Edith Cowan University
Research Online

ECU Publications Post 2013

2018

Type 2 diabetes mellitus: Integrative analysis of multiomics data for biomarker discovery

Siqi Ge
Edith Cowan University

Youxin Wang
Edith Cowan University

Song Manshu
Edith Cowan University, m.song@ecu.edu.au

Xingang Li
Edith Cowan University, xingang.li@ecu.edu.au

Xinwei Yu
Edith Cowan University

See next page for additional authors

10.1089/omi.2018.0053
Originally published as: Ge, S., Wang, Y., Song, M., Li, X., Yu, X., Wang, H., ... & Wang, W. (2018). Type 2 Diabetes Mellitus: Integrative Analysis of Multiomics Data for Biomarker Discovery. *Omics: a journal of integrative biology*, *22*(7), 514-523. Original article available here.
This Journal Article is posted at Research Online.
https://ro.ecu.edu.au/ecuworkspost2013/4539

Authors
Siqi Ge, Youxin Wang, Song Manshu, Xingang Li, Xinwei Yu, Hao Wang, Jing Wang, Qiang Zeng, and Wei Wang

This journal article is available at Research Online: https://ro.ecu.edu.au/ecuworkspost2013/4539

OMICS A Journal of Integrative Biology
Volume 22, Number 7, 2018
© Mary Ann Liebert, Inc.
DOI: 10.1089/omi.2018.0053

Type 2 Diabetes Mellitus: Integrative Analysis of Multiomics Data for Biomarker Discovery

Siqi Ge,[1,2] Youxin Wang,[1,2] Manshu Song,[1,2] Xingang Li,[2] Xinwei Yu,[1,2]
Hao Wang,[1,2] Jing Wang,[3] Qiang Zeng,[4] and Wei Wang[1,2]

Downloaded by Edith Cowan University from www.liebertpub.com at 10/23/18. For personal use only.

Abstract

Increased fasting plasma glucose (FPG) is an independent risk factor for type 2 diabetes mellitus (T2DM). The development of systems biology technologies for integration of multiomics data is crucial for predicting increased FPG levels. In this case–control study, immunoglobulin (Ig) G glycosylation profiling and genome-wide association analyses were performed on 511 participants, and among them 76 had increased FPG (aged 47.6 ± 6.14 years), and 435 had decreased or fluctuant FPG (aged 47.9 ± 6.08 years). We identified nine single nucleotide polymorphisms (SNPs) in five genes (*RPL7AP27*, *SNX30*, *SLC39A12*, *BACE2*, and *IGFL2*) that were significantly associated with increased FPG (odds ratios 1.937–2.393). Moreover, of the 24 glycan peaks (GPs), GPs 3, 8, and 11 presented positive trends with increased FPG levels, whereas GPs 4 and 14 presented negative trends. A significant improvement of predictive power was observed when adding 24 IgG GPs to 9 SNPs with the area under the curve increased from 0.75 to 0.81. This report shows that the combination of candidate SNPs with IgG glycomics offers biomarker potentials for T2DM. The substantial predictive power obtained from integrating genomics and glycomics biomarkers suggests the feasibility of applying such multiomics strategies to enable predictive, preventive, and personalized medicine for T2DM.

Keywords: T2DM, multiomics, genome-wide association study, IgG glycosylation, biomarker discovery, predictive models

Introduction

Type 2 diabetes mellitus (T2DM) has been identified as a major international health challenge which has a great worldwide impact on morbidity, premature mortality, and economic burden (DeFronzo et al., 2015; Ginter and Simko, 2012). It is a multifactorial disease with a strong genetic component as well as environmental factors contributing to its pathogenesis (Ding et al., 2015). Studies have shown that elevated fasting plasma glucose (FPG) level is an independent risk factor for T2DM development (Mozaffary et al., 2016). Moreover, the longitudinal changes of FPG level are stronger predictors of T2DM development than the measure of cross-sectional level of FPG (Liu et al., 2014).

In the past decade, genetic association studies and family linkage analysis, and genome-wide association studies (GWASs) have facilitated the identification of genetic susceptibility for FPG (Manolio, 2010). To date, 39 susceptible loci associated with the increased level of FPG have been identified in different populations (Dupuis et al., 2010; Manning et al., 2012). However, the most reported variants capture only 10% of familial aggregation of the disease susceptibility (Morris et al., 2012). The unexplained genetic variance of increased FPG level remains far from clear. Meanwhile, few of the single nucleotide polymorphisms (SNPs) identified in GWAS have clear functional implications that are relevant to pathogenesis of T2DM and its clinical process (Hindorff et al., 2009; Manolio, 2010).

Therefore, understanding the substantial roles of these candidate SNPs and their functions involved in T2DM development from the perspectives of posttranscriptional modifications (such as glycomics) is of vital importance for translational research on T2DM.

Glycosylation of proteins is one of the most common posttranslational modifications (Cummings and Pierce, 2014; Wang et al., 2016b). Unlike proteins, glycan biosynthesis does not require a template directly; rather, the structure and function of glycans could be influenced by changing environmental conditions, aging, and the presence of disease (Russell et al., 2017; Wang et al., 2016a; Yu et al., 2016). These features

[1]Beijing Key Laboratory of Clinical Epidemiology, Department of Epidemiology and Health Statistics, School of Public Health, Capital Medical University, Beijing, China.
[2]Department of Public Health, School of Medical and Health Sciences, Edith Cowan University, Perth, Australia.
[3]Department of Pathophysiology, Peking Union Medical College, China Academy of Medical Sciences, Beijing, China.
[4]Department of International Inpatient, Institute of Health Management, Chinese PLA General Hospital, Beijing, China.

Downloaded by Edith Cowan University from www.liebertpub.com at 10/23/18. For personal use only.

provide glycans the potential to serve as dynamic biomarkers in complex disease such as T2DM (Adua et al., 2017; Russell et al., 2018). Immunoglobulin (Ig) is a cluster of glycoproteins with antibody activity (including IgA, IgD, IgE, IgG, and IgM), accounting for $\sim$20% of plasma proteins. IgG is the most abundant (>80%), the major antibody in the secondary humoral response of immunity (Junqueira et al., 1998).

The glycosylation of IgG is particularly interesting, as *N*-glycans attached to the fragment crystallizable part of IgG can modulate and switch its function from pro- to anti-inflammatory and vice versa (Anthony et al., 2008; Gornik et al., 2012). To date, the levels and structures of IgG glycosylation have been found associated with the incidence and development of many inflammatory diseases and processes, such as aging, hypertension, inflammatory bowel disease, Parkinson's disease, and systemic lupus erythematosus (Russell et al., 2017; Trbojevic Akmacic et al., 2015; Vuckovic et al., 2015; Wang et al., 2016a; Yu et al., 2016).

