4 Chapter 4: The heritability of fructosamine and its genetic relationship to HbA1c.

4.1 Introduction

Hyperglycaemia (high blood glucose) is the defining characteristic of diabetes. Normal fasting glucose (FG) levels in the blood range between 4.0 to 5.4 mmol/L (72-99 mg/dL), while post-prandial levels range up to 7.8 mmol/L (140 mg/dL) two hours after eating (2hr glucose) [329]. The most common tests to diagnose diabetes are fasting glucose (FG) and 2hr glucose after an oral glucose tolerance test (OGTT) (**Fig 4.1**). Both tests require a fasting period between 8 and 12 hours. OGTT involves taking a blood sample after the fasting period and then patients are given a very sweet drink containing 75g of glucose. Another blood sample is taken after two hours and this sample is the one used for diagnosis (2hr glucose post-OGTT). In some cases, blood samples are also taken at regular intervals between the intake of the sugar drink and the 2hr blood sample. Concordance between FG and 2hr glucose is not complete [330] and interestingly, individuals diagnosed using both criteria have higher cardiovascular disease risk than those only diagnosed using FG [331], and cardiovascular and all-cause mortality are increased in individuals diagnosed using 2hr glucose when compared to individuals diagnosed using FG [332].

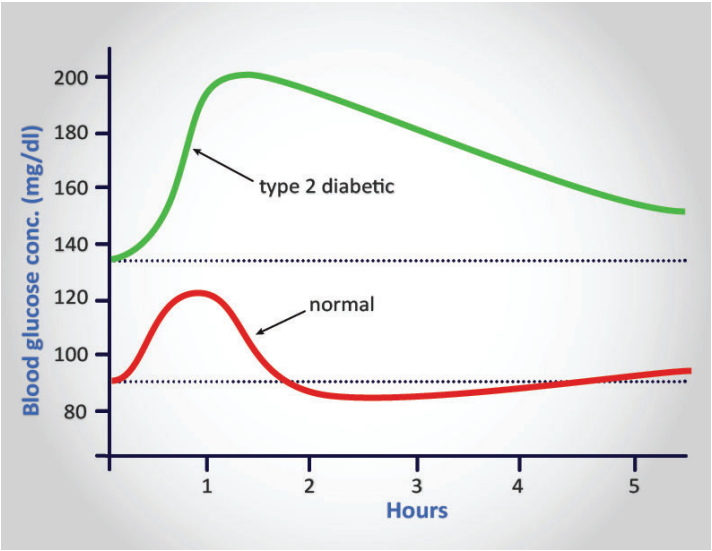

Figure 4.1: Diagnosis of type 2 diabetes. Blood glucose levels are shown for unaffected individuals (red line) and individuals with diabetes (green line) over a time course of 5 hours after glucose challenge. Fasting glucose (FG) test only measures glucose at the first time point. OGTT measures the first time point and the 2hr mark. Extracted from: https://themedicalbiochemistrypage.org/diabetes.php

Timely diagnosis of diabetes is important as uncontrolled high blood glucose levels can lead to clinical complications such as retinopathy, kidney failure or heart disease [25]. Undiagnosed diabetes can lead to damage to tissues due to hyperglycaemia which occurs over time without individuals displaying any marked symptoms. The degree of hyperglycaemia, reflecting the amount of damage to either insulin secretion or insulin response mechanisms (**Fig 4.2**), has an impact on the action course for treatment of the condition. Individuals with elevated glucose levels, but below the established threshold for diabetes diagnosis (FG=5.5 to 6.9 mmol/l (100 to 125 mg/dl), 2hr glucose=7.8 to 11.0 mmol/l (140 to 199 mg/dl)), are referred to as having prediabetes and can often manage their glucose levels by a combination of weight loss, physical activity and in rare cases oral glucose reducing medication [333]. When insulin secretion systems are severely damaged, individuals require insulin injections. This is the case for type 1 diabetes patients, who suffer from complete destruction of their beta-cells due to autoimmunity. For T2D cases there are

a host of oral treatments (e.g. metformin, sulphonylureas, thiazolidinediones (TZDs), dipeptidyl peptidase-4 (DPP-4) inhibitors and sodium glucose transporterase (SLGT-2) inhibitors) [334-336] but eventually many patients require insulin treatment.

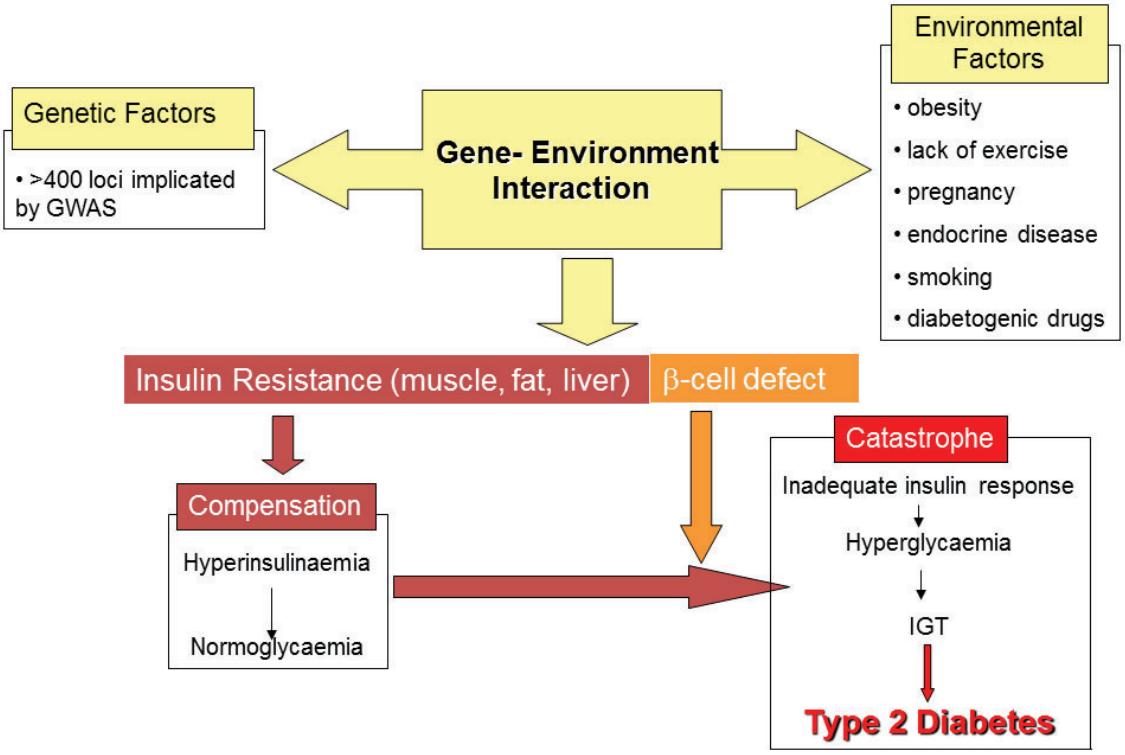

Figure 4.2: Aetiology of T2D. Interplay between genetic and environmental factors determines an individual's susceptibility to T2D. T2D arises from impaired beta-cell function and insulin resistance which occurs primarily in muscle, fat and liver tissues (i.e. insulin target organs). Under normal beta-cell function, in the setting of insulin resistance, insulin production is increased to increase the uptake of glucose by these tissues and glucose levels in the blood are kept within normal ranges. If this fails, glucose levels increase in the blood leading to impaired glucose tolerance (IGT). Individuals with untreated IGT have a high risk of developing T2D and cardiovascular disease. Image provided by Inês Barroso.

