Progress in Defining the Genetic Contribution to Type 2 Diabetes in Individuals of East Asian Ancestry

Cassandra N. Spracklen[1] · Xueling Sim[2]

Abstract

Purpose of Review Prevalence of type 2 diabetes (T2D) and progression of complications differ between worldwide populations. While obesity is a major contributing risk factor, variations in physiological manifestations, e.g., developing T2D at lower body mass index in some populations, suggest other contributing factors. Early T2D genetic associations were mostly discovered in European ancestry populations. This review describes the progression of genetic discoveries associated with T2D in individuals of East Asian ancestry in the last 10 years and highlights the shared genetic susceptibility between the population groups and additional insights into genetic contributions to T2D.

Recent Findings Through increased sample size and power, new genetic associations with T2D were discovered in East Asian ancestry populations, often with higher allele frequencies than European ancestry populations.

Summary As we continue to generate maps of T2D-associated variants across diverse populations, there will be a critical need to expand and diversify other omics resources to enable integration for clinical translation.

Keywords Type 2 diabetes (T2D) · Genome-wide association study (GWAS) · East Asian populations · Meta-analysis · Susceptibility loci

Introduction

Type 2 diabetes (T2D) is a progressive, common metabolic disorder that results in the inability to regulate blood glucose in the body. Individuals with T2D are primarily characterized by inefficient use of insulin to regulate glucose in peripheral tissues such as muscles, liver, and fats, causing additional stress on the pancreas to produce more insulin. East Asian ancestry individuals develop T2D at lower body mass index (BMI) compared to European ancestry individuals, likely attributable to higher fat deposition at the abdominal area [1, 2]. It is unclear if the higher

This article is part of the Topical Collection on *Genetics*

✉ Cassandra N. Spracklen
cspracklen@umass.edu

✉ Xueling Sim
ephsx@nus.edu.sg

1 Department of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, 715 North Pleasant Street, 429 Arnold House, Amherst, MA 01002, USA

2 Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, 12 Science Drive 2, #10-01, Tahir Foundation Building, Singapore 117549, Singapore

proportion and location of body fat storage in East Asian individuals or effect modification with environmental and/or genetic factors account for these differences [3•].

Large-scale genome-wide association studies (GWAS) have been very successful in identifying over 500 genomic regions associated with T2D, explaining up to 20% of the disease heritability [4–7]. Historically, T2D GWAS and meta-analyses, like other complex traits, were predominantly undertaken in European ancestry populations, as a substantial number of samples were required to discover and replicate these findings reliably. As more genomic data have become available in non-European ancestry populations, particularly in populations of East Asian ancestry [6••, 8•, 9], the emerging pattern of shared variant-T2D associations across multiple ancestries supports the model of shared underlying causal genetic variants before out-of-Africa migration. However, additional loci have also been identified in non-European ancestry populations, primarily as a result of increased analytic power from higher allele frequencies, but also from heterogeneous effect sizes across populations and possible gene-by-environment and gene-by-gene interactions. Additionally, differences in linkage disequilibrium across populations can be leveraged at loci shared across populations to narrow association signals and identify candidate causal variants underlying

the associations. Full characterization of the genetics of T2D in different populations can identify the allelic spectrum of genetic variation in all ancestries and elevate established T2D-associated loci toward effective therapeutic targets.

In this review, we will summarize the progress in identifying T2D genetic associations in populations of East Asian ancestry and how these studies have advanced the current repertoire of T2D genetics beyond performing larger studies in European ancestry populations. In addition, we discuss the need for additional biologic and bioinformatic resources in non-European ancestry populations to enhance our functional understanding and improve our interpretation of the results.

Discovery of Genetic Loci in East Asian Ancestry Individuals from 2008 to 2020

T2D risk is strongly influenced by multiple genetic loci. Most of the T2D-associated loci were initially discovered in populations of European ancestry [10] as early as 2007 with three population-based T2D GWAS published back-to-back in Science [11–13]. Subsequent efforts to discover additional T2D loci in European ancestry populations involved large-scale meta-analyses of individual studies [14–16], the most recent of which was published in 2018 and identified 403 distinct signals at 243 loci in nearly 900,000 individuals [5]. As T2D genetic association data emerged in non-European ancestry populations, trans-ancestry meta-analyses also revealed more shared susceptibility between global populations [4, 7, 17, 18].

The earliest T2D GWAS performed in individuals of East Asian ancestry were two parallel papers led by independent Japanese groups published simultaneously in Nature Genetics in 2008 [19, 20]. Using data from genotype arrays, both studies independently identified novel T2D associations at/near the gene KCNQ1, a potassium voltage-gated channel subfamily Q member gene, on chromosome 11 that had not yet been described in any of the European ancestry analyses. Yasuda et al. [19] reported common intronic variant rs2237892 to be associated with T2D. Unoki et al. [20] simultaneously described three additional common variants at the same locus (rs2283228, rs2237897, and rs2237895) in moderate to high LD with rs2237892 ($r[2]$ in 1000G Phase 3 East Asians: 0.23 to 0.84). The minor allele frequency for all four identified variants, which were identified in analyses of <10,000 T2D cases, occur at a much lower frequency in European ancestry populations than East Asian ancestry populations (e.g., rs2237892: EAS minor allele frequency [MAF]=35.5%, EUR MAF=6.3%; Table 1). Associations near KCNQ1 were not reported in European ancestry analyses until 2010 when the number of T2D cases included in meta-analyses increased to more than 42,000 [14]. The European lead variant, rs231362, maps 150kb away from the lead variants in East

Asian ancestry individuals [14] [LD between rs2237892 and rs231362: EAS $r[2]$ = 0.0031; EUR $r[2]$ = 0.0001]. KCNQ1 remains the strongest T2D-associated locus in East Asian ancestry populations to date, with more than 17 distinct signals identified in the latest East Asian T2D meta-analysis [6] [rs2237897 (risk allele frequency, RAF = 37%), OR= 1.31 (95% CI 1.29 – 1.33), P-value of 2.0×10^{-246}]. Following the two initial Japanese publications in 2008, three separate East Asian GWAS published in 2010 described six additional novel T2D-associated loci, including SRR [21], PTPRD [21], UBE2E2 [22], C2CD4A/B [22], and SPRY2 [23]. Associations at/near UBE2E2, C2CD4A/B, and SPRY2 were subsequently replicated in European ancestry analyses [5].

The first large-scale T2D genome-wide meta-analysis in East Asian ancestry populations assembled 6952 cases and 11,865 controls in the discovery phase from China, Japan, Korea, the Philippines, Singapore, and Taiwan, as part of the Asian Genetic Epidemiology Network (AGEN) [8]. Utilizing the HapMap2 CHB+JPT [24] imputation reference panel, the final meta-analysis sample size after in silico and de novo replications included 25,079 T2D cases and 29,611 controls and reported a total of eight novel loci at/near MAEA, GLIS3, FITM2/R3HDML/HNF4A, GCC1/PAX4, PSMD6, ZFAND3, PEPD, and KCNK16. The strongest novel association was a common variant at the MAEA locus with a risk allele frequency of 58% in East Asian ancestry populations but only 3% in European ancestry populations, further demonstrating the ease of discovery and improved power due to differences in risk allele frequencies. Of the eight novel loci, three were previously implicated in type 1 diabetes (T1D; GLIS3) [25] or maturity onset diabetes of the young (MODY; HNF4A [26] and PAX4 [27, 28]). GLIS3 plays an important role in human beta cell development and function and is associated with both T1D [25] and T2D [29] with consistent directions of effect. At the GCC1/PAX4 locus on chromosome 7, the association signal was near a cluster of genes, including GCC1, containing a GRIP domain, and PAX4 [27, 28], a transcription factor known to play a role in pancreatic islet development. Conditional and sensitivity analyses indicate that these association signals are independent of previously known MODY variants and not likely to be attributable to misclassification of T2D cases. The lead variant at the KCNK16 locus was a common intronic variant near the GLP1R gene that encodes a receptor for glucagon-like peptide 1 (GLP1), a pancreatic beta-cell receptor involved in blood sugar homeostasis and insulin secretion. GLP1R is also a target of GLP-1 receptor agonists, a class of medications used for the treatment of T2D. Less is known about the function and role of MAEA with regards to T2D pathogenesis, but PSMD6, ZFAND3, and PEPD are all related to pancreatic beta-cell function and/or insulin secretion [30–32]. As for GWAS of other common diseases, most of the T2D-associated lead variants identified in either European or East Asian ancestry analyses are located