Our previous studies have found that specific *N*-glycan structural features were significantly correlated with FPG level (Lu et al., 2011). Moreover, certain patterns of change in IgG glycans were found with FPG level, hinting the significant role of IgG glycans in T2DM pathophysiology (Keser et al., 2017; Lemmers et al., 2017). However, these studies only evidenced the correlations between IgG glycans and increased FPG levels, and could not reflect whether these are reactive processes or causative ones.

The combination of different omics is referred to as multiple omics, multiomics, or integrated omics (Hasin et al., 2017). As compared with studies of a single omics type, multiomics offers the opportunity to understand the flow of information that underlies disease, for example, T2DM. Such attempt has been made by Parks et al. (2013). By examining global gene expression in multiple tissues and metabolites in plasma, they were able to identify pathways and genes contributing to diet-induced obesity and diabetes in mice. In addition, transcriptomics and protein–protein interaction data were used to identify molecular and tissue-specific biomarker signatures of T2DM (Calimlioglu et al., 2015). However, these studies have limited generalization and application potential due to the limitations of their study designs.

Thus, this study provides a rationale for an integrative analysis on genetics and glycomics at multiomics level to decipher the pathogenesis of T2DM and to identify molecular patterns associated with increased FPG level. We hypothesized that the adding of glycomic profiling to genetic genotyping would further explain the process of T2DM from multiple prospective, including genetic risk, and environmental conditions. Consequently, the combination of glycomic and genetic biomarkers will improve the predictive ability for T2DM and the increase of FPG level. In this study, we combined the IgG glycome with GWAS, and investigated the potential glycosylation patterns of IgG and candidate genes that lead to increased FPG levels among Chinese populations. We also compared the predictive power of candidate genes and IgG glycans individually and conjunctively for increased FPG level.

Materials and Methods

Study design and participants

This was a community-based case–control study, performed on a representative sample of an urban population of Beijing, China. Participants were recruited from Beijing *XuanWu* community, who had completed three health checkups and questionnaire surveys in years 2012, 2013, and 2014. Adults (aged >18 years) with negative history of diabetes who had completed information from health checkups were eligible for inclusion. We excluded individuals diagnosed with diabetes before or during health checkups, and those with documented diabetes, pregnant women, and individuals taking medications during or 3 months before the recruitment.

A total of 511 participants were eligible for this study, with a mean age of 47.9 (25–73) years, comprising 159 males (31.1%) and 352 females (68.9%). This study was approved by the Ethics Committees of Capital Medical University, Beijing, China and Edith Cowan University, Perth, Australia. Each participant included in this study signed informed consent.

Key variables

Phenotype grouping. Taking the advantage of the longitudinal design of this study, instead of dividing the participants into diabetic/nondiabetic groups, we used the changes in FPG levels as the categorical reference, an independent risk factor with potential predictive and preventive application for T2DM (Mozaffary et al., 2016). Consequently, of 511 participants with longitudinal follow-ups, 76 had increased FPG levels (increased FPG group), and 435 had decreased or fluctuant FPG levels (no increased FPG group) over a period of 3 years.

Genome-wide genotyping and quality control. The genotyping procedures were conducted with Illumina Omni *Zhonghua* chips (Illumina, Inc., San Diego, CA, USA). The standard quality check (QC) procedures contained the following criteria: We excluded all the SNPs with an overall call rate <97%, minor allele frequency <1%, Hardy–Weinberg equilibrium *p*-value in control subjects $<1.0\times10^{-8}$. Finally, we selected 69,485 eligible SNPs for further analyses (see Supplementary Table).

IgG *N*-glycosylation. Glycan analysis was performed in *Genos* Glycoscience Research laboratory in Zagreb. IgG *N*-glycosylation analyses were conducted on all 511 participants. IgG glycan isolation, release, and labeling were executed as described previously (Yu et al., 2016).

Briefly, IgG was isolated from plasma using 96-well protein G monolithic plates, eluted with 0.1 M formic acid, and neutralized with 1 M ammonium bicarbonate. Dried IgG samples (150–200 μg) were denatured (with 30 μL sodium dodecyl sulfate (SDS)), incubated at 65°C for 10 min, and cooled to room temperature, followed by addition of 10 μL of Igepal-CA630 ($\varphi=4\%$). *N*-glycans were released with the addition of 1.2 U of PNGase F in 10 μL 5$\times$phosphate-buffered saline and incubation at 37°C for 18 h. Released *N*-glycans were labeled with 2-AB.

The samples were cleaned using hydrophilic interaction liquid chromatography solid phase extraction and stored at –20°C until ultra-performance liquid chromatography (UPLC). Finally, 24 IgG glycan peaks (GPs) were measured by Waters Acquity UPLC instrument as described previously (Pucic et al., 2011). All chromatograms were separated in the same manner into 24 peaks, and the amount of glycans in each peak was expressed as percentage of total integrated area (see Supplementary Table).

Downloaded by Edith Cowan University from www.liebertpub.com at 10/23/18. For personal use only.

Covariates. We selected and measured the known T2DM risk factors as covariates, including age, gender, systolic (SBP) and diastolic (DBP) blood pressures, body mass index (BMI), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), and waist-to-hip ratio (WHR; Wilson et al., 2007).

Statistical analyses

The distribution of the IgG GPs was tested for normality, then the transformation which performed best normality was used. Student's *t*-tests and chi-square tests were applied to test differences between two groups (increased FPG group vs. no increased FPG group) for continuous and categorical variables, respectively. Pearson's correlation analyses were applied to assess the correlations of clinical risk factors with IgG GPs. A significant reference of $p < 0.05$ was applied.

Genome-wide associations were performed and adjusted for age, gender, BMI, SBP, DBP, HDL-C, LDL-C, and WHR. The *mmscore* function of the GenABEL package in R statistical software (Aulchenko et al., 2007) was used for the association test. An association was considered statistically significant at the genome-wide level if the *p*-value for an individual SNP was $<5 \times 10^{-5}$ (based on Bonferroni correction to account for multiple testing).

Heat-map analyses were performed using *heatmap.2* function of the gplots package in R statistical software. Heat maps were used for clustering correlations between significant SNPs and IgG GPs.

The predictive assessment of SNPs and IgG glycans for increased FPG level was performed using receiver operator characteristic (ROC) analyses and the area under the curves (AUCs) analyses. We fitted logistic regression models with SPSS 23.0 (IBM Corp., New York, NY, USA) for the evaluation and comparison of the AUCs between models containing SNPs, IgG glycans, and both, with adjustment for above-mentioned clinical covariates. All continuous variables were standardized before they were included in the models for AUC calculation.

