes emerged following the Human Genome Project. The consortium discovered that most phenotypes were too complex to be explained by genetics alone. Thus, the biological importance of epigenetics was recognised; heritable modifications to gene expression rather than the genome itself. N-glycosylation is a form of epigenetic regulation known as a post-translational modification. It stabilises the immunoglobulin G structure and alters downstream responses elicited by the antibody and is extensively studied as a candidate biomarker in the post-genomic era.

The N-glycosylation of immunoglobulin G itself is complex, with glycosyltransferases and glycosylhydrolases influencing the biosynthesis of the branching structures. Moreover, altered N-glycosylation is associated with an array of phenotypes. Our research team considers the N-glycome as an interphenotype of subclinical health status; an amalgamation of genetic predisposition, environmental exposure and health behaviours over the life-course. This underscores the value of the immunoglobulin G N-glycome in the shift towards predictive, preventive and personalised medicine. However, there is still considerable heterogeneity even among individuals with the same disorder, which warrants further investigation to improve precision of the biomarker.

This thesis aimed to determine the degree the underlying genome and clinical factors explain the heterogeneity of the immunoglobulin G N-glycome. I used a subset of the cross-sectional population-based Busselton Healthy Ageing Study (n=637, 54.0% female, 46.2 to 68.3 years of age). The participants represent a highly homogenous population (99% identify as Caucasian with Caucasian parents), and all non-institutionalised ‘Baby-boomers’ (adults born between 1946 and 1964) listed on the electoral roll in the City of Busselton between 2010 and 2016 were eligible to participate.

Three studies were designed to address the thesis aim. Firstly, previous IgG-related genetic polymorphisms were successfully validated using association studies of the N-glycan features. Secondly, next-generation sequencing of leucocyte mRNA was modelled with the N-glycome. Differentially expressed genes were identified, as well as the implementation of a multivariate model to integrate the ‘omics datasets. Finally, clinical factors and health behaviours were modelled using various statistics, extending on previous research. Collectively, however, the three studies evidenced potential utility of the immunoglobulin G N-glycome in identifying cardiometabolic disorders and associated risk factors. A polymorphism with genome-wide significance

had pleiotropy to type 2 diabetes mellitus. Additionally, the clinical studies correlated cardiometabolic risk factors (central adiposity, blood pressure, C-reactive protein, triglycerides, fasting blood glucose and insulin) as well as the health behaviours excessive alcohol consumption and current smoking status (both associated with increased risk of cardiometabolic disorders) to an increase in pro-inflammatory immunoglobulin G glycoforms, thus potentiating involvement of immunoglobulin G in the pathophysiology of these phenotypes. Overall, this data-driven thesis identified several factors explaining immunoglobulin G N-glycome heterogeneity. These should be considered in subsequent translational research, to improve the precision of this complex biomarker when stratifying populations of interest.

Statement of Contribution of Others

The thesis "*Quantifying the Heterogeneity of the Immunoglobulin G N-Glycome in an Ageing Australian Population: the Busselton Healthy Ageing Study*" includes a series of publications complemented by other chapters, for the Degree of Doctor of Philosophy by Alyce Russell, School of Medical and Health Sciences, Edith Cowan University, 2019. Below is a statement of contribution of others for the publications included.

Reference of Publications	**Contribution by Alyce Russell (%)**
Chapter 1	
Russell, A.C., Adua, E., Ugrina, I., Laws, S.M. and Wang, W. (2018). Immunoglobulin G Fc *N*-Glycosylation: Unravelling the Role of Glycosylation in Altering Effector Functions. *IJMS,* 19(2), 390. DOI: 10.3390/ijms19020390	AR 80% (literature review, figure development, chapter write-up, final proof)
Chapter 4	
Russell, A.C., Ugrina, I., Strbenac, D., Kepka, A., Laws, S.M., Akmacic, I.T., Wang, X., Li, X., Hui, J., Hunter, M., Yang, J.Y.H., Lauc, G. and Wang, W. (2019). Integrating mRNA Sequencing Data with the Immunoglobulin G *N*-Glycome Using Extensions of Differential Expression Analysis and Orthogonal Two-Way Partial Least Squares. *Submitted to Glycobiology 30/07/2019 for Peer Review.*	AR 65% (ran glycan laboratory work, prepared data for analysis, wrote first draft, ran majority of the statistics, performed all gene annotation and ontologies, final proof)

Chapter 5

Russell, A.C., Kepka, A., Akmacic, I.T., Hui, J., Hunter, M., Laws, S.M., Ugrina, I., Lauc, G. and Wang, W. (2019). Higher Levels of Central Body Fat are Associated with an Increase in Pro-Inflammatory Immunoglobulin G *N*-Glycans: The Busselton Healthy Ageing Study. *Immunobiology, 224(1), 110-115.* DOI: 10.1016/j.imbio.2018.10.002	AR 75% (glycan laboratory work, prepared data for analysis, wrote first draft, ran all statistics, final proof)
Russell, A.C., Kepka, A., Akmacic, I.T., Hui, J., Hunter, M., Laws, S.M., Ugrina, I., Lauc, G. and Wang, W. (2019). Identifying clinical factors associated with the immunoglobulin G glycome: a substudy of the Busselton Healthy Ageing Study. *Submitted to OMICS 17/03/2019 for peer review.*	AR 75% (glycan laboratory work, prepared data for analysis, wrote first draft, ran all statistics, final proof)

I, *Alyce Russell*, contributed to the above listed publications at the stated level.

Signed: [redacted] Date: 30/07/2019

I, as corresponding author, endorse that this level of contribution by the candidate indicated above is appropriate.

Signed: [redacted] Date: 30/07/2019

Prof. Wei Wang (Principal Supervisor and Corresponding Author)

Other Publications During This Thesis

Accepted Publications

Adua, E., Memarian, E., **Russell, A.**, Trbojević-Akmačić, I., Gudelj, I., Jurić, J., ... & Wang, W. (2019). High throughput profiling of whole plasma N-glycans in type II diabetes mellitus patients and healthy individuals: A perspective from a Ghanaian population. *Biomark Med.*, 13(15), 1273 – 1287.

Li, X., Wang, H., **Russell, A**., Cao, W., Wang, X., Ge, S., … & Wang, W. (2019). Type 2 Diabetes Mellitus is Associated with the Immunglobulin G N-Glycome through Putative Proinflammatory Mechanisms in an Australian Population. *OMICS, Online Ahead of Print.*

Adua, E., Memarian, E., **Russell, A.**, Trbojević-Akmačić, I., Gudelj, I., Jurić, J., ... & Wang, W. (2019). High throughput profiling of whole plasma N-glycans in type II diabetes mellitus patients and healthy individuals: A perspective from a Ghanaian population. *Arch. Biochem. Biophys.*, *661*, 10-21.

Liu, J. N., Dolikun, M., Štambuk, J., Trbojević-Akmačić, I., Zhang, J., Wang, H., … **Russell, A.**, … & Liu, D. (2018). The association between subclass-specific IgG Fc N-glycosylation profiles and hypertension in the Uygur, Kazak, Kirgiz, and Tajik populations. *J Hum Hypertens.*, *32*(8), 555.

Garcia, M., Downs, J., **Russell, A.**, & Wang, W. (2018). Impact of biobanks on research outcomes in rare diseases: a systematic review. *Orphanet J Rare Dis*, *13*(1), 202.

Adua, E., **Russell, A.**, Roberts, P., Wang, Y., Song, M., & Wang, W. (2017). Innovation analysis on postgenomic biomarkers: Glycomics for chronic diseases. *OMICS*, *21*(4), 183-196.

Russell, A. C., Šimurina, M., Garcia, M. T., Novokmet, M., Wang, Y., Rudan, I., ... & Wang, W. (2017). The *N*-glycosylation of immunoglobulin G as a novel biomarker of Parkinson's disease. *Glycobiology*, *27*(5), 501-510.

Yan, N., Zhou, Y., Wang, Y., Wang, A., Yang, X., **Russell, A.**, ... & Wang, W. (2016). Association of ideal cardiovascular health and brachial–ankle pulse wave velocity: a cross-sectional study in Northern China. *J. Stroke Cerebrovasc. Dis., 25*(1), 41-48.

Wang, Y.*, Adua, E.*, **Russell, A.***, Roberts, P., Ge, S., Zeng, Q., & Wang, W. (2016). Glycomics and its application potential in precision medicine. *Science*: AAAS.
* contributed equally to this work

Other Publications Under Peer Review

Russell, A.C., Ng, Z. Y., Anto, E. O., Adua, E., Mosdell, B. A., Hunter, M., & Wang, W. (2019). Reduced physical activity associates with poorer health status in Baby-boomers: an application of the Suboptimal Health Status Questionnaire in the Busselton Healthy Ageing Study, Australia. *Revisions submitted to EPMA 31/10/2019.*

Conference Presentations

Russell, A., Kepka, A., Akmacic, I. T., Hui, J., Hunter, M., Ugrina, I., Laws, S., & Wang, W. (2017). Higher levels of abdominal body fat are associated with an increase in pro-inflammatory immunoglobulin G N-glycans: results from the Busselton Healthy Ageing Study. *Poster and oral presentation at the 12th Jenner Glycobiology and Medicine Symposium Translational Glycobiology: From Bench to Bedside, May 6-9, 2017, Dubrovnik, Croatia.*

(Attendees with Glycobiology Expertise)

Russell, A., Kepka, A., Akmacic, I. T., Hui, J., Hunter, M., Ugrina, I., Laws, S., & Wang, W. (2018). Identifying Clinical Factors Associated with Heterogeneity in the IgG Glycome. *Poster presented at Science on the Swan, May 1-3, 2018, Fremantle, Australia.*

(Attendees with Public Health Expertise)

Russell, A., Kepka, A., Akmacic, I. T., Hui, J., Hunter, M., Ugrina, I., Laws, S., & Wang, W. (2018). Identifying Clinical Factors Associated with Heterogeneity in the IgG Glycome. *Poster presented at the 10th Western Australian Young Statisticians Workshop, September 26, 2018, Perth, Australia.*

(Attendees with Statistical Expertise)

Acknowledgements

This thesis has certainly been an experience. I am forever thankful to everyone who contributed to my research journey as well as those who have supported me through my private struggles. Specifically, I would like to thank:

- The participants at the Busselton Healthy Ageing Study who continue to support medical research by allowing their data to be collected.
- Prof Gordan Lauc, Dr Irena Trbojevic Akmacic and Dr Lucija Klaric for their expert mentorship and guidance in glycomics during my month-long visit to their laboratory, Genos Ltd, in Zagreb, Croatia, whilst analysing my samples.
- The Sydney Precision Bioinformatics Group at the University of Sydney for guiding me through the RNA-Seq alignment pipeline, as well as exposing me to normalisation procedures and different statistical methods within other 'omics fields. Your time and patience with me were appreciated, and I am humbled to have been welcomed into a world-class team for this period.
- My supervisor Dr Ivo Ugrina for his commitment to solving my analysis pipeline issues, brainstorming statistical approaches and helping to keep me accountable. I know this has been a very tough journey for you too. I could not have done this without your help, so your contributions are appreciated.
- My supervisor's Dr Jennie Hui, Dr Michael Hunter and A/Prof Simon Laws for their contributions with organising samples, quotes, analyses and just generally answering any questions along the way.
- My ECU research colleagues for always offering a helping hand with their expertise when it is needed, particularly Dr Siqi Ge, Dr Xinwei Yu, Monique Garcia, Xingang Li, Xueqing Wang, Belinda Mosdell, A/Prof Manshu Song, Dr Eric Adua and Enoch Anto. I hope I can return the favour someday.
- And to my principal supervisor Prof Wei Wang, who has nurtured my crazy yet semi-focused research vision for over half a decade, and supported my trips aimed at developing my statistical programming skills and collaborations.

On a more personal note, I would like to acknowledge those closest to me who still stick by me through this crazy career path and are always by my side when I need them the most. This journey has seen a lot of personal hurdles and I could not have done it without my support group. Get. It. Carrots!

Finally, I would like to acknowledge my beautiful daughter Sophie, whom has never known her mum not to research. I hope you take from this that you can do whatever you put your mind to, no matter what hurdles life throws at you. I love you.

Page intentionally left blank

Abbreviations

ADCC	Antibody-dependent cell cytotoxicity
A/G ratio	Android/gynoid ratio
Asn	Asparagine
BCR	B-cell receptor
BHAS	Busselton Health Ageing Study
BMI	Body mass index
CCA	Canonical correlation analysis
CPM	Counts per million
CRP	C-reactive protein
DAG	Diacylglycerol
DC-SIGN	Dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin
DBP	Diastolic blood pressure
DXA	Dual-energy X-ray absorptiometry
ER	Endoplasmic reticulum
Fab	Fragment antibody-binding
FBG	Fasting blood glucose
Fc	Fragment crystallisable
FcR	Fragment crystallisable receptor
FcRn	Neonatal FcR
FFQ	Food frequency questionnaire
Fuc	Fucose
Gal	Galactose
Glc	Glucose
GlcNAc	*N*-acetylglucosamine
GO	Gene Ontology
GP	N-glycan peak
gQTL	Quantitative trait loci of N-glycosylation
GWAS	Genome-wide association study
HDL	High-density lipoprotein
HILIC-SPE	Hydrophilic interaction chromatography - solid phase extraction
IC	Immune complex
IgG	Immunoglobulin G
IL-6	Interleukin-6
IP3	Inositol 1, 4, 5-triphosphate
IPAQ	International Physical Activity Questionnaire
ITAM	Immunoreceptor tyrosine-based activation motif
ITIM	Immunoreceptor tyrosine-based inhibitory motif
LDL	Low-density lipoprotein

MAC	Membrane attack complex
Man	Mannose
MAPK	Mitogen activated protein kinase
MBL	Mannose-binding lectin
MDS	Multidimensional scaling
MET-minutes	Metabolic equivalence minutes
Neu5Ac	N-acetylneuraminic acid (aka sialic acid)
O2PLS	Orthogonal two-way partial least squares
PI-3K	Phosphatidylinositol 3-kinase
PIP2	Phosphatidylinositol 4, 5-bisphosphate
PIP3	Phosphatidylinositol-3, 4, 5-triphosphate
PKC	Protein kinase C
PLC-γ	Phospholipase C-gamma
PTEN	Tensin homologue
RNA-Seq	RNA sequencing
Ser	Serine
SH2	Src homology 2
SHIP	SH2 inositol 5-phosphate
SNP	Single nucleotide polymorphism
SBP	Systolic blood pressure
TC	Total cholesterol
TG	Triglycerides
Thr	Threonine
UPLC	Ultra-performance liquid chromatography
WCC	White cell count
WHR	Waist-to-hip ratio
WHtR	Waist-to-height ratio

Table of Contents

Page intentionally left blank

List of Tables

List of Figures

Page intentionally left blank

Chapter 1 – The Immunoglobulin G N-Glycome: An Innovative Biomarker in the Post-Genomic Era

1.1 Prologue

The study of immunoglobulin G (IgG) N-glycomics began in the late 1980s when it was noted that those with rheumatoid arthritis had markedly higher levels of agalactosylated IgG structures (1). Currently, research in this area focuses on many morbidities and phenotypes, including some of my own research (2-5). As part of my MSc project, IgG N-glycans were analysed in a case-control study whereby Parkinson's disease was the morbidity of interest (6). Although we found some interesting results, including evidence of accelerated ageing within the Parkinson's cases, and a pilot model of classifying between Parkinson's and age- and sex-matched controls (6), the study also underscored the need to understand what was driving changes within the IgG N-glycome. Similar heterogeneity has been reported elsewhere, particularly with reference to age, sex, and geographical location (7-10). Indeed, the study demonstrated that even within the cases alone, the results were highly heterogeneous (**Figure 1.1**). Thus, to address this gap in the literature, I explored the use of different types of biological data to model variation in the IgG N-glycome.

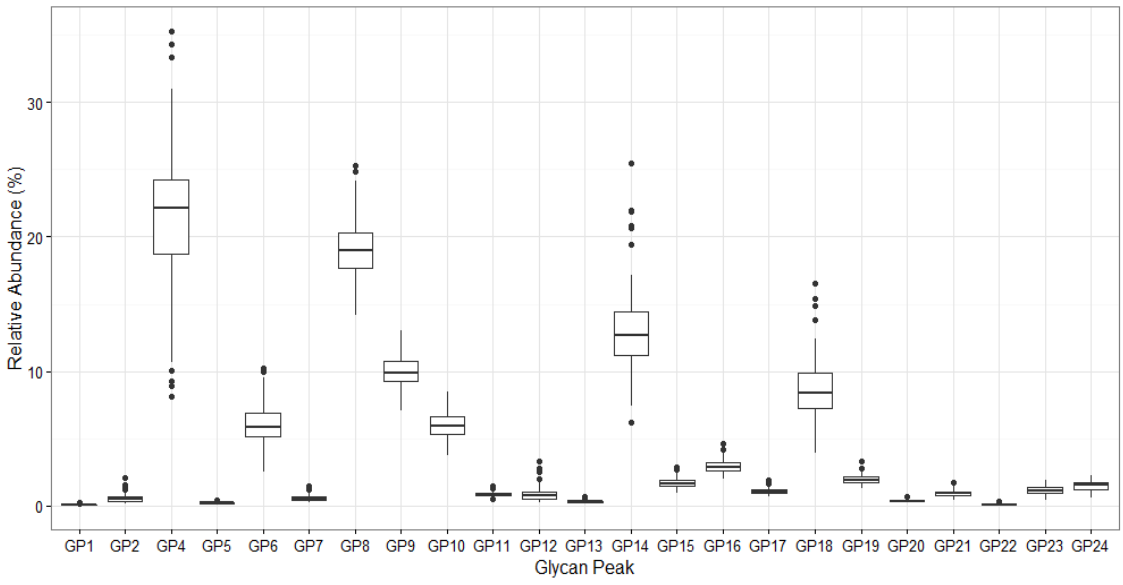


Figure 1.1: *Boxplots of the abundance of IgG N-glycans within the Parkinson's disease case of my MSc project (based on data presented in Russell et al, 2017)*

This current chapter presents the thorough evaluation of the literature. This was conducted at the inception of this thesis and subsequently published as a review paper. This review provides the foundation upon which the hypothetical framework of the thesis was developed. Subsequent chapters present the body of work undertaken throughout the course of this thesis to address the specific aims and hypotheses presented in **Section 1.3**.

1.2 Review Paper: Unravelling Immunoglobulin G Fc N-Glycosylation: A Dynamic Marker Potentiating Predictive, Preventive and Personalised Medicine

Alyce C. Russell, Eric Adua, Ivo Ugrina, Simon M. Laws, and Wei Wang (2018). *Int. J. Mol. Sci.*, 19(2), 390.

1.2.1 Abstract

Multiple factors influence IgG glycosylation, which in turn affect the glycoproteins' function on eliciting an anti-inflammatory or pro-inflammatory response. It is prudent to underscore these processes when considering the use of IgG N-glycan moieties as an indication of disease presence, progress, or response to therapeutics. It has been demonstrated that the altered expression of genes that encode enzymes involved in the biosynthesis of IgG N-glycans, receptors, or complement factors may significantly modify IgG effector response, which is important for regulating the immune system. The IgG N-glycome is highly heterogeneous; however, it is considered an interphenotype of disease (a link between genetic predisposition and environmental exposure) and so has the potential to be used as a dynamic biomarker from the perspective of predictive, preventive, and personalised medicine. Undoubtedly, a deeper understanding of how the multiple factors interact with each other to alter IgG glycosylation is crucial. Herein we review the current literature on IgG glycoprotein structure, IgG Fc glycosylation, associated receptors, and complement factors, the downstream effector functions, and the factors associated with the heterogeneity of IgG glycosylation.

1.2.2 Introduction

IgG is an important effector glycoprotein linking the innate and adaptive branches of the immune system. It has the ability to exert both anti-inflammatory and pro-inflammatory responses throughout the body, which are triggered by antigen recognition, and are dependent on its affinity for a number of different activating or inhibitory fragment crystallisable receptors (FcRs) and complement factors. This process mediates key effector functions, including pathogen clearance, antibody-dependent cell cytotoxicity (ADCC), and complement-initiated inflammation, all with both beneficial and detrimental effects depending on the premise of the IgG activity. For example, during primary bacterial infection, IgG can initiate opsonisation through

complement activation and phagocytosis of the bacterial cells by macrophages, monocytes and neutrophils, as well as neutralise endotoxins and exotoxins (11-13). These beneficial effects are well-established and have been harnessed in the form of intravenous immunoglobulin therapy in immunodeficient individuals (14). On the contrary, there are examples where these effects are detrimental, such as in rheumatoid arthritis patients (15-17). In this instance, IgG is thought to tandemly bind synovial cells and mannose-binding lectin (MBL), resulting in the initiation of the lectin complement cascade and secondary damage of surrounding tissues within the synovial joints (15-17). These immune responses are largely modulated by the Fc domain of the IgG glycoprotein (**Figure 1.2**).

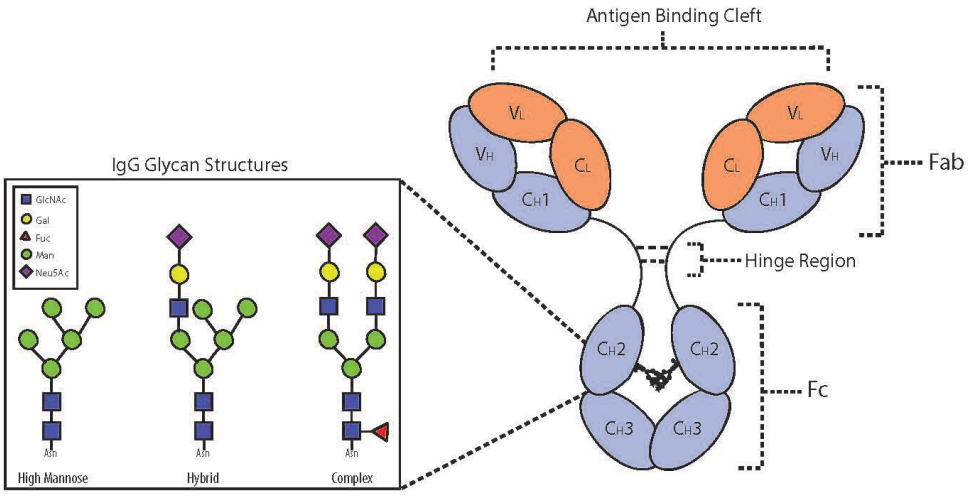


Figure 1.2: *IgG and the associated N-glycan structures. Consisting of both heavy chains ('$_H$') and light chains ('$_L$'), the IgG glycoprotein has two domains that infer different properties; the fragment antigen-binding (Fab) and fragment crystallisable (Fc) domains. The Fab and Fc are connected by a hinge region containing disulphide bonds, which differ depending on the IgG subclass. The Fab domain is responsible for recognising and binding antigen. The Fc domain contains two glycans attached to conserved regions of the C_H2. The Fc domain elicits effector functions by binding Fc receptors (FcRs) on natural killer and other inflammatory cells. Changes to the attached glycan moieties, simplified as the three glycan structural types in the side figure, can significantly alter effector functions of the IgG glycoprotein. GlcNAc—N-acetylglucosamine. Gal—galactose, Fuc—core fucose, Man—mannose, Neu5Ac—N-acetylneuraminic acid (sialic acid)*

It has been established that the IgG Fc sugar moieties, hereon known as N-glycans, affect the affinity of the Fc domain for a number of different FcRs, ultimately initiating different cellular events that lead to an array of inflammatory responses (10). The glycosylation of the IgG Fc domain may be a predesigned outcome of the producing B cell, and variation in IgG glycosylation has physiological significance (10, 18). Different inflammatory factors influence B cells during activation and differentiation, and can modulate the glycosylation of secreted IgG without disturbing the general cellular glycosylation machinery (19). Furthermore, glycosylation changes in the IgG glycome are thought to be fairly stable over short periods of time, and modifications can result from biological and chronological age (9), as well as altered cellular environment and disease status (18, 20).

Therefore, the N-glycan moieties of IgG represent intermediate phenotypes of subclinical and chronic disease. They are a link between the genetic makeup of our cells and the cellular environment, which is largely under the influence of our habits and daily routines (20). With so much interest in utilising the IgG glycome in the pursuit for predictive, preventive and personalised medicine, this begs the question: are the genetic changes more significant when it comes to elucidating aberrant IgG glycosylation or are they irrelevant when compared with the contribution of the cellular environment? This paper reviews the current literature on the IgG glycoprotein and its FcRs, as well as the factors associated with altered IgG glycosylation.

1.2.3 IgG Glycoprotein Structure and Function

The immunoglobulins are a group of plasma glycoproteins, among which IgG is the most abundant (16). The protein portion of IgG consists of four polypeptide chains, two identical light chains and two identical heavy chains (**Figure 1.2**), which are held together by inter-peptide disulphide bonds (21). There are two types of light chains, kappa and lambda, with their genes located in 2q12 and 22q11, respectively, and one type of heavy chain with its gene located in 14q32 (22). Intra-peptide disulphide bonds also exist and cause the formation of loops, which link the anti-parallel β-sheets in the tertiary structure of IgG (21). There are two functionally distinct domains of the IgG glycoprotein: the fragment antigen-binding (Fab) and the Fc domains. The Fab domain contains the variable (V_L and V_H) and constant (C_L and C_H1) regions of the light and heavy chains. These form the antigen-binding sites, with the variable regions recognising and distinguishing specific molecular structures of antigens (21). On the contrary, the Fc

domain contains constant regions (C_H2 and C_H3) and mediates key effector functions (23).

Antigens are diverse; thus, the repertoire of IgG Fab domains must also be diverse to target these ubiquitous molecular structures. This is achieved during B cell development through V(D)J shuffling (also known as somatic recombination) of the multiple gene segments scattered along the locus of a given IgG polypeptide (21). Introduced somatic mutations are also important for Fab polyclonality (24). The variable regions V_H and V_L, primarily responsible for antigen specificity, are made up of highly variable subregions involved in the antigen-binding activity (24). These are known as complementary-determining regions (CDR1 to CDR3), and are flanked by less variable regions known as framework regions (FR1 to FR4). Proximal to the hinge on the Fab domain are the constant regions C_L and C_H1. The antigen-binding sites are the culmination of these subregions. Nevertheless, even with the proper recognition of specific antigens, the IgG cannot mount an effector response without the glycosylation of the Fc domain (25).

There are four different subclasses of IgG (IgG1 to IgG4). These subclasses arise from alternative amino acid sequences translated from the singular heavy chain gene (14q32) (21) and they differ in the length of their hinge region and the number of disulphide bonds present. These variations result in an altered affinity for a number of different FcRs or complement factors and, therefore, different effector functions (24). Moreover, the effector functions are influenced by the N-glycan moieties attached to both C_H2 regions. These interact with each other to change the conformation of the Fc domain (**Figure 1.2**). It is possible that a single altered monosaccharide on an N-glycan moiety can lead to a conformational change that is inhibiting rather than activating, and may even change the type of FcR or complement factor bound by the Fc domain.

Though beyond the scope of this review, it should be noted that more recently the importance of Fab glycosylation, which is evident on 15–20% of circulating IgG, has also been reported (24, 26). Moreover, three recognition sites specific for O-glycosylation exist on the hinge region of IgG3. The IgG3 subclass constitutes approximately 8% of the IgG glycoprotein pool, with only 10–13% expected occupancy of any of these three O-glycosylation sites (27).

1.2.4 N-Glycosylation of the Fc Domain

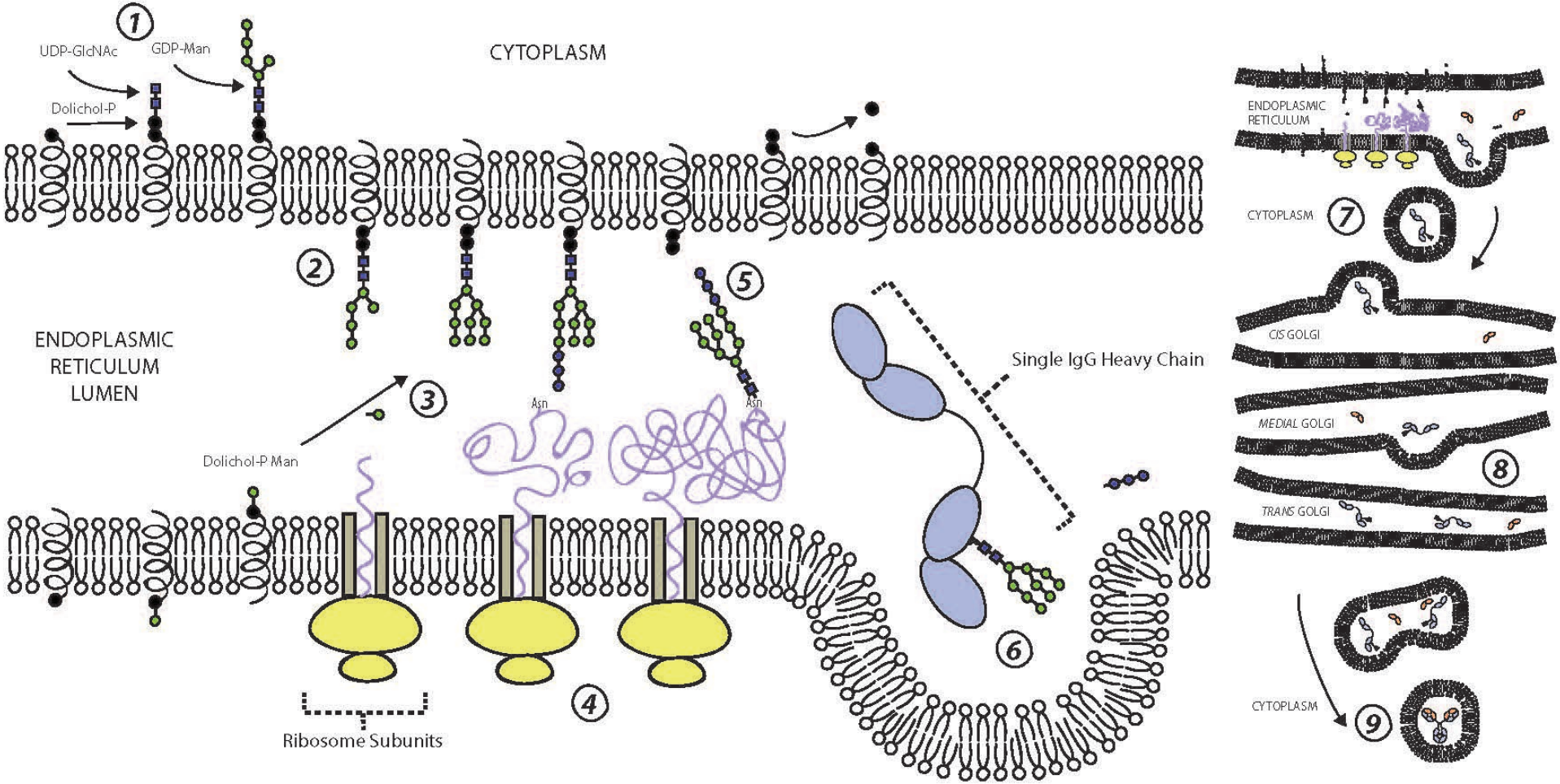
An N-glycan is covalently attached to the side chain nitrogen atom of the highly conserved asparagine (Asn)-297 of both heavy chains (**Figure 1.2**). This occurs at the

recognition sequence Asn–X–Serine (Ser)/Threonine (Thr), located in the C_H2 domains, whereby X is any amino acid except proline (16, 20). The majority of these N-glycans are complex-type biantennary structures (**Figure 1.2**); however, a high degree of heterogeneity exists due to the presence of different monosaccharides (20). In total, over 36 IgG glycoforms have been identified (18, 28, 29) and the two N-glycan moieties that attach to a single IgG in the Fc domain may vary, leading to greater variability in effector function (30). These N-glycans are thought to not only stabilise the quaternary structure of the IgG Fc, but also contribute to recognition of the target FcRs (16, 31). Importantly, IgG lacking Fc N-glycans may have a conformation which may hinder FcR binding (20, 25). This is due to the Fc glycan moieties interacting with and stabilising the C'E loop, the region of the C_H2 domains containing the recognition sequence (11). Though on the contrary, recent research has demonstrated the capacity for therapeutic aglycosylated IgG to be engineered with Fc binding affinity (32).

N-glycans are biosynthesised sequentially (**Figure 1.3**) and in tandem to protein biosynthesis, commencing on the membrane of the endoplasmic reticulum (ER). Here, sugar nucleotides assemble onto a dolichol pyrophosphate donor molecule on the cytoplasmic side of the ER membrane to form a mannose $(Man)_5$ N-acetylglucosamine $(GlcNAc)_2$oligosaccharide (24). The structure is subsequently inverted to be on the luminal side of the membrane and extended by a further four Man and three glucose (Glc) sugar nucleotides before the dolichol-bound $Glc_3Man_9GlcNAc_2$oligosaccharide, termed a precursor N-glycan, is transferred to an Asn recognition sequence by oligosaccharyltransferase (20). The HC and LC polypeptides are then folded and the nascent IgG assembled into its respective form. Following removal of the three terminating Glc via two trimming steps, the glycoprotein with attached $Man_9GlcNAc_2$oligosaccharides is transferred to the cis portion of the Golgi apparatus for further glycoprocessing (33, 34). Different glycosylhydrolases and glycosyltransferases remodel the branched IgG glycan structure throughout the glycoprocessing steps, from the cis Golgi to the trans Golgi (35). Variation in these enzymatic processes results in diversification of the IgG N-glycome composition.

The quaternary structure of the IgG Fc is thought to limit accessibility of the N-glycan moieties to glycosyltransferases in the Golgi, thus limiting the variability of galactosylation and sialylation (16). It has also been argued that intramolecular

Figure 1.3: *Biosynthesis of the IgG glycoprotein. ① N-glycan biosynthesis begins on the endoplasmic reticulum membrane where two N-acetylglucosamines (UDP-GlcNAc) and five mannose (GDP-Man) sugar nucleotides contribute the given monosacchairdes to a dolichol-phosphate (Dolichol-P) donor molecule on the cytoplasmic side. ② The whole Dolichol-P attached N-glycan is flipped so it is on the luminal side of the endoplasmic reticulum. ③ Dolichol-P flips individual sugar nucleotides from the cytoplasm to the lumen. ④ In tandem, the ribosomes biosynthesise the polypeptide structure of the IgG. ⑤ Oligosaccharyltransferase transfers the N-glycan moiety from Dolichol-P to asparagine (Asn) 297 on the growing polypeptide. ⑥ The three terminal N-acetylglucosamine sugar nucleotides are removed over two steps, following folding and assembly of the nascent IgG, and this signals the transfer of the IgG from the endoplasmic reticulum to the Golgi apparatus. ⑦,⑧ The IgG components move through the Golgi apparatus, where different glycosyltransferases and glycosylhydrolases add and remove respectively, different sugar nucleotides to the glycan moiety. ⑨ Following excretion from the trans Golgi, the final IgG glycoprotein either secreted from the B-cell lymphocyte or attached to the plasma membrane of the B-cell to become a B-cell receptor (BCR)*


1
UDP-GlcNAc
GDP-Man
Dolichol-P
CYTOPLASM
2
5
ENDOPLASMIC
RETICULUM
LUMEN
3
Asn
Asn
Dolichol-P Man
Single IgG Heavy Chain
4
6
Ribosome Subunits
ENDOPLASMIC
RETICULUM
CYTOPLASM
7
CIS GOLGI
MEDIAL GOLGI
8
TRANS GOLGI
CYTOPLASM
9

interactions, which increase as the N-glycan moieties are extended, restrict glycoprocessing (36). This constraint to diversify may be essential for the regulation of effector functions. Subsequently, the final glycosylated IgG is secreted into the blood serum or assimilated into the plasma membrane to form B cell receptors (BCRs).

The resulting N-glycans on a given IgG Fc are a predesigned outcome that depends on the relative expression of the genes that encode the specialised enzymes, particularly the glycosyltransferases, and the availability of the sugar nucleotides within the producing B cell lymphocyte (21). These are clearly affected by age, pregnancy, hormones, cytokines, bacterial DNA, and food metabolites (21). Differences in glycosylation pattern result in altered binding affinities for the many FcRs and complement factors, which influence effector response (16, 21, 25). A shift towards certain types of IgG N-glycans has been reported for a number of diseases and conditions, including rheumatoid arthritis (37, 38), metabolic syndrome and type 2 diabetes mellitus (18, 39), inflammatory bowel disease (18), systemic lupus erythematosus (40), hypertension (41), various cancers (18, 42), and more recently neurological disorders such as Alzheimer's disease and progressive mild cognitive impairment (43), multiple sclerosis (44), and Parkinson's disease (6). This makes the study of the N-glycan moieties biologically important in terms of the physiological activity of IgG in the human body. Indeed, even alterations to the IgG N-glycan moieties that appear structurally minute can significantly change its affinity to a number of different FcRs and complement factors (summarised in **Figure 1.4**).

The addition of N-acetylneuraminic acid (Neu5Ac), more commonly referred to as sialic acid, to the terminating end of the IgG Fc N-glycan has a similar anti-inflammatory effect to the addition of core fucose (Fuc), i.e., fucosylation of the Asn-attached core GlcNAc (31, 45). Sialylated IgG Fc N-glycans may have a lower affinity for FcγRIIIa on natural killer cells leading to an anti-inflammatory effect (16). Sialylated, particularly α2-6 sialylated, IgG Fc N-glycans exert their anti-inflammatory effect by binding to a type II FcR known as dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN), a C-type lectin which is present on dendritic cells (23). Sialylation results in increased conformational flexibility of IgG Fc with a preference to engage these types of FcR (23), though the target FcRs of sialylated IgG have more recently been debated (31). For example, an FcR-independent reduction in pro-inflammation also occurs due to inhibition binding of IgG Fc to C1q (24, 46). Interestingly, it has been

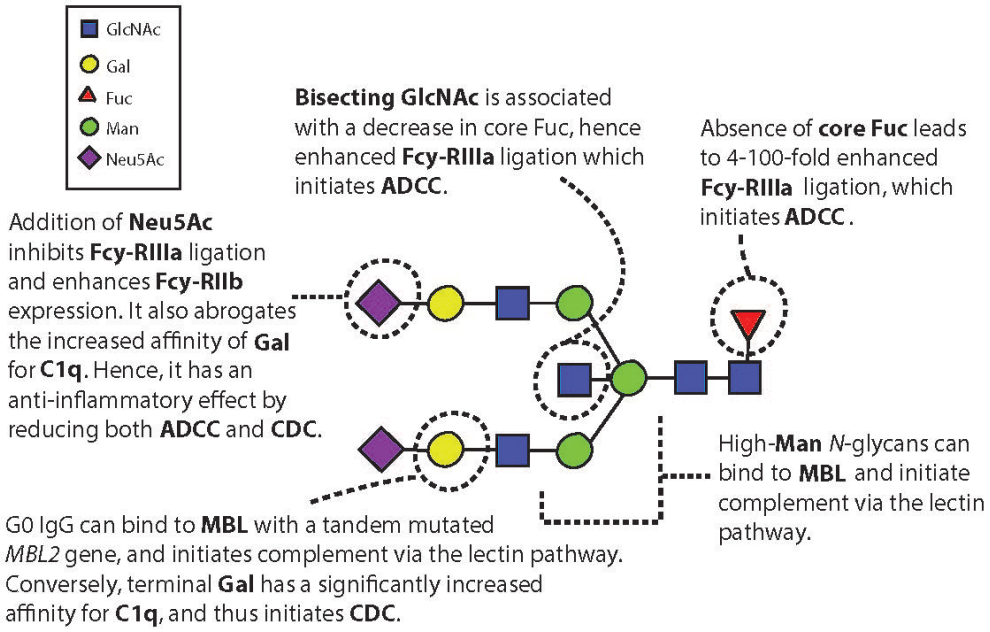


Figure 1.4: *A picture summary of altered IgG glycosylation and its downstream effects. GlcNAc—N-acetylglucosamine. Gal—galactose, Fuc—core fucose, Man—mannose, Neu5Ac—N-acetylneuraminic acid (sialic acid), ADCC—antibody-dependent cell cytotoxicity, CDC—complement-dependent cytotoxicity*

found that B-cell-independent sialylation of IgG exists, as it can be regulated by the release of ST6Gal1 from the cells lining central veins in the liver, and by circulating nucleotide sugar donor CMP-Neu5Ac, which degranulating platelets partially contribute to (47). Autoantibodies produced during autoimmune or inflammatory diseases frequently lack terminating galactose (Gal), and therefore Neu5Ac, which decreases the inhibition of ADCC (23). In vitro and in vivo studies have also demonstrated that autoantibodies lacking sialylation enhance osteoclastogenesis, providing a possible mechanism for the drastically increased levels of agalactosylated IgG structures found in rheumatoid arthritis patients (48). However, it has not been reported whether its absence leads to such a high increase in ADCC, as seen with the lack of core Fuc.

Out of all the N-glycosylation traits, core Fuc may be the most important in terms of effect modification in affinity to FcRs (35). A lack of core Fuc leads to a 4 to 100-fold increase in ADCC due to increased binding affinity for FcγRIIIa and FcγRIIIb, making these IgG more pro-inflammatory (31, 35, 49, 50). Conversely, the presence of Fuc on the Asn-attached core GlcNAc of the IgG Fc N-glycan has an anti-inflammatory effect by inhibiting the binding of type I FcRs, particularly FcγRIIIa (23, 49). FcγRIIIa are primarily expressed on the surface of natural killer cells (10) and have a role in initiating ADCC

(49). Previous studies (10, 18) have shown that the vast majority of IgG Fc are comprised of at least one core fucosylated N-glycan moiety, hence they are designed to be less efficient at ADCC activation subsequent to many IgG binding to an antigen (via the antigen-binding sites; see **Figure 1.2**). However, structural analysis of the IgG Fc/FcγRIIIa complex has demonstrated that specific N-glycans on FcγRIIIa at Asn-162 may also influence the effect of the absence of core Fuc (31, 51), due to direct glycan-glycan interactions (16). The presence of a bisecting GlcNAc has also been associated with enhanced ADCC activation via FcγRIIIa binding (52) but to a lesser degree than the absence of core Fuc (49).

It was first suggested that Man-rich N-glycans have an increased affinity for MBL, which initiates the lectin complement cascade via opsonisation of the target cells by IgG (15). Both Man and GlcNAc (GlcNAc to a lesser degree than Man) are suggested to be prominent ligands for MBL, and are usually found on the surface of pathogens, such as bacteria and viruses (53). Other monosaccharides, such as Neu5Ac and Gal, normally decorate mammalian glycoproteins and have undetectable affinity for MBL (15, 53). This enables the specific recognition of N-glycans on pathogen cell surfaces. However, this innate immunity system can malfunction, and has been implicated in many disease processes due to an increase in agalactosylated (Man-rich or GlcNAc terminating) N-glycans. Importantly, an increase in agalactosylated IgG is associated with advancing age, thus we may be more at risk for developing immune-related diseases as we age (5, 10, 54). Therefore, agalactosylated IgG are considered pro-inflammatory.

In vitro studies have demonstrated a link between agalactosylated IgG and increased MBL binding (15, 30). After multiple recent in vivo studies, however, it is still unclear whether the mechanism for complement activation is via MBL, as first hypothesised or another complement factor (16). For example, MBL-null and wild-type mice with agalactosylated IgG were found to have no difference in potential to activate inflammation (30). Indeed, evidence suggests that galactosylation in the absence of sialylation significantly increases complement-dependent cytotoxicity via increased C1q binding, activating the classical complement cascade (46, 55), highlighting another IgG glycan-modulated complement pathway. Of note, MBL2 polymorphism may also be required for the association of agalactosylated IgG and activation of the lectin complement cascade (37).

1.2.5 IgG Fc Structure and Function

The IgG Fc domain has the ability to bind to a range of different FcRs and complement factors, resulting in distinct effector and immunomodulatory pathways. Monomeric IgG

mediates these effects, particularly by binding to FcγRI (24, 56). Nevertheless, the formation of immune complexes (IC) is a prerequisite for efficient FcR binding (56). FcR binding requires the presence of the Fc N-glycan moieties; however, in some cases, the N-glycans themselves may not interact with the FcRs as in the case of type II FcRs for example (23). This suggests that N-glycan moieties modulate the FcR affinity by altering the conformational state of the Fc domain.

There are two dominant conformational states that the Fc domain may exhibit, and two structurally distinct sets of FcRs with selective abilities to engage the Fc domain. Type I FcRs belong to the immunoglobulin receptor superfamily, and include: the activating FcRs, FcγRI, FcγRIIa, FcγRIIc, FcγRIIIa, and FcγRIIIb; and the inhibitory FcR, FcγRIIb (23, 24, 57). Specifically, these FcRs bind to the Fc domain of IgG in the open conformation, neighbouring the hinge region with a stoichiometry of 1:1 (11, 23, 58). Whereas type II FcRs belong to the C-lectin receptor family, and include: CD209, which is also known as DC-SIGN; and CD23, which is an IgE FcR (FcεRII) that also has weak binding affinity for IgG (23). These FcRs bind to the Fc domain in the closed conformation, which occludes the type I FcR binding site and permits type II FcR attachment in the region where C_H2 and C_H3 join with a stoichiometry of 2:1 (23). The cellular events following binding of these FcRs are important, and the different effector functions elicited, such as the transcription of cytokines, may have a direct impact on the future glycosylation of IgG (summarised in **Table 1.2**). It should be noted that IgG Fc also ligates neonatal FcR (FcRn), which increases the half-life of serum IgG (14, 30).

1.2.5.1 Coengagement of Activating and Inhibitory FcRs

The activating and inhibitory FcRs exist on many different cell types in varying quantities, depending on the primary physiological response of the cell. IC that are composed of differing Fc conformations will simultaneously interact with either activating or inhibitory type I or type II FcRs on a single immune cell (23). Activating FcRs are present on all myeloid cells, with cross-linking resulting in sustained effector response (45). Catalysed by tyrosine kinases, activating FcRs mediate IgG immunomodulatory response by undergoing phosphorylation of an immunoreceptor tyrosine-based activation motif (ITAM) present in its cytoplasmic tail (24).

Page intentionally left blank

Table 1.1. Summary of FcRs ligated by IgG. *Regulation: either up- or down-regulation of the expression of the FcR on a given cell. Produces: the cytokines released following binding with the given FcR. Effects: summary points from the literature*

Receptor	Classification	IgG Affinity	Cell Types	Regulation	Produces	Effects	Ref
FcγRI (CD64)	Type I Activating	**High** for IgG1, 3, 4 **Low** or **No** affinity for IgG2	Mast cells, Monocytes, Macrophages, Neutrophils, Dendritic Cells	↑ IL-10, INF-γ ↓ IL-4	IL-6 TNF-α	↑ B Cell Differentiation, Immunoglobulin Production, Acute-Phase Protein Synthesis ↑ ADCC, Degranulation, Phagocytosis (depends on expressing cell) – through SRC-family kinase phosphorylation ↓ Lipo-Protein Lipase, Fcγ-RIIIb ON Neutrophils	Akira (1990) (59) Daëron (1997) (60) Fanger (1997) (61) Lu (2015) (31) Pincetic (2014) (23) Ravetch (2001) (57) Quast (2017) (24) Rogers (2006) (62) Van de Winkel (1991) (63) Woolhiser (2001) (64)
FcγRIIa1/2 (CD32a)	Type I Activating	**Low**	Mast Cells, Monocytes, Macrophages, Neutrophils, Dendritic Cells, Eosinophils, Basophils, Platelets	↓ IL-4	TNF-α	↑ ADCC, Degranulation, Phagocytosis (depends on expressing cell) – through SRC-family kinase phosphorylation	Daëron (1997) (60) Karsten (2012) (56) Pincetic (2014) (23) Ravetch (2001) (57) Van de Winkel (1991) (63)

Page intentionally left blank

Table 1.1. *cont.*

Receptor	Classification	IgG Affinity	Cell Types	Regulation	Produces	Effects	Ref
FcγRIIb1/2/3 (CD32b)	Type I Inhibitory	**Low** ↓ with F ↑ with G	T Cells, NK Cells, Immature B Cells (only FcR)	↑ IL-4		↓ BCR-Induced Ca^{2+} Mobilisation, Cell Proliferation, ITAM-Regulated FcRs, Akt Limits Autoantibody Production (on B Cells)	Karsten (2012) (56) Quast (2017) (24) Subedi (2016) (35) Unkeless (1997) (65) Vivier (1997) (66)
FcγRIIc (CD32c)	Type I Activating	**Low** ↓ with F	Monocytes, Neutrophils, NK Cells	↓ IL-4	TNF-α	↑ ADCC, Degranulation, Phagocytosis (depends on expressing cell) – through SRC-family kinase phosphorylation	Daëron (1997) (60) Karsten (2012) (56) Pincetic (2014) (23) Ravetch (2001) (57) Subedi (2016) (35) Van de Winkel (1991) (63)
FcγRIIIa (CD16a)	Type I Activating	**Medium** ↓ with F, S	NK Cells, Macrophages, Monocytes (10%)	↓ IL-4	TNF-α	↑ ADCC	Bryceson (2011) (67) Daëron (1997) (60) Pincetic (2014) (23) Ravetch (2001) (57) Rogers (2006) (62) Subedi (2016) (35) Van de Winkel (1991) (63)

Page intentionally left blank

Table 1.1. *cont.*

Receptor	Classification	IgG Affinity	Cell Types	Regulation	Produces	Effects	Ref
FcγRIIIb (CD16b)	Type I Activating	**Low** ↓ with F	T Cells, Neutrophils, Macrophages, Monocytes	↑ INF-γ ↓ TNF-α		↑ Ca^{2+} Mobilisation	Fernandes (2005) (68) Hořejší (1999) (69) Pincetic (2014) (23) Ravetch (2001) (57) Subedi (2016) (35) Wirthmueller (1992) (70)
FcεRII (CD23)	Type II Inhibitory	**?** ↑ with S	B Cells, T Cells, Monocytes, Follicular Dendritic Cells, Macrophages			↑ Production FcγRIIb - Inhibits further activating FcγR ligation	Pincetic (2014) (23) Sondermann (2013) (58)
DC-SIGN (CD209)	Type II Inhibitory	**?** ↑ with S	B Cells, Monocytes, Dendritic Cells, Macrophages	↑ IL-4, IL-33	IL-4	↑ Production FcγRIIb - Inhibits further activating FcγR ligation	Anthony (2011) (71) Pincetic (2014) (23) Sondermann (2013) (58) van Kooyk (2003) (72) van Kooyk (2008) (73)

Phosphorylated ITAM becomes a binding site for Src homology 2 (SH2) (74), and this complexation triggers a signalling cascade that begins with the activation of Syc kinases. Activated Syc kinases phosphorylate and activate phosphatidylinositol 3-kinase (PI-3K). This then stimulates other signalling intermediates such as anti-apoptotic kinase (Akt), mitogen activated protein kinase (MAPK), and phospholipase C-gamma (PLC-γ). In particular, activated PLC-γ influences the dissociation of phosphatidylinositol 4, 5-bisphosphate (PIP2) into inositol 1, 4, 5-trisphosphate (IP3) and diacylglycerol (DAG). These two metabolites perform distinct roles in the signalling pathway. While DAG activates protein kinase C (PKC), IP3 triggers Ca^{2+} mobilisation from tissues and promotes cell proliferation (74).

Conversely, inhibitory FcRs also exist and act to dominantly dampen activating the immunomodulatory response (45). Inhibitory FcRs, particularly FcγRIIb, negatively mediate this response by utilising a phosphorylated immunoreceptor tyrosine-based inhibitory motif (ITIM) on its cytoplasmic tail, triggering a downstream signalling pathway (24). This signalling is fuelled by SH2 phosphatases, such as SH2 inositol 5-phosphate (SHIP) 1 & 2, SHP-1, and phosphatase and tensin homologue (PTEN), that act to dephosphorylate certain specific proteins in the signalling cascade. For example, SHIP is involved in the hydrolysis of phosphatidylinositol-3, 4, 5-triphosphate (PIP3) into PIP2. Importantly, SHIP negatively regulates PKC and MAPKs, thereby inhibiting cell proliferation (75).

Although the inhibitory pathway is thought to be dominant, it is the ratio between the activating and inhibitory signalling pathways on a given cell that determines the immunomodulatory outcome (14, 23, 24, 45). Thus, the FcR are ‘designed’ in such a way to prevent inflammation. Coengagement of activating and inhibitory FcRs leads to tyrosine phosphorylation via LYN kinase, recruitment of SH2-domain containing SHIP, and inhibition of ITAM-triggered Ca^{2+}mobilisation and cellular proliferation (57).

1.2.6 Fc Effector Functions

As stated earlier, the Fc domain of IgG is responsible for the glycoproteins’ key effector functions. These are initially triggered following antigen recognition and the binding of one of the many FcRs or complement factors (23), and can be grouped into being either pro-inflammatory or anti-inflammatory in nature. The process mediates pathogen neutralisation, opsonization, and clearance, as well as ADCC and complement-initiated inflammation, all with both beneficial and detrimental effects depending on the premise

of the IgG activity. Moreover, more than one of these effector functions can occur in a given cell at any particular time, due to up-stream factors that promote simultaneous outcomes. Summarised in **Table 1.2** are those effector functions pertaining to malfunctions in IgG, particularly with regard to autoimmune and inflammatory-related diseases.

***Table 1.2: Summary of FcR effector responses elicited by IgG binding**. IgG N-glycosylation of the Fc domain is directly linked to changes in affinity for FcRs that either produce more pro- inflammatory responses (ADCC) or anti-inflammatory responses (immune modulation). Anti-inflammatory activity is said to be dominant*

Effector Response	Immune Cells	Inflammation	Relation to IgG	Ref
Cytokines				
- Molecules with hormone-like function	All	Both	Altered IgG glycosylation may be linked to changes in cytokine expression	Lin (1995) (76) Van de Winkel (1991) (63)
Degranulation				
- Release of antimicrobial cytotoxic agents	Mast Cells, Basophils, Neutrophils, Eosinophils, Cytotoxic T Cells, NK Cells	Pro	↑ Fcγ-RI ligation = ↑ degranulation, thus ↑ localised inflammation	Marshall (2004) (77) Woolhiser (2001) (64)
Phagocytosis				
- Recognising & engulfing large particles or cells opsonised by C3b or IC, or amassed IC	Mast Cells, Basophils, Neutrophils, Eosinophils, Macrophages	Pro	↑ Fc ligation can lead to ↑ localised inflammation	Aderem (1999) (78) Caron (1998) (79) Forni (2013) (80) Mantovani (1975) (81) Quast (2014) (16) Russell (2009) (82)
ADCC				
- Cell lysis mediated by cytotoxic granules containing perforin & granzymes	NK Cells, Macrophages, Monocytes, Neutrophils, Eosinophils	Pro	↓ core fucosylated/sialylated IgG = ↑ Fcγ-RIIIa ligation = ↑ ADCC ↑ *localised inflammation/cell apoptosis*	Broussas (2013) (83) Bryceson (2011) (67) Nimmerjahn (2005) (45) Quast (2014) (16)
Immune Modulation				
- Upregulation of Fcγ-RIIb, which dominantly inhibit activating Fc-R	All	Anti	↑ sialylated IgG = ↑ Fcγ-RIIb ligation ↑ *anti-inflammatory activity*	Anthony (2011) (71) Böhm (2012) (84) Pincetic (2014) (23) Scallon (2007) (85) Sondermann (2013) (58)

1.2.7 The Complement Cascade

The complement cascade is a major component of the innate immune system and it is believed it evolved before the development of the adaptive immune system (56). There are three initiation pathways for the complement cascade: the classical pathway, the alternative pathway, and the lectin pathway. The key component of the complement cascade is the activation of C3, either directly or indirectly. C3 harbours a reactive thioester moiety that facilitates the attachment of C3 to molecules and cell surfaces, thus allowing them to become targets for other complement receptors (56, 86). Specifically, it is the cleavage component of C3, anaphylatoxin C3b, which initiates the complement cascade. Following this cleavage, a number of events occur, including the conversion of C3-convertase to C5-convertase (16, 86). The cleavage of C5, a C3 homolog that lacks the reactive thioester, initiates inflammation through the formation of membrane-attack complexes (MACs) that target the cell surface (86). These MACs contain the C5b component, along with C6, C7, C8, and many copies of C9, which are recruited sequentially, and their role is to perforate the cell membrane and cause lysis (86).

Glycosylation of the IgG Fc domain is associated with activation of all three complement cascades. Namely, the addition of terminal Gal is associated with significantly increased affinity for the C1q component of the classical pathway (55). The addition of terminal Neu5Ac, however, abrogates the effect of increased affinity of galactosylated IgG Fc for the C1q component, leading to an FcR-independent anti-inflammatory response (46). On the contrary, a decrease in terminal Gal or increase in high-Man glycoforms may be associated with an increased affinity for MBL of the lectin pathway (56) and C3b of the alternative pathway (24).

1.2.8 Loci Associated with Aberrant IgG Glycosylation

A number of loci have been identified as being associated with aberrant glycosylation. Importantly, these have shown a clear directional effect with either increases or decreases in the relative abundance of particular N-glycan moieties. The first locus found to be associated with aberrant plasma glycosylation was hepatocyte nuclear factor 1α (*HNF1α*), a master regulator of the expression of the fucosyltransferase genes, fucosyltransferase-6 (*FUT6*) and *FUT8*, that influence multiple stages in fucosylation (18). Particularly for IgG, *FUT8* has been found to be associated with core fucosylation of the N-glycan moieties in Caucasian populations, and therefore, is important for the regulation of the immune response (18).

In addition, many other loci were identified as being associated with IgG glycosylation. Of those specifically linked to glycosylation, sialyltransferase 6 (*ST6GAL1*), β-1,4-galactosyltransferase (*B4GALT1*), and mannosyl (β-1,4)-glycoprotein β-1,4-*N*-acetylglucosaminyltransferase (*MGAT3*) encode genes for glycosyltransferases (18, 87). Whereas genes previously associated with other diseases, such as autoimmune diseases and haematological cancers, and later identified as being associated with IgG glycosylation, include *IL6ST-ANKRD55*, *IKZF1*, *ABCF2-SMARCD3*, *SUV420H1, SMARCB1-DERL3* and *SYNGR1-TAB1-MGAT3-CACNA1I* (18) (87). Though this study contained participants from four populations (Orkney Islands in the UK, Vis and Korcula Islands in Croatia, Northern Sweden, and The Netherlands), one for validation (The Netherlands), there is still a considerable gap in knowledge regarding the association of specific loci and IgG glycosylation, and to what extent this can impact the final N-glycan moiety when compared with environmental factors.

1.2.9 Environmental Factors Associated with Aberrant IgG Glycosylation

Further to the abovementioned genetic alterations, the cellular environment is associated with aberrant glycosylation, which strongly influences inflammatory properties. The IgG glycome is malleable as it is reliant on the expression levels of enzymes such as glycosyltransferases and glycosylhydrolases, and the abundance of sugar nucleotide donors, which in turn are epigenetically regulated within the producing B cells. Further, the IgG N-glycome is considered a link between the genetic makeup of cells and the cellular environment. Therefore, in theory, one can change their IgG N-glycan composition through modification of lifestyle choices, such as participating in certain activities (i.e., reduced/no smoking and alcohol, and increased exercise) and eating healthy.

Aside from disease presence, altered plasma protein glycosylation has been linked to gender, age, smoking status, body mass index (BMI), plasma lipid profile parameters (high-density lipoprotein (HDL), low-density lipoprotein (LDL), total cholesterol (TC) and triglyceride (TG) levels), blood pressure, fasting blood glucose (FBG), certain medications and diet (88, 89). Several factors have been further explored in association with IgG glycosylation, which could drastically affect the affinity of IgG Fc for the aforementioned FcRs and complement factors.

As stated earlier, one of the most profound factors associated with IgG glycosylation, particularly increasing agalactosylation, is ageing. The IgG glycome explains between

23.3–58% of the variance in age (5, 9, 90). Numerous “GlycanAge” concept studies have been able to explain age in different populations (5, 7, 9, 54) using either blood stains or plasma (9, 90, 91). They have the potential to not only inform individuals of their “biological age”, but also give incentive to improve overall health. Although the concept of ageing can be the culmination of unfavourable levels of multiple factors, the translation of glycomics (i.e., the system-wide study of the relative abundance of glycan moieties) for use in predictive, preventive and personalised medicine is becoming a reality (20).

Gender (5, 9) and hormone levels (92, 93) are also associated with notable changes to the IgG Fc glycome. Particularly, these factors affect IgG Fc galactosylation and sialylation, with evidence of cyclical changes, such as in the menstrual cycle (92). Although not the focus of this review, it should be noted that IgG Fab glycosylation is also associated with changing hormones in pregnancy (94). It has been suggested that oestrogens may be responsible for modulating IgG Fc galactosylation in both women and men, with the oestrodial aromatised from testosterone responsible for these changes in men (93). Taken together, these represent factors that should be controlled for in studies utilising IgG N-glycans.

Aside from hormones, many other blood factors are associated with IgG glycosylation. Extracellular Glc is associated with in vitro changes to IgG galactosylation and sialylation (95), with the increased availability of Gal sugar nucleotide donors (UDP-Gal) proposed as a mechanism (96). The association of FBG to IgG glycosylation is also seen in vivo in multiple populations (5, 9). Other clinical traits found to be associated with IgG glycosylation after correcting for age include lipid profile parameters (5, 9), blood pressure (5, 9, 41), insulin (9), the liver markers alanine aminotransferase and asparate aminotransferase (5), uric acid and urea (5, 9), and fibrinogen, calcium and glycosylated haemoglobin (9).

More recently, increases in various body fat parameters were found to correlate with the increased pro-inflammatory potential of IgG. Increases in BMI, often used as a proxy to body fat, are associated with increases in IgG agalactosylation (9, 97). Further, waist circumference (5, 9), the waist to hip and waist to height ratios, and dual-energy X-ray absorptiometry body fat parameters are associated with altered IgG glycosylation, with the latter explaining the most variation in the IgG glycome (98). The importance of these findings will be validated in longitudinal studies if it is found that reducing body fat via exercise, diet or medication leads to positive changes in the IgG glycome.

Medications are associated with overall plasma glycosylation (88) and IgG-specific glycosylation (29, 88). Moreover, the effect of statin use, prevalent in many populations,

was associated with IgG glycosylation (29). Although both studies implicate medications as affecting the relative abundance of certain IgG N-glycan moieties, they also present inconclusive results, suggesting that this effect is so small as to not have a significant effect on the IgG glycome.

Overall, these associations may directly influence the activity of the producing B cell or alter expression of a number of “glycogenes” that encode glycosyltransferases and glycosylhydrolases (42). In addition to the biomarkers associated with IgG glycosylation, we would expect other biomarkers within the plasma that are yet to be explored in terms of their effect on the plasma or IgG glycome. Thus, although there has been a considerable increase in our knowledge of endogenous and exogenous factors associated with altered IgG glycosylation, further investigation is still warranted.

1.2.10 Conclusions

It is established that genetic and other factors influence IgG glycosylation, which in turn can affect whether the IgG glycoproteins will elicit an anti-inflammatory or pro-inflammatory response. It is important to underscore these processes when considering the use of IgG N-glycan moieties as an indication of disease presence, progress or response to therapeutics. The integral next steps will be the undertaking of large scale and longitudinal studies that aim to further the understanding of how allelic changes, the expression of genes and the presence of other biomolecules within the cellular environment impact the IgG glycome. Indeed, altered expression of genes within any of the pathways involved in the biosynthesis of N-glycans, IgG FcRs or complement factors may significantly alter IgG response. Moreover, changes along the FcR or complement pathways can alter the transcription of cytokines or other biomolecules that in turn act as mediators within the immune system, such that they can regulate whether an anti-inflammatory or pro-inflammatory response is required. The translation of this knowledge will be the ability to decipher whether aberrant IgG N-glycan moieties within a biomarker study, particularly those of age-related chronic diseases, are indeed related to the pathophysiology of the disease or are within the limits of the natural heterogeneity of IgG glycosylation.

