The Journal of Clinical Endocrinology & Metabolism, 2020, Vol. 105, No. 12, 1–11
doi:10.1210/clinem/dgaa658
Meta-Analysis

Meta-Analysis

Genome-Wide Association for HbA1c in Malay Identified Deletion on *SLC4A1* that Influences HbA1c Independent of Glycemia

Jin-Fang Chai,[1,*] Shih-Ling Kao,[2,*] Chaolong Wang,[3,4] Victor Jun-Yu Lim,[1] Ing Wei Khor,[2] Jinzhuang Dou,[3,4] Anna I. Podgornaia,[5] Sonia Chothani,[4] Ching-Yu Cheng,[6,7,8] Charumathi Sabanayagam,[6,7] Tien-Yin Wong,[6,7,8] Rob M. van Dam,[1,2,9] Jianjun Liu,[2,4] Dermot F. Reilly,[5,10] Andrew D. Paterson,[11,12] and Xueling Sim[1]

[1]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore 117549, Singapore; [2]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore 117597, Singapore; [3]Department of Epidemiology and Biostatistics, Key Laboratory for Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei 430030, China; [4]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore 138672; [5]Merck Research Laboratories, Kenilworth, New Jersey 07033; [6]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore 169856, Singapore; [7]Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore 169857, Singapore; [8]Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore 119228, Singapore; [9]Department of Nutrition, Harvard T.H Chan School of Public Health, Boston, Massachusetts 02115; [10]Janssen Pharmaceuticals Inc, Titusville, New Jersey 08560; [11]Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, M5G 1X8, Canada; and [12]Divisions of Epidemiology and Biostatistics, Dalla Lana School of Public Health, University of Toronto, M5T 3M7, Canada

ORCiD numbers: 0000-0003-3770-1137 (J.-F. Chai); 0000-0003-3945-1012 (C. Wang); 0000-0002-1233-7642 (X. Sim).

*These two authors contributed equally.

Abbreviations: AF, allele frequency; FG, fasting glucose; GWAS, genome-wide association study; HbA1c, glycated hemoglobin A1c; HWE, Hardy-Weinberg equilibrium; LD, linkage disequilibrium; MAC, minor allele count; MAF, minor allele frequency; MEC, Multi-Ethnic Cohort; QC, quality control; RG, random glucose; SAO, Southeast Asian ovalocytosis; SH2012, Singapore Health 2012; SiMES, Singapore Malay Eye Study; T2D, type 2 diabetes; WES, whole-exome sequence.

Received: 8 July 2020; Accepted: 15 September 2020; First Published Online: 16 September 2020; Corrected and Typeset: 19 October 2020.

Abstract

Context: Glycated hemoglobin A1c (HbA1c) level is used to screen and diagnose diabetes. Genetic determinants of HbA1c can vary across populations and many of the genetic variants influencing HbA1c level were specific to populations.

Objective: To discover genetic variants associated with HbA1c level in nondiabetic Malay individuals.

ISSN Print 0021-972X ISSN Online 1945-7197
Printed in USA

Design and Participants: We conducted a genome-wide association study (GWAS) analysis for HbA1c using 2 Malay studies, the Singapore Malay Eye Study (SiMES, N = 1721 on GWAS array) and the Living Biobank study (N = 983 on GWAS array and whole-exome sequenced). We built a Malay-specific reference panel to impute ethnic-specific variants and validate the associations with HbA1c at ethnic-specific variants.
Results: Meta-analysis of the 1000 Genomes imputed array data identified 4 loci at genome-wide significance ($P < 5 \times 10^{-8}$). Of the 4 loci, 3 (*ADAM15*, *LINC02226*, *JUP*) were novel for HbA1c associations. At the previously reported HbA1c locus *ATXN7L3-G6PC3,* association analysis using the exome data fine-mapped the HbA1c associations to a 27-bp deletion (rs769664228) at *SLC4A1* that reduced HbA1c by 0.38 ± 0.06% ($P = 3.5 \times 10^{-10}$). Further imputation of this variant in SiMES confirmed the association with HbA1c at *SLC4A1*. We also showed that these genetic variants influence HbA1c level independent of glucose level.
Conclusion: We identified a deletion at *SLC4A1* associated with HbA1c in Malay. The nonglycemic lowering of HbA1c at rs769664228 might cause individuals carrying this variant to be underdiagnosed for diabetes or prediabetes when HbA1c is used as the only diagnostic test for diabetes.

Key Words: HbA1c, type 2 diabetes, genome-wide association study, ethnic variations, Southeast Asian ovalocytosis

Glycated hemoglobin A1c (HbA1c), produced by nonenzymatic glycation of hemoglobin, is increasingly used as a standalone screening and diagnostic test for type 2 diabetes (T2D) (1, 2). As the average lifespan of the human erythrocyte is 2 to 3 months, HbA1c is also used as an indicator for glycemic control in patients with T2D.

HbA1c level is influenced by glycemic and nonglycemic pathways. Changes in HbA1c level can reflect blood glucose control responses to insulin resistance and reduced beta-cell function, both defining characteristics of T2D (3), or nonglycemic influences, such as erythrocyte function and lifespan, that may affect the clinical accuracy of the HbA1c level in the screening and monitoring of T2D. Cohen et al demonstrated that variability in erythrocyte lifespan could result in clinically significant variation in HbA1c levels in normal individuals (4). Erythrocyte lifespan is affected by blood cell conditions such as iron deficiency anemia and hemolytic anemia, resulting in nonglycemic increase and decrease in HbA1c levels, respectively (5).

Several genetic variants have been associated with nonglycemic influences on HbA1c levels in large-scale genome-wide association studies (GWAS), with different variants predominating in different populations (6-9). For example, the common missense variant *G6PD*-Asahi (rs1050828) has been associated with a 0.81% lowering of HbA1c levels in males via the erythrocyte pathway and is one of the major causes of variation in HbA1c levels in African American individuals (8). More recently, at least 3 other ancestry-specific, HbA1c-associated *G6PD* variants have been identified in Hispanic/Latino (10) and Asian individuals (11). These included *G6PD*-Canton, observed only in Chinese individuals, and *G6PD*-Viangchan, observed only in Malay individuals (11). By lowering HbA1c via the nonglycemic pathway, these variants may cause a delayed diagnosis or a missed diagnosis of T2D.