Table 1. Characteristics of Participants With/Without Increased Fasting Plasma Glucose Level at Baseline

Clinical factors	*Increased FPG*[a] (n = *76*)	*Decreased or fluctuant FPG*[a] (n = *435*)	p
Age	47.62 ± 6.14	47.93 ± 6.08	0.380
Gender (male/female)	23/53	136/299	0.080
BMI	24.63 ± 3.06	24.51 ± 3.08	0.948
SBP	119.79 ± 14.34	117.62 ± 13.95	0.213
DBP	80.30 ± 11.09	78.94 ± 10.41	0.309
WHR	0.80 ± 0.07	0.81 ± 0.07	0.670
HDL-C	1.73 ± 0.34	1.63 ± 0.33	0.288
LDL-C	2.76 ± 0.72	2.78 ± 0.71	0.494

Unless stated otherwise, mean (±SD) are given.

[a]Indicating participants with increased FPG levels and participants with decreased or fluctuant FPG levels.

BMI, body mass index; DBP, diastolic blood pressure; FPG, fasting plasma glucose; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; SBP, systolic blood pressure; SD, standard deviation; WHR, waist-to-hip ratio.

Results

Cohort characteristics

The characteristics of 511 participants, where 76 had increased FPG level (aged 47.6 ± 6.14 years), and 435 had decreased or fluctuant FPG levels (aged 47.9 ± 6.08 years), >3 years are shown in Table 1. No significant differences were shown between these two groups in age, gender, BMI, SBP, DBP, WHR, HDL-C, and LDL-C ($p > 0.05$; Table 1).

FPG level associated with IgG N-glycosylation

In total, 24 IgG GPs were observed (Table 2). Correlations of 24 IgG GPs with increased FPG levels are shown in Table 2. Five IgG GPs (GPs 3, 4, 8, 11, and 14) showed significant associations with the increased FPG level. GPs 3, 8, and 11 were found to present positive trends with increased FPG (*p*-values range from 0.003 to 0.014), whereas GPs 4 and 14 presented negative trends of associations (*p*-values 0.007–0.041).

GWAS analyses for increased FPG level

The distribution of *p*-values for the association of SNPs with increased FPG levels is shown in Fig. 1. Consequently, among the 69,485 eligible SNPs, 199 were found significantly associated with increased FPG levels ($p < 0.001$). Nine of these SNPs (rs7686384, rs10739371, rs2153839, rs10817396, rs2497753, rs9418383, rs4802289, rs2252576, and rs2837981) located in five respective chromosomal regions (4q35.2, 9q32, 10p12.33, 19q13.32, and 21q22.3) met the predefined criteria, and showed strong signals with increased FPG levels (*p*-values 3.35×10^{-6}–7.71×10^{-7}) (Table 3). After adjusting for age, gender, BMI, SBP, DBP, HDL-C, LDL-C, and WHR, 9 SNPs showed significantly increased risk with odds ratios (ORs) ranging from 1.937 to 2.393.

Heat-map analysis

By applying clustering heat-map analysis, we analyzed the correlation between the 9 SNPs and 24 IgG GPs, and found significant associations and explicit clustering patterns within both SNPs and IgG GPs (Fig. 2).

The risk allele "A" of rs10739371, "G" of rs2153839, and "A" of rs10817396 presented the same correlation patterns with decreased galactosylated glycan structures (GPs 11, 12, 14, and 15) and increased agalactosylated glycan structures (GPs 1 and 3–5) (Fig. 2). Rs2497753 and rs9418383 presented significant increased correlation with neutral digalactosylated structure (GPs 14 and 15), and decreased correlation with sialylated digalactosylated structure (Fig. 2). We also found that two SNPs (rs2252576 and rs2837981) were correlated with increase in neutral glycan structures (GPs 4 and 8) and decrease in sialylated glycan structures (GPs 16 and 20–22) (Fig. 2).

Area under the ROC curves for increased FPG level

In terms of predictive power for increased FPG level, the mean AUC for the 9 SNPs without 24 IgG GPs was 0.75, and the mean AUC for the 9 SNPs with 24 IgG GPs was increased to 0.81. As shown in Table 4, adding the 24 IgG GPs to 9 SNPs resulted in significant improvement of AUC in all models. Figure 3 shows the comparisons of ROC curves for

Table 2. Correlations of 24 Immunoglobulin G Glycan Peaks with Fasting Plasma Glucose Levels

Glycan	*Description*	*Structure*	*Increased FPG*
GP1	*The percentage of FA1 glycan in total IgG glycans*		NS
GP2	*The percentage of A2 glycan in total IgG glycans*		NS
GP3	*The percentage of A2B glycan in total IgG glycans*		↑
GP4	*The percentage of FA2 glycan in total IgG glycans*		↓
GP5	*The percentage of M5 glycan in total IgG glycans*		NS
GP6	*The percentage of FA2B glycan in total IgG glycans*		NS
GP7	*The percentage of A2G1 glycan in total IgG glycans*		NS
GP8	*The percentage of FA2[6]G1 glycan in total IgG glycans*		↑
GP9	*The percentage of FA2[3]G1 glycan in total IgG glycans*		NS
GP10	*The percentage of FA2[6]BG1 glycan in total IgG glycans*		NS
GP11	*The percentage of FA2[3]BG1 glycan in total IgG glycans*		↑
GP12	*The percentage of A2G2 glycan in total IgG glycans*		NS
GP13	*The percentage of A2BG2 glycan in total IgG glycans*		NS
GP14	*The percentage of FA2G2 glycan in total IgG glycans*		↓
GP15	*The percentage of FA2BG2 glycan in total IgG glycans*		NS
GP16	*The percentage of FA2G1S1 glycan in total IgG glycans*		NS
GP17	*The percentage of A2G2S1 glycan in total IgG glycans*		NS
GP18	*The percentage of FA2G2S1 glycan in total IgG glycans*		NS
GP19	*The percentage of FA2BG2S1 glycan in total IgG glycans*		NS
GP20	*The percentage of FA2FG2S1 glycan in total IgG glycans*		NS
GP21	*The percentage of A2G2S2 glycan in total IgG glycans*		NS
GP22	*The percentage of A2BG2S2 glycan in total IgG glycans*		NS
GP23	*The percentage of FA2G2S2 glycan in total IgG glycans*		NS
GP24	*The percentage of FA2BG2S2 glycan in total IgG glycans*		NS

↑ indicated IgG glycan peaks were higher with the increase of FPG level, while ↓ indicated a negative correlation. NS, not significant. Structure abbreviations: F, α-1,6-linked core fucose; A, number of antennas; B, bisecting GlcNac β1–4 linked to β1–3 mannose; M, number of mannose residues; Gx, number of β1–4 linked galactoses; [3]G1, galactose on the antenna of the α1–3 linked mannose; [6]G1, galactose on the antenna of the α1–6 linked mannose; S, sialic acid linked to galactose. *Blue square*, GlcNac; *green circle*, mannose; *red triangle*, core fucose; *yellow circle*, galactose; *purple rhomb*, sialic acid.