Table 1 Summary of T2D genetic association studies in East Asian individuals

Year	Publication	Variant density	Sample size	New loci	Lead variant-risk allele	1000G Phase 3 risk allele frequency	
						EAS	EUR
2008	Yasuda K, Miyake K, Horikawa Y, et al. Variants in *KCNQ1* are associated with susceptibility to type 2 diabetes mellitus. Nat Genet. 2008 Sep;40(9):1092-7.	100,000 (JSNP Genome Scan)	*Discovery* in Japanese: 1612 cases + 1424 controls; *Replication* in trans-ethnic samples: 9569 cases + 10,361 controls	*KCNQ1*	rs2237892-C	0.645	0.937
2008	Unoki H, Takahashi A, Kawaguchi T, et al. SNPs in *KCNQ1* are associated with susceptibility to type 2 diabetes in East Asian and European populations. Nat Genet. 2008 Sep;40(9):1098-102.	100,000 (JSNP Genome Scan)	*Multi-stage* in Japanese: 5118 cases + 4176 controls; *Replication* in trans-ethnic samples: 8846 cases + 7943 controls	*KCNQ1*	rs2283228-A rs2237897-C rs2237895-C	0.620 0.647 0.330	0.932 0.953 0.419
2010	Tsai FJ, Yang CF, Chen CC, et al. A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. PLoS Genet. 2010 Feb 19;6(2):e1000847.	Illumina 550	*Discovery* in Taiwanese: 995 cases + 894 controls; *Replication* in Taiwanese: 1803 cases + 1473 controls	*SRR* *PTPRD*	rs391300-G rs17584499-T	0.679 0.139	0.614 0.184
2010	Yamauchi T, Hara K, Maeda S, et al. A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at *UBE2E2* and *C2CD4A-C2CD4B*. Nat Genet. 2010 Oct;42(10):864-8.	HapMap imputed	*Multi-stage* in Japanese: 10,978 cases + 8514 controls; *Replication* in trans-ethnic samples : 11,164 cases + 12,769 controls	*UBE2E2* *C2CD4A-C2CD4B*	rs7612463-C rs7172432-A	0.809 0.629	0.868 0.588
	Shu XO, Long J, Cai Q, et al. Identification of new genetic risk variants for type 2 diabetes. PLoS Genet. 2010 Sep 16;6(9):e1001127.	HapMap imputed	*Discovery* in Chinese: 1019 cases + 1710 controls; *Replication* in trans-ethnic samples: 8775 cases + 12,905 controls	*SPRY2*	rs1359790-G	0.722	0.733
2011	Cho YS, Chen CH, Hu C, et al. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat Genet. 2011 Dec 11;44(1):67-72.	HapMap imputed	*Discovery*: 6952 cases + 11,865 controls; *In silico replication*: 5843 cases + 4574 controls; *De novo replication*: 12,284 cases + 13,172 controls	*MAEA* *GLIS3* *HNF4A* *PAX4* *PSMD6* *ZFAND3* *PEPD* *KCNK16*	rs6815464-C rs7041847-A rs6017317-G rs6467136-G rs831571-C rs9470794-C rs3786897-A rs1535500-T	0.575 0.501 0.434 0.795 0.632 0.290 0.597 0.467	0.974 0.533 0.163 0.501 0.792 0.094 0.586 0.493
2013	Hara K, Fujita H, Johnson TA, et al. Genome-wide association study identifies three novel loci for type 2 diabetes. Hum Mol Genet. 2014 Jan 1;23(1):239-46.	1000G Phase 1 EAS	*Discovery*: 5976 cases + 20,829 controls; *In silico replication*: 2799 cases + 3793 controls; *De novo replication*: 10,319 cases + 6795 controls	*MIR129-LEP* *GPSM1* *SLC16A13*	rs791595-A rs11787792-A rs312457-G	0.125 0.944 0.099	0.178 0.673 0.019
2016	Imamura M, Takahashi A, Yamauchi T, et al. Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes. Nat Commun. 2016 Jan 28;7:10531.	1000G Phase 1 EAS	*Discovery* in Japanese: 15,463 cases + 26,183 controls; *Replication* in trans-ethnic samples: 80,314 cases + 158,030 controls	*CCDC85A* *FAM60A* *DMRTA1* *ASB3* *ATP8B2* *MIR4686*	rs1116357-G rs147538848-A rs1575972-T rs9309245-G rs67156297-A rs7107784-G	0.281 0.189 0.958 0.204 0.912 0.121	0.535 0.005 0.968 0.353 0.250 0.280

Table 1 (continued)

Year Publication	Variant density	Sample size	New loci	Lead variant-risk allele	1000G Phase 3 risk allele frequency	
					EAS	EUR
2019 Suzuki K, Akiyama M, Ishigaki K, et al. Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. Nat Genet. 2019 Mar;51(3):379-386.	1000G Phase 3 ALL	*Discovery* in Japanese: 36,614 cases + 155,150 controls	INAFM2	rs67839313-C	0.248	0.114
			USP48	rs1825307-A	0.264	0.254
			FANCL/BCL11A	rs13417036-A	0.188	0.161
			SCTR	rs3731600-C	0.988	1.00
			EPC2	rs16829174-A	0.501	0.806
			DGKD	rs838720-G	0.363	0.371
			CASR/CD86	rs2134223-A	0.489	0.869
			GNPDA2/GABRG1	rs10938398-A	0.298	0.420
			GRSF1/MOB1B	rs12649012-G	0.804	0.044
			AGPAT9/NKX6-1	rs201597274-T	0.980	0.993
			PARP8/EMB	rs16884025-C	0.943	1.00
			ITGA1	rs11410487-A	0.153	0.069
			FGFR4/ZNF346	rs3135911-A	0.466	0.291
			NUS1	rs80196932-T	0.169	0.801
			ETV1/ARL4A	rs56805921-C	0.403	0.395
			AUTS2	rs6947395-T	0.209	0.137
			C10orf11/ZNF503	rs7900112-T	0.436	0.401
			WDR11/FGFR2	rs10886863-C	0.611	0.968
			SLC1A2	rs2421897-G	0.125	0.984
			ETS1	rs11819995-T	0.186	0.222
			KSR2	rs79310463-T	0.215	0.122
			DLEU7/KCNRG	rs123378-G	0.162	0.339
			GP2	rs11267808-A	0.064	0.020
			ZNF257	rs14816037-T	0.953	1.000
			NFATC2	rs602276-T	0.327	0.340
			CNKSR2	rs663421-A	0.702	0.829
			TIMM17B/PCSK1N	rs782100977-C	0.601	0.436
			SPIN2A/FAAH2	rs144226500-AT	0.387	0.275
			IL13RA1	rs11390176-T	0.461	0.341
2020 Spracklen CN, Horikoshi M, Kim YJ, et al. Identification of type 2 diabetes loci in 433,540 East Asian individuals	1000G Phase 3 ALL	*Discovery* in East Asians: 77,418 cases and 356,122 controls	VWA5B1	rs60573766-C	0.595	0.884
			MAST2	rs562138031-C	0.725	0.735
			PGM1	rs2269245-G	0.791	0.767
			TSEN15	rs1327123-C	0.450	0.327
			MDM4	rs201297151-CAAAAAAAA	0.418	0.713
			SIX3	rs12712928-C	0.445	0.133
			IKZF2	rs75179644-T	0.962	0.999
			ZBTB20	rs6806156-T	0.567	0.120
			TFRC	rs9866168-T	0.696	0.426
			RANBP3L	rs16902871-G	0.202	0.000
			PCSK1	rs6556925-C	0.467	0.610
			REPS1	rs9376382-C	0.662	0.291
			HIVEP2	rs9390022-T	0.778	0.602
			ZNF713	rs56050730-GA	0.298	0.250

Table 1 (continued)