Malay is an ethnic group of Austronesian-speaking people comprising an estimated 23.5 million people worldwide, most of whom reside in the Southeast Asian countries of Indonesia, Malaysia, Thailand, and Singapore. In Singapore, prevalence of T2D is higher in the Malay and Indian populations compared with the Chinese population (12). While the Malay population shows a considerable degree of genetic similarity with the Chinese, imputation of Malay individuals to the cosmopolitan 1000 Genomes reference panel, which lacks Austronesian representation, may result in the loss of variants specific to the Malay population (13). For example, the rs12603404 genetic variant on chromosome 17 was found to be associated with HbA1c only in Malay individuals (N = 1735) and not in other populations, despite similar allele frequencies (7, 8). We hypothesized that we could more effectively discover genetic variants associated with HbA1c in nondiabetic Malay individuals by increasing the sample size and augmenting the genome-wide array data with whole-exome sequencing data. Here, we conducted a genome-wide association analysis for HbA1c using 2 Malay studies, the Living Biobank study (N = 983 on GWAS array and whole-exome sequenced) and the Singapore Malay Eye Study (SiMES, N = 1721 on GWAS array).

Materials and Methods

Study cohorts

We performed association analyses on the data from 2 studies in Singapore. The first study is the Living Biobank study that included Malay and Chinese individuals aged 18 to 75 years, sampled from 2 population-based cohorts in Singapore: the Multi-Ethnic Cohort (MEC) (14) and the Singapore Health 2012 (SH2012) (15). MEC was a population-based cohort initiated in 2007 to investigate the genetic and lifestyle factors that affect the risk of developing chronic diseases, such as diabetes and cardiovascular outcomes, in the 3 ethnic groups (Chinese, Malay, and Indians). SH2012 was a population-based, cross-sectional survey conducted between 2012 and 2013 in the 3 ethnic groups (Chinese, Malay, and Indian) with oversampling in the Malay and Indian populations. The second study is SiMES, a population-based, cross-sectional study involving 3280 Malay, aged 40 to 79 years and living in 15 southwestern residential districts in Singapore. The primary study aim was to determine the prevalence and risk factors of major eye diseases in the Singapore Malay population (16). We will henceforth refer to the 2 studies as Living Biobank and SiMES.

Glycemic measurements and exclusion criteria

HbA1c (%) was measured from whole blood using National Glycoprotein Standardization Program (NGSP) certified methods (17). Fasting glucose (FG) was measured using an enzymatic colorimetry assay, from the Living Biobank samples, where individuals were asked to fast overnight for 8 to 12 hours. Random glucose (RG) was measured from SiMES, as fasting was not required at the time of the study.

We excluded individuals with self-reported physician-diagnosed diabetes, use of diabetes medication, or undiagnosed diabetes (FG ≥7 mmol/l or RG ≥11.1 mmol/l or HbA1c ≥6.5% [48 mmol/mol], where available). Summary characteristics on the glycemic traits are in Supplementary Table 1. All supplementary tables and figures were deposited online (18).

Array genotyping, quality control, and imputation

Both Living Biobank and SiMES collected array genotype data. For Living Biobank, while the primary focus was on Malay, the array genotyping, quality control (QC) and imputation procedures were performed centrally for both Malay and Chinese. A total of 2500 samples (1208 Malay and 1292 Chinese) were genotyped on the Illumina HumanOmniExpress array. We first performed variant QC to exclude variants that were unmapped to hg19 human reference genome, or had overall call rates of <95%, or failed Hardy-Weinberg equilibrium (HWE, $P < 10^{-5}$). The pseudo-cleaned set of variants was used for subsequent sample QC. In sample QC, we excluded 13 samples with call rate <95% or in excess heterozygosity (heterozygosity > median + 5*interquartile range), and 8 samples with discordant clinical and genetic sex. To assess cryptic relationships among the samples, we used KING version 1.4 (19) to estimate pairwise kinship and exclude 5 samples that were detected as genetic duplicates. To ensure that the reported ethnicity matched the genetic ethnicity, we applied principal component analysis, using local reference samples from the Singapore Genome Variation Project (SGVP) (20), on our genotype data. We excluded 9 Malay with discordant ethnicity (2 clustered with SGVP Indians; 7 clustered with SGVP Chinese) (Supplementary Fig. 1A) (18). A total of 2452 samples (1189 Malay, 1263 Chinese) passed the genotype QC. For SiMES, the Malay samples were genotyped using an Illumina HumanHap 610Quad array. Similar genotype QC was performed (21), and a total of 2542 samples passed the genotype QC.

Before imputation, a second round of variant QC https://www.well.ox.ac.uk/~wrayner/tools/index.html#Checking was performed for each study to ensure our genotype data matched the 1000 Genome Phase 3 reference panel. We excluded variants with position mismatched, not on forward strand, palindromic with minor allele frequency (MAF) > 0.4, or discrepant allele frequencies (AF) from 1000 Genomes East Asians (AF difference >0.3 in Malay or >0.2 in Chinese). The final set of genotypes was then phased with SHAPEIT (22) and imputed to the 1000 Genomes Phase 3 combined panel (23) using the Michigan Imputation Server (24). A total of 47 095 002 variants were imputed into each study.

Living Biobank whole-exome sequencing and QC

In addition to the whole-genome array data in these 2 studies, the Living Biobank study also has whole-exome sequence (WES) data. WES data processing for Living Biobank can be generalized into 3 main steps: (i) sequences generation and reads mapping, (ii) genotype calling, and (iii) calls data QC.

Whole-exome sequences were generated from Nimblegen SeqCap EZ Human Exome Library v3 (Roche cat no: 06465692001) and sequenced on Illumina HiSeq2000 instruments (125 bp paired reads). BWA-MEM v0.7.12 (25) was used to map the raw FASTQ reads to human reference genome (GRCh37). We removed polymerase chain reaction duplicates, and poorly mapped reads (read mapping quality <20 or base quality <20). Base quality was then recalibrated using GATK v3.6 (26). We

used verifyBamID (27) to check sample contaminations in the resultant BAM files. Sample contaminations arise when there are sample swaps during sample preparation and sequencing steps. We removed samples with contamination rate >0.08, or mean sequencing depth <10× across target regions. After these pre-processing steps, a total of 2527 samples with mean depth of ~28× in target regions and ~1.2× in off-target regions, were available for genotype calling.

We applied a WES calling pipeline (28) to improve genotype calling accuracy for exome sequences in both target and off-target regions. Briefly, the pipeline combined variants detected within the targeted regions with all the biallelic single nucleotide polymorphisms (MAF >0.01) in the 1000 Genomes Phase 3 panel (23) to form a union set of variants for joint calling. Joint calling was performed using the GotCloud pipeline (29), adjusted for contamination rates (30).