Downloaded by Edith Cowan University from www.liebertpub.com at 10/23/18. For personal use only.

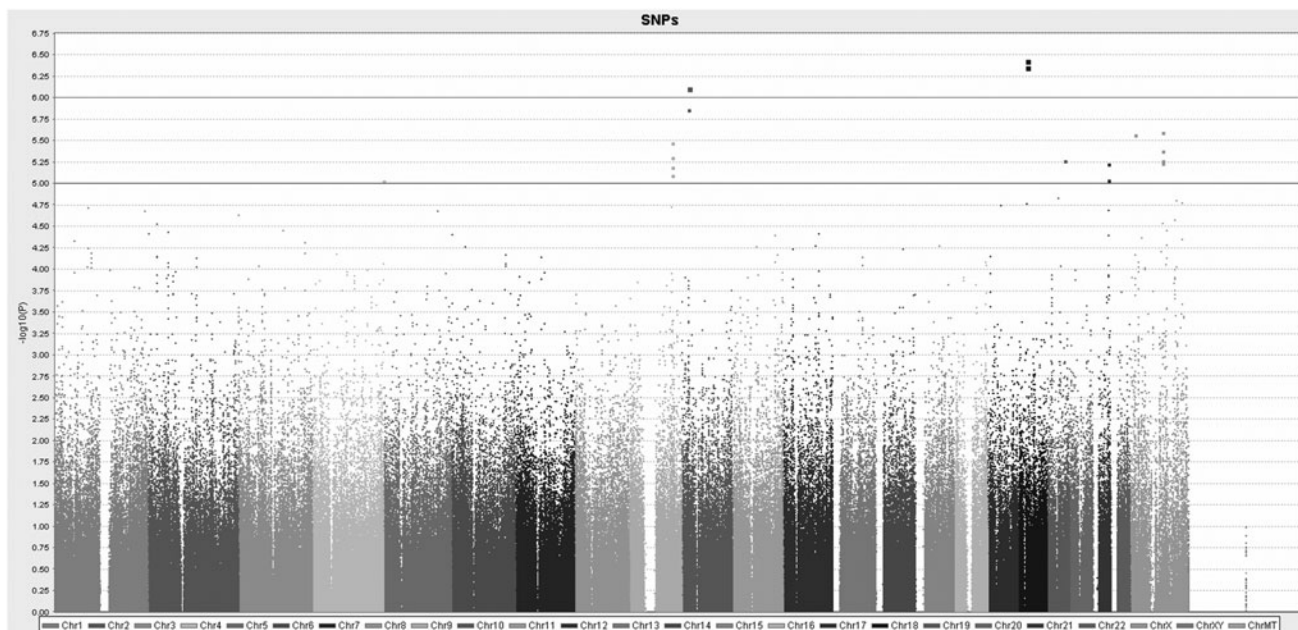

FIG. 1. Signal-intensity plots showing the association of SNPs with increased FPG levels. The –log *p*-values are for the association of each SNP with increase FPG levels. Only SNPs of sufficient quality are shown. FPG, fasting plasma glucose; SNPs, single nucleotide polymorphisms.

Downloaded by Edith Cowan University from www.liebertpub.com at 10/23/18. For personal use only.

the predictive ability of the 9 SNPs with or without 24 IgG GPs for increased FPG level.

Discussion

T2DM is a multifactorial disorder in which genetic, environmental, and metabolic influences interact at multiple levels. The increase of FPG level is characterized not only as one of the symptoms of T2DM but also as an independent risk factor for T2DM development (Mozaffary et al., 2016). We conducted GWAS and IgG glycosylation profiling on the 511 Chinese participants to investigate the association and predictive potential of candidate genes and IgG *N*-glycans for increased FPG level.

We identified nine SNPs in five genetic loci (*RPL7AP27*, *SNX30*, *SLC39A12*, *BACE2*, and *IGFL2*), which considerably affect the increase of FPG level. Of the studied variants, the SNP at 4q35.2 (rs7686384, with a *p*-value of 9.41×10^{-6} for the association with increased FPG level) is located in the gene of ribosomal protein L7a pseudogene 27 (*RPL7AP27*), which has been identified as one of the obesity-related traits (Comuzzie et al., 2012). The risk allele (A) for rs7686384 has a prevalence of 24.3% in participants with increased FPG level and an OR of 1.937, after adjusted for covariates (age, gender, BMI, SBP, DBP, HDL-C, LDL-C, and WHR).

Elevated FPG and obesity are both recognized as main features of metabolic syndrome, and the increase in the prevalence of T2DM is closely linked to the upsurge in obesity (Hossain et al., 2007). However, although T2DM and obesity are highly interrelated from both epidemiological and pathophysiological viewpoints, the shared genetic background by identifying common variants is limited (Grarup

Table 3. Genetic Biomarkers Associated with Increased Fasting Plasma Glucose Level

Chr.	*SNP rs#*	*Nonrisk/ risk allele*	*Frequency of risk allele*		*Adjusted OR for risk allele (95% CI)*[b]	p	*Gene*
			Increased FPG[a] (n = 76)	*No increased FPG*[a] (n = 435)			
4q35.2	rs7686384	A/G	0.2433	0.1424	1.937 (1.592–2.282)	9.41×10^{-06}	*RPL7AP27*
9q32	rs10739371	G/A	0.2367	0.1365	1.961 (1.54–2.382)	8.05×10^{-06}	*SNX30*
9q32	rs2153839	A/G	0.24	0.1383	1.968 (1.395–2.541)	6.48×10^{-06}	*SNX30*
9q32	rs10817396	G/A	0.2416	0.1367	2.011 (1.234–2.788)	3.35×10^{-06}	*SNX30*
10p12.33	rs2497753	A/G	0.1733	0.08382	2.292 (1.556–3.028)	1.39×10^{-06}	*SLC39A12*
10p12.33	rs9418383	A/C	0.1633	0.07544	2.393 (1.957–2.829)	7.71×10^{-07}	*SLC39A12*
19q13.32	rs4802289	G/A	0.1367	0.06243	2.377 (2.022–2.732)	5.46×10^{-06}	*CTCF*
21q22.3	rs2252576	A/G	0.1667	0.0846	2.164 (1.712–2.616)	9.23×10^{-06}	*BACE2*
21q22.3	rs2837981	G/A	0.1767	0.09043	2.158 (1.824–2.492)	5.96×10^{-06}	*BACE2*

[a]Increased FPG indicates participants with increased FPG level; no increased FPG indicates participants with decreased or fluctuant FPG level.
[b]Adjusted for age, gender, BMI, SBP, DBP, HDL-C, LDL-C, and WHR.
CI, confidence interval; OR, odds ratio; SNP, single nucleotide polymorphism.