Year Publication	Variant density	Sample size	New loci	Lead variant-risk allele	1000G Phase 3 risk allele frequency	
					EAS	EUR
			STEAP1	rs62469016-C	0.191	0.501
			CALCR	rs2074120-A	0.290	0.521
			GRM8/PAX4	rs11737118-G	0.110	0.000
			ASAH1	rs34642578-T	0.056	0.020
			ZNF703	rs4739515-G	0.504	0.944
			FGFR1	rs328301-T	0.367	0.355
			KCNB2	rs349359-C	0.246	0.129
			GDAP1	rs149265787-G	0.017	0.018
			TRIB1	rs60089934-A	0.432	0.020
			EFR3A	rs73708054-C	0.317	0.105
			DMRT2	rs1016565-A	0.434	0.334
			PTCH1	rs113154802-C	0.892	0.902
			ABCA1	rs201375651-CA	0.394	0.851
			PTF1A	rs77065181-A	0.051	0.000
			ARID5B	rs141583966-G	0.934	0.960
			JMJD1C	rs148928116-T	0.738	0.982
			ARHGAP19	rs10736116-C	0.289	0.364
			BBIP1	rs7895872-T	0.534	0.765
			BDNF	rs4922793-A	0.511	0.765
			FAIM2	rs77978149-T	0.081	0.003
			ALDH2	rs149212747-A	0.848	1.000
			RBM19	rs7307263-G	0.516	0.561
			FGF9	rs9316706-A	0.335	0.816
			NYNRIN	rs12437434-C	0.722	0.989
			LRRC74A	rs58524310-G	0.297	0.143
			DLK1/MEG3/miRNA cluster	rs73347525-A	0.728	0.813
			TRAF3	rs55700915-A	0.398	0.229
			HERC2	rs76704029-T	0.703	0.894
			MYO5C	rs149336329-G	0.961	0.976
			RGMA	rs61021634-A	0.381	0.906
			IGF1R	rs79826452-A	0.864	0.000
			PKD1L3	rs12600132-T	0.448	0.369
			ZFHX3	rs6416749-C	0.385	0.324
			SUMO2	rs35559984-CA	0.686	0.534
			ZNF799	rs4804181-A	0.460	0.765
			ZNRF3	rs147413364-T	0.331	0.296
			WNT7B	rs28637892-T	0.183	0.126

in the non-coding regions of the genome, resulting in a more difficult path to identifying candidate causal variants and genes and/or disease biology.

Advances in DNA sequencing technologies and the availability of denser reference panels for imputation are two important developments in human genetics that, along with increasing sample sizes, have allowed for refinement of existing association signals and discovery of new association signals. With the first release of the 1000 Genomes Project Phase 1 reference panel and increase in sample sizes [33], at least 10 additional T2D-associated loci were reported in Japanese populations [34, 35]. In 2013, Hara et al. [34] reported three novel loci at *MIR219-LEP* (within 1Mb from *PAX4*), *GPSM1*, and *SLC16A13* in a discovery cohort of 5976 Japanese T2D cases and 20,829 controls [34]. *LEP* encodes leptin, a hormone that regulates appetite [36], and increased leptin levels are associated with obesity and T2D [37]. At the *SLC16A13* locus, the strongest association was at a common intronic variant, rs312457. Just 5 kb away, a meta-analysis of Mexican and other Latin American ancestry populations reported a T2D association at a common synonymous variant, rs13342232, at *SLC16A11* [38]. High LD between the two lead variants (1000G EAS r [2]=0.93 and AMR r [2]=0.80) suggests the two ancestries share the same association signal. A risk haplotype at *SLC16A11* accounting for ~20% of the increased prevalence in T2D in Mexican American individuals [38] and further functional experiments [39] suggest *SLC16A11* as the more likely causal candidate gene at the locus. In a separate analysis, the discovery samples from Hara et al. [34] were subsequently combined with an additional 9817 Japanese T2D cases and 6763 controls, bringing the discovery meta-analysis to >15,000 T2D cases and >26,000 controls [35]. An additional seven new loci were reported, including *FAM60A* (now known as *SINHCAF*), a master transcriptional repressor that regulates a response to hypoxia and influences an in vitro model of angiogenesis [40], and *INS/IGF2/TH*, which is independent of *KCNQ1* (*INS/IGF2* located ~620kb upstream).

The T2D-GENES consortium (Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples) focused on the protein-coding sequence in five major ancestry groups (African American, East Asian, European, Hispanic, and South Asian) [41•, 42]. Association analyses were performed for each ancestry, and ancestry-specific association results were combined into a trans-ancestry meta-analysis. Across all single-ancestry analyses, only one missense variant reached genome-wide significance: Arg192His within *PAX4* identified in 2184 East Asian ancestry individuals. The lead variant, rs2233580, is common in individuals of East Asian ancestry (MAF ~10%) but almost absent in other ancestries [43]. The rs2233580 variant was not well imputed and not in LD with the lead T2D variant at the previously reported *GCC1/PAX4* locus from the Cho et al. [8] AGEN meta-

analysis (1000G EAS r [2]=0.028). The association between rs2233580 and T2D was subsequently validated in other East Asian studies using either the exome array [44] or exome sequencing [45]. To date, this missense *PAX4* variant remains associated with T2D only in East Asian ancestry populations as they are monomorphic in other ancestry populations. Though *PAX4* has been implicated in MODY, the consistent and replicable associations in multiple East Asian studies and lack of association with age of diagnosis suggest multiple variants at/near *PAX4* that are likely playing a role in the pathogenesis of T2D among East Asian ancestry individuals [41•, 45].

As large biobanks mature in East Asia, available sample sizes for genome-wide meta-analyses are greatly surpassing previous efforts. In 2019, a Japanese T2D meta-analysis combined 36,614 T2D cases and 155,150 controls from Biobank Japan and identified 88 T2D-associated loci, of which 28 were novel [9]. Lead variants at 68 loci were common in both the Japanese population and European ancestry samples included in the 1000G Phase 3 reference panel. The meta-analysis also identified 28 missense variants within 21 genes that were in high LD (EAS r [2]≥0.80) with the lead T2D variants. Among the 28 missense variants, at least 8 missense variants were less frequent (0.01<MAF<0.05) in Japanese populations and rare (MAF<0.01) or monomorphic in European ancestry individuals. For instance, at one of the novel loci, *SCTR*, rs3731600 (p.Ala122Pro; MAF=0.05) is polymorphic only in individuals of East Asian ancestry. In a second example, a missense variant rs3765467 (p.Arg131Gln), at the previously reported fasting glucose-associated locus *GLP1R* [46, 47], is common in East Asian ancestry individuals (MAF=0.23) but is extremely rare in European ancestry populations (MAF=0.001) [48]. The *GLP1R* association signal is independent of associated variants at the neighboring *KCNK16* locus [8], and the protective allele is associated with increased GLP1-induced insulin secretion [49]. Trans-ancestry comparisons of molecular pathway analyses highlight shared (e.g., MODY and beta cell) and heterogenous (e.g., NOTCH signaling and insulin secretion) molecular pathways in Japanese and European ancestry populations.

The advent of biobanks in China (China Kadoorie Biobank CKB) [50], Japan (BioBank Japan BBJ) [9], and South Korea (Korea Biobank Array, KBA) [51], along with continually denser reference panels (e.g., 1000G Phase 3) allowed for continued improvement in revealing the genetic architecture of T2D in East Asian ancestry populations. Most recently, the AGEN consortium performed a T2D genome-wide meta-analysis with data imputed to the denser and larger 1000 Genomes Phase 3 reference panel [48] and included up to 433,530 individuals (77,418 T2D cases and 356,122 controls) from 23 studies across six East Asian countries [6]. Association models were performed with and without adjustment for BMI to account for obesity, one of the biggest risk

factors for T2D. A total of 301 T2D association signals at 183 loci were identified, of which 61 were novel for any ancestry after accounting for previously published loci [5, 9]. Association effect sizes were also highly correlated between East Asian and European ancestry individuals [5] at T2D-associated loci, supporting evidence of substantial shared T2D susceptibility between the two ancestries. Loci with large differences in effect sizes between the two ancestries are generally rare (MAF ≤0.1%) in European ancestry populations but common (e.g., *PAX4*, *RANBP3L*) or less frequent (e.g., *ZNF257*, *DGKD*) in East Asian ancestry populations (Table 1). In T2D association analyses adjusted for BMI, we identified four additional novel loci, including *NID2*, which exhibits other associations consistent with a lipodystrophy phenotype of increased T2D risk but lower BMI and higher triglycerides in East Asian ancestry individuals [52, 53]. The substantial increase in sample size also allowed for sex-stratified analyses, revealing a male-specific T2D association at the *ALDH2* locus with the strongest evidence of heterogeneity between sexes. Furthermore, conditional analyses established independence of multiple signals at many of the T2D-associated loci, for example, seven independent signals at the *PAX4* locus including the adjacent *LEP* signal first reported in Hara et al. [34], and two independent signals mapping to *GLP1R* and *KCNK16* on chromosome 6.