1.2.11 Acknowledgments

This work was supported by the Joint Project of the Australian National Health and Medical Research Council and the National Natural Science Foundation of China

(NHMRC APP1112767-NSFC 81561128020), the National Natural Science Foundation of China (81370083, 81673247, 81273170, and 81573215), FP7 European Union MIMOmics (305280), and the Cooperative Research Centre (CRC) for Mental Health (20100104), an Australian Government Initiative.

1.2.12 Author Contributions

Alyce Russell conceived and wrote the paper; Wei Wang conceived and edited the paper; Eric Adua, Ivo Ugrina and Simon Laws edited the ideas presented.

1.2.13 Conflicts of Interest

The authors declare no conflict of interest.

1.3 Epilogue

The premise of this project was to answer the research question: to what degree does a person's underlying genetics and environmental exposures over their life-course contribute to the heterogeneity of the IgG N-glycome?

A total of three studies were coined to address this, with the third combining three sub-studies. The specific aims and hypotheses for each study were conceived according to whether it was validating previous research or providing new evidence to answer the research question. The purpose and significance of these studies are outlined in the executive summary (**Section 1.3.2**).

1.3.1 Specific Aims and Hypotheses of the Thesis

Study 1: Validating Quantitative Trait Loci of N-Glycosylation (gQTLs) of IgG in an Australian Population

The aim of study 1 was to validate existing associations of genetic loci and aberrant IgG N-glycosylation from previous studies in Caucasians.

Study 2: Further Exploring the Role of Genetics in IgG N-Glycosylation by Implementing RNA Sequencing (RNA-Seq)

The aim of study 2 was to model IgG N-glycome heterogeneity using differing counts of given gene transcripts isolated from leucocytes and build on previous genetic research. It was hypothesised that the gene transcripts would elucidate new genes associated with directional changes to the IgG glycoforms.

Study 3: Clinical Measures and Health Behaviours Associated with IgG N-Glycosylation

The aim of study 3 was to do a thorough analysis with both univariate and multivariate statistics, with an array of clinical measures, including health behaviours. It was hypothesised that the clinical measures detrimental to health would have a consistent association to pro-inflammatory IgG glycoforms.

1.3.2 Executive Summary

This thesis was divided into three clear studies, which address specific aims used to answer the overall research question. Herewith I summarise the significance of each chapter to the research question and clarify important facets within the body of work.

Before any analyses begun, a comparison of two different normalisation procedures was warranted. It illustrated that an alternative normalisation method, namely median quotient normalisation (**Section 2.5.3**), was as effective and valid as the most commonly used method. The gold standard for the normalisation of liquid chromatography N-glycan data at inception of this thesis was area normalisation (**Section 2.5.2**). Its popularity is due to the ease of interpreting the area normalised IgG N-glycan peaks (GPs). This method converts the area under each given GP on the resulting chromatogram to a percentage or relative abundance. It does this by dividing each GP by the total area and then multiplying by 100. Thus, the data are normalised so the sum of all the GPs equals 100. In doing so, however, area normalisation introduces correlation structure into the data, and this multicollinearity impedes the use of multivariate analyses. This has limited the type of valid statistics to univariate analyses and has meant that previous studies may have overestimated given effect sizes. Hence, before proceeding with the three studies of this thesis, I showed that median quotient normalisation performed as well as area normalisation and was a suitable substitute for IgG N-glycome normalisation when multivariate statistics were to be utilised (**Section 2.5.4**).

Study 1 was vital for this thesis. It had the purpose of validating previous genetic associations to demonstrate the clinical relevance of my research to the Caucasian population at large, not just to the Australian population (**Chapter 3**). Genome-wide association studies (GWAS), controlling for age, sex and population stratification, were performed on all the IgG glycoforms. Though the study was underpowered, the two loci with genome-wide significance to several IgG N-glycan features were previously found in four isolated European populations. Thus, these Caucasian Australians were suggested to be genetically homogeneous to the Caucasians of European, at least in terms of IgG N-glycosylation. The validation of these genetic loci was important, given the novelty of the other studies in terms of the data and statistical methods proposed. It further ensures the data are derived and pre-processed correctly.

Study 2 combined the IgG N-glycome with next-generation RNA-Seq of expressed genes quantified from whole blood leucocytes (**Chapter 4**). This was the first-time gene expression data had been used to explain altered N-glycosylation. The purpose of this study was to explore pathways linked with different glyco-genes or FcRs,

potentially identifying previously unknown genetic associations. Linear regression models, controlling for age and sex, were performed to determine genes differentially expressed in association to the abundance of the IgG GPs. The most highly associated gene transcripts were identified and used to derive gene ontologies, which describe potential functions of the collection of differentially expressed genes. GP10 had the most significant differently expressed genes and was the focus for much of this study. I also experimented with integrating the two 'omics datasets using orthogonal two-way partial least squares regression (O2PLS); a novel multivariate method. The resulting model quantified the percentage of variation of IgG N-glycosylation explained by the expressed leucocyte gene transcripts.

Study 3 was the amalgamation of three sub-studies exploring the association of various clinical measures and the IgG N-glycome using different statistical approaches (**Chapter 5**). The first sub-study compared different measures of body fat distribution (**Section 5.2**). Its purpose was to validate a previous weak association between BMI and agalactosylated N-glycans, and potentially improve on this association by using more accurate measures of adiposity. Obesity is considered the most preventable modifiable risk factor in the world today. I not only validated the previous weak association but evidenced that measures of central adiposity were associated with a total of 17 GPs, not just agalactosylated N-glycans. Moreover, the direction of the derived effect sizes indicated that pro-inflammatory IgG glycoforms were associated with increased central adiposity.

The second sub-study analysed a collection of clinical measures, including the body fat distribution variables (**Section 5.3**). Importantly, this analysis extended on previous research by implementing a multivariate model to associate the clinical factors to the IgG GPs, controlling for age and sex and retaining all the variables in a single model. The novelty was the ability to estimate the intrinsic correlations, explicating the independent associations of single variables to the other dataset while controlling the effect of the other variables. Notably, I extended on the central adiposity finding of the first sub-study by determining that all the cardiometabolic risk factors were significantly associated with increased pro-inflammatory IgG glycoforms.

Lastly, I briefly analysed the health behaviours excessive alcohol consumption, smoking and physical inactivity against the IgG N-glycome (**Section 5.4**). The foundation for proposing this sub-study was the clinical tests of biological age already available in Europe, and now emerging in China. It is prudent to underscore that if we are to measure and provide results of an individuals' biological age, then we have a responsibility to

knowledgeably prescribe lifestyle changes that may improve unfavourable discordance between biological and chronological age. Physical activity was not associated with any of the IgG GPs; however, it was thought this may be due to the nature of the collected data (that is, 7-day recall of physical activity). Since the IgG N-glycome is known to be stable over periods longer than 6 months (99), these self-reported recall questions would not capture more consistent efforts to stay physically healthy. Both excessive alcohol intake and binge drinking were associated with a few IgG GPs. Finally, the IgG N-glycome of current smokers but not ex-smokers was significantly different to those who had never smoked in comparisons of nine IgG GPs. The health behaviours, however, need to be further explored using a longitudinal cohort, since cross-sectional analyses cannot detect causal effects due to increasing exercise, lowering alcohol intake, or quitting smoking. These inferences will be important as individuals may be advised to alter health behaviours to modify the IgG N-glycome with the purpose of improving biological age; a proxy of subclinical health status.

Overall, this data-driven thesis presents a thorough investigation of how numerous genetic and clinical variables associate with heterogeneity of the IgG N-glycome. Notably, I experimented with data not previously assimilated into the study of the IgG N-glycome and I implemented multivariate models to extend on current evidence. The purpose was to bridge the gap in knowledge and determine factors that may improve the precision of IgG N-glycosylation in stratifying individuals within future studies.

1.3.3 Study Population: A Subset of the Busselton Healthy Ageing Study

A subset of The Busselton Healthy Ageing Study (BHAS) cohort was used. The overall aim of the BHAS was to enhance the understanding of the relationship between ageing and chronic disease processes (100). To this end, the study of how IgG *N*-glycan moieties, which are associated with age and often implicated in studies of chronic and autoimmune disease, fits in with the scope of the overarching BHAS objective.

The BHAS contains a highly homogenous population; 99% identifying as Caucasian with Caucasian parents. Hence, the results of this study may be generalisable to other Caucasian populations. All non-institutionalised Baby-boomers (adults born between 1946 and 1964) listed on the electoral roll in the City of Busselton between 2010 and 2016, were eligible to participate. At completion of recruitment in December 2016, there were a total of 5107 individuals who had participated in the 2010/2016 BHAS project, with a participation rate of approximately 80%.

As part of routine data collection, the BHAS participants completed a detailed questionnaire in a 4 hour clinical assessment as well as provided samples for biobanking (100). Blood serum was acquired from the BHAS for deriving the IgG N-glycan abundance, work completed as part of this thesis (**Chapter 2**). The BHAS participants also involved in the 1994/1995 Busselton Health Study were selected since genotyping had already been performed on this earlier study (**Chapter 3**). Further, extracted RNA samples were obtained from participants within this subset that attended in the first two years (**Chapter 4**). This was due to proposed longitudinal RNA-Seq analyses to be undertaken at completion of the currently active Busselton Baby Boomer Study; the 5-year follow-up of the BHAS. Anthropometric, biochemical and other clinical data routinely tested within the BHAS were also requested (**Chapter 5**). Significantly, the profiles of each subset within this thesis are comparable to the larger BHAS cohort (**Table 1.3**).

***Table 1.3**: Summary data for the whole BHAS cohort and subsets used in each study*

	Entire BHAS Cohort	**GWAS Subset (Study 1)**	**RNA-Seq Subset (Study 2)**	**Clinical Study Subset (Study 3)**
n	5107	614	303	637
Male $n\ (\%)$	2307 (45.2%)	282 (45.9%)	134 (44.2%)	293 (46.0%)
Female $n\ (\%)$	2800 (54.8%)	332 (54.1%)	169 (55.8%)	344 (54.0%)
Age* $(mix - max)$	45.4 – 69.8	46.2 – 68.3	46.2 – 66.3	46.2 – 68.3
Age* $\bar{x}\ (s^2)$	58.0 (5.8)	57.5 (5.2)	57.0 (5.2)	57.6 (5.2)

** Age expressed in years*

Page intentionally left blank

Chapter 2 – IgG N-Glycome Derivation, Data Pre-Processing and Normalisation

2.1 Prologue

Though usually implemented as explanatory variables, the IgG N-glycans were the phenotype of interest in this thesis. Hence, we were attempting to explain the heterogeneity of the IgG N-glycome utilising the other biomolecules and factors of interest in the three studies. A crucial step with any work involving IgG N-glycans is correctly processing the data for downstream analyses. These steps are laborious, both in the wet and dry laboratory.

The wet laboratory methods presented here were the standardised protocols I followed during my month-long internship at one of the world-leading glycobiology laboratories; Genos-Glyco Ltd in Zagreb, Croatia. They are a research-intensive small to medium enterprise specialising in liquid chromatography and mass spectrometry methods of N-glycome analysis in various tissues. They perform contract research, analyses and services for universities, hospitals and private individuals within Europe as well as internationally. Importantly, their protocols have been optimised (10, 18, 89).

Herewith are the details of the experimental design, including the wet laboratory methods I used to isolate, label and analyse the IgG N-glycans using ultra-performance liquid chromatography (UPLC). Further, I detail the dry laboratory methods used while implementing data pre-processing and normalisation, including a brief exploration into the use of median quotient normalisation instead of area normalisation in cases where multivariate statistics are proposed.

2.2 Experimental Design

One often overlooked aspect of high-throughput data analysis are batch effects, defined as "…sub-groups of measurements that have qualitatively different behaviour across conditions and are unrelated to the biological or scientific variables in the study" (101). They exist because laboratory conditions, reagents and technicians vary between different batches, and can compound the biological effects within the data (101). Blocking can overcome this shortfall. Briefly, it involves organising samples for high-throughput analysis so covariates, such as age and sex, are evenly distributed on the plates (i.e. plates are comparable). To this end, any significant variation between plates may be

assumed to be random technical variation, therefore permitting the use of batch correction methods.

The chromatography experiment and processing steps in the laboratory were done on 96-well plates. To control for possible batch effects, the samples need to be blocked and randomised onto these plates. There should be at least one blank allowed to check for contamination, as well as six standards (which are all the same samples) and six duplicates (randomly chosen from the plate). This step can often be an overlooked, arduous phase of any 'omics analysis on a similar platform. The steps used for designing the experimental plates are summarised below and performed in R version 3.1.3 (102).

The factors of interest blocked in this thesis were age and sex. Median age was found for each sex (males = 58.3 years, females = 57.9 years), creating four blocks of samples (**Table 2.1**). Given the above minimum required control wells (blank, standards and duplicates) on a single 96 well plate, there were at least 83 positions free for samples on each plate. Conveniently, the 664 serum samples originally obtained from the BHAS randomised perfectly onto eight plates.

After separating the summary data into the four blocks, these blocks were randomised onto the given plates. Six samples were then randomly selected from each plate to be duplicated. Finally, the order samples were loaded onto the plates was randomised and pasted onto experimental templates, ready for processing in the laboratory.

Table 2.1: *Counts of participants in each block*

	Male	Female
< median age	157	180
> median age	151	176

2.3 Wet Laboratory Protocol to Derive the IgG N-Glycans

The serum aliquots were first arranged according to the order on the experimental templates. They were subsequently left to thaw at room temperature on a shaking plate, briefly vortexed then pulse centrifuged, before been pipetted onto the 96-well preparation plates and subsequently frozen. These preparation plates remained frozen until required and were sequentially thawed on a shaker for processing in batches of two. It took a total

of three days to prepare two plates for N-glycan analysis, and three days to run N-glycan analysis on a single plate using one UPLC machine, though two were available for this experiment.

2.3.1 IgG Isolation

Serum samples (100µL) were diluted ten times with binding buffer and applied to the 96-well Protein G monolithic plates (BIA Separations, Ajdovščina, Slovenia; Pucic 2011, Akmacic 2017). These plates were washed five times with five column volumes of binding buffer (1X PBS, pH 7.4) to remove unbound proteins. IgG was released from the protein G monoliths using five column volumes of elution solvent (0.1M formic acid, pH 2.5). Eluates were collected in 96-deep-well plates and immediately neutralised to pH 7.0 with neutralisation buffer (1M ammonium bicarbonate) to maintain IgG stability.

2.3.2 N-Glycan Release and Labelling

Isolated IgG samples were dried in a vacuum centrifuge. After drying, proteins were denatured with the addition of 30µL 1.33% sodium dodecyl sulfate (w/v) (Invitrogen, Carlsbad, CA, USA) and incubated at 60°C for 10min. Subsequently, 10µL 4% Igepal-CA630 (Sigma-Aldrich, St. Louis, MO, USA) and 0.5mU of PNGase F (ProMega, Hayward, CA) in 10µL 5 × PBS were added to the samples. The samples were incubated overnight at 37°C to release the N-glycans from the IgG glycoprotein.

The released N-glycans were subsequently labelled with 2-aminobenzamide. The labelling mixture was freshly prepared by dissolving 2-aminobenzamide (Sigma-Aldrich, St. Louis, MO, USA) in a dimethyl sulphoxide (DMSO; Sigma-Aldrich, St. Louis, MO, USA) and glacial acetic acid (Merck, Darmstadt, Germany) mixture (17:3 (v/v)) to a final concentration of 48mg/mL. A volume of 25µL of labelling mixture was added to each released N-glycan sample in the 96-well plate. Then 25µL of freshly prepared reducing agent solution (2-picoline borane (Sigma-Aldrich, St. Louis, MO, USA) in DMSO – concentration of 106.96mg/ml) was added and the plates were sealed using adhesive tape. Mixing was achieved by shaking for 10min at room temperature, followed by incubation at 65°C for 2hr. Samples were brought to 80% acetonitrile (ACN) (v/v) by adding 400µL of ACN (J.T. Baker, Phillipsburg, NJ).

2.3.3 Cleaning and Elution of Labelled N-Glycans using Hydophilic Interaction Chromatography – Solid Phase Extraction (HILIC-SPE)

Free label and reducing agent were removed from the samples using HILIC-SPE. An amount of 200µL of a 0.1g/mL suspension of microcrystalline cellulose (Merck, Darmstadt, Germany) in water was applied to each well of a 0.45µm GHP filter plate (Pall Corporation, Ann Arbor, MI, USA). Solvent was removed by application of vacuum using a vacuum manifold (Millipore Corporation, Billerica, MA, USA). All wells were prewashed five times using 200µL water, followed by equilibration three times using 200µL acetonitrile/water (80:20, v/v). The samples were loaded into the wells and subsequently washed seven times using 200µL of acetonitrile/water (80:20, v/v). N-glycans were eluted two times with 90µL of water and combined eluates stored at –20°C until analysis with UPLC.

2.3.4 N-Glycan Analysis with UPLC

Fluorescently-labelled N-glycans were separated by UPLC on a Waters Acquity UPLC instrument (Waters, Milford, MA) consisting of a quaternary solvent manager, sample manager and a FLR fluorescence detector set with excitation and emission wavelengths of 250nm and 428nm, respectively. The UPLC instrument utilised Empower 3 software (Waters, Milford, MA). Labelled *N*-glycans were separated on a Waters BEH Glycan chromatography column, 100 × 2.1mm i.d., 1.7µm BEH particles, with 100mM ammonium formate, pH 4.4, as solvent A and acetonitrile as solvent B.

The UPLC separation method used a linear gradient of 75–62% acetonitrile at a flow rate of 0.4ml/min in a 25min analytical run. Samples were maintained at 5°C before injection, and the separation temperature was set at 60°C. The system was calibrated using an external standard of hydrolysed and 2-aminobenzamide labelled glucose oligomers from which the retention times for the individual glycans were converted to glucose units.

Data processing was performed using an automatic method with a traditional integration algorithm, after which each chromatogram was manually examined to ensure the same 24 IgG GPs of integration existed for all samples. Below is a typical chromatogram resulting from UPLC (**Figure 2.1**) as well as the IgG N-glycans prevalent in the 24 IgG GPs (**Figure 2.2)**.

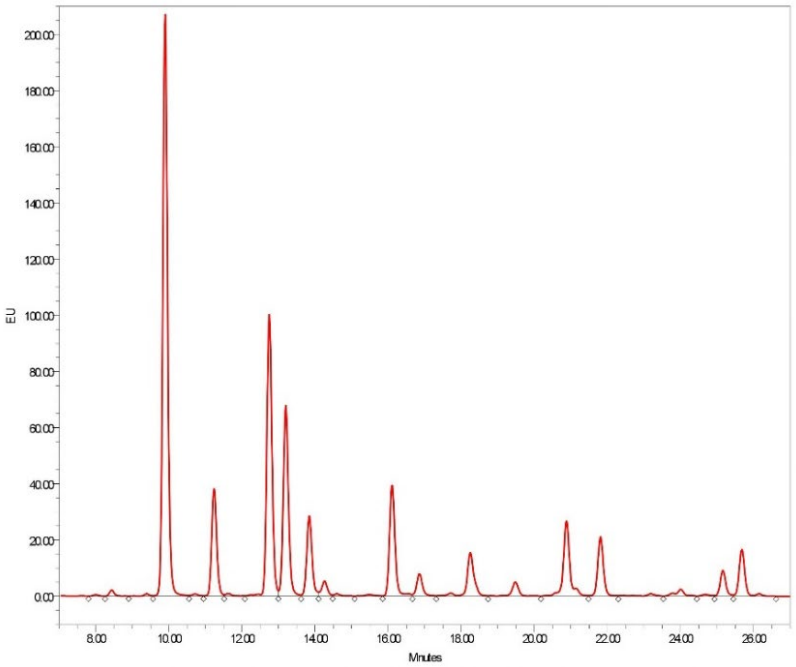


Figure 2.1: *A typical chromatogram of the 24 IgG GPs following UPLC*

GP	Main structure	Abbreviation	GP	Main structure	Abbreviation	GP	Main structure	Abbreviation
GP1		FA1	GP9		FA2[3]G1	GP17		A2G2S1
GP2		A2	GP10		FA2[6]BG1	GP18		A2BG2S1
GP3		A2B	GP11		FA2[3]BG1			FA2G2S1
GP4		FA2	GP12		A2G2	GP19		FA2BG2S1
GP5		M5	GP13		A2BG2	GP20		FA2FG2S1
GP6		FA2B	GP14		FA2G2	GP21		A2G2S2
GP7		A2[3]G1	GP15		FA2BG2	GP22		A2BG2S2
GP8		A2BG1	GP16		FA2[6]G1S1	GP23		FA2G2S2
		FA2[6]G1			FA2[3]G1S1	GP24		FA2BG2S2

Figure 2.2: *The N-glycan moieties prevalent in the 24 IgG GPs derived using UPLC (103) F – core fucosylation; A2 – biantennary; B – bisecting N-acetylglucosamine; Gx – galactosylation; Sx - sialylation*

2.4 Data Pre-Processing

After integration, these data were extracted from Empower 3 and the procedures outlined in **Figure 2.3** were performed in R version 3.1.3 (102).

The extracted data gave the area under each given GP. These were dependent on sample signal quality and the size of the peak itself, thus they need to be normalised in some manner. Following normalisation, however, batch correction was performed to adjust the plate signals. This was due to each plate been analysed over many days, both during N-glycan isolation and labelling, and UPLC.

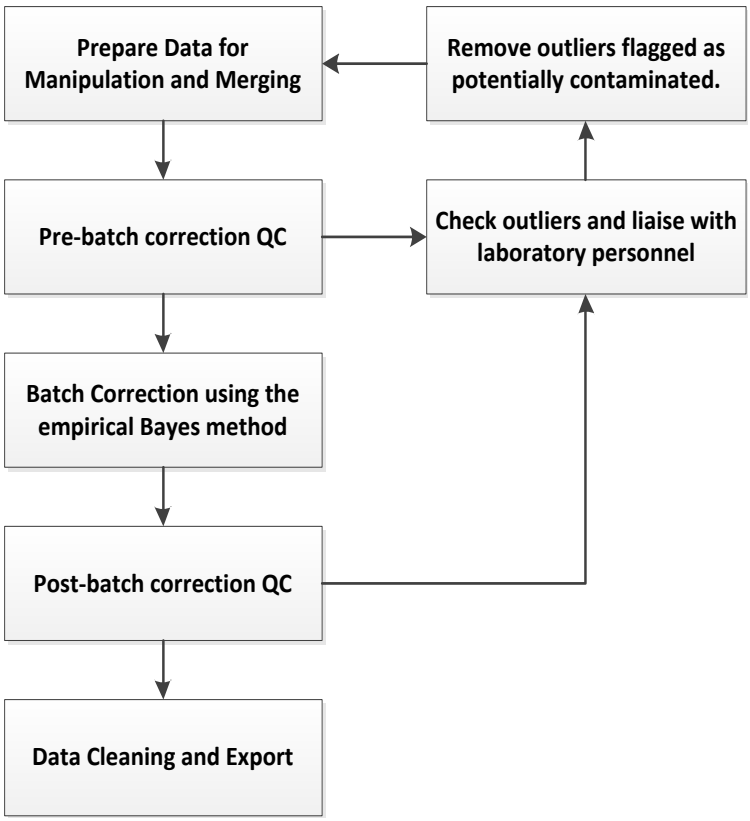


Figure 2.3: *Flow chart of data pre-processing steps*

Outlier detection was performed before and after batch correction. Importantly, the first quality control step developed a list of samples that were potential outliers. These were checked by looking at the wet laboratory records for the listed outlying samples (i.e. IgG quality, UPLC signal, any other laboratory notes taken during the sample preparation, isolation and labelling stages). Participants self-reported medical history was also checked for evidence of morbidities associated with marked changes to the IgG N-

glycome, which are classified as biological outliers. Certainly, if these samples passed quality checks in the wet laboratory, they were maintained in the dataset. This thesis was interested in explaining heterogeneity within the IgG N-glycome, and so biological outliers were pertinent to the research question.

Batch correction was performed using the empirical Bayes method, implemented in R as the ComBat procedure; a function within the sva package. The batch correction was completed following normalisation, discussed below.

2.5 Normalisation

Data normalisation helps to remove unwanted noise in the data, common in large experiments utilising many batches to analyse a whole study sample. The most widely used form of IgG N-glycan UPLC normalisation is area normalisation. However, I wanted to assess how it performed when compared with another emerging method for normalising N-glycome data; median quotient normalisation (104). A thorough review of other normalisation methods was beyond the scope of this thesis and thus has not been included.

2.5.1 Noise in the Raw UPLC Intensities

The raw IgG N-glycan UPLC-derived data was a measure of the intensity of fluorescents of each N-glycan structure within the sample. The GPs are derived by manually integrating the intensity peaks (**Figure 2.1**) and deriving an area from the Empower3 software. Multidimensional scaling (MDS) plots were used to visualise the levels of similarity or distances between individual samples.

A pattern can be seen in the MDS plot on all raw data and appears to mimic a volcano plot, starting on the right-hand side (**Figure 2.4**). There are, however, no obvious clusters due to the plate of analysis. It should be noted that only data that passed quality control checks were considered in the following MDS plots.

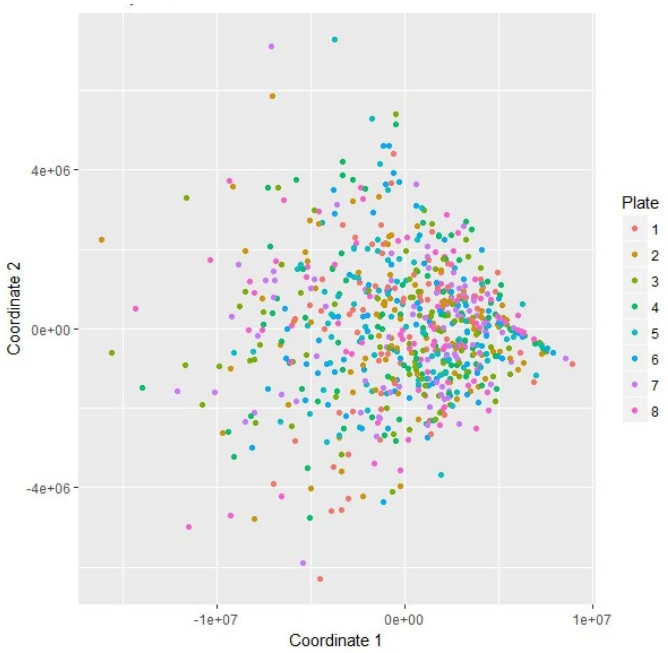


Figure 2.4: *MDS plot of raw IgG GP intensities, derived using UPLC*

2.5.2 Area Normalised and Batch Corrected IgG GPs

Area normalisation (more formally known as integral normalisation) is the most widely implemented method of normalisation in N-glycome studies (6, 18, 103) as well as other fields (104). Its wide usage is due to ease of interpretation (i.e. each GP is interpreted as relative abundance or percentage of the total IgG N-glycome).

Area normalisation involves taking the area under a given peak and dividing it by the total area under all peak intensities, per sample. Following area normalisation and batch correction, there was a reduction in the unwanted noise seen in the raw intensities, which can be noted by the change in shape of the cluster of IgG N-glycan samples (**Figure 2.5**). There appears to be several outliers in the dataset following batch correction and area normalisation. As mentioned earlier, however, these were biological outliers and were of interest for the purpose of exploring IgG N-glycome heterogeneity in this thesis. That is, after consulting the clinically collected data of these participants, it was ascertained these varied profiles were biological and not due to some laboratory error, and these differences warranted inclusion in these analyses.

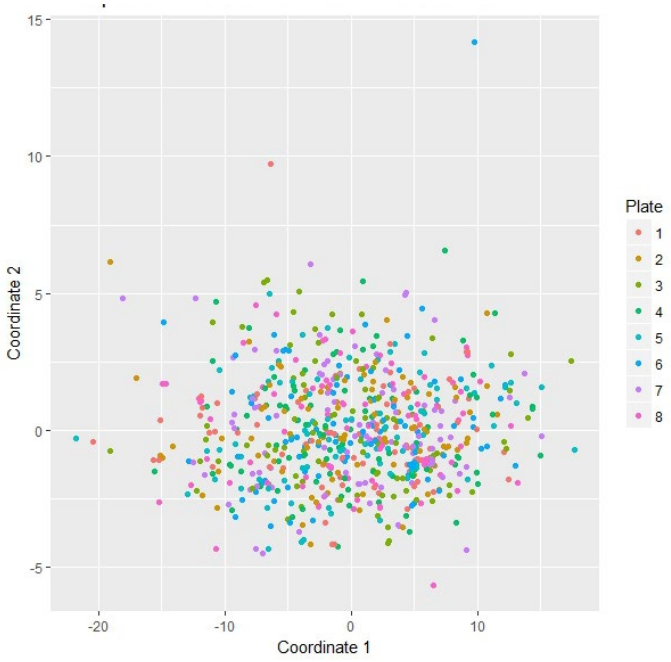


Figure 2.5: *MDS plot of area normalised IgG GP intensities, derived using UPLC*

Although it is easy to interpret the data by normalising the GPs to relative abundances or percentages of the total IgG N-glycome, there is also the introduction of correlation structure; as one GP percentage decreases in a population, another will naturally increase/decrease to "make up the total area" in the sample. For this reason, area normalisation is not recommended when planning analyses where multicollinearity would violate model assumptions.

2.5.3 Median Quotient Normalised and Batch Corrected IgG GPs

The use of median quotient normalisation may remedy this issue (105). Median quotient normalisation was originally employed to normalise mass spectrometry and nuclear magnetic resonance spectroscopy data, especially from metabolomics experiments (104). It assumes that biologically-interesting changes in intensity influence only parts of the chromatography spectrum, while dilution effects will affect all intensities (106).

The procedure involves first performing area normalisation (**Eq. 1**) then calculating a reference spectrum from the study samples (in this case, the median spectrum), calculating the quotients of all the variables of interest of the sample spectrums to the reference spectrum, calculating the median of the quotients, and finally dividing all variables of the sample spectrums by the median quotient (104). The median quotient

normalised data clustered in a similar manner to the area normalised data and showed corresponding biological outliers on the same plates (**Figure 2.6**).

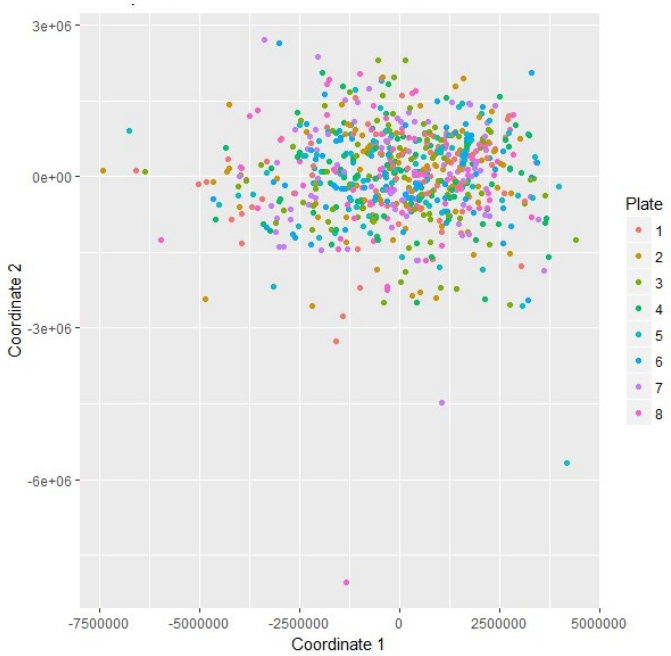


Figure 2.6: *MDS plot of median quotient normalised IgG GP intensities, derived using UPLC*

2.5.4 Comparing Normalisation Methods

To compare the two normalisation methods, models were run against a known associated factor, age (5, 9, 91), controlling for the covariate sex. The model metrics were extracted (**Table 2.2**) and demonstrated neither normalisation method drastically outperforms the other. This was common for other markers tested (excluded here). Thus, the decision on what method to normalise the UPLC-derived data for a study was informed by the planned analyses. Both types are used in this thesis. Specifically, area normalised IgG GPs were used when validating previous research, while median quotient was specifically used for proposed multivariate analyses. These are explicitly stated in each chapter.

Table 2.2: *Comparison of modelling age using the area normalised and median quotient normalised IgG N-glycan data*

	Area Normalised N-Glycan Data					Median Quotient Normalised N-Glycan Data				
	β	**SE**	**p**	**MSE**	**r2**	**β**	**SE**	**p**	**MSE**	**r2**
GP1	1.541	0.195	4.94E-14	24.258	0.091	1.425	0.198	8.39E-12	24.642	0.076
GP2	1.026	0.199	1.03E-06	25.575	0.041	1.063	0.200	3.98E-07	25.510	0.044
GP3	1.667	0.198	1.28E-15	23.970	0.101	1.439	0.201	8.39E-12	24.647	0.076
GP4	1.751	0.191	4.31E-18	23.533	0.118	1.482	0.198	1.56E-12	24.489	0.082
GP5	0.979	0.212	9.97E-06	25.777	0.034	0.901	0.205	2.59E-05	25.858	0.031
GP6	2.178	0.190	2.31E-26	22.077	0.172	2.235	0.197	8.43E-26	22.166	0.169
GP7	0.338	0.200	1.11E-01	26.527	0.006	0.376	0.201	7.06E-02	26.500	0.007
GP8	-0.638	0.200	2.42E-03	26.226	0.017	-0.521	0.208	1.53E-02	26.386	0.011
GP9	-0.747	0.204	5.37E-04	26.093	0.022	-0.665	0.204	1.87E-03	26.206	0.018
GP10	0.722	0.203	7.15E-04	26.123	0.021	0.748	0.208	6.67E-04	26.116	0.021
GP11	0.959	0.197	3.75E-06	25.685	0.037	0.920	0.198	9.51E-06	25.766	0.034
GP12	-0.629	0.199	2.45E-03	26.232	0.017	-0.683	0.198	1.02E-03	26.154	0.020
GP13	-0.353	0.205	1.09E-01	26.522	0.006	-0.551	0.202	9.11E-03	26.338	0.013
GP14	-2.034	0.188	4.71E-24	22.496	0.157	-2.062	0.185	1.68E-25	22.263	0.165
GP15	-0.963	0.200	4.59E-06	25.709	0.036	-1.138	0.195	2.73E-08	25.293	0.052
GP16	-0.598	0.201	4.38E-03	26.281	0.015	-0.599	0.200	4.25E-03	26.274	0.015
GP17	-0.365	0.206	1.03E-01	26.515	0.006	-0.539	0.202	9.77E-03	26.349	0.012
GP18	-1.900	0.193	1.98E-20	23.120	0.133	-1.944	0.188	2.40E-22	22.803	0.145
GP19	-0.097	0.206	6.67E-01	26.637	0.002	-0.193	0.202	3.55E-01	26.608	0.003
GP20	-0.731	0.206	7.15E-04	26.127	0.021	-0.901	0.203	2.29E-05	25.841	0.031
GP21	-0.341	0.207	1.14E-01	26.532	0.005	-0.552	0.204	9.11E-03	26.340	0.013
GP22	0.072	0.205	7.27E-01	26.641	0.001	0.004	0.203	9.83E-01	26.646	0.001
GP23	-1.178	0.203	3.25E-08	25.296	0.052	-1.196	0.198	8.94E-09	25.196	0.056
GP24	-0.162	0.203	4.63E-01	26.619	0.002	-0.268	0.201	2.00E-01	26.572	0.004

Page intentionally left blank

Chapter 3 – Validating gQTLs of IgG in an Australian Population

3.1 Introduction

The premise of the Human Genome Project, a collaborative international effort launched in the 1990s, was to explore and identify causes for phenotypes, and disorders (107). Although it was discovered that most phenotypes were too complex to be explained by genetics alone, the Human Genome Project set precedence for ethical considerations and biological importance of more discoveries in the medical sciences. Importantly, there were not as many genes as first thought (approximately 24,000 instead of the hypothesised 100,000). These conclusions lend support to the emerging field of epigenetics and the importance of post-translational modifications, among others, in the construction of the complex phenotypes in the human repertoire.

N-glycosylation is considered a complex phenotype and many glyco-enzymes, termed glycosyltransferases and glycosylhydrolases, act on the branching IgG N-glycan moieties to add and remove sugar nucleotides, respectively (108). However, several loci, termed gQTLs, are associated with aberrant IgG N-glycosylation. These have shown a clear directional effect with either increases or decreases in the relative abundance of GPs or N-glycosylation features. These N-glycosylation features are derived from the measured IgG GPs to represent different groups of structures, such as the abundance of digalactosylated N-glycans.

The first gQTL found to be associated with aberrant plasma glycosylation was *HNF1α* [MIM 142410], a gene whose product is hepatocyte nuclear factor 1α (HNF1a). HNF1a is a master regulator of the expression of *FUT6* [MIM 136836] and *FUT8* [MIM 602589], which are translated to fucosyltransferases that influence multiple stages in fucosylation (19). Particularly for IgG, *FUT8* associates with core fucosylation in European Caucasian populations, thought to be important for the regulation of the immune response (19).

Many other gQTLs were identified as being associated with IgG N-glycosylation. Of those specifically linked to N-glycosylation, the *ST6GAL1* [MIM 109675], *B4GALT1* [MIM 137060], and *MGAT3* [MIM 604621] genes encode for glycosyltransferases that add Neu5Ac, Gal and bisecting GlcNAc, respectively, to the branching structures (18, 19, 94). Recently, a multivariate modification to GWAS was performed, which identified and replicated a further five novel loci of IgG N-glycosylation (105).

The aim of Study 1 was to verify gQTLs previously found in European Caucasian populations, though performed in an Australian population for the first time. To this end, GWAS were implemented for each IgG N-glycan feature (**Appendix 1**), ultimately leading to the identification of gQTLs; a term coined by our team to describe genes associated with altered N-glycosylation.

3.2 Methods

3.2.1 Participants

Of the 637 BHAS participants with IgG N-glycome data that passed QC (**Chapter 2**), 614 (96.4%) had genotype data. Summary data and further details about the participants can be found in **Section 1.3.2** and **Table 1.3**.

This study conformed to the ethical guidelines of the 1975 Declaration of Helsinki.

3.2.2 Genotyping

Genotyping was previously done for a subset of BHAS participants who also participated in the earlier 94/95 Busselton Health Study. This was performed on two different Illumina micro-arrays at two different time points. In 2009, approximately 1500 participant samples from the 94/95 Busselton Health Study were genotyped using the Illumina Human610-Quad BeadChip micro-array. In 2010, the remaining participants (approximately 3000) from the 94/95 Busselton Health Study were genotyped using the Illumina Human660W-Quad BeadChip micro-array. The resulting single nucleotide polymorphisms (SNPs) measured were amalgamated using imputation, giving a final total number of 28,071,904 SNPs.

Unfortunately, the chromosome 5 file was corrupt when we received the genotype data. Another genotype dataset was eventually located; however, due to time constraints, I present the original GWAS results. Thus, only 21 autosomes were analysed.

3.2.3 Imputation of the Genotype Data

Imputation aims to predict genotypes for the SNPs that were not directly genotyped with either platform, thus increasing the density of available genotypes by several orders of magnitude (109). Imputation can be utilised to control for missingness of SNPs within and between datasets, especially when combining genotype data derived on different

platforms, and increase the number of SNPs available for GWAS. The genotype data was imputed using the 1000 Genomes Project (all ancestries panel, March 2012), with phasing completed using MACH1 (version 1.0.18) and imputations completed using Minimac (version 2012.10.3).

3.2.4 Genotype QC

QC was performed on the data following genotyping. This consisted of checking the imputed genotype data for several allowable parameters. Specifically, individuals with a call rate of less than 95% were removed from the analysis, as well as SNPs with a call rate of less than 95%, minor allele frequency (MAF) of less than 1%, and a Hardy-Weinberg equilibrium (HWE) *p*-value less than 1×10^{-4}. We expected some deviation from the HWE if there was any bias introduced by population stratification or participant selection, which was most probably the case for BHAS. Thus, as well as controlling for age and sex, we also controlled for relatedness by constructing a kinship matrix.

3.2.5 Controlling for Population Stratification

The data was tested for population stratification to determine whether any extreme outliers existed. A kinship matrix was formulated to control for relatedness, specifically population specific identity by state, within the cohort. Though previous GWAS using the BHAS genotype data have assumed independent individuals, there are known relatives within the BHAS cohort. Furthermore, population stratification occurs in Caucasian populations, though this was assumed to only have an effect on GWAS in extreme cases (110). However, given the complexity of IgG N-glycosylation this was controlled for in the GWAS modelling of the N-glycosylation features, as done in previous such research (18). This kinship matrix was generated using the “ibs” function of the GenABEL package in R, with the option weight=“freq” (111).

3.2.6 GWAS Modelling

Linear mixed modelling methods were utilised for the GWAS in this study, adjusting for the covariates age and sex, as well as the kinship matrix. Since this was a validation study, the IgG GPs were normalised using area normalisation (**Section 2.5**). These data were processed with standard formulas to give relative abundance of N-glycosylation features; termed derived traits (**Appendix 1**). Thus, GWAS was completed on 24 IgG GPs and an additional 32 derived traits. It should be noted that not all the previously

characterised derived traits were explored in this study due to perceived redundancy. However, the naming convention proposed in (18) was maintained for comparative purposes.

Transformation with inverse-variance weighted transformation converted the N-glycan variables into z-scores before the analysis, as done previously (18). The DatABEL package (111) was used to convert file formats into manageable file sizes and the ProbABEL package (112) to implement the models. The resulting analyses were the given associations between the different SNPs and the individual N-glycan features.

We used a P-value of 2.27×10^{-9} to indicate genome-wide significance, as previously described (18). A P-value of 5×10^{-8} was considered strongly suggestive of an association and warranting further investigation.

The GWAS were performed in R version 3.3.3 (102).

3.2.7 GWAS Annotation and Visualisation

GWAS results were presented overall in Manhattan plots for each IgG N-glycan variable (**Appendix 2**), which allowed for the visualisation of the gQTLs. The strongly suggestive P-value of 5×10^{-8} were represented by a red horizontal line on the Manhattan plot at $-\log_{10}(5 \times 10^{-8}) \sim 7.3$, with higher points indicating a stronger association between the given SNPs on the plot and the N-glycan feature of the given GWAS.

Data mining of the GWAS results was required to determine the significance of any gQTLs and a given N-glycan variable. NCBIs' dsSNP was used to derive specific gene and function information from each SNP, including: gene name, chromosomal position, predicted function, and association with complex diseases. Gene references are indicated with the OMIM convention throughout or NCBIs GeneID when this was missing. Moreover, SNiPA (113) further explored the associated gQLTs by zooming into specific gene regions indicated on the Manhattan plots, illustrating the location of the associated SNPs in more detail. SNiPA was also used to plot linkage disequilibrium, which may be used to link SNPs within the non-coding genomic regions to flanking genes or regulatory elements.

3.3 Results

A total of 56 GWAS were performed to identify gQTLs of IgG N-glycosylation, one for each of the IgG N-glycan features (**Appendix 1**). The Manhattan plots of these GWAS are in **Appendix 2** and referenced throughout.

3.3.1 SNPs with Genome-Wide Significance

The associations of two gQTLs reached genome-wide significance (2.27×10^{-9}) among a total of nine N-glycosylation features (**Table 3.1**). Additionally, there were a further four *N*-glycosylation features suggestively associated with these gQTLs. These gQTLs contained two genes coding glycosyltransferases, namely the previously mentioned *ST6GAL1* and *MGAT3*, which add terminal Neu5Ac or a bisecting GlcNAc, respectively. There were, however, three genes in the latter locus in strong linkage disequilibrium; *SYNGR1-TAB1-MGAT3*.

Table 3.1: *Summary of genetic markers with genome-wide significant (p<2.27x10-9) associations with the N-glycosylation of IgG*

Chr.	SNP with Lowest P-Value	Lowest P-value	Effect Size (s.e.)	MAF	R^2	Genes in Interval of Identified SNPs
3	rs6764279	6.09E-36	0.716 (0.057)	0.276	0.999	ST6GAL1
22	rs113200473	2.22E-16	-0.536 (0.064)	0.280	0.869	SYNGR1-TAB1-MGAT3

The SNP identified with the lowest *P*-value within *ST6GAL1* (rs6764279), located at locus 3q27.3, consistently had the strongest association among the N-glycosylation features (**Figure 3.1**). It is centrally located within *ST6GAL1*, and the N-glycosylation features it associates with were all sialylated structures, though most were describing the abundance of monosialylation. Using the Edinburgh naming convention, the N-glycan feature with the highest association was IGP29 (**Figure A2.30**); a derived trait of the relative abundance of fucosylated digalactosylated monosialylated N-glycans among all fucosylated digalactosylated N-glycans within the IgG N-glycome; though six other N-glycan features were also associated with the gene (**Table 3.2**). The polymorphism itself has strong linkage disequilibrium to many SNPs within *ST6GAL1*, but no flanking genes were implicated (**Figure 3.2**).

Table 3.2: *The N-glycosylation features associated with locus 3q27.3, containing ST6GAL1*

Edinburgh Code	N-Glycan Feature Details	Genome-Wide Significance	Manhattan Plot
GP16	FA2G1S1	Yes	Fig A2.16
GP23	FA2G2S2	Yes	Fig A2.23
IGP24	FGS/(FG+FGS)	Yes	Fig A2.25
IGP28	FG1S1/(FG1+FG1S1)	Yes	Fig A2.29
IGP29*	FG2S1/(FG2+FG2S1+FG2S2)	Yes	Fig A2.30
IGP32	FBG2S2/(FBG2+FBG2S1+FBG2S2)	Yes	Fig A2.33
IGP35	FBG1/FBG2	Yes	Fig A2.36

* denotes feature with the highest association

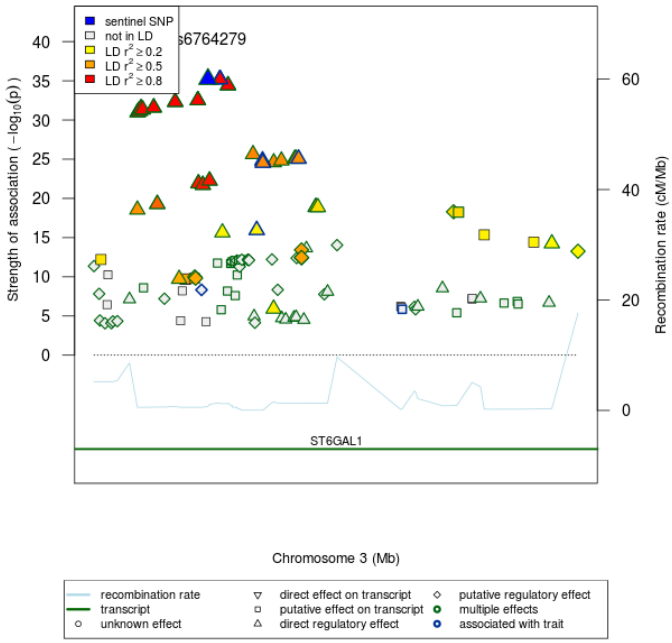


Figure 3.1: *Regional association plot of GWAS of the mono-sialylation of fucosylated digalactosylated N-glycans, derived using the formula FG2S1/(FG2+FG2S1+FG2S2). Further derivation details are given in* ***Appendix 1***

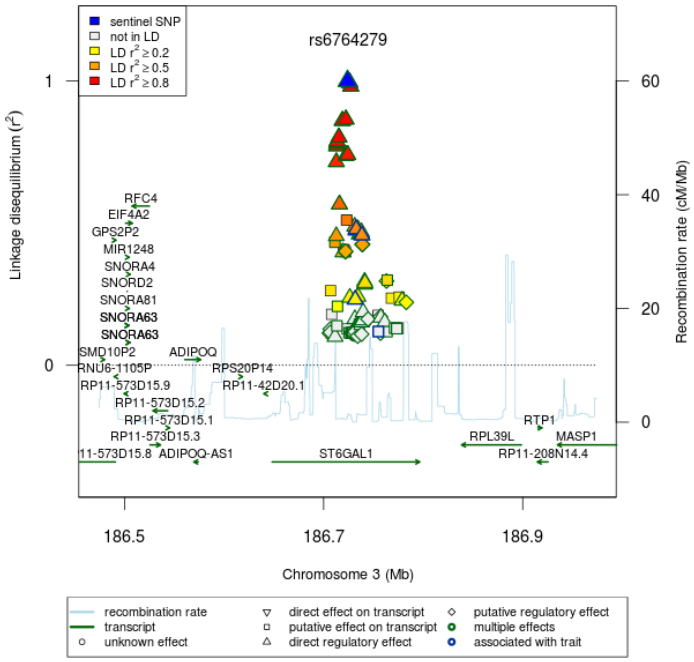


Figure 3.2: *Linkage disequilibrium plot of SNP with genome-wide significance to the mono-sialylation of fucosylated digalactosylated N-glycans, derived using the formula FG2S1/(FG2+FG2S1+FG2S2). Further derivation details are given in* ***Appendix 1***

The second gQTL identified (22q13.1) had three genes presenting SNPs with genome-wide significance; *SYNGR1* [MIM 60392], *TAB1* [MIM 602615], and *MGAT3*. There were six N-glycan features associated with this locus, though only two reached genome-wide significance (**Table 3.3**). The SNP with the highest association (rs113200473) was within the coding region of *MGAT3*. The feature with the highest association to this SNP was IGP40 (**Figure A2.39**); a derived trait of the abundance of bisection by GlcNAc among fucosylated disialylated N-glycans (**Figure 3.3**). The SNP with the highest association was shown to be in linkage disequilibrium with many SNPs, including those from the flanking genes *SYNGR1* and *TAB1*, both also having many hits from these N-glycan features (**Figure 3.4**).

Table 3.3: *The N-glycosylation features associated with the locus 22q13.1, containing SYNGR1-TAB1-MGAT3*

Edinburgh Code	N-Glycan Feature Details	Genome-Wide Significance	Manhattan Plot
IGP40*	FBS2/(FS2+FBS2)	Yes	Fig A2.39
IGP62	Fn	No	Fig A2.47
IGP66	FBn	No	Fig A2.51
IGP67	FBG0n/G0n	Yes	Fig A2.52
IGP70	FBn/Fn	No	Fig A2.55
IGP71	FBn/Fn total	No	Fig A2.56

* denotes feature with the highest association

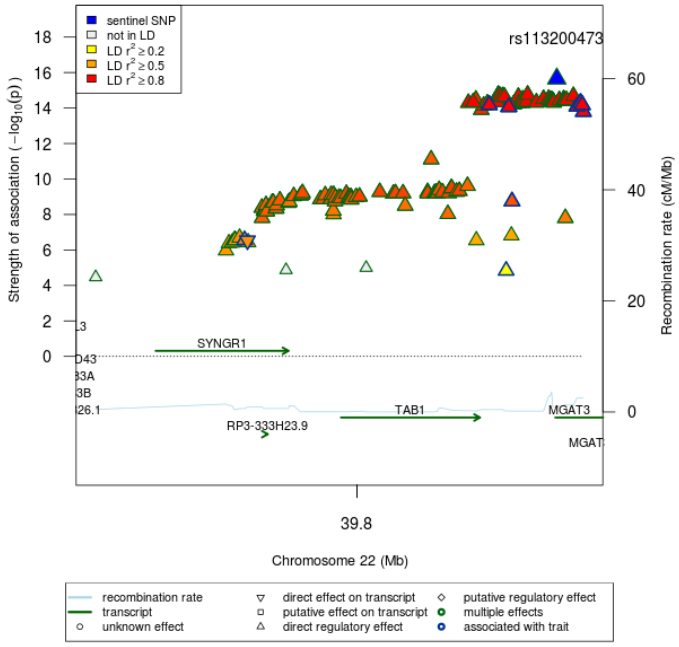


Figure 3.3: *Regional association plot of GWAS of the bisection of fucosylated disialylated N-glycans with GlcNAc, derived using the formula FBS1/(FS2+FBS2). Further derivation details are given in* ***Appendix 1***

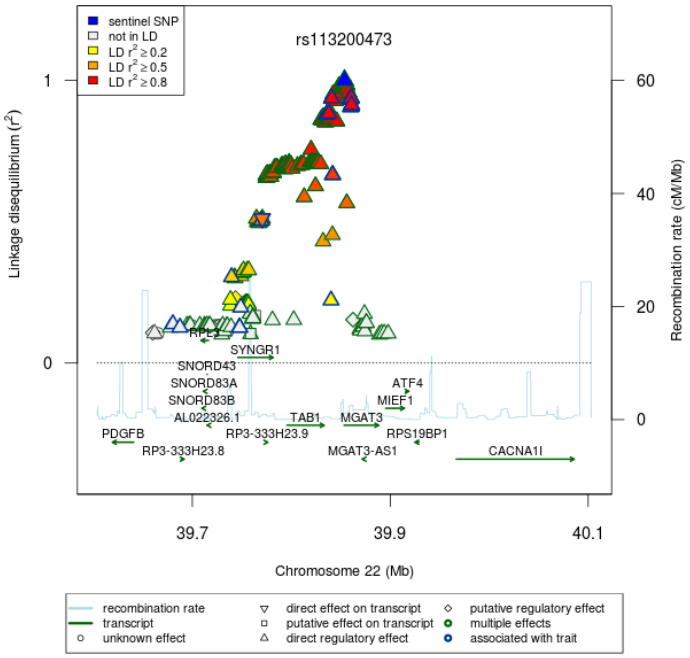


Figure 3.4: *Linkage disequilibrium plot of SNP with genome-wide significance to the abundance of bisection by GlcNAc among all fucosylated disialylated N-glycans, derived using the formula FBS1/(FS2+FBS2). Further derivation details are given in* ***Appendix 1***

3.3.2 SNPs with Strongly Suggestive Associations

A further eight gQTLs demonstrated strongly suggestive associations ($2.27 \times 10^{-9} < p < 5 \times 10^{-8}$) among a total of 11 N-glycosylation features (**Table 3.4**). Although none of the associated genes encoded glycosyltransferases, one did contain a hydroxylase for Neu5Ac, though this is inactive in humans (*CMAHP* [MIM 603209]).

The first suggestively associated SNP lies within the 2kb region upstream of *miRNA 4425*, in the locus 1p36.11 (**Figure A2.53**). This gene was previously characterised as part of a deep sequencing study analysing normal and malignant B-cell small RNAs (114), and found to be negatively associated with IGP68; a derived feature of the abundance of fucosylation (with bisecting GlcNAc) among monogalactosylated N-glycans. It was the only SNP within the region with an association though; therefore, it may be a spurious find.

Table 3.4: *Summary of genetic markers with suggestive ($2.27x10^{-9}<p<5x10^{-8}$) associations with the N-glycosylation of IgG*

Chr.	SNP with Lowest P-Value	Lowest P-value	Effect Size (s.e.)	MAF	R^2	Genes in Interval of Identified SNPs
1	rs193294272	4.52E-08	-1.558 (0.282)	0.012	0.929	miRNA4425
6	rs9461096	1.76E-08	0.607 (0.107)	0.063	0.986	CMAHP
7	rs176	2.32E-08	0.321 (0.057)	0.371	0.997	OSBPL3*
8	rs71514186	1.35E-08	-0.449 (0.078)	0.216	0.729	ERICH1
14	rs140499943	2.57E-10	0.651 (0.101)	0.097	0.871	
15	rs57331048	4.35E-08	0.345 (0.062)	0.392	0.891	SNRPN*
17	rs7214156	1.63E-08	-0.416 (0.073)	0.272	0.659	TBKBP1-TBX21*
18	rs141768436	3.30E-08	-1.323 (0.237)	0.018	0.715	SKOR2

* denotes gene in linkage disequilibrium with SNP, which was in a flanking non-coding region

GP5, a high-Man N-glycan, was suggested to be associated with *CMAHP*, within the locus 6p22.3-22.2 (**Figure A2.5**). Although it encodes a type of Neu5Ac abundant in most mammals, its expression is not detectable in humans. Thus, it is considered a pseudogene.

IGP35, a ratio of monosialylation over disialylation among fucosylated N-glycans with a bisecting GlcNAc, had a suggestive association with a SNP (rs176) within the non-coding genomic region of the locus 7p15.3 (**Figure A2.36**). It was, however, in strong linkage disequilibrium to SNPs of the flanking gene *OSBPL3* [MIM 606732] (**Figure 3.5**). The expression of this gene is ubiquitous in human tissues.

IGP32 (**Figure A2.33**) was positively associated, whereas IGP34 (**Figure A2.35**) was negatively associated, with a SNP (rs71514186) within the locus 8p23.3. Though it was the only suggestive SNP association within the locus, there were many SNPs that were almost suggestive with strong linkage disequilibrium (**Figure 3.6**). These SNPs span *ERICH1* [NCBI GeneID 157697].

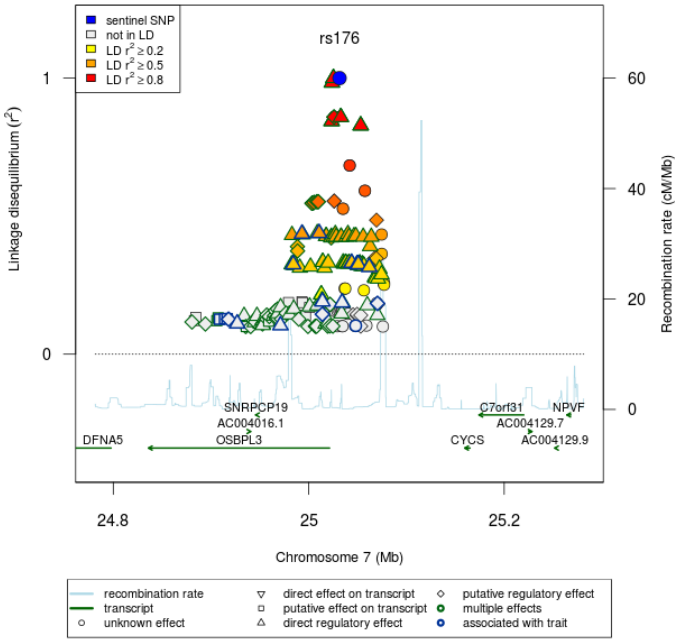


Figure 3.5: *Linkage disequilibrium plot of SNP with a suggestive association to the ratio of monosialylation over disialylation of fucosylated bisected N-glycans, derived using the formula FBS1/FBS2. Further derivation details are given in* ***Appendix 1***

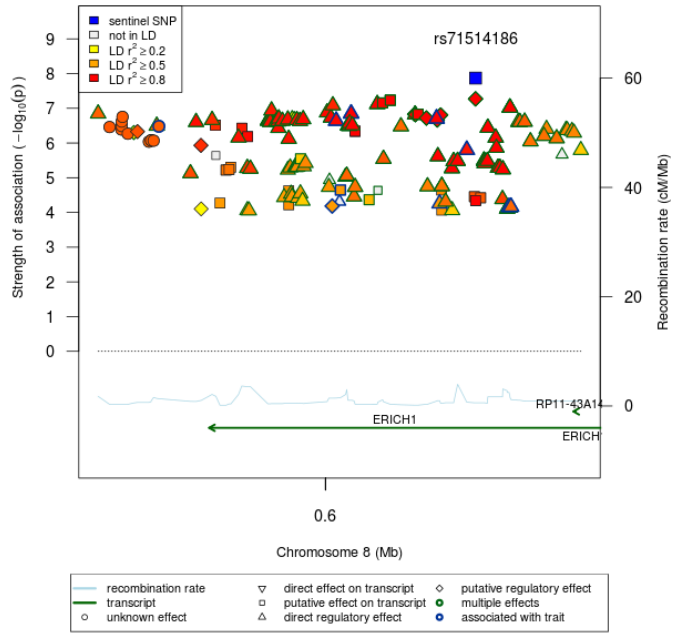


Figure 3.6: *Regional association plot of GWAS of the abundance of fucosylated digalactosylate disialylated N-glycans with bisecting GlcNAc over all fucosylated bisected digalactosylated structures, derived using the formula FBG2S2/(FBG2+FBG2S1+FBG2S2). Further derivation details are given in* ***Appendix 1***

This was in stark contrast to the SNP (rs140499943), suggestively associated with three N-glycan features; positively associated with IGP30 (**Figure A2.31**), and negatively associated with IGP33 (**Figure A2.34**) and IGP34 (**Figure A2.35**). It was however located within a non-coding region of DNA and not in linkage disequilibrium with flanking genes (**Figure 3.7**). IGP30 was a representation of the abundance of disialylation among fucosylated diagalactosylated N-glycans with bisecting GlcNAc, whereas IGP33 and IGP34 are ratios of monosialylation over disialylated fucosylated N-glycans. Therefore, this polymorphism may be linked to the regulation of disialylation.

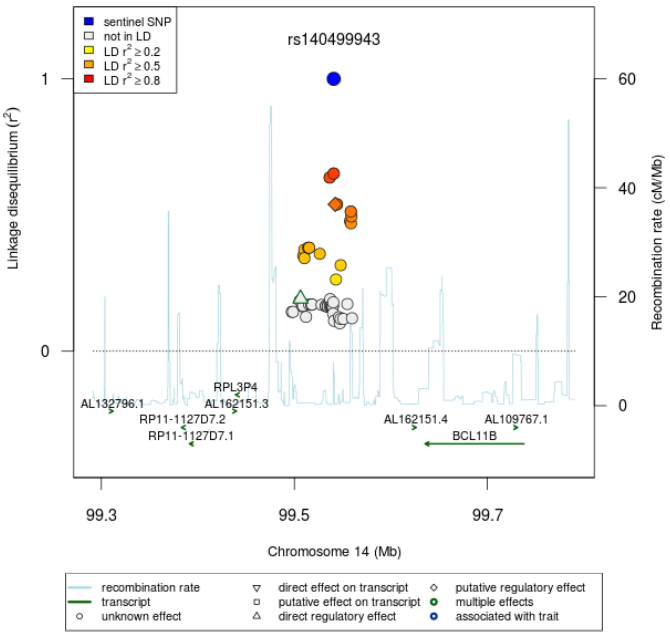


Figure 3.7: *Linkage disequilibrium plot of SNP with a suggestive association to IGP30, IGP33 and IGP34. Further derivation details are given in* ***Appendix 1***

A suggestive positive association (rs57331048) was indicated within the locus 15q11.2 for IGP58 (**Figure A2.43**); the abundance of total fucosylated structures. Although within a non-coding genomic region, the polymorphism was in moderate linkage disequilibrium to nearby *SNRPN* [MIM 182279] (**Figure 3.8**). It was also the only significant SNP within the region, thus may represent a spurious association.

GP4, a GP containing fucosylated agalactosylated N-glycan structures, had a positive suggestive association with a SNP (rs7214156) within the locus 17q21.32 (**Figure A2.4**). Though located within the non-coding genomic region, this polymorphism was in weak linkage disequilibrium to neighbouring genes *TBKBP1* [MIM 608476] and *TBX21* [MIM 604895] (**Figure 3.9**). This locus also had a suggestive negative association with IGP57, digalactosylated N-glycans (**Figure A2.42**).