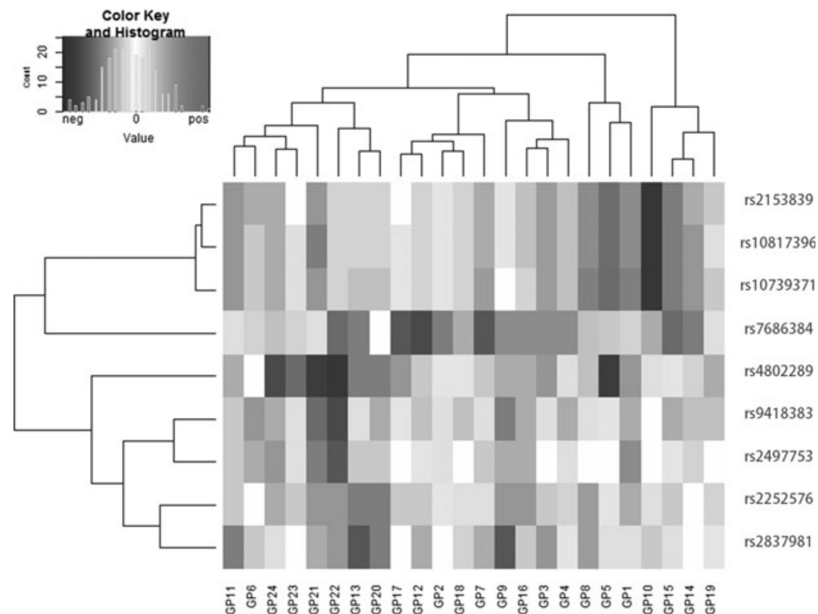

FIG. 2. Significant correlations between 24 IgG glycans and 9 SNPs. Heat map coding the standardized *r*-value in *dark*. Cells with nonsignificant correlations are displayed in *light*. In total, 9 SNPs were clustered into 5 groups (A, B, C, D, and E), while 24 GPs were grouped into 6 categories (a, b, c, d, e, and f). IgG, immunoglobulin G.

Downloaded by Edith Cowan University from www.liebertpub.com at 10/23/18. For personal use only.

et al., 2014). Here, we found that the risk allele ''A'' of rs7686384 is not only related with obesity but also positively associate with increased FPG level ($p < 0.001$), indicating that gene *RPL7AP27* (4q35.2) may have contributions in the pathogenesis of T2DM and obesity.

Moreover, rs10739371, rs2153839, and rs10817396 were located in the intron region of gene *SNX30* (sorting nexin family member 30) in 9q32. Gene *SNX30* encodes sorting nexin-30 protein, a member of sorting nexins, which are a large group of proteins that are localized in the cytoplasm and have the potential for membrane association through their phospholipid-binding domain. The ability of these proteins to bind specific phospholipids, as well as their propensity to form protein–protein complexes, points to a role of these proteins in membrane trafficking and protein sorting (Worby and Dixon, 2002). To date, no association has yet been found between *SNX30* and T2DM susceptibility or increased FPG level.

TABLE 4. COMPARISONS BETWEEN AREA UNDER THE CURVES FOR INCREASED FASTING PLASMA GLUCOSE LEVEL

Model	*AUC without 24 IgG glycans*	*AUC with 24 IgG glycans*	p
No adjustment			
Nine SNPs alone	0.740	0.805	0.0004*
Adjusted			
For age, gender	0.747	0.809	0.0005*
For age, gender, BMI	0.757	0.814	0.0006*
For age, gender, BMI, SBP, DBP, WHR, HDL-C, LDL-C	0.763	0.823	0.0004*

*A *p*-value <0.05 was considered significant.
AUC, area under the curve.

However, *SNX30* has been found associated with age-related hearing impairment and changes in SBP levels (Edwards et al., 2014; Fransen et al., 2015). Our finding demonstrated that the risk allele ''A'' of rs10739371, ''G'' of rs2153839, and ''A'' of rs10817396 have effect on the increased FPG level. In addition, the combined genetic-glycomic analyses found that these three SNPs presented same correlation patterns with decreased galactosylated glycan structures (GPs 11, 12, 14, and 15) and increased agalactosylated glycan structures (GPs 1 and 3–5) (Fig. 2).

Two SNPs (rs2497753 and rs9418383) in the intron region of gene *SLC39A12* at 10p12.33 (Solute Carrier Family 39 Member 12, also known as Zinc Transporter Member 12) showed significant effect on increased FPG level, where the risk allele ''C'' of rs9418383 had the highest OR of 2.393 among all SNPs (Table 3). *SLC39A12* belongs to a subfamily of proteins that show structural characteristics of zinc (Zn) transporters. Zn is an essential cofactor for hundreds of enzymes. It is involved in protein, nucleic acid, glycans, and lipid metabolism (Taylor and Nicholson, 2003). Zn is required for galactosyltransferase activity, but it has been demonstrated that excess Zn, on the contrary, decreases galactosyltransferase activity (Permyakov et al., 1993).

Galactosyltransferase is an enzyme that adds galactose to *N*-acetylglucosamine residues, and the posttranslational modifications of the β4-galactosyltransferase-1 may lead to the decrease of galactosylation, which is associated with a proinflammatory state of IgG (Dall'Olio et al., 2013). This overall proinflammatory state of IgG was found associated with T2DM and biological aging process (Gornik et al., 2012; Lu et al., 2011). Our results showed that polymorphisms in *SLC39A12* presented significant correlation with both increased FPG level and IgG glycosylation (Table 3, Fig. 2), indicating that polymorphisms in *SLC39A12* may have potential influences in mediating the activity of β4-galactosyltransferase-1 through the expression of Zn transporters, and affect the FPG level eventually.