Biologic and Mechanistic Insights into T2D Pathophysiology Identified from East Asian Analyses

GWAS and meta-analyses have been an efficient tool for identifying loci associated with T2D in all populations, elucidating novel mechanisms for disease development and effective therapeutic targets. Even in the early days of GWAS, analyses in East Asian ancestry populations identified loci not observed in European ancestry populations that yielded new biological insights into T2D risk. We highlight five examples below.

Locus with Multiple Association Signals with Effects Mediated Through Multiple Tissues (Shared Susceptibility Across Ancestries)

A variant within the *ANK1* locus, rs515071, was initially identified as a T2D-associated locus in a combined meta-analysis of East Asian and European ancestry individuals in 2012 [54]. Associated variants in and near *ANK1* and nearby *NKX6-3* were subsequently replicated in numerous meta-analyses of single- and trans-ancestry populations [5, 9, 15, 17, 22, 35], including multiple distinct association signals [5, 9]. Skeletal muscle *cis*-eQTL analyses initially identified *ANK1* as the effector transcript at this locus (EUR lead GWAS variant: rs508419) [55]; this result was also observed in subcutaneous

adipose tissue [56]. Conversely, Mahajan et al. [5] used credible set and pancreatic islet *cis*-eQTL results to suggest *NXK6-3* as the effector transcript at this locus (EUR lead GWAS variant: rs13262861). However, in the most recent meta-analysis of East Asian ancestry individuals, two of three distinct association signals (also identified in European ancestry individuals [5]) were found to co-localize with *cis*-eQTLs for *NKX6-3* in pancreatic islets [57] (EAS lead GWAS variant: rs33981001) and *ANK1* in both subcutaneous adipose tissue [56] and skeletal muscle [55] (EAS lead GWAS variant: rs62508166) (Table 2), providing evidence that variants at this locus are acting on different genes in different tissues [6].

Loci with Variants Present at Higher Frequencies in East Asian Individuals Compared to Non-East Asian Ancestry Populations

At the *ZNF257* locus, rs142395395 was recently reported to be associated with T2D in East Asian ancestry individuals (OR 1.24, 95% CI 1.19–1.29) [6, 9]. With a MAF around 4%, this variant tags a previously described inversion of 415 kb that disrupts the coding sequence and expression of *ZNF257*, as well as lymphoblastoid expression of 81 downstream genes and transcripts [58]. However, rs142395395 has only been reported twice in 29,828 non-East Asian ancestry individuals from the gnomAD database [43]. While the exact biologic mechanism and/or effector gene underlying the locus is still to be determined, disrupting the expression of a transcription factor with a substantial number of downstream targets could impact T2D risk in multiple ways.

Variants at/near the *PAX4* locus have consistently demonstrated association with T2D in East Asian ancestry populations [6, 8, 9, 34, 35]. This locus was originally identified as a T2D susceptibility gene in a candidate gene study of Japanese individuals [59]. Recent GWAS meta-analyses in individuals of East Asian ancestry expanded the locus to include multiple signals both upstream, near *GCC1* [6, 8, 9], and downstream, near *LEP* [6, 9, 34, 35], of *PAX4*. In the largest East Asian ancestry T2D GWAS meta-analyses, two missense variants in codon 192 of *PAX4*, rs2233580 (Arg192His; RAF=8.6%; OR 1.31, 95% CI 1.28–1.34; $P=3.4\times10^{-93}$) and rs3824004 (Arg192Ser; RAF=3.4%; OR 1.24, 95% CI 1.19–1.28; $P=1.1\times10^{-30}$), are the strongest association signals in the region [6, 9, 45]. Both variants are extremely rare in non-East Asian ancestry populations [60]. *PAX4* is a pancreatic islet transcription factor that is known to play a critical role in differentiating embryonic pancreatic progenitor cells into beta-cells [61]. Additionally, consequences of targeted disruption of *pax4* in mice include a fatal absence of mature b-cells that results in severe insulin deficient diabetes and death within the first few days of life [62]. Downstream of *PAX4*, *LEP* encodes the leptin protein, a hormone secreted by white adipose tissue [63] that plays a key role in glucose metabolism

Table 2   Colocalization between expression quantitative trait loci and loci with evidence of association with T2D in East Asian individuals