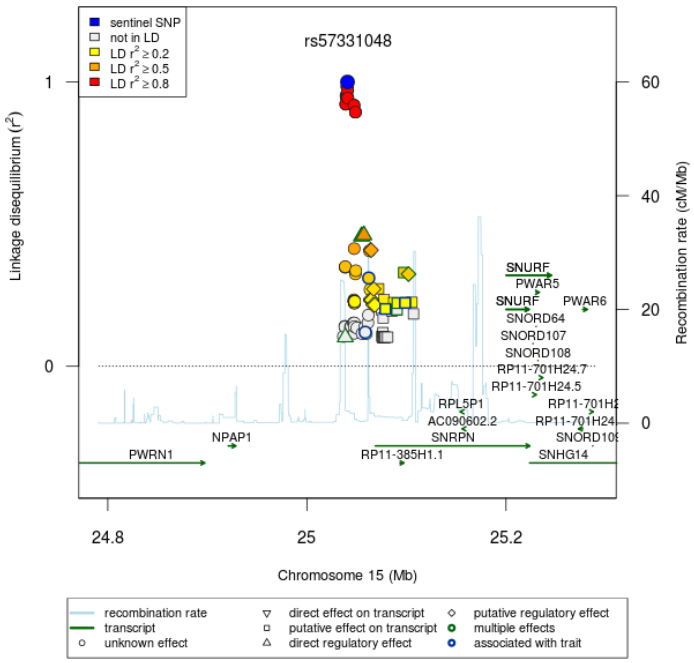


Figure 3.8: *Linkage disequilibrium plot of SNP with a suggestive association to fucosylated N-glycans. Further derivation details are given in* ***Appendix 1***

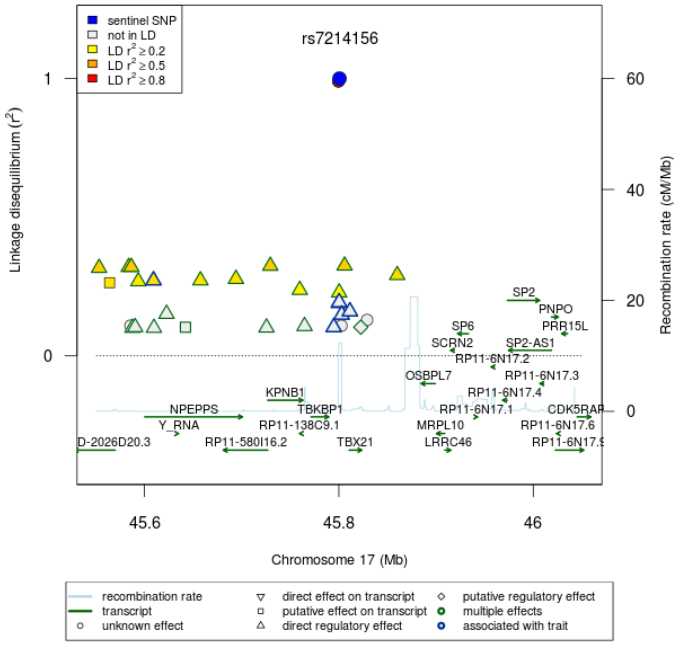


Figure 3.9: *Linkage disequilibrium plot of SNP with a suggestive association to GP4, fucosylated agalactosylated N-glycans*

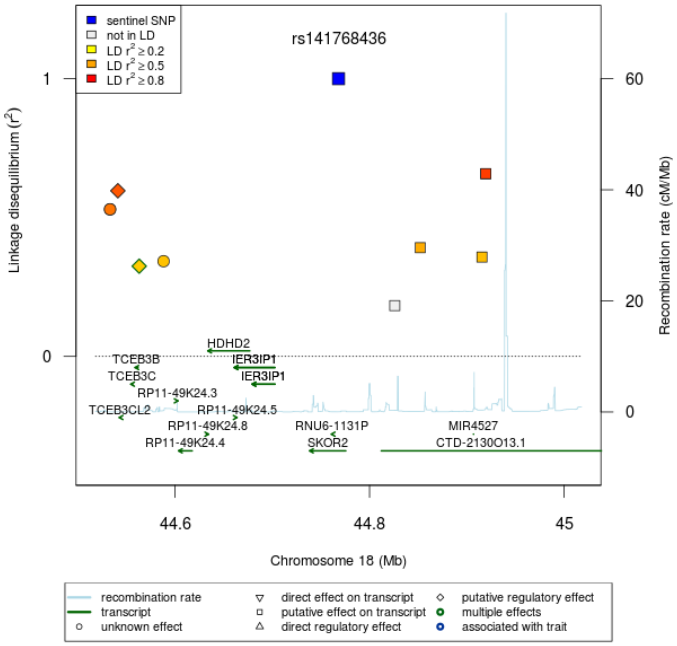


Figure 3.10: *Linkage disequilibrium plot of SNP with a suggestive association to GP8, fucosylated monogalactosylated N-glycans*

Lastly, GP8, a GP containing fucosylated monosialylated N-glycans, was suggestively negatively associated with a SNP (rs141768436) within the locus 18q21.1 (**Figure A2.8**). The polymorphism lies within the coding region of *SKOR2* [MIM 617138]. While it was the only hit within this locus, and so may represent a spurious association, it was in linkage disequilibrium to SNPs of flanking genes (**Figure 3.10**).

3.4 Discussion

This study identified gQTLs of IgG. The aim was to validate previous GWAS findings for IgG N-glycosylation. Importantly, this was the first GWAS of IgG N-glycosylation features in an Australian population. The N-glycans were quantitated using UPLC, which were found to perform better in GWAS when compared with N-glycans measured by mass spectrometry (18). Though a file was corrupt for chromosome 5, we were able to provide GWAS based on the other 21 autosomes in a total of 614 BHAS participants.

3.4.1 Validating gQTLs of IgG

Out of 56 GWAS, there were nine N-glycosylation features found to have genome-wide significance with two gQTLs. Importantly, *ST6GAL1* and *MGAT3* were identified in a previous study of Caucasians from four European populations: two islands in Croatia (Vis and Korcula), the Orkney Islands in the UK, and Sweden (18). The identified genes within these gQTLs are glycosyltransferases, validating previously identified gQTLs specific to IgG glycoforms.

ST6GAL1 encodes sialyltransferase 6 that catalyses the transfer of Neu5Ac to the N-glycan moieties (115). The enzyme is localised in the membrane of the Golgi apparatus and is specific to B cells in the later stages, following differentiation. Mutations within the *ST6GAL1* are consistently identified in GWAS of IgG N-glycosylation (18, 105, 116). This locus had genome-wide significance to seven IgG N-glycosylation features, six of which represented abundance of various monosialylated N-glycans. The IgG N-glycan feature with the highest association to *ST6GAL1* was IGP29, which was previously identified as the most associated feature (18, 116). Polymorphisms within *ST6GAL1* have pleiotropy to type 2 diabetes mellitus (T2DM) (18). Interestingly, altered sialylation of IgG glycoforms have recently been implicated to occur in T2DM among European Caucasians (39). Thus, this finding may have biological significance.

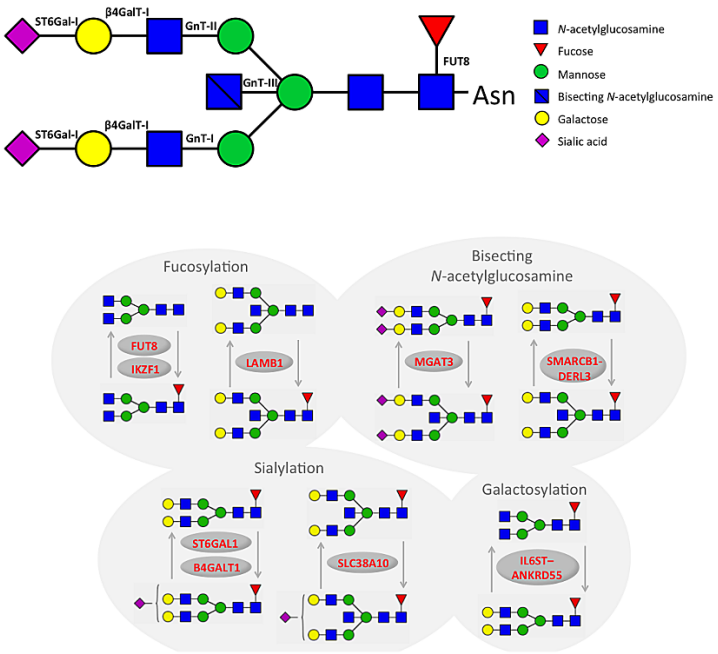


Figure 3.11: *Genes whose glyco-enzymes function to add and remove specific monosaccharides from branching structures (reproduced from Lauc et al 2013)*

The gQTL containing *SYNGR1-TAB1-MGAT3* had many polymorphisms with genome-wide significant associations; another consistently identified gQTL among GWAS (18, 105, 116). The SNP with the highest association was within the *MGAT3* genomic region, a gene that encodes mannosyl (β-1,4)-glycoprotein β-1,4-*N*-acetylglucosaminyltransferase, involved in the addition of a bisecting GlcNAc. This N-glycosylation feature is associated with E-cadherin, EGF-, Wnt- and integrin- cancer-associated signalling pathways (117). Altered MGAT3 has pleiotropy with Alzheimer's disease (118) and inflammatory bowel disease (119). *SYNGR1* and *TAB1* were shown to be in linkage disequilibrium with *MGAT3*. They encode synaptogyrin 1 and TGF-beta activated kinase 1, respectively. Polymorphisms within *SYNGR1* have pleiotropy with schizophrenia and bipolar disorder (120, 121), whereas TAB1 may have pleiotropy with colorectal cancer (122).

3.4.2 Potentially Spurious Strongly Suggestive gQTLs

None of the strongly suggestive gQTLs have been implicated in previous GWAS on IgG glycoforms. This may indicate true associations unique to the Australian population, or most probably by chance associations inherent in multiple testing. It is for this reason

that the *p*-value indicating genome-wide significance, as derived previously in similar data (18), is very small.

GP4, abundance of fucosylated agalactosylated N-glycans, had a positive suggestive association with a SNP located in a non-coding DNA region, however in moderate linkage disequilibrium with flanking genes *TBKBP1* and *TBX21* (**Figure 3.9**). Interestingly, both genes encode proteins within immune-modulating pathways. *TBKBP1* encodes an adapter protein that binds TBK1 and is part of the tumour necrosis factor (TNF)/nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB) pathway (123). Moreover, *TBX21* encodes a T helper 1 (Th1)-specific T-box transcription factor that controls the expression of the Th1 cytokine, interferon-gamma (IFNγ) (124). Increased levels of GP4 are hypothesised to infer an increase in IgG glycoforms that are pro-inflammatory in nature (108). Therefore, although not specifically related to the biosynthesis pathway of GP4, the positive association found may indicate increased inflammation in these individuals and a complex network between these IgG glycoforms and the given immune-modulating pathways.

Several of the strongly suggestive gQTLs have been implicated in health research. It was found that in conjunction with long non-coding RNA HCG11 and MTA3, *miRNA4425* encoded a miRNA that suppressed growth of gliomas (125). There was a negative association between IGP68, a ratio of bisecting GlcNAc among fucosylated monogalactosylated N-glycan moieties, and the *miRNA4425* polymorphism. Bisecting GlcNAc associates with a decrease in core fucosylation, therefore enhanced initiation of ADCC (108). *SKOR2* encodes SKI family transcription corepressor 2, a highly expressed in the central nervous system (126). Lastly, *SNRPN* encodes small nuclear ribonucleoprotein polypeptide N and is implicated in the genetic disorders Prader-Willi syndrome and Angelmen syndrome (127). The polymorphism within *SNRPN* in our study was positively associated with IGP58, the abundance of core fucosylation among asialylated N-glycans.

Two of the strongly suggestive gQTLs contained genes that are ubiquitously expressed in human tissues. *OSBPLS3* encodes oxysterol binding protein like 3, involved in regulating cell adhesion and organisation of actin cytoskeleton (128), as well as coordinating transport, signalling and metabolism of sterol and phosphoinositide lipids (129). *ERICH1* is expressed ubiquitously in human tissues though not a great deal of research has been conducted for this gene (NCBI gene ID157697). Finally, *CMAHP* is a pseudogene, which is expressed in mammals but not humans. It is inactivated in humans by a deletion that renders its enzyme non-functional (130)

3.4.4 Limitations

Given the sample size, this study was underpowered, so the detection of novel polymorphisms was not expected. This power would have been increased through implementing meta-analysis with other cohorts, though this was beyond the scope of the thesis. Finally, the chromosome 5 file was corrupt. However, haplotype reference consortium (HRC) imputed genotype data has been obtained and will be used to rerun analyses when time permits. Through this we may be able to validate potentially spurious associations of these GWAS.

3.4.5 Conclusion

The aim of this study was to verify gQTLs previously found in European Caucasian populations. GWAS were implemented for each IgG N-glycan feature (**Appendix 1**), ultimately leading to the identification of gQTLs of IgG glycoforms within this Australian population. Two gQTLs were validated; *ST6GAL1* and *MGAT3*. Therefore, even with the underpowered GWAS, this study suggested genetic homogeneity between Caucasians of European and Australian origin. It is prudent to underscore the importance of doing such, given the novelty of the other studies in terms of the data and statistical methods proposed. This study demonstrates that gQTLs do not fully explain heterogeneity within the IgG N-glycome, thus warranting further investigation.

Chapter 4 – Further Exploring the Role of Genetics in IgG N-Glycosylation by Implementing RNA-Seq

4.1 Prologue

The study of the gene expression profiles pertaining to some phenotype of interest is biologically important since not all genes are expressed in a given tissue (e.g. blood in this case). Moreover, not all genetic polymorphisms associated with IgG N-glycosylation perpetuate in a given biological system. Thus, gene expression data may evidence novel gene associations.

Although there has been varied types of IgG N-glycan genetic analyses, to date there has been no work using gene expression data. One of the main barriers to this has been the associated costs of running such analyses on a large enough sample to achieve sufficient statistical power. Indeed, this was also a consideration during the experimental design process with selecting how many samples we wished to analyse. Further considerations were needed when using one 'omics dataset to explain another.

As a means of high-throughput analysis of gene expression, RNA-Seq was used to analyse fractionated poly(A)+ RNA from isolated leucocytes (mRNA-Seq). It is prudent to underscore the benefits RNA-Seq provides when compared with microarray analyses. Unlike microarrays, RNA-Seq is not limited to detecting transcripts that correspond to existing knowledge of genome sequences or the probes selected to interrogate the cDNA sample (131, 132). Additionally, RNA-Seq has very low background noise relative to microarrays, and is superior in detecting low abundance transcripts as well as differentiating between biologically critical isoforms (131, 132). Finally, RNA-Seq and microarray platforms using the same set of samples will show gene expression results that are highly correlated (132). Therefore, gene expression analysis utilising RNA-Seq was the optimal choice for this study.

Herewith I present the research paper describing this work. The median quotient normalised IgG N-glycan dataset was used in this study due to implementing multivariate statistics to integrate the two 'omics datasets.

4.2 Research Paper: Integrating mRNA Sequencing Data with the Immunoglobulin G *N*-Glycome Using Extensions of Differential Expression Analysis and Orthogonal Two-Way Partial Least Squares

Section 4.2 is not available in this version of the thesis.

4.3 Epilogue

RNA-Seq was implemented to further explore the genetic influence on the heterogeneity of the IgG N-glycome, in terms of the transcribed genes and not just gQTLs (**Chapter 3**). It was used to quantify the abundance of mRNA within leucocytes. The decision to explore total leucocyte mRNA rather than targeting B-cell lymphocytes transpired since both endogenous and exogenous factors are known to alter IgG N-glycome composition. Termed the "cellular environment" in this thesis, it refers to the collection of factors that may influence N-glycosylation and in turn impact the inflammatory properties of the IgG effector glycoproteins. Indeed, the IgG N-glycome is malleable as it is reliant on the expression levels of the glycosyltransferases and glycosylhydrolases within the B-cell, as well as the abundance of sugar nucleotide donors, which act to remodel the branching IgG N-glycan structure during biosynthesis (**Section 1.2.4**).

It was resolved that almost 60% of the heterogeneity of IgG N-glycans could be explained by the differential expression of leucocyte genes in this Australian population, using a multivariate model to assimilate all variables from both datasets. This evidenced strong genetic influence in the overall N-glycosylation of IgG. It further supports the notion that variation in the expression of glyco-enzymes results in diversification of the IgG glycoforms. Expression of these is clearly affected by exogeneous factors during B-cell activation and maturation (153). Hence, there exists a complex interplay in the epigenetic regulation of IgG N-glycosylation.

Age- and sex-adjusted linear regression models were applied to determine genes whose differential expression was associated with the altered abundance of the IgG GPs. I predicted that targeting these GPs as separate phenotypes would allow for the identification of mRNA that were associated with inclusions to the branching structures unique to each IgG N-glycan moiety, much like the methods used in previous GWAS and in **Chapter 3**. After controlling for a 5% false discovery rate, I found only one gene transcript to be differentially expressed in terms of GP6, 105 differentially expressed in relation to GP23, and 372 gene transcripts associated with GP10 (**Appendix 3**). Since the greater number of differentially expressed genes was thought to better inform GO terms, GP10 remained the focus for the rest of this study. The full list of GO enrichment terms for GP10 and GP23 are in **Appendix 4**.

GP10 consists of core fucosylated biantennary monogalactosylated N-glycan moieties with bisecting GlcNAc. Though it was expected an increased expression of cytokines, glyco-enzymes and other immune-related factors would be detected, the

identified differentially expressed genes were mostly associated with the protein portion of the IgG glycoprotein (**Table 4.1**). After further interrogating these results, many of the related GO terms were Fc domain effector responses; that is, FcR and complement binding (**Figure 4.2**). These immune-related GO terms also tended to decrease as GP10 abundance increased. Present bisecting GlcNAc and core Fuc within an IgG glycoform, both N-glycosylation features of GP10, are associated with a decrease in Fcγ-RIIIa activity (**Figure 1.4**); an FcR that elicits ADCC (23, 108). This underscores a potential biological explanation for these negative associations. It was also indicative that epigenetic factors influencing IgG biosynthesis may be shared between the peptide and N-glycan portions of the glycoprotein.

The lack of identified differentially expressed genes within the other GPs may have been due to an underpowered study. Like the GWAS performed in **Chapter 3**, however, there were existent constraints. Namely, RNA-Seq is expensive, though it has improved in recent years. For this reason, 350 extracted RNA samples were tested for integrity before RNA-Seq was performed on 305, with a final 303 sequences passing pre-processing procedures and included in the analyses. As this was the first use of integrating mRNA transcriptome data with IgG N-glycan moieties, the given results still add to the growing IgG N-glycomics literature. They may also be used to calculate power and set precedence for study design in future such studies.

Chapter 5 – Clinical Measures and Health Behaviours Associated with IgG N-Glycosylation

5.1 Prologue

Clinical measures are arguably the most published factors to contribute to IgG N-glycome heterogeneity, with many studies arising from my own research team (2, 5, 38, 41) and others (9, 29, 93, 154). Indeed, genetics plays a key role in the baseline risk of chronic and the relative abundance of IgG N-glycan features. However, the biosynthesis of the branching N-glycans within IgG Fc are regulated by several endogenous and exogenous factors, including those found in the blood or routinely measured in the clinical health settings. In this sense, the IgG N-glycome is malleable as it is reliant on the expression levels of the glyco-enzymes as well as the abundance of sugar nucleotide donors during biosynthesis.

I explored existing as well as less commonly analysed clinical measures and health behaviours in this chapter. Contained are two published manuscripts, plus further unpublished analyses. The two publications highlight associations with several measured clinical factors, whereas the latter underscores the potential to prescribe a change in health behaviours to modify the IgG N-glycome, with the aim of improving overall health.

Firstly, I explored whether more of the IgG N-glycome heterogeneity could be explained through other specific or accurate measures of body fat distribution other than BMI, such as measures targeting central adiposity or using sophisticated equipment (**Section 5.2**). A previous study (97) explored the pairwise-associations of the IgG N-glycan features (measured IgG GPs and the derived N-glycosylation traits) and BMI using area normalised N-glycome data. Subsequently, to explore whether I could validate or improve on previous findings, I used the area normalised IgG N-glycome data to explore the relationship to the distribution of adiposity (see **Section 2.5**).

The novelty of this study was three-fold: it identified the adiposity measures associated with IgG N-glycan features; it explored how these associations differ by the type of adiposity measure (that is, differential association according to total or central adiposity measures); and, it verified the reported pairwise-association of the relative abundance of agalactosylated IgG glycoforms and BMI. It should be noted that Perkovic and colleagues (97) only found a single statistically significant but weak pairwise-association in their study; BMI explaining only 2% of the variation of

agalactosylated IgG glycoforms. Thus, I expected at least an increase in the number of IgG N-glycan features to be associated with the other body adiposity measures.

The second paper was a multivariate analysis of a collection of health indicators and the IgG N-glycome (**Section 5.3**). Certainly, a lot of research exists that relates health indicators to the IgG N-glycome in several populations. These previous findings, however, have been derived using univariate statistics and thus may overestimate given associations. The novelty of this paper was the implementation of multivariate statistics.

The model used was an extension of Pearson's correlation coefficient, namely Canonical Correlation Analysis (CCA), which evaluated the simultaneous associations between the set of predictor variables (the collection of health indicators) and the set of outcome variables (the IgG N-glycan features). The model reduces the number of dimensions in the dataset while also maximising the correlation coefficient between the new latent variables, highlighting which variables in each of the two datasets tended to be the most strongly related to the other dataset. The median quotient normalised IgG N-glycan data (**Section 2.5**) were used for this publication since area normalisation introduces collinearity into the data, which would be an undesirable trait and violate the assumptions of this statistical model.

Lastly, I analysed the association of IgG N-glycome composition with the modifiable risk factors physical activity, smoking and alcohol consumption (**Section 5.4**). The precedence for this final part of the chapter was the nature of the IgG N-glycan features themselves; that is, my team hypothesises they are an interphenotype of health and disease. Therefore, if in the future we can specify and indicate an individuals' discordance from a desirable IgG N-glycan profile, then ethically we should be able to prescribe modifications to their health behaviours to improve these profiles. Thus, these modifiable risk factors may be simple lifestyle changes, if they are indeed associated with the IgG N-glycome. I used the median quotient normalised IgG N-glycome data for this sub-study (**Section 2.5**).

5.2 Research Paper: Increased Central Adiposity is Associated with Pro-Inflammatory IgG N-Glycans

5.3 Research Paper: Increases in Cardiometabolic Factors Tend to Associate with a Larger Pro-Inflammatory Fraction of the Immunoglobulin G N-Glycome: A Multivariate Analysis

Sections 5.2 and 5.3 are not included in this version of the thesis.

Page intentionally left blank

5.4 Modifiable Health Behaviours and Aberrant IgG N-Glycosylation

5.4.1 Introduction

Proposed to represent interphenotypes of health status, the IgG N-glycans are an emerging biomarker integrated into research focusing on the health paradigm shift from reactive medicine towards predictive, preventive and personalised medicine. IgG is an important effector glycoprotein within the immune system, which exerts both anti-inflammatory and pro-inflammatory responses (108). These immune responses are largely modulated by the Fc domain through its affinity for several FcRs and complement factors (**Figure 1.4**). The N-glycans contribute to this altered affinity, ultimately initiating different cellular events that induce an array of inflammatory responses (23, 24, 50, 108).

Variation among several clinical factors, particularly those related to cardiometabolic health, associate with the heterogeneity of IgG N-glycosylation (**Section 5.2** and **Section 5.3**). Importantly, it was evident that poor health status as measured by central adiposity as well as CRP, insulin, TG, FBG, and SBP, tended to be associated with increased pro-inflammatory IgG glycoforms. These identified exogenous factors are thought to represent instrumental variables of health behaviour and environmental exposure over the life-course. These cellular environmental factors may act to alter the expression of the glyco-enzymes, affecting the IgG N-glycan elongation and trimming stages of biosynthesis. Thus, I propose that an adjustment in health behaviours aimed at improving overall health status may have a downstream impact on IgG N-glycome and be pertinent to the pathophysiology of existent subclinical or clinical phenotypes.

Behaviours that may be altered for the purposes of improving overall health status, known as modifiable risk factors, were the focus of this study. Aside from obesity, other prominent modifiable risk factors include physical inactivity, smoking and excessive alcohol consumption (184). Importantly, physical inactivity and smoking were recently identified among the most influential factors contributing to poor health status (185).

5.4.2 Methods

5.4.2.1 Participants

A total of 637 participants were included in this study. Summary data and further details about the participants can be found in **Section 1.3.2** and **Table 1.3**.

This study conformed to the ethical guidelines of the 1975 Declaration of Helsinki.

5.4.2.2 IgG N-Glycan Derivation and Pre-Processing

The procedures for the UPLC-derived IgG N-glycans are described in **Chapter 2**, as are the pre-processing steps used to create the final dataset. The median quotient normalised IgG N-glycan data was used for the below analyses (**Section 2.5**).

5.4.2.3 Estimating Physical Activity

Physical activity was assessed using the self-administered, short-form of the International Physical Activity Questionnaire (IPAQ), which retrospectively collects information on different levels of physical activity carried out over the previous week, as well as time spent sitting (186). The IPAQ activities were converted to continuous scales of MET-minutes per week (a measure of metabolic equivalence between the different levels of vigorous, moderate or walking physical activity during the week proceeding assessment). Categorical levels were derived (Highly Active, Moderately Active, or Insufficiently Active) according to the guidelines set out previously for High, Moderate and Low activity levels, respectively (186).

5.4.2.4 Assessing Smoking Status

Smoking status was assessed through asking the participants about their current smoking status and frequency of cigarettes consumed per day, on average. Categories for Never Smoked, Ex-Smoker and Current Smoker were derived from these data.

5.4.2.5 Calculating Alcohol Consumption

The BHAS implement a version of the Food Frequency Questionnaire (FFQ) alcohol questions, which may be used to not only analyse alcohol in terms of reported intake but also frequency. These questions collected an overall quantity of alcohol intake, then frequencies of alcohol consumption by beverage type. The inclusion of an overall estimate of alcoholic drinks per day rather than beverage-specific intake is asked. Thus, some individuals may have drunk more than one type of alcoholic beverage on any given day, while others may only drink one type per day. To account for this, a weighted approach was used to derive frequency of alcohol consumption.

Alcohol consumption days were first defined as the maximum weekly frequency by summing all alcoholic beverage types (termed ‘max of days’), addressing the first assumption. Then the frequency of drinking may be no more than 7 days under the

second assumption (termed ‘sum of days’). The beverage-specific frequencies were, therefore, proportionately reduced to a total 7 days (that is, ‘sum of days’ = ‘max of days’ divided by the mean of ‘max of days’ and multiplied by 7). The ‘max of days’ and ‘sum of days’ calculations were weighted under the assumption that individuals who drink more than one type of alcohol were about twice as likely to do so on different days (187). Subsequently, the weighted frequency of alcohol intake was calculated by the summation of two thirds of ‘sum of days’ and one third of ‘max of days’. This weighted frequency was then multiplied by the reported number of drinks consumed per day, on average, to give estimated alcohol consumption per week.

Using the Australian alcohol consumption guidelines, an adult should not consume more than two standard drinks per day to avoid lifetime risk of alcohol-related disease or injury (188, 189). To assess regular drinking to excess, termed ‘excessive weekly consumption’, participants were categorised according to whether they drank more than 14 beverages over the span of a week. Moreover, an adult should not consume more than four standard drinks per day, termed ‘binge drinking’ (189). This was categorised by whether a participant reported drinking more than four drinks on a usual day.

5.4.2.6 Statistical Analysis

Chi-square tests were used to compare sex-specific proportions of the modifiable risk factor categories. Spearman’s correlation coefficient was used to assess the association between the different types of IPAQ continuous scales and the IgG GPs. Multivariate linear regression models, controlling for age and sex, were used to analyse the association of IPAQ categorisation, excessive alcohol consumption, binge drinking and smoking status to the IgG GPs. The false discovery rate was controlled for using the Benjamini-Hochberg method (190). A *p*-value less than 0.05 after adjustment provided sufficient evidence at the 5% level of significance. R version 3.6.1 (191) was used.

5.4.3 Results

The sex-specific summary statistics for the modifiable risk factors are given in **Table 5.4**. All the modifiable risk factors were statistically disproportionate between the sexes in this population. Hence, sex was controlled in regression analyses. A total of 26.8% of the participants were classified as insufficiently active, 15.1% participated in binge drinking, 26.8% drank to excess over a week and 8.5% were current smokers.

Table 5.4: *Sex-specific summary statistics of the modifiable risk factors*

	Females	Males	Statistic	*p*-value
n (%)	344 (54.0)	293 (46.0)	4.083[#]	0.043
Age, years $\bar{x}\ (SD)$	57.4 (5.2)	57.8 (5.2)	0.867[t]	0.387
Physical Activity *n (%)*				
Insufficient	102 (29.7)	69 (23.5)		
Moderate	138 (40.1)	80 (27.3)		
High	104 (30.2)	144 (49.1)	24.324[#]	5.22E-06
Alcohol Consumption *n (%)*				
Binge Drinking	14 (4.1)	82 (28.0)	63.752[#]	1.41E-15
Excessive Weekly Consumption	34 (9.9)	137 (46.8)	100.67[#]	1.09E-23
Smoking *n (%)*				
Never Smoked	220 (64.0)	161 (54.9)		
Ex-Smoker	101 (29.4)	101 (34.5)		
Current Smoker	23 (6.7)	31 (10.6)	6.279[#]	0.043

[#] chi-square statistic used to test proportions

[t] t-test used to compare means of groups

5.4.3.1 Physical Inactivity

Physical activity was assessed according to the MET-minutes spent in each activity type, on average. There were no evident associations between the IgG GPs and these IPAQ continuous scales (**Table 5.5** and **Figure 5.5**). Additionally, there were no associations between the categories of physical activity and the IgG GPs, controlling for age and sex (**Table 5.6**).

Table 5.5: *Spearman's correlation of different types of physical activity (in MET-minutes) and the IgG GPs. All p-values are adjusted for the 5% false discovery rate*

	Activity Type					
	Vigorous ρ, adj *p*-value		Moderate ρ, adj *p*-value		Walking ρ, adj *p*-value	
GP1	0.0228	0.941	0.0611	0.634	0.0222	0.941
GP2	-0.0008	0.997	0.0281	0.941	-0.0328	0.941
GP3	0.0354	0.941	0.0613	0.634	0.0151	0.941
GP4	-0.0056	0.941	0.0371	0.941	0.0308	0.941
GP5	-0.0182	0.941	0.0111	0.941	-0.0251	0.941
GP6	0.0000	0.999	0.0047	0.941	0.0319	0.941
GP7	0.0228	0.941	0.0132	0.941	-0.0332	0.941
GP8	0.0043	0.941	-0.0057	0.941	-0.0239	0.941
GP9	0.0200	0.941	0.0262	0.941	-0.0147	0.941
GP10	0.0181	0.941	0.0116	0.941	0.0310	0.941
GP11	0.0078	0.941	0.0216	0.941	0.0053	0.941
GP12	0.0430	0.941	0.0104	0.941	-0.0244	0.941
GP13	0.0070	0.941	-0.0426	0.941	-0.0470	0.941
GP14	0.0223	0.941	-0.0128	0.941	0.0046	0.941
GP15	0.0241	0.941	-0.0057	0.941	0.0234	0.941
GP16	-0.0395	0.941	0.0170	0.941	-0.0107	0.941
GP17	-0.0216	0.941	-0.0346	0.941	-0.0856	0.432
GP18	-0.0122	0.941	-0.0257	0.941	-0.0099	0.941
GP19	-0.0882	0.432	-0.0756	0.443	0.0065	0.941
GP20	-0.0292	0.941	-0.0784	0.432	-0.0165	0.941
GP21	-0.0672	0.589	-0.1017	0.253	-0.0786	0.432
GP22	-0.0652	0.601	-0.0811	0.432	-0.0242	0.941
GP23	-0.0741	0.443	-0.0578	0.696	-0.0220	0.941
GP24	-0.1138	0.253	-0.1013	0.253	0.0054	0.941

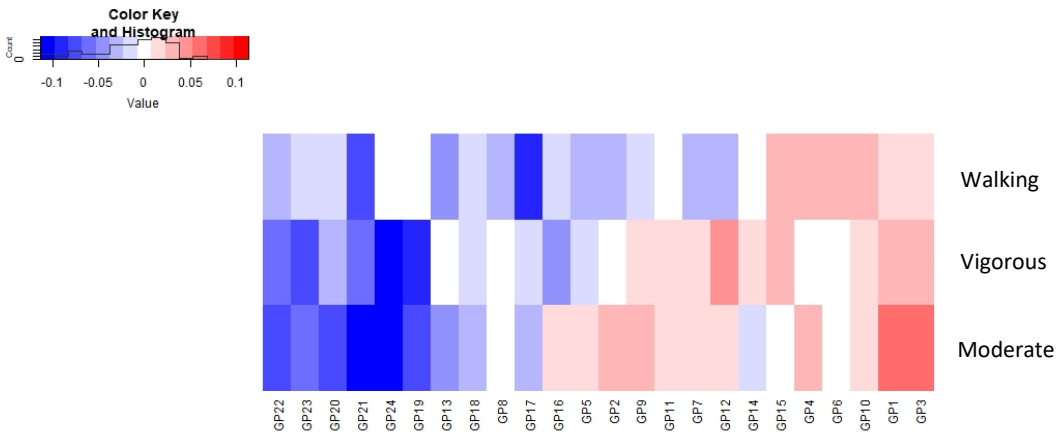


Figure 5.5: *Heatmap of the association between the IgG GPs and different classes of physical activity, correlated using Spearman's Rho*

Table 5.6: *Linear regression models of level of physical activity (ref = "insufficient") and the IgG GPs, controlling for age and sex. All p-values are adjusted for the 5% false discovery rate*

	Physical Activity	B	SE	t	Adj. *p*-value
GP1	**Moderate**	0.002	0.003	0.591	0.856
	High	0.002	0.003	0.538	0.856
GP2	**Moderate**	0.045	0.047	0.961	0.856
	High	0.004	0.046	0.082	0.975
GP3	**Moderate**	0.002	0.003	0.691	0.856
	High	0.002	0.003	0.624	0.856
GP4	**Moderate**	0.405	0.530	0.764	0.856
	High	0.272	0.521	0.521	0.856
GP5	**Moderate**	-0.001	0.003	-0.519	0.856
	High	-0.003	0.003	-1.082	0.856
GP6	**Moderate**	-0.011	0.150	-0.074	0.975
	High	0.080	0.148	0.543	0.856
GP7	**Moderate**	0.011	0.024	0.470	0.856
	High	-0.001	0.024	-0.032	0.975
GP8	**Moderate**	-0.116	0.174	-0.669	0.856
	High	0.031	0.171	0.184	0.954
GP9	**Moderate**	0.190	0.145	1.317	0.856
	High	0.011	0.142	0.079	0.975
GP10	**Moderate**	-0.056	0.109	-0.513	0.856
	High	0.138	0.107	1.286	0.856
GP11	**Moderate**	-0.006	0.014	-0.452	0.856
	High	0.000	0.014	-0.032	0.975
GP12	**Moderate**	-0.040	0.059	-0.685	0.856
	High	-0.048	0.058	-0.825	0.856
GP13	**Moderate**	-0.003	0.006	-0.440	0.856
	High	-0.007	0.006	-1.101	0.856
GP14	**Moderate**	-0.213	0.297	-0.719	0.856
	High	0.053	0.292	0.183	0.954
GP15	**Moderate**	-0.013	0.036	-0.352	0.916
	High	0.023	0.036	0.637	0.856

Table 5.6: *Cont.*

	Physical Activity	B	SE	t	Adj. *p*-value
GP16	**Moderate**	0.052	0.049	1.066	0.856
	High	-0.036	0.048	-0.739	0.856
GP17	**Moderate**	-0.017	0.027	-0.632	0.856
	High	-0.042	0.026	-1.617	0.856
GP18	**Moderate**	-0.106	0.198	-0.535	0.856
	High	-0.101	0.195	-0.518	0.856
GP19	**Moderate**	0.026	0.038	0.681	0.856
	High	-0.038	0.038	-1.006	0.856
GP20	**Moderate**	0.002	0.007	0.285	0.950
	High	-0.009	0.007	-1.270	0.856
GP21	**Moderate**	-0.006	0.011	-0.502	0.856
	High	-0.024	0.011	-2.243	0.744
GP22	**Moderate**	0.004	0.005	0.837	0.856
	High	-0.006	0.005	-1.196	0.856
GP23	**Moderate**	-0.011	0.041	-0.264	0.950
	High	-0.088	0.041	-2.163	0.744
GP24	**Moderate**	0.009	0.042	0.221	0.954
	High	-0.061	0.041	-1.475	0.856

5.4.3.2 Excessive Alcohol Consumption

Excessive alcohol consumption was analysed according to the current Australian guidelines; binge drinking (more than four drinks in one sitting) and excessive drinking (more than fourteen drinks in one week). Binge drinking was positively associated with GP4, and negatively associated with GP13 and GP14 (**Table 5.7**), while excessive drinking was negatively associated with GP7, GP12 and GP13 (**Table 5.8**). Although neither significantly explained more IgG GPs than the other, this finding is in line with current literature; that is, binge drinking and prolonged excessive consumption of alcohol impact health status (192).

GP4 is the most prevalent agalactosylated structure in the N-glycome and is proposed to elicit pro-inflammatory immune responses (**Figure 1.4**). Moreover, GP7 is monogalactosylated and GP12, GP13 and GP14 are digalactosylated, decreased abundance of the latter is also proposed to be linked to pro-inflammatory responses.

Indeed, the direction of the effect sizes given by the models (**Table 5.7** and **Table 5.8**) suggests that both binge drinking and the excessive regular consumption of alcohol were associated with an increase in pro-inflammatory IgG glycoforms.

Table 5.7: *Linear regression models assessing the effect of binge drinking (ref = normal levels of alcohol consumption) on the IgG GPs, controlling for age and sex. All p-values are adjusted for the 5% false discovery rate*

	B	SE	t	Adj. *p*-value
GP1	-0.001	0.004	-0.282	0.849
GP2	-0.002	0.053	-0.037	0.970
GP3	0.005	0.003	1.602	0.220
GP4	1.740	0.616	2.823	**0.040**
GP5	-0.004	0.004	-0.992	0.453
GP6	0.468	0.178	2.621	0.054
GP7	-0.044	0.028	-1.560	0.220
GP8	-0.186	0.203	-0.914	0.481
GP9	-0.123	0.174	-0.706	0.608
GP10	0.178	0.130	1.369	0.285
GP11	0.038	0.016	2.302	0.088
GP12	-0.150	0.069	-2.158	0.102
GP13	-0.024	0.007	-3.251	**0.012**
GP14	-1.112	0.347	-3.207	**0.012**
GP15	0.072	0.043	-1.670	0.207
GP16	0.071	0.059	1.201	0.345
GP17	-0.062	0.032	-1.963	0.133
GP18	-0.553	0.234	-2.367	0.086
GP19	0.016	0.045	0.347	0.845
GP20	-0.018	0.009	-2.130	0.102
GP21	-0.004	0.013	-0.334	0.845
GP22	0.000	0.006	-0.084	0.970
GP23	-0.088	0.049	-1.815	0.168
GP24	0.065	0.048	1.348	0.285

Table 5.8: *Linear regression models of excessive weekly drinking (ref = normal levels of alcohol consumption) and the IgG GPs, controlling for age and sex. All p-values are adjusted for the 5% false discovery rate*

	B	SE	t	Adj. *p*-value
GP1	-0.003	0.003	-0.997	0.477
GP2	-0.061	0.045	-1.366	0.375
GP3	0.002	0.003	0.780	0.573
GP4	0.965	0.525	1.839	0.275
GP5	-0.006	0.003	-1.824	0.275
GP6	0.146	0.152	0.959	0.477
GP7	-0.073	0.024	-3.098	**0.025**
GP8	-0.081	0.172	-0.468	0.668
GP9	-0.111	0.148	-0.749	0.573
GP10	0.110	0.110	0.998	0.477
GP11	0.017	0.014	1.194	0.430
GP12	-0.171	0.059	-2.920	**0.029**
GP13	-0.022	0.006	-3.578	**0.009**
GP14	-0.458	0.296	-1.545	0.328
GP15	-0.043	0.036	-1.181	0.430
GP16	0.084	0.050	1.684	0.318
GP17	-0.064	0.027	-2.386	0.104
GP18	-0.081	0.199	-0.406	0.685
GP19	0.019	0.038	0.500	0.668
GP20	-0.011	0.007	-1.553	0.328
GP21	-0.006	0.011	-0.584	0.668
GP22	-0.005	0.005	-1.149	0.430
GP23	-0.021	0.041	-0.515	0.668
GP24	0.061	0.041	1.484	0.332

5.4.3.3 Smoking Status

Current smoker status was positively associated with seven IgG GPs (GP2, GP6, GP7, GP10, GP11, GP15 and GP24), and negatively associated with GP9 and GP23 (**Table 5.9**). However, participants who identified as ex-smokers were not statistically different to those who have never smoked. This could indicate potential to alter the IgG N-glycome by quitting smoking, since any associations appear to be due to current exposure. Thus, smoking may have more of an impact on IgG N-glycosylation compared with the other two health behaviours. There were only 54 cases of current smokers (**Table 5.4**); thus, a study of the effect of the number of cigarettes per day on the IgG GPs would have been underpowered and so was omitted.

Table 5.9: *Linear regression models of smoking status (ref = "never smoked") and the IgG GPs, controlling for age and sex. All p-values are adjusted for the 5% false discovery rate*

	Smoking Status	B	SE	t	Adj. *p*-value
GP1	**Ex**	-0.001	0.003	-0.247	0.859
	Current	-0.004	0.004	-0.940	0.617
GP2	**Ex**	-0.004	0.040	-0.097	0.942
	Current	0.223	0.066	3.357	**0.007**
GP3	**Ex**	-0.002	0.003	-0.859	0.637
	Current	-0.002	0.004	-0.500	0.732
GP4	**Ex**	-0.544	0.454	-1.197	0.552
	Current	-0.473	0.759	-0.623	0.712
GP5	**Ex**	0.000	0.003	0.136	0.930
	Current	0.010	0.004	2.331	0.087
GP6	**Ex**	0.049	0.126	0.387	0.799
	Current	1.087	0.211	5.152	**0.000**
GP7	**Ex**	-0.018	0.021	-0.885	0.637
	Current	0.094	0.035	2.707	**0.042**
GP8	**Ex**	-0.126	0.149	-0.846	0.637
	Current	-0.151	0.249	-0.607	0.712
GP9	**Ex**	-0.008	0.123	-0.068	0.946
	Current	-0.759	0.205	-3.698	**0.002**
GP10	**Ex**	0.215	0.091	2.362	0.087
	Current	0.988	0.152	6.516	**0.000**
GP11	**Ex**	0.024	0.012	2.082	0.121
	Current	0.074	0.020	3.771	**0.002**
GP12	**Ex**	-0.038	0.050	-0.755	0.675
	Current	0.086	0.084	1.019	0.573
GP13	**Ex**	-0.003	0.005	-0.600	0.712
	Current	0.011	0.009	1.300	0.490
GP14	**Ex**	0.290	0.253	1.144	0.552
	Current	-0.757	0.423	-1.789	0.209
GP15	**Ex**	0.069	0.031	2.256	0.087
	Current	0.166	0.051	3.232	**0.009**

Table 5.9: *Cont.*

	Smoking Status	B	SE	t	Adj. *p*-value
GP16	**Ex**	0.021	0.042	0.489	0.732
	Current	-0.072	0.071	-1.016	0.573
GP17	**Ex**	-0.026	0.023	-1.155	0.552
	Current	0.020	0.038	0.524	0.732
GP18	**Ex**	0.176	0.169	1.037	0.573
	Current	-0.481	0.283	-1.701	0.238
GP19	**Ex**	0.009	0.033	0.272	0.857
	Current	0.125	0.055	2.280	0.087
GP20	**Ex**	0.004	0.006	0.621	0.712
	Current	-0.003	0.011	-0.308	0.846
GP21	**Ex**	0.007	0.010	0.733	0.675
	Current	0.032	0.016	2.034	0.127
GP22	**Ex**	0.003	0.004	0.732	0.675
	Current	0.016	0.007	2.240	0.087
GP23	**Ex**	-0.019	0.035	-0.549	0.732
	Current	-0.308	0.058	-5.282	**0.000**
GP24	**Ex**	0.039	0.036	1.086	0.573
	Current	0.158	0.060	2.649	**0.044**

5.4.4 Discussion

The premise of this sub-study was to further explore clinical factors in terms of health behaviours, as potential contributors to IgG N-glycome heterogeneity. Aside from obesity (**Section 5.2**), other heavily publicised modifiable risk factors in health campaigns include physical inactivity, smoking and excessive alcohol consumption. Further, these factors contribute to decline in cardiometabolic health, especially in middle-aged and elderly individuals (184). Increasing physical activity, lowering alcohol consumption or quitting smoking may provide a potential window for consciously altering IgG N-glycosylation, which is considered malleable. Indeed, modifications to health behaviours may be prescribed to improve IgG effector responses or more generally, overall health status. The latter is particularly important with the advent of tests utilising the IgG N-glycans to model biological age. Biological age may have been impacted by several health behaviours and environmental exposures over the life-course. Certainly, testing

biological age is an important initiative amongst the shift towards predictive, preventive and personalised medicine.

It remains to be determined how individuals will truly benefit from the knowledge of subclinical morbidity risk or discordant biological age, particularly if much higher than chronological age. There needs to exist a body of research into how prescribed changes in health behaviours may improve biological age, and by extension overall health status. Changes to physical activity, alcohol consumption or smoking make an ideal target since they are the simplest to modify in the literal sense; that is, you make a conscious effort to alter your health behaviours. Nevertheless, it should be noted interplay between these health behaviours, mental status (particularly untreated mental disorders) and addiction complicates this prospect (193-195).

Physical activity is ubiquitously promoted as improving health status, mental well-being, mobility, and even patient's response and ability to tolerate therapies (196-198). In fact, it is emerging as an effective complementary treatment in cancer therapies (199, 200). The estimated prevalence of insufficient physical activity in this population reflected the 2014/2015 National Health Survey (26.8% vs 29.7% estimated by the National Health Survey). However, physical activity did not explain any of the variation in the IgG GPs, even after controlling for age and sex.

The self-report IPAQ was used to gather information on physical activity within the 7 days prior to assessment. Also, adherence to physical activity is low (201). Therefore, the lack of explanatory potential may be due to inconsistent participation in physical activity. Indeed, consistent increased levels of physical activity bring several health benefits, while concurrently acting as a preventative measure for cardiometabolic disorders and associated risk factors, cognitive decline, depression, falls and fractures, and osteoporosis (197, 202-204).

The excessive consumption of alcohol presents itself as another public health challenge (192). Importantly, it is not only consistent excessive intake that is a health concern, but binge drinking has been found to be as detrimental to health (205). Both were associated with a few IgG GPs, though these significant associations were weak. Nonetheless, the analysis sets precedence to further explore the impact of lowering alcohol intake on IgG N-glycome composition over time.

Participants indicating a current smoking status were found to have several significantly different IgG GPs when compared with those who have never smoked in this population. Smoking was previous implicated in a European-based study (116). They identified the

potential mechanism of the association of smoking and IgG N-glycosylation to be partially mediated by altered DNA methylation at several loci. It was also shown in the current study that there was no difference in IgG N-glycosylation when comparing those that have never smoked to ex-smokers (**Table 5.9**). This may indicate the effect of quitting smoking on improving health status, reflected in other health research (206).

It is important to note that smoking tends to be a chronic addiction, whereas alcohol consumption may fluctuate, as does keeping up with a healthy exercise regime. Therefore, taken together, the cross-sectional study design may not have been ideal for detecting associations. Certainly, the implementation of longitudinal studies looking at the long-term effects of consistent physical activity, smoking cessation and reduced alcohol consumption are required to measure the causal impact on the IgG N-glycome, and potential improvement to predicted biological age.

5.5 Epilogue

In this chapter, I attempted to explore the association of several clinical factors and health behaviours to explain IgG N-glycome heterogeneity. Indeed, it is suggested the IgG N-glycome is malleable. Though almost 60% of the heterogeneity of IgG N-glycans was explained by the differential expression of leucocyte genes (**Chapter 4**), evidencing strong genetic influence, the expression of glyco-enzymes is clearly under epigenetic regulation. Thus, several exogenous factors would be expected to contribute to the diversification of the IgG N-glycome. The factors studied in this chapter were assumed to be instrumental variables of environmental exposure since they are dynamic markers routinely measured to analyse overall health and known to be impacted by many health behaviours.

Firstly, I present an analysis of the adiposity measures to potentially improve on previously published results correlating BMI to agalactosylated IgG glycoforms (**Section 5.2**). Obesity is considered the most preventable modifiable risk factor today. I validated the previously reported weak association and evidenced that measures of central adiposity were of importance. Moreover, the univariate analyses indicated that pro-inflammatory IgG glycoforms were associated with increased central adiposity.

Secondly, I conducted an analysis on a collection of clinical measures, including the adiposity variables from the first sub-study of this chapter (**Section 5.3**). The CCA amalgamates the intrinsic correlations between both the clinical and IgG N-glycan datasets, elucidating the independent associations of single variables while controlling the effect of the other variables in the same dataset. It makes sense to construct such models from a biological standpoint since multiple factors may constantly be interacting. I conclude by extending on the central adiposity finding, demonstrating that all the cardiometabolic risk factors were significantly associated with increased pro-inflammatory IgG glycoforms. This provides evidence for the efficacy of modelling the IgG N-glycome to stratify individuals at risk of cardiometabolic disorders, which is an emerging area of research (2, 3, 39, 41, 176).

The IgG N-glycome is considered a link between the genetic makeup of cells and the cellular environment. Therefore, in theory, one may possibly change their IgG N-glycan composition through modifying health behaviours. The final sub-study of this chapter explored associations of physical activity, alcohol consumption and smoking status to the IgG N-glycome. It was evident, however, longitudinal study designs will need to be implemented to fully explore these associations and to confirm whether modifying health behaviours can cause change to IgG N-glycome composition.

Chapter 6 – General Discussion

Using a data-driven approach, this thesis presents a thorough investigation of the IgG N-glycome and the explanatory potential of several different datasets. Specifically, I attempted to explain the heterogeneity using genotype, gene expression, clinical and health behaviour data. The IgG N-glycan data was newly derived during this thesis (**Chapter 2**), as was the gene expression data (**Chapter 4**), with all other data available for my use upon application to the BHAS cohort.

There contained analyses that served the purpose of validating previous findings in our Australian sample; that is, the GWAS on the IgG N-glycan features (**Chapter 3**) and the analysis of BMI in exploring the association of adiposity to IgG N-glycosylation (**Section 5.2**). These validation studies were implemented to demonstrate homogeneity of the Australian population when compared with other Caucasian populations, at least in terms of IgG N-glycosylation. In doing so, I additionally verify the processing of these data and I am confident in other novel results presented throughout. Indeed, this is important as I incorporated new types of data and applied multivariate statistics not commonly found in the literature pertaining to IgG N-glycosylation.

This thesis proceeded to determine the degree the underlying genome and various clinical factors, assumed to be instrumental variables of environmental exposure over the life-course, contributed to the heterogeneity of the IgG N-glycome. The N-glycosylation of IgG is emerging as an innovative dynamic biomarker of health status in the post-genomic era (20). However, even the best models struggle to account for the variability of the biomarker in comparable groups. This is what I found in my MSc research project, whereby there was considerable variability in each of the case-control groups (**Section 1.1**). Therefore, I intended to identify potential cofounding factors that may improve precision of the modelled IgG N-glycans in stratifying participants of interest in future clinical research.

To address this, previous genetic findings were validated using GWAS of each IgG N-glycosylation feature (**Chapter 3**). Subsequently, the potential of differential gene expression to explain variability in the N-glycome was studied for the first time. This was accomplished using univariate models analysing differential gene expression and determining their collective function, as well as implementing a multivariate model to integrate the ‘omics datasets (**Chapter 4**). Lastly, clinical data were modelled using univariate and multivariate statistics, extending on previous research (**Chapter 5**).

Overall, factors with potential clinical importance in the regulation of the IgG N-glycome were described and these may improve the precision of this complex biomarker in stratifying populations of interest.

6.1 Recapping IgG N-Glycome Heterogeneity

IgG is an important effector glycoprotein in the immune system. It acts to exert both anti-inflammatory and pro-inflammatory responses, dependent on its affinity for several different activating or inhibitory FcR and complement factors. Thus, IgG glycoproteins mediate key effector functions, and these are largely modulated by the N-glycosylation of the recognition sequence within the Fc domain (**Figure 1.4**).

Complex-type biantennary N-glycans constitute much of the IgG N-glycome, though a high degree of heterogeneity exists (20). Moreover, two different N-glycan moieties may be within an IgG glycoform, leading to greater variability in effector functions (30). The N-glycosylation of IgG itself is complex. The N-glycans are biosynthesised sequentially along with the elongating polypeptide (**Figure 1.3**). Biosynthesis begins in the ER and progresses through to the Golgi apparatus, where glyco-enzymes alter composition of the branching N-glycan structures. The resulting N-glycome is hypothesised to be a predesigned outcome that depends on the relative expression of the genes that encode the specialised enzymes and the availability of the sugar nucleotides within the producing B cell lymphocyte (21). Thus, this gives rise to the diversification of the IgG N-glycome.

Although many glyco-enzymes have been characterised, polymorphisms in the genome alone cannot fully explain the heterogeneity, even in very large association studies. The expression of these specialised enzymes is under the influence of epigenetic regulation, and variation may be attributable to a multitude of endogenous and exogenous factors. Further, N-glycosylation is the most prevalent post-translational modification; a form of epigenetic regulation. Consequently, it is not only epigenetically regulated but an epigenetic regulator of downstream responses.

Henceforth, the IgG N-glycome is hypothesised to represent an interphenotype and may elucidate subclinical phenotypes on the spectrum from health to phenotypic disease. It is invariable over short periods, with current estimates over six months (99). Therefore, it may serve as a stable stratification biomarker for: elucidating overall subclinical health, identifying risk developing disorders, tracking progression of phenotypic disorders, and indicating who may respond to certain therapies.

6.2 Implementation of IgG N-Glycomics in the Post-Genomic-Era

The use of IgG N-glycomics to study chronic non-communicable disorders and other complex phenotypes emerged in the wake of the Human Genome Project. The consortium discovered that most phenotypes were too complex to be explained by genetics alone. Given its potential to identify and monitor health status, the IgG N-glycome has been implemented in several population-based studies in the post-genomic-era. To date, research has been published in the areas of rheumatoid arthritis (37, 38), metabolic syndrome and T2DM (18, 39), inflammatory bowel disease (18), systemic lupus erythematosus (40), hypertension (3, 41), various cancers (18, 42), and neurological disorders such as Alzheimer's disease and progressive mild cognitive impairment (43), multiple sclerosis (44), and Parkinson's disease (6). These studies implicate biological importance of N-glycans in terms of the physiological activity of IgG in several morbidities.

Currently there exists clinical tests in Europe that propose to quantify an individual's biological age using the IgG N-glycome (known as "GlycanAge"; https://glycanage.com/), with efforts expanding into China. The premise of GlycanAge is to predict biological age, which would have been influenced by several health behaviours and exposures over the life-course. Therefore, an individual can use this test, an interphenotype of overall health, to identify potential discordance from their chronological age. It is an important initiative amongst the shift towards predictive, preventive and personalised medicine. However, it is yet to be determined how this information will benefit an individual and what may be done to improve a person's biological age.

The Human Glycome Project was recently initiated (207). It aims to follow in the footsteps of the Human Genome Project and elucidate further biological regulation within the human body attributed to the N-glycome by the collaborative effort of world leaders in glycobiology from several countries. It is known that N-glycans are involved in the physiology of every major disorder; thus, the timing of the Human Glycome Project is crucial to advancing our understanding on health status (207).

6.3 Validating the Contribution of Genetics

Through the GWAS of the various IgG N-glycosylation features (**Chapter 3**), I evidenced the genetic homogeneity of this Australian population to the previously studied European Caucasian populations, at least in terms of the gQTLs associated with the IgG N-glycome. This validation study, along with validating an association between BMI and N-

glycosylation features, were of importance since other studies within this thesis included novel data and analyses. I first had to demonstrate that previous analyses could be replicated in this isolated population before proceeding, thus evidencing the integrity of the UPLC-derived data.

I identified gQTLs encoding glycosyltransferases that were previously annotated as being associated with the N-glycosylation of IgG. Indeed, even though the study was underpowered, the two identified gQTLs with genome-wide significance were previously found in four isolated European populations (18, 208). The study also evidences that gQTLs do not fully explain heterogeneity within the IgG N-glycome, thus warranting the further investigation undertaken to complete this thesis.

6.4 Epigenetic Control of IgG N-Glycosylation

Both endogenous and exogenous factors are known to alter IgG N-glycome composition. Termed the "cellular environment", it refers to the collection of factors that may influence N-glycosylation and in turn the inflammatory properties of the IgG. I showed that almost 60% of the heterogeneity in the IgG N-glycome is attributed to influence of the transcribed leucocyte genes (**Chapter 4**), themselves under epigenetic control. The O2PLS model suggested that this effect is mostly in the direction from expressed genes to the N-glycosylation of IgG, rather than the epigenetic effects of modulated effector responses to overall gene expression amongst the leucocytes. It is apparent though that there exists a complex interplay among factors in the epigenetic regulation of N-glycosylation. This underscores the importance of further multivariate analyses in larger population-based studies to elucidate unidentified mechanisms. Importantly, though this study was also underpowered, it provides data to be used to correctly power future research with these data types.

6.5 Implications on Cardiometabolic Health

An important conclusion that is drawn in this thesis is the potential for the IgG N-glycome to be assimilated into prediction and prevention models of cardiometabolic health status; an emerging area of research with N-glycome data (2, 3, 39, 41, 176). I present evidence that more unfavourable results in cardiometabolic risk factors are associated with an increased abundance of pro-inflammatory IgG glycoforms (**Section 5.2** and **Section 5.3**). Further, *ST6GAL1* had genome-wide significance in this thesis and is a gene encoding a glycotransferase controlling terminal sialylation, with pleiotropy to T2DM (**Chapter 3**), a prevalent cardiometabolic disorder. Indeed, reduced sialylation of the IgG

has been indicated among T2DM cases in Europe (39). This not only supports the further exploration into the utility of the IgG N-glycome in cardiometabolic-based research, as suggested by genetic and clinical factors, but also generates hypotheses on the activity of the IgG itself in these phenotypic disorders.

Though longitudinal studies are needed to explore causation, I provided further support for IgG playing a role in the pathophysiology of cardiometabolic disorders. My clinical studies (**Chapter 5**) also underline the need to control for an array of clinical factors in risk models, such as done recently when associating the IgG N-glycome to cardiovascular risk (176). Certainly, multiple clinical factors should be considered when building risk models integrating the IgG N-glycome, particularly those related to cardiometabolic health.

It was evident that poor health status as measured by central adiposity as well as CRP, insulin, TG, FBG, and SBP, tended to be associated with increased pro-inflammatory IgG glycoforms (**Section 5.2** and **Section 5.3**). These factors were assumed to represent instrumental variables of exposure over the life-course; that is, they are the amalgamation of unfavourable health behaviours over time, which may alter IgG N-glycosylation. If the association of these exogenous factors is causative, perhaps they can be modified to improve health status. I proposed that an adjustment in health behaviours may have a downstream impact on IgG N-glycome and be pertinent to the pathophysiology of existent subclinical and clinical phenotypes.

Three health behaviours were explored to determine whether variation among them could account for heterogeneity of the IgG N-glycome (**Section 5.4**); excessive alcohol consumption, current smoking status and physical inactivity. Importantly, these health behaviours are listed as risk factors of poor cardiometabolic health (188, 189). No difference was detected in the IgG GPs according to physical activity. However, binge drinking and excessive alcohol consumption were associated with an increased pro-inflammatory fraction of IgG glycoforms. Moreover, current smoking status was associated with variation in nine IgG GPs not reflected in ex-smokers. Overall, this sub-study suggests that intermittent changes to health behaviours may not be captured in a cross-sectional study design. Thus, it was evident that longitudinal studies are critical for exploring the causative effect of modifying these health behaviours on the IgG N-glycome. These studies are necessary in terms of prevention since the IgG N-glycans are used for modelling biological age and by proxy, subclinical health status. Thus, how an individual can modify health behaviours to improve biological age is pertinent to this translational research.

6.6 Limitations of this Thesis

This thesis was not without its limitations. Firstly, the subset of the BHAS utilised for this research were selected due to previous participation in a Busselton-based health study. Though this may lend itself to selection bias, it was necessary given they were already genotyped, significantly reducing the economic and time costs of this thesis. The population itself is in a regional city 250km south-west of Perth, Western Australia, and 99% of the participants were Caucasian with Caucasian parents. Therefore, although its homogeneity to other European Caucasian populations was suggested through the validation studies, it remains to be determined whether these results can have utility in improving precision of models integrating IgG N-glycosylation profiles in other Caucasian populations. BHAS participants were within the age range of 45 to 70 years. While this population is a target of much research, a wider adult age range may have identified important factors antecedent to age-related N-glycosylation changes. The study design was cross-sectional in nature, so no assumptions on causality could be ascertained. Rather the thesis presents a thorough investigation of associations that may occur contemporaneously to IgG N-glycome modifications. Lastly, the GWAS and perhaps the RNA-Seq analyses were certainly underpowered. The size of these subsets, however, was largely dictated by availability of data and funding resources.

6.7 Future Prospects

Integration of the multiple datasets to explain variability in the IgG N-glycome was the aim from the outset of this thesis. This research will be continued through collaborations with other research teams. Indeed, the integration of multiple 'omics datasets is an ongoing area of growth, and new models may have utility with describing these data. These newer models will be used to include all data in a single model to study joint variation. There are plans to extend on the RNA-Seq study with a much larger cohort containing IgG N-glycan and genotype data. Briefly, we may be able to impute *cis*- and *trans*-regulating gQTLs associated with altered gene expression and correlated with the genotypes and N-glycans, through implementing a transcriptome-wide association study (209). In turn, our collaborators may use our estimates to impute RNA-Seq derived associations and further explore this in their larger population consortium.

6.8 Concluding Remarks

Though this thesis attempts to further understand the contribution of genes as well as the cellular environment to the heterogeneity of IgG N-glycosylation, research exploring

causation as well as gene-environment interactions must proceed. Efforts are currently underway to use the already derived data to explore emerging novel methods of multiple-omics data integration. Other research within our team is exploring causation through *in vitro* modelling as well as through Mendelian randomisation. By amalgamating the research within this thesis with that conducted by other colleagues, our team is contributing to the efforts of the Human Glycome Project and the future of predictive, preventive and personalised medicine in our ageing population.