We further filtered variants and samples in the called data. For variants QC, we first applied a support vector machine (SVM) classifier https://github.com/statgen/topmed_freeze3_calling to exclude low-quality variants (SVM score < -0.5), which corresponds to 99.1% sensitivity and 99.5% specificity. For each variant, we performed linkage disequilibrium (LD)-based genotype calling using BEAGLE v4.1 (31) to refine genotype likelihoods, with and without the 1000 Genomes Phase 3 reference panel. If the variant was present in the 1000 Genomes reference panel, genotypes from the refinement set with 1000 Genomes as reference were used; otherwise, genotypes from the other refinement set were used. After genotype refinement, we further excluded variants with minor allele count (MAC) < 1, dosage r^2 < 0.5, HWE $P < 10^{-5}$, mean depth >300 in targeted regions or >50 in off-targeted regions, within 5 bp of indels reported by 1000 Genomes Phase 3, or >5% of samples having maximum genotype probability <0.9. For sample QC, we excluded 9 samples with cryptic relationship among the 2527 samples from kinship analysis, and 11 samples with discordant ethnicity from principal component analysis (3 Malay and Chinese clustered with SGVP Indians; 8 Malay clustered with SGVP Chinese) (Supplementary Fig. 1B) (18). The cleaned exome sequence data consisted of 2507 samples (1216 Malay, 1291 Chinese) and 7 664 993 autosomal variants (1 162 550 in target regions, 6 502 443 in off-target regions). Concordance rate with array genotype data at heterozygotes is 99.93% in the target regions and 98.32% in the off-target regions.

For Living Biobank, subsequent association analyses were performed on a common set of 983 nondiabetic Malay that passed both array and exome QC (Supplementary Table 2) (18). A summary of the study samples and genetic data is presented in Supplementary Fig. 2 (18).

Association tests and meta-analysis

For association analyses in both the imputed and WES data, we performed single-variant genome-wide association test for HbA1c in Living Biobank Malay and SiMES, separately. We regressed HbA1c on age, age^2, and sex to obtain residuals. Inverse-normalized residuals were then regressed against the variants using an additive linear mixed model implemented in Efficient and Parallelizable Association Container Toolbox (EPACTS, http://genome.sph.umich.edu/wiki/EPACTS). We applied the EMMAX (32) kinship to account for population structure and sample relatedness. To obtain effect size change, we repeated the association test using untransformed trait values. We excluded variants with imputation quality Rsq <0.3, MAC <5, or standard error >10 for the imputed data, and variants with call rates <95%, HWE $P < 1 \times 10^{-6}$, or standard error >10 for the whole-exome data. To combine association test statistics across the 2 studies, we performed fixed-effect sample-size–weighted meta-analysis using METAL (33). Double genomic control was applied to correct inflations in test statistics in each study and meta-analysis. Variants with HbA1c association in only one study were removed after meta-analysis.

Genome-wide significance was defined at $P < 5 \times 10^{-8}$. We assigned all genome-wide significant associations into independent loci by assigning variant with the smallest *P* value as lead variant, and variants within 500kb of it into same locus. To refine associations in the loci, we performed conditional analysis by including the lead variant as an additional covariate in the regression model. Regional association results were visualized using LocusZoom plots (34).

To determine whether the associations were independent of glycemia, we performed association tests with (i) FG/RG and (ii) HbA1c adjusted for FG/RG, where available, using both inverse-normalized and untransformed traits. In association analysis with FG, body mass index was included as additional covariate in the regression model.

Malay-specific reference panel

As the Malay population is not included in the 1000 Genomes Phase 3 reference panel, imputation performance of the 1000 Genomes panel on Malay genotype data might be suboptimal. Specifically, the 27-bp *SLC4A1* deletion on chromosome 17, which was identified from the Living Biobank exome data, was not observed in the 1000 Genomes reference panel. To verify this HbA1c association in SiMES, we built a Malay-specific reference panel to impute the deletion in the SiMES array data for association testing with HbA1c.

The Malay-specific reference panel consists of exome and array genotype from 1175 Living Biobank Malay. We first excluded variants with call rates <95% or HWE $P < 1 \times 10^{-6}$ from the exomes. For variants present in both exome and array, we computed their genotype concordance, prioritized array genotype calls and excluded variants with discordant genotypes (>5%). For variants present in 1000 Genomes reference panel, we compared the variants and excluded variants with alleles mismatches. After QC, the Malay-specific reference panel consisted of 6 649 101 variants (6 639 936 single nucleotide variations, 3395 insertions, and 5770 deletions). We used SHAPEIT (22) to phase variants in the panel.

SiMES array data consists of 549 947 variants and 2542 samples. Prior to imputation, we checked the SiMES array data against the Malay-specific reference panel and excluded variants with strand mismatches and variants present in SiMES but absent in the reference panel. We then phased and imputed SiMES to the Malay-specific reference panel using SHAPEIT (22) and Minimac3 (35), respectively. We repeated the association test for HbA1c using the SiMES array data that was imputed on the Malay-specific reference panel.

Results

Summary of HbA1c associations from 2 Malay studies

To discover HbA1c associations in Malay, we first performed a genome-wide meta-analysis combining ~47M variants imputed from 1000 Genomes Phase 3 panel in 2704 nondiabetic Malay (1721 from SiMES, 983 from Living Biobank). The genomic inflation factors were 1.006 for SiMES, 0.996 for Living Biobank Malay, and 1.012 for the meta-analysis. At genome-wide significance ($P < 5 \times 10^{-8}$), we identified 95 genetic variants that were associated with HbA1c (Supplementary Table 3) (18). By iteratively assigning the variants that were located within 500kb of the lead associated variant into the same locus, we clustered the associations into 4 main loci (*ADAM15* on chromosome 1, *LINC02226* on chromosome 5, as well as *JUP* and *SLC4A1* (previously known as *ATXN7L3-G6PC3* locus, and we further show below that *SLC4A1* is the likely causal gene) on chromosome 17 (Supplementary Table 4, Supplementary Figs 3, 4) (18). Conditional analyses at each of the 4 loci indicated the absence of secondary signals ($P > 1 \times 10^{-4}$). Compared with 58 variants previously associated with HbA1c in a trans-ethnic meta-analysis (8), this Malay association meta-analysis exhibited a highly consistent direction of effect between different populations: 83.02% concordance for variants reported in European, and 84.62% concordance for variants reported in East Asian (Supplementary Table 5) (18).

In the meta-analysis, we found that HbA1c was most strongly associated with the *SLC4A1* locus on chromosome 17 (Supplementary Fig. 4D) (18). The lead variant from this meta-analysis, rs2285951 (effect allele frequency [EAF] = 1.32%, $\beta = -0.67 \pm 0.05\%$ [-7.3 ± 0.5 mmol/mol], $P = 2.27 \times 10^{-35}$), is in high LD (EAS $r^2 = 0.31$) with the previously reported lead variant rs12603404 in Malay (7). Conditional analysis on rs2285951 abolished the association at rs12603404 with HbA1c ($P_{conditional} = 0.57$). The remaining 3 loci (*ADAM15*, *LINC02226*, *JUP*) were novel for HbA1c. These loci have been linked with magnesium level (36), gut microbiome measurement (37), and esophageal cancer (38), respectively (Supplementary Figs 4A–C) (18). Of the 3 lead variants, the 2 lead variants at *ADAM15* (rs569036944 MAF = 0.03) and *JUP* (rs11871109 MAF = 0.02) loci were low frequency, while the last lead variant (at the *LINC02226* locus) was common in Malay individuals. These variants are mostly rare in other major continental populations (39).