Lead variant	Chr	Position (hg19)	Nearest protein-coding gene (kb)	# of genes within 100 kb	Coding variants (EAS $r[2]>0.80$)	Colocalized eQTL gene (tissue, P-value)
rs11205766	1	51,191,935	FAF1 (0.0)	1	-	CDKN2C (Sk M, 10^{-10})
rs1327123	1	184,014,593	TSEN15 (0.7)	2	TSEN15 (Q59H)	TSEN15 (S Ad, 10^{-22})
rs238763	1	229,642,499	NUP133 (0.0)	4	-	ABCB10 (Islets, 10^{-5})
rs1260326	2	27,730,940	GCKR (0.0)	7	-	EMLIN1 (S Ad, 10^{-6}); ZNF513 (WB, 10^{-9})
rs12712928	2	45,192,080	SIX3 (18.8)	3	-	SIX3 (Islets, 10^{-23}; Pan, 10^{-12}); SIX3-AS1 (Islets, 10^{-9}; Pan, 10^{-10}); SIX2 (Islets, 10^{-9}); RP11-89K21.1 (Islets, 10^{-13}); AC012354.6 (Islets, 10^{-20}; Pan, 10^{-11})
rs67114627	3	63,904,715	ATXN7 (0.0)	4	-	SCAANT1 (Liv, 10^{-8}; WB, 10^{-14})
rs6806156	3	114,968,018	ZBTB20 (101.9)	1	-	ZBTB20 (Islets, 10^{-36}; V Ad, 10^{-7}; WB, 10^{-5})
rs13092876	3	185,495,320	IGF2BP2 (0.0)	1	-	IGF2BP2 (S Ad, 10^{-4})
rs11332772	3	186,649,931	ST6GAL1 (0.0)	2	-	ST6GAL1 (Pan, 10^{-13})
rs9998835	4	6,293,237	WFS1 (0.0)	3	WFS1 (R228R; V395V; N500K; N500N; K811N; K811K; S855S)	WFS1 (Sk M, 10^{-18}; S Ad, 10^{-12}; V Ad, 10^{-9})
rs28599782	4	71,844,118	MOB1B (0.0)	2	-	DCK (S Ad, 10^{-10}); M0B1B (WB, 10^{-13})
rs2408653	5	56,276,198	MIER3 (28.2)	2	-	MIER3 (WB, 10^{-7})
rs3135911	5	176,513,896	FGFR4 (0.02)	4	FGFR4 (V10I, G388R*)	FGFR4 (S Ad 10^{-8}; V Ad, 10^{-9})
rs572817245	6	32,442,964	HLA-DRA (31.2)	5	-	HLA-DRB1 (Islets, 10^{-10})
rs4711389	6	34,214,670	SMIM29 (0.0)	5	-	NUDT3 (Liv, 10^{-7})
rs35099099	6	39,284,857	KCNK16 (0.0)	4	KCNK16 (N786N; N786K)	KCNK17 (Islets, 10^{-11}; S Ad, 10^{-10}l; WB, 10^{-33}; PBC, 10^{-17})
rs80196932	6	117,996,631	NUS1 (0.0)	2	-	NUS1 (Islets, 10^{-20}; Pan, 10^{-43}; Sk M, 10^{-7})
rs2800734	6	127,417,035	RSPO3 (23.3)	1	-	RNF146 (Sk M, 10^{-5})
rs7739842	6	131,954,797	ENPP3 (0.0)	4	-	MED23 (Sk M, 10^{-15}; WB, 10^{-160}; PBMC, 10^{-7})
rs9390022	6	143,056,556	HIVEP2 (16.0)	1	-	HIVEP2 (S Ad, 10^{-4})
rs7787720	7	13,886,654	ETV1 (44.2)	1	-	ETV1 (Islets, 10^{-5})
rs10950550	7	15,063,923	DGKB (49.4)	1	-	DGKB (Islets, 10^{-15})
rs12698877	7	69,696,905	AUTS2 (0.0)	1	-	AUTS2 (Islets, 10^{-6})
rs56687477	8	36,832,310	KCNU1 (38.8)	1	-	ERLIN2 (S Ad, 10^{-6}); ZNF703 (S Ad, 10^{-14}); RP11-419C23.1 (S Ad, 10^{-8})
rs328301	8	38,343,012	FGFR1 (17.6)	3	-	RP11-350N15.4 (WB, 10^{-10})
rs62508166	8	41,537,318	ANK1 (0.0)	3	-	ANK1 (Sk M, 10^{-9}; S Ad, 10^{-10})
rs896852	8	95,960,886	NDUFAF6, TP53INP1 (0.0)	4	-	NDUFAF6 (S Ad, 10^{-11}); TP53INP1 (WB, 10^{-225}); RP11-347C18.3 (WB, 10^{-5})
rs4237150	9	4,290,085	GLIS3 (0.0)	1	-	INSL4 (S Ad, 10^{-4})
rs113154802	9	98,278,413	PTCH1 (0.0)	1	-	PTCH1 (S Ad, 10^{-8}; WB, 10^{-252})
rs529565	9	136,149,500	ABO (0.0)	7	-	ABO (Islets, 10^{-59}); RP11-430N14.4 (Islets, 10^{-55})
rs78270318	9	139,245,389	GPSM1 (0.0)	8	-	CARD9 (Islets, 10^{-24}); DNLZ (S Ad, 10^{-56}; Hyp, 10^{-14}; Islets, 10^{-22}; Liv, 10^{-13}; Pan, 10^{-35}; Sk M, 10^{-90}; V Ad, 10^{-38}; WB, 10^{-310}); SDCCAG3 (Sk M, 10^{-6})
rs10870136	9	139,302,900	SDCCAG3, ENTR1 (0.0)	10	INPP5E (T416T; G428G; L1259L); ENTR1 (G374G)	INPP5E (Hyp, 10^{-12}; Pan, 10^{-5}; WB, 10^{-310}); SEC16A (Islets, 10^{-10}; WB, 10^{-77})
rs11257657	10	12,309,139	CDC123 (16.8)	4	-	CAMK1D (PBMC, 10^{-9})
rs7900112	10	77,314,617	LRMDA (0.0)	1	-	KAT6B (S Ad, 10^{-4})
rs35906730	10	94,435,673	HHEX (14.1)	2	-	EIF2S2P3 (WB, 10^{-23}); KIF11 (WB, 10^{-13})
rs10748694	10	99,056,190	ARHGAP19 (3.8)	4	-	ARHGAP19 (Islets, 10^{-5}; S Ad, 10^{-11}); SLIT1 (WB, 10^{-52})
rs7895872	10	112,678,657	BBIP1 (0.0)	4	-	RP11-313D6.3 (WB, 10^{-8}); BBIP1 (WB, 10^{-23})
rs7901695	10	114,754,088	TC7L2 (0.0)	1	-	TCF7L2 (Islets, 10^{-7})
rs112820281	10	124,150,342	PLEKHA1 (0.0)	4	-	PLEKHA1 (S Ad, 10^{-10}; V Ad, 10^{-9})
rs808964	11	2,386,600	CD81 (12.2)	7	-	CD81 (S Ad, 10^{-14})
rs7925578	11	2,729,947	KCNQ1 (0.0)	1	-	KCNQ1OT1 (PBMC, 10^{-5})
rs4148646	11	17,415,190	ABCC8 (0.0)	4	KCNJ11 (V337I, K23E); ABCC8 (A1369S)	NCR3LG1 (S Ad; 10^{-16}); KCNJ11 (WB, 10^{-37})
rs7109575	11	72,463,435	ARAP1 (0.0)	5	-	PDE2A (Islets, 10^{-6}); STARD10 (Islets, 10^{-32})
rs10830963	11	92,708,710	MTNR1B (0.0)	2	-	MTNR1B (Islets, 10^{-10})
rs3217854	12	4,398,349	CCND2 (0.0)	3	-	CCND2 (WB, 10^{-15}); TIGAR (WB, 10^{-14})
rs3751236	12	27,963,402	KLHL42 (12.3)	3	-	KLHL42 (Sk M, 10^{-5}; WB, 10^{-217}; PBMC, 10^{-6}); RP11-860B13. (Islets, 10^{-7})
rs7313668	12	71,449,521	TSPAN8 (69.6)	1	-	CTD-2021H9.3 (Sk M, 10^{-24})
rs1616443	12	120,922,892	DYNLL1 (0.0)	8	-	GATC (S Ad, 10^{-128}; Islets, 10^{-5}; Pan, 10^{-12}; WB, 10^{-287}); POP5 (Hyp, 10^{-6}; S Ad 10^{-8}); COQ5 (Sk M, 10^{-37})
rs6492947	15	40,761,549	BAHD1 (3.2)	8	-	BAHD1 (WB, 10^{-26}); CHST14 (WB, 10^{-22})
rs149336329	15	52,587,740	MYO5C (0.0)	2	-	

Table 2 (continued)

Lead variant	Chr	Position (hg19)	Nearest protein-coding gene (kb)	# of genes within 100 kb	Coding variants (EAS r [2]>0.80)	Colocalized eQTL gene (tissue, P-value)
rs7171507	15	75,737,287	*SIN3A* (0.0)	4	-	*MYO5C* (Islets, 10^{-15}); Liv, 10^{-16}; Pan, 10^{-6}; Sk M, 10^{-9}; WB, 10^{-270}); *MAN2C1* (S Ad, 10^{-177}; Hyp, 10^{-18}; Liv, 10^{-18}; Pan, 10^{-78}; Sk M, 10^{-20}; V Ad, 10^{-144}; WB, 10^{-185}; PBMC, 10^{-11}); *RP11-817O13.8* (Liv, 10^{-8}; Pan, 10^{-5}; Sk M, 10^{-7}; S Ad, 10^{-51}; V Ad, 10^{-31}; WB, 10^{-30}; PBMC, 10^{-10}); *SNUPN* (Pan, 10^{-8}; WB, 10^{-310}); *IMP3* (Pan, 10^{-4}; Sk M, 10^{-5}; WB, 10^{-55}); *NEIL1* (S Ad, 10^{-10}; V Ad, 10^{-7}); *RP11-817O13.6* (WB, 10^{-23}); *PTPN9* (WB, 10^{-9})
rs952471	15	77,776,498	*HMG20A* (0.0)	2	-	*HMG20A* (S Ad, 10^{-16}; Islets, 10^{-15})
rs10852123	15	90,428,894	*AP3S2* (0.0)	3	-	*AP3S2* (Islets, 10^{-59}; Hyp, 10^{-7}; Liv, 10^{-28}; Pan, 10^{-50}; Sk M, 10^{-50}; S Ad, 10^{-96}; V Ad, 10^{-79}; WB, 10^{-310}); *ARPIN-AP3S2* (Islets, 10^{-63}); *C15orf38 (S Ad, 10-10; WB, 10-14)*; *RP11-493E3.1* (WB, 10^{-7})
rs8026714	15	91,522,253	*PRC1* (0.0)	8	*PRC1* (T282T)	*RCCD1* (S Ad, 10^{-44}; Islets, 10^{-11}; Hyp, 10^{-10}; Pan, 10^{-10}; S Ad, 10^{-58}; V Ad, 10^{-58}; WB, 10^{-15}); *PRC1-AS1* (Pan, 10^{-5}; S Ad, 10^{-13}; V Ad, 10^{-13})
rs2925979	16	81,534,790	*CMIP* (0.0)	1	-	*CMIP* (S Ad, 10^{-8})
rs7502556	17	29,642,430	*NF1* (0.0)	3	-	*EVI2A* (S Ad, 10^{-9}; V Ad, 10^{-8}; WB, 10^{-310}); *NF1* (Islets, 10^{-5}; S Ad, 10^{-7}; Sk M, 10^{-4}); *RP11-218M11.1* (Sk M, 10^{-4})
rs77706698	17	65,953,348	*BPTF* (0.0)	3	-	*C17orf58* (S Ad, 10^{-13})
rs143488340	19	46,413,789	*NANOS2* (4.2)	8	-	*AC093503.2* (Islets, 10^{-5})
rs6021276	20	50,155,386	*NFATC2* (0.0)	2	-	*NFATC2* (Islets, 10^{-8})
rs28691713	22	50,356,302	*PIM3* (0.0)	5	*IL17REL* (L333P)	*PIM3* (WB, 10^{-215})