Page intentionally left blank

Chapter 7 – Thesis References

1. Rademacher TW, Parekh RB, Dwek RA, Isenberg D, Rook G, Axford JS, et al. The role of IgG glycoforms in the pathogenesis of rheumatoid arthritis. Springer Semin Immun. 1988;10(2-3):231-49.
2. Liu D, Zhao Z, Wang A, Ge S, Wang H, Zhang X, et al. Ischemic stroke is associated with the pro-inflammatory potential of N-glycosylated immunoglobulin G. J Neuroinflammation. 2018;15(1):123.
3. Liu JN, Dolikun M, Štambuk J, Trbojević-Akmačić I, Zhang J, Wang H, et al. The association between subclass-specific IgG Fc N-glycosylation profiles and hypertension in the Uygur, Kazak, Kirgiz, and Tajik populations. J Hum Hypertens. 2018.
4. Lu J-P, Knezevic A, Wang Y-X, Rudan I, Campbell H, Zou Z-K, et al. Screening novel biomarkers for metabolic syndrome by profiling human plasma N-glycans in Chinese Han and Croatian populations. J Proteome Res. 2011;10(11):4959-69.
5. Yu X, Wang Y, Kristic J, Dong J, Chu X, Ge S, et al. Profiling IgG N-glycans as potential biomarker of chronological and biological ages: A community-based study in a Han Chinese population. Medicine. 2016;95(28).
6. Russell A, Šimurina M, Garcia M, Novokmet M, Wang Y, Rudan I, et al. The N-glycosylation of immunoglobulin G as a novel biomarker of Parkinson's disease. Glycobiology. 2017;27(5):501-10.
7. Vanhooren V, Dewaele S, Libert C, Engelborghs S, De Deyn PP, Toussaint O, et al. Serum N-glycan profile shift during human ageing. Exp Gerontol. 2010;45(10):738-43.
8. Chen G, Wang Y, Qiu L, Qin X, Liu H, Wang X, et al. Human IgG Fc-glycosylation profiling reveals associations with age, sex, female sex hormones and thyroid cancer. J Proteomics. 2012;75(10):2824-34.
9. Krištić J, Vučković F, Menni C, Klarić L, Keser T, Beceheli I, et al. Glycans are a novel biomarker of chronological and biological ages. J Gerontol A Biol Sci Med Sci. 2014;69(7):779-89.
10. Pučić M, Knežević A, Vidič J, Adamczyk B, Novokmet M, Polašek O, et al. High throughput isolation and glycosylation analysis of IgG–variability and heritability of the IgG glycome in three isolated human populations. Mol Cell Prot. 2011;10(10):M111.010090.
11. Subedi GP, Barb AW. The structural role of antibody N-glycosylation in receptor interactions. Structure. 2015;23(9):1573-83.
12. Krause I, Wu R, Sherer Y, Patanik M, Peter J, Shoenfeld Y. In vitro antiviral and antibacterial activity of commercial intravenous immunoglobulin preparations–a potential role for adjuvant intravenous immunoglobulin therapy in infectious diseases*. Transfusion Med 2002;12(2):133-9.
13. Ioan-Facsinay A, de Kimpe SJ, Hellwig SMM, van Lent PL, Hofhuis FMA, van Ojik HH, et al. FcγRI (CD64) Contributes Substantially to Severity of Arthritis, Hypersensitivity Responses, and Protection from Bacterial Infection. Immunity. 2002;16(3):391-402.
14. Schwab I, Nimmerjahn F. Intravenous immunoglobulin therapy: how does IgG modulate the immune system? Nat Rev Immunol. 2013;13(3):176-89.
15. Fujita T. Evolution of the lectin–complement pathway and its role in innate immunity. Nat Rev Immunol. 2002;2(5):346-53.
16. Quast I, Lünemann JD. Fc glycan-modulated immunoglobulin G effector functions. J Clin Immunol. 2014;34(1):51-5.
17. Malhotra R, Wormald MR, Rudd PM, Fischer PB, Dwek RA, Sim RB. Glycosylation changes of IgG associated with rheumatoid arthritis can activate complement via the mannose-binding protein. Nat Med. 1995;1(3):237-43.

18. Lauc G, Huffman JE, Pučić M, Zgaga L, Adamczyk B, Mužinić A, et al. Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. PLoS One. 2013;9(1):e1003225.
19. Wang J, Balog CI, Stavenhagen K, Koeleman CA, Scherer HU, Selman MH, et al. Fc-glycosylation of IgG1 is modulated by B-cell stimuli. Mol Cell Prot. 2011;10(5):M110. 004655.
20. Adua E, Russell A, Roberts P, Wang Y, Song M, Wang W. Innovation Analysis on Postgenomic Biomarkers: Glycomics for Chronic Diseases. OMICS. 2017;21(4):183-96.
21. Vidarsson G, Dekkers G, Rispens T. IgG subclasses and allotypes: from structure to effector functions. Front Immunol. 2014;5.
22. Lai E, Wilson RK, Hood LE. Physical maps of the mouse and human immunoglobulin-like loci. Adv Immunol. 1989;46:1-59.
23. Pincetic A, Bournazos S, DiLillo DJ, Maamary J, Wang TT, Dahan R, et al. Type I and type II Fc receptors regulate innate and adaptive immunity. Nat Immunol. 2014;15(8):707-16.
24. Quast I, Peschke B, Lünemann JD. Regulation of antibody effector functions through IgG Fc N-glycosylation. Cell Mol Life Sci. 2017;74(5):837-47.
25. Krapp S, Mimura Y, Jefferis R, Huber R, Sondermann P. Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. J Mol Biol. 2003;325(5):979-89.
26. van de Bovenkamp FS, Hafkenscheid L, Rispens T, Rombouts Y. The emerging importance of IgG Fab glycosylation in immunity. J Immunol. 2016;196(4):1435-41.
27. Plomp R, Dekkers G, Rombouts Y, Visser R, Koeleman CAM, Kammeijer GSM, et al. Hinge-Region O-Glycosylation of Human Immunoglobulin G3 (IgG3). MCP. 2015;14(5):1373-84.
28. Arnold JN, Wormald MR, Sim RB, Rudd PM, Dwek RA. The impact of glycosylation on the biological function and structure of human immunoglobulins. Annu Rev Immunol. 2007;25:21-50.
29. Keser T, Vučković F, Barrios C, Zierer J, Wahl A, Akinkuolie AO, et al. Effects of statins on the immunoglobulin G glycome. Biochim Biophys Acta Gen Subj. 2017;1861(5):1152-8.
30. Nimmerjahn F, Anthony RM, Ravetch JV. Agalactosylated IgG antibodies depend on cellular Fc receptors for in vivo activity. PNAS. 2007;104(20):8433-7.
31. Lu J, Sun PD. Structural mechanism of high affinity FcγRI recognition of immunoglobulin G. Immunol Rev. 2015;268(1):192-200.
32. Chen TF, Sazinsky SL, Houde D, DiLillo DJ, Bird J, Li KK, et al. Engineering Aglycosylated IgG Variants with Wild-Type or Improved Binding Affinity to Human Fc Gamma RIIA and Fc Gamma RIIIAs. Journal of Molecular Biology. 2017;429(16):2528-41.
33. Tomiya N, Narang S, Lee YC, Betenbaugh MJ. Comparing N-glycan processing in mammalian cell lines to native and engineered lepidopteran insect cell lines. Glycoconjugate J. 2004;21(6):343-60.
34. Trombetta ES. The contribution of N-glycans and their processing in the endoplasmic reticulum to glycoprotein biosynthesis. Glycobiology. 2003;13(9):77R-91R.
35. Subedi GP, Barb AW, editors. The immunoglobulin G1 N-glycan composition affects binding to each low affinity Fc γ receptor. mAbs; 2016: Taylor & Francis.
36. Barb AW. Intramolecular N-Glycan/Polypeptide Interactions Observed at Multiple N-Glycan Remodeling Steps through [13C,15N]-N-Acetylglucosamine Labeling of Immunoglobulin G1. Biochemistry. 2015;54(2):313-22.
37. Troelsen LN, Jacobsen S, Abrahams JL, Royle L, Rudd PM, Narvestad E, et al. IgG glycosylation changes and MBL2 polymorphisms: associations with markers of

systemic inflammation and joint destruction in rheumatoid arthritis. J Rheum. 2012;39(3):463-9.
38. Sebastian A, Alzain MA, Asweto CO, Song H, Cui L, Yu X, et al. Glycan biomarkers for rheumatoid arthritis and its remission status in han Chinese patients. OMICS. 2016;20(6):343-51.
39. Lemmers RF, Vilaj M, Urda D, Agakov F, Šimurina M, Klaric L, et al. IgG glycan patterns are associated with type 2 diabetes in independent European populations. Biochim Biophys Acta Gen Subj. 2017;1861(9):2240-9.
40. Vučković F, Krištić J, Gudelj I, Teruel M, Keser T, Pezer M, et al. Association of Systemic Lupus Erythematosus With Decreased Immunosuppressive Potential of the IgG Glycome. Arthritis Rheumatol. 2015;67(11):2978-89.
41. Wang Y, Klaric L, Yu X, Thaqi K, Dong J, Novokmet M, et al. The Association Between Glycosylation of Immunoglobulin G and Hypertension: A Multiple Ethnic Cross-Sectional Study. Medicine. 2016;95(17):e3379.
42. Meany DL, Chan DW. Aberrant glycosylation associated with enzymes as cancer biomarkers. Clin Proteomics. 2011;8(7):10.1186.
43. Lundström SL, Yang H, Lyutvinskiy Y, Rutishauser D, Herukka S-K, Soininen H, et al. Blood plasma IgG Fc glycans are significantly altered in Alzheimer's disease and progressive mild cognitive impairment. J Alzheimers Dis. 2013;38(3):567-79.
44. Wuhrer M, Selman MH, McDonnell LA, Kümpfel T, Derfuss T, Khademi M, et al. Pro-inflammatory pattern of IgG1 Fc glycosylation in multiple sclerosis cerebrospinal fluid. J Neuroinflamm. 2015;12(1):235.
45. Nimmerjahn F, Ravetch JV. Divergent immunoglobulin g subclass activity through selective Fc receptor binding. Science. 2005;310(5753):1510-2.
46. Quast I, Keller CW, Maurer MA, Giddens JP, Tackenberg B, Wang L-X, et al. Sialylation of IgG Fc domain impairs complement-dependent cytotoxicity. J Clin Invest. 2015;125(11):4160-70.
47. Jones MB, Oswald DM, Joshi S, Whiteheart SW, Orlando R, Cobb BA. B-cell–independent sialylation of IgG. PNAS. 2016:201523968.
48. Harre U, Lang SC, Pfeifle R, Rombouts Y, Frühbeißer S, Amara K, et al. Glycosylation of immunoglobulin G determines osteoclast differentiation and bone loss. Nat Commun. 2015;6.
49. Shields RL, Lai J, Keck R, O'Connell LY, Hong K, Meng YG, et al. Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human FcγRIII and antibody-dependent cellular toxicity. J Biol Chem. 2002;277(30):26733-40.
50. Dekkers G, Treffers L, Plomp R, Bentlage AEH, de Boer M, Koeleman CAM, et al. Decoding the Human Immunoglobulin G-Glycan Repertoire Reveals a Spectrum of Fc-Receptor- and Complement-Mediated-Effector Activities. Frontiers in Immunology. 2017;8(877).
51. Ferrara C, Grau S, Jäger C, Sondermann P, Brünker P, Waldhauer I, et al. Unique carbohydrate–carbohydrate interactions are required for high affinity binding between FcγRIII and antibodies lacking core fucose. PNAS. 2011;108(31):12669-74.
52. Zou G, Ochiai H, Huang W, Yang Q, Li C, Wang L-X. Chemoenzymatic synthesis and Fcγ receptor binding of homogeneous glycoforms of antibody Fc domain. Presence of a bisecting sugar moiety enhances the affinity of Fc to FcγIIIa receptor. Journal of the American Chemical Society. 2011;133(46):18975-91.
53. Maupin KA, Liden D, Haab BB. The fine specificity of mannose-binding and galactose-binding lectins revealed using outlier motif analysis of glycan array data. Glycobiology. 2012;22(1):160-9.
54. Ruhaak LR, Uh H-W, Beekman M, Koeleman C, Hokke CH, Westendorp R, et al. Decreased levels of bisecting GlcNAc glycoforms of IgG are associated with human longevity. PLoS One. 2010;5(9):e12566.
55. Peschke B, Keller CW, Weber P, Quast I, Lunemann JD. Fc-galactosylation of human IgG isotypes improves C1q binding and enhances complement-dependent cytotoxicity. Front Immunol. 2017;8:646.

56. Karsten CM, Köhl J. The immunoglobulin, IgG Fc receptor and complement triangle in autoimmune diseases. Immunobiol. 2012;217(11):1067-79.
57. Ravetch JV, Bolland S. Igg fc receptors. Annu Rev Immunol. 2001;19(1):275-90.
58. Sondermann P, Pincetic A, Maamary J, Lammens K, Ravetch JV. General mechanism for modulating immunoglobulin effector function. PNAS. 2013;110(24):9868-72.
59. Akira S, Hirano T, Taga T, Kishimoto T. Biology of multifunctional cytokines: IL 6 and related molecules (IL 1 and TNF). The FASEB journal. 1990;4(11):2860-7.
60. Daëron M. Fc receptor biology. Annu Rev Immunol. 1997;15(1):203-34.
61. Fanger NA, Voigtlaender D, Liu C, Swink S, Wardwell K, Fisher J, et al. Characterization of expression, cytokine regulation, and effector function of the high affinity IgG receptor Fc gamma RI (CD64) expressed on human blood dendritic cells. J Immunol. 1997;158(7):3090-8.
62. Rogers KA, Scinicariello F, Attanasio R. IgG Fc receptor III homologues in nonhuman primate species: genetic characterization and ligand interactions. J Immunol. 2006;177(6):3848.
63. Van de Winkel J, Anderson CL. Biology of human immunoglobulin G Fc receptors. J Leukoc Biol. 1991;49(5):511-24.
64. Woolhiser MR, Okayama Y, Gilfillan AM, Metcalfe DD. IgG-dependent activation of human mast cells following up-regulation of FcγRI by IFN-γ. Eur J Immunol. 2001;31(11):3298-307.
65. Unkeless JC, Jin J. Inhibitory receptors, ITIM sequences and phosphatases. Curr Opin Immunol. 1997;9(3):338-43.
66. Vivier E, Daëron M. Immunoreceptor tyrosine-based inhibition motifs. Immunol Today. 1997;18(6):286-91.
67. Bryceson YT, Chiang SC, Darmanin S, Fauriat C, Schlums H, Theorell J, et al. Molecular mechanisms of natural killer cell activation. J Innate Immun. 2011;3(3):216-26.
68. Fernandes MJ, Lachance G, Paré G, Rollet-Labelle E, Naccache PH. Signaling through CD16b in human neutrophils involves the Tec family of tyrosine kinases. J Leukoc Biol. 2005;78(2):524-32.
69. Hořejší V, Drbal K, Cebecauer M, Černý J, Brdička T, Angelisová P, et al. GPI-microdomains: a role in signalling via immunoreceptors. Immunol Today. 1999;20(8):356-61.
70. Wirthmueller U, Kurosaki T, Murakami M, Ravetch J. Signal transduction by Fc gamma RIII (CD16) is mediated through the gamma chain. J Exp Med. 1992;175(5):1381-90.
71. Anthony RM, Nimmerjahn F. The role of differential IgG glycosylation in the interaction of antibodies with FcgammaRs in vivo. Current opinion in organ transplantation. 2011;16(1):7-14.
72. van Kooyk Y, Geijtenbeek TB. DC-SIGN: escape mechanism for pathogens. Nat Rev Immunol. 2003;3(9):697-709.
73. van Kooyk Y, Rabinovich GA. Protein-glycan interactions in the control of innate and adaptive immune responses. Nat Immunol. 2008;9(6):593-601.
74. Cooney DS, Phee H, Jacob A, Coggeshall KM. Signal transduction by human-restricted FcγRIIa involves three distinct cytoplasmic kinase families leading to phagocytosis. J Immunol. 2001;167(2):844-54.
75. Brauweiler A, Tamir I, Marschner S, Helgason CD, Cambier JC. Partially distinct molecular mechanisms mediate inhibitory FcγRIIB signaling in resting and activated B cells. J Immunol. 2001;167(1):204-11.
76. Lin J-X, Migone T-S, Tseng M, Friedmann M, Weatherbee JA, Zhou L, et al. The role of shared receptor motifs and common Stat proteins in the generation of cytokine pleiotropy and redundancy by IL-2, IL-4, IL-7, IL-13, and IL-15. Immunity. 1995;2(4):331-9.

77. Marshall JS. Mast-cell responses to pathogens. Nat Rev Immunol. 2004;4(10):787-99.
78. Aderem A, Underhill DM. Mechanisms of phagocytosis in macrophages. Annu Rev Immunol. 1999;17(1):593-623.
79. Caron E, Hall A. Identification of two distinct mechanisms of phagocytosis controlled by different Rho GTPases. Science. 1998;282(5394):1717-21.
80. Forni GL, Pinto V, Musso M, Mori M, Girelli D, Caldarelli I, et al. Transferrin-immune complex disease: A potentially overlooked gammopathy mediated by IgM and IgG. Am J Hematol. 2013;88(12):1045-9.
81. Mantovani B. Different roles of IgG and complement receptors in phagocytosis by polymorphonuclear leukocytes. J Immunol. 1975;115(1):15-7.
82. Russell DG, VanderVen BC, Glennie S, Mwandumba H, Heyderman RS. The macrophage marches on its phagosome: dynamic assays of phagosome function. Nat Rev Immunol. 2009;9(8):594-600.
83. Broussas M, Broyer L, Goetsch L. Evaluation of antibody-dependent cell cytotoxicity using lactate dehydrogenase (LDH) measurement. Glycosylation Engineering of Biopharmaceuticals: Springer; 2013. p. 305-17.
84. Böhm S, Schwab I, Lux A, Nimmerjahn F, editors. The role of sialic acid as a modulator of the anti-inflammatory activity of IgG. Semin Immunopathol

2012: Springer.
85. Scallon BJ, Tam SH, McCarthy SG, Cai AN, Raju TS. Higher levels of sialylated Fc glycans in immunoglobulin G molecules can adversely impact functionality. Mol Immunol. 2007;44(7):1524-34.
86. de Cordoba SR, Tortajada A, Harris CL, Morgan BP. Complement dysregulation and disease: from genes and proteins to diagnostics and drugs. Immunobiol. 2012;217(11):1034-46.
87. Menni C, Keser T, Mangino M, Bell JT, Erte I, Akmačić I, et al. Glycosylation of immunoglobulin g: role of genetic and epigenetic influences. PLoS One. 2013;8(12):e82558.
88. Saldova R, Huffman JE, Adamczyk B, Mužinić A, Kattla JJ, Pučić M, et al. Association of medication with the human plasma N-glycome. J Proteome Res. 2012;11(3):1821-31.
89. Knežević A, Gornik O, Polašek O, Pučić M, Redžić I, Novokmet M, et al. Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans. Glycobiology. 2010:cwq051.
90. Gudelj I, Keser T, Vučković F, Škaro V, Goreta SŠ, Pavić T, et al. Estimation of human age using N-glycan profiles from bloodstains. Int J Legal Med. 2015;129(5):955-61.
91. Catera M, Borelli V, Malagolini N, Chiricolo M, Venturi G, Reis CA, et al. Identification of novel plasma glycosylation-associated markers of aging. Oncotarget. 2016;7(7):7455-68.
92. Jurić J, Peng H, Song M, Šimunović J, Trbojević-Akmačić I, Hanić M, et al., editors. Immunoglobulin G Glycosylation in Menstrual Cycle 12th Jenner Glycobiology and Medicine Symposium on Translational Glycobiology: From Bench to Bedside; 2017.
93. Ercan A, Kohrt WM, Cui J, Deane KD, Pezer M, Yu EW, et al. Estrogens regulate glycosylation of IgG in women and men. JCI Insight. 2017;2(4).
94. Bondt A, Rombouts Y, Selman MHJ, Hensbergen PJ, Reiding KR, Hazes JMW, et al. Immunoglobulin G (IgG) Fab Glycosylation Analysis Using a New Mass Spectrometric High-throughput Profiling Method Reveals Pregnancy-associated Changes. Mol Cell Prot. 2014;13(11):3029-39.
95. Liu B, Spearman M, Doering J, Lattová E, Perreault H, Butler M. The availability of glucose to CHO cells affects the intracellular lipid-linked oligosaccharide distribution,

site occupancy and the N-glycosylation profile of a monoclonal antibody. J Biotech. 2014;170:17-27.
96. Fan Y, Jimenez Del Val I, Müller C, Wagtberg Sen J, Rasmussen SK, Kontoravdi C, et al. Amino acid and glucose metabolism in fed-batch CHO cell culture affects antibody production and glycosylation. Biotechnol Bioeng. 2015;112(3):521-35.
97. Perkovic MN, Bakovic MP, Kristic J, Novokmet M, Huffman JE, Vitart V, et al. The association between galactosylation of immunoglobulin G and body mass index. Prog Neuro-Psychoph. 2014;48:20-5.
98. Russell A, Kepka A, Trbojević-Akmačić I, Hui J, Hunter M, Ugrina I, et al., editors. Higher levels of abdominal body fat are associated with an increase in pro-inflammatory immunoglobulin G N-glycans: results from the Busselton Healthy Ageing Study. 12th Jenner Glycobiology and Medicine Symposium on Translational Glycobiology: From Bench to Bedside; 2017 May 6-9, 2017; Dubrovnik, Croatia.
99. Adua E, Anto EO, Roberts P, Kantanka OS, Aboagye E, Wang W. The potential of N-glycosylation profiles as biomarkers for monitoring the progression of Type II diabetes mellitus towards diabetic kidney disease. J Diabetes Metab Disord. 2018;17(2):233-46.
100. James A, Hunter M, Straker L, Beilby J, Bucks R, Davis T, et al. Rationale, design and methods for a community-based study of clustering and cumulative effects of chronic disease processes and their effects on ageing: the Busselton healthy ageing study. BMC Public Health. 2013;13(1):936.
101. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics. 2010;11(10):733-9.
102. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.
103. Gudelj I, Lauc G, Pezer M. Immunoglobulin G glycosylation in aging and diseases. Cellular Immunology. 2018.
104. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1H NMR Metabonomics. Analytical Chemistry. 2006;78(13):4281-90.
105. Shen X, Klarić L, Sharapov S, Mangino M, Ning Z, Wu D, et al. Multivariate discovery and replication of five novel loci associated with Immunoglobulin G N-glycosylation. Nature Communications. 2017;8(1):447.
106. Kohl SM, Klein MS, Hochrein J, Oefner PJ, Spang R, Gronwald W. State-of-the art data normalization methods improve NMR-based metabolomic analysis. Metabolomics. 2012;8(Suppl 1):146-60.
107. Collins FS, McKusick VA. Implications of the Human Genome Project for Medical Science. JAMA. 2001;285(5):540-4.
108. Russell A, Adua E, Ugrina I, Laws S, Wang W. Unravelling Immunoglobulin G Fc N-Glycosylation: A Dynamic Marker Potentiating Predictive, Preventive and Personalised Medicine. Int J Mol Sci. 2018;19(2):390.
109. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nature Reviews Genetics. 2010;11(7):499-511.
110. Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. Journal of the National Cancer Institute. 2000;92(14):1151-8.
111. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 2007;23(10):1294-6.
112. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. BMC bioinformatics. 2010;11:134.
113. Arnold M, Raffler J, Pfeufer A, Suhre K, Kastenmüller G. SNiPA: an interactive, genetic variant-centered annotation browser. Bioinformatics. 2014:btu779.

114. Jima DD, Zhang J, Jacobs C, Richards KL, Dunphy CH, Choi WW, et al. Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs. Blood. 2010;116(23):e118-27.
115. Kuhn B, Benz J, Greif M, Engel AM, Sobek H, Rudolph MG. The structure of human alpha-2,6-sialyltransferase reveals the binding mode of complex glycans. Acta crystallographica Section D, Biological crystallography. 2013;69(Pt 9):1826-38.
116. Wahl A, van den Akker E, Klaric L, Štambuk J, Benedetti E, Plomp R, et al. Genome-Wide Association Study on Immunoglobulin G Glycosylation Patterns. Front Immunol. 2018;9(277).
117. Kohler RS, Anugraham M, López MN, Xiao C, Schoetzau A, Hettich T, et al. Epigenetic activation of MGAT3 and corresponding bisecting GlcNAc shortens the survival of cancer patients. Oncotarget. 2016;7(32):51674-86.
118. Fiala M, Mahanian M, Rosenthal M, Mizwicki MT, Tse E, Cho T, et al. MGAT3 mRNA: a biomarker for prognosis and therapy of Alzheimer's disease by vitamin D and curcuminoids. J Alzheimers Dis. 2011;25(1):135-44.
119. Klasic M, Markulin D, Vojta A, Samarzija I, Birus I, Dobrinic P, et al. Promoter methylation of the MGAT3 and BACH2 genes correlates with the composition of the immunoglobulin G glycome in inflammatory bowel disease. Clin Epigenetics. 2018;10:75.
120. Verma R, Kubendran S, Das SK, Jain S, Brahmachari SK. SYNGR1 is associated with schizophrenia and bipolar disorder in southern India. Journal of human genetics. 2005;50(12):635-40.
121. Iatropoulos P, Gardella R, Valsecchi P, Magri C, Ratti C, Podavini D, et al. Association study and mutational screening of SYNGR1 as a candidate susceptibility gene for schizophrenia. Psychiatric genetics. 2009;19(5):237-43.
122. Gong H, Fang L, Li Y, Du J, Zhou B, Wang X, et al. miR873 inhibits colorectal cancer cell proliferation by targeting TRAF5 and TAB1. Oncology reports. 2018;39(3):1090-8.
123. Bouwmeester T, Bauch A, Ruffner H, Angrand P-O, Bergamini G, Croughton K, et al. A physical and functional map of the human TNF-α/NF-κB signal transduction pathway. Nature Cell Biology. 2004;6(2):97-105.
124. Szabo SJ, Kim ST, Costa GL, Zhang X, Fathman CG, Glimcher LH. A Novel Transcription Factor, T-bet, Directs Th1 Lineage Commitment. Cell. 2000;100(6):655-69.
125. Zhang L, Cao Y, Kou X, Che L, Zhou X, Chen G, et al. Long non-coding RNA HCG11 suppresses the growth of glioma by cooperating with the miR-4425/MTA3 axis. The journal of gene medicine. 2019;21(4):e3074.
126. Arndt S, Poser I, Schubert T, Moser M, Bosserhoff A-K. Cloning and functional characterization of a new Ski homolog, Fussel-18, specifically expressed in neuronal tissues. Laboratory Investigation. 2005;85(11):1330-41.
127. Rodriguez-Jato S, Nicholls RD, Driscoll DJ, Yang TP. Characterization of cis - and trans -acting elements in the imprinted human SNURF-SNRPN locus. Nucleic Acids Research. 2005;33(15):4740-53.
128. Lehto M, Mayranpaa MI, Pellinen T, Ihalmo P, Lehtonen S, Kovanen PT, et al. The R-Ras interaction partner ORP3 regulates cell adhesion. Journal of cell science. 2008;121(Pt 5):695-705.
129. Olkkonen VM, Li S. Oxysterol-binding proteins: sterol and phosphoinositide sensors coordinating transport, signaling and metabolism. Progress in lipid research. 2013;52(4):529-38.
130. Irie A, Koyama S, Kozutsumi Y, Kawasaki T, Suzuki A. The Molecular Basis for the Absence ofN-Glycolylneuraminic Acid in Humans. 1998;273(25):15866-71.
131. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics. 2009;10(1):57-63.
132. Zhao S, Fung-Leung W-P, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One. 2014;9(1).

133. Ahmed AA, Giddens J, Pincetic A, Lomino JV, Ravetch JV, Wang L-X, et al. Structural characterization of anti-inflammatory immunoglobulin G Fc proteins. J Mol Biol. 2014;426(18):3166-79.
134. Horvat T, Zoldoš V, Lauc G. Evolutional and clinical implications of the epigenetic regulation of protein glycosylation. Clin Epigenetics. 2011;2(2):425-32.
135. Gornik O, Wagner J, Pučić M, Knežević A, Redžić I, Lauc G. Stability of N-glycan profiles in human plasma. Glycobiology. 2009;19(12):1547-53.
136. Johnson JL, Jones MB, Ryan SO, Cobb BAJTii. The regulatory power of glycans and their binding partners in immunity. Trends Immunol. 2013;34(6):290-8.
137. Rabinovich Gabriel A, Croci Diego O. Regulatory Circuits Mediated by Lectin-Glycan Interactions in Autoimmunity and Cancer. Immunity. 2012;36(3):322-35.
138. Trbojevic-Akmacic I, Ugrina I, Lauc G. Comparative Analysis and Validation of Different Steps in Glycomics Studies. Methods in enzymology. 2017;586:37-55.
139. el Bouhaddani S, Houwing-Duistermaat J, Salo P, Perola M, Jongbloed G, Uh H-W. Evaluation of O2PLS in Omics data integration. BMC bioinformatics. 2016;17:S11.
140. el Bouhaddani S, Uh H-W, Jongbloed G, Hayward C, Klarić L, Kiełbasa SM, et al. Integrating omics datasets with the OmicsPLS package. BMC bioinformatics. 2018;19(1):371.
141. Bylesjö M, Eriksson D, Kusano M, Moritz T, Trygg J. Data integration in plant biology: the O2PLS method for combined modeling of transcript and metabolite data. Plant J. 2007;52(6):1181-91.
142. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Methodol. 1995:289-300.
143. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC bioinformatics. 2013;14(1):128.
144. Lefranc M-P. Nomenclature of the human immunoglobulin lambda (IGL) genes. Exp Clin Immunogenet. 2001;18(4):242-54.
145. Lefranc M-P. Nomenclature of the human immunoglobulin kappa (IGK) genes. Exp Clin Immunogenet. 2001;18(3):161-74.
146. Lefranc M-P. Nomenclature of the human immunoglobulin heavy (IGH) genes. Exp Clin Immunogenet. 2001;18(2):100-16.
147. Schroeder HW, Jr., Cavacini L. Structure and function of immunoglobulins. J Allergy Clin Immunol. 2010;125(2 Suppl 2):S41-S52.
148. Visser A, Hamza N, Kroese FG, Bos NA. Acquiring new N-glycosylation sites in variable regions of immunoglobulin genes by somatic hypermutation is a common feature of autoimmune diseases. Ann Rheum Dis. 2018;77(10):e69-e.
149. Keser T, Gornik I, Vučković F, Selak N, Pavić T, Lukić E, et al. Increased plasma N-glycome complexity is associated with higher risk of type 2 diabetes. Diabetologia. 2017;60(12):2352-60.
150. van de Bovenkamp FS, Derksen NIL, Ooijevaar-de Heer P, van Schie KA, Kruithof S, Berkowska MA, et al. Adaptive antibody diversification through N-linked glycosylation of the immunoglobulin variable region. PNAS. 2018;115(8):1901.
151. Rosenbaum M, Andreani V, Kapoor T, Herp S, Flach H, Duchniewicz M, et al. MZB1 is a GRP94 cochaperone that enables proper immunoglobulin heavy chain biosynthesis upon ER stress. Gene Dev. 2014;28(11):1165-78.
152. Sanderson RD, Lalor P, Bernfield M. B lymphocytes express and lose syndecan at specific stages of differentiation. Cell regulation. 1989;1(1):27-35.
153. LeBien TW, Tedder TF. B lymphocytes: how they develop and function. Blood. 2008;112(5):1570-80.
154. Baković MP, Selman MHJ, Hoffmann M, Rudan I, Campbell H, Deelder AM, et al. High-Throughput IgG Fc N-Glycosylation Profiling by Mass Spectrometry of Glycopeptides. J Proteome Res. 2013;12(2):821-31.

155. Festa A, D'Agostino Jr R, Williams K, Karter A, Mayer-Davis E, Tracy R, et al. The relation of body fat mass and distribution to markers of chronic inflammation. Int J Obes Relat Metab Disord. 2001;25(10).
156. Panagiotakos DB, Pitsavos C, Yannakoulia M, Chrysohoou C, Stefanadis C. The implication of obesity and central fat on markers of chronic inflammation: The ATTICA study. Atherosclerosis. 2005;183(2):308-15.
157. Crowson CS, Matteson EL, Davis JM, Gabriel SE. Contribution of obesity to the rise in incidence of rheumatoid arthritis. Arthritis Care Res. 2013;65(1):71-7.
158. Tedeschi SK, Barbhaiya M, Malspeis S, Lu B, Sparks JA, Karlson EW, et al. Obesity and the risk of systemic lupus erythematosus among women in the Nurses' Health Studies. Semin Arthritis Rheum. 2017;47(3):376-83.
159. Geiss LS, Wang J, Cheng YJ, Thompson TJ, Barker L, Li Y, et al. Prevalence and incidence trends for diagnosed diabetes among adults aged 20 to 79 years, United States, 1980-2012. JAMA. 2014;312(12):1218-26.
160. Gupta N, Goel K, Shah P, Misra A. Childhood obesity in developing countries: epidemiology, determinants, and prevention. Endocr Rev. 2012;33(1):48-70.
161. Alissa EM, Maisa'a M, Alama NA, Ferns GA. Role of omentin-1 and C-reactive protein in obese subjects with subclinical inflammation. Journal of clinical & translational endocrinology. 2016;3:7-11.
162. Fontana L, Eagon JC, Trujillo ME, Scherer PE, Klein S. Visceral fat adipokine secretion is associated with systemic inflammation in obese humans. Diabetes. 2007;56(4):1010-3.
163. Koster A, Stenholm S, Alley DE, Kim LJ, Simonsick EM, Kanaya AM, et al. Body fat distribution and inflammation among obese older adults with and without metabolic syndrome. Obesity. 2010;18(12):2354-61.
164. Gaens KH, Ferreira I, Van De Waarenburg MP, van Greevenbroek MM, Van Der Kallen CJ, Dekker JM, et al. Protein-Bound Plasma Nε-(Carboxymethyl) lysine Is Inversely Associated With Central Obesity and Inflammation and Significantly Explain a Part of the Central Obesity–Related Increase in Inflammation: The Hoorn and CODAM Studies. Arteriosclerosis, thrombosis, and vascular biology. 2015;35(12):2707-13.
165. Maratha A, Stockmann H, Coss KP, Estela Rubio-Gozalbo M, Knerr I, Fitzgibbon M, et al. Classical galactosaemia: novel insights in IgG N-glycosylation and N-glycan biosynthesis. European Journal of Human Genetics. 2016;24(7):976-84.
166. Lauc G, Pezer M, Rudan I, Campbell H. Mechanisms of disease: The human N-glycome. Biochimica et Biophysica Acta (BBA) - General Subjects. 2016;1860(8):1574-82.
167. Krištić J, Zoldoš V, Lauc G, Seeberger PH, Hart GW, Wong CH, et al. Complex Genetics of Protein N-Glycosylation. Glycoscience: Biology and Medicine. 2014:1-7.
168. Engdahl C, Raufer J, Harre U, Bondt A, Pfeifle R, Krönke G, et al. SAT0019 Estrogen influences the sialylation profile and inflammatory properties of antibodies – a potential explanation for the sex differences and increased risk for ra in postmenopausal women. Ann Rheum Dis. 2017;76(Suppl 2):775-.
169. Dekkers G, Rispens T, Vidarsson G. Novel Concepts of Altered Immunoglobulin G Galactosylation in Autoimmune Diseases. Frontiers in Immunology. 2018;9(553).
170. Steiger JH. Tests for comparing elements of a correlation matrix. Psychol Bull. 1980;87(2):245.
171. Chen G, Wang Y, Qin X, Li H, Guo Y, Wang Y, et al. Change in IgG1 Fc N-linked glycosylation in human lung cancer: Age-and sex-related diagnostic potential. Electrophoresis. 2013;34(16):2407-16.
172. Shen S, Lu Y, Qi H, Li F, Shen Z, Wu L, et al. Waist-to-height ratio is an effective indicator for comprehensive cardiovascular health. Sci Rep. 2017;7.
173. Whitmer R, Gustafson D, Barrett-Connor E, Haan M, Gunderson E, Yaffe K. Central obesity and increased risk of dementia more than three decades later. Neurology. 2008;71(14):1057-64.

174. Hsieh C-J, Wang P-W, Chen T-Y. The relationship between regional abdominal fat distribution and both insulin resistance and subclinical chronic inflammation in non-diabetic adults. Diabetology & Metabolic Syndrome. 2014;6(1):49.
175. Wang Y, Adua E, Russell A, Roberts P, Ge S, Zeng Q, et al. Glycomics and its application potential in precision medicine. Science. 2016;354(6319):36-9.
176. Menni C, Gudelj I, MacDonald-Dunlop E, Mangino M, Zierer J, Bešić E, et al. Glycosylation Profile of Immunoglobulin G Is Cross-Sectionally Associated with Cardiovascular Disease Risk Score and Subclinical Atherosclerosis in Two Independent Cohorts. Circulation research. 2018:CIRCRESAHA. 117.312174.
177. Plomp R, Ruhaak LR, Uh H-W, Reiding KR, Selman M, Houwing-Duistermaat JJ, et al. Subclass-specific IgG glycosylation is associated with markers of inflammation and metabolic health. 2017;7(1):12325.
178. Russell AC, Kepka A, Trbojević-Akmačić I, Ugrina I, Song M, Hui J, et al. Increased central adiposity is associated with pro-inflammatory immunoglobulin G N-glycans. Immunobiology. 2018.
179. Ge S, Wang Y, Song M, Li X, Yu X, Wang H, et al. Type 2 diabetes mellitus: Integrative analysis of multiomics data for biomarker discovery. Omics: a journal of integrative biology. 2018;22(7):514-23.
180. Knuiman MW, Hung J, Divitini ML, Davis TM, Beilby JP. Utility of the metabolic syndrome and its components in the prediction of incident cardiovascular disease: a prospective cohort study. 2009;16(2):235-41.
181. Segman RH, Stein MB. C-reactive protein: a stress diathesis marker at the crossroads of maladaptive behavioral and cardiometabolic sequelae. Am Psychiatric Assoc; 2015.
182. Bondt A, Selman MHJ, Deelder AM, Hazes JMW, Willemsen SP, Wuhrer M, et al. Association between Galactosylation of Immunoglobulin G and Improvement of Rheumatoid Arthritis during Pregnancy Is Independent of Sialylation. J Proteome Res. 2013;12(10):4522-31.
183. Novokmet M, Lukić E, Vučković F, –Durić Ž, Keser T, Rajšl K, et al. Changes in IgG and total plasma protein glycomes in acute systemic inflammation. Scientific Reports. 2014;4:4347.
184. Díaz-Redondo A, Giráldez-García C, Carrillo L, Serrano R, García-Soidán FJ, Artola S, et al. Modifiable risk factors associated with prediabetes in men and women: a cross-sectional analysis of the cohort study in primary health care on the evolution of patients with prediabetes (PREDAPS-Study). BMC Family Practice. 2015;16(1):5.
185. Byrne DW, Rolando LA, Aliyu MH, McGown PW, Connor LR, Awalt BM, et al. Modifiable Healthy Lifestyle Behaviors: 10-Year Health Outcomes From a Health Promotion Program. American Journal of Preventive Medicine. 2016;51(6):1027-37.
186. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Medicine & science in sports & exercise. 2003;35(8):1381-95.
187. Clemens S, Matthews S. Comparison of a food-frequency questionnaire method and a quantity-frequency method to classify risky alcohol consumption in women2008. 223-9 p.
188. Coomber K, Jones SC, Martino F, Miller PG. Predictors of awareness of standard drink labelling and drinking guidelines to reduce negative health effects among Australian drinkers. Drug and alcohol review. 2017;36(2):200-9.
189. Callinan S, Livingston M, Room R, Dietze P. Drinking contexts and alcohol consumption: how much alcohol is consumed in different Australian locations? Journal of studies on alcohol and drugs. 2016;77(4):612-9.
190. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society Series B (Methodological). 1995:289-300.

191. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2016 [Available from: http://www.R-project.org.
192. Shield KD, Gmel G, Gmel G, Mäkelä P, Probst C, Room R, et al. Life-time risk of mortality due to different levels of alcohol consumption in seven European countries: implications for low-risk drinking guidelines. Addiction. 2017;112(9):1535-44.
193. Verdurmen J, Monshouwer K, Dorsselaer Sv, Bogt Tt, Vollebergh W. Alcohol use and mental health in adolescents: interactions with age and gender-findings from the Dutch 2001 Health Behaviour in School-Aged Children survey. Journal of Studies on Alcohol. 2005;66(5):605-9.
194. Nady el-Guebaly, M.D. ,, Janice Cathcart, B.S.N., M.Ed. ,, Shawn Currie, Ph.D. ,, Diane Brown, R.N. , and, Susan Gloster, R.N., B.N. Public Health and Therapeutic Aspects of Smoking Bans in Mental Health and Addiction Settings. Psychiatric Services. 2002;53(12):1617-22.
195. Teesson M, Hall W, Slade T, Mills K, Grove R, Mewton L, et al. Prevalence and correlates of DSM-IV alcohol abuse and dependence in Australia: findings of the 2007 National Survey of Mental Health and Wellbeing. Addiction. 2010;105(12):2085-94.
196. Floegel TA, Perez GA. An integrative review of physical activity/exercise intervention effects on function and health-related quality of life in older adults with heart failure. Geriatric Nursing. 2016;37(5):340-7.
197. Cairney J, Faught BE, Hay J, Wade TJ, Corna LM. Physical activity and depressive symptoms in older adults. Journal of Physical Activity and Health. 2005;2(1):98-114.
198. Paterson DH, Jones GR, Rice CL. Ageing and physical activity: evidence to develop exercise recommendations for older adults. Applied physiology, nutrition, and metabolism. 2007;32(S2E):S69-S108.
199. Cormie P, Zopf EM, editors. Exercise medicine for the management of androgen deprivation therapy-related side effects in prostate cancer. Urologic Oncology: Seminars and Original Investigations; 2018: Elsevier.
200. Spry NA, Kennedy MA, Singh F, Galvão D, Taaffe D, Chee R, et al., editors. Exercise medicine is the new radiotherapy adjunct–when co-located and timetabled with treatment. Radiotherapy and Oncology; 2018: ELSEVIER IRELAND LTD ELSEVIER HOUSE, BROOKVALE PLAZA, EAST PARK SHANNON, CO ….
201. Sperandei S, Vieira MC, Reis AC. Adherence to physical activity in an unsupervised setting: Explanatory variables for high attrition rates among fitness center members. Journal of Science and Medicine in Sport. 2016;19(11):916-20.
202. Marengoni A, Rizzuto D, Fratiglioni L, Antikainen R, Laatikainen T, Lehtisalo J, et al. The effect of a 2-year intervention consisting of diet, physical exercise, cognitive training, and monitoring of vascular risk on chronic morbidity—the FINGER randomized controlled trial. Journal of the American Medical Directors Association. 2018;19(4):355-60. e1.
203. Tomas-Carus P, Leite N, Raimundo A. Effects of Physical Exercise on the Quality of Life of Type 2 Diabetes Patients. Quality of Life: IntechOpen; 2019.
204. Bamman MM, Wick TM, Carmona-Moran CA, Bridges SL, Jr. Exercise Medicine for Osteoarthritis: Research Strategies to Maximize Effectiveness. Arthritis Care Res (Hoboken). 2016;68(3):288-91.
205. Castillo JM, Jivraj S, Ng Fat L. The regional geography of alcohol consumption in England: Comparing drinking frequency and binge drinking. Health & Place. 2017;43:33-40.
206. Knuchel-Takano A, Hunt D, Jaccard A, Bhimjiyani A, Brown M, Retat L, et al. Modelling the implications of reducing smoking prevalence: the benefits of increasing the UK tobacco duty escalator to public health and economic outcomes. Tobacco Control. 2018;27(e2):e124-e9.
207. Bennett H. Life is sweet. New Scientist. 2019;241(3223):34-7.

208. Sharapov SZ, Tsepilov YA, Klaric L, Mangino M, Thareja G, Shadrina AS, et al. Defining the genetic control of human blood plasma N-glycome using genome-wide association study. Human Molecular Genetics. 2019;28(12):2062-77.
209. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nature Genetics. 2016;48:245.

Appendix 1 – Description of Derived IgG N-Glycosylation Features

This table depicts all the calculations used for describing the glycosylation features within the paper. These are based on those presented in Lauc et al. (2013) *PLoS One* and Russell et al. (2017) *Glycobiology*. The IgG GPs are derived using the ultra-performance liquid chromatography (UPLC; **Section 2.3**) and used to calculate the total derived N-glycosylation features. The neutral IgG GPs are those containing N-glycans without terminating Neu5Ac. Although not explored in this thesis, the neutral IgG GPs are used for calculating the neutral derived N-glycosylation features. Thus, there were an additional 32 derived N-glycosylation features explored.

Page intentionally left blank

Table A1.1: *The UPLC-measured IgG GPs, also depicted in Figure 2.2*

FEATURE	DESCRIPTION	FORMULA
IgG GPs		
GP1	The percentage of FA1 glycan in total IgG glycans	GP1 / GP* 100
GP2	The percentage of A2 glycan in total IgG glycans	GP2 / GP* 100
GP3	The percentage of A2B glycan in total IgG glycans	GP3 / GP* 100
GP4	The percentage of FA2 glycan in total IgG glycans	GP4 / GP* 100
GP5	The percentage of M5 glycan in total IgG glycans	GP5 / GP* 100
GP6	The percentage of FA2B glycan in total IgG glycans	GP6 / GP* 100
GP7	The percentage of A2G1 glycan in total IgG glycans	GP7 / GP* 100
GP8	The percentage of FA2[6]G1 glycan in total IgG glycans	GP8 / GP* 100
GP9	The percentage of FA2[3]G1 glycan in total IgG glycans	GP9 / GP* 100
GP10	The percentage of FA2[6]BG1 glycan in total IgG glycans	GP10 / GP* 100
GP11	The percentage of FA2[3]BG1 glycan in total IgG glycans	GP11 / GP* 100
GP12	The percentage of A2G2 glycan in total IgG glycans	GP12 / GP* 100
GP13	The percentage of A2BG2 glycan in total IgG glycans	GP13 / GP* 100
GP14	The percentage of FA2G2 glycan in total IgG glycans	GP14 / GP* 100
GP15	The percentage of FA2BG2 glycan in total IgG glycans	GP15 / GP* 100
GP16	The percentage of FA2G1S1 glycan in total IgG glycans	GP16 / GP * 100
GP17	The percentage of A2G2S1 glycan in total IgG glycans	GP17/ GP * 100
GP18	The percentage of FA2G2S1 glycan in total IgG glycans	GP18 / GP * 100
GP19	The percentage of FA2BG2S1 glycan in total IgG glycans	GP19 / GP * 100
GP20	The percentage of FA2FG2S1 glycan in total IgG glycans	GP20 / GP * 100
GP21	The percentage of A2G2S2 glycan in total IgG glycans	GP21 / GP * 100
GP22	The percentage of A2BG2S2 glycan in total IgG glycans	GP22 / GP * 100
GP23	The percentage of FA2G2S2 glycan in total IgG glycans	GP23 / GP * 100
GP24	The percentage of FA2BG2S2 glycan in total IgG glycans	GP24 / GP * 100
	Total area	GP = SUM(GP1:GP24)=100

F – core fucosylation; A2 – biantennary; B – bisecting N-acetylglucosamine Gx – galactosylation; Sx - sialylation

Page intentionally left blank

Table A1.2: *N-glycosylation features derived using the IgG GPs*

FEATURE	DESCRIPTION	FORMULA
Total Derived N-Glycosylation Features		
FGS/(FG+FGS)	The percentage of sialylation of fucosylated galactosylated structures without bisecting GlcNAc in total IgG glycans	SUM(GP16 + GP18 + GP23) / SUM(GP16 + GP18 + GP23 + G 100
FBGS/(FBG+FBGS)	The percentage of sialylation of fucosylated galactosylated structures with bisecting GlcNAc in total IgG glycans	SUM(GP19 + GP24) / SUM(GP19 + GP24 + GP10 + GP11 + G
FGS/(F+FG+FGS)	The percentage of sialylation of all fucosylated structures without bisecting GlcNAc in total IgG glycans	SUM(GP16 + GP18 + GP23) / SUM(GP16 + GP18 + GP23 + G GP14)* 100
FBGS/(FB+FBG+FBGS)	The percentage of sialylation of all fucosylated structures with bisecting GlcNAc in total IgG glycans	SUM(GP19 + GP24) / SUM(GP19 + GP24 + GP6 + GP10 + GI
FG1S1/(FG1+FG1S1)	The percentage of monosialylation of fucosylated monogalactosylated structures in total IgG glycans	GP16 / SUM(GP16 + GP8 + GP9)* 100
FG2S1/(FG2+FG2S1+FG2S2)	The percentage of monosialylation of fucosylated digalactosylated structures in total IgG glycans	GP18 / SUM(GP18 + GP14 + GP23)* 100
FG2S2/(FG2+FG2S1+FG2S2)	The percentage of disialylation of fucosylated digalactosylated structures in total IgG glycans	GP23 / SUM(GP23 + GP14 + GP18)* 100
FBG2S1/(FBG2+FBG2S1+FBG2S2)	The percentage of monosialylation of fucosylated digalactosylated structures with bisecting GlcNAc in total IgG glycans	GP19 / SUM(GP19 + GP15 + GP24)* 100
FBG2S2/(FBG2+FBG2S1+FBG2S2)	The percentage of disialylation of fucosylated digalactosylated structures with bisecting GlcNAc in total IgG glycans	GP24 / SUM(GP24 + GP15 + GP19)* 100
FtotalS1/FtotalS2	Ratio of all fucosylated (+/- bisecting GlyNAc) monosialylated and disialylated structures in total IgG glycans	SUM(GP16 + GP18 + GP19) / SUM(GP23 + GP24)
FS1/FS2	Ratio of fucosylated (without bisecting GlcNAc) monosialylated and disialylated structures in total IgG glycans	SUM(GP16 + GP18) / GP23
FBS1/FBS2	Ratio of fucosylated (with bisecting GlcNAc) monosialylated and disialylated structures in total IgG glycans	GP19 / GP24
FBStotal/FStotal	Ratio of all fucosylated sialylated structures with and without bisecting GlcNAc	SUM(GP19 + GP24) / SUM(GP16 + GP18 + GP23)
FBS1/(FS1+FBS1)	The incidence of bisecting GlcNAc in all fucosylated monosialylated structures in total IgG glycans	GP19 / SUM(GP16 + GP18 + GP19)
FBS2/(FS2+FBS2)	The incidence of bisecting GlcNAc in all fucosylated disialylated structures in total IgG glycans	GP24 / SUM(GP23 + GP24)A4:A6:C17

Page intentionally left blank

Table A1.3: *The neutral UPLC-measured IgG GPs; those lacking terminal sialylation. These are used to derive the N-glycosylation features (Table A1.4)*

FEATURE	DESCRIPTION	FORMULA
Neutral IgG GPs		
GP1n	The percentage of FA1 glycan in total neutral IgG glycans (GPn)	GP1 / GPn* 100
GP2n	The percentage of A2 glycan in total neutral IgG glycans (GPn)	GP2 / GPn* 100
GP3n	The percentage of A2B glycan in total IgG glycans	GP3 / GP* 100
GP4n	The percentage of FA2 glycan in total neutral IgG glycans (GPn)	GP4 / GPn* 100
GP5n	The percentage of M5 glycan in total neutral IgG glycans (GPn)	GP5 / GPn* 100
GP6n	The percentage of FA2B glycan in total neutral IgG glycans (GPn)	GP6 / GPn* 100
GP7n	The percentage of A2G1 glycan in total Ineutral IgG glycans (GPn)	GP7 / GPn* 100
GP8n	The percentage of FA2[6]G1 glycan in total neutral IgG glycans (GPn)	GP8 / GPn* 100
GP9n	The percentage of FA2[3]G1 glycan in total neutral IgG glycans (GPn)	GP9 / GPn* 100
GP10n	The percentage of FA2[6]BG1 glycan in total neutral IgG glycans (GPn)	GP10 / GPn* 100
GP11n	The percentage of FA2[3]BG1 glycan in total neutral IgG glycans (GPn)	GP11 / GPn* 100
GP12n	The percentage of A2G2 glycan in total neutral IgG glycans (GPn)	GP12 / GPn* 100
GP13n	The percentage of A2BG2 glycan in total neutral IgG glycans (GPn)	GP13 / GPn* 100
GP14n	The percentage of FA2G2 glycan in total neutral IgG glycans (GPn)	GP14 / GPn* 100
GP15n	The percentage of FA2BG2 glycan in total neutral IgG glycans (GPn)	GP15 / GPn* 100
	Neutral area	GPn = SUM(GP1n:GP15n)

F – core fucosylation; A2 – biantennary; B – bisecting N-acetylglucosamine Gx – galactosylation; Sx - sialylation

Page intentionally left blank

Table A1.4: *N-glycosylation features derived using the neutral IgG GPs*

FEATURE	DESCRIPTION	FORMULA
Neutral N-Glycosylation Features		
G0n	The percentage of agalactosylated structures in total neutral IgG glycans	SUM(GP1n: GP6n)
G1n	The percentage of monogalactosylated structures in total neutral IgG glycans	SUM(GP7n: GP11n)
G2n	The percentage of digalactosylated structures in total neutral IgG glycans	SUM(GP12n: GP15n)
Fn total	The percentage of all fucosylated (+/- bisecting GlcNAc) structures in total neutral IgG glycans	SUM(GP1n+ GP4n+ GP5n+ GP6n+ GP8n+ GP9n+ GP10n+ GP11n+ GP14n+ GP15n)
FG0n total/G0n	The percentage of fucosylation of agalactosylated structures	SUM(GP1n+ GP4n+ GP5n+ GP6n) / G0n * 100
FG1n total/G1n	The percentage of fucosylation of monogalactosylated structures	SUM(GP8n+ GP9n+ GP10n+ GP11n) / G1n * 100
FG2n total /G2n	The percentage of fucosylation of digalactosylated structures	SUM(GP14n+ GP15) / G2n * 100
Fn	The percentage of fucosylated (without bisecting GlcNAc) structures in total neutral IgG glycans	SUM(GP1n+ GP4n+ GP5n+ GP8n+ GP9n+ GP14n)
FG0n/G0n	The percentage of fucosylation (without bisecting GlcNAc) of agalactosylated structures	SUM(GP1n+ GP4n+ GP5n) / G0n * 100
FG1n/G1n	The percentage of fucosylation (without bisecting GlcNAc) of monogalactosylated structures	SUM(GP8n+ GP9n) / G1n * 100
FG2n/G2n	The percentage of fucosylation (without bisecting GlcNAc) of digalactosylated structures	GP14n/ G2n * 100
FBn	The percentage of fucosylated (with bisecting GlcNAc) structures in total neutral IgG glycans	SUM(GP6n + GP10n + GP11n + GP15n)
FBG0n/G0n	The percentage of fucosylation (with bisecting GlcNAc) of agalactosylated structures	GP6n/ G0n * 100
FBG1n/G1n	The percentage of fucosylation (with bisecting GlcNAc) of monogalactosylated structures	SUM(GP10n + GP11n) / G1n * 100
FBG2n/G2n	The percentage of fucosylation (with bisecting GlcNAc) of digalactosylated structures	GP15) / G2n * 100
FBn/Fn	Ratio of fucosylated structures with and without bisecting GlcNAc	FBn/ Fn * 100
FBn/Fn total	The incidence of bisecting GlcNAc in all fucosylated structures in total neutral IgG glycans	FBn/ Fn total * 100

Page intentionally left blank

Appendix 2 – Manhattan Plots of GWAS

Below are the Manhattan plots from the GWAS in **Chapter 3**. In total, there were 24 IgG GPs, which were used to calculate a further 32 derived traits (**Appendix 1**). These derived traits are thought to summarise grouped N-glycan features that may demonstrate stronger genetic control. Please refer to **Chapter 3** for full discussion on these plots.

NOTE: The genotype file for chromosome 5 was corrupt so it was omitted from the GWAS.