To facilitate the discovery of HbA1c-associated coding variants at newly associated or previously known loci, we further analyzed Living Biobank Malay whole-exome sequence data for HbA1c associations. Three additional variants, 2 synonymous and 1 in-frame deletion at the *SLC4A1* locus, achieved genome-wide significance (Table 1). As these variants were not present in the 1000 Genomes Phase 3 reference panel, they were not imputed. The strongest association with HbA1c was observed at a synonymous variant rs369762319 on *GPATCH8* (EAF = 1.42%, $\beta = -0.42 \pm 0.06\%$ [-4.6 ± 0.7 mmol/mol], $P = 3.1 \times 10^{-11}$). As the sample size for WES data was smaller, HbA1c associations at the other 3 loci identified from the 1000 Genomes imputed meta-analysis (*ADAM15*, *LINC02226*, *JUP*) did not attain genome-wide significance. The minimum *P* values for HbA1c associations in the exome data were 4.14×10^{-6} at the *ADAM15* locus, 1.79×10^{-4} at the *LINC02226* locus, and 1.46×10^{-3} at the *JUP* locus (Supplementary Figs 4E–G) (18).

Analysis of exome sequence data revealed a novel HbA1c-associated *SLC4A1* variant rs769664228 linked to Southeast Asian ovalocytosis (SAO)

We combined association results from both 1000G-imputed and exome association analyses in Living Biobank Malay individuals, prioritizing the observed exome data at overlapping variants (Supplementary Fig. 5) (18). The most significant variant in the combined data was the lead variant

Table 1. Summary of Genome-Wide Significant Associations with HbA1c in Living Biobank Malay Exome Sequence Data

Locus	Variant ID	Chr	Position	EA/NEA	Annotation	Gene	Unconditional					Conditioned on *SLC4A1* deletion (rs769664228)				
							N	Counts	EAF	Beta	SE	*P* value	N	EAF	Beta	SE	*P* value
SLC4A1	rs769664228	17	42335410	A/ACGGCAGCC AGGACCTGGGG GCTGAATG	In-frame deletion	*SLC4A1*	983	953/30/0	0.02	-0.38	0.059	3.50×10^{-10}	-	-	-	-	-
SLC4A1	rs375378814	17	42432346	G/C	Synonymous	*FAM171A2*	983	950/33/0	0.02	-0.34	0.056	3.00×10^{-9}	983	0.02	-0.150.105		0.1824
SLC4A1	rs369762319	17	42477360	C/T	Synonymous	*GPATCH8*	983	955/28/0	0.01	-0.42	0.061	3.10×10^{-11}	983	0.01	-0.280.117		0.0232

Physical positions are based on hg19. Annotation was obtained from Variant Effect Predictor (VEP) v82. Counts are in the form of 0 copies of EA, 1 copy of EA, 2 copies of EA. Beta estimates reflected per allele effects of variants on untransformed HbA1c values. *P* values were obtained from sample-size–weighted fixed-effect meta-analysis on inverse-transformed HbA1c values.

Abbreviations: Chr, chromosome; EA, effect allele; EAF, effect allele frequency; N, sample size; NEA, non-effect allele; SE, standard error.

from exome analysis rs369762319. When we conditioned the analysis on rs369762319 in the exome data, no secondary signal was observed at this locus (Supplementary Fig. 6A) (18).

Next, we conditioned on the 27-bp in-frame coding deletion in *SLC4A1*, rs769664228 (in LD with rs369762319 in Living Biobank Malay: r^2 = 0.74, *D'* = 0.89, Table 1), which was a causal variant for SAO [OMIM #166900] (40). No causal evidence for blood cell trait has been reported for the other 2 exome variants that reached genome-wide significance for HbA1c associations at the *SLC4A1* locus. The conditional analysis indicated that the deletion was the single variant responsible for the HbA1c associations of exome variants in the locus (Supplementary Fig. 6B, Table 1) (18). By conditioning the imputed association signals on rs769664228, we fine-mapped the associations with HbA1c in the locus to rs769664228 ($P_{conditional} > 1 \times 10^{-4}$; Supplementary Fig. 6C) (18). This variant accounts for 4.1% of phenotypic HbA1c variance in Malay. Among 983 nondiabetic Living Biobank Malay, we observed 30 heterozygotes (HWE *P* = 1.0, Fig. 1A) and no homozygotes for rs769664228. We did not observe this variant in the Living Biobank Chinese.

HbA1c association with rs769664228 was replicated in SiMES data imputed to the Malay-specific reference panel

To impute Malay-specific variants that might be missing in the 1000G reference panel, we assembled a Malay-specific reference panel using Living Biobank whole-exome and array data. We imputed SiMES array data using the Malay-specific reference panel. Specifically, we observed 76 heterozygotes of rs769664228 among 2542 SiMES individuals, with MAF of 1.44% and imputation quality score of 0.85. Of the 76 carriers in SiMES, 56 were nondiabetic and included in subsequent association tests for HbA1c (Fig. 1B).

We performed a meta-analysis to combine the HbA1c associations from Living Biobank Malay exomes (N = 983) and SiMES (N = 1721) imputed to the Malay-specific reference panel. A total of 78 variants achieved genome-wide significance (Supplementary Table 6) (18). The strongest association with HbA1c was observed at the *SLC4A1* rs769664228 (EAF = 1.55%, β = –0.55 ± 0.04% [–6 ± 0.4 mmol/mol], $P = 6.58 \times 10^{-38}$, Fig. 2 and Supplementary Fig. 7) (18). In addition, we also observed significant association with HbA1c at rs5036, a Band 3 Memphis missense variant in *SLC4A1* that has been linked to SAO (41) (EAF = 94.82%, β = 0.18 ± 0.02% [2.0 ± 0.2 mmol/mol], $P = 1.54 \times 10^{-15}$; LD with rs769664228 in Living Biobank Malay: r^2 = 0.27, *D'* = 1). Conditional analysis on rs769664228 indicated no secondary signal at the locus.