1000G Phase 3 was used as the reference panel to identify coding variants in linkage disequilibrium (LD) r [2] >0.80 (East Asians) with the index T2D variant. T2D-associated variants were considered to be colocalized with an eQTL if the lead T2D-associated variant was in LD r [2]>0.80 with the lead eQTL variant (1000Gp3; ancestry matched the eQTL dataset). eQTL data were used from the following sources: BioBank Japan (unfractioned peripheral blood; http://jenger.riken.jp/en/result) [86]; eQTLGen (whole blood) [86]; GTEx (hypothalamus, liver, pancreas, skeletal muscle, subcutaneous adipose, visceral adipose, whole blood) [82]; InsPIRE (islets) [85]; METSIM (Subcutaenous Adipose) [56]; Parker Lab Islet Browser (islets) [84]

Abbreviations: chr, chromosome; EAS, East Asian; Hyp, hypothalamus; kb, kilobases; Liv, liver; Pan, pancreas; S Ad, subcutaneous adipose; Sk M, skeletal muscle; V Ad, visceral adipose; WB, whole blood

[64]. Mice with leptin deficiencies are characterized by both insulin resistance and diabetes [65], and epidemiological studies have consistently demonstrated leptin to be associated with risk for T2D [66, 67]. It is currently unclear if the different association signals are acting independently or in tandem, but both *PAX4* and *LEP* are strong candidates for effector genes at this East Asian–specific locus.

The associations between variants near *SIX3/SIX2* and T2D and related glycemic traits were first reported in East Asian ancestry populations. The variant rs895636 (risk allele frequency (RAF) =42.8%) was reported to be associated with fasting plasma glucose, a T2D-related trait, in a GWAS of >17,000 Korean and Japanese individuals [68] and subsequently replicated in other East Asian ancestry GWAS and meta-analyses [69, 70]. Additionally, a variant in strong LD, rs12712928 (r [2]=0.91; RAF=40.2%) was significantly associated with HbA1c, another T2D-related trait, in a recent glycemic trait meta-analysis of East Asian ancestry individuals [69]. Furthermore, rs12712928 was recently identified in the Spracklen et al. [6] AGEN T2D meta-analysis. However, despite the common allele frequency of rs12712928 in other ancestry populations (European RAF=13.3%, African American and Hispanic RAF=24.2%, South Asian RAF=26.7%), no associations have yet been observed between variants near *SIX3/SIX2* and T2D in non-East Asian ancestry populations [4, 5, 7, 69]. *SIX3* and *SIX2* encode pancreatic islet transcription factors that are known to play roles in islet beta-cell functions [71] by coordinately regulating the expression of target genes that govern mature beta-cell function (e.g., insulin production) and maintain beta-cell fate [72].

Locus with Environmental Interactions

The *ALDH2* locus spans almost 2Mb, reflecting a recently selected sweep in individuals of East Asian ancestry where carriers of the alternate allele experience discomfort such as flushing, nausea, and headache following alcohol consumption [73]. The causal variant in *ALDH2* that alters alcohol metabolism, rs671 (Glu504Lys), is polymorphic only in East Asian ancestry individuals and has also been associated with cardiometabolic traits including blood pressure, BMI, HDL-cholesterol, LDL cholesterol, and cardiovascular disease risk

[74–76]. Furthermore, a Mendelian randomization study in Chinese individuals suggests evidence of a non-causal effect of moderate alcohol drinking on the risk of myocardial infarction [77]. The association with T2D at this locus is heterogeneous between sexes, highly significant only in males (P=5.8×10^{-27}) and not in females (P=0.19) [6]. The strong sex-heterogeneity observed suggests possible gene-environment interactions with alcohol consumption patterns between sexes or alcohol effects on BMI and/or insulin sensitivity [78]. The highly pleiotropic nature of rs671 and important detoxification role of alcohol dehydrogenases suggest *ALDH2* as a possible therapeutic target for T2D and other cardiometabolic diseases [79].

Identifying Biologic Mechanisms at Loci Found in Non-European Ancestry Populations

While the process for identifying functional variants, effector transcripts, and directions of effect can be complex for loci identified within any ancestry, it can be especially difficult for non-European ancestry populations for several reasons. First, while bioinformatic resources exist for determining the variant effect on biological pathways (e.g., PASCAL, DEPICT) [80, 81], the base programs rely on linkage disequilibrium (LD) calculations derived from a European ancestry reference population, and users are required to input custom reference genotype information for other ancestries. Second, target genes can be predicted by colocalizing GWAS variants with gene expression (eQTL) in disease-relevant tissues, DNA methylation that regulates gene expression (mQTL), or protein abundance (pQTL) data. However, the majority of these omics data generated to date are in populations of European ancestry, which can exhibit limited transferability to other ancestry populations when LD structure and/or allele frequencies are heterogenous.

As an illustrative example, we interrogated gene expression datasets for hypothalamus, liver, pancreas, pancreatic islets, peripheral blood, skeletal muscle, subcutaneous and visceral adipose, and whole blood tissues for colocalization (LD r [2] between the lead GWAS and lead eQTL variant >0.80 in European ancestry populations as eQTL data were generated from European ancestry individuals) at all 301 AGEN T2D association signals and lead eQTL variants (Table 2) [56, 82–86]. Overall, 60 (19.9%) of the T2D GWAS variants colocalized with at least one gene in at least one tissue. However, 43 (14.3%) of the variants were not tested for association with gene expression in any of the datasets used for colocalization, likely because these variants failed quality metrics or MAF thresholds in the European ancestry eQTL studies due to allele frequency differences between the two ancestral populations (more frequent in EAS but less frequent or rare in EUR).

Similarly, with the Genotype-Tissue Expression Project (GTEx) [87], the most widely used eQTL dataset in the field, we were unable to test 95 (31.6%) of the variants identified in the AGEN T2D analysis. The colocalization analyses suggest that T2D in East Asian ancestry individuals appears to be more related to insulin resistance (e.g., adipose tissue, skeletal muscle) mechanisms rather than insulin secretion (pancreas). To enable translation of more of the genetic loci identified in East Asian ancestry populations into disease insights and therapeutic targets, additional non-European ancestry bioinformatic and genomic resources need to be generated and available for use.

Conclusion

In summary, GWAS across multiple population groups have identified more than 500 T2D-associated loci over the past 15 years, providing mechanistic insights into T2D pathophysiology. While analyses in European ancestry populations now include >1 million individuals [7], analyses in non-European ancestry individuals, including East Asian populations, continue to identify unique genetic signals due to allele frequency differences. Unfortunately, the differences in allele frequencies across ancestry groups also indicate that European ancestry–based resources for identifying associated variants (e.g., reference panels), effector transcripts (e.g., eQTLs, meQTLs, pQTLs), and biological pathways are insufficient for non-European ancestry populations. There is a clear need for denser reference panels comprised of deeply sequenced East Asian ancestry individuals to refine association signals and identify rare variant associations. Additionally, other omics data and bioinformatics resources (e.g., transcriptomics, epigenomics, proteomics, metabolomics) generated in East Asian ancestry individuals are necessary to further translate genetic findings to effective therapeutic targets.

Declarations

References

Papers of particular interest, published recently, have been highlighted as:

• Of importance

•• Of major importance

1. Odegaard AO, Koh WP, Vazquez G, et al. BMI and diabetes risk in Singaporean Chinese. Diabetes Care. 2009;32:1104–6. https://doi.org/10.2337/dc08-1674.

2. Hsu WC, Araneta MR, Kanaya AM, et al. BMI cut points to identify at-risk Asian Americans for type 2 diabetes screening. Diabetes Care. 2015;38:150–8. 2014. https://doi.org/10.2337/dc14-2391.

3.• Ma RC, Chan JC. Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. Ann N Y Acad Sci. 2013;1281:64–91. https://doi.org/10.1111/nyas.12098 **This review provides a detailed comparison on the epidemiology of type 2 diabetes and its complications between East Asian and European populations.**

4. Mahajan A, Spracklen CN, Zhang W, et al.. Trans-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. medRxiv 2020: 2020.2009.2022.20198937. https://doi.org/10.1101/2020.09.22.20198937.

5. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet. 2018;50:1505–13. https://doi.org/10.1038/s41588-018-0241-6.