Another measure of glucose levels in the blood is glycated haemoglobin (HbA1c), which is the proportion of haemoglobin in the blood that has been glycated, and reflects average glucose levels over the life-span of an erythrocyte (~3 months). HbA1c is widely used to assess glycaemic control in patients with diabetes [337, 338]. As a diagnostic tool, HbA1c has a lower sensitivity than FG, but its negative predictive value is high, suggesting that low HbA1c levels provide strong evidence to discard a diabetes diagnosis [339]. However, HbA1c

levels as a diagnostic tool can be influenced by conditions that affect the lifespan of erythrocytes such as sickle cell trait and anaemia [340, 341], or by ethnic differences [342] (**Figure 4.3**). For example, ethnic minorities in the US such as Hispanics, Asians, American Indians and blacks have on average higher HbA1c levels than whites after adjusting for factors affecting glycaemia. This could affect the utility of HbA1c for T2D diagnosis, especially in populations of different ancestry [343]. Twin studies have estimated that heritability for HbA1c is high ranging from 62% to 75% [344, 345]. Multiple GWAS of HbA1c have looked into the genetic component of this trait, with a total of 60 loci identified to date [121, 346-350]. Lookups of association results of HbA1c-associated loci, with publicly available summary statistics for additional glycaemic (FG, 2h glucose and fasting insulin) and blood cell traits, in addition to conditional analyses adjusting for glycaemic traits (FG, 2hr glucose) or blood cell traits (haemoglobin levels, mean corpuscular volume, mean corpuscular haemoglobin) classified these loci as those mostly influencing HbA1c through glycaemic, erythrocytic, or unclassifiable pathways [121]. Understanding the pathway through which these variants affect HbA1c levels is important as it may influence their effect on T2D risk, diagnosis and treatment. For example, in a recent study from the MAGIC investigators [121] the authors described a missense variant in *G6PD* that lowers HbA1c levels through non-glycaemic pathways. This means that the lower HbA1c levels in *G6PD* variant carriers no longer reflect ambient glycaemia and therefore individuals with this variant could remain undiagnosed for T2D if this information were not taken into account and if they were screened by a single HbA1c measurement (see **Chapter 1, Section 1.2.2**). Understanding HbA1c genetics, which is studied in healthy non diabetic individuals to avoid confounding by diabetes and its treatment, could therefore help improve T2D diagnosis in populations of different ancestry.

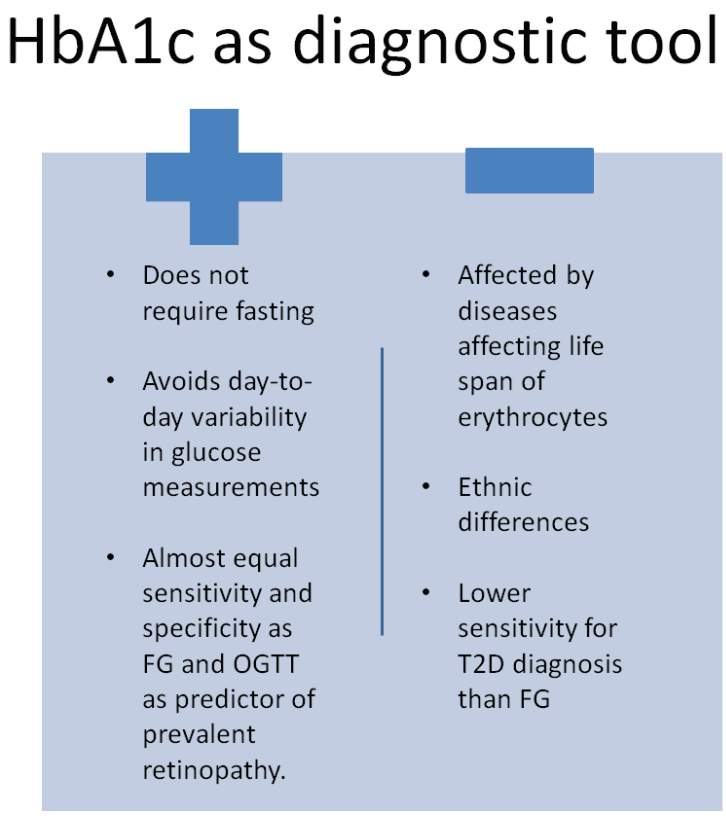

Figure 4.3: Advantages and disadvantages of HbA1c as a diagnostic tool.

Fructosamine is a measurement of glycation of total serum proteins. Since the most abundant serum protein is albumin, fructosamine normally reflects glycation of albumin [351]. In contrast to HbA1c, fructosamine measures short-term glycaemic control (from two to three weeks) and has been suggested as a useful marker for monitoring quick changes in glycaemic levels after treatment in individuals with diabetes [352]. As it is independent of haemoglobin, fructosamine levels are not affected by blood disorders and therefore is less likely to be influenced by erythrocytic traits, making it a viable alternative to HbA1c in the presence of anaemia or other blood disorders to monitor glycaemic levels [353]. Despite its potential advantages, it is not widely used as a measure of glucose control and has not been as standardised as HbA1c [354]. This lack of standardisation is problematic if fructosamine is

to be implemented as a diagnostic tool as accurate cutoffs need to be defined making sure variability within and between different labs is due to individual differences and are not assay or laboratory dependent. Nevertheless, studies have found association of fructosamine levels with diabetes incidence [355], retinopathy and chronic kidney disease [356], independently of baseline fasting glucose and HbA1c. Furthermore, there is a high correlation of fructosamine levels with HbA1c levels in patients with diabetes (r=0.7-0.8; [356-359]). The discordance in individuals between HbA1c levels and those levels predicted by its regression on fructosamine has been termed "glycation gap (GG)" [360]. Differences in FG and HbA1c as T2D diagnostic tools can be influenced by the GG as shown in a study where individuals were classified into three groups based on GG tertiles: low, medium and high glycators. Individuals diagnosed with diabetes by FG/2hr glucose and diagnosed as normoglycaemic by HbA1c had a significant depletion of individuals in the upper tertile of the GG suggesting individuals with low propensity for haemoglobin glycation are less likely to be diagnosed as diabetic by HbA1c criteria. In fact, the optimal HbA1c cutoff (i.e, the value misclassifying fewest patients) for low glycators was lower than that of high glycators (5.75% vs 6.25%) and had reduced sensitivity (54% vs 70%) [361]. To date, only one small study has looked at heritability of fructosamine in twins (40 monozygotic and 46 dizygotic) concluding it was not significantly inherited although a model including additive genetic effects and unique environmental influences could not be excluded [362]. A recent fructosamine GWAS performed on 8,951 mostly normoglycaemic white individuals ($N_{discovery}$=7,647, $N_{replication}$=1,304) found one single replicating association near *RCN3* (rs34459162, p=5.3x10^{-9}) after meta-analysis. This study did not examine the heritability of fructosamine and there was no significant evidence of genetic correlation with FG or HbA1c although there was some evidence of association for three established FG and/or HbA1c loci

(*TCF7L2*, *GCK* and *SLC2A2*) [363]. Elucidating non-glycaemic genetic influences of measurements used in T2D diagnosis, such as HbA1c and fructosamine, can help improve the diagnostic accuracy of these tests as well as provide insights into the different mechanisms by which these traits increase risk of diabetes comorbidities such as cardiovascular disease or chronic kidney disease independently of other glycaemic traits.

In this chapter, I performed a fructosamine GWAS to gain further insights into the genetic influences on fructosamine levels, and explore its genetic relationship with HbA1c and other glycaemic traits.