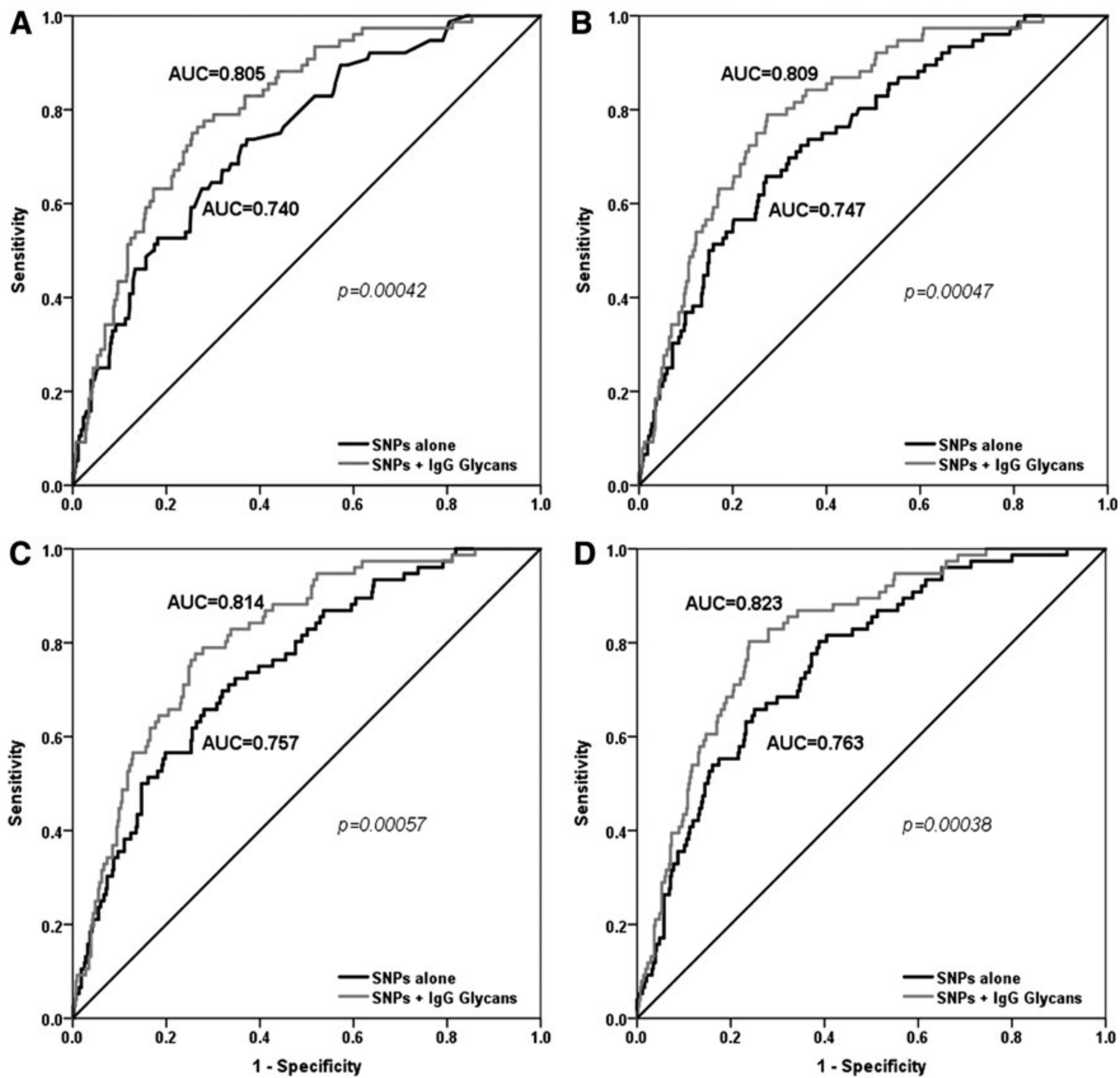

FIG. 3. ROC curves for increased FPG level. The AUCs are shown for **(A)** crude model (no adjustment for covariates); **(B)** adjusted for age and gender; **(C)** adjusted for age, gender, and BMI; and **(D)** adjusted for age, gender, BMI, SBP, DBP, HDL-C, LDL-C, and WHR. AUC, area under the curve; BMI, body mass index; DBP, diastolic blood pressure; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; ROC, receiver operator characteristic; SBP, systolic blood pressure; WHR, waist-to-hip ratio.

Downloaded by Edith Cowan University from www.liebertpub.com at 10/23/18. For personal use only.

BACE2 (Beta-Site APP-Cleaving Enzyme 2, 21q22.3) encodes an integral membrane glycoprotein that functions as an aspartic protease, and this encoded protein cleaves amyloid precursor protein into amyloid beta peptide, which is a critical step in the etiology of Alzheimer's disease and Down syndrome (Lange et al., 2015; Mok et al., 2014). It has been reported that protein Bace2 is a β cell-enriched protease, and the intake of Bace2 inhibitor displays augmented β cell mass and improved control of glucose homeostasis due to increased insulin levels in mice (Esterhazy et al., 2011).

Here, we found that two SNPs (rs2252576 and rs2837981) in the intron region of *BACE2* that had significant effect on increased FPG level were correlated with increase in neutral glycan structures (GPs 4 and 8) and decrease in sialylated glycan structures (GPs 16 and 20–22) (Fig. 2). Sialylation plays an important role in the inflammatory potential of IgG. Addition of sialic acid to IgG would decrease its binding to Fcγ receptors, and converts the function from pro- to anti-inflammatory (Kaneko et al., 2006). Based on the association with increased FPG level and decreased sialylation in IgG glycosylation, the polymorphisms in *BACE2* may participate in the inflammatory and immune pathway.

The risk allele ''A'' of rs4802289, located in the intron region of gene *IGFL2* (Insulin Growth Factor-Like Family Member 2, 19q13.32), was also significantly associated with increased FPG level. *IGFL2* belongs to the insulin-like growth factor (IGF) family of signaling molecules that play critical roles in cellular energy metabolism and in growth and development, especially prenatal growth (Emtage et al., 2006). We also found that rs4802289 was correlated with the decrease of sialylated glycan structures (GPs 21–24), which would increase IgG binding to Fcγ receptors and converts the function of IgG from pro- to anti-inflammatory (Kaneko et al., 2006).

Downloaded by Edith Cowan University from www.liebertpub.com at 10/23/18. For personal use only.

As expected, several SNPs identified from this study were located in the introns of the related genes, and mechanisms for these SNPs to regulate the expression of IgG *N*-glycosylation remain unclear. However, in contrast to genomics and proteomics, IgG *N*-glycosylation is a nontemplate-driven process. As we have discussed, all of the identified genes (*RPL7AP27*, *SNX30*, *SLC39A12*, *BACE2*, and *IGFL2*) are not directly regulating the IgG glycosylation, but somehow involved in the metabolic pathway of galactosylation, fucosylation, or sialylation of IgG. With these biological roles of glycosylation, these genes are shown to be strongly associated with the relative proportions of IgG *N*-glycans. Under these circumstances, further validation study on protein levels is needed to reveal the underlying molecular mechanisms.

T2DM is characterized by a proinflammatory state and elevated levels of inflammatory markers, such as C-reactive protein and interleukin-6, that have been associated with risk of developing T2DM (Wang et al., 2013). Evidence has shown that plasma *N*-glycan and IgG *N*-glycan were correlated with T2DM in previous studies (Keser et al., 2017; Lemmers et al., 2017). However, these researches were based on the prevalence of T2DM or the cross-sectional study on the level of FPG, which may limit the generalization and application of the glycomics markers to be predictive biomarkers for T2DM or pre-T2DM.