6.•• Spracklen CN, Horikoshi M, Kim YJ, et al. Identification of type 2 diabetes loci in 433,540 East Asian individuals. Nature. 2020. https://doi.org/10.1038/s41586-020-2263-3 **This article, to date, is the largest T2D genome-wide meta-analysis in East Asian populations that combined more than 77K cases and 356K controls.**

7. Vujkovic M, Keaton JM, Lynch JA, et al. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. Nat Genet. 2020;52:680–91. https://doi.org/10.1038/s41588-020-0637-y.

8.• Cho YS, Chen CH, Hu C, et al. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat Genet. 2011;44:67–72. https://doi.org/10.1038/ng.1019 **This article is the first large-scale T2D genome-wide meta-analysis in East Asian populations.**

9. Suzuki K, Akiyama M, Ishigaki K, et al. Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. Nat Genet. 2019;51:379–86. https://doi.org/10.1038/s41588-018-0332-4.

10. Flannick J, Florez JC, et al. Nat Rev Genet. 2016;17:535–49. https://doi.org/10.1038/nrg.2016.56.

11. Diabetes Genetics Initiative of Broad Institute of H, Mit LU, Novartis Institutes of BioMedical R, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science. 2007;316:1331–6. https://doi.org/10.1126/science.1142358.

12. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science, 2007;316:1341–5. https://doi.org/10.1126/science.1142382.

13. Zeggini E, Weedon MN, Lindgren CM, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science. 2007;316:1336–41. https://doi.org/10.1126/science.1142364.

14. Voight BF, Scott LJ, Steinthorsdottir V, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010;42:579–89. https://doi.org/10.1038/ng.609.

15. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012;44:981–90. https://doi.org/10.1038/ng.2383.

16. Mahajan A, Wessel J, Willems SM, et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. Nat Genet. 2018;50:559–71. https://doi.org/10.1038/s41588-018-0084-1.

17. Replication DIG, Meta-analysis C, Asian Genetic Epidemiology Network Type 2 Diabetes C, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet. 2014;46:234–44. https://doi.org/10.1038/ng.2897.

18. Zhao W, Rasheed A, Tikkanen E, et al. Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. Nat Genet. 2017;49:1450–7. https://doi.org/10.1038/ng.3943.

19. Yasuda K, Miyake K, Horikawa Y, et al. Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. Nat Genet. 2008;40:1092–7. https://doi.org/10.1038/ng.207.

20. Unoki H, Takahashi A, Kawaguchi T, et al. SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. Nat Genet. 2008;40:1098–102. https://doi.org/10.1038/ng.208.

21. Tsai FJ, Yang CF, Chen CC, et al. A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. PLoS Genet. 2010;6:e1000847. https://doi.org/10.1371/journal.pgen.1000847.

22. Yamauchi T, Hara K, Maeda S, et al. A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. Nat Genet. 2010;42:864–8. https://doi.org/10.1038/ng.660.

23. Shu XO, Long J, Cai Q, et al. Identification of new genetic risk variants for type 2 diabetes. PLoS Genet. 2010;6:e1001127. https://doi.org/10.1371/journal.pgen.1001127.

24. International HapMap C, Frazer KA, Ballinger DG, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449:851–61. https://doi.org/10.1038/nature06258.

25. Barrett JC, Clayton DG, Concannon P, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet. 2009;41:703–7. https://doi.org/10.1038/ng.381.

26. Nakajima H, Yoshiuchi I, Hamaguchi T, et al. Hepatocyte nuclear factor-4 alpha gene mutations in Japanese non-insulin dependent diabetes mellitus (NIDDM) patients. Res Commun Mol Pathol Pharmacol. 1996;94:327–30.

27. Plengvidhya N, Kooptiwut S, Songtawee N, et al. PAX4 mutations in Thais with maturity onset diabetes of the young. J Clin Endocrinol Metab. 2007;92:2821–6. https://doi.org/10.1210/jc.2006-1927.

28. Jo W, Endo M, Ishizu K, et al. A novel PAX4 mutation in a Japanese patient with maturity-onset diabetes of the young. Tohoku J Exp Med. 2011;223:113–8. https://doi.org/10.1620/tjem.223.113.

29. Aylward A, Chiou J, Okino ML, et al. Shared genetic risk contributes to type 1 and type 2 diabetes etiology. Hum Mol Genet. 2018. https://doi.org/10.1093/hmg/ddy314.

30. Rende D, Baysal N, Kirdar B. Complex disease interventions from a network model for type 2 diabetes. PLoS One. 2013;8:e65854. https://doi.org/10.1371/journal.pone.0065854.

31. Zheng JS, Huang T, Li K, et al. Modulation of the association between the PEPD variant and the risk of type 2 diabetes by n-3

fatty acids in Chinese Hans. J Nutrigenet Nutrigenomics. 2015;8: 36–43. https://doi.org/10.1159/000381348.

32. Ndiaye FK, Ortalli A, Canouil M, et al. Expression and functional assessment of candidate type 2 diabetes susceptibility genes identify four new genes contributing to human insulin secretion. Mol Metab. 2017;6:459–70. https://doi.org/10.1016/j.molmet.2017.03. 011.

33. Genomes Project C, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491:56–65. https://doi.org/10.1038/nature11632.

34. Hara K, Fujita H, Johnson TA, et al. Genome-wide association study identifies three novel loci for type 2 diabetes. Hum Mol Genet. 2014;23:239–46. https://doi.org/10.1093/hmg/ddt399.

35. Imamura M, Takahashi A, Yamauchi T, et al. Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes. Nat Commun. 2016;7:10531. https://doi.org/10. 1038/ncomms10531.

36. Klok MD, Jakobsdottir S, Drent ML. The role of leptin and ghrelin in the regulation of food intake and body weight in humans: a review. Obes Rev. 2007;8:21–34. https://doi.org/10.1111/j.1467-789X.2006.00270.x.

37. Rasmussen-Torvik LJ, Wassel CL, Ding J, et al. Associations of body mass index and insulin resistance with leptin, adiponectin, and the leptin-to-adiponectin ratio across ethnic groups: the Multi-Ethnic Study of Atherosclerosis (MESA). Ann Epidemiol. 2012;22:705–9. https://doi.org/10.1016/j.annepidem.2012.07.011.

38. Consortium STD, Williams AL, Jacobs SB, et al. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. Nature. 2014;506:97–101. https://doi.org/10.1038/nature12828.

39. Rusu V, Hoch E, Mercader JM, et al. Type 2 diabetes variants disrupt function of SLC16A11 through two distinct mechanisms. Cell. 2017;170:199–212 e120. https://doi.org/10.1016/j.cell.2017. 06.011.

40. Biddlestone J, Batie M, Bandarra D, et al. SINHCAF/FAM60A and SIN3A specifically repress HIF-2alpha expression. Biochem J. 2018;475:2073–90. https://doi.org/10.1042/BCJ20170945.

41.• Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. Nature. 2016;536:41–7. https://doi.org/10.1038/nature18642. **This article is the first trans-ancestry T2D meta-analysis for whole-exome sequencing data that included populations of African-American, East Asian, European, Hispanic, and South Asian ancestries.**

42. Flannick J, Mercader JM, Fuchsberger C, et al. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. Nature. 2019;570:71–6. https://doi.org/10.1038/s41586-019-1231-2.

43. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536: 285–91. https://doi.org/10.1038/nature19057.

44. Cheung CY, Tang CS, Xu A, et al. Exome-chip association analysis reveals an Asian-specific missense variant in PAX4 associated with type 2 diabetes in Chinese individuals. Diabetologia. 2017;60:107–15. https://doi.org/10.1007/s00125-016-4132-z.

45. Kwak SH, Chae J, Lee S, et al. Nonsynonymous variants in PAX4 and GLP1R are associated with type 2 diabetes in an East Asian population. Diabetes. 2018;67:1892–902. https://doi.org/10.2337/db18-0361.

46. Wessel J, Chu AY, Willems SM, et al. Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. Nat Commun. 2015;6:5897. https://doi.org/10. 1038/ncomms6897.

47. Mahajan A, Sim X, Ng HJ, et al. Identification and functional characterization of G6PC2 coding variants influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. PLoS Genet. 2015;11:e1004876. https://doi.org/10.1371/journal. pgen.1004876.

48. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. Nature. 2015;526:68–74. https://doi. org/10.1038/nature15393.