4.2 Chapter aims

The overall aim of this chapter is to explore the genetic basis of fructosamine. I use genome-wide genetic data available on up to ~19M SNPs on 24,586 individuals from the INTERVAL cohort to:

I. Assess the heritability of fructosamine.
II. Find novel loci associated with fructosamine.
III. Explore the genetic correlation of fructosamine with other glycaemic traits.
IV. Explore the effects of established glycaemic loci on fructosamine.

4.3 Methods

4.3.1 Participants

Work in this chapter was done using the INTERVAL cohort which consists of 47,394 predominantly healthy blood donors in the UK (more details in **Chapter 3 Methods 3.3.1**).

4.3.2 Genotyping, variant quality control and imputation

Genotyping, variant quality control and imputation on the INTERVAL cohort were performed by collaborators and full details can be found in Astle et al 2017 [293]. INTERVAL participants (N_{total}=48,813) were genotyped in ten batches on the UK Biobank Affymetrix Axiom Array. Standard QC procedures were implemented by Affymetrix during the genotype calling pipeline. Samples were excluded if signal intensity was poor (dish QC <0.82) or if call rate was low (<97%). Variants were excluded if a) call rate was low (<95%), b) there were more than three genotype clusters, c) cluster statistics (Fisher's linear discriminant, heterozygous cluster strength, homozygote ratio offset) indicated poor quality or d) they were complicated multi-allelic variants . Extra QC steps were performed by Tao Jiang and Heather Elding. Variants from a batch were failed if: a) fewer than ten minor allele homozygotes were called, b) the cluster plot contained at least one sample with an intensity at least twice as far from the origin as the next most extreme sample, c) the outlying sample(s) had an extreme polar angle (< 15° or > 75°) in the direction of the minor allele. Next, duplicate and non-European samples were excluded. Non-Europeans were defined based on PC1 and PC2 score thresholds defined after visual inspection of a PCA with 1000G major ancestry populations. Within batch variant QC was then performed discarding variants based on deviation from HWE ($p < 5x10^{-6}$) and a low call rate (<97%). If variants failed any of these last two filters or any of the Affymetrix filters in four out of ten batches, variants were discarded across all batches. After merging all batches, sample contamination was estimated using a contamination estimate based on allele frequency and probeset intensity [297]. Samples were excluded if this estimate was >10% or >3% if the sample also had more than ten first- or second-degree relatives (PI_HAT > 0.1875). Samples were then

excluded if they were heterozygosity outliers (>3 SD from mean), they had missing phenotypic sex or if supplied sex mismatched genetically inferred sex. Variants were removed if they had a MAF range >0.05 across all batches, if they were monomorphic in one or more batches and MAF > 0.01 in another batch, or if they had differing minor allele between batches (for variants with max MAF <0.475). An extra 69 across-batch duplicates were removed after merging batches. A global HWE filter ($p < 5x10^{-6}$) and a stringent call rate filter (>99% on non-failed batches and >75% globally) were applied to select variants used for imputation. Dataset was phased using SHAPEIT3 [364] and then imputed to a combined UK10K-1000G Phase III imputation reference panel using the Sanger Imputation Server [127].

4.3.3 Phenotyping

Phenotyping was performed by Star-SHL lab (http://www.star-shl.nl/). Fructosamine was measured using a colorimetric assay (Roche/Hitachi MODULAR P analyser system) and HbA1c was measured using a high performance liquid chromatography assay (Tosoh Automated Glycohemoglobin Analyser HLC-723G8 system) using serum collected on the 2 year follow-up visit. Fructosamine and HbA1c measurements as well as questionnaire data, technical variables and blood cell trait measurements were provided by the data administrator (University of Cambridge). I performed phenotype quality control in R to prepare the data for association analysis.

4.3.4 Association analysis, heritability and genetic correlation

Residuals obtained after phenotype quality control (**Results 4.4.1**) for fructosamine and HbA1c were used in this analysis. BOLT-LMM [227] was used to run genome-wide

association analysis on 19,100,024 variants with MAF > 0.1% and INFO score >0.4. LD score regression [42] was used to establish the heritability and genetic correlation of both traits. LD score regression was also used to compute genetic correlation of fructosamine and HbA1c with blood cell traits. A subset of 1,142,170 of the 1,217,312 HapMap3 SNPs with non-missing betas was kept in each dataset to perform these analyses. HbA1c summary statistics for European individuals from Wheeler et al 2017 [121] were obtained from the MAGIC consortium website (https://www.magicinvestigators.org/downloads/). Blood cell traits summary statistics were downloaded from http://www.bloodcellgenetics.org/. Genetic correlation analyses with glycaemic traits and albumin was performed using LD Hub [231].

4.3.5 Fructosamine discovery GWAS

LD score regression results showed no signs of inflation so no genomic correction was performed (LD intercept=1.01). I performed clumping as implemented in PLINK [223] to establish unique loci. Variants were clumped if they were 250kb away from the lead signal and if r^2>0.1. Conditional analysis was also used to identify distinct signals within a locus after clumping. To compare effect sizes of the lead variant near *RCN3* (rs34459162) found in a previous study [363], I reran normalisation on fructosamine matching the transformation done in that study and correcting for sex, donation centre, height, weight, processing date, number of donations and attendance date. SNPTEST v2.5.2 was used to rerun association analysis under an additive model and used for reciprocal conditional analyses. To estimate the significance of the difference in effect sizes I used a Z-test. SNPTEST v2.5.2 was also

used to test for association of the lead signal near *RCN3* in this study with serum albumin levels.

4.3.6 Lookup of established glycaemic loci

A list of established glycaemic loci was obtained from Eleanor Wheeler. This list was curated by Eleanor Wheeler and Gaëlle Marenne and last updated in March 2018. I first removed from the list chromosome X variants, variants monomorphic in the European population, and one variant not present in the INTERVAL data. To extract index variants per locus, I extracted LD information from European individuals in 1000G using LDlink (https://ldlink.nci.nih.gov/) [365]. For variants that were not biallelic in 1000G, I calculated LD in the INTERVAL dataset. For each pair of variants with $r^2 > 0.1$, I kept variants that had the lowest p-value in any of the association analysis with fasting glucose, 2 hr glucose, fasting insulin or HbA1c performed in European individuals in the latest trans-ethnic MAGIC unpublished analyses. Association data was provided by Ji Chen. The full list of index SNPs is presented in **Table 4.1**. Significance threshold for association was established by dividing 0.05 by the number of loci tested ($0.05/142 = 3.5 \times 10^{-4}$). To assess if there was an enrichment of directionally consistent and nominally significant signals in fructosamine that have been previously associated with glycaemic traits I performed a binomial test. Of the set of variants previously associated with glycaemic traits that were nominally significant in fructosamine, signals near *USP4* and *ANK1* were removed in this test since their association status with HbA1c is through non-glycaemic pathways and I chose to focus only on those influencing HbA1c through glycaemic pathways [121]. I also removed rs9727115 and rs150781447 as these were only associated with proinsulin levels adjusted for fasting glucose and late-phase proinsulin levels and the expected relationship between directions

of effect of a variant affecting proinsulin and fructosamine is not obvious. Direction of effect was determined using the HbA1c raising allele as reference (i.e we expect HbA1c raising variants to also raise fructosamine). If the variant was not associated with HbA1c, I used FG as reference and for rs9884482, which was not associated with HbA1c or FG, I used FI.