We investigated the differences of 24 IgG GP levels among the patients with increased FPG levels and patients with no increased FPG levels. Consequently, GPs 3, 8, and 11 presented positive trends with increased FPG levels, whereas GPs 4 and 14 showed negative trends of associations (Table 2). Through the clustering heat-map analyses, we found a clear interacting association between the 9 candidate SNPs and 24 IgG GPs, with obvious clustering patterns.

Twenty-four IgG GPs were clustered into six groups, and within each group same correlation patterns with genetic markers were shown (Fig. 2), indicating the involvement of genetic factor in the generation and function of IgG glycosylation. Specifically, agalactosylated glycan structures (GPs 1–5), which is associated with a proinflammatory state of IgG (Dall'Olio et al., 2013), were shown to have positive correlations with most of the SNPs. Among the 24 GPs, GPs 21–24 are the GPs with two sialic acids linked to galactose residues (Kaneko et al., 2006), which presented most negative associative patterns with the SNPs (rs4802289, rs2252576, and rs2837981).

In this study, we observed a significant improvement in predicting the increase of FPG level when 24 IgG glycans were added to the models with 9 SNPs in ROC analysis (Table 4). Furthermore, significant improvements remained to the same degree after adjusting for clinical covariates (age, gender, BMI, SBP, DBP, HDL-C, LDL-C, and WHR) in the ROC analysis (Fig. 3). The combination of 9 SNPs and 24 IgG GPs provided a remarkable AUC of 0.823 in predicting the increase of FPG level, showing the potential for the combination of genomics and glycomics as a predictive biomarker of the complex inflammatory pathophysiological changes in T2DM and increased FPG levels.

Genomics is the most matured techniques of the omics fields, where GWAS is a successful approach that has been used to identify thousands of genetic variants associated with complex diseases in multiple populations (GWAS catalog https://www.ebi.ac.uk/gwas/home). Glycomics reflects the genetic, environmental, and metabolic influences, and their interaction at multiple levels, which makes glycans attractive biomarkers for multifactorial diseases (Hasin et al., 2017). The strength of this study is the integration of genomics and glycomics to predict the increased FPG level, which is considered as an independent risk factor for T2DM.

By combining genomics with glycomics, we obtained a predictive biomarker for T2DM and FPG level. Moreover, the addition of glycomic profiling to genetic genotyping resulted in significant improvement of the predictive power for the changes in FPG level. The integration of genomics and glycomics could be recognized as a multidimensional interaction between internal and external risk factors, including genetic background, age, gender, and genetic-environmental interaction.

There are also limitations in this study. The validation is going on in another independent cohort to further confirm our predictive model constructed by genomics and glycomics biomarkers for generic application in general population.

To conclude, we found significant associations of nine genetic loci located in five genes (*RPL7AP27*, *SNX30*, SLC39A12, *BACE2*, and *IGFL2*) with increased FPG level. We also found that IgG GPs 3, 8, and 11 presented positive trends with increased FPG level, whereas GPs 4 and 14 showed negative trends with increased FPG levels. The predictive power of combined genetic and glycomic biomarkers is substantial with an AUC of 0.823 for increased FPG level. This study also implied the potential for the integration of genomics and glycomics as a biomarker of the complex inflammatory pathophysiological changes in T2DM development.

Acknowledgments

The authors thank all participants for their enrollment in the study. This work was supported by Australia–China International Collaborative Grant (NH&MRC-APP1112767-NSFC81561128020) and National Natural Science Foundation of China (Grant Nos. 81773527, 81273170, 81370083, and 81573215). Y.W. was supported by Beijing Nova Program (Grant No. Z141107001814058). S.G., X.Y., and H.W. were supported by China Scholarship Council (CSC-2015, CSC-2017).

Author Disclosure Statement

The authors declare that no conflicting financial interests exist.

References

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

Anthony RM, Nimmerjahn F, Ashline DJ, Reinhold VN, Paulson JC, and Ravetch JV. (2008). Recapitulation of IVIG anti-inflammatory activity with a recombinant IgG Fc. Science 320, 373–376.

Aulchenko YS, Ripke S, Isaacs A, and van Duijn CM. (2007). GenABEL: An R library for genome-wide association analysis. Bioinformatics 23, 1294–1296.

Calimlioglu B, Karagoz K, Sevimoglu T, Kilic E, Gov E, and Arga KY. (2015). Tissue-specific molecular biomarker signatures of