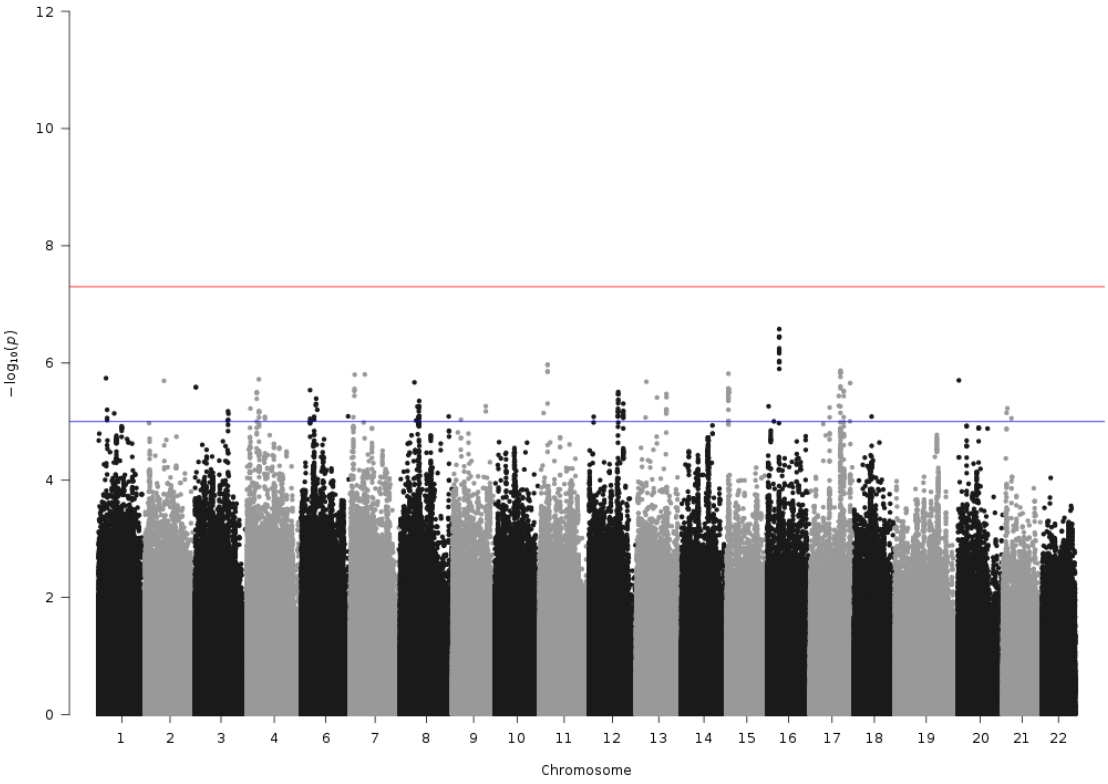


Figure A2.1: *Manhattan Plot of GWAS of GP1*

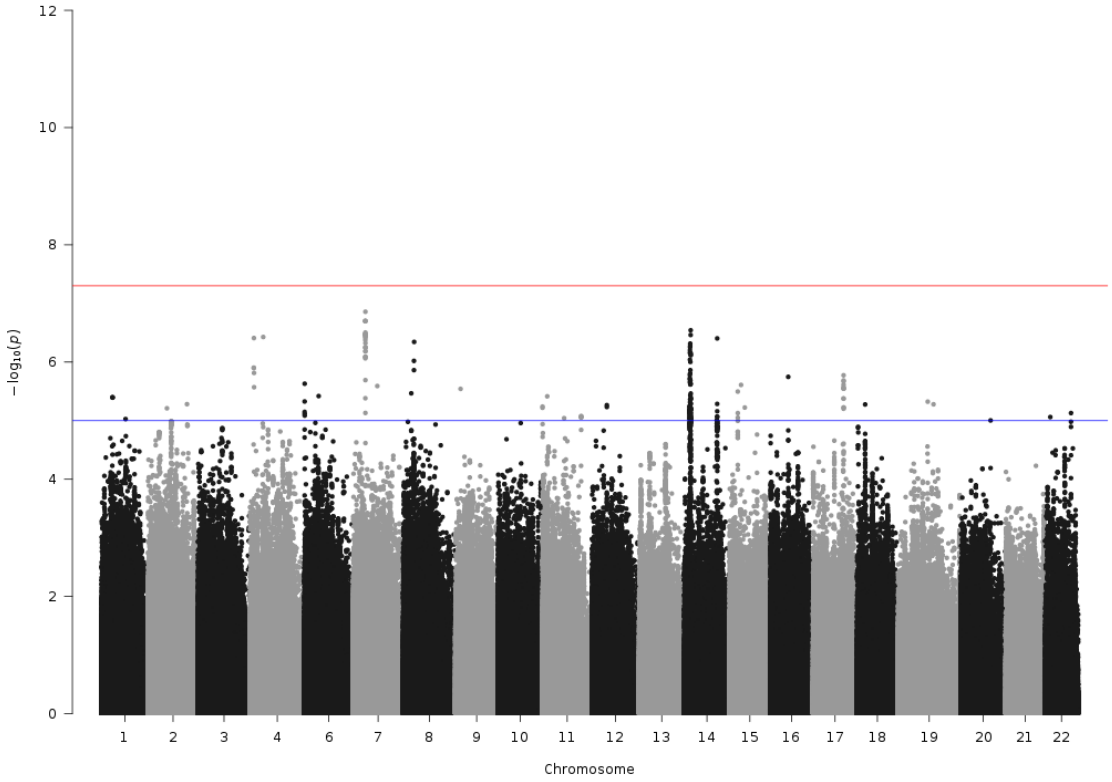


Figure A2.2: *Manhattan Plot of GWAS of GP2*

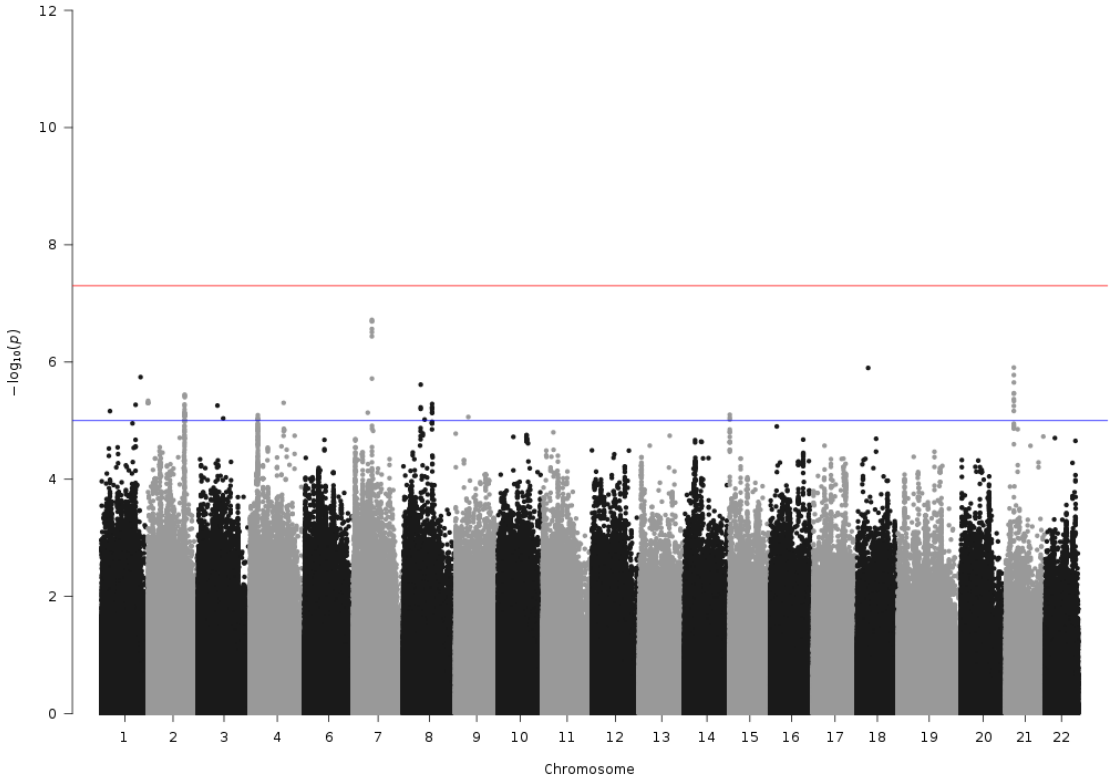


Figure A2.3: *Manhattan Plot of GWAS of GP3*

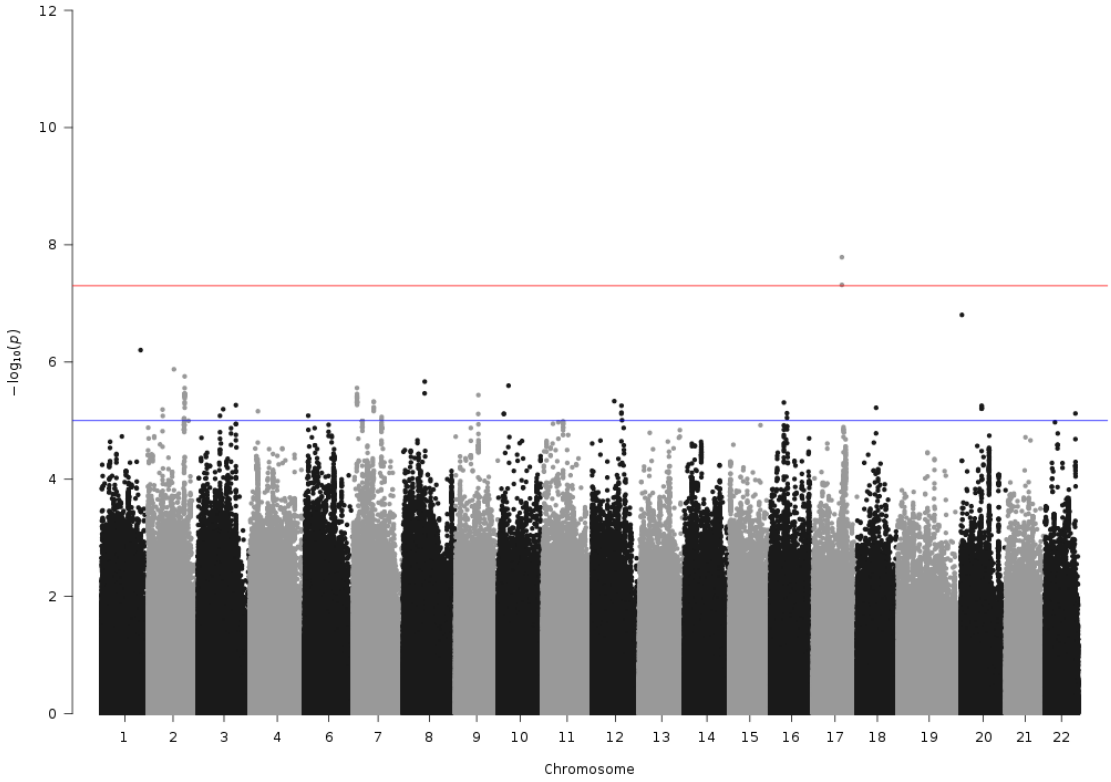


Figure A2.4: *Manhattan Plot of GWAS of GP4*

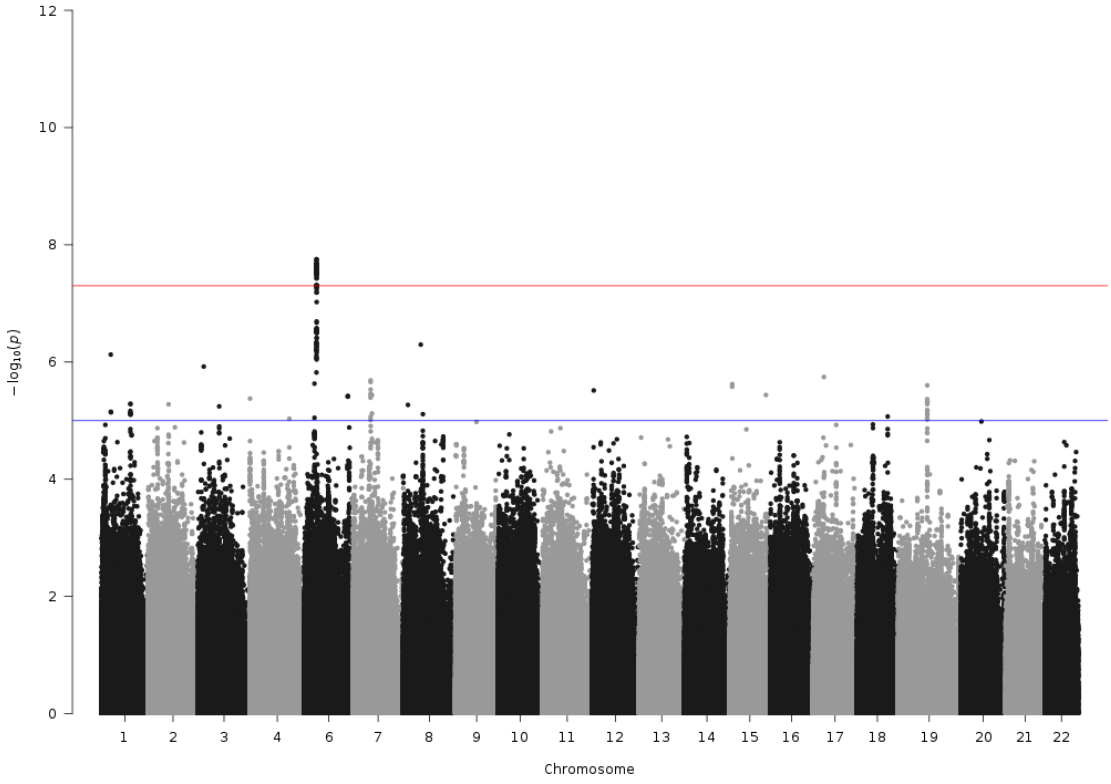


Figure A2.5: *Manhattan Plot of GWAS of GP5*

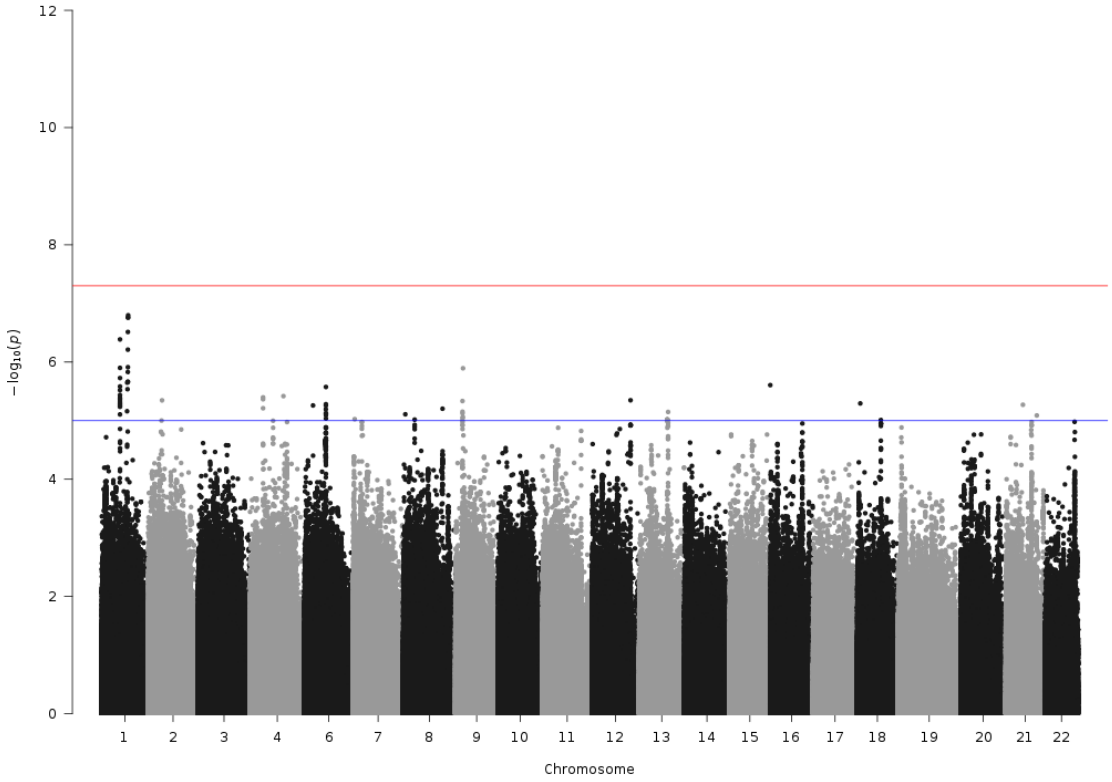


Figure A2.6: *Manhattan Plot of GWAS of GP6*

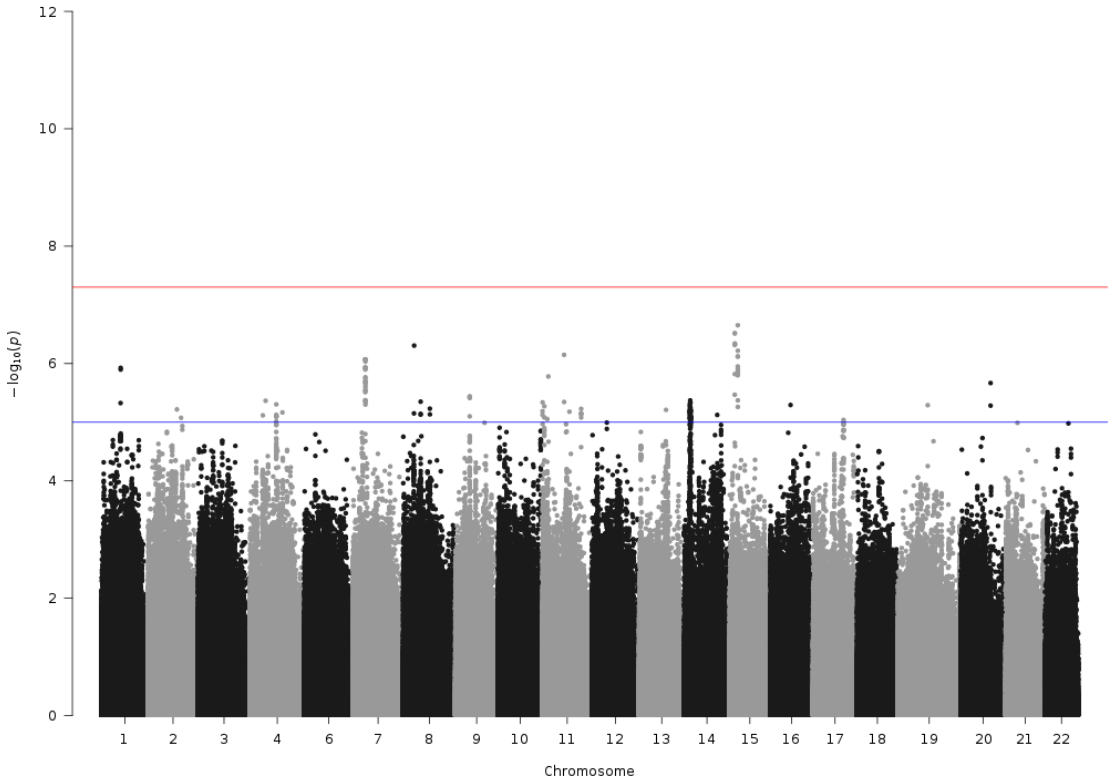


Figure A2.7: *Manhattan Plot of GWAS of GP7*

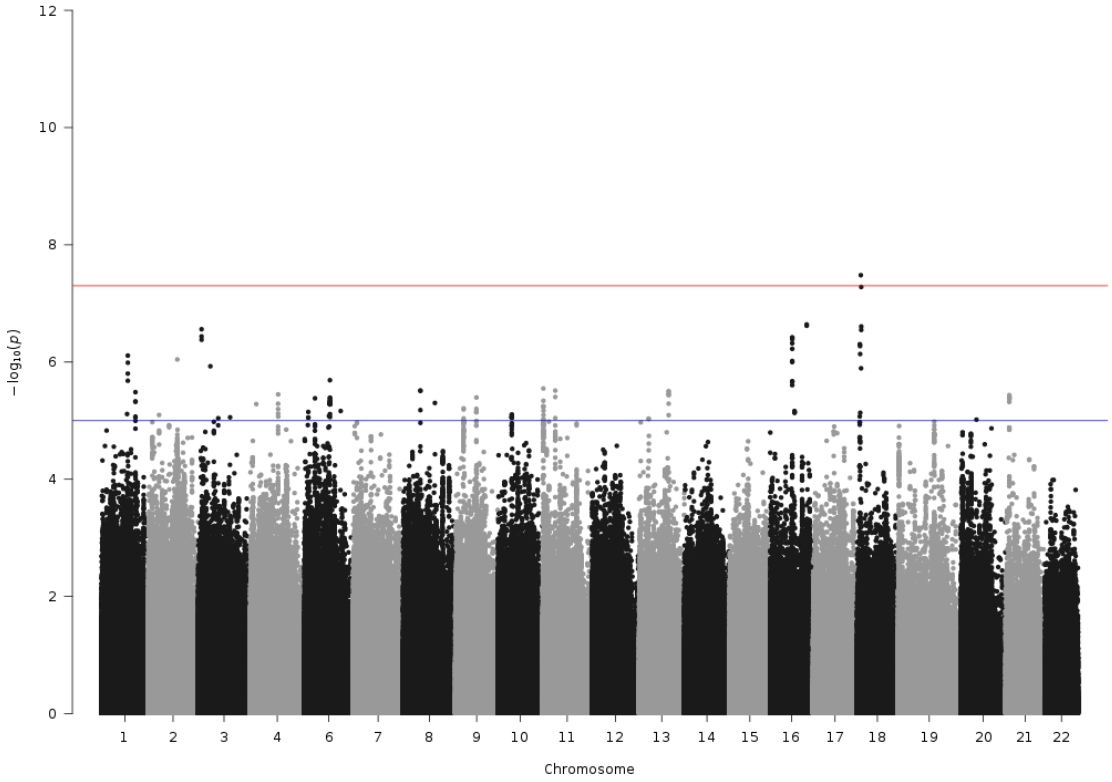


Figure A2.8: *Manhattan Plot of GWAS of GP8*

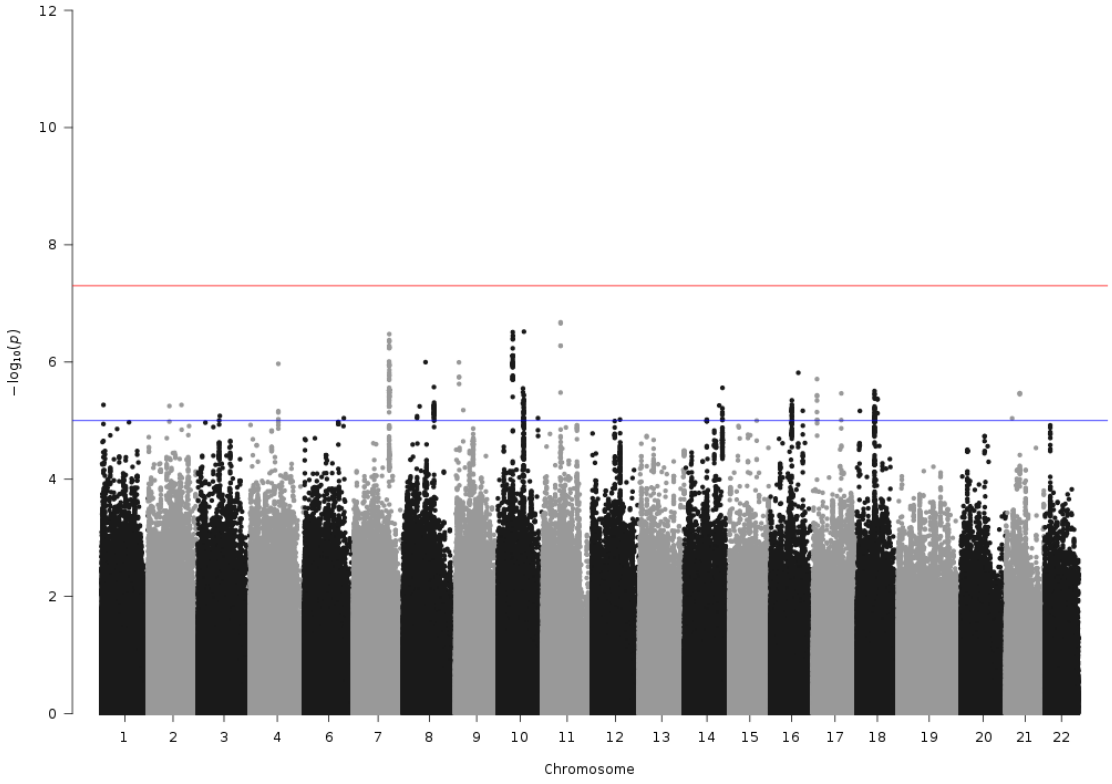


Figure A2.9: *Manhattan Plot of GWAS of GP9*

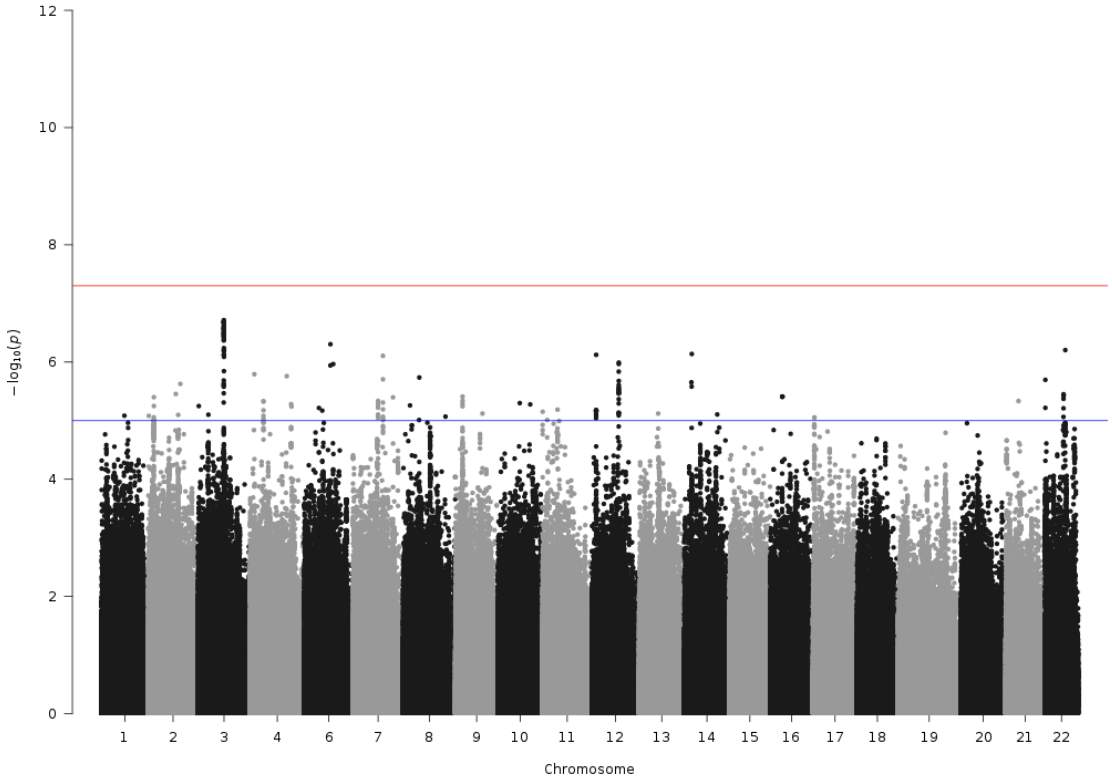


Figure A2.10: *Manhattan Plot of GWAS of GP10*

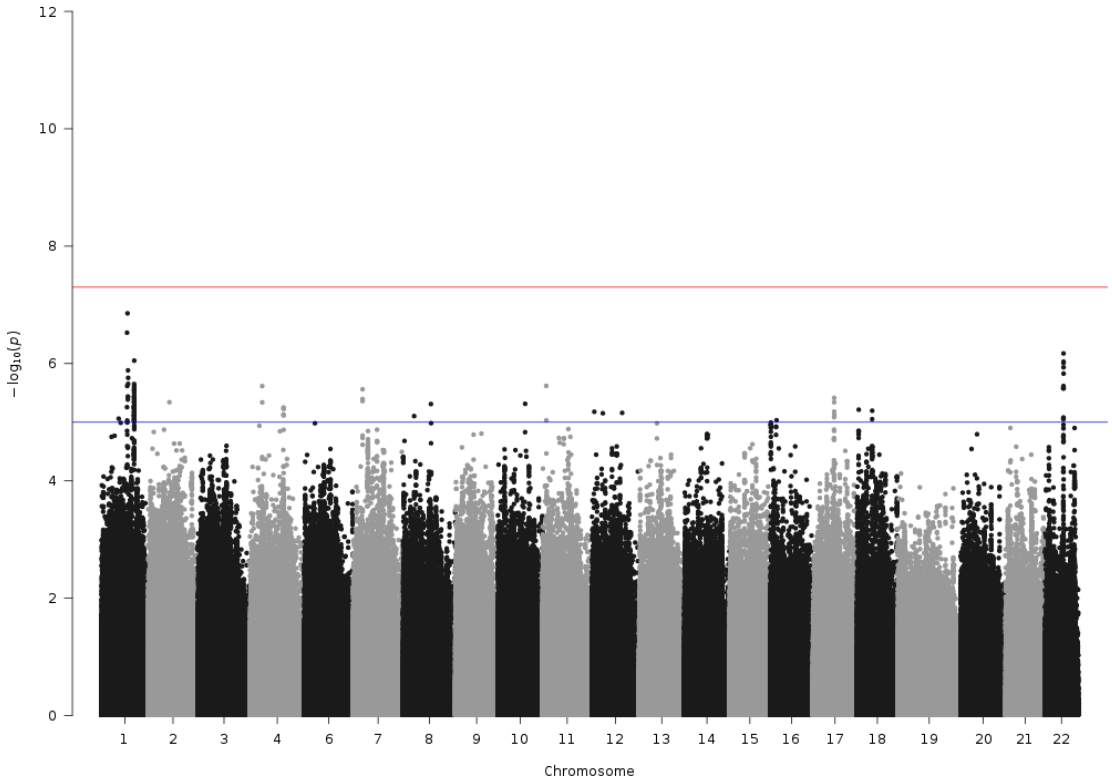


Figure A2.11: *Manhattan Plot of GWAS of GP11*

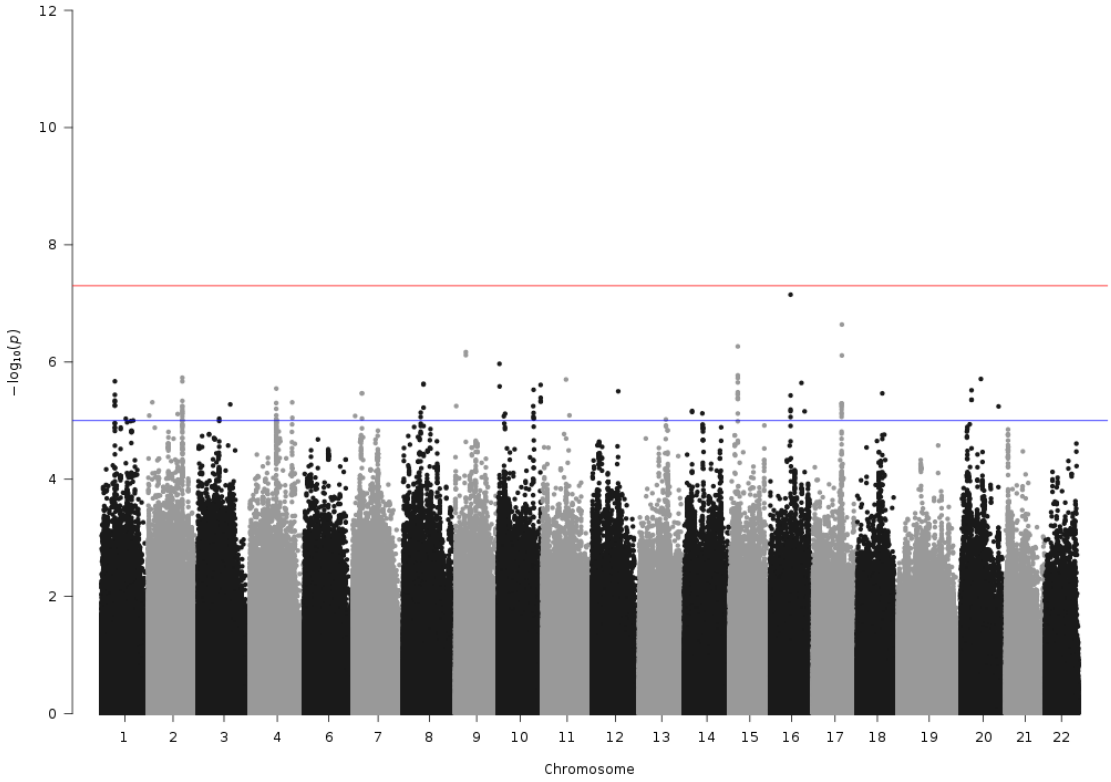


Figure A2.12: *Manhattan Plot of GWAS of GP12*

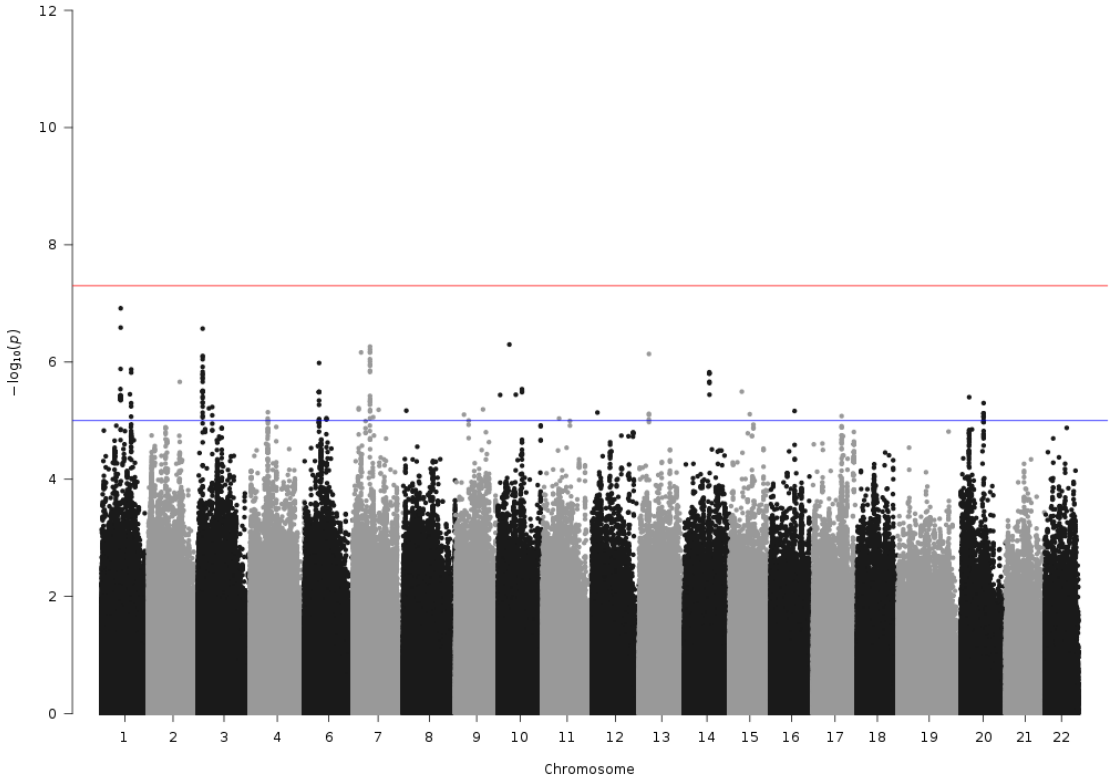


Figure A2.13: *Manhattan Plot of GWAS of GP13*

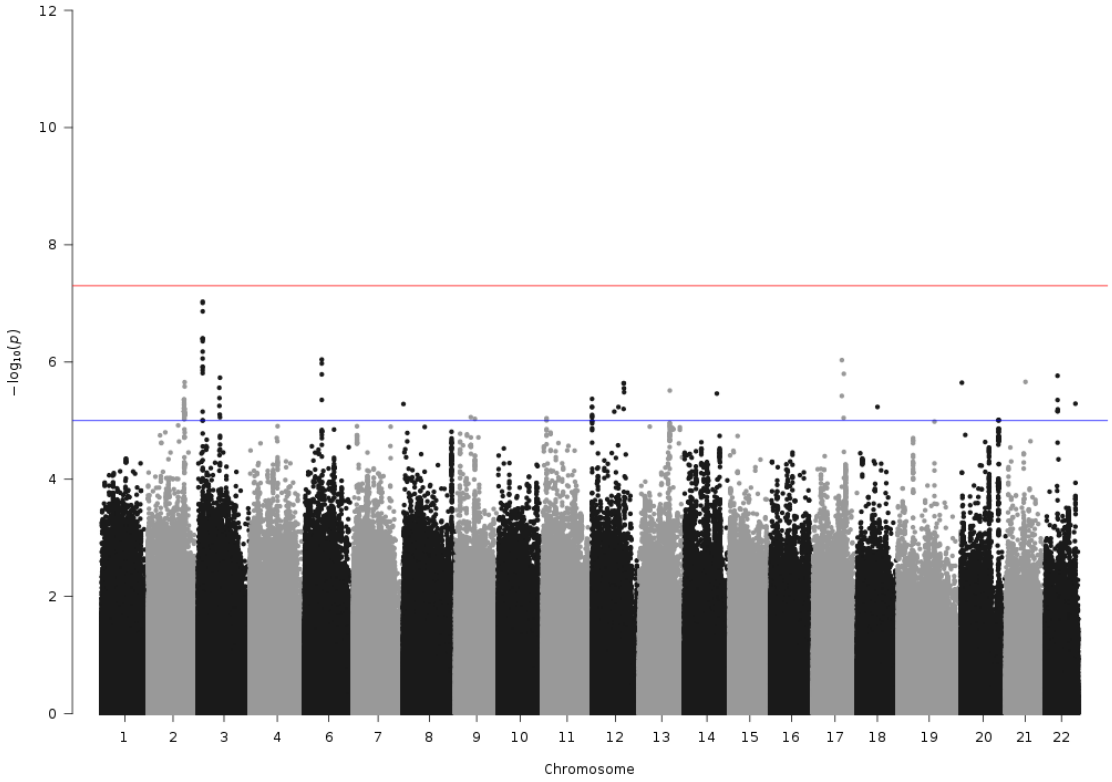


Figure A2.14: *Manhattan Plot of GWAS of GP14*

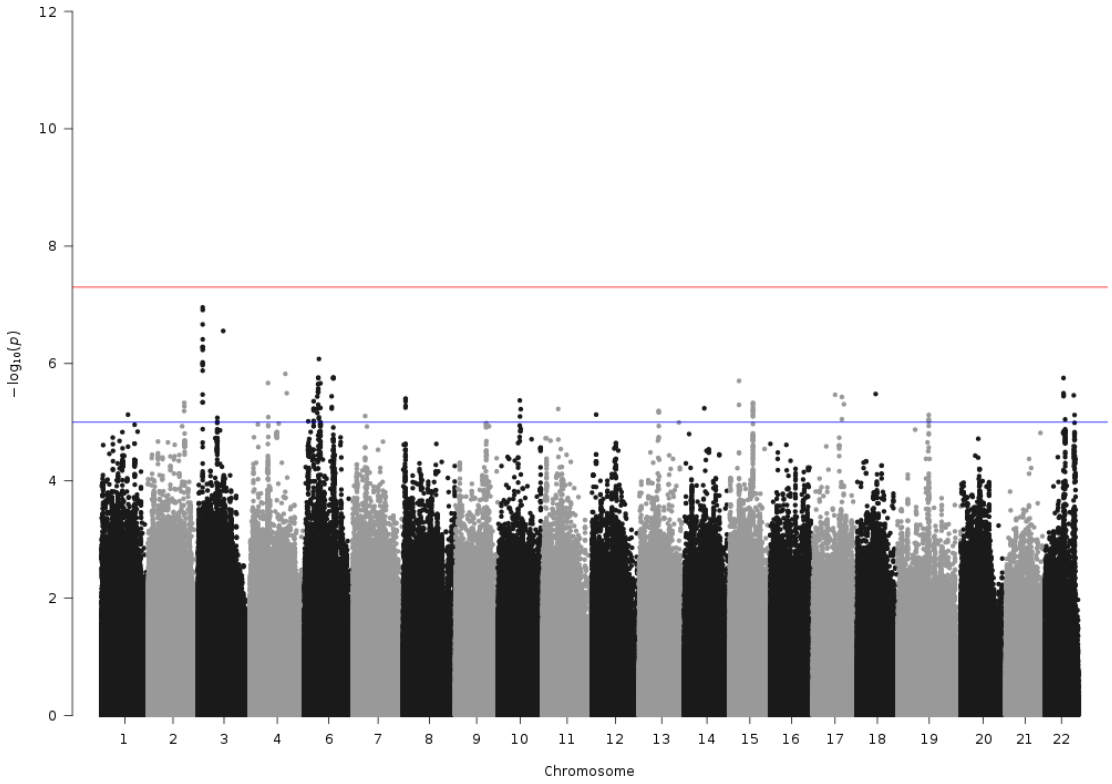


Figure A2.15: *Manhattan Plot of GWAS of GP15*

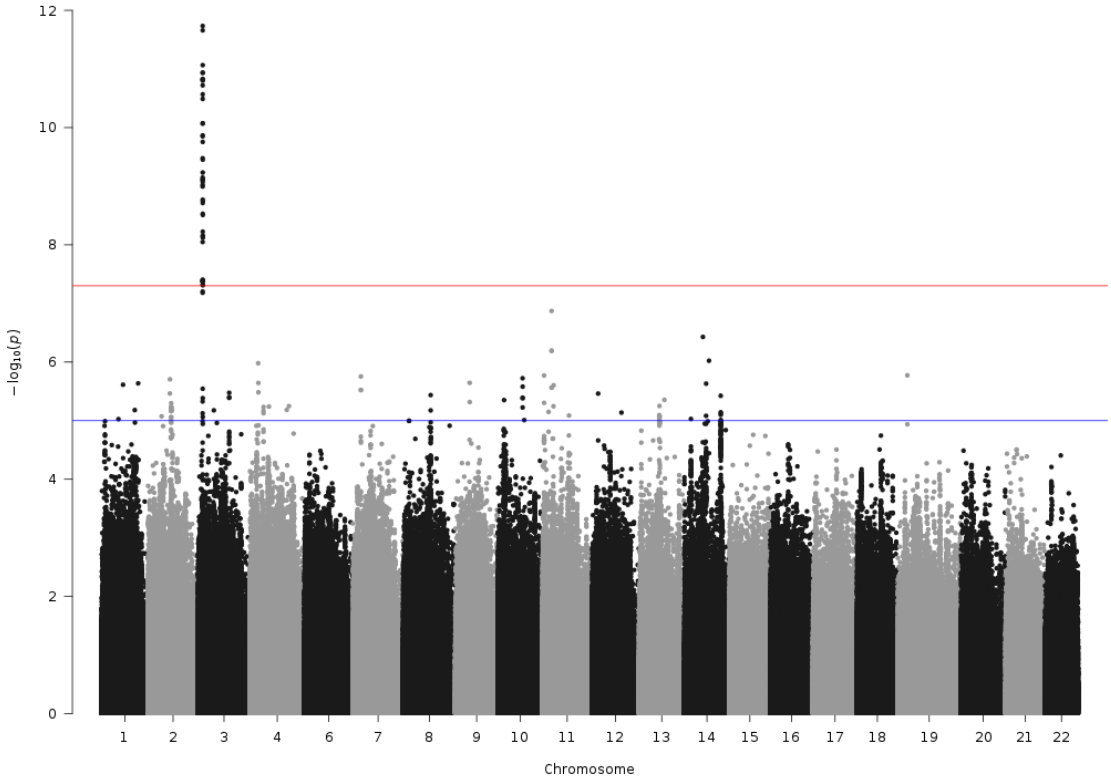


Figure A2.16: *Manhattan Plot of GWAS of GP16*

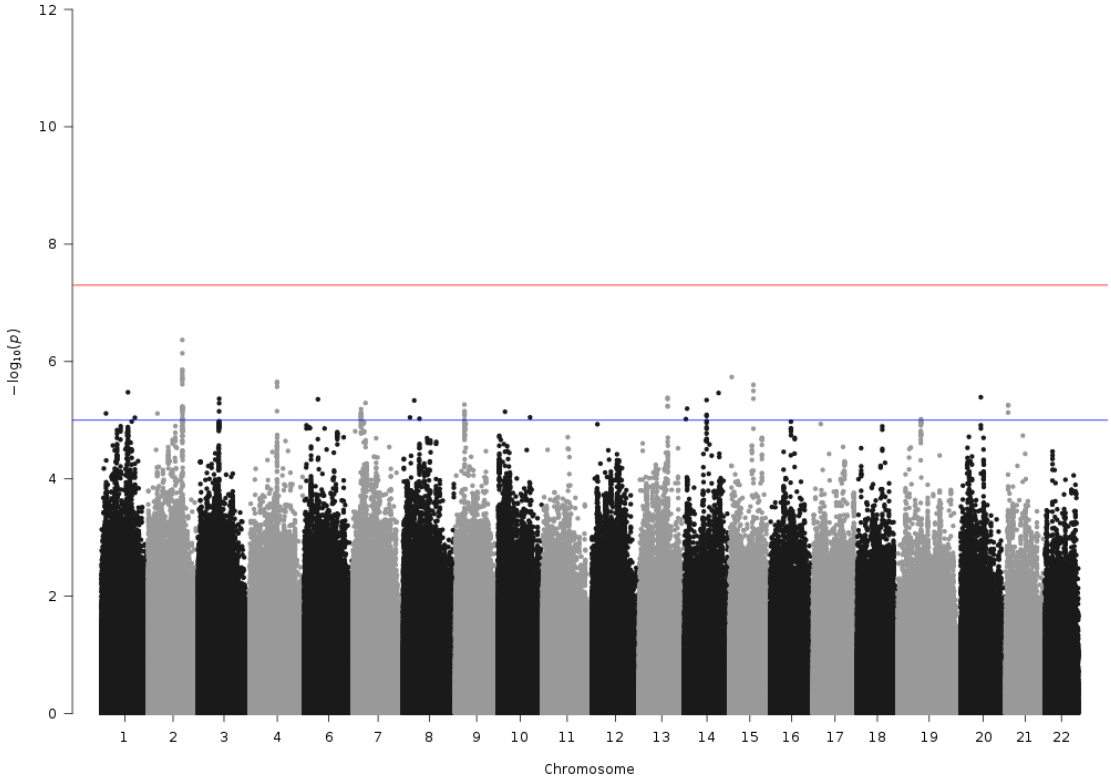


Figure A2.17: *Manhattan Plot of GWAS of GP17*

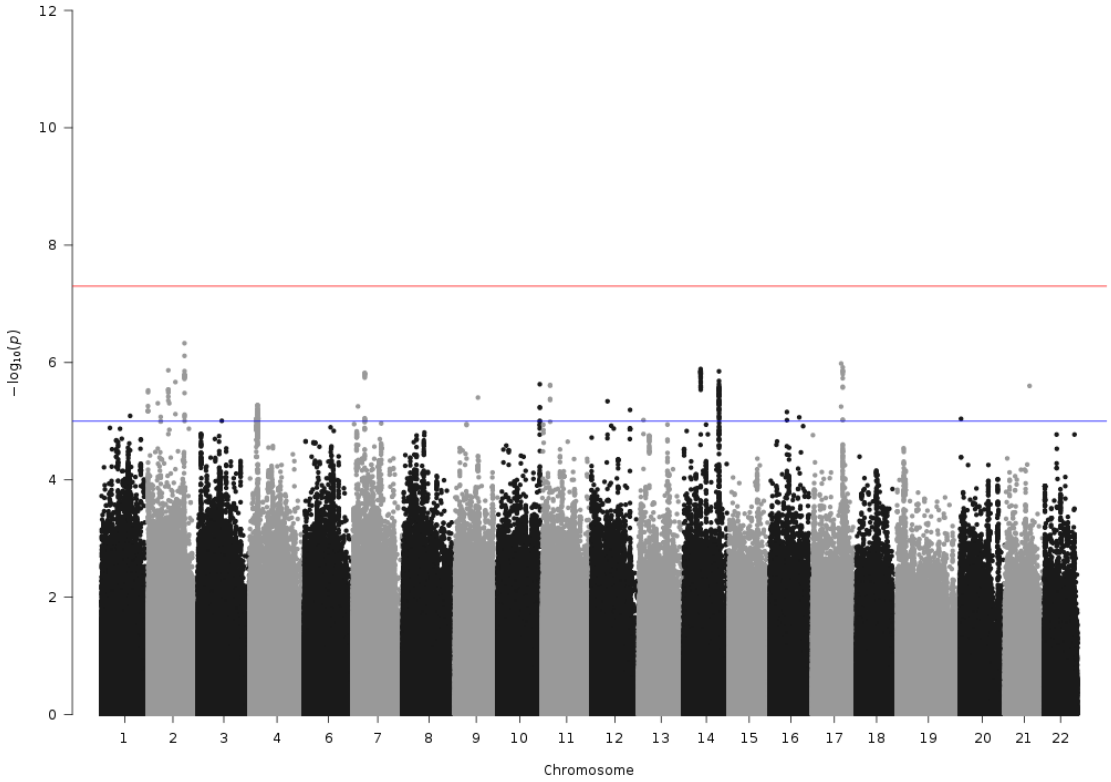


Figure A2.18: *Manhattan Plot of GWAS of GP18*

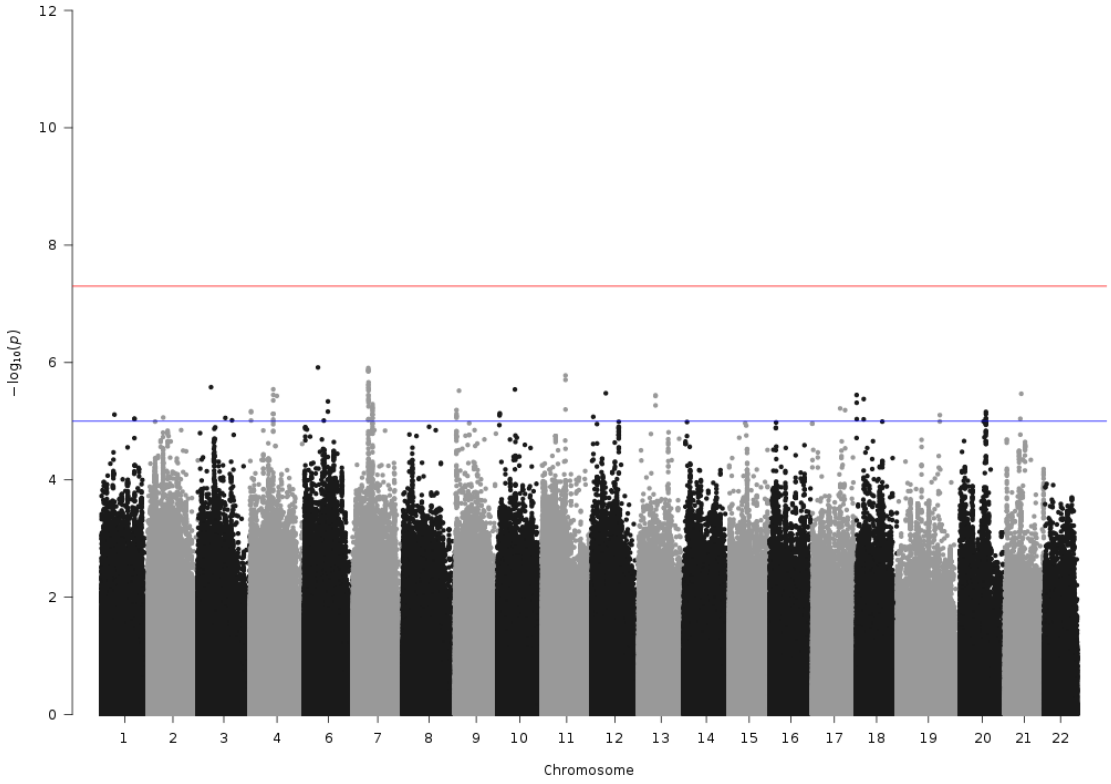


Figure A2.19: *Manhattan Plot of GWAS of GP19*

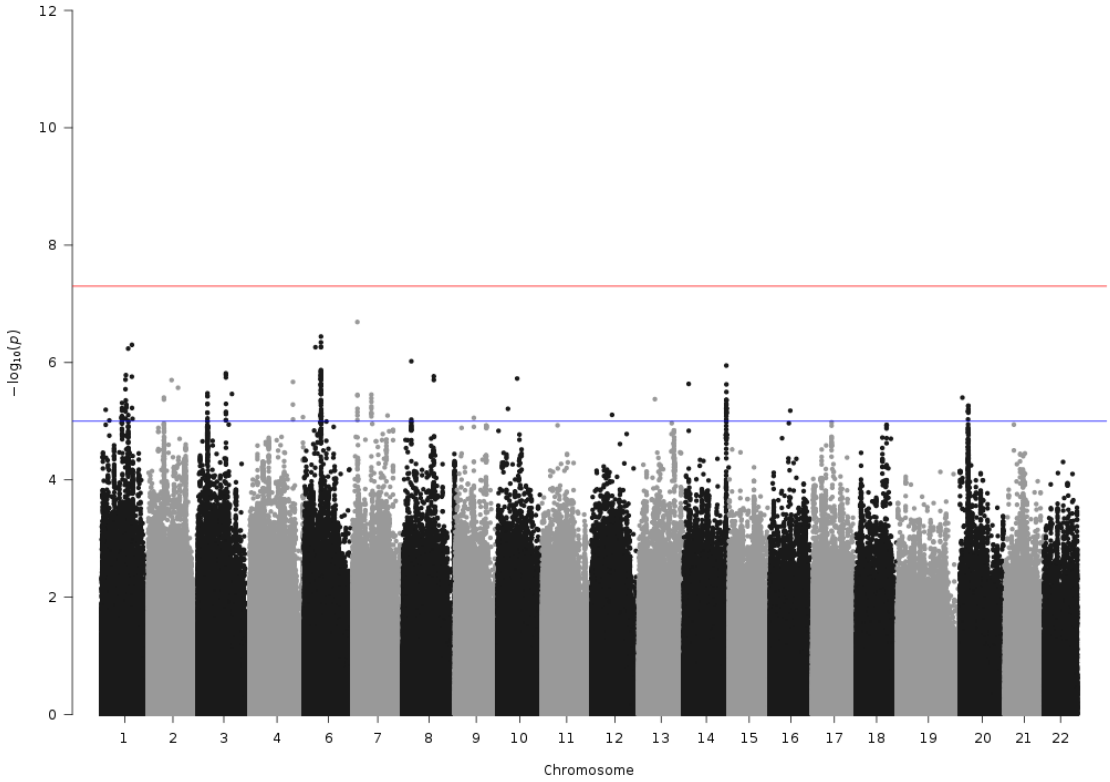


Figure A2.20: *Manhattan Plot of GWAS of GP20*

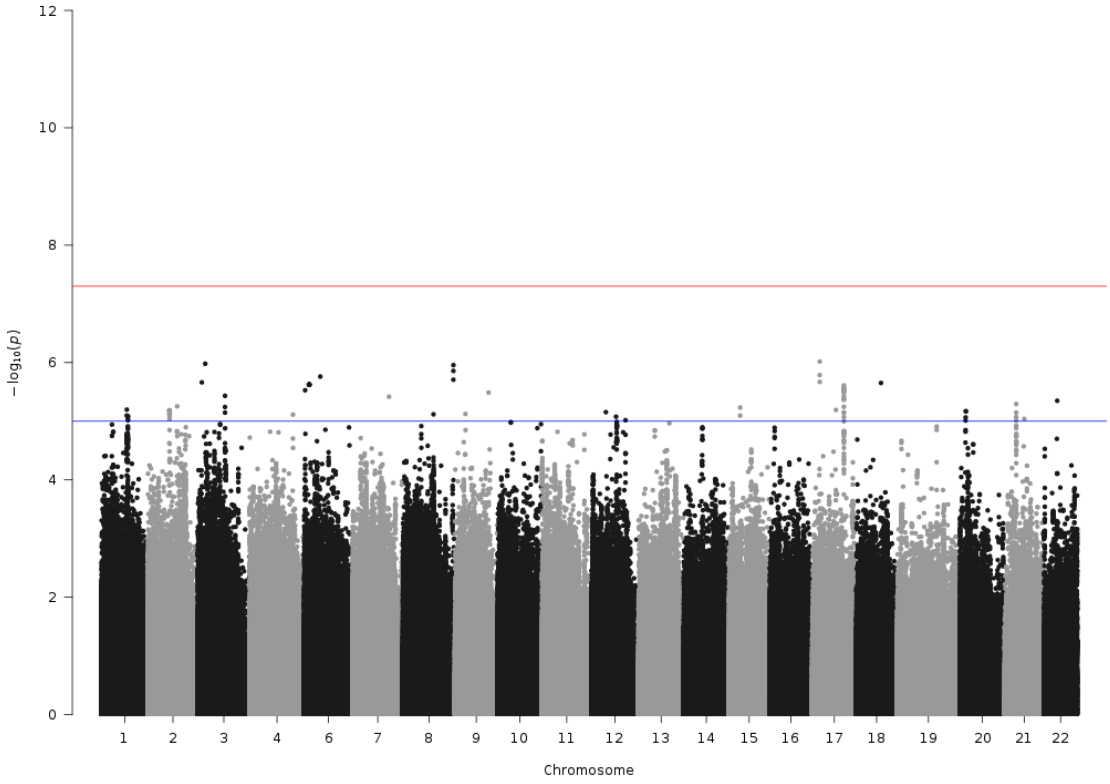


Figure A2.21: *Manhattan Plot of GWAS of GP21*

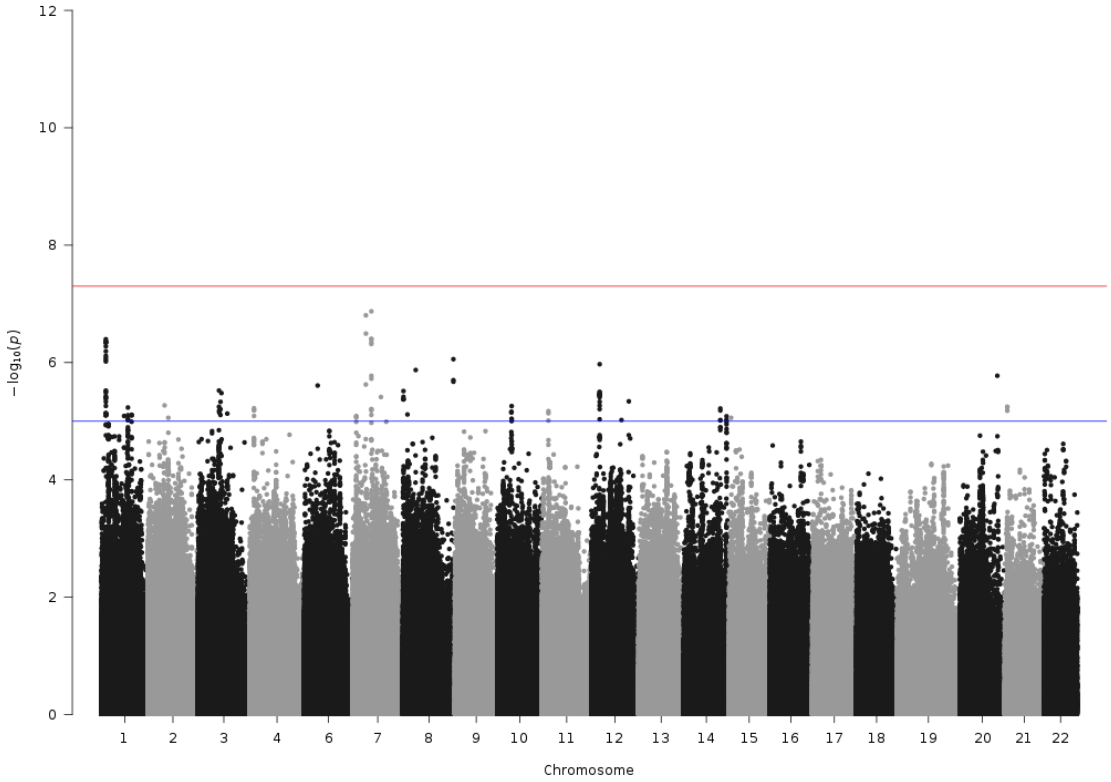


Figure A2.22: *Manhattan Plot of GWAS of GP22*

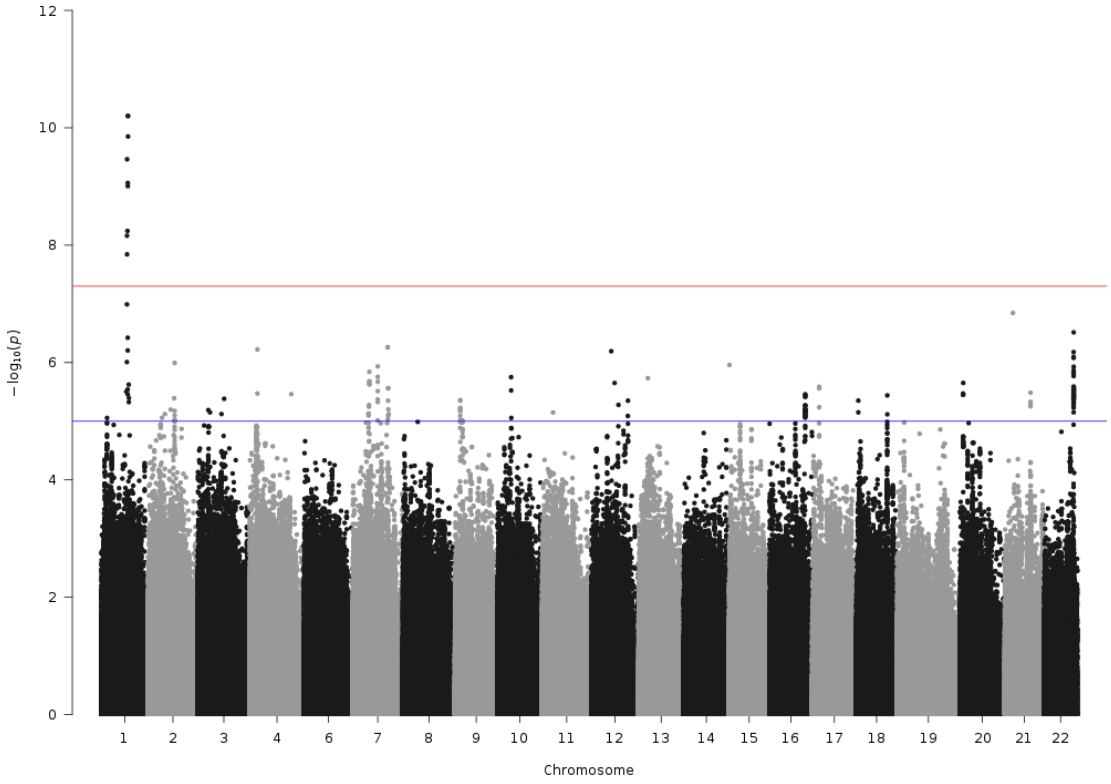


Figure A2.23: *Manhattan Plot of GWAS of GP23*

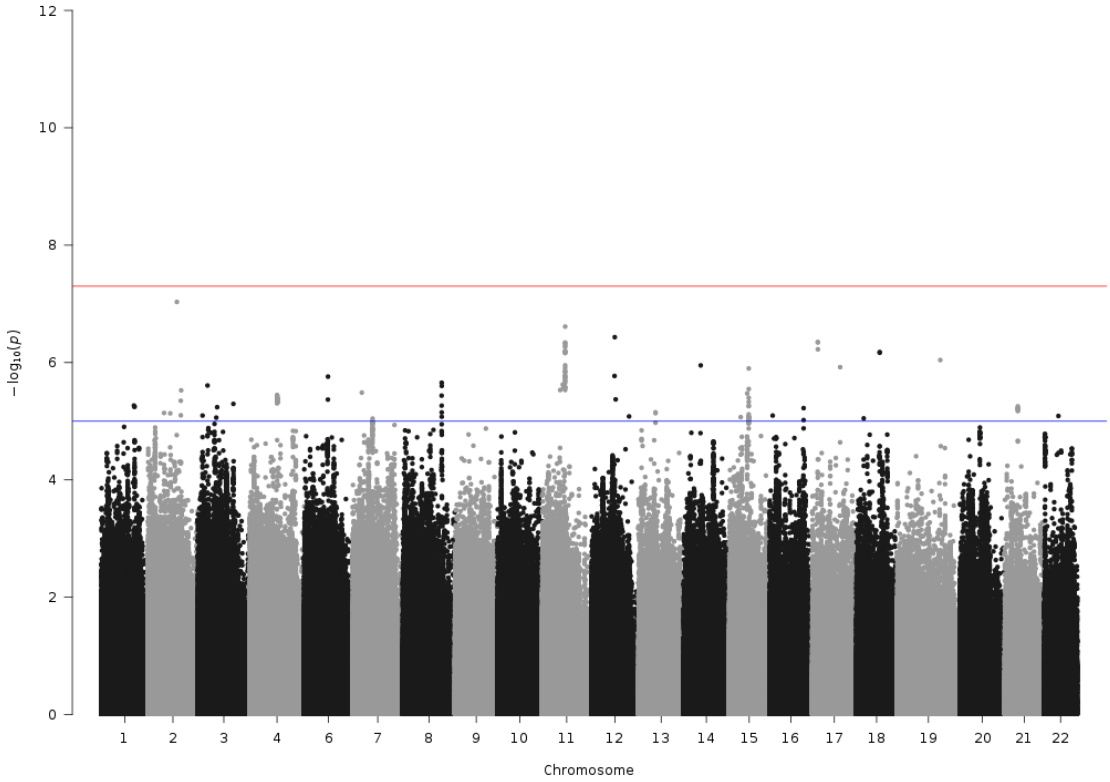


Figure A2.24: *Manhattan Plot of GWAS of GP24*

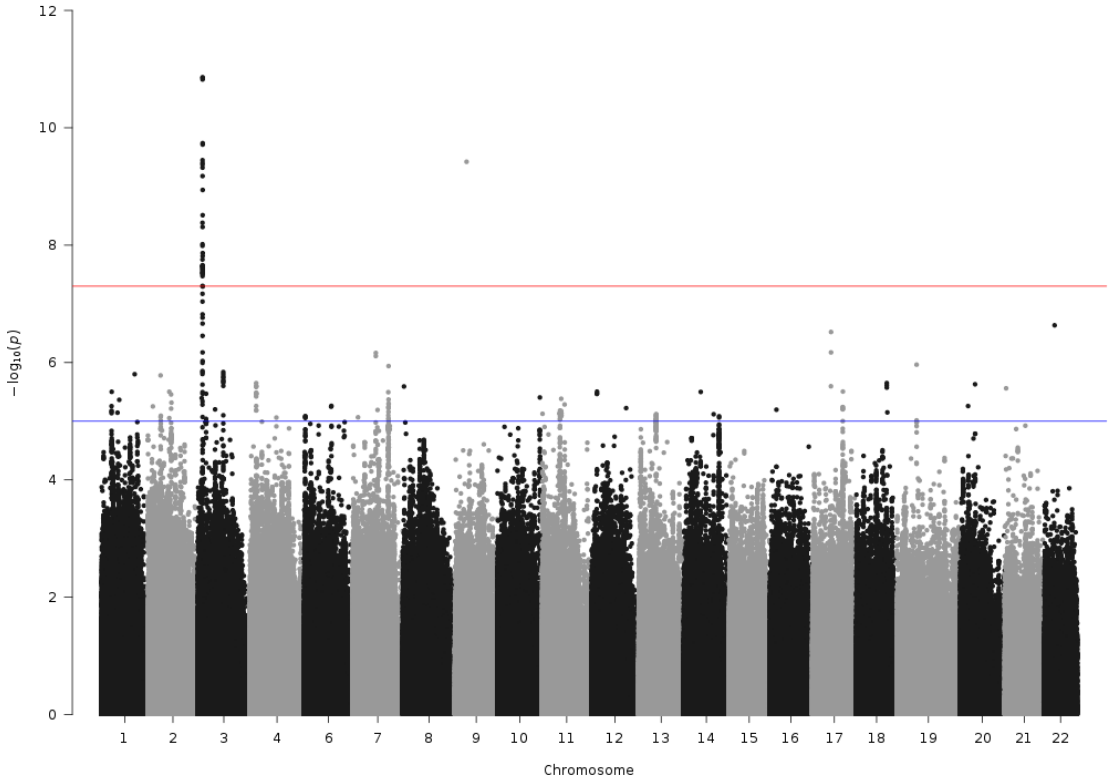


Figure A2.25: *Manhattan Plot of GWAS of IGP24, FGS/(FG+FGS)*

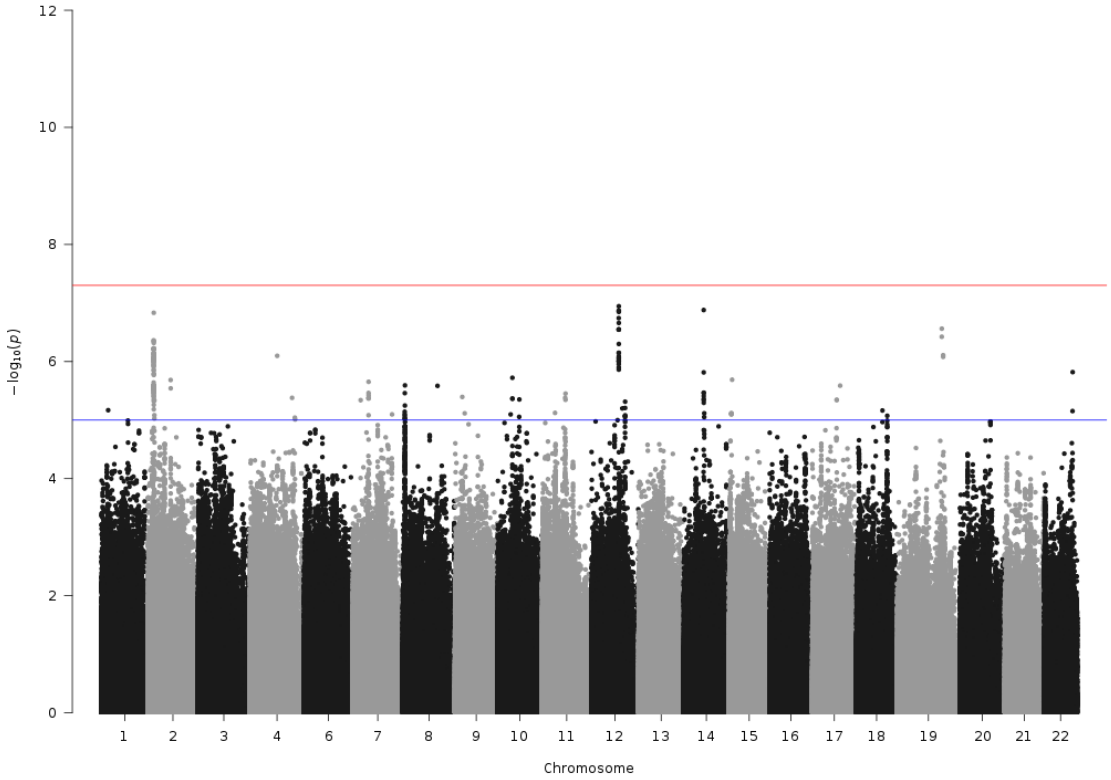


Figure A2.26: *Manhattan Plot of GWAS of IGP25, FBGS/(FBG+FBGS)*

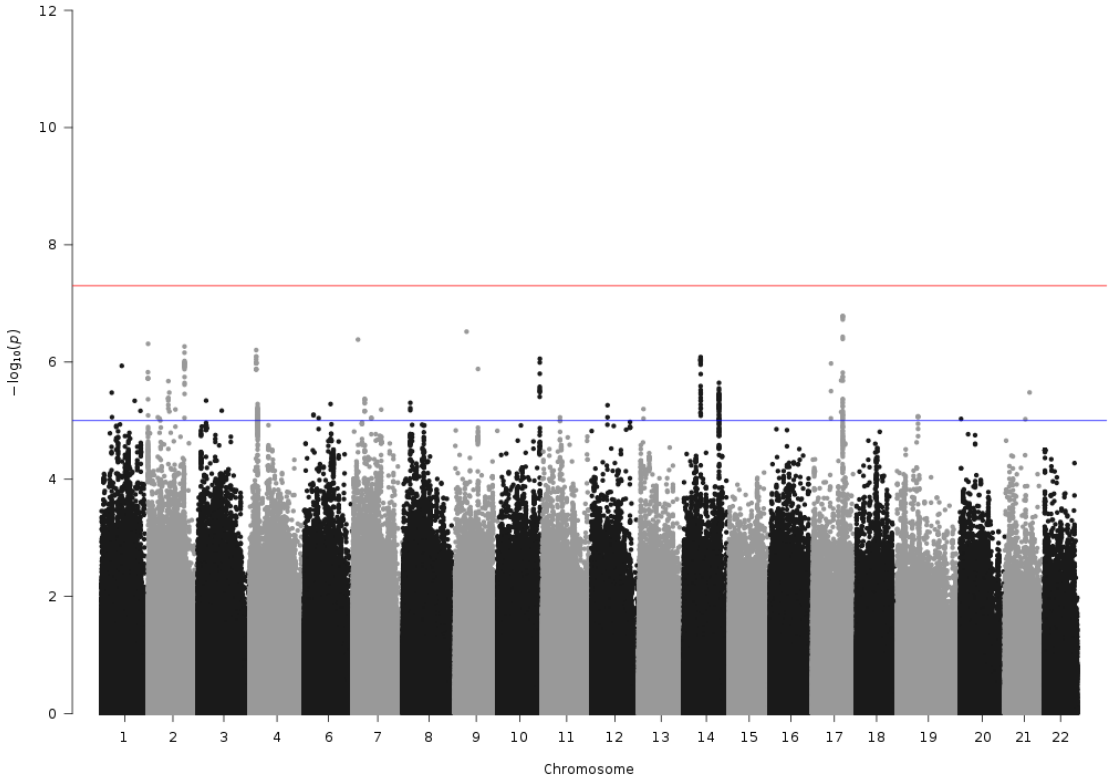


Figure A2. 27: *Manhattan Plot of GWAS of IGP26, FGS/(F+FG+FGS)*

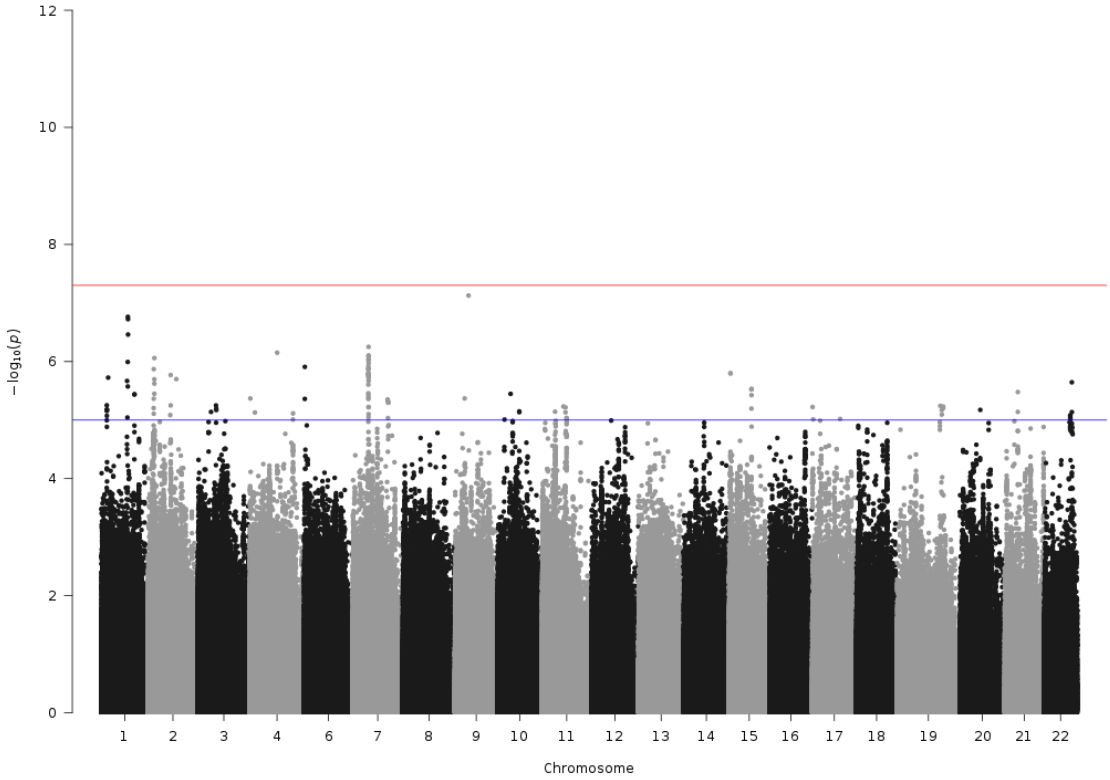


Figure A2.28: *Manhattan Plot of GWAS of IGP27, FBGS/(FB+FG+FGS)*

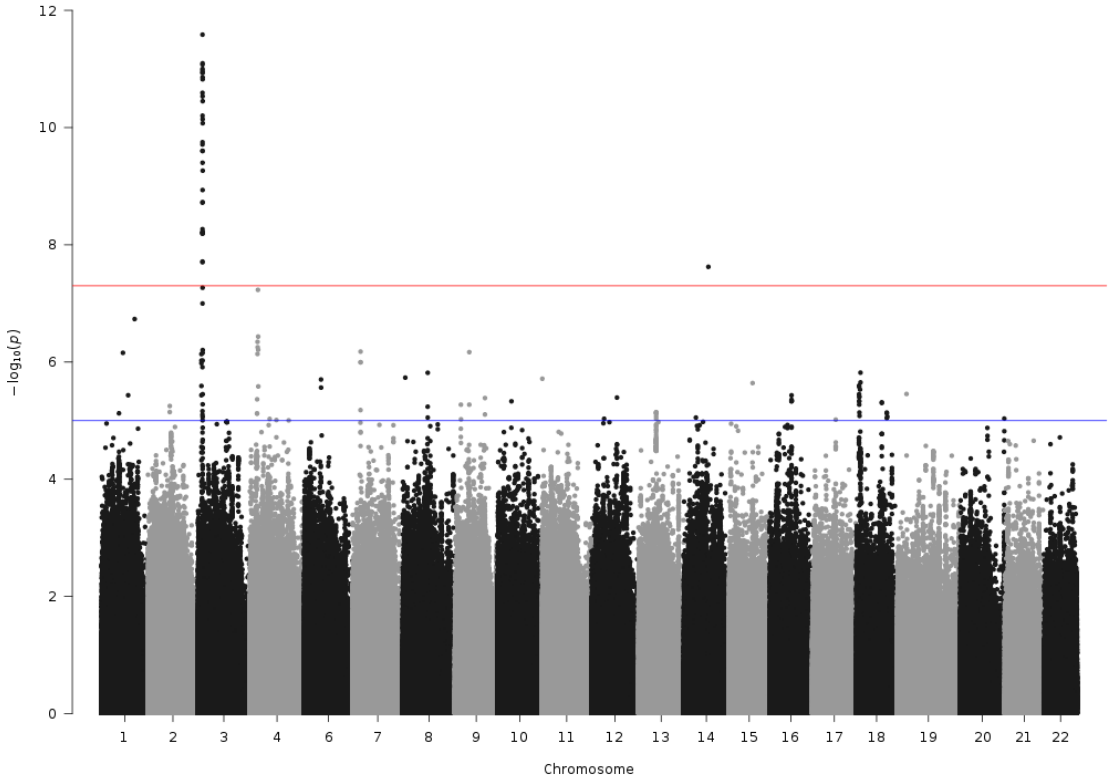


Figure A2.29: *Manhattan Plot of GWAS of IGP28, FG1S1/(FG1+FG1S1)*

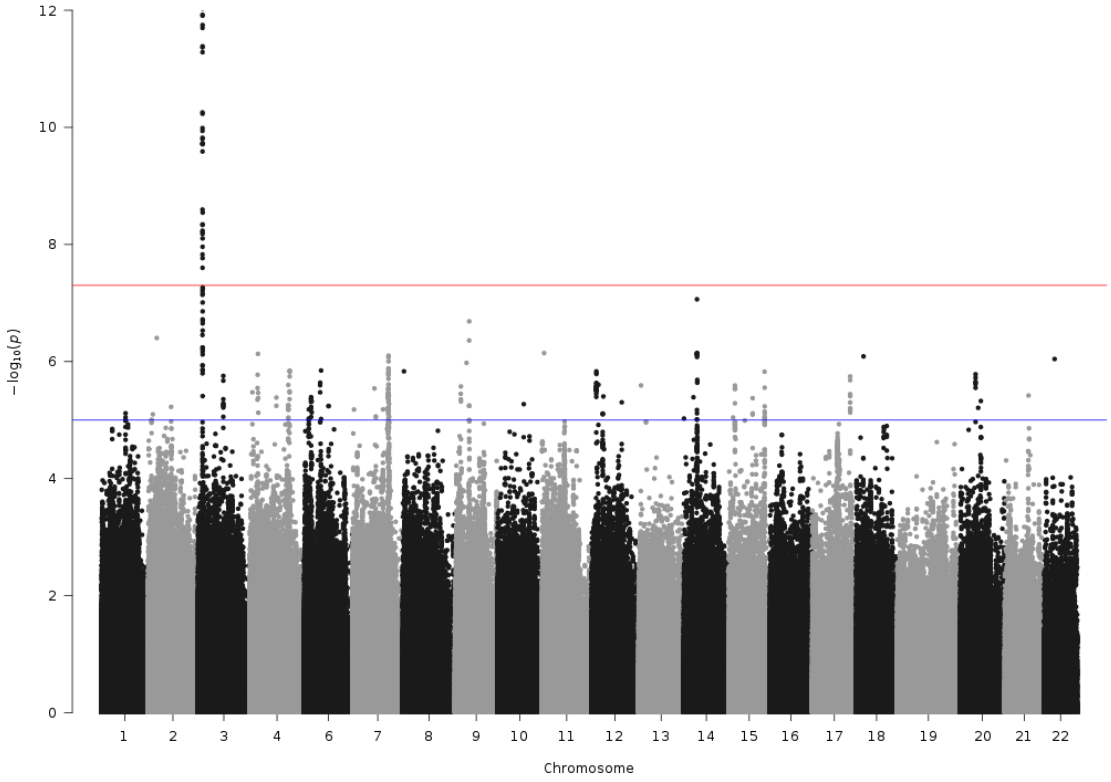


Figure A2.30: *Manhattan Plot of GWAS of IGP29, FG2S1/(FG2+FG2S1+FG2S2)*

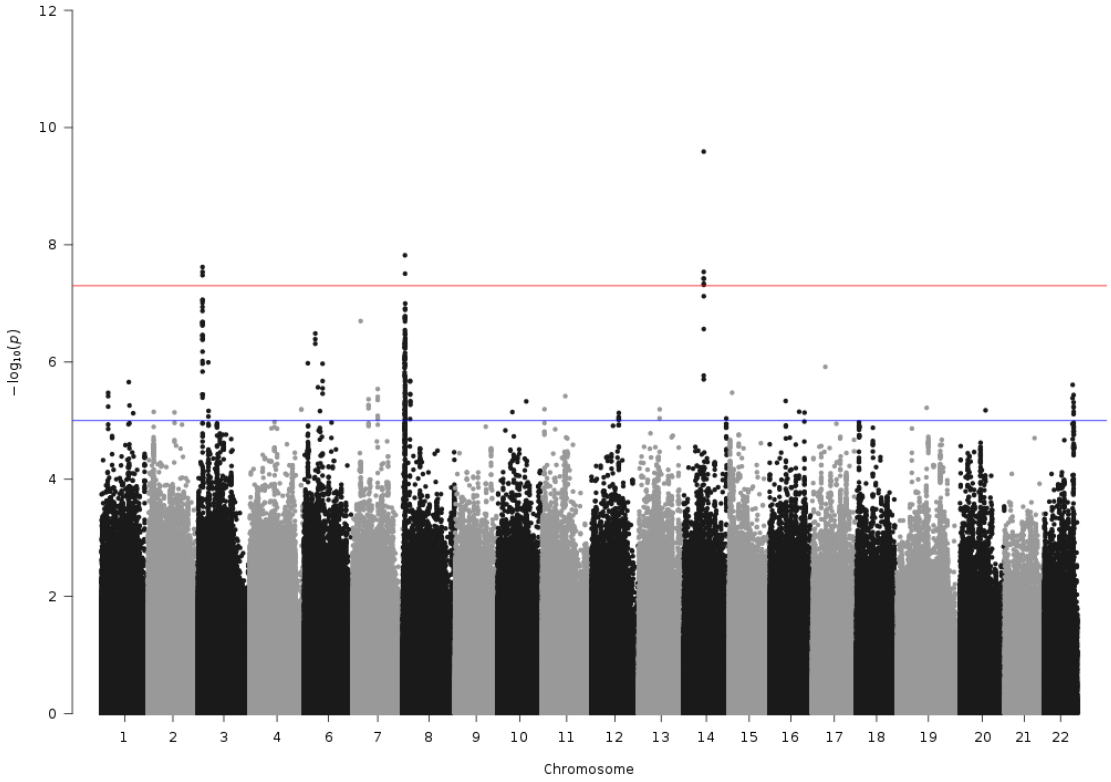


Figure A2.31: *Manhattan Plot of GWAS of IGP30, FG2S2/(FG2+FG2S1+FG2S2)*

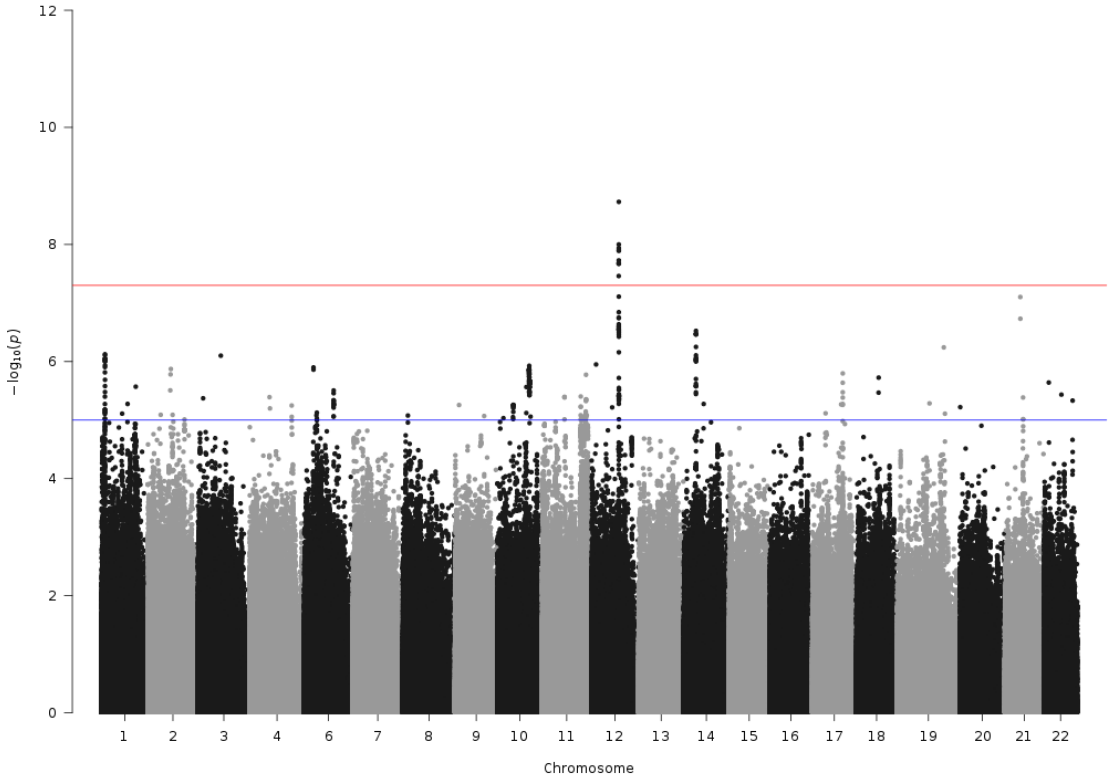


Figure A2.32: *Manhattan Plot of GWAS of IGP31, FBG2S1/(FBG2+FBG2S1+ FBG2S2)*

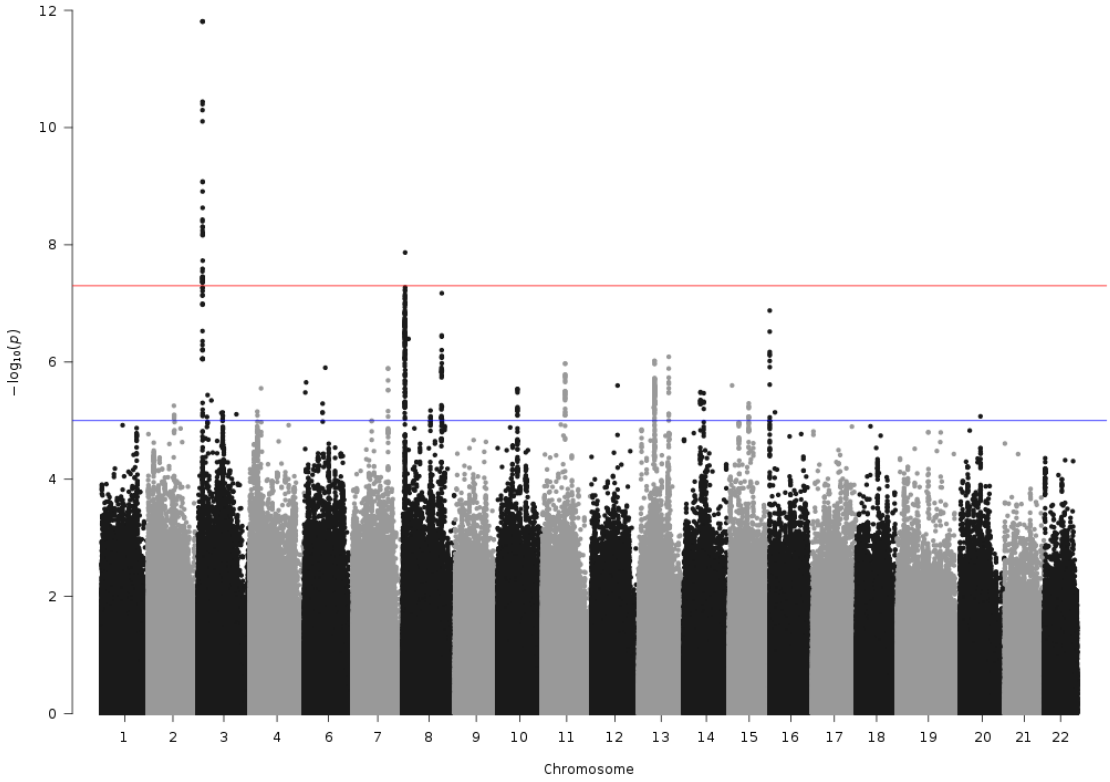


Figure A2.33: *Manhattan Plot of GWAS of IGP32, FBG2S2/(FBG2+FBG2S1+ FBG2S2)*

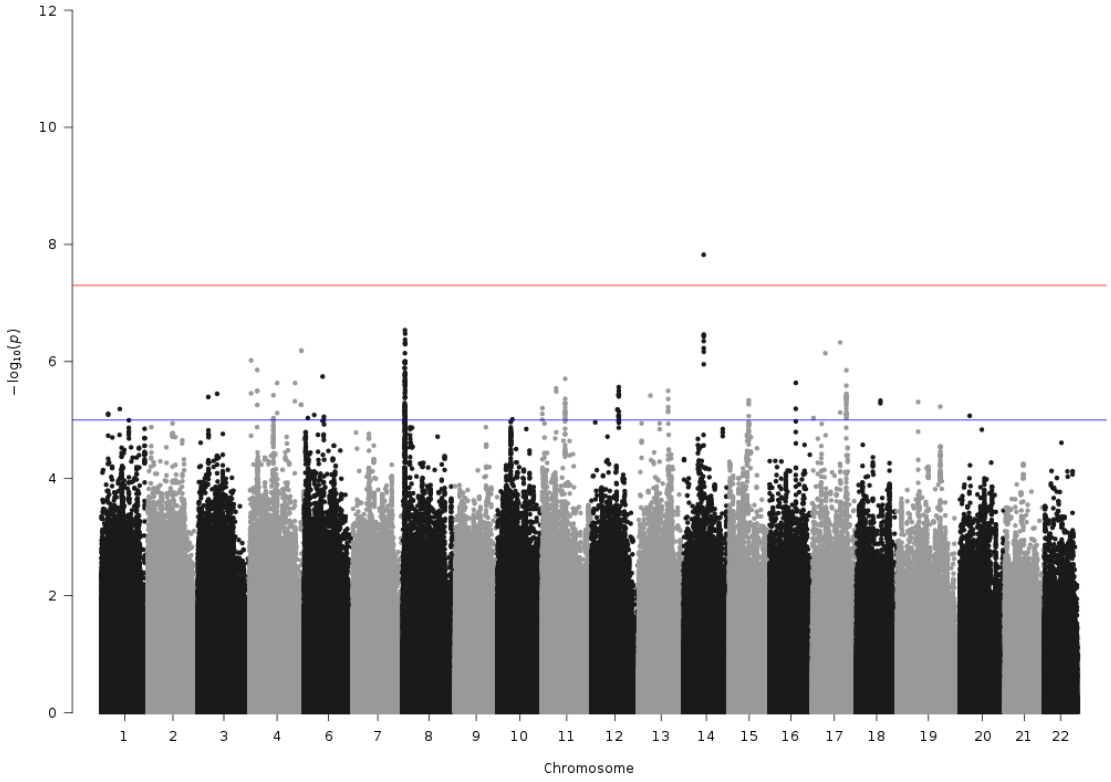


Figure A2.34: *Manhattan Plot of GWAS of IGP33, $F^{total}S1/ F^{total}S2$*

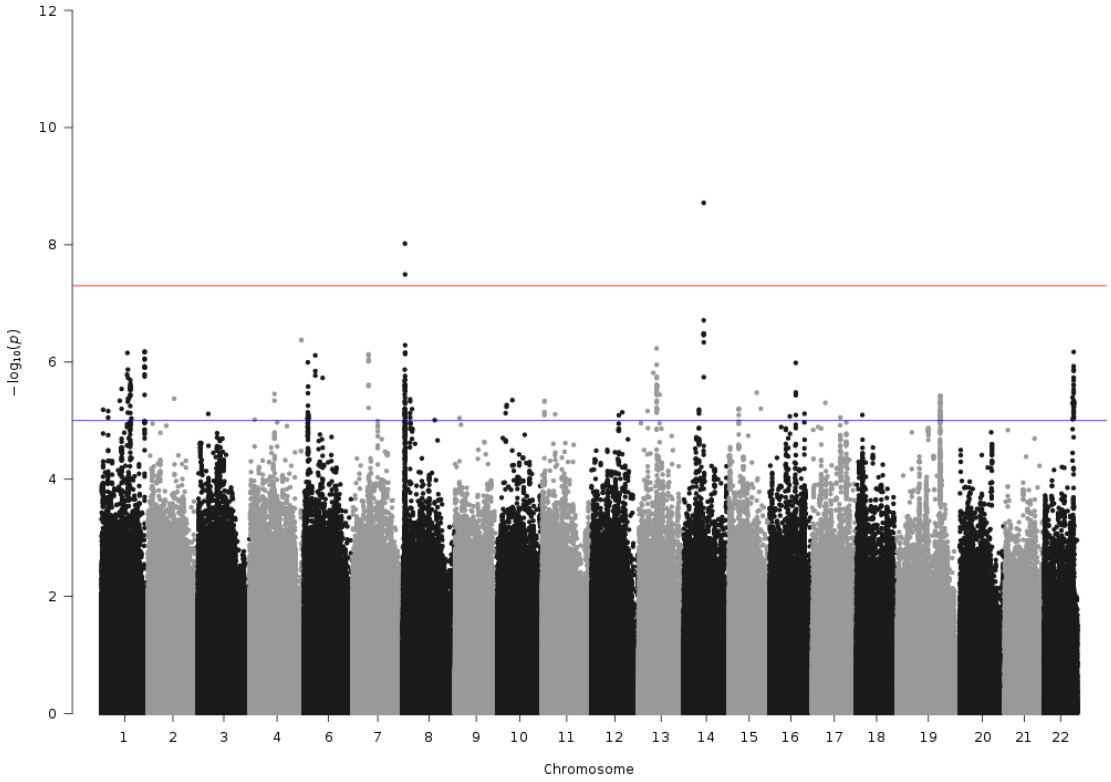


Figure A2.35: *Manhattan Plot of GWAS of IGP34, FS1/ FS2*

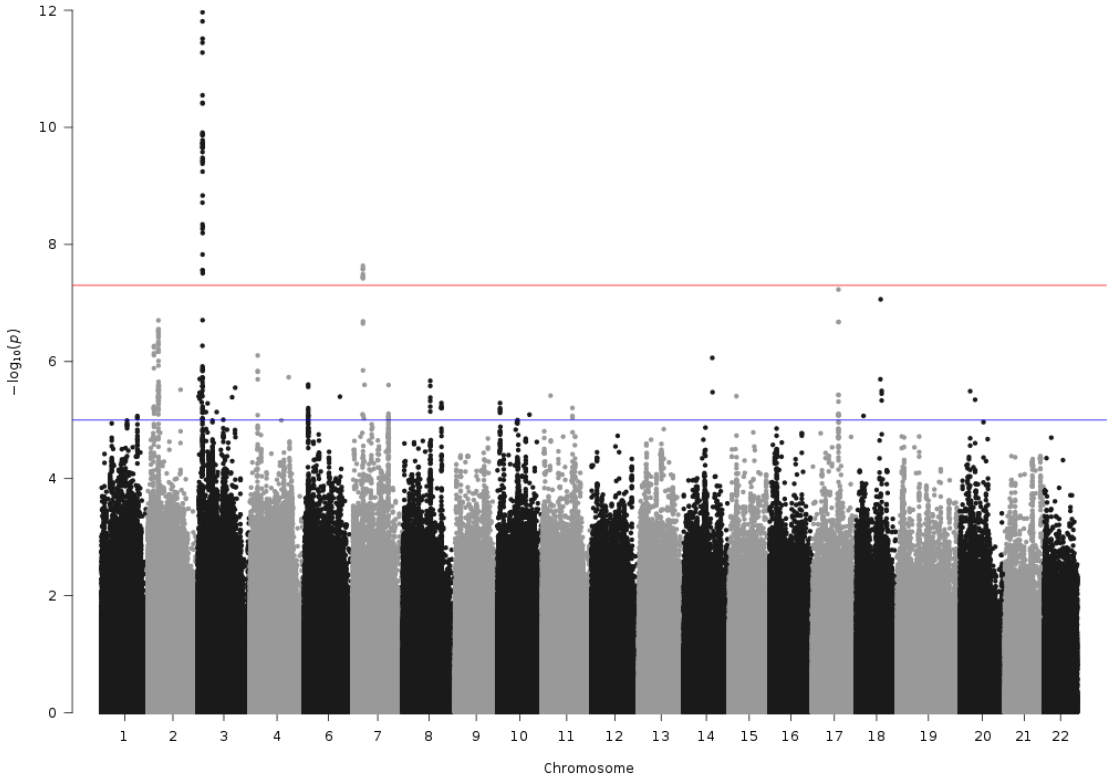


Figure A2.36: *Manhattan Plot of GWAS of IGP35, FBS1/ FBS2*

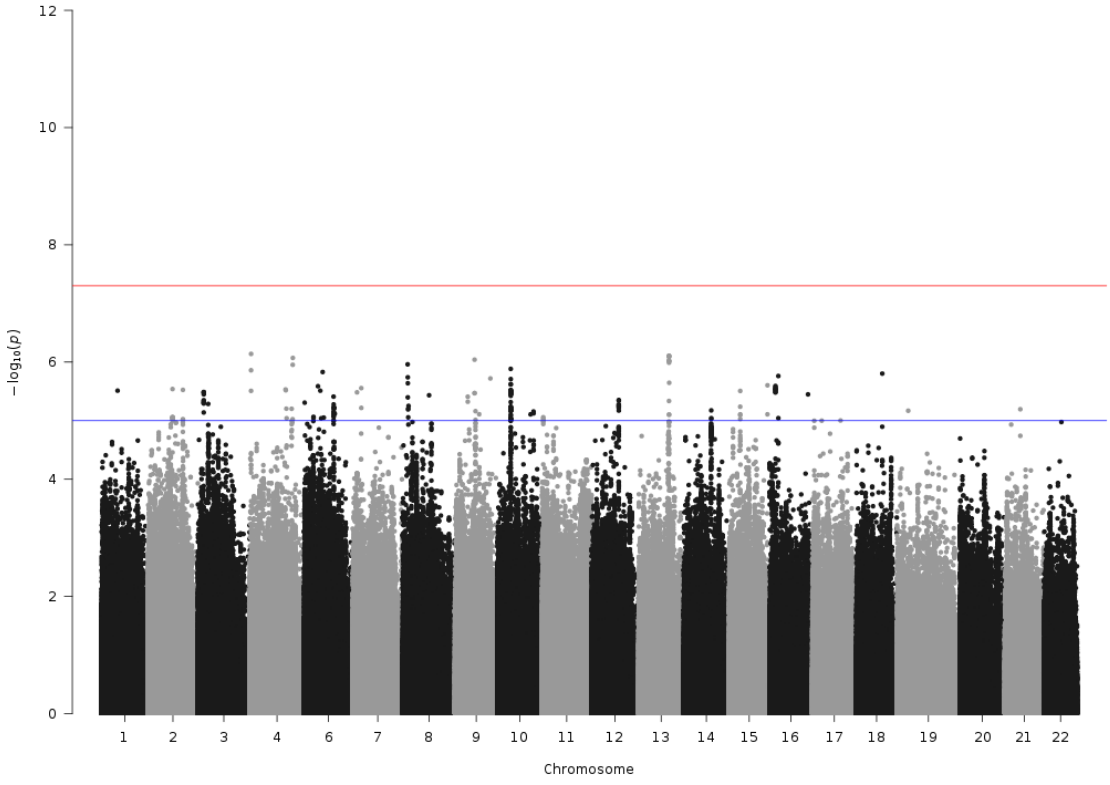


Figure A2.37: *Manhattan Plot of GWAS of IGP36, FBS^{total}/FS^{total}*

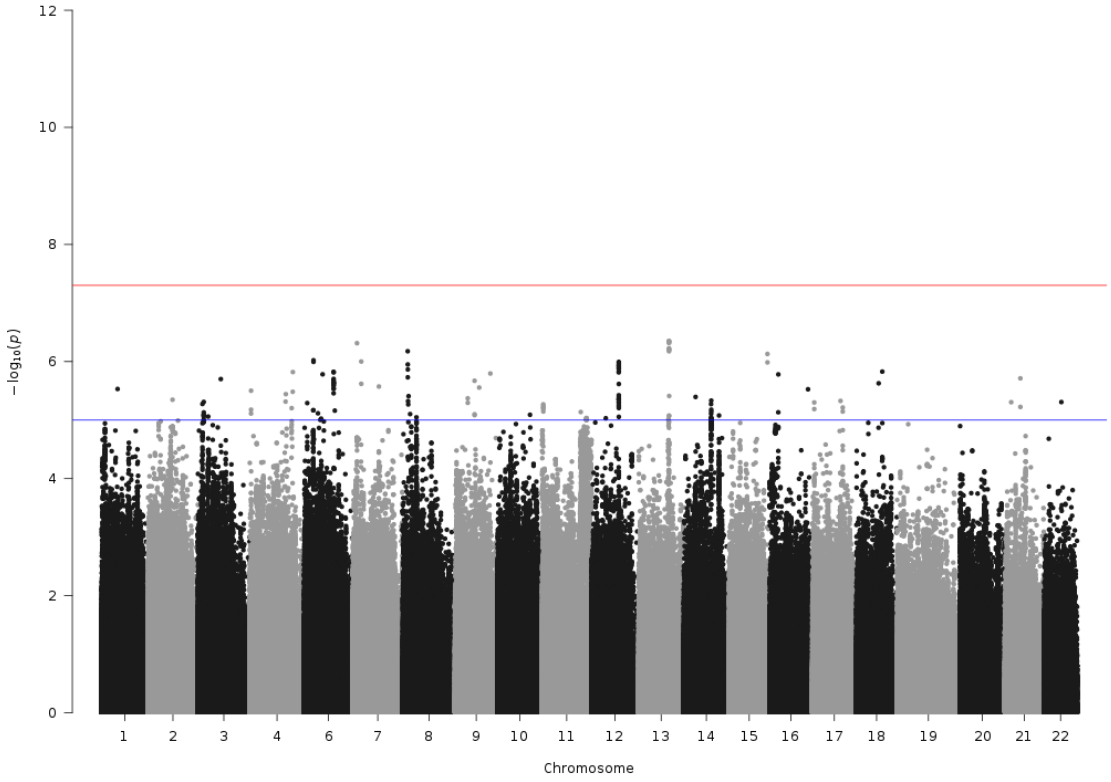


Figure A2.38: *Manhattan Plot of GWAS of IGP38, FBS1/(FS1+FBS1)*

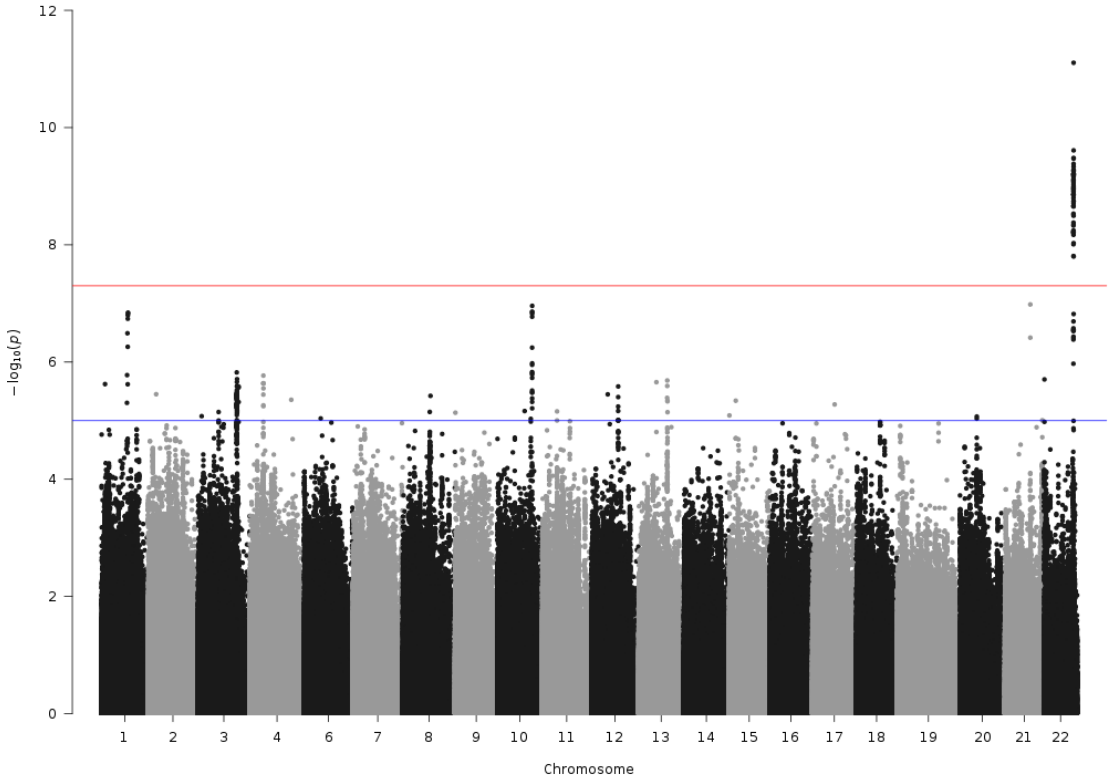


Figure A2.39: *Manhattan Plot of GWAS of IGP40, FBS2/(FS2+FBS2)*

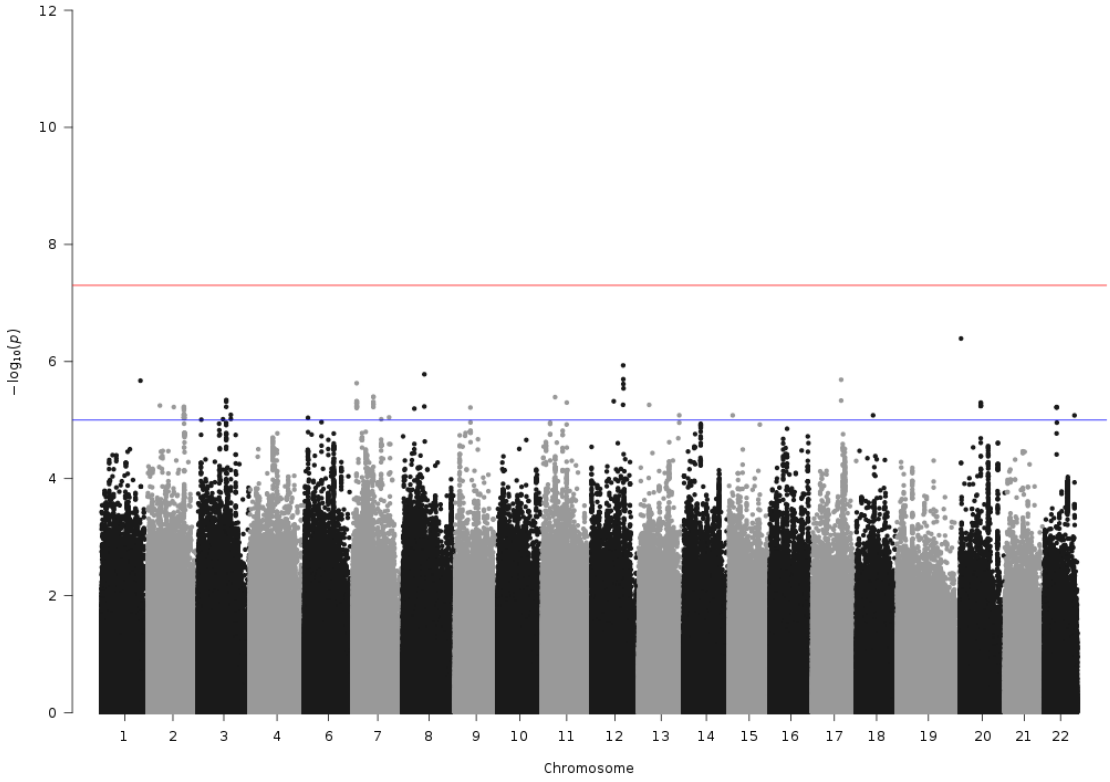


Figure A2.40: *Manhattan Plot of GWAS of IGP55, G0^n*

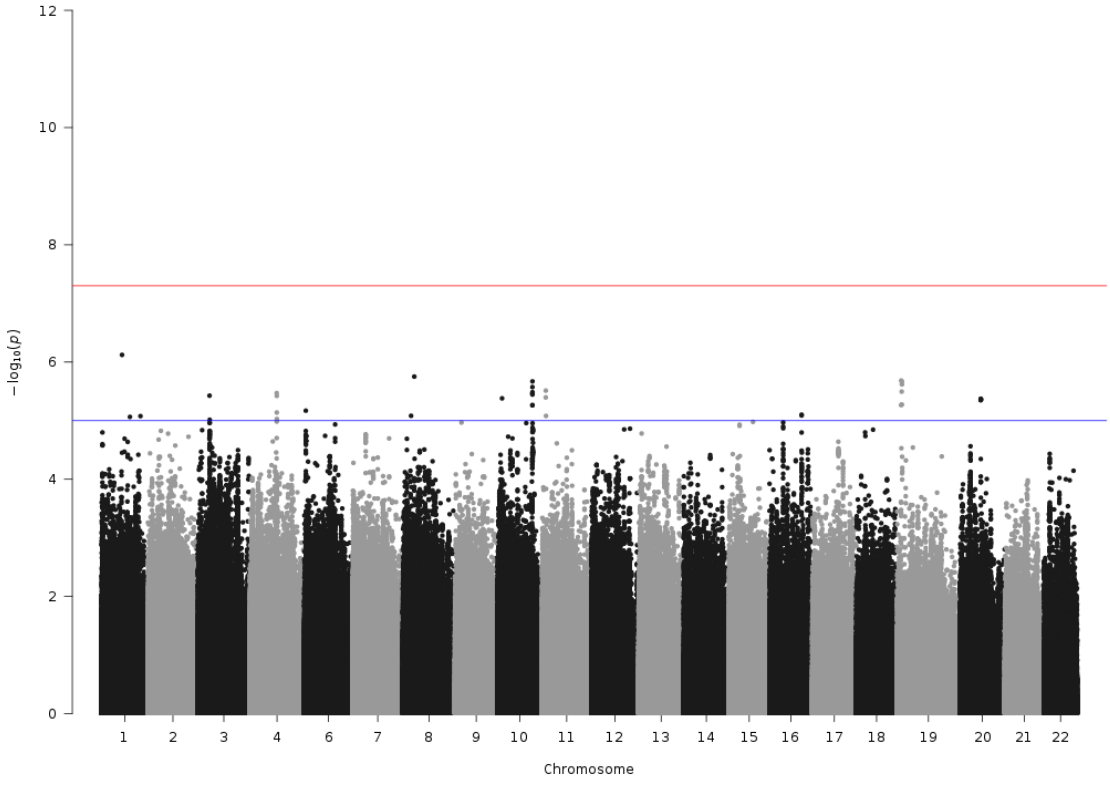


Figure A2.41: *Manhattan Plot of GWAS of IGP56, G1^n*

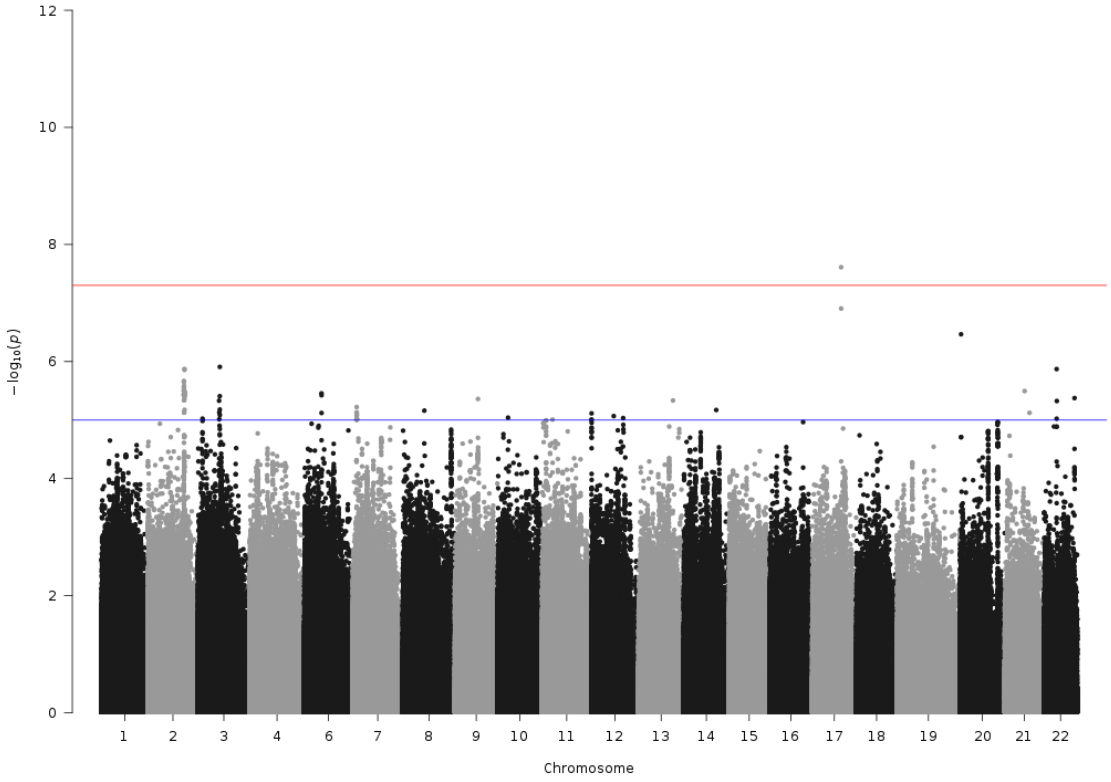


Figure A2.42: *Manhattan Plot of GWAS of IGP57, $G2^n$*

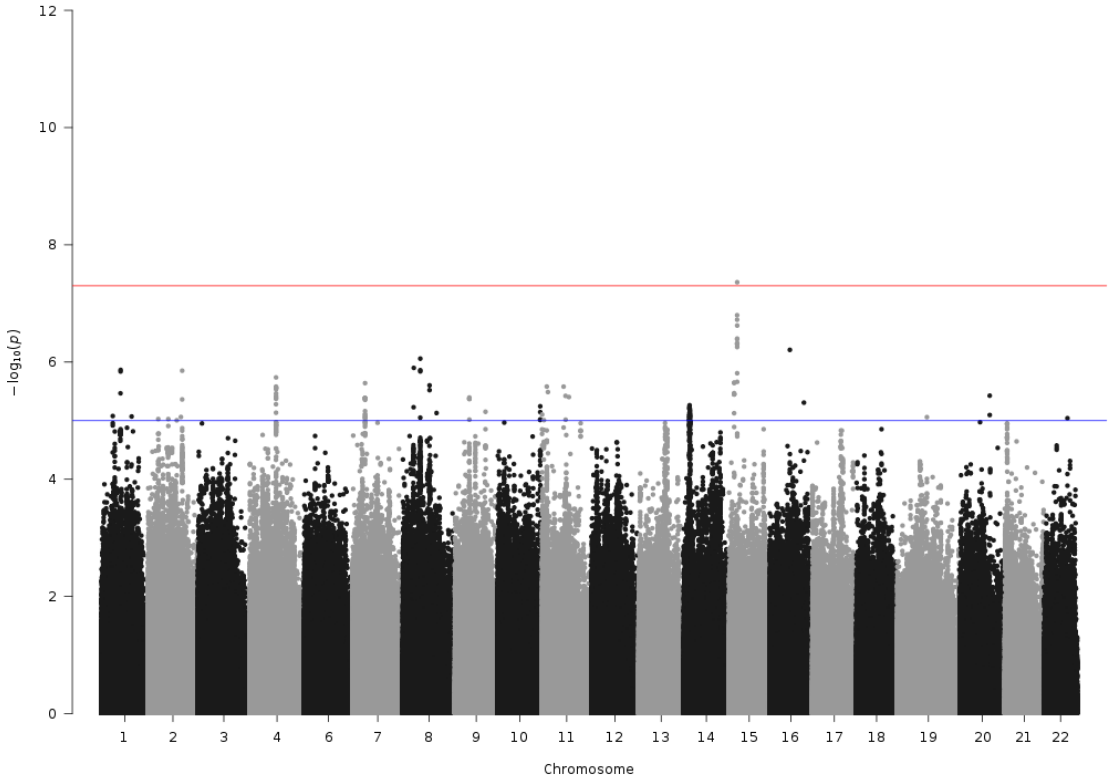


Figure A2.43: *Manhattan Plot of GWAS of IGP58, $F^{n\ total}$*

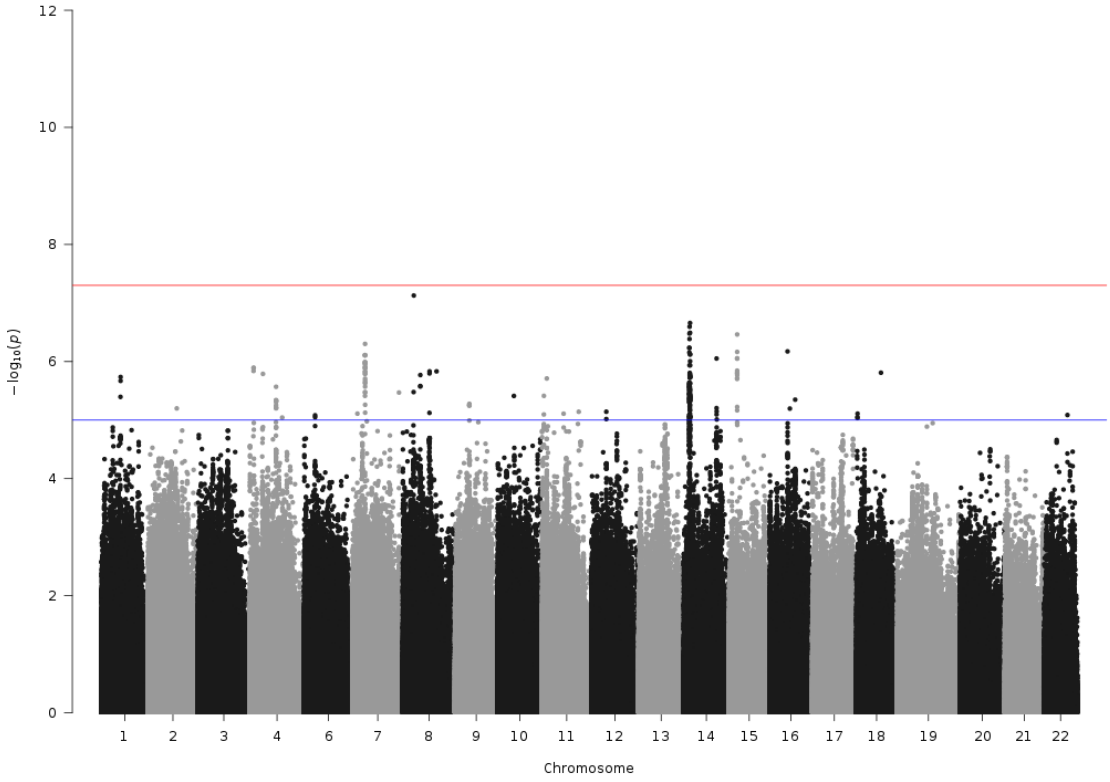


Figure A2.44: *Manhattan Plot of GWAS of IGP59, $FG0^{n\ total}/G0^{n}$*

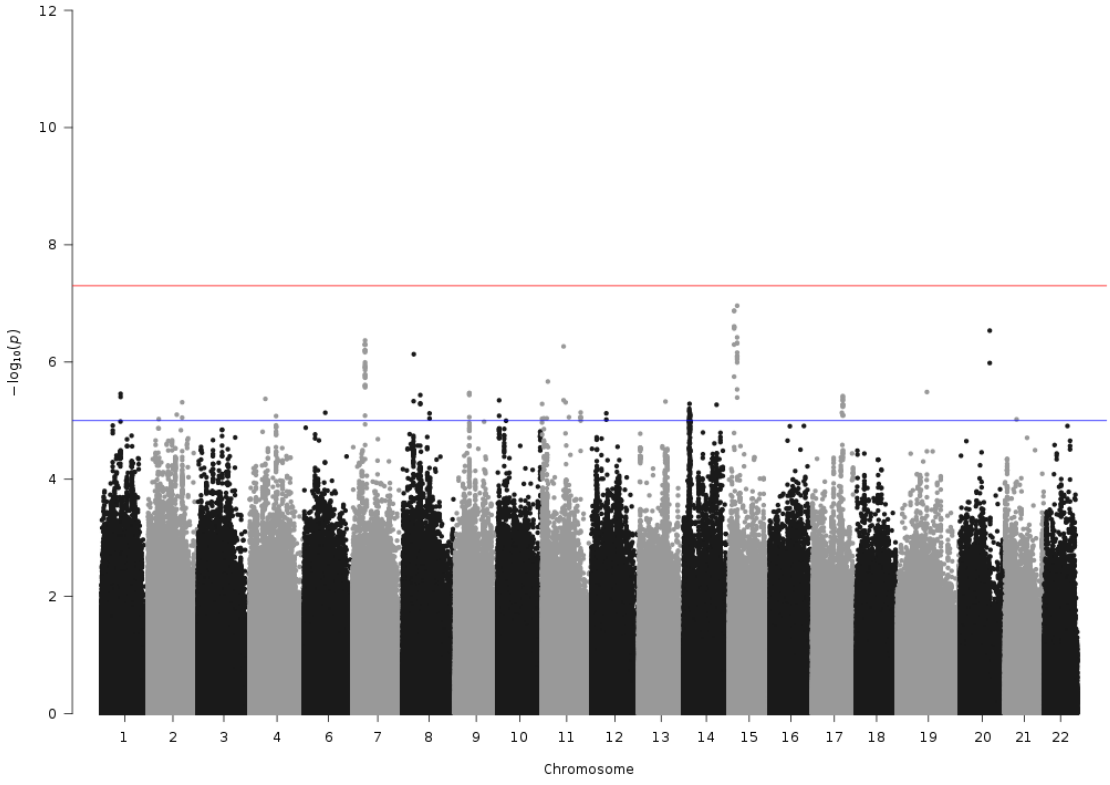


Figure A2.45: *Manhattan Plot of GWAS of IGP60, $FG1^{n\ total}/G1^{n}$*

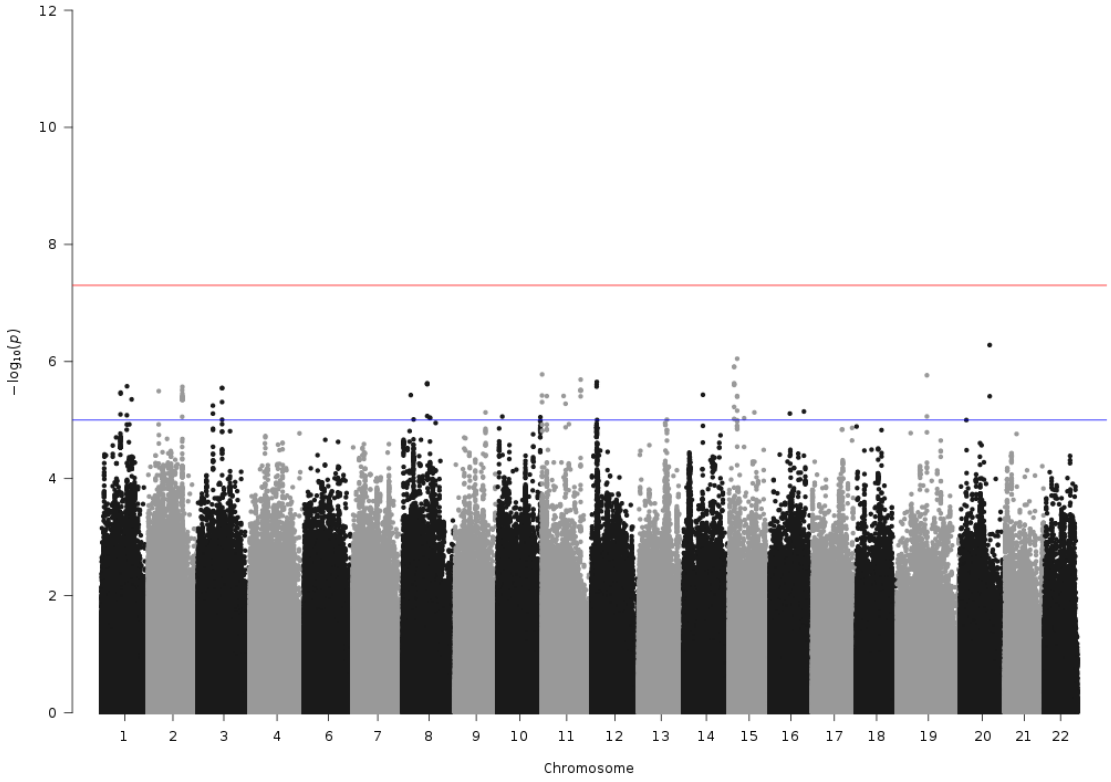


Figure A2.46: *Manhattan Plot of GWAS of IGP61, $FG2^{n\ total}/G2^{n}$*

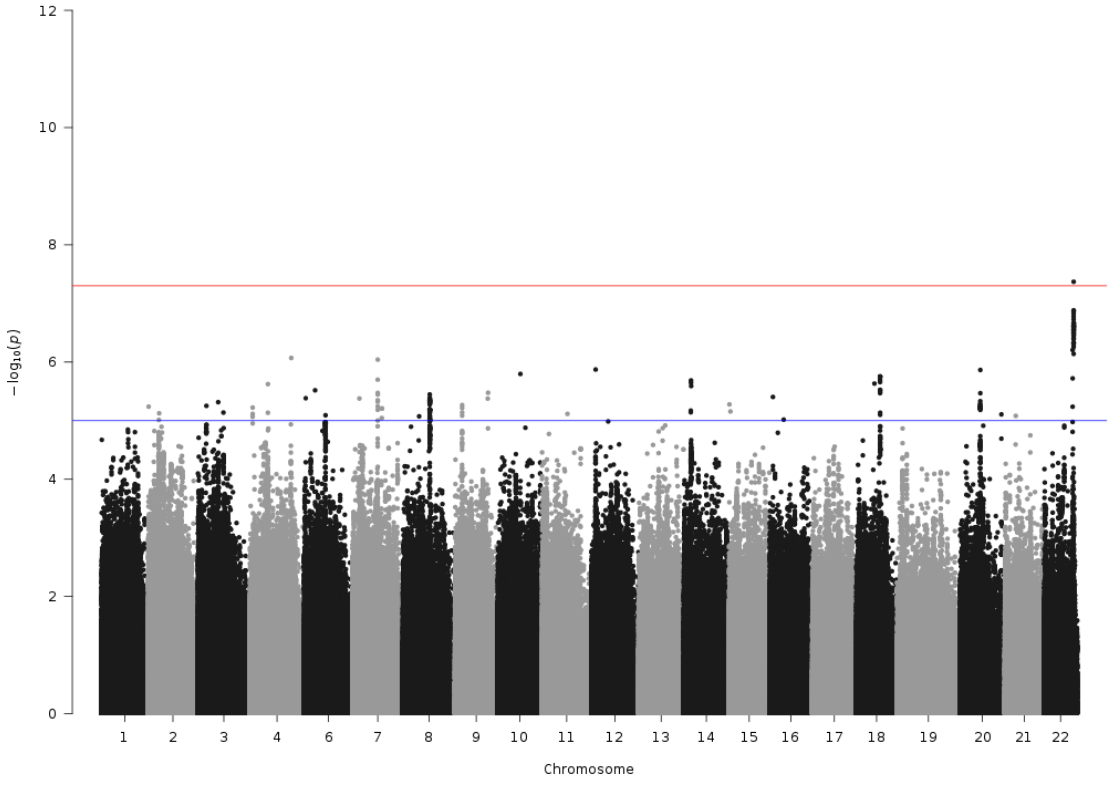


Figure A2.47: *Manhattan Plot of GWAS of IGP62, F^{n}*

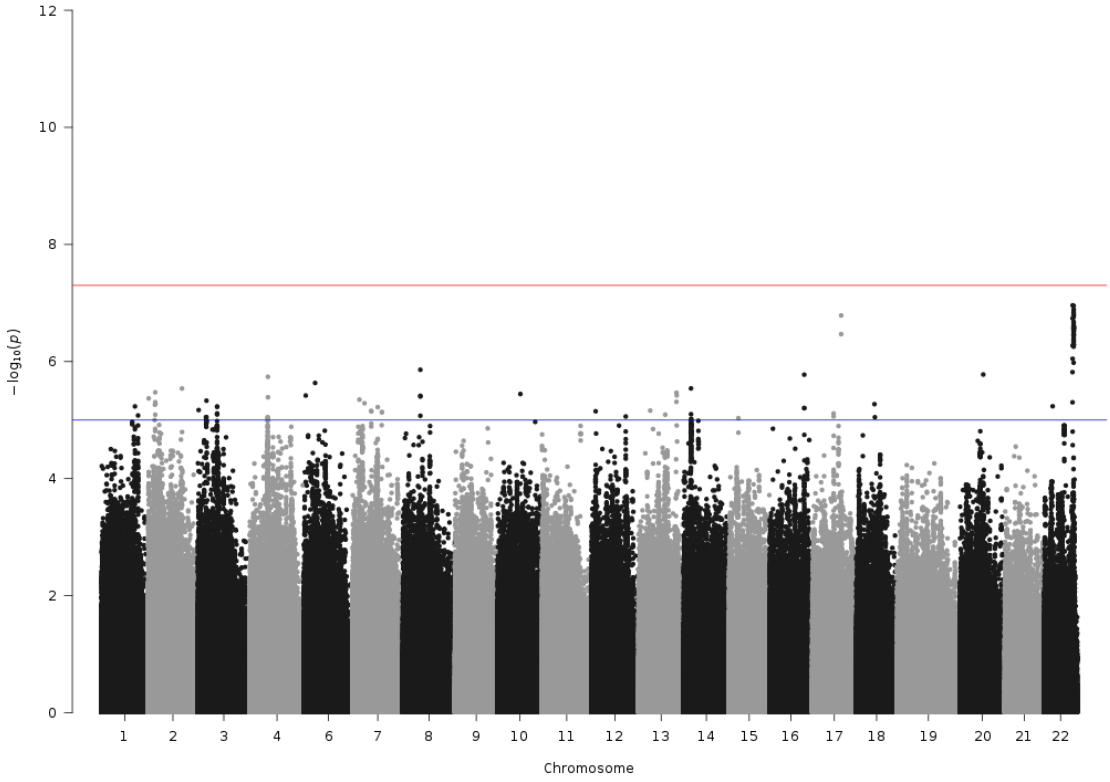


Figure A2.48: *Manhattan Plot of GWAS of IGP63, FG0^{n}/G0^{n}*

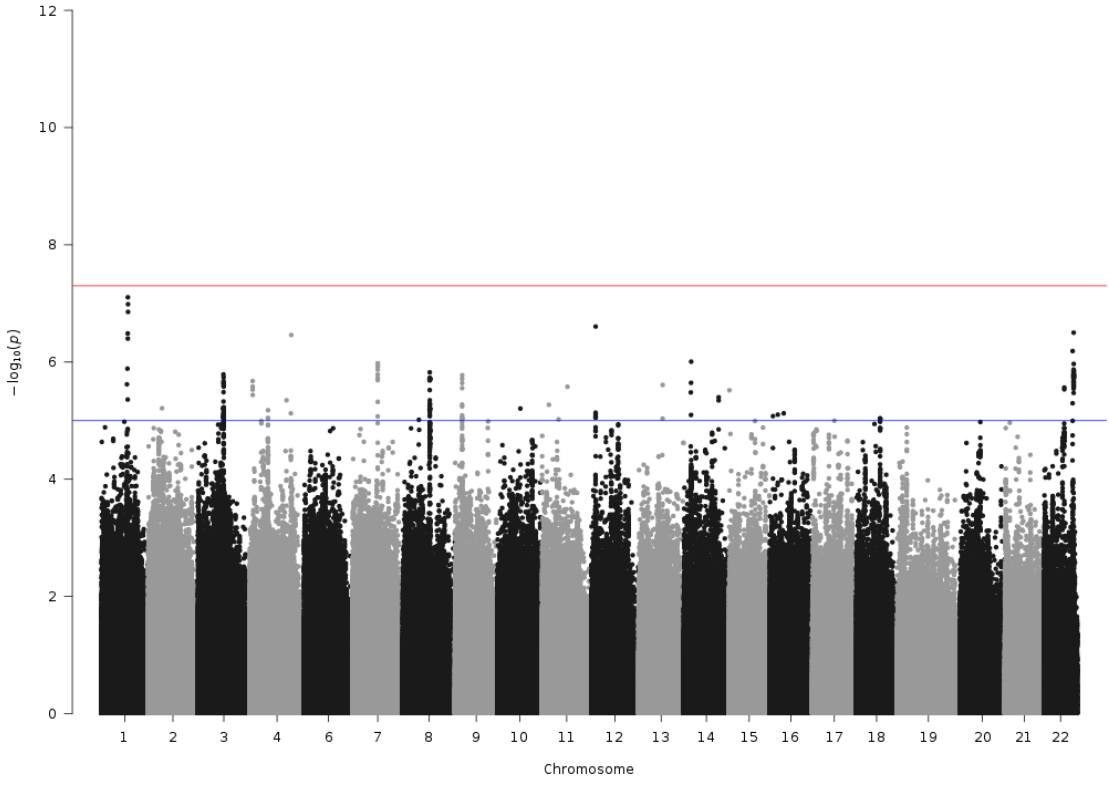


Figure A2.49: *Manhattan Plot of GWAS of IGP64, FG1^{n}/G1^{n}*

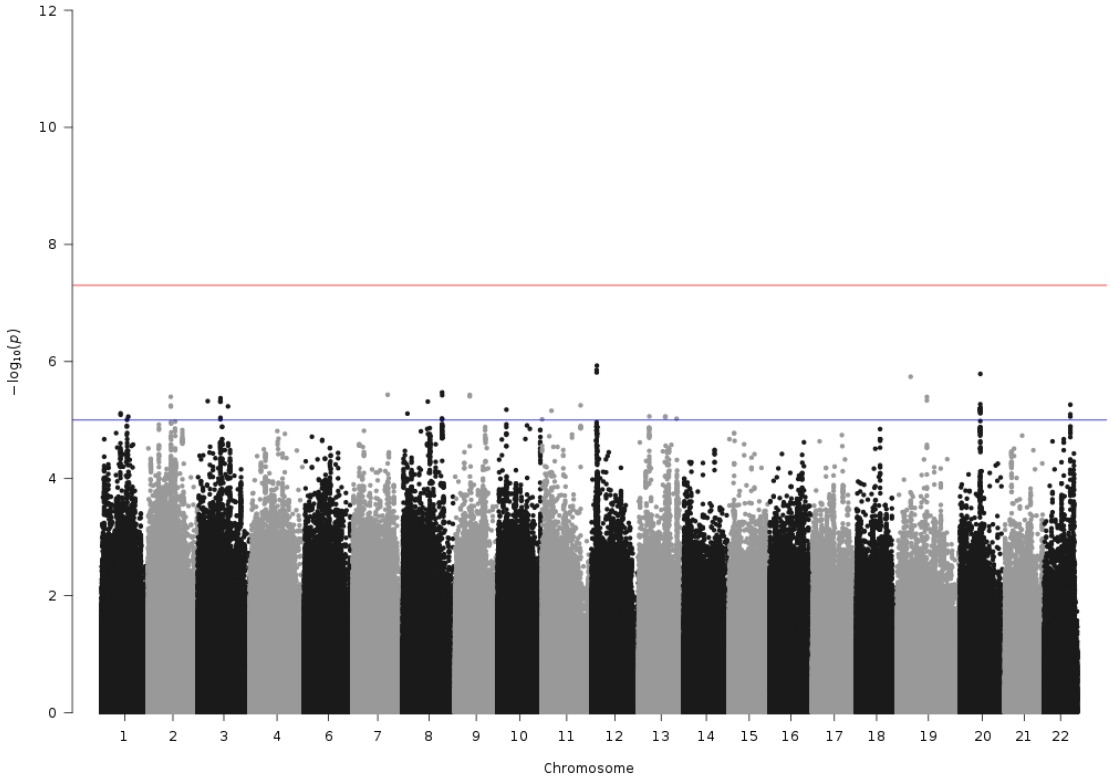


Figure A2.50: *Manhattan Plot of GWAS of IGP65, $FG2^n/G2^n$*

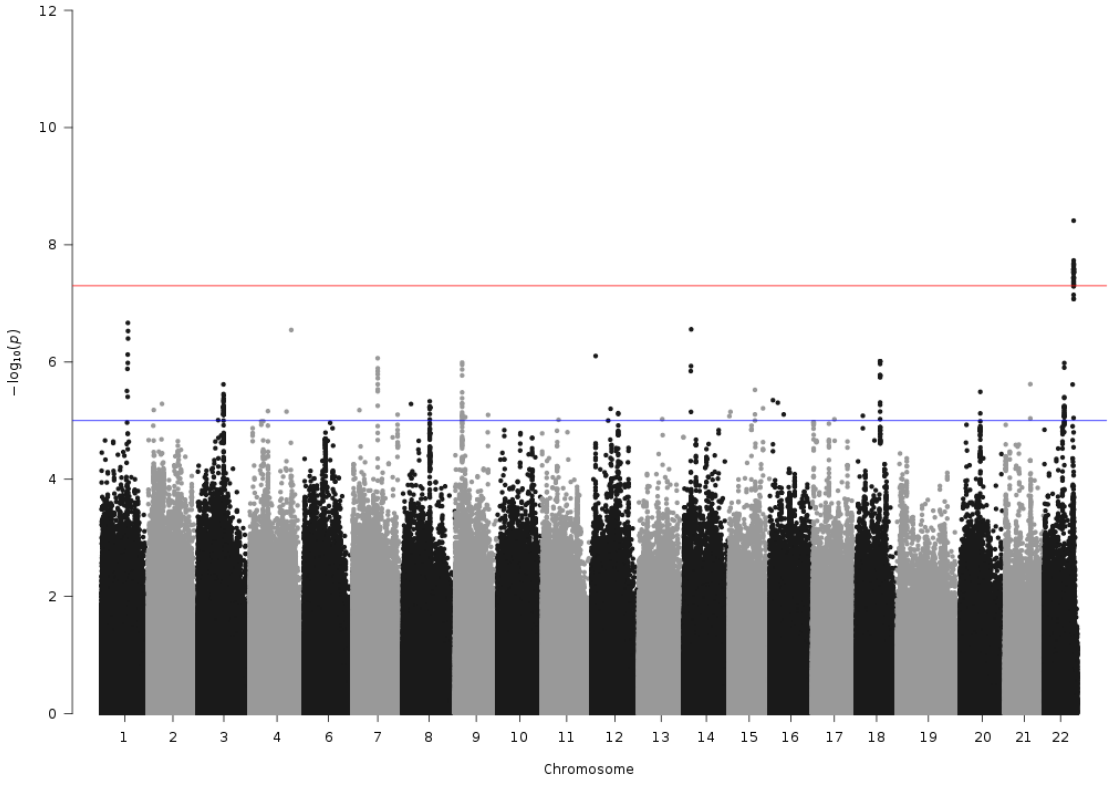


Figure A2.51: *Manhattan Plot of GWAS of IGP66, FB^n*

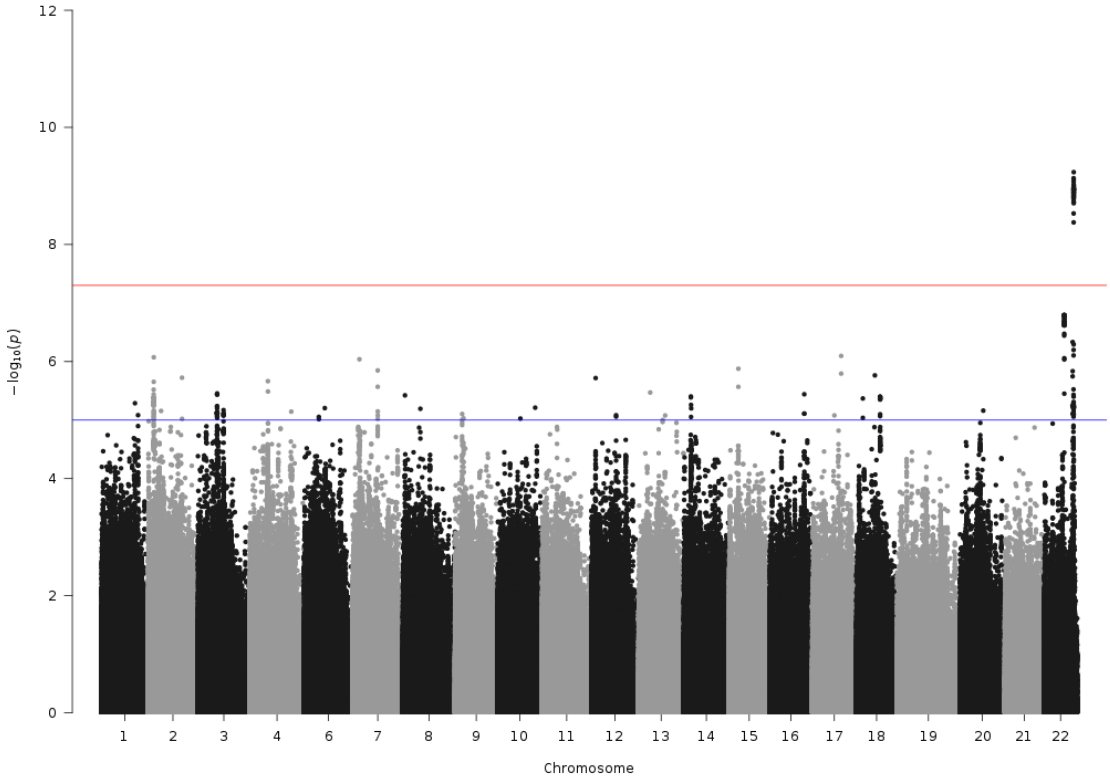


Figure A2.52: *Manhattan Plot of GWAS of IGP67, FBG0^n/G0^n*

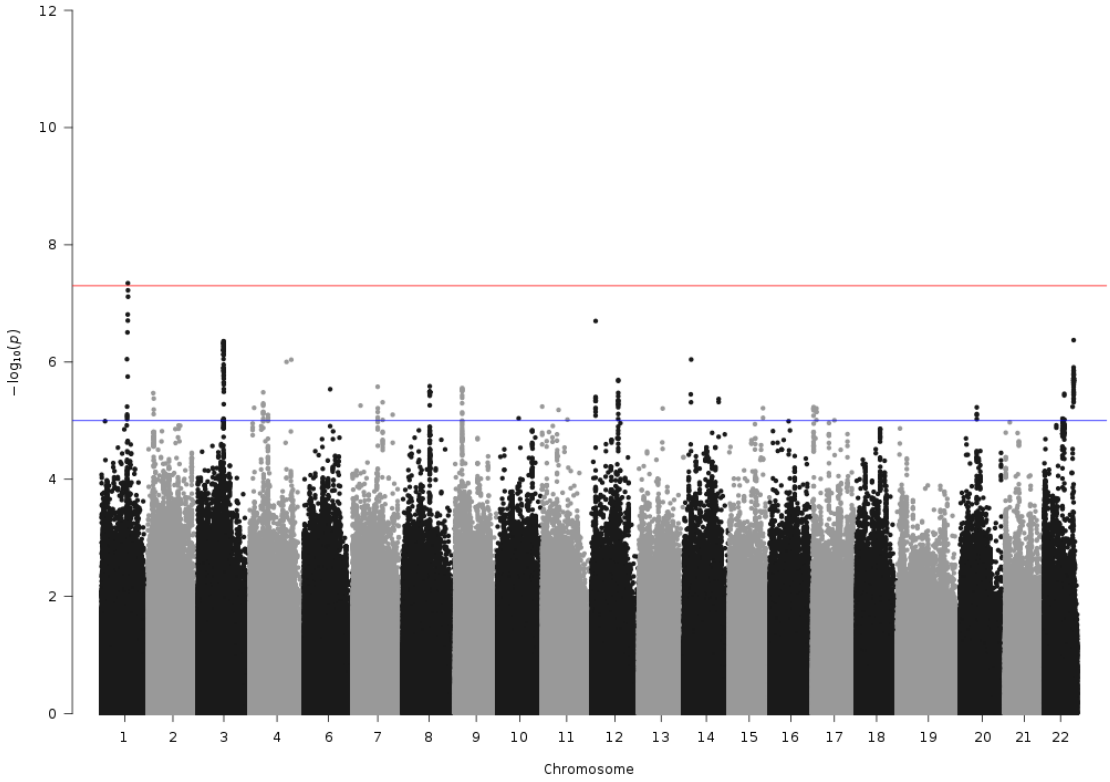


Figure A2.53: *Manhattan Plot of GWAS of IGP68, FBG1^n/G1^n*

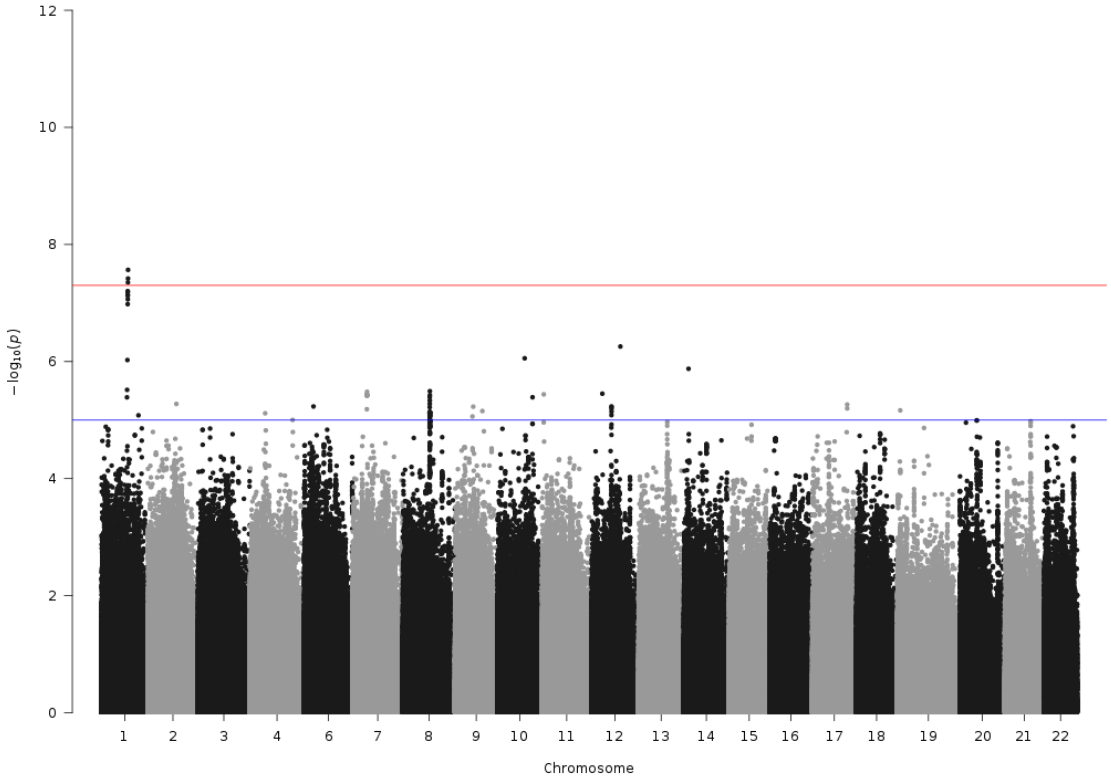


Figure A2.54: *Manhattan Plot of GWAS of IGP69, FBG2^n/G2^n*

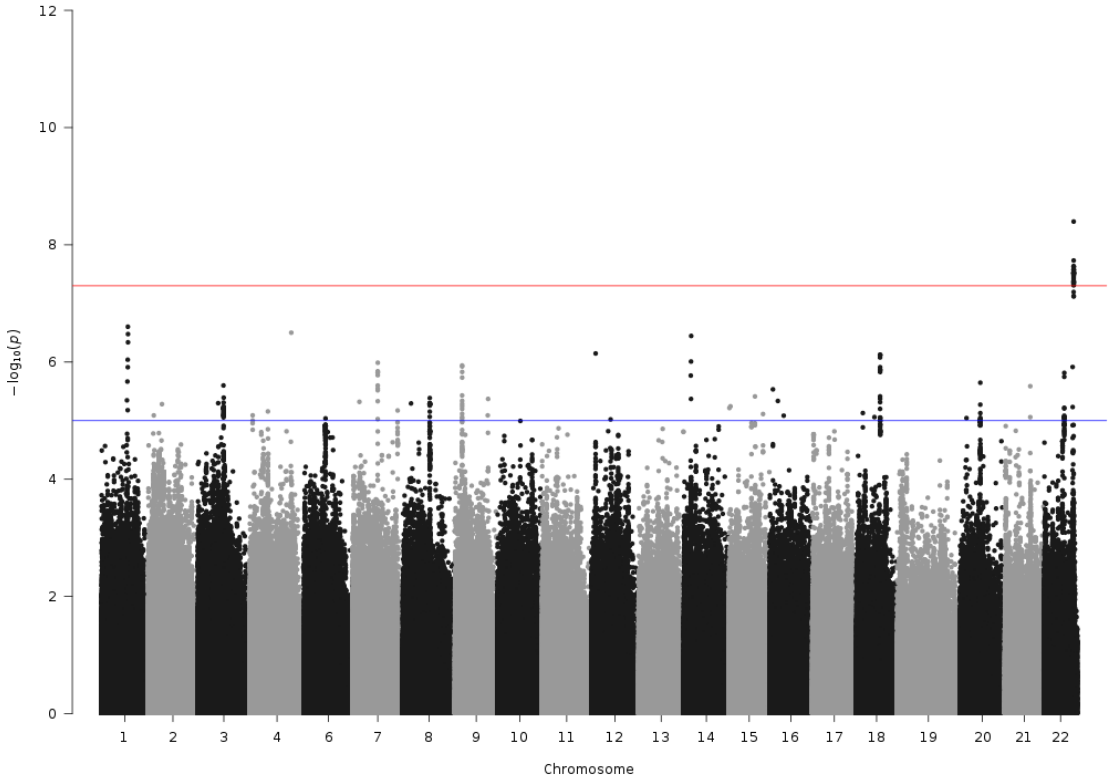


Figure A2.55: *Manhattan Plot of GWAS of IGP70, FB^n/F^n*

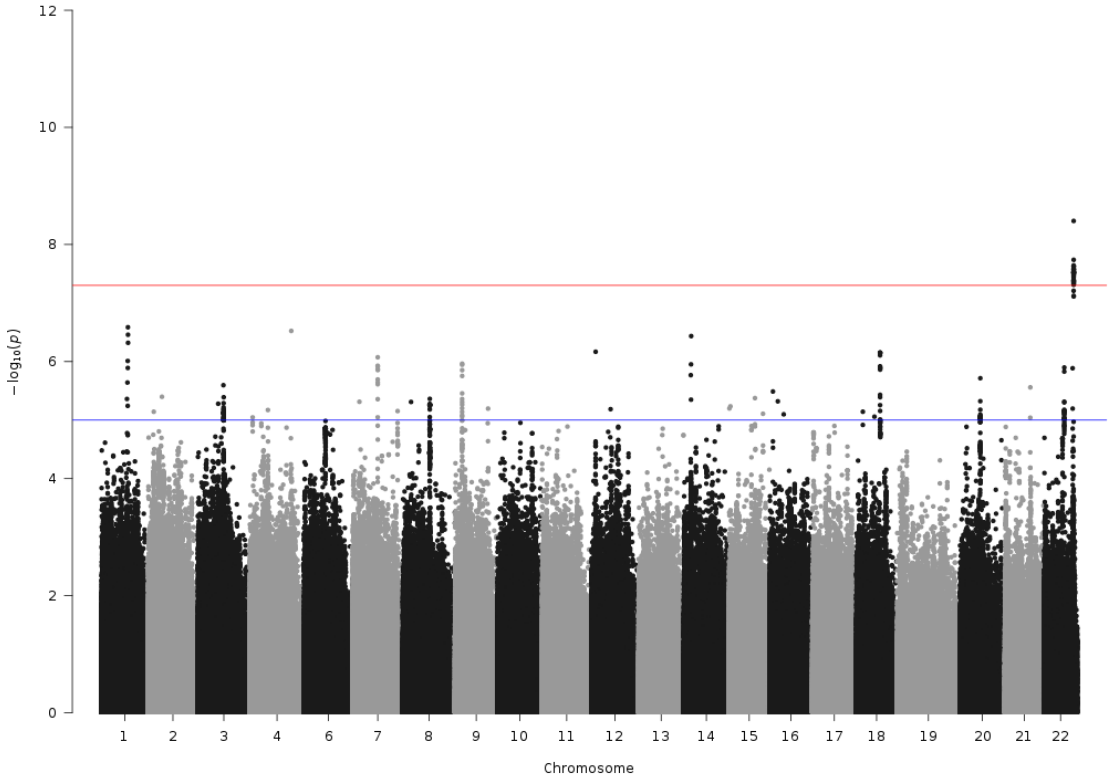


Figure A2.56: *Manhattan Plot of GWAS of IGP71, $FB^n/F^{n\ total}$*

Page intentionally left blank

Appendix 3 – Tables of Differentially Expressed Genes

Full list of identified differentially expressed genes associated with GP6, GP10 and GP23, all returning significant effects after adjusting for the false discovery rate (**Chapter 4**).

***Table A3.1**: Identified differentially expressed genes of IgG GP6*

Gene	Effect	S.E.	t	*p*-value	Adj. *p*-value
CRTAP	-1.06373	0.215326	-4.94009	1.30E-06	0.018544

***Table A3.2**: Identified differentially expressed genes of IgG GP10*

Gene	Effect	S.E.	t	*p*-value	Adj. *p*-value
IGKV1-6	-0.28498	0.055322	-5.15135	4.70E-07	0.002989
IGLL5	-0.29276	0.057335	-5.10611	5.86E-07	0.002989
IGLC1	-0.28236	0.055457	-5.09156	6.29E-07	0.002989
IGHG2	-0.26084	0.05345	-4.88011	1.73E-06	0.005454
MZB1	-0.3696	0.076436	-4.83546	2.13E-06	0.005454
IGLJ1	-0.28435	0.059005	-4.81912	2.30E-06	0.005454
IGKC	-0.31056	0.066208	-4.6906	4.15E-06	0.007945
IGKV1-5	-0.28663	0.061611	-4.65219	4.94E-06	0.007945
IGKV1-9	-0.26744	0.057933	-4.61633	5.81E-06	0.007945
IGLV2-18	-0.21286	0.046389	-4.58855	6.58E-06	0.007945
IGKV3D-20	-0.23662	0.05164	-4.58204	6.77E-06	0.007945
ANK3	0.471086	0.102919	4.57727	6.92E-06	0.007945
IGKJ1	-0.33714	0.073826	-4.56669	7.25E-06	0.007945
FA2H	-0.26649	0.05881	-4.53131	8.48E-06	0.008455
TMEM53	-0.70521	0.156004	-4.52043	8.90E-06	0.008455
IGKV1-16	-0.2725	0.060552	-4.50033	9.73E-06	0.008662
IGKV3-15	-0.28499	0.064064	-4.44851	1.22E-05	0.010045
GLDC	-0.18584	0.041859	-4.43964	1.27E-05	0.010045
TPPP3	-0.41439	0.093972	-4.40978	1.45E-05	0.010835
IGKJ4	-0.27624	0.063424	-4.35548	1.83E-05	0.012589
SRRM1	1.222925	0.281018	4.351771	1.86E-05	0.012589
TXNDC5	-0.4101	0.09473	-4.32912	2.04E-05	0.013203
BLOC1S5-TXNDC5	-0.42064	0.097483	-4.31494	2.17E-05	0.013203
IGKV2-24	-0.24405	0.056633	-4.30938	2.22E-05	0.013203
IGKV4-1	-0.29276	0.068133	-4.29691	2.35E-05	0.013365
SDC1	-0.18163	0.04257	-4.26662	2.67E-05	0.013451
CCDC189	-0.77132	0.180813	-4.26585	2.67E-05	0.013451

***Table A3.2**: cont.*

Gene	Effect	S.E.	t	*p*-value	Adj. *p*-value
IGLC7	-0.21475	0.050378	-4.26285	2.71E-05	0.013451
CSRP1	-1.24533	0.292967	-4.25076	2.85E-05	0.013451
C20orf96	-0.67106	0.157993	-4.24739	2.89E-05	0.013451
IGKV3-7	-0.24114	0.056813	-4.24448	2.93E-05	0.013451
PTPMT1	-0.90152	0.214198	-4.20881	3.40E-05	0.014728
S100A10	-0.74107	0.176113	-4.20792	3.41E-05	0.014728
IGLV1-51	-0.25181	0.059972	-4.19884	3.54E-05	0.014846
IGLV1-44	-0.22852	0.054723	-4.17602	3.90E-05	0.015858
IGKV3-20	-0.27704	0.066679	-4.15482	4.25E-05	0.016833
IGLV2-11	-0.23126	0.055829	-4.14239	4.48E-05	0.016867
COA3	-0.76871	0.185854	-4.13608	4.60E-05	0.016867
ZDHHC1	-0.45169	0.109383	-4.12948	4.72E-05	0.016867
IGLV1-40	-0.26033	0.063052	-4.12878	4.74E-05	0.016867
IGKV3-11	-0.26956	0.065521	-4.11413	5.03E-05	0.017072
MPDU1	-0.83907	0.204487	-4.10327	5.26E-05	0.017072
IGKV1-17	-0.23783	0.057964	-4.10305	5.26E-05	0.017072
URB2	0.793813	0.19349	4.102604	5.27E-05	0.017072
WRN	0.654725	0.159895	4.094726	5.45E-05	0.017239
HJURP	-0.30476	0.074601	-4.08516	5.66E-05	0.017536
IGKV2-30	-0.25517	0.062655	-4.0726	5.96E-05	0.018063
IGHA1	-0.19133	0.047679	-4.01281	7.59E-05	0.022525
CDKN2C	-0.64083	0.160491	-3.99296	8.22E-05	0.023894
IGHV3-49	-0.24706	0.062375	-3.96091	9.34E-05	0.026612
GPRC5D	-0.16236	0.041182	-3.94254	0.0001	0.027374
LRRN3	0.211603	0.053676	3.942235	0.000101	0.027374
PPIB	-0.73568	0.187225	-3.9294	0.000106	0.027374
IGKJ2	-0.2791	0.071188	-3.92065	0.00011	0.027374
NF1	0.659968	0.168428	3.91839	0.000111	0.027374
HMGN2P5	-0.77014	0.196689	-3.91553	0.000112	0.027374
AC096579.15	-0.27929	0.07137	-3.91333	0.000113	0.027374
JCHAIN	-0.19897	0.050882	-3.91051	0.000114	0.027374
IGLV2-23	-0.24703	0.06319	-3.90927	0.000115	0.027374
AP1S1	-0.78342	0.200482	-3.9077	0.000115	0.027374
UTP20	0.644352	0.165517	3.892963	0.000122	0.02853
IGKV3D-15	-0.16237	0.041994	-3.86655	0.000135	0.030928
COX5A	-0.7086	0.18353	-3.86096	0.000138	0.030928
IGHV4-34	-0.25111	0.065053	-3.86005	0.000139	0.030928
IGKV1-12	-0.21717	0.056356	-3.85357	0.000142	0.031183
TMED3	-0.77031	0.200159	-3.84847	0.000145	0.031183
LY75	0.456558	0.118705	3.846163	0.000147	0.031183
CDT1	-0.31852	0.082926	-3.84105	0.00015	0.031249
IGHG4	-0.14274	0.037288	-3.8281	0.000157	0.031249
MT1F	-0.48312	0.126229	-3.82733	0.000158	0.031249

***Table A3.2**: cont.*

Gene	Effect	S.E.	t	*p*-value	Adj. *p*-value
IGHGP	-0.17559	0.04589	-3.82634	0.000158	0.031249
EHBP1	0.841203	0.21986	3.826089	0.000159	0.031249
BMP8B	-0.40961	0.10713	-3.8235	0.00016	0.031249
SLC50A1	-1.03257	0.270961	-3.81078	0.000168	0.03144
ZDHHC4	-0.98614	0.259024	-3.80713	0.000171	0.03144
C11orf98	-0.75537	0.198427	-3.80678	0.000171	0.03144
OVOL2	-0.90822	0.238664	-3.80545	0.000172	0.03144
RP11-405M12.4	-0.79526	0.209018	-3.80476	0.000172	0.03144
IGHA2	-0.15604	0.041106	-3.79605	0.000178	0.031703
TK1	-0.32789	0.086378	-3.79601	0.000178	0.031703
PTTG1	-0.4351	0.114816	-3.78958	0.000182	0.032095
FOXM1	-0.30322	0.080254	-3.77825	0.000191	0.032962
IGKV1-27	-0.251	0.0665	-3.77445	0.000193	0.032962