snp	chr	pos	gene	trait
rs340874	1	214159256	PROX1-AS1	FG [83], FG_adjBMI [366]
rs1886686	1	67390468	WDR78	FG_adjBMI *
rs2820436	1	219640680	LYPLAL1	FI [367], FI_adjBMI [367]
rs141203811	1	229772141	URB2	FI_adjBMI [368]
rs2375278	1	25529038	SYF2	HbA1c [121]
rs267738	1	150940625	CERS2	HbA1c [121], HbA1c_adjBMI *
rs6684514	1	156255456	TMEM79	HbA1c [346]
rs857725	1	158607935	SPTA1	HbA1c_adjBMI *
rs9727115	1	99177253	SNX7	Proinsulin_adjFG [369]
rs1260326	2	27730940	GCKR	2hG_adjBMI [370], FG_adjBMI [367], FI_adjBMI [367]
rs895636	2	45188353	SIX3	FG [371], FG_adjBMI [372]
rs1371614	2	27152874	DPYSL5	FG_adjBMI [366], FG_BMI30 [366]
rs3736594	2	27995781	MRPL33	FG_adjBMI [366]
rs35720761	2	43519977	THADA	FG_adjBMI *, HbA1c_adjBMI *
rs138726309	2	169763262	G6PC2	FG_adjBMI [368]
rs2232323	2	169764141	G6PC2	FG_adjBMI *, HbA1c_adjBMI *
rs146779637	2	169764368	G6PC2	FG_adjBMI *, HbA1c_adjBMI *
rs552976	2	169791438	ABCB11	FG_adjBMI [367], HbA1c [367]
rs733331	2	173546313	RAPGEF4-AS1	FG_adjBMI [372]
rs1530559	2	135755629	MAP3K19	FI [367], FI_adjBMI [367]
rs10195252	2	165513091	COBLL1	FI [367], FI_adjBMI [367]
rs1983210	2	220421417	OBSL1	FI_adjBMI *
rs2943645	2	227099180	LOC646736	FI_adjBMI [367]
rs17509001	2	24021231	ATAD2B	HbA1c [121]
rs12621844	2	48414735	FOXN2	HbA1c [121]
rs3755157	2	169792171	ABCB11	HbA1c [346]
rs17256082	2	175292364	SCRN3	HbA1c [121]
rs11708067	3	123065778	ADCY5	2hG_adjBMI [367], FG [83], FG_adjBMI [366], HbA1c [121]
rs7651090	3	185513392	IGF2BP2	2hG_adjBMI [367], FG [367], FG_adjBMI [367]
rs17036328	3	12390484	PPARG	FI_adjBMI [367]
rs7616006	3	12267648	SYN2	HbA1c [121]
rs9818758	3	49382925	USP4	HbA1c [121]
rs8192675	3	170724883	SLC2A2	HbA1c [121]

snp	chr	pos	gene	trait
rs4894799	3	171795540	FNDC3B	HbA1c [121]
rs35726701	3	49740895	RNF123	HbA1c_adjBMI *
rs9884482	4	106081636	TET2	FI [367], FI_adjBMI [367]
rs3822072	4	89741269	FAM13A	FI_adjBMI [367]
rs6822892	4	157734675	PDGFC	FI_adjBMI [367]
rs17046216	4	166255704	MSMO1	FI_adjBMI [378], HOMA-IR [378]
rs13134327	4	144659795	FREM3	HbA1c [121]
rs2237051	4	110901198	EGF	HbA1c_adjBMI *
rs1019503	5	96254817	ERAP2	2hG [367], 2hG_adjBMI [367]
rs146886108	5	14751305	ANKH	FG_adjBMI *
rs7708285	5	76425867	ZBED3-AS1	FG_adjBMI [367]
rs7713317	5	95716722	PCSK1	FG_adjBMI [367]
rs4865796	5	53272664	ARL15	FI [367], FI_adjBMI [367]
rs6450057	5	51647364	PELO	FI_adjBMI [373]
rs459193	5	55806751	LOC101928448	FI_adjBMI [367]
rs31244	5	75594743	SV2C	HbA1c_adjBMI *
rs35658696	5	102338811	PAM	Insulinogenic index [374]
rs10305492	6	39046794	GLP1R	FG [375], FG_adjBMI [368]
rs17762454	6	7213200	RREB1	FG_adjBMI [367]
rs35742417	6	7247344	RREB1	FG_adjBMI [368], HbA1c_adjBMI *
rs2745353	6	127452935	RSPO3	FI [367], FI_adjBMI [367]
rs6912327	6	34764922	UHRF1BP1	FI_adjBMI [367]
rs7756992	6	20679709	CDKAL1	HbA1c [121]
rs1800562	6	26093141	HFE	HbA1c [367], HbA1c_adjBMI *
rs11964178	6	109562035	LOC100996634	HbA1c [121]
rs9399137	6	135419018	HBS1L	HbA1c [346]
rs592423	6	139840693	LOC645434	HbA1c [121]
rs1799945	6	26091179	HFE	HbA1c_adjBMI *
rs1799884	7	44229068	GCK	1hG [376], 2hG [376], FG [376], HbA1c [350]
rs2191349	7	15064309	AGMO	FG [83], FG_adjBMI [366], HbA1c [121]
rs6943153	7	50791579	GRB10	FG [367], FG_adjBMI [367]
rs6947345	7	101071933	COL26A1	FG [377]
rs194524	7	89861832	STEAP2	FG_adjBMI *
rs1167800	7	75176196	HIP1	FI [367], FI_adjBMI [367]
rs35332062	7	73012042	MLXIPL	HbA1c_adjBMI *
rs11558471	8	118185733	SLC30A8	FG [367], FG_adjBMI [366], HbA1c [121], Proinsulin [369]
rs4841132	8	9183596	LOC157273	FG_adjBMI [366], FG_BMI30 [366], FI_adjBMI [366], FI_BMI30 [366]
rs4737009	8	41630405	ANK1	HbA1c [367]
rs6980507	8	42383084	SLC20A2	HbA1c [121]
rs34664882	8	41543675	ANK1	HbA1c_adjBMI *
rs7034200	9	4289050	GLIS3	FG [83], FG_adjBMI [366]
rs10811661	9	22134094	CDKN2B-AS1	FG [367], FG_adjBMI [366]