type 2 diabetes: An integrative analysis of transcriptomics and protein-protein interaction data. OMICS 19, 563–573.
Comuzzie AG, Cole SA, Laston SL, et al. (2012). Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. PLoS One 7, e51954.
Cummings RD, and Pierce JM. (2014). The challenge and promise of glycomics. Chem Biol 21, 1–15.
Dall'Olio F, Vanhooren V, Chen CC, Slagboom PE, Wuhrer M, and Franceschi C. (2013). *N*-glycomic biomarkers of biological aging and longevity: A link with inflammaging. Ageing Res Rev 12, 685–698.
DeFronzo RA, Ferrannini E, Groop L, et al. (2015). Type 2 diabetes mellitus. Nat Rev Dis Primers 1, 15019.
Ding D, Chong S, Jalaludin B, Comino E, and Bauman AE. (2015). Risk factors of incident type 2-diabetes mellitus over a 3-year follow-up: Results from a large Australian sample. Diabetes Res Clin Pract 108, 306–315.
Dupuis J, Langenberg C, Prokopenko I, et al. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 42, 105–116.
Edwards JS, Atlas SR, Wilson SM, Cooper CF, Luo L, and Stidley CA. (2014). Integrated statistical and pathway approach to next-generation sequencing analysis: A family-based study of hypertension. BMC Proc 8, S104.
Emtage P, Vatta P, Arterburn M, et al. (2006). IGFL: A secreted family with conserved cysteine residues and similarities to the IGF superfamily. Genomics 88, 513–520.
Esterhazy D, Stutzer I, Wang H, et al. (2011). Bace2 is a beta cell-enriched protease that regulates pancreatic beta cell function and mass. Cell Metab 14, 365–377.
Fransen E, Bonneux S, Corneveaux JJ, et al. (2015). Genome-wide association analysis demonstrates the highly polygenic character of age-related hearing impairment. Eur J Hum Genet 23, 110–115.
Ginter E, and Simko V. (2012). Type 2 diabetes mellitus, pandemic in 21st century. Adv Exp Med Biol 771, 42–50.
Gornik O, Pavic T, and Lauc G. (2012). Alternative glycosylation modulates function of IgG and other proteins—Implications on evolution and disease. Biochim Biophys Acta 1820, 1318–1326.
Grarup N, Sandholt CH, Hansen T, and Pedersen O. (2014). Genetic susceptibility to type 2 diabetes and obesity: From genome-wide association studies to rare variants and beyond. Diabetologia 57, 1528–1541.
Hasin Y, Seldin M, and Lusis A. (2017). Multi-omics approaches to disease. Genome Biol 18, 83.
Hindorff LA, Sethupathy P, Junkins HA, et al. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. PNAS 106, 9362–9367.
Hossain P, Kawar B, and El Nahas M. (2007). Obesity and diabetes in the developing world—A growing challenge. N Engl J Med 356, 213–215.
Junqueira LC, Carneiro J, and Kelley RO. (1998). *Basic Histology*, 9th ed. Stamford, NY: McGraw-Hill/Appleton & Lange.
Kaneko Y, Nimmerjahn F, and Ravetch JV. (2006). Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. Science 313, 670–673.
Keser T, Gornik I, Vuckovic F, et al. (2017). Increased plasma *N*-glycome complexity is associated with higher risk of type 2 diabetes. Diabetologia 60, 2352–2360.
Lange J, Lunde KA, Sletten C, et al. (2015). Association of a BACE1 gene polymorphism with Parkinson's disease in a Norwegian population. Parkinsons Dis 2015, 817–820.
Lemmers RFH, Vilaj M, Urda D, et al. (2017). IgG glycan patterns are associated with type 2 diabetes in independent European populations. Biochim Biophys Acta 1861, 2240–2249.
Liu R, Christoffel KK, Brickman WJ, et al. (2014). Do static and dynamic insulin resistance indices perform similarly in predicting pre-diabetes and type 2 diabetes? Diabetes Res Clin Pract 105, 245–250.
Lu JP, Knezevic A, Wang YX, et al. (2011). Screening novel biomarkers for metabolic syndrome by profiling human plasma *N*-glycans in Chinese Han and Croatian populations. J Proteome Res 10, 4959–4969.
Manning AK, Hivert MF, Scott RA, et al. (2012). A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet 44, 659–669.
Manolio TA. (2010). Genomewide association studies and assessment of the risk of disease. N Engl J Med 363, 166–176.
Mok KY, Jones EL, Hanney M, et al. (2014). Polymorphisms in BACE2 may affect the age of onset Alzheimer's dementia in Down syndrome. Neurobiol Aging 35, 1513.e1–1513.e5.
Morris AP, Voight BF, Teslovich TM, et al. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 44, 981–990.
Mozaffary A, Asgari S, Tohidi M, Kazempour-Ardebili S, Azizi F, and Hadaegh F. (2016). Change in fasting plasma glucose and incident type 2 diabetes mellitus: Results from a prospective cohort study. BMJ Open 6, e010889.
Parks BW, Nam E, Org E, et al. (2013). Genetic control of obesity and gut microbiota composition in response to high-fat, highsucrose diet in mice. Cell Metab 17, 141–152.
Permyakov EA, Reyzer IL, and Berliner LJ. (1993). Effects of Zn(II) on galactosyltransferase activity. J Protein Chem 12, 633–638.
Pucic M, Knezevic A, Vidic J, et al. (2011). High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations. Mol Cell Proteomics 10, M111010090.
Russell A, Adua E, Ugrina I, Laws S, and Wang W. (2018). Unravelling immunoglobulin G Fc *N*-glycosylation: A dynamic marker potentiating predictive, preventive and personalised medicine. Int J Mol Sci 19, E390.
Russell AC, Simurina M, Garcia MT, et al. (2017). The *N*-glycosylation of immunoglobulin G as a novel biomarker of Parkinson's disease. Glycobiology 27, 501–510.
Taylor KM, and Nicholson RI. (2003). The LZT proteins; the LIV-1 subfamily of zinc transporters. Biochim Biophys Acta 1611, 16–30.
Trbojevic Akmacic I, Ventham NT, Theodoratou E, et al. (2015). Inflammatory bowel disease associates with proinflammatory potential of the immunoglobulin G glycome. Inflamm Bowel Dis 21, 1237–1247.
Vuckovic F, Kristic J, Gudelj I, et al. (2015). Association of systemic lupus erythematosus with decreased immunosuppressive potential of the IgG glycome. Arthritis Rheumatol 67, 2978–2989.

Downloaded by Edith Cowan University from www.liebertpub.com at 10/23/18. For personal use only.

Wang X, Bao W, Liu J, et al. (2013). Inflammatory markers and risk of type 2 diabetes: A systematic review and meta-analysis. Diabetes Care 36, 166–175.

Wang Y, Klaric L, Yu X, et al. (2016a). The association between glycosylation of immunoglobulin G and hypertension: A multiple ethnic cross-sectional study. Medicine (Baltimore) 95, e3379.

Wang YX, Adua E, Russell A, et al. (2016b). Glycomics and its application potential in precision medicine, Science 354, 36–39.

Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, and D'Agostino RB Sr. (2007). Prediction of incident diabetes mellitus in middle-aged adults: The Framingham Offspring Study. Arch Intern Med 167, 1068–1074.

Worby CA, and Dixon JE. (2002). Sorting out the cellular functions of sorting nexins. Nat Rev Mol Cell Biol 3, 919–931.

Yu XW, Wang YX, Kristic J, et al. (2016). Profiling IgG *N*-glycans as potential biomarker of chronological and biological ages: A community-based study in a Han Chinese population. Medicine (Baltimore) 95, e4112.

Address correspondence to:
Wei Wang
Department of Public Health
School of Medical and Health Sciences
Edith Cowan University
Perth WA 6027
Australia

E-mail: wei.wang@ecu.edu.au

Qiang Zeng, MD, PhD
Department of International Inpatient
Institute of Health Management
Chinese PLA General Hospital
Beijing 100853
China

E-mail: zq301@126.com

Abbreviations Used

AUC = area under the curve
BMI = body mass index
CI = confidence interval
DBP = diastolic blood pressure
FPG = fasting plasma glucose
GP = glycan peak
GWAS = genome-wide association study
HDL-C = high-density lipoprotein cholesterol
IgG = immunoglobulin G
IGF = insulin-like growth factor
LDL-C = low-density lipoprotein cholesterol
OR = odds ratio
ROC = receiver operator characteristic
SBP = systolic blood pressure
SD = standard deviation
SNP = single nucleotide polymorphism
T2DM = type 2 diabetes mellitus
UPLC = ultra-performance liquid chromatography
WHR = waist-to-hip ratio

Downloaded by Edith Cowan University from www.liebertpub.com at 10/23/18. For personal use only.