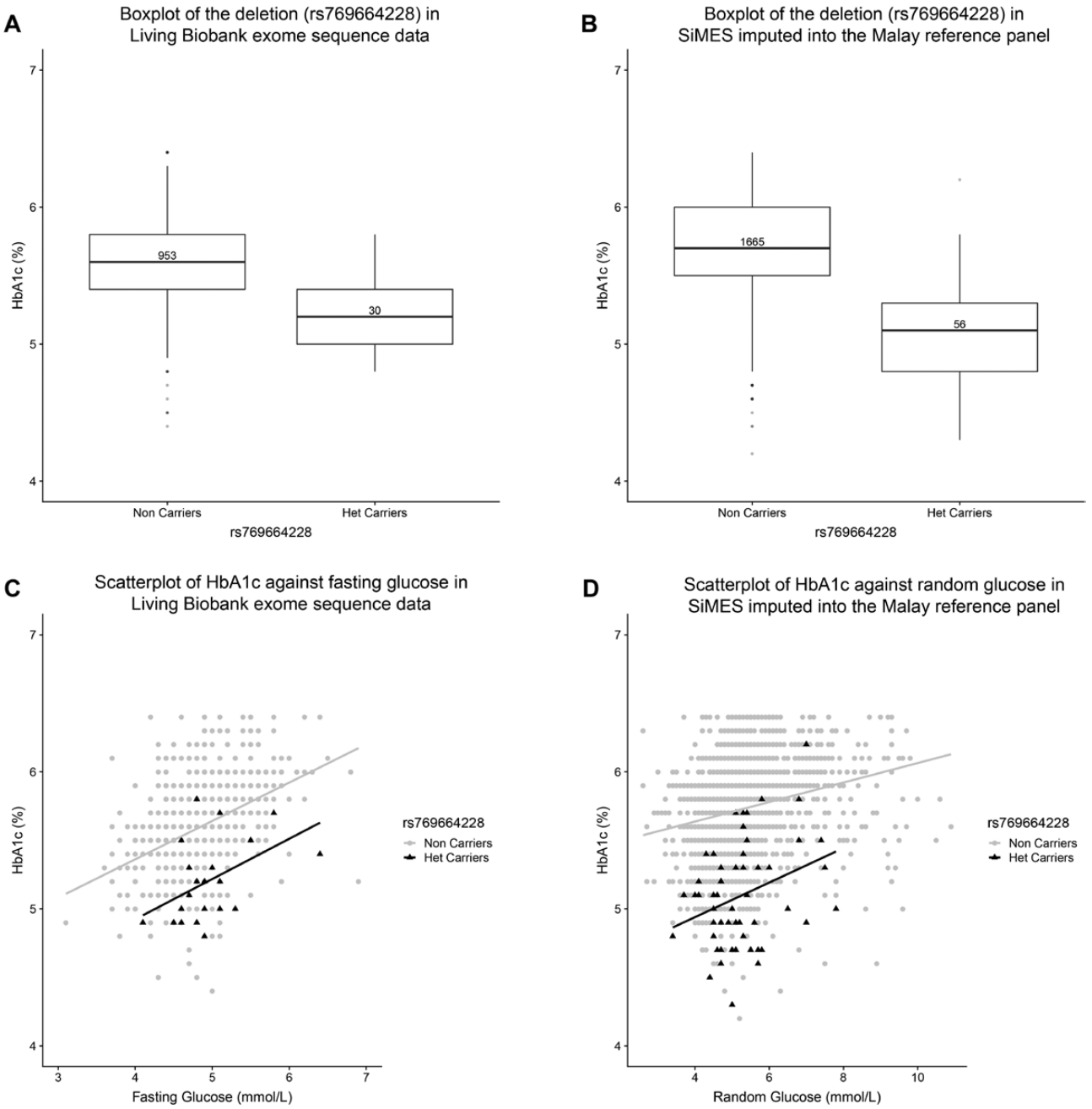

Figure 1. Distribution of HbA1c (%) by *SLC4A1* 27-bp deletion rs769664228 carrier status and glucose. Box-and-whisker plots of HbA1c by rs769664228 in (A) Living Biobank exome, and (B) SiMES imputed to the Malay-specific reference panel. Numbers in the boxplot above the median denote sample size. Scatterplot of HbA1c against (C) fasting glucose in Living Biobank exome, and (D) random glucose in SiMES imputed to the Malay-specific reference panel. In each scatterplot, regression lines of HbA1c on glucose were fitted for noncarriers and carriers of rs769664228.

No association with glucose at rs769664228, suggesting that HbA1c effect occurs through a nonglycemic mode of action

We tested association with other glycemic traits, including fasting glucose (FG) in Living Biobank Malay (N = 879) and random glucose (RG) in SiMES (N = 1721) (Supplementary Table 7) (18). No association with FG or RG was observed at any of our identified HbA1c loci (minimum P: 4.60×10^{-4} at *ADAM15*, 3.79×10^{-4} at *LINC02226*, 2.88×10^{-3} at *JUP*, 1.44×10^{-3} at *SLC4A1*). In addition, neither adjustment with FG or RG in the regression models showed any significant impact on the HbA1c effect estimates, suggesting that the identified genetic associations with HbA1c were driven by nonglycemic mechanisms (Supplementary Table 8) (18). We generated scatterplots of HbA1c on FG (for Living Biobank) and RG (for SiMES). HbA1c was