snp	chr	pos	gene	trait
rs16913693	9	111680359	IKBKAP	FG [367], FG_adjBMI [367]
rs651007	9	136153875	ABO	FG [375]
rs3829109	9	139256766	DNLZ	FG [367], FG_adjBMI [367]
rs7040409	9	91503236	C9orf47	HbA1c [121]
rs1467311	9	110536932	KLF4	HbA1c [121]
rs11557154	9	34107505	DCAF12	HbA1c_adjBMI *
rs3824420	9	712766	KANK1	Proinsulin AUC 0-30 [374]
rs7903146	10	114758349	TCF7L2	2hG [367], 2hG_adjBMI [367], FG [83], FG_adjBMI [366], FI [367], FI_adjBMI [366], HbA1c [367], Proinsulin [369]
rs701865	10	95381773	PDE6C	FG_adjBMI *
rs11195502	10	113039667	ADRA2A	FG_adjBMI [367]
rs7077836	10	132751498	MIR378C	FI_adjBMI [378], HOMA-IR [378]
rs16926246	10	71093392	HK1	HbA1c [350]
rs906220	10	71060610	HK1	HbA1c_adjBMI *
rs11605924	11	45873091	CRY2	FG [83], FG_adjBMI [366]
rs11603334	11	72432985	ARAP1	FG [367], FG_adjBMI [366], FG_BMI30 [366], HbA1c [121], Proinsulin [369]
rs1387153	11	92673828	MTNR1B	FG_adjBMI [367], HbA1c [367]
rs3782123	11	205198	BET1L	HbA1c [121]
rs2237896	11	2858440	KCNQ1	HbA1c [121]
rs174577	11	61604814	FADS2	HbA1c [121]
rs11224302	11	100456604	ARHGAP42	HbA1c [121]
rs415895	11	9769562	SWAP70	HbA1c_adjBMI *
rs117706710	11	10508903	AMPD3	HbA1c_adjBMI *
rs643788	11	118967758	DPAGT1	HbA1c_adjBMI *
rs10501320	11	47293799	MADD	Proinsulin [369]
rs10838687	11	47312892	MADD	Proinsulin [369]
rs17331697	12	97868906	RMST	FG [377]
rs10747083	12	133041618	FBRSL1	FG [367], FG_adjBMI [367]
rs2657879	12	56865338	GLS2	FG_adjBMI [367]
rs145878042	12	48143315	RAPGEF3	FI_adjBMI *
rs860598	12	102898446	IGF1	FI_adjBMI [367]
rs2110073	12	7075882	PHB2	HbA1c [121]
rs2408955	12	48499131	SENP1	HbA1c [121]
rs3184504	12	111884608	SH2B3	HbA1c_adjBMI *
rs2650000	12	121388962	HNF1A-AS1	Insulinogenic index [374]
rs150781447	12	65224220	TBC1D30	Proinsulin AUC 30-120 [374]
rs11619319	13	28487599	PDX1-AS1	FG [367], FG_adjBMI [367], HbA1c [121]
rs576674	13	33554302	KL	FG [367], FG_adjBMI [367], HbA1c [121]
rs282587	13	113351662	ATP11A	HbA1c [121]
rs9604573	13	114542858	GAS6, GAS6-AS1	HbA1c [121]
rs3783347	14	100839261	WARS	FG [367], FG_adjBMI [367]

snp	chr	pos	gene	trait
rs229587	14	65263300	SPTB	HbA1c_adjBMI *
rs4502156	15	62383155	C2CD4A	2hG_adjBMI [367], FG_adjBMI [367], Proinsulin [369]
rs2018860	15	99258710	IGF1R	FG_adjBMI [372]
rs1549318	15	71109147	LARP6	Proinsulin [369]
rs11248914	16	293562	ITFG3	HbA1c [121]
rs1558902	16	53803574	FTO	HbA1c [121]
rs4783565	16	68750190	CDH3	HbA1c [121]
rs837763	16	88853729	PIEZO1	HbA1c [121]
rs3747481	16	30666367	PRR14	HbA1c_adjBMI *
rs201226914	16	88798919	PIEZO1, LOC100289580	HbA1c_adjBMI *
rs72839768	17	7129898	DVL2	2hG_adjBMI *
rs61741902	17	2282779	SGSM2	Fasting proinsulin [374]
rs9914988	17	27183104	ERAL1	HbA1c [121]
rs12602486	17	42241929	C17orf53	HbA1c [346]
rs2073285	17	76117361	TMC6	HbA1c [121]
rs1046896	17	80685533	FN3KRP	HbA1c [367]
rs2748427	17	76121864	TMC6	HbA1c_adjBMI *
rs7225887	17	80904844	B3GNTL1	HbA1c_adjBMI *
rs4790333	17	2262703	SGSM2	Proinsulin [369]
rs1800437	19	46181392	GIPR	2hG_adjBMI *
rs731839	19	33899065	PEPD	FI [367], FI_adjBMI [367]
rs11667918	19	17232499	MYO9B	HbA1c [346]
rs17533903	19	17256523	MYO9B	HbA1c [121]
rs35413309	19	33167837	RGS9BP	HbA1c_adjBMI *
rs6113722	20	22557099	LINC00261	FG [367], FG_adjBMI [367]
rs17265513	20	39832628	ZHX3	FG_adjBMI [368]
rs855791	22	37462936	TMPRSS6	HbA1c [367], HbA1c_adjBMI *

Table 4.1: Index variants for established glycaemic loci per trait. 1hG= 1 hr Glucose. 2hG_adjBMI= 2 hr glucose adjusted for BMI. FG=Fasting glucose. FG_adjBMI=Fasting glucose adjusted for BMI. FI=Fasting insulin. FI_adjBMI=Fasting insulin adjusted for BMI. HbA1c=Glycated haemoglobin. HbA1c_adjBMI=Glycated haemoglobin adjusted for BMI. HOMA-IR= Insulin resistance homeostasis model assessment. Proinsulin AUC 0-30= Early phase proinsulin. Proinsulin AUC 30-120= Late phase proinsulin. Proinsulin_adjFG=Proinsulin adjusted for fasting glucose. Sources=Publication where the index SNP in the table was first associated with its respective trait. *MAGIC unpublished: based on European results from meta-analysis of data genotyped on the ExomeChip array.

4.4 Results

4.4.1 Phenotype quality control

Biological measurements can be susceptible to technical variation. To prepare the data for association analysis I performed quality control to assess if there were any effects of

technical variables on the fructosamine and HbA1c measurements. Fructosamine was measured on 28,310 individuals of the INTERVAL cohort and HbA1c was measured on 5,811 individuals out of which 5,420 had both measurements.

First, I inspected if measured values were concordant with what is expected based on available literature on the subject. Median fructosamine levels were high (294 µmol/L) compared to the normal expected range in healthy individuals (202-285 µmol/L) [379]. In contrast, HbA1c median levels were within range (median=35 mmol/mol, expected value=31-42 mmol/mol) [380]. Correlation between fructosamine and HbA1c was lower than expected (r=0.11 (this study) vs r=0.61[360], **Fig 4.4**).

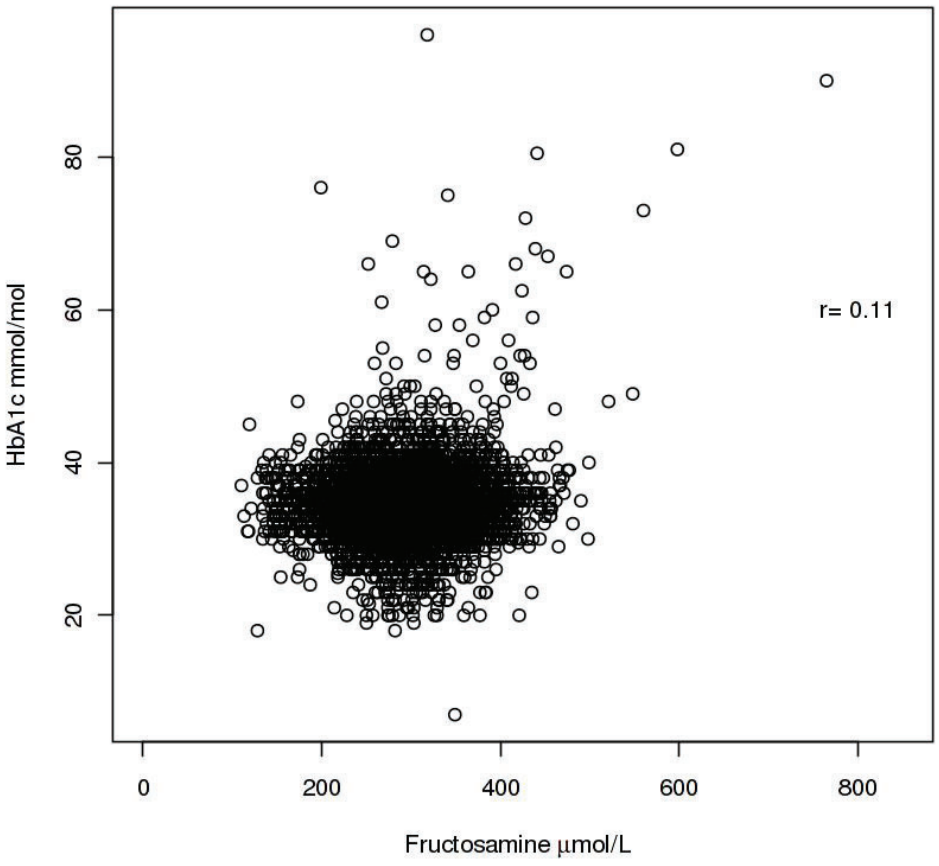

Figure 4.4: Correlation between fructosamine and HbA1c levels. r=correlation between fructosamine and HbA1c.