ANAPC15	-0.65702	0.17413	-3.77314	0.000194	0.032962
IGLJ7	-0.17574	0.046631	-3.7687	0.000198	0.03304
SYNRG	0.802704	0.213279	3.76364	0.000202	0.03304
USP34	0.655747	0.174461	3.758694	0.000205	0.03304
IGHG3	-0.19011	0.050658	-3.75276	0.00021	0.03304
IGKJ5	-0.2606	0.06954	-3.74741	0.000214	0.03304
CACFD1	-0.61144	0.16336	-3.74289	0.000218	0.03304
CDK13	1.011308	0.270375	3.740386	0.00022	0.03304
L3MBTL3	0.557375	0.149076	3.73887	0.000221	0.03304
IGHV3-23	-0.24576	0.065782	-3.73597	0.000224	0.03304
IGLV3-25	-0.21945	0.058769	-3.73409	0.000226	0.03304
HMGN2	-1.16862	0.31324	-3.73075	0.000228	0.03304
IGLV1-47	-0.2277	0.061059	-3.72915	0.00023	0.03304
H1FX-AS1	-0.51908	0.139323	-3.72573	0.000233	0.03304
IGLV1-36	-0.17053	0.0458	-3.72335	0.000235	0.03304
PRKDC	0.567734	0.152566	3.72124	0.000237	0.03304
MRPS7	-0.79269	0.213055	-3.72058	0.000237	0.03304
CENPM	-0.33396	0.089786	-3.71951	0.000238	0.03304
FOXJ3	0.802497	0.215833	3.718131	0.00024	0.03304
OXA1L	-1.05903	0.284967	-3.71631	0.000241	0.03304
DYRK1A	0.709758	0.191018	3.715653	0.000242	0.03304
TOR3A	-0.95051	0.255938	-3.71382	0.000244	0.03304
STAC	-0.1823	0.049152	-3.70877	0.000248	0.033359
VPS25	-0.79099	0.214157	-3.69352	0.000263	0.035002
MED17	1.128738	0.307302	3.673054	0.000284	0.036861
TTPAL	1.069479	0.291797	3.665147	0.000292	0.036861
CDC20	-0.23027	0.062856	-3.6634	0.000294	0.036861
SBNO1	0.875582	0.239168	3.660946	0.000297	0.036861
TMEM258	-0.65349	0.178687	-3.65716	0.000301	0.036861
SEC24B	0.706563	0.193201	3.657142	0.000301	0.036861
AKAP9	0.635762	0.173841	3.657136	0.000301	0.036861

***Table A3.2**: cont.*

Gene	Effect	S.E.	t	*p*-value	Adj. *p*-value
ATP6V1F	-0.70269	0.1923	-3.65412	0.000305	0.036861
SLC35B1	-1.11932	0.306425	-3.65284	0.000306	0.036861
ARIH1	0.800647	0.219294	3.65103	0.000308	0.036861
YTHDC1	1.039382	0.285142	3.645146	0.000315	0.036861
DDX6	0.776046	0.212993	3.643519	0.000317	0.036861
HMGB3	-0.46007	0.126528	-3.63612	0.000326	0.036861
IGKV2-28	-0.24971	0.068689	-3.63544	0.000327	0.036861
DCPS	-0.79103	0.217844	-3.63115	0.000332	0.036861
RPS6KA5	0.54262	0.149482	3.630012	0.000333	0.036861
WDR7	0.595862	0.164232	3.628176	0.000336	0.036861
IGLJ3	-0.21849	0.060233	-3.62733	0.000337	0.036861
ATP5G3	-0.76107	0.209864	-3.62648	0.000338	0.036861
LGALS1	-0.48082	0.132595	-3.62625	0.000338	0.036861
IGLJ6	-0.17174	0.047378	-3.62486	0.00034	0.036861
RANBP2	0.550631	0.151997	3.622646	0.000343	0.036861
USF3	0.487972	0.134781	3.62048	0.000345	0.036861
ANAPC1	0.658091	0.181921	3.617452	0.000349	0.036861
SYNE2	0.542589	0.150029	3.616563	0.00035	0.036861
TADA3	-0.87164	0.241048	-3.61605	0.000351	0.036861
TMSB10	-0.62368	0.172477	-3.61603	0.000351	0.036861
FRYL	0.55468	0.153431	3.615183	0.000352	0.036861
HNRNPU	0.81101	0.224486	3.612737	0.000355	0.036861
MED19	-0.63046	0.174539	-3.61215	0.000356	0.036861
PAQR4	-0.56784	0.157244	-3.6112	0.000357	0.036861
USP42	0.934485	0.258912	3.60928	0.00036	0.036861
TNRC6C	0.560574	0.155392	3.607491	0.000362	0.036861
TMEM101	-0.87889	0.243775	-3.60535	0.000365	0.03689
HECTD1	0.687272	0.190802	3.602016	0.00037	0.037083
MIF4GD	-0.75186	0.20911	-3.59552	0.000379	0.037587
IGHV3-64	-0.1515	0.042166	-3.59291	0.000382	0.037587
ZC3H13	0.622966	0.173424	3.592153	0.000383	0.037587
MRPL37	-0.83395	0.232245	-3.59081	0.000385	0.037587
EMSY	0.720725	0.200905	3.587386	0.00039	0.037634
SMARCA2	0.795748	0.221857	3.586761	0.000391	0.037634
CDK5	-0.71078	0.198362	-3.58324	0.000396	0.037708
MYBL2	-0.24402	0.068166	-3.57978	0.000401	0.037708
SRI	-1.00518	0.280838	-3.57922	0.000402	0.037708
PFDN2	-0.58482	0.163405	-3.57896	0.000402	0.037708
COQ9	-1.31146	0.36733	-3.57024	0.000415	0.038358
IGLJ2	-0.25066	0.070225	-3.56939	0.000417	0.038358
ELMSAN1	0.706705	0.198014	3.568962	0.000417	0.038358
E2F1	-0.40188	0.112805	-3.5626	0.000427	0.038552
CHD7	0.551552	0.15487	3.561392	0.000429	0.038552
TNFRSF17	-0.20861	0.058577	-3.56125	0.000429	0.038552

***Table A3.2**: cont.*

Gene	Effect	S.E.	t	*p*-value	Adj. *p*-value
MRPL16	-1.1323	0.318101	-3.55957	0.000432	0.038552
LRBA	0.528146	0.148401	3.558897	0.000433	0.038552
ARFGEF1	0.633668	0.178214	3.555655	0.000438	0.038597
C11orf24	-0.85607	0.240795	-3.55518	0.000439	0.038597
MPV17L2	-0.67979	0.19149	-3.54999	0.000447	0.03885
ACSF2	-0.69279	0.195168	-3.54972	0.000448	0.03885
RP11-831H9.11	-0.79471	0.223967	-3.54835	0.00045	0.03885
FXYD6-FXYD2	-0.49197	0.138992	-3.53953	0.000465	0.039517
IGLV2-5	-0.22633	0.063969	-3.53816	0.000467	0.039517
NUP153	0.648414	0.183577	3.532112	0.000477	0.039517
TMEM9	-0.61127	0.173069	-3.53196	0.000478	0.039517
FABP5	-0.41157	0.116564	-3.5309	0.000479	0.039517
EP400	0.721835	0.204523	3.52935	0.000482	0.039517
NDUFS3	-0.67241	0.190524	-3.52927	0.000482	0.039517
BIK	-0.40054	0.113532	-3.52804	0.000484	0.039517
MRPL27	-0.64315	0.182306	-3.52786	0.000485	0.039517
DOCK10	0.606897	0.17213	3.525812	0.000488	0.039517
ZNF136	0.637201	0.18083	3.523756	0.000492	0.039517
RECQL4	-0.47893	0.135919	-3.52365	0.000492	0.039517
POGZ	1.057242	0.300116	3.522782	0.000494	0.039517
BIRC5	-0.25022	0.071068	-3.52086	0.000497	0.03957
VSTM1	-0.25139	0.071493	-3.51629	0.000505	0.040006
B3GALT4	-0.73751	0.209932	-3.51308	0.000511	0.04025
MTR	0.511508	0.145668	3.511454	0.000514	0.040264
MIR611	-0.55254	0.157544	-3.50718	0.000522	0.040459
GOLGA6L9	0.273932	0.07827	3.49985	0.000536	0.040459
RAB34	-0.49302	0.140992	-3.49677	0.000542	0.040459
LYPD2	-0.16845	0.048174	-3.49673	0.000542	0.040459
TNNT1	-0.17577	0.050293	-3.49493	0.000546	0.040459
NEO1	0.39859	0.114106	3.493147	0.000549	0.040459
IGLC2	-0.21946	0.062834	-3.49262	0.000551	0.040459
FTX	0.406499	0.116453	3.490658	0.000554	0.040459
COMMD9	-0.75604	0.216685	-3.4891	0.000558	0.040459
IGHV3-21	-0.24006	0.068813	-3.48858	0.000559	0.040459
NDUFA8	-0.77295	0.221659	-3.48713	0.000562	0.040459
MYDGF	-0.68281	0.195812	-3.48707	0.000562	0.040459
TCTA	-0.96237	0.276244	-3.48378	0.000568	0.040459
NDUFV2	-0.7527	0.216115	-3.48289	0.00057	0.040459
ADNP	0.879265	0.2525	3.482239	0.000571	0.040459
PER3	0.429383	0.12334	3.481307	0.000573	0.040459
RRM2	-0.26206	0.075278	-3.48115	0.000574	0.040459
DHRS4	-0.64014	0.183913	-3.48064	0.000575	0.040459
IGLV3-27	-0.15556	0.044739	-3.47719	0.000582	0.040459
ARID2	0.565668	0.162687	3.477032	0.000582	0.040459

***Table A3.2**: cont.*

Gene	Effect	S.E.	t	*p*-value	Adj. *p*-value
STK4	0.633454	0.182252	3.475711	0.000585	0.040459
SLC26A6	-0.82131	0.236472	-3.47319	0.00059	0.040459
HEATR5B	0.666614	0.192066	3.470748	0.000595	0.040459
SCAF11	0.54414	0.156826	3.469698	0.000598	0.040459
RASSF4	-0.63479	0.18296	-3.46953	0.000598	0.040459
IGHV4-55	-0.16322	0.047047	-3.46938	0.000598	0.040459
FOXO1	0.614434	0.177196	3.467541	0.000602	0.040459
IGKV3D-11	-0.1805	0.052062	-3.46708	0.000603	0.040459
LBHD1	-0.80776	0.233033	-3.46628	0.000605	0.040459
PSMB7	-0.73197	0.211221	-3.46542	0.000607	0.040459
CEP350	0.480284	0.138736	3.461866	0.000615	0.040459
RP11-121C2.2	0.451603	0.130462	3.461577	0.000615	0.040459
OAF	-0.58015	0.167616	-3.46122	0.000616	0.040459
IGLV2-14	-0.22165	0.06404	-3.4611	0.000616	0.040459
MRPL21	-0.55823	0.161343	-3.45989	0.000619	0.040459
CCNB1	-0.431	0.124572	-3.45984	0.000619	0.040459
COPZ2	-0.35474	0.102598	-3.45759	0.000624	0.0406
MRPS11	-0.99239	0.287137	-3.45617	0.000627	0.040621
AQR	0.768035	0.222628	3.449849	0.000642	0.041209
SNIP1	1.062156	0.308005	3.448501	0.000645	0.041209
IGKV1D-39	-0.13502	0.039167	-3.44714	0.000648	0.041209
DNASE1L1	-0.82607	0.239686	-3.44646	0.000649	0.041209
SETX	0.641819	0.18626	3.445822	0.000651	0.041209
BATF3	-0.41172	0.11961	-3.44222	0.000659	0.041556
MBD5	0.960611	0.279273	3.439683	0.000665	0.041747
TIMM23	-0.86546	0.251758	-3.43766	0.00067	0.041755
RBM12	1.073279	0.312257	3.437164	0.000671	0.041755
PCLAF	-0.2381	0.069374	-3.43212	0.000683	0.042207
MRPL51	-0.61805	0.180136	-3.43103	0.000686	0.042207
ATF7IP	0.571185	0.166504	3.430457	0.000687	0.042207
REV3L	0.504956	0.147269	3.428794	0.000691	0.042274
C7orf49	-1.20252	0.351215	-3.42388	0.000704	0.042791
MEA1	-0.64235	0.187721	-3.42181	0.000709	0.042791
PFDN1	-0.71749	0.209685	-3.42175	0.000709	0.042791
TNFSF12-TNFSF13	-0.61798	0.180672	-3.42047	0.000712	0.042804
RABGAP1	0.922632	0.269908	3.418327	0.000718	0.042948
NDUFB9	-0.64712	0.18946	-3.41561	0.000724	0.043181
SCAF8	0.839318	0.245908	3.413135	0.000731	0.043379
METTL16	0.799004	0.234254	3.410852	0.000737	0.043548
TAF4	0.721069	0.211619	3.407392	0.000746	0.0439
TMEM141	-0.6457	0.189565	-3.40623	0.000749	0.0439
WAPL	0.639819	0.188139	3.400777	0.000763	0.044453
PER2	0.753184	0.221512	3.400194	0.000765	0.044453
NOM1	0.654083	0.192434	3.399005	0.000768	0.044453

***Table A3.2**: cont.*

Gene	Effect	S.E.	t	*p*-value	Adj. *p*-value
CTSH	-0.49083	0.144446	-3.39804	0.000771	0.044453
NUP133	0.879212	0.258931	3.395548	0.000778	0.044538
DOCK11	0.54228	0.159722	3.395142	0.000779	0.044538
PSMB6	-0.6192	0.182437	-3.39406	0.000782	0.044538
TMEM147	-0.6508	0.19197	-3.39009	0.000793	0.044727
NIPBL	0.520251	0.153618	3.386642	0.000802	0.044727
CYC1	-0.65434	0.193262	-3.38577	0.000805	0.044727
NUBP1	-0.79466	0.234841	-3.38383	0.00081	0.044727
PIK3R4	0.75462	0.223045	3.383257	0.000812	0.044727
C8orf82	-0.58918	0.174209	-3.38204	0.000815	0.044727
PRRC2C	0.54406	0.160897	3.381424	0.000817	0.044727
RP11-182J1.3	-0.61947	0.183204	-3.38133	0.000817	0.044727
RP11-21J18.1	-0.78628	0.232567	-3.38086	0.000819	0.044727
RP5-994D16.12	-0.85369	0.252583	-3.37985	0.000822	0.044727
GOLGA4	0.532321	0.157511	3.379574	0.000822	0.044727
DCTN3	-0.86041	0.254794	-3.37689	0.00083	0.044727
NDUFA7	-0.61989	0.183578	-3.3767	0.000831	0.044727
ORC1	-0.41772	0.123719	-3.3764	0.000832	0.044727
SHCBP1	-0.36769	0.108909	-3.37607	0.000833	0.044727
TUBG1	-0.71242	0.211235	-3.37265	0.000843	0.044727
XIAP	0.551601	0.163588	3.371882	0.000845	0.044727
SON	0.673855	0.199948	3.370145	0.00085	0.044727
YLPM1	0.643349	0.190902	3.370042	0.00085	0.044727
RP4-785G19.5	-0.63586	0.188694	-3.36976	0.000851	0.044727
MYCBP2	0.53007	0.157327	3.369215	0.000853	0.044727
IGKJ3	-0.25911	0.076926	-3.36832	0.000855	0.044727
ATP5G1	-0.59238	0.175899	-3.36774	0.000857	0.044727
ZNF100	0.339426	0.100875	3.364806	0.000866	0.044892
PUM1	0.612618	0.182078	3.364592	0.000867	0.044892
KIAA1429	0.772574	0.229811	3.361782	0.000875	0.045091
MTFR1L	-1.08009	0.321383	-3.36077	0.000878	0.045091
CLIP1	0.558289	0.166168	3.359776	0.000881	0.045091
EIF4EBP1	-0.49813	0.14829	-3.35918	0.000883	0.045091
BAX	-0.78936	0.235148	-3.35688	0.00089	0.045292
SPTY2D1	0.546106	0.162897	3.352459	0.000904	0.045831
IGKV2D-29	-0.17223	0.051453	-3.34738	0.00092	0.046329
MELK	-0.25668	0.076683	-3.34731	0.00092	0.046329
NDUFB4	-0.67118	0.200662	-3.34483	0.000928	0.046485
C3AR1	-0.32561	0.097362	-3.34432	0.00093	0.046485