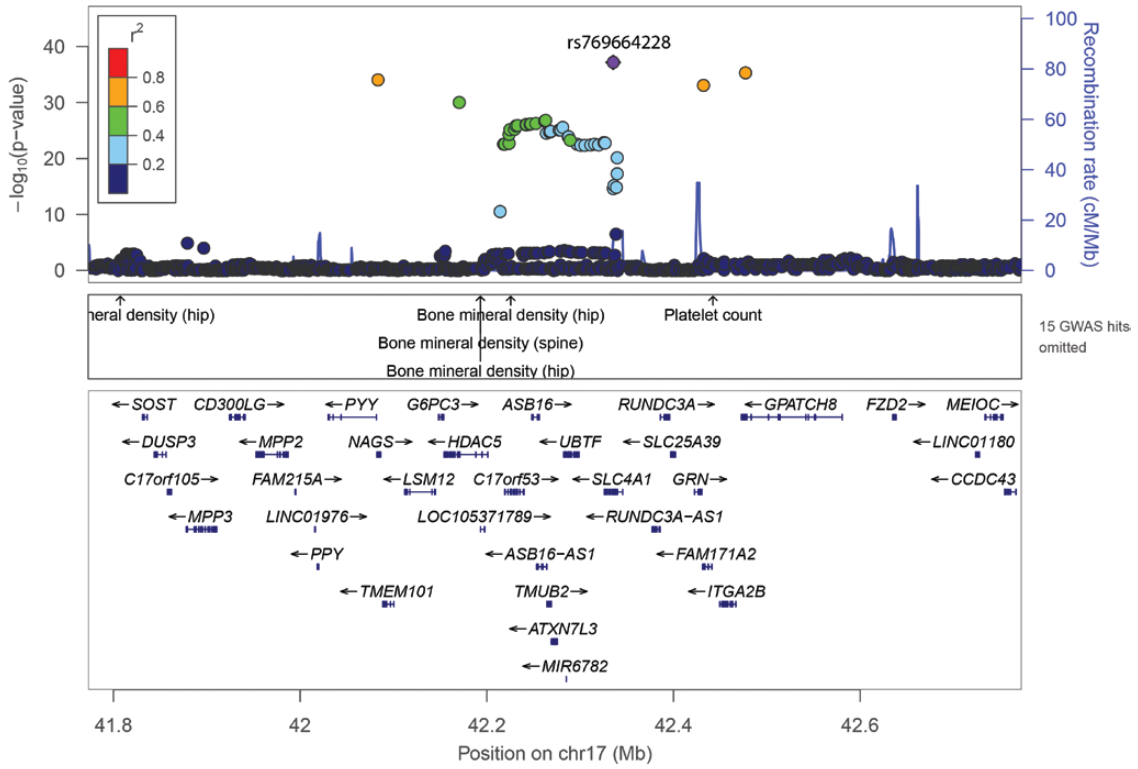

Figure 2. Regional association plot at *SLC4A1* locus on chromosome 17 showing HbA1c associations from meta-analysis of the Living Biobank Malay exome and SiMES imputed into the Malay-specific reference panel. Purple diamond indicates the lead variant in the locus. Variants were colored based on the Malay-specific reference panel LD with the lead variant.

positively correlated with FG (Pearson's $r = 0.16$) and RG (Pearson's $r = 0.05$), independent of the *SLC4A1* deletion (rs769664228) genotype (Fig. 1C and D).

The strongest association with FG was observed at missense variant rs2232326 at *G6PC2* (Living Biobank Malay exome: EAF = 11.9%, β = -0.20 ± 0.03 [-2.2 ± 0.3 mmol/mol], $P = 6.49 \times 10^{-11}$, Supplementary Fig. 8) (18). This FG association was previously reported as a secondary signal in East Asian Chinese individuals (42). In Living Biobank Chinese, the association was suggestive (exome: EAF = 5%, β = -0.16 ± 0.04 [-1.7 ± 0.4 mmol/mol], $P = 9.57 \times 10^{-7}$), and consistent direction of effect with the association in Malay individuals. No association was observed with HbA1c at this locus (minimum $P = 1.7 \times 10^{-3}$).

Discussion

In this study, we conducted a genome-wide association analysis to discover genetic variants associated with HbA1c in 2704 nondiabetic Malay individuals in Singapore. We identified 3 novel loci that have not been previously associated with HbA1c. At a previously reported HbA1c locus *SLC4A1* in Malay individuals, we replicated the association with HbA1c in an additional independent study of Malay individuals and used exome sequence data to fine map the association to a 27-bp deletion rs769664228 in the *SLC4A1* gene. As rs769664228 is missing in the 1000 Genomes Phase 3 reference panel, we demonstrated that the Malay-specific reference panel is critical for detecting and validating ethnic-specific variants. We showed that the genetic variants influence HbA1c levels independent of glucose levels.

SLC4A1 encodes the erythrocyte membrane protein 3, an erythrocyte anion-exchange protein. Rs769664228 at *SLC4A1* results in a 9–amino acid deletion in the protein localized in the boundary between the cytoplasmic and erythrocyte membrane domains. This deletion leads to a rigid erythrocyte membrane, changing the erythrocyte into an ovalocyte (40, 43). The deletion has also been linked with the Band 3 Memphis polymorphism known to affect erythrocyte membrane characteristics (41). Rs769664228 is a known causal variant for Southeast Asian ovalocytosis (SAO), a condition marked by hemolysis and hyperbilirubinemia that predominantly manifests during the neonatal period, but asymptomatic in adults (44). Based on the role of *SLC4A1* in erythrocyte biology and the fact that rs769664228 is a causal variant of SAO, rs769664228 likely lowers HbA1c level by reducing erythrocyte lifespan. However, we could not test this theory because of the lack of erythrocyte measures in our cohorts.

To our knowledge, this is the first study to link HbA1c to SAO. High prevalence of SAO was reported in the Malay (13.2%) and the Melanesians (22.4%) (45, 46). SAO exists predominantly in heterozygous state, as the homozygous state results in a severe disease that is usually embryonically fatal (47). SAO heterozygosity has been observed in Southeast Asia and the Pacific Rim (Indonesia, Papua New Guinea, Malaysia, Philippines, Brunei, Cambodia, and southern Thailand), the Torres Strait islands of Australia, and Cape Coloured in South Africa (48). In Living Biobank Malay, 30 of 983 (3.05%) normoglycemic individuals are heterozygotes for the SAO causal variant rs769664228.

HbA1c has been increasingly used as a screening and diagnostic test for T2D (1, 2). However, various factors influence HbA1c levels, some of which are ethnic-specific. To account for such ethnic differences, population-specific cutoffs for HbA1c have been proposed (2). For example, in Malaysia, a lower cutoff of HbA1c of ≥6.3% is used for the diagnosis of T2D (49). Furthermore, a study on multi-ethnic populations had proposed the use of HbA1c as first-step screening of T2D, but a combination of HbA1c and FG test to improve the classification of diabetes and prediabetes (50). We have now identified another factor, SAO carrier status, which is associated with nonglycemic lowering of HbA1c levels and may thus reduce the diagnostic sensitivity of HbA1c for diabetes or prediabetes. We recognize that it is not an essential practice to screen individuals for the SAO causal variant rs769664228. Faced with a patient of unknown SAO carrier status, clinicians should consider using a combination of HbA1c and FG tests to ensure a more accurate diagnosis of T2D, especially in populations in which rs769664228 is common.

We also demonstrated that a population-specific reference panel is more efficient for genotype imputation than a cosmopolitan panel such as the 1000 Genomes reference panel, when the population is absent or under-represented in the 1000 Genomes Project. The HbA1c-associated and SAO causal variant rs769664228 was observed in gnomAD v2.1.1, but very rare in major populations (MAF < 0.04%) and absent in the 1000 Genomes Phase 3 reference panel. We therefore theorize that rs769664228 is a Malay-specific variant. We generated a Malay-specific reference panel that includes rs769664228 and used it as a reference to impute rs769664228 in SiMES array data, obtaining a respectable imputation quality score of 0.85. Subsequent analysis using the imputed data showed stronger evidence of association with HbA1c at rs769664228 (EAF = 1.55%, $\beta = -0.55 \pm 0.04\%$ [-6 ± 0.4 mmol/mol], $P = 6.58 \times 10^{-38}$), supported the HbA1c finding at rs769664228 in the discovery analysis.