Next, I performed linear regression to determine which biometric and technical variables were significantly associated with fructosamine and HbA1c measurements. I determined sex, donation centre, use of glucose medication, height, weight, processing date and number of donations were significantly associated with both; while age, number of low haemoglobin deferrals, use of lipid lowering medication and use of blood pressure medication were associated with HbA1c exclusively and attendance date with fructosamine (**Table 4.2**). Individuals that reported use of glucose medication were subsequently removed.

Variable	Fructosamine	HbA1c
Age		X
Sex	X	X
Height	X	X
Weight	X	X
Attendance date	X	
Processing date	X	X
Donation centre	X	X
Number of donations	X	X
Number of low haemoglobin deferrals		X
Use of glucose medication	X	X
Use of lipid medication		X
Use of blood pressure medication		X

Table 4.2: Variables significantly associated with fructosamine and HbA1c. An X marks variables significantly associated (p<0.05 in linear regression).

After adjusting for relevant biometric and technical variables, residuals were extracted and inverse rank normalised. Correlation between fructosamine and HbA1c remained unchanged after adjusting for covariates (**Fig 4.5**).

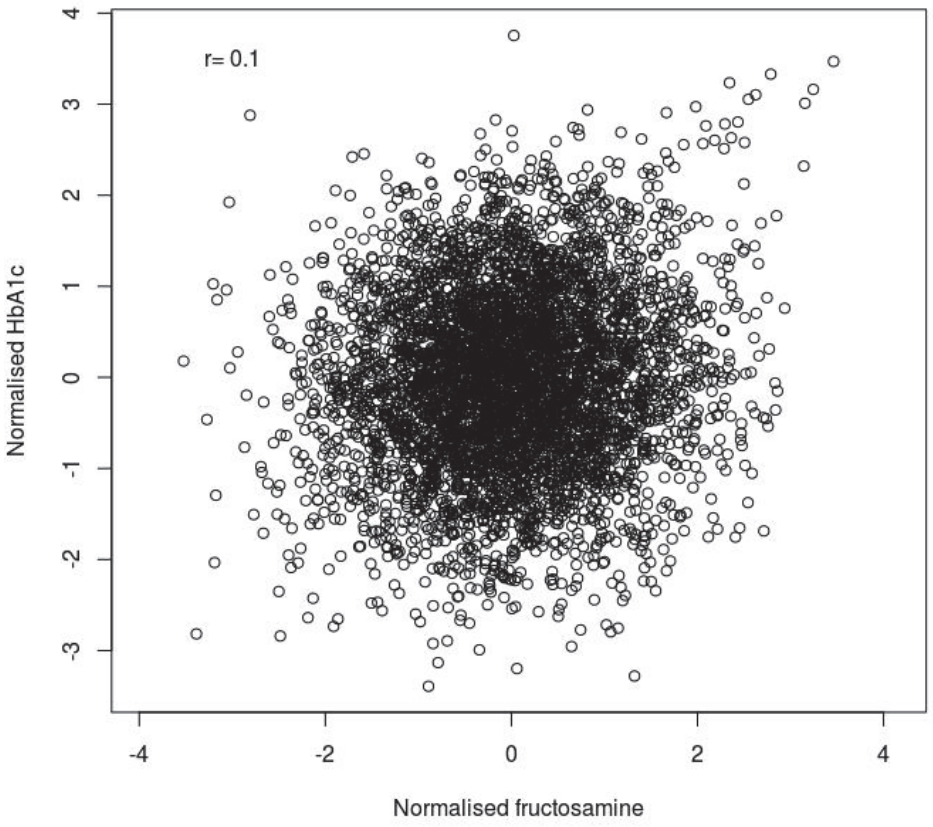

Figure 4.5: Correlation between normalised fructosamine and HbA1c levels after adjusting for biometric and technical variables. r=correlation between fructosamine and HbA1c.

4.4.2 Heritability of fructosamine and genetic correlation results

After inverse rank normalisation, 24,586 individuals with fructosamine measurements and 5,153 with HbA1c measurements were mapped to the genetic data. I used this data to run genetic association analyses with BOLT-LMM. LD score regression was used on association results to estimate common SNP (MAF>5%) heritability for both traits. The heritability estimate for fructosamine was very low (2% (95% CI -2%-5%)) which is in contrast to that for HbA1c (17% (95% CI 0%-35%)). No evidence of genetic correlation was observed between the two traits (Genetic correlation (RG)=0.40, SE=0.63, *p*=0.52). Using LD Hub, I estimated genetic correlation with other glycaemic traits. Fructosamine was not significantly

correlated with any other glycaemic trait (p>0.05, **Table 4.3**). HbA1c was significantly correlated with T2D (RG=0.47, p=0.01) and FG (RG=0.45, p=3.6x10^{-3}, **Table 4.3**). There was also significant and almost complete genetic correlation between HbA1c in this study and HbA1c results from previous MAGIC efforts [121] (RG=0.88, SE=0.20, p=1.34x10^{-5}) supporting the reliability of HbA1c measurements in this study. I also calculated genetic correlation between fructosamine and serum albumin and found no evidence of genetic correlation. Finally, given the established role of HbA1c-associated variants with some blood cell traits [121], I also calculated genetic correlation with twelve blood cell traits (**Table 4.3**). I found a positive genetic correlation of HbA1c with mean corpuscular haemoglobin concentration (MCHC, RG=0.37, SE=0.14, p=6.1x10^{-3}), mean corpuscular volume (MCV, RG=0.21, SE=0.11,p=0.04) and mean corpuscular haemoglobin (MCH,RG=0.28, SE=0.11, p=0.01) and a negative genetic correlation with red cell distribution width (RDW, RG=-0.24, SE=0.11, p=0.04).

Trait	Fructosamine			HbA1c		
	RG	SE	*P*	RG	SE	*p*
T2D	-0.04	0.19	0.88	0.47	0.18	0.01
FG_adjBMI	0.60	0.35	0.09	0.45	0.15	3.6x10^{-3}
FI_adjBMI	-0.43	0.32	0.17	-0.15	0.17	0.36
Fasting proinsulin	-0.10	0.37	0.78	-0.01	0.30	0.98
2hG_adjBMI	-0.76	0.49	0.12	0.05	0.28	0.87
Albumin	0.49	0.49	0.32	0.18	0.30	0.55
HCT	0.52	0.29	0.07	0.13	0.11	0.25
HGP	0.45	0.26	0.08	0.22	0.12	0.07
HLR	0.30	0.22	0.18	0.15	0.11	0.18
HLR%	0.23	0.20	0.24	0.17	0.11	0.14
IRF	0.30	0.24	0.21	0.01	0.12	0.90
MCHC	-0.18	0.21	0.41	0.37	0.14	6.1x10^{-3}
MCH	-0.03	0.14	0.83	0.28	0.11	0.01
MCV	0.04	0.14	0.75	0.21	0.11	0.04
RBC	0.35	0.21	0.09	-0.07	0.11	0.54
RDW	-0.18	0.16	0.27	-0.24	0.11	0.04
RET	0.25	0.19	0.20	0.19	0.12	0.10