CTD-2550O8.5	-0.56833	0.170072	-3.34172	0.000938	0.046648
EIF2S3L	-0.16774	0.050221	-3.33998	0.000944	0.046648
FUOM	-0.49384	0.147862	-3.3399	0.000944	0.046648
ZNF264	0.547815	0.164051	3.339294	0.000946	0.046648

***Table A3.2**: cont.*

Gene	Effect	S.E.	t	*p*-value	Adj. *p*-value
RASL11A	-0.41339	0.123879	-3.33708	0.000954	0.046761
VPS13D	0.685113	0.205333	3.336599	0.000955	0.046761
MAML2	0.530869	0.159152	3.335619	0.000958	0.046761
RBBP9	0.686755	0.206149	3.331357	0.000973	0.047295
PCM1	0.685522	0.20606	3.326814	0.000988	0.047758
ELK3	0.627718	0.188807	3.32465	0.000996	0.047758
PAPOLG	0.528103	0.158902	3.323452	0.001	0.047758
SLC2A8	-0.51785	0.155889	-3.3219	0.001005	0.047758
ACOT8	-0.95707	0.288162	-3.32129	0.001007	0.047758
BBS9	0.786321	0.23678	3.320888	0.001009	0.047758
GPANK1	-0.84824	0.255439	-3.32069	0.001009	0.047758
ATG2B	0.581592	0.175172	3.320121	0.001011	0.047758
SECISBP2L	0.382261	0.115196	3.318353	0.001017	0.047758
PIKFYVE	0.392031	0.118148	3.31815	0.001018	0.047758
AFF4	0.406682	0.122573	3.317876	0.001019	0.047758
PDE12	0.624626	0.188385	3.315693	0.001027	0.047962
RAD50	0.620264	0.187347	3.310777	0.001044	0.048147
PSMD8	-0.70677	0.213552	-3.30957	0.001049	0.048147
IGKV6-21	-0.17299	0.052297	-3.30796	0.001055	0.048147
AAK1	0.604517	0.182778	3.307383	0.001057	0.048147
SNHG11	-0.70941	0.214526	-3.30687	0.001059	0.048147
ATM	0.346482	0.104813	3.305699	0.001063	0.048147
SSR4	-0.57828	0.174973	-3.30498	0.001065	0.048147
ATR	0.631521	0.191126	3.304216	0.001068	0.048147
RASGRF2	0.468193	0.141724	3.303549	0.001071	0.048147
NDUFB10	-0.61961	0.187575	-3.30326	0.001072	0.048147
SLC35B2	-0.94145	0.28509	-3.30228	0.001075	0.048147
INTS2	0.417011	0.126285	3.302136	0.001076	0.048147
UBQLN1	0.784399	0.237615	3.30114	0.00108	0.048147
MTMR12	0.764106	0.231493	3.300768	0.001081	0.048147
PUM2	0.546591	0.165605	3.30057	0.001082	0.048147
TYMS	-0.27757	0.084172	-3.29765	0.001093	0.048147
FAM208B	0.547258	0.165962	3.297495	0.001093	0.048147
RAPGEF6	0.40151	0.121779	3.297046	0.001095	0.048147
PPP4R3A	0.790001	0.239654	3.296429	0.001097	0.048147
UFC1	-0.66298	0.20114	-3.29612	0.001098	0.048147
MRPS18A	-0.68428	0.207793	-3.29308	0.00111	0.048426
AC116366.7	0.843963	0.256441	3.291057	0.001118	0.048426
CDCA5	-0.29265	0.088987	-3.28862	0.001127	0.048426
MPST	-0.60759	0.184787	-3.28805	0.001129	0.048426
STAG1	0.678139	0.206285	3.287385	0.001132	0.048426
DERL3	-0.36135	0.10994	-3.28677	0.001134	0.048426
BIRC6	0.577596	0.175769	3.286103	0.001137	0.048426
ITM2C	-0.42776	0.130194	-3.28556	0.001139	0.048426

***Table A3.2**: cont.*

Gene	Effect	S.E.	t	*p*-value	Adj. *p*-value
ERP29	-0.82708	0.25186	-3.28388	0.001145	0.048426
LATS1	0.403689	0.122936	3.283727	0.001146	0.048426
PDCL3	-0.76342	0.232493	-3.28364	0.001146	0.048426
LY75-CD302	0.428393	0.130512	3.282413	0.001151	0.048426
CAMK1	-0.39457	0.120215	-3.28218	0.001152	0.048426
ZNF347	0.397768	0.121192	3.282128	0.001152	0.048426
KAT6B	0.535267	0.163246	3.278901	0.001165	0.048709
CNN3	0.329366	0.100456	3.278708	0.001166	0.048709
IGHV3-7	-0.21696	0.066222	-3.27619	0.001176	0.048973
LIX1L	0.732084	0.223519	3.275271	0.00118	0.048973
TNFSF13	-0.53157	0.162333	-3.27456	0.001183	0.048973
CUL4A	1.023683	0.312786	3.272789	0.00119	0.049008
UBR3	0.518334	0.158384	3.272653	0.00119	0.049008
DCP1A	0.821481	0.251139	3.271025	0.001197	0.049023
UBR5	0.617876	0.188971	3.269683	0.001202	0.049023
GOSR1	1.179774	0.361037	3.267736	0.00121	0.049023
NHLRC4	-0.53862	0.164848	-3.26739	0.001212	0.049023
HEIH	-0.81516	0.24951	-3.26704	0.001213	0.049023
KMT5B	0.818927	0.250823	3.264955	0.001222	0.049023
FAM168B	0.954356	0.292329	3.264668	0.001223	0.049023
NUP160	0.648584	0.198689	3.264322	0.001225	0.049023
AUP1	-0.88954	0.272537	-3.26393	0.001226	0.049023
ZNF486	0.387827	0.118846	3.263263	0.001229	0.049023
NDUFA9	-0.92028	0.282042	-3.26292	0.00123	0.049023
NDUFB8	-0.79247	0.242897	-3.26256	0.001232	0.049023
CLEC3B	-0.62826	0.192655	-3.26104	0.001238	0.04914
REST	0.400114	0.122839	3.257218	0.001255	0.049533
CDCA3	-0.34816	0.1069	-3.25684	0.001256	0.049533
NPAT	0.523105	0.160646	3.256258	0.001259	0.049533
IGLV3-10	-0.13838	0.042547	-3.2525	0.001275	0.049657
KIAA0895L	-0.66695	0.205069	-3.25233	0.001276	0.049657
RP11-216L13.17	-0.61025	0.187643	-3.25218	0.001276	0.049657
GAPVD1	0.635716	0.195517	3.251455	0.001279	0.049657
AKR1A1	-0.6909	0.212546	-3.2506	0.001283	0.049657
RALGAPB	0.786649	0.242044	3.250029	0.001286	0.049657
UQCRFS1	-0.70757	0.217721	-3.24988	0.001286	0.049657
NT5DC2	-0.34757	0.106998	-3.24839	0.001293	0.049776
AL928654.7	-0.44115	0.135861	-3.24704	0.001299	0.049869
NEK9	1.06174	0.32714	3.245524	0.001305	0.049991

***Table A3.3**: Identified differentially expressed genes of IgG GP23*

Gene	Effect	S.E.	t	*p*-value	Adj. *p*-value
IGKV3-7	0.297334	0.054884	5.417505	1.24E-07	0.001722
IGKV1-5	0.308935	0.060129	5.137829	5.02E-07	0.001722
IGKV1-6	0.277773	0.054432	5.103151	5.95E-07	0.001722
IGLV1-36	0.224765	0.044188	5.086535	6.45E-07	0.001722
IGLC7	0.247339	0.048966	5.051217	7.65E-07	0.001722
IGKV3D-20	0.253752	0.050431	5.03165	8.41E-07	0.001722
IGLL5	0.282475	0.05646	5.003134	9.64E-07	0.001722
IGKC	0.321723	0.064828	4.962686	1.17E-06	0.001722
IGKJ1	0.357794	0.072163	4.958108	1.20E-06	0.001722
IGKV3D-11	0.2487	0.050184	4.955741	1.21E-06	0.001722
IGLC1	0.268886	0.054669	4.918471	1.44E-06	0.001727
MZB1	0.369022	0.075057	4.91655	1.46E-06	0.001727
IGLJ6	0.223201	0.045809	4.87244	1.79E-06	0.001905
IGLJ1	0.281921	0.057973	4.862959	1.87E-06	0.001905
IGHA1	0.224186	0.046342	4.837601	2.11E-06	0.00194
IGLJ7	0.218298	0.045192	4.830466	2.18E-06	0.00194
SDC1	0.199966	0.041528	4.815195	2.34E-06	0.00196
JCHAIN	0.232858	0.049488	4.705297	3.88E-06	0.003006
IGLC2	0.285734	0.060818	4.698145	4.01E-06	0.003006
IGKV3D-15	0.190236	0.040851	4.65682	4.84E-06	0.003445
IGLJ2	0.316216	0.068085	4.644427	5.12E-06	0.00347
IGKV4-1	0.308029	0.066685	4.619168	5.73E-06	0.003667
IGKJ4	0.286549	0.062132	4.611969	5.92E-06	0.003667
IGKV1-12	0.25047	0.054887	4.563358	7.36E-06	0.003801
IGLV2-23	0.280911	0.061587	4.561172	7.43E-06	0.003801
FA2H	0.263302	0.0578	4.555401	7.62E-06	0.003801
IGLC3	0.234603	0.051524	4.553304	7.70E-06	0.003801
IGKV1-9	0.259478	0.05701	4.551438	7.76E-06	0.003801
IGKV3-20	0.296578	0.065204	4.548446	7.86E-06	0.003801
TXNDC5	0.421947	0.092857	4.544056	8.02E-06	0.003801
BLOC1S5-TXNDC5	0.4335	0.095546	4.537062	8.27E-06	0.003801
IGLV3-25	0.258154	0.057195	4.513566	9.18E-06	0.004086
TNFRSF17	0.254785	0.056923	4.475986	1.08E-05	0.004674
IGKV2-29	0.19546	0.043799	4.462612	1.15E-05	0.00481
IGLV1-51	0.261534	0.05876	4.450911	1.21E-05	0.004918
IGLV2-5	0.276089	0.062177	4.440355	1.27E-05	0.005007
IGLV2-11	0.242133	0.054677	4.428424	1.33E-05	0.005131
IGLV2-14	0.274993	0.06221	4.42043	1.38E-05	0.005173
GPRC5D	0.177599	0.040237	4.413785	1.42E-05	0.005188
IGKV3-11	0.281419	0.064185	4.384496	1.61E-05	0.005742
IGKV1-17	0.24812	0.056786	4.369418	1.72E-05	0.005977
IGLV1-47	0.259071	0.059555	4.350141	1.87E-05	0.006339
IGLV3-1	0.25635	0.059501	4.308288	2.23E-05	0.007403
IGLV7-43	0.233544	0.054619	4.275893	2.56E-05	0.008301

***Table A3.3**: Cont.*

Gene	Effect	S.E.	t	p-value	Adj. p-value
IGLV3-21	0.199463	0.04676	4.265689	2.68E-05	0.008474
PCLAF	0.285491	0.067548	4.226493	3.16E-05	0.009776
IGLV3-19	0.245858	0.05868	4.189829	3.68E-05	0.011034
IGKV3-15	0.264677	0.06321	4.18729	3.72E-05	0.011034
IGKV1-16	0.249534	0.059798	4.172915	3.95E-05	0.011474
IGHV3-15	0.271957	0.065599	4.145744	4.42E-05	0.012583
IGKV1-8	0.281357	0.068253	4.12229	4.86E-05	0.013586
IGLJ3	0.242056	0.058866	4.111955	5.07E-05	0.013902
IGLV1-40	0.253816	0.06202	4.092496	5.50E-05	0.014771
IGHA2	0.16371	0.040278	4.064466	6.16E-05	0.01625
BIRC5	0.281426	0.069423	4.053782	6.43E-05	0.016497
IGKV1-33	0.2524	0.062294	4.051778	6.49E-05	0.016497
IGKJ2	0.282277	0.069885	4.039183	6.82E-05	0.017055
AC096579.15	0.28225	0.070066	4.028347	7.13E-05	0.017396
IGKV2-28	0.270574	0.067211	4.025733	7.20E-05	0.017396
PHLDB2	-0.41251	0.103174	-3.9982	8.05E-05	0.019052
IGKV6-21	0.203755	0.051005	3.994814	8.16E-05	0.019052
CDC20	0.245369	0.061555	3.986191	8.44E-05	0.019403
IGHJ4	0.315612	0.079417	3.974111	8.86E-05	0.019971
IGHV3-72	0.203241	0.051181	3.971003	8.97E-05	0.019971
TPPP3	0.367082	0.092953	3.949132	9.79E-05	0.021451
IGLC6	0.219965	0.055817	3.940852	0.000101	0.021832
CENPM	0.345772	0.088051	3.926961	0.000107	0.022721
IGKV2-24	0.219052	0.055966	3.913992	0.000112	0.023225
IGHV3-49	0.240163	0.061361	3.913944	0.000112	0.023225
IGKJ3	0.293668	0.075158	3.907374	0.000115	0.023281
CCNB1	0.475498	0.121837	3.902755	0.000118	0.023281
IGKV1-27	0.254746	0.065277	3.902524	0.000118	0.023281
IGLV2-8	0.210098	0.053964	3.893326	0.000122	0.023806
IGLV1-44	0.209984	0.053999	3.888655	0.000124	0.023919
IGKJ5	0.264974	0.068257	3.881982	0.000128	0.024225
TTPAL	-1.10961	0.28615	-3.87772	0.00013	0.024309
IGLV10-54	0.136676	0.035375	3.863585	0.000137	0.025354
SHCBP1	0.410514	0.106483	3.855218	0.000142	0.025859
MBD5	-1.05041	0.273279	-3.84373	0.000148	0.026698
IGHV3-23	0.246852	0.064608	3.82078	0.000162	0.028816
IGHV4-59	0.223422	0.058634	3.810447	0.000168	0.029295
CCNB2	0.279049	0.073239	3.810113	0.000169	0.029295
HJURP	0.278943	0.073617	3.789093	0.000183	0.03138
IGKV2D-29	0.1903	0.050337	3.780542	0.000189	0.03204
GINS2	0.336424	0.089273	3.76847	0.000198	0.033161
SRRM1	-1.04752	0.278385	-3.76284	0.000202	0.033488
VPS13D	-0.75281	0.200936	-3.7465	0.000215	0.03523
MIR650	0.199402	0.053274	3.742919	0.000218	0.035307

***Table A3.3**: Cont.*

Gene	Effect	S.E.	t	p-value	Adj. p-value
SYNRG	-0.7806	0.209794	-3.72078	0.000237	0.037968
S100A10	0.647705	0.174218	3.717773	0.00024	0.037976
IGLV2-18	0.171273	0.046135	3.712433	0.000245	0.038324
IGKV2-30	0.229074	0.061883	3.701696	0.000255	0.039332
ANAPC1	-0.6611	0.178685	-3.69979	0.000257	0.039332
BSCL2	0.568637	0.153878	3.695369	0.000261	0.039367
GLDC	0.153486	0.041551	3.693896	0.000263	0.039367
STMN1	0.624958	0.169392	3.689418	0.000267	0.039619
RRM2	0.271523	0.073844	3.676961	0.00028	0.041087
IGKV1-39	0.228469	0.062222	3.671855	0.000285	0.041453
CDT1	0.299449	0.081701	3.665174	0.000292	0.042073
SLC50A1	0.976059	0.266886	3.657213	0.000301	0.042909
BMP8B	0.385449	0.105541	3.652113	0.000307	0.0433
ELMSAN1	-0.70838	0.194516	-3.64178	0.000319	0.044558
SMARCA2	-0.79225	0.218001	-3.63416	0.000328	0.045392
LINC00649	-0.35531	0.098029	-3.62453	0.00034	0.046591
NEO1	-0.40377	0.112044	-3.60368	0.000367	0.049844

Appendix 4 – Full List of GO Enrichment Terms

Although the focus of **Chapter 4** was on GP10 and only the top 25 GO enrichment terms for biological processes and molecular function were presented, below is the full list of GO terms identified for GP10 and GP23. Details of their derivation are in **Section 4.2.3.8**.