Finally, at the *G6PC2* locus, our result replicated a previous report of the East Asian-specific, FG-lowering missense variant rs2232326 (42) in Malay individuals. The variant is more common among the Malay population (MAF = 12%) than in other populations (East Asian Chinese: MAF = 4%, other populations: MAF < 0.4%) (51). This variant encodes a serine-to-proline substitution at position 324, which is predicted to be "damaging" by SIFT (52). In Ensembl GRCh37 release 97 (53), rs2232326 is only annotated as missense for a minor transcript (ENST00000375363.3), whereas it was annotated as a 3' untranslated region and a nonsense mediated decay transcript for the most abundant transcript ENST00000282075.4. These *G6PC2* transcripts were mainly expressed in the pancreas (Genotype-Tissue Expression [GTEx] Portal v8 on June 30, 2020), which is responsible for the regulation of blood glucose level.

In summary, we identified genetic variants in Malay individuals that influence HbA1c level via a nonglycemic pathway. The association with HbA1c at the SAO causal variant rs769664228 highlights the potentially significant impact that ethnic-specific influences on HbA1c might have on the accuracy of T2D diagnosis. In populations in which SAO is prevalent, a combination of both HbA1c and FG tests should be considered when establishing a T2D diagnosis or monitoring the diabetes treatment response.

Acknowledgments

The authors thank all investigators, staff members, and study participants for their contributions in this study.

Weblinks

KING, http://people.virginia.edu/~wc9c/KING/
Pre-imputation variant QC, http://www.well.ox.ac.uk/~wrayner/tools/
Michigan Imputation Server, https://imputationserver.sph.umich.edu/index.html#!pages/home
BWA-MEM, http://bio-bwa.sourceforge.net/
GATK, https://gatk.broadinstitute.org/hc/en-us
VerifyBamID, https://genome.sph.umich.edu/wiki/VerifyBamID
GotCloud variant calling pipeline, https://genome.sph.umich.edu/wiki/GotCloud:_Variant_Calling_Pipeline
SVM classifier, https://github.com/statgen/topmed_freeze3_calling
BEAGLE, https://faculty.washington.edu/browning/beagle/beagle_4.1_21Jan17.pdf
EPACTS, https://genome.sph.umich.edu/wiki/EPACTS
METAL, https://genome.sph.umich.edu/wiki/METAL
GWAS Catalog (accessed on 17/04/2020), https://www.ebi.ac.uk/gwas/home

Grants and Fellowships

Financial Support: The MEC study was funded by the Biomedical Research Council, National Medical Research Council, National Research Foundation, Singapore Ministry of Health, Na-

tional University of Singapore and National University Health System. The SH2012 study was funded by the Singapore Ministry of Health, National University of Singapore and National University Health System. Genome Institute of Singapore provided services for genotyping. Genotyping and whole-exome sequencing for Living Biobank was jointly funded the Agency for Science, Technology and Research, Singapore (https://www.a-star.edu.sg/) and Merck Sharp & Dohme Corp., Whitehouse Station, NJ USA (http://www.merck.com). The Singapore Malay Eye Study (SiMES) is supported by the National Medical Research Council (NMRC), Singapore (grants 0796/2003, 1176/2008, 1149/2008, STaR/0003/2008, 1249/2010, CG/SERI/2010, CIRG/1371/2013, and CIRG/1417/2015), and Biomedical Research Council (BMRC), Singapore (08/1/35/19/550 and 09/1/35/19/616).

Additional Information

Correspondence and Reprint Requests: Xueling Sim, Saw Swee Hock School of Public Health, Tahir Foundation Building, 12 Science Drive 2 #09-01G, Singapore 117549. E-mail: ephsx@nus.edu.sg.

Disclosure Summary: All authors have no competing interests for this work to disclose. D.F.R. was an employee of Merck Sharp & Dohme Corp. at the time of data generation and is currently with Janssen Inc.

Data Availability: The datasets generated during and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request. All summary statistics are available online (18).

References

1. American Diabetes Association. Classification and diagnosis of diabetes: standards of medical care in diabetes-2019. *Diabetes Care*. 2019;**42**(Suppl 1):S13-S28.
2. Bennett CM, Guo M, Dharmage SC. HbA(1c) as a screening tool for detection of Type 2 diabetes: a systematic review. *Diabet Med*. 2007;**24**(4):333-343.
3. Hou X, Liu J, Song J, et al. Relationship of hemoglobin A1c with β cell function and insulin resistance in newly diagnosed and drug naive Type 2 diabetes patients. *J Diabetes Res*. 2016;**2016**:8797316.
4. Cohen RM, Franco RS, Khera PK, et al. Red cell life span heterogeneity in hematologically normal people is sufficient to alter HbA1c. *Blood*. 2008;**112**(10):4284-4291.
5. English E, Idris I, Smith G, Dhatariya K, Kilpatrick ES, John WG. The effect of anaemia and abnormalities of erythrocyte indices on HbA1c analysis: a systematic review. *Diabetologia*. 2015;**58**(7):1409-1421.
6. Soranzo N, Sanna S, Wheeler E, et al.; WTCCC. Common variants at 10 genomic loci influence hemoglobin A_1(C) levels via glycemic and nonglycemic pathways. *Diabetes*. 2010;**59**(12):3229-3239.
7. Chen P, Ong RT, Tay WT, et al. A study assessing the association of glycated hemoglobin A1C (HbA1C) associated variants with HbA1C, chronic kidney disease and diabetic retinopathy in populations of Asian ancestry. *PLoS One*. 2013;**8**(11):e79767.
8. Wheeler E, Leong A, Liu CT, et al.; EPIC-CVD Consortium; EPIC-InterAct Consortium; Lifelines Cohort Study. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *Plos Med*. 2017;**14**(9):e1002383.
9. Ng NHJ, Willems SM, Fernandez J, et al. Tissue-specific alteration of metabolic pathways influences glycemic regulation. *bioRxiv*. Deposited October 3, 2019. https://doi.org/10.1101/790618.
10. Moon JY, Louie TL, Jain D, et al. A Genome-wide association study identifies blood disorder-related variants influencing hemoglobin A1c with implications for glycemic status in U.S. hispanics/latinos. *Diabetes Care*. 2019;**42**(9):1784-1791.
11. Leong A, Lim VJY, Wang C, et al. Association of G6PD variants with hemoglobin A1c and impact on diabetes diagnosis in East Asian individuals. *BMJ Open Diabetes Res Care*. 2020;**8**(1):e001091.
12. Ministry of Health Singapore. *National Health Survey 2010*. 2010. https://www.moh.gov.sg/resources-statistics/reports/national-health-survey-2010
13. Wong LP, Ong RT, Poh WT, et al. Deep whole-genome sequencing of 100 southeast Asian Malays. *Am J Hum Genet*. 2013;**92**(1):52-66.
14. Tan KHX, Tan LWL, Sim X, et al. Cohort profile: the Singapore Multi-Ethnic Cohort (MEC) study. *Int J Epidemiol*. 2018;**47**(3):699-699j.
15. Win AM, Yen LW, Tan KH, Lim RB, Chia KS, Mueller-Riemenschneider F. Patterns of physical activity and sedentary behavior in a representative sample of a multi-ethnic South-East Asian population: a cross-sectional study. *BMC Public Health*. 2015;**15**:318.
16. Foong AW, Saw SM, Loo JL, et al. Rationale and methodology for a population-based study of eye diseases in Malay people: The Singapore Malay eye study (SiMES). *Ophthalmic Epidemiol*. 2007;**14**(1):25-35.
17. Little RR, Rohlfing CL, Sacks DB, National Glycohemoglobin Standardization Program Steering Committee. Status of hemoglobin A1c measurement and goals for improvement: from chaos to order for improving diabetes care. *Clin Chem*. 2011;**57**(2):205-214.
18. Jin-Fang Chai S-LK, Chaolong W, Lim VJ-Y, et al. Data from: Genome-wide association for HbA1c in Malay identified deletion on *SLC4A1* that influences HbA1c independent of glycemia. *Asian Genetic Epidemiology Network*. https://blog.nus.edu.sg/agen/summary-statistics/hba1c-malay-slc4a1-2020/.
19. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 2010;**26**(22):2867-2873.
20. Teo YY, Sim X, Ong RT, et al. Singapore Genome Variation Project: a haplotype map of three Southeast Asian populations. *Genome Res*. 2009;**19**(11):2154-2162.
21. Sim X, Ong RT, Suo C, et al. Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. *Plos Genet*. 2011;**7**(4):e1001363.
22. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2012;**9**(2):179-181.
23. Auton A, Brooks LD, Durbin RM, et al.; Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015;**526**(7571):68-74.

24. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;**48**(10):1284-1287.
25. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*:13033997 [q-bioGN]. 2013. https://arxiv.org/abs/1303.3997
26. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;**43**(5):491-498.
27. Jun G, Flickinger M, Hetrick KN, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet*. 2012;**91**(5):839-848.
28. Dou J, Wu D, Ding L, et al. Using off-target data from whole-exome sequencing to improve genotyping accuracy, association analysis and polygenic risk prediction. [Published online ahead of print June 17, 2020]. *Brief Bioinform*. 2020;bbaa084. Doi: 10.1093/bib/bbaa084
29. Jun G, Wing MK, Abecasis GR, Kang HM. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res*. 2015;**25**(6):918-925.
30. Flickinger M, Jun G, Abecasis GR, Boehnke M, Kang HM. Correcting for sample contamination in genotype calling of DNA sequence data. *Am J Hum Genet*. 2015;**97**(2):284-290.
31. Browning BL, Yu Z. Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. *Am J Hum Genet*. 2009;**85**(6):847-861.
32. Kang HM, Sul JH, Service SK, et al. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet*. 2010;**42**(4):348-354.
33. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;**26**(17):2190-2191.
34. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;**26**(18):2336-2337.
35. Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation. *Bioinformatics*. 2015;**31**(5):782-784.
36. Chang X, Li J, Guo Y, et al. Genome-wide association study of serum minerals levels in children of different ethnic background. *PLoS One*. 2015;**10**(4):e0123499.
37. Bonder MJ, Kurilshikov A, Tigchelaar EF, et al. The effect of host genetics on the gut microbiome. *Nat Genet*. 2016;**48**(11):1407-1412.
38. Wu C, Kraft P, Zhai K, et al. Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. *Nat Genet*. 2012;**44**(10):1090-1097.
39. Wang Q, Pierce-Hoffman E, Cummings BB, et al.; Genome Aggregation Database Production Team; Genome Aggregation Database Consortium. Landscape of multi-nucleotide variants in 125 748 human exomes and 15 708 genomes. *Nat Commun*. 2020;**11**(1):2539.
40. Jarolim P, Palek J, Amato D, et al. Deletion in erythrocyte band 3 gene in malaria-resistant Southeast Asian ovalocytosis. *Proc Natl Acad Sci U S A*. 1991;**88**(24):11022-11026.
41. Jarolim P, Rubin HL, Zhai S, et al. Band 3 Memphis: a widespread polymorphism with abnormal electrophoretic mobility of erythrocyte band 3 protein caused by substitution AAG—GAG (Lys—Glu) in codon 56. *Blood*. 1992;**80**(6):1592-1598.
42. Spracklen CN, Shi J, Vadlamudi S, et al. Identification and functional analysis of glycemic trait loci in the China Health and Nutrition Survey. *Plos Genet*. 2018;**14**(4):e1007275.
43. Liu SC, Palek J, Yi SJ, et al. Molecular basis of altered red blood cell membrane properties in Southeast Asian ovalocytosis: role of the mutant band 3 protein in band 3 oligomerization and retention by the membrane skeleton. *Blood*. 1995;**86**(1):349-358.
44. Laosombat V, Viprakasit V, Dissaneevate S, et al. Natural history of Southeast Asian Ovalocytosis during the first 3 years of life. *Blood Cells Mol Dis*. 2010;**45**(1):29-32.
45. Ganesan J, George R, Lie-Injo LE. Abnormal haemoglobins and hereditary ovalocytosis in the Ulu Jempul District of Kuala Pilah, West Malaysia. *Southeast Asian J Trop Med Public Health*. 1976;**7**(3):430-433.
46. Amato D, Booth PB. Hereditary ovalocytosis in Melanesians. 1977. *Papua New Guinea Med J*. 2005;**48**(1-2):102-108.
47. Picard V, Proust A, Eveillard M, et al. Homozygous Southeast Asian ovalocytosis is a severe dyserythropoietic anemia associated with distal renal tubular acidosis. *Blood*. 2014;**123**(12):1963-1965.
48. Garnett C, Bain BJ. South-East Asian ovalocytosis. *Am J Hematol*. 2013;**88**(4):328.
49. *Management of Type 2 Diabetes Mellitus*. Malaysia: Ministry of Health; 2015. https://www.moh.gov.my/moh/resources/Penerbitan/CPG/Endocrine/3a.pdf
50. Lim WY, Ma S, Heng D, Tai ES, Khoo CM, Loh TP. Screening for diabetes with HbA1c: Test performance of HbA1c compared to fasting plasma glucose among Chinese, Malay and Indian community residents in Singapore. *Sci Rep*. 2018;**8**(1):12419.
51. Lek M, Karczewski KJ, Minikel EV, et al.; Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60 706 humans. *Nature*. 2016;**536**(7616):285-291.
52. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;**31**(13):3812-3814.
53. Zerbino DR, Achuthan P, Akanni W, et al. Ensembl 2018. *Nucleic Acids Res*. 2018;**46**(D1):D754-D761.