Trait	Fructosamine			HbA1c		
	RG	SE	*P*	RG	SE	*p*
RET%	0.18	0.17	0.30	0.21	0.12	0.07

Table 4.3: Genetic correlation results for fructosamine and HbA1c. RG=genetic correlation. SE=standard error. P=p-value FG_adjBMI=Fasting glucose adjusted by BMI, FI_adjBMI=Fasting insulin adjusted by BMI, 2hG_adjBMI=2hr glucose adjusted by BMI. HCT=Haematocrit. HGB=Haemoglobin concentration. HLR=High light scatter reticulocyte count. HLR%=High light scatter percentage of red cells. IRF=Immature fraction of reticulocytes. MCHC=Mean corpuscular haemoglobin concentration. MCH=Mean corpuscular haemoglobin. MCV=Mean corpuscular volume. RBC=Red blood cell count. RDW=Red cell distribution width. RET=Reticulocyte count. RET%=Reticulocyte fraction of red cells. Highlighted in yellow, significant genetic correlation estimates. Genetic correlation analyses with glycaemic traits and albumin was performed using LD Hub [231]. Blood cell traits summary statistics for genetic correlation obtained from Astle et al 2017 [291].

4.4.3 Discovery of novel loci associated with fructosamine

Fructosamine association analysis yielded two associated loci at genome-wide significance ($p<5x10^{-8}$). The first association signal was rs853777 near *G6PC2* (beta=0.06, $p=1.7x10^{-10}$). This locus is also associated with HbA1c and FG [83]. The lead SNP in Dupuis et al 2010 [83] is rs560887 (r^2 with rs853777=0.63) with an effect size of 0.032 (%) for HbA1c and 0.075 (mmol/L) for FG. The effect size for rs853777 on the untransformed fructosamine values was 2.54 µmol/L. The second association signal was rs111476047 near *RCN3* (beta=0.09, $p=4.8x10^{-14}$). This locus was also previously associated with fructosamine in Loomis et al 2018 [363]. The lead signal in their study, rs34459162, is in moderate LD with my index variant (rs111476047 r^2= 0.28), and has the same direction of effect. To compare effect sizes, I repeated the association analysis for rs34459162 transforming the fructosamine measurements using natural log transformation as done in the previous study instead of inverse rank normalisation. The effect size was smaller in this study but the difference was not significant ($beta_{INTERVAL}$=0.015, $beta_{reported}$=0.02, p_{diff}=0.24). Reciprocal conditional analysis suggested that the lead signal found in this study was more tightly linked to the true causal variant than the previously reported signal (**Table 4.4**).

rsid	p-value	beta	conditioned p-value	conditioned beta	conditioned on
rs111476047*+	$2.19x10^{-11}$	0.014	$3.42x10^{-6}$	0.011	rs34459162**
rs34459162+**	$2.08x10^{-7}$	0.015	$5.35x10^{-2}$	0.007	rs111476047*
rs111476047*	$5.23x10^{-14}$	0.087	$8.90x10^{-9}$	0.074	rs739347***
rs739347***	$7.61x10^{-8}$	0.079	$2.09x10^{-2}$	0.038	rs111476047*
rs111476047*	$5.23x10^{-14}$	0.087	$1.51x10^{-12}$	0.083	rs34010237****
rs34010237****	$1.75x10^{-4}$	0.047	$6.24x10^{-3}$	0.034	rs111476047*

Table 4.4: Reciprocal conditional analysis of lead variant near *RCN3*. r^2=0.28.+Analysis was performed using log transformed fructosamine values. *Lead signal in this study. **Lead signal in Loomis et al 2018[363]. ***Lead signal for albumin GWAS in Franceschini et al (2012) [381]. ****Lead signal for albumin GWAS in Kanai et al 2018[382]. Numbers might differ slightly from main text due to difference in software use for association analysis (**Methods 4.3.5**).

Two additional SNPs mapping near the *RCN3* locus, rs739347 (r2=0.25 with rs111476047) and rs34010237 (r^2=0.02 with rs111476047), have been previously associated with serum albumin levels in European [381] and Japanese individuals [382], respectively. Both rs739347 and rs34010237 variants were significantly associated with fructosamine levels in this study (p=6.9x10^{-8} and p=2.2x10^{-4}, respectively) though reciprocal conditional analyses showed the rs739347 signal was heavily attenuated after conditioning on the lead signal in this study suggesting this association was mostly driven by the lead signal in this study (**Table 4.4**). Finally, to assess the effect of my index variant (rs111476047) on serum albumin levels, I used NMR measurements available from the first visit (**Chapter 3 Methods 3.3.1**), and found a significant and directionally consistent association with albumin levels in these data (beta=0.02, p=6.2x10^{-3}).

4.4.4 Evaluation of the effects of established glycaemic loci on fructosamine levels

Finally, I explored the influence of established glycaemic loci on fructosamine levels. For this analysis I used a list curated by Eleanor Wheeler and Gaëlle Marenne (**Table 4.1**). In total,

142 unique SNPs associated with at least one glycaemic trait were extracted. I found significant associations (Bonferroni-corrected $p < 3.5 \times 10^{-4}$, **Methods 4.3.6, Table 4.5**) in four loci: *ADCY5*, *GCK*, *G6PC2* and *MTNR1B*.

rsid	Gene	Chr	Pos	EA	NEA	EAF	B	SE	P
rs11708067	*ADCY5*	3	123065778	A	G	0.76	0.04	0.01	1.5×10^{-4}
rs730497	*GCK*	7	44223721	T	C	0.18	0.04	0.01	1.4×10^{-4}
rs1387153	*MTNR1B*	11	92673828	T	C	0.29	0.04	0.01	1.9×10^{-4}
rs552976	*G6PC2**	2	169791438	G	A	0.64	0.06	0.01	8.3×10^{-9}

Table 4.5: Associations of established glycaemic loci on fructosamine. Chr=chromosome. Pos = position in GRCH37. EA=Effect allele. NEA=Non-effect allele. EAF=Effect allele frequency. B=effect size SE=standard error of effect. P=p-value. *Nearest gene is *ABCB11*, but *G6PC2* is known to be the effector transcript at the locus [368].

I also found an enrichment of nominally significant and directionally consistent glycaemic signals in the fructosamine association results suggesting that these loci also have an effect on fructosamine (binomial $p=5.6 \times 10^{-3}$, **Table 4.6**).

rsid	Chr	Pos	EA-FR	EAF	B	SE	P	Gene	Trait	EA-T
rs10811661	9	22134094	T	0.83	0.04	0.01	1.9×10^{-3}	*CDKN2B-AS1*	HbA1c	T
rs11603334	11	72432985	G	0.85	0.03	0.01	7.9×10^{-3}	*ARAP1*	HbA1c	G
rs11708067	3	123065778	A	0.76	0.04	0.01	1.5×10^{-4}	*ADCY5*	HbA1c	A
rs1387153	11	92673828	T	0.29	0.04	0.01	1.4×10^{-4}	*MTNR1B*	HbA1c	T
rs17265513	20	39832628	C	0.19	0.03	0.01	1.0×10^{-2}	*ZHX3*	FG	C
rs1799884	7	44229068	T	0.18	0.04	0.01	1.4×10^{-5}	*GCK*	HbA1c	T
rs2232323	2	169764141	A	0.99	0.13	0.06	2.3×10^{-2}	*G6PC2*	HbA1c	A
rs3829109	9	139256766	G	0.73	0.02	0.01	3.3×10^{-2}	*DNLZ*	HbA1c	G
rs552976	2	169791438	G	0.64	0.06	0.01	8.3×10^{-9}	*ABCB11*	HbA1c	G
rs7651090	3	185513392	G	0.32	0.03	0.01	1.2×10^{-3}	*IGF2BP2*	HbA1c	G
rs7708285	5	76425867	G	0.31	0.02	0.01	2.2×10^{-2}	*ZBED3-AS1*	HbA1c	G
rs9884482	4	106081636	C	0.37	0.02	0.01	2.4×10^{-2}	*TET2*	FI	C