49. Sathananthan A, Man CD, Micheletto F, et al. Common genetic variation in GLP1R and insulin secretion in response to exogenous GLP-1 in nondiabetic subjects: a pilot study. Diabetes Care. 2010;33:2074–6. https://doi.org/10.2337/dc10-0200.

50. Chen Z, Chen J, Collins R, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol. 2011;40:1652–66. https://doi.org/10.1093/ije/dyr120.

51. Moon S, Kim YJ, Han S, et al. The Korea Biobank Array: design and identification of coding variants associated with blood biochemical traits. Sci Rep. 2019;9:1382. https://doi.org/10.1038/s41598-018-37832-9.

52. Wen W, Zheng W, Okada Y, et al. Meta-analysis of genome-wide association studies in East Asian-ancestry populations identifies four new loci for body mass index. Hum Mol Genet. 2014;23: 5492–504. https://doi.org/10.1093/hmg/ddu248.

53. Akiyama M, Okada Y, Kanai M, et al. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. Nat Genet. 2017;49:1458–67. https://doi.org/10.1038/ng.3951.

54. Imamura M, Maeda S, Yamauchi T, et al. A single-nucleotide polymorphism in ANK1 is associated with susceptibility to type 2 diabetes in Japanese populations. Hum Mol Genet. 2012;21:3042–9. https://doi.org/10.1093/hmg/dds113.

55. Scott LJ, Erdos MR, Huyghe JR, et al. The genetic regulatory signature of type 2 diabetes in human skeletal muscle. Nat Commun. 2016;7:11764. https://doi.org/10.1038/ncomms11764.

56. Civelek M, Wu Y, Pan C, et al. Genetic regulation of adipose gene expression and cardio-metabolic traits. Am J Hum Genet. 2017;100:428–43. https://doi.org/10.1016/j.ajhg.2017.01.027.

57. van de Bunt M, Manning Fox JE, Dai X, et al. Transcript expression data from human islets links regulatory signals from genome-wide association studies for type 2 diabetes and glycemic traits to their downstream effectors. PLoS Genet. 2015;11:e1005694. https://doi. org/10.1371/journal.pgen.1005694.

58. Puig M, Castellano D, Pantano L, et al. Functional impact and evolution of a novel human polymorphic inversion that disrupts a gene and creates a fusion transcript. PLoS Genet. 2015;11: e1005495. https://doi.org/10.1371/journal.pgen.1005495.

59. Shimajiri Y, Sanke T, Furuta H, et al. A missense mutation of Pax4 gene (R121W) is associated with type 2 diabetes in Japanese. Diabetes. 2001;50:2864–9. https://doi.org/10.2337/diabetes.50.12. 2864.

60. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434–43. https://doi.org/10.1038/s41586-020-2308-7.

61. Mauvais-Jarvis F, Smith SB, Le May C, et al. PAX4 gene variations predispose to ketosis-prone diabetes. Hum Mol Genet. 2004;13: 3151–9. https://doi.org/10.1093/hmg/ddh341.

62. Sosa-Pineda B, Chowdhury K, Torres M, et al. The Pax4 gene is essential for differentiation of insulin-producing beta cells in the mammalian pancreas. Nature. 1997;386:399–402. https://doi.org/10.1038/386399a0.

63. Cammisotto PG, Bendayan M. Leptin secretion by white adipose tissue and gastric mucosa. Histol Histopathol. 2007;22:199–210. https://doi.org/10.14670/hh-22.199.

64. Meek TH, Morton GJ. The role of leptin in diabetes: metabolic effects. Diabetologia. 2016;59:928–32. https://doi.org/10.1007/s00125-016-3898-3.

65. Dubuc PU. The development of obesity, hyperinsulinemia, and hyperglycemia in ob/ob mice. Metabolism. 1976;25:1567–74. https://doi.org/10.1016/0026-0495(76)90109-8.

66. Bidulescu A, Dinh PC Jr, Sarwary S, et al. Associations of leptin and adiponectin with incident type 2 diabetes and interactions among African Americans: the Jackson heart study. BMC Endocr Disord. 2020;20:31. https://doi.org/10.1186/s12902-020-0511-z.

67. Schmidt MI, Duncan BB, Vigo A, et al. Leptin and incident type 2 diabetes: risk or protection? Diabetologia. 2006;49:2086–96. https://doi.org/10.1007/s00125-006-0351-z.

68. Kim YJ, Go MJ, Hu C, et al. Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. Nat Genet. 2011;43:990–5. https://doi.org/10.1038/ng.939.

69. Chen J, Spracklen CN, Marenne G, et al.. The trans-ancestral genomic architecture of glycaemic traits. bioRxiv 2020: 2020.2007.2023.217646. 10.1101/2020.07.23.217646.

70. 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:e1007275. https://doi.org/10.1371/journal.pgen.1007275.

71. Arda HE, Li L, Tsai J, et al. Age-dependent pancreatic gene regulation reveals mechanisms governing human β cell function. Cell Metab. 2016;23:909–20. https://doi.org/10.1016/j.cmet.2016.04.002.

72. Bevacqua RJ, Lam JY, Peiris H, et al.. SIX2 and SIX3 coordinately regulate functional maturity and fate of human pancreatic β cells. bioRxiv 2020: 2020.2012.2003.411033. https://doi.org/10.1101/2020.12.03.411033.

73. Okada Y, Momozawa Y, Sakaue S, et al. Deep whole-genome sequencing reveals recent selection signatures linked to evolution and disease risk of Japanese. Nat Commun. 2018;9:1631. https://doi.org/10.1038/s41467-018-03274-0.

74. Kato N, Takeuchi F, Tabara Y, et al. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. Nat Genet. 2011;43:531–8. https://doi.org/10.1038/ng.834.

75. Takeuchi F, Isono M, Nabika T, et al. Confirmation of ALDH2 as a major locus of drinking behavior and of its variants regulating multiple metabolic phenotypes in a Japanese population. Circ J. 2011;75:911–8. https://doi.org/10.1253/circj.cj-10-0774.

76. Xu F, Chen Y, Lv R, et al. ALDH2 genetic polymorphism and the risk of type II diabetes mellitus in CAD patients. Hypertens Res. 2010;33:49–55. https://doi.org/10.1038/hr.2009.178.

77. Millwood IY, Walters RG, Mei XW, et al. Conventional and genetic evidence on alcohol and vascular disease aetiology: a prospective study of 500 000 men and women in China. Lancet. 2019;393: 1831–42. https://doi.org/10.1016/S0140-6736(18)31772-0.

78. Schrieks IC, Heil AL, Hendriks HF, et al. The effect of alcohol consumption on insulin sensitivity and glycemic status: a systematic review and meta-analysis of intervention studies. Diabetes Care. 2015;38:723–32. https://doi.org/10.2337/dc14-1556.

79. Chen CH, Ferreira JC, Gross ER, et al. Targeting aldehyde dehydrogenase 2: new therapeutic opportunities. Physiol Rev. 2014;94: 1–34. https://doi.org/10.1152/physrev.00017.2013.

80. Lamparter D, Marbach D, Rueedi R, et al. Fast and rigorous computation of gene and pathway scores from SNP-based summary statistics. PLoS Comput Biol. 2016;12:e1004714. https://doi.org/10.1371/journal.pcbi.1004714.

81. Pers TH, Karjalainen JM, Chan Y, et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun. 2015;6:5890. https://doi.org/10.1038/ncomms6890.

82. Gamazon ER, Segrè AV, van de Bunt M, et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. Nat Genet. 2018;50:956–67. https://doi.org/10.1038/s41588-018-0154-4.

83.• Ishigaki K, Kochi Y, Suzuki A, et al. Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid arthritis. Nat Genet. 2017;49:1120–5. https://doi.org/10.1038/ng.3885 **This article describes the Biobank Japan expression quantitative trait locus data in peripheral blood, the only known/publicly available East Asian ancestry eQTL dataset for any tissue.**

84. Varshney A, Scott LJ, Welch RP, et al. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. Proc Natl Acad Sci U S A. 2017;114:2301–6. https://doi.org/10.1073/pnas.1621192114.

85. Viñuela A, Varshney A, van de Bunt M, et al. Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. Nat Commun. 2020;11:4912. https://doi.org/10.1038/s41467-020-18581-8.

86. Võsa U, Claringbould A, Westra H-J, et al.. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. bioRxiv 2018: 447367. 10.1101/447367.

87. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 2020; 369: 1318-1330. 2. https://doi.org/10.1126/science.aaz1776.