Page intentionally left blank

***Table A4.1**: GO enrichment terms for GP10 biological processes. Up-regulated GO terms denote those of positively-associated differentially expressed genes, whereas down-regulated GO terms are of negatively-associated genes. Only GO terms with adjusted p-value < 0.05 are listed. Combined score used both the result of the Fisher's exact test and the p-value*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
complement activation, classical pathway (GO:0006958)	Down	1.18E-41	2.80E-38	-3.099	291.998
regulation of activation of membrane attack complex (GO:0001969)	Down	9.87E-35	7.79E-32	-3.086	241.626
regulation of complement activation, classical pathway (GO:0030450)	Down	9.87E-35	7.79E-32	-3.083	241.376
regulation of complement activation (GO:0030449)	Down	2.59E-34	1.53E-31	-3.159	244.345
negative regulation of complement activation (GO:0045916)	Down	1.62E-33	7.67E-31	-3.193	241.054
positive regulation of complement activation (GO:0045917)	Down	3.89E-33	1.53E-30	-3.218	240.147
regulation of complement activation, alternative pathway (GO:0030451)	Down	5.18E-33	1.53E-30	-3.172	235.837
regulation of complement activation, lectin pathway (GO:0001868)	Down	5.18E-33	1.53E-30	-3.169	235.618
receptor-mediated endocytosis of virus by host cell (GO:0019065)	Down	5.76E-31	1.52E-28	-3.267	227.471
receptor-mediated endocytosis involved in cholesterol transport (GO:0090118)	Down	1.49E-30	3.52E-28	-3.311	227.433
Fc-gamma receptor signaling pathway involved in phagocytosis (GO:0038096)	Down	3.77E-30	8.11E-28	-3.216	217.867
clathrin-dependent endocytosis (GO:0072583)	Down	7.29E-30	1.44E-27	-3.235	217.074
receptor internalization (GO:0031623)	Down	1.37E-28	2.49E-26	-3.610	231.603
receptor-mediated endocytosis (GO:0006898)	Down	3.57E-26	6.04E-24	-3.604	211.184
Fc-epsilon receptor signaling pathway (GO:0038095)	Down	7.75E-26	1.22E-23	-3.392	196.143
positive regulation of mast cell activation by Fc-epsilon receptor signaling pathway (GO:0038097)	Down	9.17E-26	1.36E-23	-3.455	199.171
regulation of antigen receptor-mediated signaling pathway (GO:0050854)	Down	6.49E-22	9.04E-20	-3.602	175.743
regulation of Fc receptor mediated stimulatory signaling pathway (GO:0060368)	Down	8.53E-22	1.01E-19	-3.551	172.288
regulation of basophil degranulation (GO:1903581)	Down	1.07E-21	1.01E-19	-3.426	165.404
regulation of CD4-positive, CD25-positive, alpha-beta regulatory T cell differentiation involved in immune response (GO:0032832)	Down	1.07E-21	1.01E-19	-3.424	165.337
regulation of eosinophil degranulation (GO:0043309)	Down	1.07E-21	1.01E-19	-3.422	165.244
regulation of gamma-delta T cell activation involved in immune response (GO:2001191)	Down	1.07E-21	1.01E-19	-3.420	165.143

Page intentionally left blank

***Table A4.1**: cont.*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
regulation of immune response (GO:0050776)	Down	1.07E-21	1.01E-19	-3.416	164.958
regulation of immune response to tumor cell (GO:0002837)	Down	1.07E-21	1.01E-19	-3.415	164.909
regulation of type 2 immune response (GO:0002828)	Down	1.07E-21	1.01E-19	-3.414	164.827
regulation of humoral immune response (GO:0002920)	Down	1.24E-21	1.05E-19	-3.472	167.157
regulation of T-helper cell differentiation (GO:0045622)	Down	1.24E-21	1.05E-19	-3.460	166.578
regulation of neutrophil degranulation (GO:0043313)	Down	1.24E-21	1.05E-19	-3.451	166.160
regulation of natural killer cell proliferation involved in immune response (GO:0032820)	Down	1.43E-21	1.09E-19	-3.489	167.486
regulation of natural killer cell differentiation involved in immune response (GO:0032826)	Down	1.43E-21	1.09E-19	-3.468	166.454
regulation of mast cell activation involved in immune response (GO:0033006)	Down	1.43E-21	1.09E-19	-3.465	166.322
regulation of memory T cell differentiation (GO:0043380)	Down	1.65E-21	1.15E-19	-3.482	166.656
regulation of plasma cell differentiation (GO:1900098)	Down	1.65E-21	1.15E-19	-3.471	166.108
regulation of adaptive immune response (GO:0002819)	Down	1.65E-21	1.15E-19	-3.447	164.961
positive regulation of immune response (GO:0050778)	Down	1.90E-21	1.28E-19	-3.488	166.413
negative regulation of immune response (GO:0050777)	Down	2.51E-21	1.65E-19	-3.549	168.331
regulation of cytokine production involved in immune response (GO:0002718)	Down	2.88E-21	1.84E-19	-3.511	166.054
regulation of T cell activation via T cell receptor contact with antigen bound to MHC molecule on antigen presenting cell (GO:2001188)	Down	3.31E-21	2.06E-19	-3.534	166.660
regulation of innate immune response (GO:0045088)	Down	6.52E-21	3.96E-19	-3.547	164.843
regulation of inflammatory response to antigenic stimulus (GO:0002861)	Down	2.35E-19	1.39E-17	-3.722	159.654
positive regulation of B cell activation (GO:0050871)	Down	1.85E-13	1.07E-11	-2.667	78.197
engulfment of apoptotic cell (GO:0043652)	Down	2.83E-13	1.59E-11	-2.774	80.145
membrane reorganization involved in phagocytosis, engulfment (GO:0060098)	Down	1.53E-12	8.24E-11	-2.706	73.614
positive regulation of B cell differentiation (GO:0045579)	Down	1.53E-12	8.24E-11	-2.701	73.483
positive regulation of isotype switching (GO:0045830)	Down	1.86E-12	9.76E-11	-2.616	70.673
positive regulation of B cell proliferation (GO:0030890)	Down	2.66E-12	1.37E-10	-2.917	77.740
B cell receptor signaling pathway (GO:0050853)	Down	1.49E-11	7.51E-10	-2.801	69.824

Page intentionally left blank

***Table A4.1**: cont.*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
B cell receptor apoptotic signaling pathway (GO:1990117)	Down	4.39E-11	2.12E-09	-3.125	74.530
phagocytosis, engulfment (GO:0006911)	Down	4.31E-11	2.12E-09	-2.746	65.541
antibacterial humoral response (GO:0019731)	Down	4.99E-10	2.36E-08	-3.204	68.616
mitochondrial respiratory chain complex I assembly (GO:0032981)	Down	4.91E-09	2.28E-07	-2.639	50.495
mitochondrial electron transport, NADH to ubiquinone (GO:0006120)	Down	1.72E-08	7.82E-07	-2.546	45.531
cytoskeletal rearrangement involved in phagocytosis, engulfment (GO:0060097)	Down	3.82E-08	1.71E-06	-3.147	53.758
DNA damage induced protein phosphorylation (GO:0006975)	Up	3.70E-09	9.56E-06	-5.123	99.455
defense response to bacterium (GO:0042742)	Down	4.29E-07	1.75E-05	-2.929	42.942
defense response to bacterium, incompatible interaction (GO:0009816)	Down	4.29E-07	1.75E-05	-2.927	42.912
induced systemic resistance (GO:0009682)	Down	4.29E-07	1.75E-05	-2.926	42.898
male-specific defense response to bacterium (GO:0050831)	Down	4.29E-07	1.75E-05	-2.923	42.856
peptidoglycan recognition protein signaling pathway (GO:0061057)	Down	4.29E-07	1.75E-05	-2.921	42.821
neutrophil mediated killing of bacterium (GO:0070944)	Down	4.63E-07	1.86E-05	-2.924	42.650
mitochondrial translational initiation (GO:0070124)	Down	2.52E-06	9.95E-05	-2.399	30.927
defense response to Gram-negative bacterium (GO:0050829)	Down	3.51E-06	1.32E-04	-3.203	40.229
mitochondrial translational elongation (GO:0070125)	Down	3.53E-06	1.32E-04	-2.764	34.702
mitochondrial translational termination (GO:0070126)	Down	3.53E-06	1.32E-04	-2.662	33.427
defense response to Gram-positive bacterium (GO:0050830)	Down	4.99E-06	1.85E-04	-3.150	38.452
glomerular filtration (GO:0003094)	Down	5.75E-06	2.09E-04	-2.003	24.167
histone H4 acetylation involved in response to DNA damage stimulus (GO:2000776)	Up	2.09E-07	2.33E-04	-3.680	56.605
cellular response to DNA damage stimulus (GO:0006974)	Up	5.14E-07	2.33E-04	-3.424	49.586
histone H3-K56 acetylation in response to DNA damage (GO:0097044)	Up	5.14E-07	2.33E-04	-3.422	49.551
SOS response (GO:0009432)	Up	5.14E-07	2.33E-04	-3.419	49.514
regulation of transcription from RNA polymerase II promoter in response to DNA damage (GO:1990248)	Up	5.41E-07	2.33E-04	-3.444	49.703
signal transduction in response to DNA damage (GO:0042770)	Up	6.92E-07	2.55E-04	-3.463	49.122

Page intentionally left blank

***Table A4.1**: cont.*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
regulation of gene silencing by miRNA (GO:0060964)	Up	8.17E-07	2.64E-04	-2.945	41.289
DNA damage checkpoint (GO:0000077)	Up	1.06E-06	3.04E-04	-3.539	48.686
telomere maintenance in response to DNA damage (GO:0043247)	Up	1.66E-06	4.27E-04	-3.595	47.842
regulation of miRNA mediated inhibition of translation (GO:1905616)	Up	2.04E-06	4.27E-04	-2.672	35.016
regulation of production of miRNAs involved in gene silencing by miRNA (GO:1903798)	Up	2.04E-06	4.27E-04	-2.669	34.971
intrinsic apoptotic signaling pathway in response to DNA damage (GO:0008630)	Up	2.15E-06	4.27E-04	-3.602	47.006
positive regulation of gene silencing by miRNA (GO:2000637)	Up	3.77E-06	6.96E-04	-2.794	34.894
lamin depolymerization (GO:0007078)	Up	1.37E-05	2.16E-03	-2.954	33.073
negative regulation of cellular response to heat (GO:1900035)	Up	1.47E-05	2.16E-03	-3.131	34.831
positive regulation of cellular response to heat (GO:1900036)	Up	1.47E-05	2.16E-03	-3.123	34.751
mitotic nuclear envelope disassembly (GO:0007077)	Up	1.53E-05	2.16E-03	-2.792	30.948
regulation of cellular response to heat (GO:1900034)	Up	1.59E-05	2.16E-03	-3.091	34.158
regulation of mRNA export from nucleus in response to heat stress (GO:2000728)	Up	1.97E-05	2.54E-03	-3.084	33.417
chemical homeostasis within retina (GO:0048876)	Down	7.85E-05	2.73E-03	-2.169	20.500
retina blood vessel maintenance (GO:0097601)	Down	7.85E-05	2.73E-03	-2.167	20.484
retina homeostasis (GO:0001895)	Down	7.85E-05	2.73E-03	-2.164	20.457
DNA repair (GO:0006281)	Up	2.30E-05	2.76E-03	-3.975	42.461
mitotic sister chromatid cohesion, telomeric (GO:0099404)	Up	2.56E-05	2.76E-03	-3.097	32.737
establishment of mitotic sister chromatid cohesion (GO:0034087)	Up	2.56E-05	2.76E-03	-3.057	32.314
mitotic sister chromatid cohesion, arms (GO:0071961)	Up	2.56E-05	2.76E-03	-2.926	30.935
homeostasis of number of retina cells (GO:0048877)	Down	8.92E-05	3.06E-03	-2.209	20.601
mitotic sister chromatid cohesion, centromeric (GO:0071962)	Up	3.64E-05	3.76E-03	-3.203	32.731
tRNA aminoacylation for mitochondrial protein translation (GO:0070127)	Down	1.14E-04	3.87E-03	-2.578	23.400
negative regulation of gene silencing by miRNA (GO:0060965)	Up	4.88E-05	4.85E-03	-3.079	30.564
protein phosphorylation involved in DNA double-strand break processing (GO:1990802)	Up	5.91E-05	5.65E-03	-4.267	41.545

Page intentionally left blank

***Table A4.1**: cont.*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
maintenance of mitotic sister chromatid cohesion (GO:0034088)	Up	6.61E-05	6.10E-03	-2.696	25.947
regulation of glucose transport (GO:0010827)	Up	7.39E-05	6.15E-03	-2.860	27.206
regulation of intestinal D-glucose absorption (GO:1903985)	Up	7.39E-05	6.15E-03	-2.858	27.191
tRNA re-export from nucleus (GO:0071528)	Up	7.39E-05	6.15E-03	-2.736	26.024
regulation of glucose import (GO:0046324)	Up	1.05E-04	7.67E-03	-2.826	25.898
positive regulation of glucose transport (GO:0010828)	Up	1.17E-04	7.67E-03	-2.822	25.555
protein phosphorylation involved in double-strand break repair via nonhomologous end joining (GO:1990804)	Up	1.26E-04	7.67E-03	-4.309	38.693
negative regulation of glucose transport (GO:0010829)	Up	1.44E-04	7.67E-03	-2.823	24.978
cap snatching (GO:0075526)	Up	1.59E-04	7.67E-03	-2.962	25.910
latent virus replication (GO:0019045)	Up	1.59E-04	7.67E-03	-2.957	25.863
maintenance of viral latency (GO:0019044)	Up	1.59E-04	7.67E-03	-2.953	25.829
recruitment of helicase-primase complex to DNA lesions (GO:0046799)	Up	1.59E-04	7.67E-03	-2.948	25.788
pore formation by virus in membrane of host cell (GO:0039707)	Up	1.59E-04	7.67E-03	-2.940	25.713
transport of virus (GO:0046794)	Up	1.59E-04	7.67E-03	-2.937	25.687
superinfection exclusion (GO:0098669)	Up	1.59E-04	7.67E-03	-2.934	25.662
viral gene expression (GO:0019080)	Up	1.59E-04	7.67E-03	-2.924	25.572
viral genome circularization (GO:0099009)	Up	1.59E-04	7.67E-03	-2.922	25.556
viral process (GO:0016032)	Up	1.59E-04	7.67E-03	-2.921	25.545
viral latency (GO:0019042)	Up	1.59E-04	7.67E-03	-2.921	25.545
viral scaffold assembly and maintenance (GO:0046807)	Up	1.59E-04	7.67E-03	-2.917	25.517
viral tropism switching (GO:0098678)	Up	1.59E-04	7.67E-03	-2.916	25.505
virus tail fiber assembly (GO:0098004)	Up	1.59E-04	7.67E-03	-2.906	25.418
virus baseplate assembly (GO:0098045)	Up	1.59E-04	7.67E-03	-2.900	25.362
peptidyl-serine phosphorylation (GO:0018105)	Up	1.59E-04	7.67E-03	-4.455	38.954
ribosomal skipping (GO:0075524)	Up	1.75E-04	7.67E-03	-2.928	25.324

Page intentionally left blank

***Table A4.1**: cont.*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
viral translation (GO:0019081)	Up	1.75E-04	7.67E-03	-2.915	25.215
viral translational frameshifting (GO:0075523)	Up	1.75E-04	7.67E-03	-2.908	25.148
viral protein processing (GO:0019082)	Up	1.75E-04	7.67E-03	-2.899	25.077
uncoating of virus (GO:0019061)	Up	1.75E-04	7.67E-03	-2.868	24.809
establishment of viral latency (GO:0019043)	Up	1.75E-04	7.67E-03	-2.839	24.556
transmission of virus (GO:0019089)	Up	1.75E-04	7.67E-03	-2.819	24.386
viral capsid secondary envelopment (GO:0046745)	Up	1.75E-04	7.67E-03	-2.800	24.219
protein autophosphorylation (GO:0046777)	Up	1.83E-04	7.86E-03	-4.652	40.044
virion maturation (GO:0098677)	Up	1.93E-04	8.03E-03	-3.014	25.779
release from viral latency (GO:0019046)	Up	1.93E-04	8.03E-03	-2.977	25.462
peptidyl-arginine phosphorylation (GO:0018109)	Up	2.17E-04	8.22E-03	-4.016	33.871
peptidyl-aspartic acid phosphorylation (GO:0018217)	Up	2.17E-04	8.22E-03	-4.014	33.856
peptidyl-cysteine phosphorylation (GO:0018218)	Up	2.17E-04	8.22E-03	-4.012	33.836
peptidyl-histidine phosphorylation (GO:0018106)	Up	2.17E-04	8.22E-03	-4.010	33.822
protein phosphorylation involved in protein localization to spindle microtubule (GO:1990803)	Up	2.17E-04	8.22E-03	-4.009	33.807
regulation of translational initiation by eIF2 alpha phosphorylation (GO:0010998)	Up	2.17E-04	8.22E-03	-4.006	33.781
inhibitory G-protein coupled receptor phosphorylation (GO:0002030)	Up	2.23E-04	8.22E-03	-4.020	33.803
actin phosphorylation (GO:0031289)	Up	2.23E-04	8.22E-03	-4.002	33.651
phosphorylation of RNA polymerase II C-terminal domain (GO:0070816)	Up	2.34E-04	8.48E-03	-5.625	47.023
DNA double-strand break processing (GO:0000729)	Up	2.37E-04	8.48E-03	-2.674	22.326
histone phosphorylation (GO:0016572)	Up	2.46E-04	8.58E-03	-4.027	33.469
common-partner SMAD protein phosphorylation (GO:0007182)	Up	2.46E-04	8.58E-03	-4.010	33.323
DNA strand renaturation (GO:0000733)	Down	2.64E-04	8.67E-03	-2.652	21.854
male-specific antibacterial humoral response (GO:0006962)	Down	2.64E-04	8.67E-03	-2.251	18.553
latency-replication decision (GO:0098689)	Up	2.52E-04	8.69E-03	-2.653	21.976

Page intentionally left blank

***Table A4.1**: cont.*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
negative regulation of protein kinase activity by protein phosphorylation (GO:0100002)	Up	2.59E-04	8.81E-03	-4.403	36.356
double-strand break repair via nonhomologous end joining (GO:0006303)	Up	2.65E-04	8.89E-03	-2.697	22.207
viral translational termination-reinitiation (GO:0075525)	Up	2.75E-04	9.07E-03	-2.765	22.672
pathway-restricted SMAD protein phosphorylation (GO:0060389)	Up	2.77E-04	9.07E-03	-4.093	33.521
protein phosphorylation involved in cellular protein catabolic process (GO:1902002)	Up	2.91E-04	9.39E-03	-4.021	32.737
tRNA export from nucleus (GO:0006409)	Up	2.96E-04	9.42E-03	-2.772	22.517
modulation by virus of host morphology or physiology (GO:0019048)	Up	2.99E-04	9.42E-03	-3.009	24.417
antibacterial peptide production (GO:0002778)	Down	3.05E-04	9.88E-03	-2.194	17.766
IRES-dependent viral translational initiation (GO:0075522)	Up	3.25E-04	1.01E-02	-2.844	22.848
positive regulation of respiratory burst (GO:0060267)	Down	3.26E-04	1.04E-02	-2.151	17.270
negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle (GO:0051436)	Down	3.36E-04	1.06E-02	-2.595	20.758
peptidyl-threonine phosphorylation (GO:0018107)	Up	3.67E-04	1.13E-02	-4.162	32.926
protein phosphorylation involved in mitotic spindle assembly (GO:1990801)	Up	3.93E-04	1.18E-02	-4.175	32.743
viral capsid assembly (GO:0019069)	Up	4.10E-04	1.18E-02	-2.757	21.498
viral genome maturation (GO:0019070)	Up	4.10E-04	1.18E-02	-2.753	21.469
viral genome packaging (GO:0019072)	Up	4.10E-04	1.18E-02	-2.751	21.452
viral head-tail joining (GO:0098005)	Up	4.10E-04	1.18E-02	-2.746	21.412
viral tail assembly (GO:0098003)	Up	4.10E-04	1.18E-02	-2.741	21.377
peptidyl-tyrosine phosphorylation (GO:0018108)	Up	4.29E-04	1.22E-02	-4.295	33.303
JUN phosphorylation (GO:0007258)	Up	4.39E-04	1.23E-02	-4.209	32.543
induction of bacterial agglutination (GO:0043152)	Down	4.00E-04	1.25E-02	-2.211	17.294
genome retention in viral capsid (GO:0046815)	Up	4.76E-04	1.29E-02	-2.715	20.771
double-strand break repair via single-strand annealing (GO:0045002)	Up	4.76E-04	1.29E-02	-2.459	18.812
mitochondrial double-strand break repair (GO:0097551)	Up	4.76E-04	1.29E-02	-2.452	18.762
viral replication complex formation and maintenance (GO:0046786)	Up	5.11E-04	1.32E-02	-2.742	20.784

Page intentionally left blank

***Table A4.1**: cont.*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
actin-dependent intracellular transport of virus (GO:0075520)	Up	5.11E-04	1.32E-02	-2.633	19.959
microtubule-dependent intracellular transport of viral material (GO:0075519)	Up	5.11E-04	1.32E-02	-2.628	19.917
transport of viral material towards nucleus (GO:0075606)	Up	5.11E-04	1.32E-02	-2.626	19.899
viral penetration into host nucleus (GO:0075732)	Up	5.11E-04	1.32E-02	-2.622	19.869
lysis of host organelle involved in viral entry into host cell (GO:0039664)	Up	5.87E-04	1.49E-02	-2.801	20.838
permeabilization of host organelle membrane involved in viral entry into host cell (GO:0039665)	Up	5.87E-04	1.49E-02	-2.797	20.811
viral budding (GO:0046755)	Up	6.28E-04	1.57E-02	-2.605	19.208
I-kappaB phosphorylation (GO:0007252)	Up	6.44E-04	1.59E-02	-4.347	31.943
positive regulation of mRNA splicing, via spliceosome (GO:0048026)	Up	6.48E-04	1.59E-02	-2.220	16.300
poly(A)+ mRNA export from nucleus (GO:0016973)	Up	7.49E-04	1.82E-02	-2.753	19.816
negative regulation of transcription involved in G1/S transition of mitotic cell cycle (GO:0071930)	Up	7.70E-04	1.86E-02	-4.818	34.543
mitotic sister chromatid cohesion (GO:0007064)	Up	8.15E-04	1.94E-02	-2.482	17.653
viral DNA repair (GO:0046787)	Up	8.28E-04	1.94E-02	-3.090	21.929
positive regulation of GTPase activity (GO:0043547)	Up	8.28E-04	1.94E-02	-3.040	21.570
exit of virus from host cell nucleus (GO:0039674)	Up	9.74E-04	2.27E-02	-2.587	17.942
gap filling involved in double-strand break repair via nonhomologous end joining (GO:0061674)	Up	1.09E-03	2.45E-02	-2.597	17.708
regulation of defense response to virus by virus (GO:0050690)	Up	1.16E-03	2.45E-02	-2.593	17.534
positive regulation of apoptotic process in other organism (GO:0044533)	Up	1.17E-03	2.45E-02	-3.708	25.048
positive regulation of apoptotic process involved in development (GO:1904747)	Up	1.17E-03	2.45E-02	-3.705	25.026
positive regulation of compound eye retinal cell apoptotic process (GO:1901694)	Up	1.17E-03	2.45E-02	-3.704	25.018
positive regulation of epithelial cell apoptotic process (GO:1904037)	Up	1.17E-03	2.45E-02	-3.701	25.000
positive regulation of fat cell apoptotic process (GO:1904651)	Up	1.17E-03	2.45E-02	-3.698	24.979
positive regulation of glial cell apoptotic process (GO:0034352)	Up	1.17E-03	2.45E-02	-3.694	24.953
positive regulation of leukocyte apoptotic process (GO:2000108)	Up	1.17E-03	2.45E-02	-3.692	24.938

Page intentionally left blank

***Table A4.1**: cont.*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
positive regulation of mesenchymal cell apoptotic process (GO:2001055)	Up	1.17E-03	2.45E-02	-3.691	24.930
positive regulation of muscle cell apoptotic process (GO:0010661)	Up	1.17E-03	2.45E-02	-3.687	24.906
positive regulation of myofibroblast cell apoptotic process (GO:1904522)	Up	1.17E-03	2.45E-02	-3.687	24.905
positive regulation of myeloid cell apoptotic process (GO:0033034)	Up	1.19E-03	2.49E-02	-3.703	24.922
protein autosumoylation (GO:1990466)	Up	1.22E-03	2.51E-02	-2.410	16.164
DNA unwinding involved in DNA replication (GO:0006268)	Up	1.23E-03	2.51E-02	-2.192	14.691
positive regulation of voltage-gated potassium channel activity (GO:1903818)	Up	1.27E-03	2.51E-02	-0.375	2.498
B cell homeostasis (GO:0001782)	Up	1.27E-03	2.51E-02	-0.371	2.477
marginal zone B cell differentiation (GO:0002315)	Up	1.27E-03	2.51E-02	-0.341	2.277
negative regulation of protection from non-homologous end joining at telomere (GO:1905765)	Up	1.27E-03	2.51E-02	-0.169	1.124
mRNA export from nucleus (GO:0006406)	Up	1.27E-03	2.51E-02	-2.787	18.582
positive regulation of fibroblast apoptotic process (GO:2000271)	Up	1.28E-03	2.51E-02	-3.667	24.415
mitochondrial translation (GO:0032543)	Down	8.24E-04	2.53E-02	-3.125	22.193
DNA-dependent DNA replication (GO:0006261)	Up	1.32E-03	2.55E-02	-2.881	19.098
positive regulation of anoikis (GO:2000210)	Up	1.35E-03	2.55E-02	-3.702	24.466
viral genome integration into host DNA (GO:0044826)	Up	1.36E-03	2.55E-02	-2.633	17.373
viral life cycle (GO:0019058)	Up	1.36E-03	2.55E-02	-2.630	17.358
virus maturation (GO:0019075)	Up	1.36E-03	2.55E-02	-2.626	17.330
negative regulation of GTPase activity (GO:0034260)	Up	1.36E-03	2.55E-02	-2.226	14.691
positive regulation of execution phase of apoptosis (GO:1900119)	Up	1.38E-03	2.56E-02	-3.738	24.615
intracellular transport of virus (GO:0075733)	Up	1.44E-03	2.65E-02	-2.518	16.486
protein sumoylation (GO:0016925)	Up	1.51E-03	2.76E-02	-2.469	16.035
viral release from host cell (GO:0019076)	Up	1.59E-03	2.76E-02	-2.693	17.353
regulation of transcription involved in G1/S transition of mitotic cell cycle (GO:0000083)	Up	1.59E-03	2.76E-02	-4.925	31.730
activation of GTPase activity (GO:0090630)	Up	1.61E-03	2.76E-02	-3.179	20.454

Page intentionally left blank

***Table A4.1**: cont.*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
DNA synthesis involved in double-strand break repair via single-strand annealing (GO:0043151)	Up	1.61E-03	2.76E-02	-2.253	14.492
B-1 B cell homeostasis (GO:0001922)	Up	1.62E-03	2.76E-02	-0.350	2.248
negative regulation of telomere capping (GO:1904354)	Up	1.62E-03	2.76E-02	-0.293	1.881
negative regulation of chromosome segregation (GO:0051985)	Up	1.62E-03	2.76E-02	-0.266	1.707
regulation of chromosome segregation (GO:0051983)	Up	1.62E-03	2.76E-02	-0.210	1.351
regulation of mRNA cis splicing, via spliceosome (GO:1905744)	Up	1.62E-03	2.76E-02	-0.204	1.308
regulation of spindle elongation (GO:0032887)	Up	1.62E-03	2.76E-02	-0.193	1.238
positive regulation of apoptotic signaling pathway (GO:2001235)	Up	1.63E-03	2.76E-02	-3.764	24.169
mRNA export from nucleus in response to heat stress (GO:0031990)	Up	1.70E-03	2.86E-02	-2.858	18.218
regulation of signal transduction by p53 class mediator (GO:1901796)	Up	1.70E-03	2.86E-02	-2.814	17.937
regulation of translation (GO:0006417)	Up	1.75E-03	2.90E-02	-2.149	13.639
cognition (GO:0050890)	Up	1.75E-03	2.90E-02	-1.801	11.428
protein phosphorylation (GO:0006468)	Up	1.80E-03	2.96E-02	-4.456	28.163
regulation of intrinsic apoptotic signaling pathway by p53 class mediator (GO:1902253)	Up	1.83E-03	2.99E-02	-2.810	17.718
virion assembly (GO:0019068)	Up	1.85E-03	3.00E-02	-2.624	16.518
positive regulation of ubiquitin-protein ligase activity involved in regulation of mitotic cell cycle transition (GO:0051437)	Down	1.00E-03	3.04E-02	-2.651	18.308
negative regulation of signal transduction by p53 class mediator (GO:1901797)	Up	1.89E-03	3.05E-02	-2.806	17.594
ubiquitin-dependent protein catabolic process (GO:0006511)	Up	1.94E-03	3.11E-02	-2.456	15.343
positive regulation of large conductance calcium-activated potassium channel activity (GO:1902608)	Up	2.02E-03	3.13E-02	-0.697	4.327
regulation of multicellular organism growth (GO:0040014)	Up	2.02E-03	3.13E-02	-0.619	3.841
mature B cell apoptotic process (GO:0002901)	Up	2.02E-03	3.13E-02	-0.539	3.345
replicative senescence (GO:0090399)	Up	2.02E-03	3.13E-02	-0.424	2.630
positive regulation of cysteine-type endopeptidase activity involved in apoptotic process (GO:0043280)	Up	2.03E-03	3.13E-02	-3.733	23.134
telomere capping (GO:0016233)	Up	2.06E-03	3.13E-02	-2.243	13.875

Page intentionally left blank

***Table A4.1**: cont.*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
mRNA splice site selection (GO:0006376)	Up	2.06E-03	3.13E-02	-2.115	13.079
regulation of multivesicular body size involved in ubiquitin-dependent protein catabolism (GO:0010798)	Up	2.13E-03	3.13E-02	-2.459	15.129
chromatin organization involved in negative regulation of transcription (GO:0097549)	Up	2.16E-03	3.13E-02	-4.559	27.979
negative regulation of antisense RNA transcription (GO:0060195)	Up	2.16E-03	3.13E-02	-4.555	27.956
negative regulation of cellular carbohydrate metabolic process by negative regulation of transcription, DNA-templated (GO:0010678)	Up	2.16E-03	3.13E-02	-4.553	27.945
negative regulation of mating-type specific transcription, DNA-templated (GO:0045894)	Up	2.16E-03	3.13E-02	-4.549	27.921
negative regulation of ncRNA transcription associated with protein coding gene TSS/TES (GO:1904388)	Up	2.16E-03	3.13E-02	-4.548	27.914
negative regulation of transcription by pheromones (GO:0045996)	Up	2.16E-03	3.13E-02	-4.546	27.898
negative regulation of transcription during mitotic cell cycle (GO:0007068)	Up	2.16E-03	3.13E-02	-4.541	27.871
negative regulation of transcription from RNA polymerase V promoter (GO:1904280)	Up	2.16E-03	3.13E-02	-4.539	27.860
nitrogen catabolite repression of transcription (GO:0090295)	Up	2.16E-03	3.13E-02	-4.536	27.841
negative regulation of transcription by transcription factor catabolism (GO:0010620)	Up	2.20E-03	3.13E-02	-4.555	27.882
carbon catabolite repression of transcription (GO:0045013)	Up	2.20E-03	3.13E-02	-4.544	27.813
negative regulation of transcription involved in G2/M transition of mitotic cell cycle (GO:0090419)	Up	2.20E-03	3.13E-02	-4.542	27.802
negative regulation of DNA-templated transcription, termination (GO:0060567)	Up	2.23E-03	3.17E-02	-4.570	27.900
negative regulation of transcription by competitive promoter binding (GO:0010944)	Up	2.31E-03	3.22E-02	-4.555	27.659
regulation of DNA damage response, signal transduction by p53 class mediator (GO:0043516)	Up	2.31E-03	3.22E-02	-2.796	16.973
positive regulation of signal transduction by p53 class mediator (GO:1901798)	Up	2.31E-03	3.22E-02	-2.774	16.838
negative regulation of transcription from RNA polymerase I promoter (GO:0016479)	Up	2.42E-03	3.26E-02	-4.567	27.513
regulation of synapse organization by posttranscriptional regulation of gene expression (GO:1904739)	Up	2.46E-03	3.26E-02	-0.615	3.698
positive regulation of miRNA mediated inhibition of translation (GO:1905618)	Up	2.46E-03	3.26E-02	-0.513	3.084
cell differentiation involved in metanephros development (GO:0072202)	Up	2.46E-03	3.26E-02	-0.427	2.564
cell-cell signaling involved in metanephros development (GO:0072204)	Up	2.46E-03	3.26E-02	-0.385	2.313

Page intentionally left blank

***Table A4.1**: cont.*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
metanephric pyramids development (GO:0072211)	Up	2.46E-03	3.26E-02	-0.379	2.275
metanephric juxtaglomerular apparatus development (GO:0072206)	Up	2.46E-03	3.26E-02	-0.372	2.237
metanephric nephron development (GO:0072210)	Up	2.46E-03	3.26E-02	-0.372	2.234
regulation of metanephric glomerulus development (GO:0072298)	Up	2.46E-03	3.26E-02	-0.339	2.035
regulation of transcription from RNA polymerase II promoter involved in metanephros development (GO:0072212)	Up	2.46E-03	3.26E-02	-0.319	1.918
negative regulation of DNA-templated transcription, elongation (GO:0032785)	Up	2.54E-03	3.34E-02	-4.560	27.252
viral genome replication (GO:0019079)	Up	2.55E-03	3.34E-02	-2.458	14.679
negative regulation of transcription from RNA polymerase III promoter (GO:0016480)	Up	2.58E-03	3.36E-02	-4.602	27.433
protein ubiquitination involved in ubiquitin-dependent protein catabolic process (GO:0042787)	Up	2.59E-03	3.36E-02	-3.885	23.140
negative regulation of DNA-templated transcription, initiation (GO:2000143)	Up	2.62E-03	3.37E-02	-4.582	27.242
positive regulation of neuron apoptotic process (GO:0043525)	Up	2.62E-03	3.37E-02	-3.763	22.365
ubiquitin-dependent SMAD protein catabolic process (GO:0030579)	Up	2.78E-03	3.55E-02	-2.431	14.309
histone-serine phosphorylation (GO:0035404)	Up	2.79E-03	3.55E-02	-2.799	16.460
positive regulation of histone deacetylation (GO:0031065)	Up	2.94E-03	3.61E-02	-0.934	5.446
regulation of homologous chromosome segregation (GO:0060629)	Up	2.94E-03	3.61E-02	-0.673	3.921
regulation of BMP signaling pathway involved in heart jogging (GO:2000223)	Up	2.94E-03	3.61E-02	-0.669	3.899
mesenchymal cell apoptotic process involved in metanephros development (GO:1900200)	Up	2.94E-03	3.61E-02	-0.628	3.662
cell migration involved in metanephros development (GO:0035788)	Up	2.94E-03	3.61E-02	-0.462	2.695
metanephric capsule development (GO:0072213)	Up	2.94E-03	3.61E-02	-0.459	2.676
pattern specification involved in metanephros development (GO:0072268)	Up	2.94E-03	3.61E-02	-0.453	2.638
negative regulation of transcription by transcription factor localization (GO:0010621)	Up	2.97E-03	3.63E-02	-4.501	26.202
canonical Wnt signaling pathway involved in positive regulation of apoptotic process (GO:0044337)	Up	3.15E-03	3.83E-02	-3.824	22.034
peptidyl-aspartic acid autophosphorylation (GO:1990938)	Up	3.25E-03	3.88E-02	-3.011	17.252
peptidyl-histidine autophosphorylation (GO:0090593)	Up	3.25E-03	3.88E-02	-3.008	17.238

Page intentionally left blank

***Table A4.1**: cont.*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
protein cis-autophosphorylation (GO:0036291)	Up	3.25E-03	3.88E-02	-3.006	17.227
protein trans-autophosphorylation (GO:0036290)	Up	3.25E-03	3.88E-02	-3.004	17.213
DNA synthesis involved in UV-damage excision repair (GO:1904161)	Up	3.39E-03	4.02E-02	-2.135	12.140
positive regulation of DNA damage response, signal transduction by p53 class mediator resulting in transcription of p21 class mediator (GO:1902164)	Up	3.46E-03	4.02E-02	-0.768	4.355
metanephric cortex development (GO:0072214)	Up	3.46E-03	4.02E-02	-0.729	4.133
positive regulation of telomere maintenance via telomere lengthening (GO:1904358)	Up	3.46E-03	4.02E-02	-0.656	3.717
metanephric smooth muscle tissue development (GO:0072208)	Up	3.46E-03	4.02E-02	-0.649	3.679
negative regulation of circadian sleep/wake cycle, sleep (GO:0042321)	Up	3.46E-03	4.02E-02	-0.616	3.493
ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway (GO:0043162)	Up	3.55E-03	4.11E-02	-2.466	13.909
mRNA localization resulting in posttranscriptional regulation of gene expression (GO:0010609)	Up	4.02E-03	4.61E-02	-1.097	6.051
positive regulation of chromosome segregation (GO:0051984)	Up	4.02E-03	4.61E-02	-0.866	4.779
double-strand break repair via classical nonhomologous end joining (GO:0097680)	Up	4.09E-03	4.65E-02	-2.073	11.402
double-strand break repair via alternative nonhomologous end joining (GO:0097681)	Up	4.09E-03	4.65E-02	-2.031	11.166
positive regulation of apoptotic process (GO:0043065)	Up	4.29E-03	4.86E-02	-4.254	23.191
protein arginylation (GO:0016598)	Up	4.31E-03	4.86E-02	-2.809	15.301

Page intentionally left blank

***Table A4.2**: GO enrichment terms for GP10 molecular function. Up-regulated GO terms denote those of positively-associated differentially expressed genes, whereas down-regulated GO terms are of negatively-associated genes. Only GO terms with adjusted p-value < 0.05 are listed. Combined score used both the result of the Fisher's exact test and the p-value*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
serine-type endopeptidase activity (GO:0004252)	Down	2.35E-24	1.38E-21	-2.814	153.091
Fc-gamma receptor I complex binding (GO:0034988)	Down	6.97E-15	1.36E-12	-2.247	73.263
immunoglobulin receptor binding (GO:0034987)	Down	5.40E-15	1.36E-12	-2.149	70.595
NADH dehydrogenase (ubiquinone) activity (GO:0008137)	Down	1.17E-09	1.71E-07	-1.835	37.737
mRNA binding (GO:0003729)	Up	9.54E-05	1.21E-03	-3.855	35.688
RNA polymerase II carboxy-terminal domain kinase activity (GO:0008353)	Up	4.89E-05	1.21E-03	-3.189	31.652
cyclin-dependent protein serine/threonine kinase activity (GO:0004693)	Up	6.23E-05	1.21E-03	-3.142	30.427
21U-RNA binding (GO:0034583)	Up	1.94E-04	1.21E-03	-3.526	30.135
alpha-aminoacyl-tRNA binding (GO:1904678)	Up	1.94E-04	1.21E-03	-3.520	30.084
base pairing with RNA (GO:0000498)	Up	1.94E-04	1.21E-03	-3.513	30.024
7S RNA binding (GO:0008312)	Up	1.96E-04	1.21E-03	-3.514	29.999
BRE binding (GO:0042835)	Up	1.96E-04	1.21E-03	-3.509	29.956
GU repeat RNA binding (GO:1990605)	Up	1.94E-04	1.21E-03	-3.504	29.946
G-quadruplex RNA binding (GO:0002151)	Up	1.98E-04	1.21E-03	-3.506	29.889
histone pre-mRNA DCP binding (GO:0071208)	Up	1.96E-04	1.21E-03	-3.501	29.888
misfolded RNA binding (GO:0034336)	Up	1.94E-04	1.21E-03	-3.495	29.875
N6-methyladenosine-containing RNA binding (GO:1990247)	Up	1.94E-04	1.21E-03	-3.492	29.848
pre-mRNA binding (GO:0036002)	Up	1.96E-04	1.21E-03	-3.492	29.805
pre-miRNA binding (GO:0070883)	Up	1.98E-04	1.21E-03	-3.495	29.797
regulatory region RNA binding (GO:0001069)	Up	1.94E-04	1.21E-03	-3.485	29.784
snRNA binding (GO:0017069)	Up	2.01E-04	1.21E-03	-3.498	29.783
RNA stem-loop binding (GO:0035613)	Up	1.96E-04	1.21E-03	-3.487	29.765

Page intentionally left blank

***Table A4.2**: cont.*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
piRNA binding (GO:0034584)	Up	2.01E-04	1.21E-03	-3.495	29.757
ribonuclease P RNA binding (GO:0033204)	Up	1.94E-04	1.21E-03	-3.477	29.721
RNA cap binding (GO:0000339)	Up	1.98E-04	1.21E-03	-3.486	29.717
telomeric repeat-containing RNA binding (GO:0061752)	Up	1.96E-04	1.21E-03	-3.477	29.680
RNA binding (GO:0003723)	Up	1.94E-04	1.21E-03	-3.468	29.639
AU-rich element binding (GO:0017091)	Up	2.03E-04	1.21E-03	-3.461	29.425
miRNA binding (GO:0035198)	Up	2.03E-04	1.21E-03	-3.444	29.285
primary miRNA binding (GO:0070878)	Up	2.05E-04	1.21E-03	-3.437	29.185
siRNA binding (GO:0035197)	Up	2.03E-04	1.21E-03	-3.427	29.143
telomerase RNA binding (GO:0070034)	Up	2.07E-04	1.21E-03	-3.404	28.872
snoRNA binding (GO:0030515)	Up	2.10E-04	1.21E-03	-3.406	28.853
rRNA binding (GO:0019843)	Up	2.07E-04	1.21E-03	-3.398	28.815
translation factor activity, RNA binding (GO:0008135)	Up	2.19E-04	1.21E-03	-3.368	28.382
3-phosphoinositide-dependent protein kinase activity (GO:0004676)	Up	2.02E-04	1.21E-03	-2.808	23.888
cyclic nucleotide-dependent protein kinase activity (GO:0004690)	Up	2.02E-04	1.21E-03	-2.800	23.824
Fas-activated serine/threonine kinase activity (GO:0033867)	Up	2.02E-04	1.21E-03	-2.791	23.744
Goodpasture-antigen-binding protein kinase activity (GO:0033868)	Up	2.02E-04	1.21E-03	-2.784	23.684
GTP-dependent protein kinase activity (GO:0034211)	Up	2.02E-04	1.21E-03	-2.775	23.613
Rho-dependent protein serine/threonine kinase activity (GO:0072518)	Up	2.02E-04	1.21E-03	-2.768	23.553
ribosomal protein S6 kinase activity (GO:0004711)	Up	2.02E-04	1.21E-03	-2.764	23.517
protein serine/threonine kinase activity (GO:0004674)	Up	2.13E-04	1.21E-03	-2.780	23.495
calcium-dependent protein serine/threonine kinase activity (GO:0009931)	Up	2.07E-04	1.21E-03	-2.769	23.487
G-protein coupled receptor kinase activity (GO:0004703)	Up	2.07E-04	1.21E-03	-2.763	23.442
eukaryotic translation initiation factor 2alpha kinase activity (GO:0004694)	Up	2.07E-04	1.21E-03	-2.758	23.396
histone threonine kinase activity (GO:0035184)	Up	2.07E-04	1.21E-03	-2.752	23.345

Page intentionally left blank

***Table A4.2**: cont.*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
transmembrane receptor protein serine/threonine kinase activity (GO:0004675)	Up	2.07E-04	1.21E-03	-2.748	23.310
DNA-dependent protein kinase activity (GO:0004677)	Up	2.12E-04	1.21E-03	-2.745	23.217
protein kinase C activity (GO:0004697)	Up	2.12E-04	1.21E-03	-2.740	23.180
histone serine kinase activity (GO:0035174)	Up	2.17E-04	1.21E-03	-2.726	22.989
AMP-activated protein kinase activity (GO:0004679)	Up	2.17E-04	1.21E-03	-2.725	22.982
[acetyl-CoA carboxylase] kinase activity (GO:0050405)	Up	1.66E-04	1.21E-03	-2.632	22.902
[hydroxymethylglutaryl-CoA reductase (NADPH)] kinase activity (GO:0047322)	Up	1.66E-04	1.21E-03	-2.623	22.824
[isocitrate dehydrogenase (NADP+)] kinase activity (GO:0008772)	Up	1.66E-04	1.21E-03	-2.620	22.794
myosin light chain kinase activity (GO:0004687)	Up	2.23E-04	1.21E-03	-2.698	22.685
[tyrosine 3-monooxygenase] kinase activity (GO:0050369)	Up	1.66E-04	1.21E-03	-2.606	22.672
calcium-dependent protein kinase activity (GO:0010857)	Up	1.66E-04	1.21E-03	-2.601	22.636
caldesmon kinase activity (GO:0047764)	Up	1.66E-04	1.21E-03	-2.596	22.586
[3-methyl-2-oxobutanoate dehydrogenase (acetyl-transferring)] kinase activity (GO:0047323)	Up	1.73E-04	1.21E-03	-2.601	22.537
dephospho-[reductase kinase] kinase activity (GO:0047848)	Up	1.66E-04	1.21E-03	-2.588	22.522
low-density-lipoprotein particle receptor kinase activity (GO:0050063)	Up	1.66E-04	1.21E-03	-2.578	22.433
beta-adrenergic receptor kinase activity (GO:0047696)	Up	1.79E-04	1.21E-03	-2.598	22.421
protein arginine kinase activity (GO:1990424)	Up	1.66E-04	1.21E-03	-2.576	22.415
protein histidine kinase activity (GO:0004673)	Up	1.66E-04	1.21E-03	-2.568	22.346
cyclin-dependent protein kinase activity (GO:0097472)	Up	1.79E-04	1.21E-03	-2.585	22.309
protein kinase activity involved in regulation of protein localization to cell division site involved in cytokinesis (GO:1901916)	Up	1.66E-04	1.21E-03	-2.562	22.293
rhodopsin kinase activity (GO:0050254)	Up	1.79E-04	1.21E-03	-2.582	22.277
histone kinase activity (GO:0035173)	Up	1.79E-04	1.21E-03	-2.580	22.263
pyruvate dehydrogenase (acetyl-transferring) kinase activity (GO:0004740)	Up	1.66E-04	1.21E-03	-2.556	22.241
transmembrane receptor protein kinase activity (GO:0019199)	Up	1.66E-04	1.21E-03	-2.551	22.194

Page intentionally left blank

***Table A4.2**: cont.*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
tropomyosin kinase activity (GO:0050359)	Up	1.66E-04	1.21E-03	-2.543	22.124
protein serine/threonine/tyrosine kinase activity (GO:0004712)	Up	2.21E-04	1.21E-03	-2.331	19.622
tau-protein kinase activity (GO:0050321)	Up	2.14E-04	1.21E-03	-2.271	19.194
single-stranded RNA binding (GO:0003727)	Up	2.34E-04	1.25E-03	-3.359	28.083
tRNA binding (GO:0000049)	Up	2.41E-04	1.28E-03	-3.349	27.894
double-stranded RNA binding (GO:0003725)	Up	2.52E-04	1.31E-03	-3.322	27.524
calmodulin-dependent protein kinase activity (GO:0004683)	Up	2.64E-04	1.36E-03	-2.637	21.722
cyclin-dependent protein kinase activating kinase activity (GO:0019912)	Up	3.29E-04	1.67E-03	-2.299	18.437
protein kinase activity (GO:0004672)	Up	5.21E-04	2.61E-03	-2.379	17.982
ATP-dependent DNA helicase activity (GO:0004003)	Up	7.63E-04	3.73E-03	-1.671	11.997
protein tyrosine kinase activity (GO:0004713)	Up	2.27E-03	1.08E-02	-2.387	14.536
mRNA 3'-UTR binding (GO:0003730)	Up	2.79E-03	1.31E-02	-2.031	11.944
ubiquitin-ubiquitin ligase activity (GO:0034450)	Up	3.54E-03	1.63E-02	-1.973	11.131
magnesium ion binding (GO:0000287)	Up	3.54E-03	1.63E-02	-1.966	11.093
transcription coactivator activity (GO:0003713)	Up	4.50E-03	2.04E-02	-2.203	11.905
ligand-dependent nuclear receptor transcription coactivator activity (GO:0030374)	Up	7.87E-03	3.19E-02	-2.174	10.531
RNA polymerase II transcription coactivator activity (GO:0001105)	Up	9.10E-03	3.54E-02	-2.144	10.073
NADH dehydrogenase (quinone) activity (GO:0050136)	Down	4.20E-04	4.92E-02	-1.564	12.158
purine-rich negative regulatory element binding (GO:0032422)	Up	1.41E-02	4.97E-02	-2.065	8.802

Page intentionally left blank

***Table A4.3**: GO enrichment terms for GP23 biological processes. Up-regulated GO terms denote those of positively-associated differentially expressed genes, whereas down-regulated GO terms are of negatively-associated genes. Only GO terms with adjusted p-value < 0.05 are listed. Combined score used both the result of the Fisher's exact test and the p-value*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
complement activation, classical pathway (GO:0006958)	Up	1.10E-67	5.59E-64	70.350	10846.694
humoral immune response mediated by circulating immunoglobulin (GO:0002455)	Up	2.40E-67	6.14E-64	69.233	10620.104
regulation of protein activation cascade (GO:2000257)	Up	1.79E-52	3.04E-49	64.415	7675.267
regulation of complement activation (GO:0030449)	Up	2.55E-52	3.25E-49	63.830	7582.913
regulation of humoral immune response (GO:0002920)	Up	1.01E-51	1.03E-48	61.590	7232.023
regulation of immune effector process (GO:0002697)	Up	1.41E-51	1.20E-48	61.055	7148.666
regulation of acute inflammatory response (GO:0002673)	Up	1.35E-50	9.81E-48	57.551	6608.792
regulation of protein processing (GO:0070613)	Up	1.10E-49	7.05E-47	54.429	6135.556
receptor-mediated endocytosis (GO:0006898)	Up	3.05E-49	1.73E-46	40.527	4527.283
Fc-gamma receptor signaling pathway involved in phagocytosis (GO:0038096)	Up	4.59E-49	2.34E-46	52.398	5832.010
Fc-gamma receptor signaling pathway (GO:0038094)	Up	6.06E-49	2.81E-46	52.009	5774.385
Fc receptor mediated stimulatory signaling pathway (GO:0002431)	Up	7.97E-49	3.39E-46	51.627	5717.730
Fc-epsilon receptor signaling pathway (GO:0038095)	Up	3.98E-44	1.56E-41	38.368	3834.183
Fc receptor signaling pathway (GO:0038093)	Up	4.83E-44	1.76E-41	38.159	3805.916
endocytosis (GO:0006897)	Up	1.20E-43	4.08E-41	29.014	2867.389
regulation of immune response (GO:0050776)	Up	3.01E-39	9.60E-37	27.862	2471.346
regulation of B cell activation (GO:0050864)	Up	3.61E-19	1.08E-16	33.245	1411.782
phagocytosis, engulfment (GO:0006911)	Up	6.07E-17	1.48E-14	28.369	1059.294
plasma membrane invagination (GO:0099024)	Up	5.28E-17	1.50E-14	28.642	1073.471
positive regulation of lymphocyte activation (GO:0051251)	Up	6.07E-17	1.55E-14	28.369	1059.294
B cell receptor signaling pathway (GO:0050853)	Up	6.07E-17	1.63E-14	28.369	1059.294
phagocytosis (GO:0006909)	Up	8.21E-17	1.90E-14	23.467	869.171

Page intentionally left blank

***Table A4.3**: cont.*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
positive regulation of B cell activation (GO:0050871)	Up	4.19E-16	9.29E-14	24.823	878.947
defense response to bacterium (GO:0042742)	Up	2.55E-14	5.42E-12	14.067	440.321
antigen receptor-mediated signaling pathway (GO:0050851)	Up	1.73E-11	3.53E-09	11.545	286.118
glomerular filtration (GO:0003094)	Up	9.40E-08	1.85E-05	85.106	1377.001
renal filtration (GO:0097205)	Up	1.47E-07	2.78E-05	77.369	1217.128
positive regulation of respiratory burst (GO:0060267)	Up	5.53E-06	1.01E-03	79.787	965.779
regulation of respiratory burst (GO:0060263)	Up	1.18E-05	2.07E-03	63.830	724.412
antibacterial humoral response (GO:0019731)	Up	2.40E-05	4.09E-03	23.641	251.459
retina homeostasis (GO:0001895)	Up	3.32E-05	5.46E-03	21.822	225.066
positive regulation of metabolic process (GO:0009893)	Up	7.79E-05	1.24E-02	35.461	335.442

Page intentionally left blank

***Table A4.4**: GO enrichment terms for GP23 molecular functions. Up-regulated GO terms denote those of positively-associated differentially expressed genes, whereas down-regulated GO terms are of negatively-associated genes. Only GO terms with adjusted p-value < 0.05 are listed. Combined score used both the result of the Fisher's exact test and the p-value*

GO Enrichment Term	Regulation	P-value	Adjusted P-value	Z-score	Combined Score
serine-type endopeptidase activity (GO:0004252)	Up	5.48E-43	6.31E-40	35.641	3468.217
serine-type peptidase activity (GO:0008236)	Up	3.14E-41	1.81E-38	31.770	2962.985
endopeptidase activity (GO:0004175)	Up	3.48E-34	1.33E-31	19.778	1523.760
immunoglobulin receptor binding (GO:0034987)	Up	2.89E-20	8.33E-18	38.921	1750.994

Page intentionally left blank

Appendix 5 – Univariate Correlation Analyses of Various Clinical Factors and IgG N-Glycosylation

A supplement to the research paper described in **Section 5.3**, edited here for addition into this thesis. Correlation analysis using Spearman's Rho and including *p*-values adjusted for the 5% false discovery rate.

Page intentionally left blank

***Table A5.1**: Associations of age- and sex-corrected IgG GPs to several clinical factors using Spearman's correlation coefficient*

	Height		Weight		Waist		Hip		WCC		TC		TG	
GP1	0.0004	0.9968	-0.0349	0.6140	-0.0230	0.7794	-0.0217	0.7962	-0.0446	0.5013	0.0891	0.1338	0.0159	0.8350
GP2	-0.0084	0.9097	-0.0063	0.9330	0.0295	0.6891	-0.0203	0.8018	-0.0412	0.5375	0.0698	0.2533	0.0532	0.4068
GP3	-0.0085	0.9097	0.0777	0.2068	0.1099	0.0519	0.0867	0.1496	0.0063	0.9330	0.0180	0.8085	0.0836	0.1646
GP4	-0.0186	0.8085	0.0921	0.1242	0.1264	**0.0287**	0.1138	**0.0424**	-0.0182	0.8085	0.0872	0.1480	0.1480	**0.0094**
GP5	-0.0336	0.6298	-0.0656	0.2855	-0.0714	0.2448	-0.0388	0.5600	0.0023	0.9744	0.1234	**0.0287**	0.0198	0.8078
GP6	-0.0643	0.2979	0.0470	0.4723	0.0985	0.0950	0.0613	0.3164	-0.0219	0.7954	0.1166	**0.0372**	0.1615	**0.0036**
GP7	-0.0368	0.5877	-0.0898	0.1309	-0.0664	0.2834	-0.0995	0.0904	-0.0441	0.5070	0.0276	0.7198	-0.0458	0.4831
GP8	0.0120	0.8754	0.0597	0.3316	0.0273	0.7198	0.0626	0.3089	-0.1027	0.0785	0.0284	0.7066	-0.0060	0.9330
GP9	-0.0070	0.9277	0.0606	0.3227	0.0573	0.3625	0.0525	0.4145	0.0249	0.7520	-0.0454	0.4877	0.0130	0.8697
GP10	-0.0215	0.7962	0.0672	0.2765	0.0701	0.2533	0.0600	0.3312	-0.0210	0.8000	-0.0048	0.9464	0.0611	0.3178
GP11	-0.0706	0.2502	0.0904	0.1295	0.1230	**0.0287**	0.1027	0.0785	0.0607	0.3227	0.0215	0.7962	0.1392	**0.0146**
GP12	-0.0180	0.8085	-0.0817	0.1793	-0.0839	0.1633	-0.0945	0.1132	-0.0113	0.8843	-0.0730	0.2318	-0.1238	**0.0287**
GP13	-0.0010	0.9901	-0.0870	0.1482	-0.1226	**0.0287**	-0.0713	0.2448	-0.0389	0.5600	-0.0407	0.5443	-0.1639	**0.0031**
GP14	0.0286	0.7044	-0.0226	0.7859	-0.0638	0.2988	-0.0511	0.4263	-0.0071	0.9273	-0.1053	0.0677	-0.1375	**0.0160**
GP15	0.0207	0.8000	-0.0119	0.8754	-0.0291	0.6958	-0.0423	0.5283	0.0170	0.8185	-0.0802	0.1902	-0.0756	0.2158
GP16	0.0267	0.7280	0.0770	0.2115	0.0794	0.1937	0.0943	0.1132	0.0427	0.5239	-0.0382	0.5675	0.0438	0.5095
GP17	0.0212	0.7995	-0.0639	0.2988	-0.0763	0.2136	-0.0682	0.2682	0.0630	0.3062	-0.0598	0.3314	-0.1331	**0.0196**
GP18	0.0387	0.5600	-0.0273	0.7198	-0.0642	0.2981	-0.0295	0.6891	-0.0058	0.9330	-0.1193	0.0321	-0.1105	0.0504
GP19	-0.0188	0.8085	-0.0192	0.8085	-0.0274	0.7198	-0.0072	0.9268	0.0483	0.4526	0.0108	0.8875	-0.0118	0.8754
GP20	0.0244	0.7560	-0.0228	0.7830	-0.0618	0.3125	0.0022	0.9744	-0.0556	0.3837	-0.0369	0.5877	-0.0709	0.2468
GP21	0.0188	0.8085	0.0270	0.7244	0.0153	0.8404	-0.0040	0.9591	-0.0051	0.9426	0.0514	0.4235	-0.0377	0.5768
GP22	0.0373	0.5836	-0.0400	0.5482	-0.0549	0.3894	-0.0858	0.1550	0.0275	0.7198	0.0215	0.7962	-0.0259	0.7384
GP23	0.0203	0.8018	0.0118	0.8754	-0.0060	0.9330	0.0307	0.6760	0.0024	0.9744	-0.0105	0.8914	-0.0257	0.7388
GP24	-0.0172	0.8183	-0.0068	0.9292	-0.0010	0.9901	0.0148	0.8482	0.0257	0.7388	-0.0026	0.9744	-0.0062	0.9330

Page intentionally left blank

Table A5.1 *cont.*

	FBG		CRP		HDL		LDL		W2H		TotalFat%		A/G	
GP1	0.0143	0.8521	0.0421	0.5310	-0.0171	0.8183	0.1201	**0.0312**	-0.0314	0.6675	0.0104	0.8914	0.0120	0.8754
GP2	0.0493	0.4500	0.0093	0.9037	-0.0149	0.8469	0.0489	0.4500	0.0552	0.3867	-0.0123	0.8751	0.0344	0.6205
GP3	0.0947	0.1132	0.1699	**0.0029**	-0.0845	0.1590	0.0480	0.4555	0.0545	0.3917	0.0922	0.1242	0.1210	**0.0305**
GP4	0.0906	0.1295	0.2008	**0.0002**	-0.0904	0.1295	0.1090	0.0546	0.0430	0.5217	0.1372	**0.0160**	0.1449	**0.0111**
GP5	0.0025	0.9744	0.0698	0.2533	0.0157	0.8366	0.1234	**0.0287**	-0.0747	0.2214	0.0139	0.8538	-0.0405	0.5458
GP6	0.1240	**0.0287**	0.1233	**0.0287**	-0.0658	0.2848	0.1131	**0.0433**	0.0682	0.2682	0.1204	**0.0312**	0.1170	**0.0372**
GP7	-0.0143	0.8521	-0.0815	0.1806	0.0324	0.6501	0.0077	0.9217	0.0006	0.9958	-0.0677	0.2722	-0.0623	0.3089
GP8	-0.0743	0.2216	0.0406	0.5445	0.0099	0.8953	0.0396	0.5509	-0.0064	0.9330	0.0529	0.4087	0.0157	0.8366
GP9	-0.0195	0.8085	0.0628	0.3080	-0.0745	0.2214	-0.0129	0.8697	0.0203	0.8018	0.0500	0.4397	0.0931	0.1207
GP10	0.0514	0.4235	-0.0312	0.6681	-0.0486	0.4515	-0.0037	0.9609	0.0658	0.2848	0.0697	0.2533	0.0645	0.2972
GP11	0.0857	0.1550	0.0779	0.2067	-0.1007	0.0849	0.0353	0.6094	0.0515	0.4235	0.1199	**0.0312**	0.1262	**0.0287**
GP12	-0.0590	0.3385	-0.1397	**0.0146**	0.0429	0.5221	-0.0826	0.1732	-0.0236	0.7727	-0.0942	0.1132	-0.1231	**0.0287**
GP13	-0.1026	0.0785	-0.1071	0.0617	0.0900	0.1309	-0.0459	0.4831	-0.0921	0.1242	-0.0755	0.2158	-0.1234	**0.0287**
GP14	-0.1272	**0.0284**	-0.1520	**0.0079**	0.0399	0.5489	-0.0898	0.1309	-0.0186	0.8085	-0.0895	0.1317	-0.0764	0.2136
GP15	-0.0504	0.4349	-0.1534	**0.0075**	-0.0041	0.9578	-0.0732	0.2314	0.0198	0.8078	-0.0680	0.2693	-0.0520	0.4217
GP16	-0.0186	0.8085	0.0546	0.3914	-0.0516	0.4235	-0.0276	0.7198	0.0136	0.8583	0.0388	0.5600	0.0740	0.2238
GP17	-0.0565	0.3702	-0.0489	0.4500	0.0416	0.5330	-0.0619	0.3121	-0.0400	0.5482	-0.1126	**0.0442**	-0.1166	**0.0372**
GP18	-0.1356	**0.0164**	-0.1453	**0.0111**	0.0359	0.6002	-0.1133	**0.0432**	-0.0491	0.4500	-0.0796	0.1937	-0.0956	0.1105
GP19	-0.0348	0.6140	0.0123	0.8751	0.0297	0.6890	-0.0051	0.9426	-0.0350	0.6140	0.0003	0.9968	-0.0188	0.8085
GP20	-0.0457	0.4848	-0.0598	0.3314	0.0652	0.2899	-0.0404	0.5458	-0.0732	0.2314	-0.0330	0.6407	-0.0394	0.5532
GP21	-0.0465	0.4761	0.0368	0.5877	0.0162	0.8315	0.0483	0.4526	0.0338	0.6277	0.0087	0.9081	-0.0102	0.8929
GP22	-0.0127	0.8714	-0.0662	0.2840	0.0783	0.2045	-0.0230	0.7794	0.0194	0.8085	-0.0945	0.1132	-0.0671	0.2765
GP23	-0.0703	0.2522	0.0416	0.5330	0.0533	0.4068	-0.0183	0.8085	-0.0464	0.4763	0.0176	0.8135	-0.0206	0.8000
GP24	-0.0057	0.9330	0.0504	0.4349	0.0590	0.3385	-0.0260	0.7384	-0.0182	0.8085	0.0305	0.6784	-0.0098	0.8953

Page intentionally left blank

Table A5.1 *cont.*

	TotalBone%		TotalLean%		BMI		Insulin		W2He		SBP		DBP	
GP1	-0.0251	0.7487	-0.0091	0.9055	-0.0533	0.4068	0.0441	0.5070	-0.0320	0.6576	-0.0058	0.9330	-0.0235	0.7727
GP2	-0.0554	0.3853	0.0168	0.8214	0.0181	0.8085	0.0571	0.3625	0.0346	0.6166	0.0153	0.8404	0.0197	0.8078
GP3	-0.1000	0.0879	-0.0915	0.1267	0.0849	0.1570	0.1505	**0.0079**	0.1154	**0.0387**	0.0397	0.5503	0.0244	0.7560
GP4	-0.1303	**0.0227**	-0.1365	**0.0163**	0.1219	**0.0289**	0.2162	**0.0000**	0.1429	**0.0126**	0.1021	0.0797	0.0697	0.2533
GP5	-0.0635	0.3028	-0.0110	0.8875	-0.0777	0.2068	-0.0419	0.5330	-0.0714	0.2448	-0.0427	0.5239	-0.0625	0.3089
GP6	-0.1663	**0.0031**	-0.1155	**0.0387**	0.1069	0.0617	0.1729	**0.0026**	0.1322	**0.0205**	0.0664	0.2834	0.0529	0.4087
GP7	0.0037	0.9609	0.0713	0.2448	-0.0674	0.2753	-0.0485	0.4515	-0.0581	0.3523	-0.0495	0.4481	-0.0413	0.5360
GP8	-0.0443	0.5058	-0.0542	0.3950	0.0508	0.4299	-0.0125	0.8740	0.0210	0.8000	0.0259	0.7384	0.0417	0.5330
GP9	-0.0311	0.6695	-0.0515	0.4235	0.0631	0.3062	0.0436	0.5128	0.0624	0.3089	0.0812	0.1825	0.0762	0.2136
GP10	-0.0765	0.2136	-0.0695	0.2541	0.1014	0.0818	0.0315	0.6657	0.0856	0.1550	0.0108	0.8875	0.0140	0.8538
GP11	-0.1221	**0.0289**	-0.1190	**0.0323**	0.1405	**0.0143**	0.1228	**0.0287**	0.1510	**0.0079**	0.0757	0.2158	0.0490	0.4500
GP12	0.0770	0.2115	0.0944	0.1132	-0.0756	0.2158	-0.1248	**0.0287**	-0.0851	0.1570	-0.1020	0.0797	-0.0822	0.1760
GP13	0.0724	0.2383	0.0748	0.2214	-0.1141	**0.0422**	-0.1651	**0.0031**	-0.1318	**0.0206**	-0.0968	0.1043	-0.0910	0.1292
GP14	0.1225	**0.0287**	0.0861	0.1540	-0.0473	0.4671	-0.1814	**0.0013**	-0.0764	0.2136	-0.0617	0.3125	-0.0221	0.7912
GP15	0.0798	0.1930	0.0661	0.2840	-0.0245	0.7560	-0.1361	**0.0163**	-0.0356	0.6051	-0.0851	0.1570	-0.0465	0.4761
GP16	0.0072	0.9268	-0.0416	0.5330	0.0649	0.2919	0.0487	0.4515	0.0752	0.2190	0.0844	0.1590	0.0638	0.2988
GP17	0.1066	0.0623	0.1117	0.0467	-0.0909	0.1292	-0.1261	**0.0287**	-0.0920	0.1242	-0.1162	**0.0375**	-0.0958	0.1104
GP18	0.1408	**0.0143**	0.0746	0.2214	-0.0548	0.3904	-0.1664	**0.0031**	-0.0793	0.1937	-0.0538	0.4007	-0.0324	0.6501
GP19	-0.0339	0.6275	0.0014	0.9862	-0.0167	0.8225	-0.0145	0.8521	-0.0207	0.8000	0.0140	0.8538	-0.0016	0.9836
GP20	0.0716	0.2448	0.0298	0.6890	-0.0803	0.1902	-0.0657	0.2848	-0.0886	0.1368	-0.0300	0.6869	-0.0466	0.4761
GP21	0.0096	0.8977	-0.0101	0.8929	0.0029	0.9737	0.0185	0.8085	0.0072	0.9268	-0.0022	0.9744	-0.0181	0.8085
GP22	0.0366	0.5903	0.0976	0.1003	-0.0572	0.3625	-0.0622	0.3099	-0.0712	0.2448	-0.0565	0.3702	-0.0222	0.7912
GP23	0.0401	0.5482	-0.0208	0.8000	-0.0003	0.9968	-0.0266	0.7292	-0.0108	0.8875	0.0196	0.8085	0.0002	0.9968
GP24	-0.0360	0.6002	-0.0296	0.6890	0.0028	0.9737	0.0175	0.8135	0.0087	0.9081	0.0185	0.8085	0.0087	0.9081

Page intentionally left blank