Table 4.6: Nominally significant and directionally consistent established glycaemic loci. Table legend: Chr=chromosome. Pos=position in GRCH37. EA-FR=effect allele in fructosamine. B=effect size in fructosamine. SE=standard error of effect size in fructosamine. P=p-value in fructosamine. Gene=nearest gene. Trait=Trait where the association of the SNP was previously reported. EA-T=Effect allele of the associated trait. Binomial $p=5.6 \times 10^{-3}$.

4.5 Discussion

In this chapter, as is standard for studies of glycaemic measures, I examined the genetic influences on fructosamine levels in a healthy population where one can explore these influences in a non-diabetic setting where measures are unaffected by disease and its treatment. Specifically, I sought to quantify the heritability of fructosamine, identify loci affecting the trait and explore its genetic relationship with other glycaemic traits. Overall, my results show that, in contrast with HbA1c, the heritability for fructosamine is low, despite some evidence for shared genetic aetiology. Results also highlight a variant potentially regulating fructosamine levels through pathways that also regulate circulating albumin.

Firstly I established that, in agreement with previous twin studies [362], fructosamine appears to be a lowly heritable trait (2% (95% CI -2%-5%)) suggesting most of the variation of the trait in this population is due to environmental factors which is not surprising given the fact that this is a trait normally used to measure short term changes in glycaemia after treatment . Fructosamine also does not show evidence of significant genetic correlation with other glycaemic traits including HbA1c (p>0.05) which is somewhat surprising given the fact that both traits normally have a high phenotypic correlation (~0.61[360]) and reflect similar biological processes, namely, the glycation of serum proteins. This lack of genetic correlation was also observed in Loomis et al 2018 [363]. The HbA1c heritability estimate in this study (17% (95% CI 0%-35%) was higher than those reported in LD Hub (7% (95% CI 4%-9%))[231] and the one obtained using summary statistics from the latest published MAGIC effort [121] (6% (95% CI 5%-8%)) but this difference was not statistically significant due to the wide confidence intervals in this study. It is likely though, that the estimate from this

study is inflated and the actual heritability estimate is closer to the one obtained from the MAGIC HbA1c data as this dataset has a much larger sample size (>20 times as large) and therefore is better powered to obtain a more accurate estimate. HbA1c was genetically correlated with both glycaemic traits (FG and T2D) and erythrocytic traits (MCHC, MCH, MCV and RDW), which is consistent with what is known about the biology of HbA1c [121].

Despite the low heritability, I was able to detect two loci associated with fructosamine levels at genome-wide significance ($p<5x10^{-8}$). The first locus was rs853777, near *G6PC2*. *G6PC2* codes for Glucose-6-Phosphatase Catalytic Subunit 2 and it is a well-established locus in HbA1c and fasting glucose metabolism [83]. This protein is produced specifically in islet beta cells and is involved in regulation of insulin secretion [383]. A mouse knockout of this gene exhibits mild metabolic phenotypes (reduction of blood glucose with no impact on cholesterol, glycerol, insulin and glucagon concentrations or body weight) and enhanced islet responsiveness to blood glucose levels[384, 385], making it a feasible therapeutic target given that no deleterious consequences were observed after the knockout. Interestingly, this locus was only nominally associated with fructosamine levels in Loomis et al 2018 [363] (rs1402837, $p=0.016$, r^2 with rs853777=0.17). This therefore represents a novel association of this locus with fructosamine levels.

The other associated SNP was rs111476047 located downstream of *RCN3*. *RCN3* codes for Reticulocalbin 3, which is an EF-hand calcium-binding protein of poorly understood function [386]. This locus was previously associated with fructosamine levels and conditional analysis shows that the locus found in this study is possible more tightly linked to the true causal variant (**Table 4.4**). There was no expression information for the lead signal in GTEx but rs113886122, the second strongest SNP in this locus ($p=7.3x10^{-14}$, r^2 with

rs111476047=0.62) is an eQTL for *FCGRT* in tibial nerve, subcutaneous adipose, transverse colon, skin and transformed fibroblasts tissues [387]. *FCGRT* codes for Fc Fragment of IgG Receptor and Transporter which plays a role in maintenance of albumin levels protecting it from degradation [388]. In addition to this, mouse studies have shown that hepatic levels of this protein regulate albumin homeostasis and susceptibility to liver injury [389]. These results combined with the suggestive evidence of association with albumin levels (p=6.2x10^{-3}) suggest that the locus found in this study could influence fructosamine levels through pathways that also regulate albumin.

Finally, lookups of previously established glycaemic loci suggest that factors affecting other glycaemic traits such as HbA1c, fasting glucose and fasting insulin also influence fructosamine levels reflecting a shared genetic aetiology for these traits. As sample sizes increase, it is likely some of these signals will reach genome-wide significance.

The results in this chapter need further exploration given a few limitations. Firstly, fructosamine levels were unusually high (median=294 µmol/L) compared to established reference ranges (202-285 µmol/L). Secondly, the correlation of fructosamine levels and HbA1c was unexpectedly low (r=0.1) and this discrepancy appears to not be driven by unreliability of HbA1c measurements as these fell within expected ranges and were supported by genetic correlation results. Phenotype quality control did not address either of these issues which suggests that there might be other factors influencing these observations such as machine calibration issues. Nevertheless, fructosamine measurements seem reliable enough to produce biologically plausible association results.

Future studies on the genetic architecture of fructosamine will shed more light into the different glycaemic and non-glycaemic mechanisms that can affect fructosamine levels, and

potentially identify mechanisms that affect risk of comorbidities independently of other glycaemic traits such as fasting glucose and HbA1c. Furthermore, increasing the number of individuals with both HbA1c and fructosamine measured could help identify variants associated with protein glycation by examining the genetic influences of the glycation gap. Screening for these variants could be potentially useful in the clinic when testing for T2D using HbA1c as a diagnostic tool.

4.6 Future directions

While working on this analysis, the first GWAS on fructosamine levels was published [363]. This presents an opportunity to use summary statistics from that study for meta-analysis with my data. To my knowledge, ARIC and CARDIA, the cohorts used in this previous study, are the only cohorts with available fructosamine and genetic data therefore the only other dataset that can be combined with mine.

To further explore the influence of established glycaemic loci on fructosamine levels, I will build a GRS score for T2D, FG, albumin and HbA1c and test them on the fructosamine dataset for association. In addition to this, I can also explore whether there is an enrichment of rare variant associations in known glycaemic loci using the WES and WGS data in the INTERVAL cohort ($N_{WES+WGS}$=5,874).

Another possible avenue to explore is to perform multi-trait analysis with fructosamine and HbA1c, or T2D, to identify pleiotropic effects or to boost power in identification of loci affecting fructosamine levels exclusively. This can be achieved using a method that uses summary statistics as input such as MTAG [390] so I can combine summary statistics from this study with data from the MAGIC and DIAGRAM consortia.