bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Identification of type 2 diabetes loci in 433,540 East Asian individuals

Cassandra N Spracklen[1,83], Momoko Horikoshi[2,83], Young Jin Kim[3,83], Kuang Lin[4,83], Fiona Bragg[4], Sanghoon Moon[3], Ken Suzuki[2,5,6,7], Claudia HT Tam[8,9], Yasuharu Tabara[10], Soo-Heon Kwak[11], Fumihiko Takeuchi[12], Jirong Long[13], Victor JY Lim[14], Jin-Fang Chai[14], Chien-Hsiun Chen[15], Masahiro Nakatochi[16], Jie Yao[17], Hyeok Sun Choi[18], Apoorva K Iyengar[1], Hannah J Perrin[1], Sarah M Brotman[1], Martijn van de Bunt[19,20], Anna L Gloyn[19,20,21], Jennifer E Below[22,23], Michael Boehnke[24], Donald W Bowden[25,26], John C Chambers[27,28,29,30,31], Anubha Mahajan[19,20], Mark I McCarthy[19,20,21,*], Maggie CY Ng[22,25], Lauren E Petty[22,23], Weihua Zhang[28,29], Andrew P Morris[20,32,33], Linda S Adair[34], Zheng Bian[35], Juliana CN Chan[8,9,36,37], Li-Ching Chang[15], Miao-Li Chee[38], Yii-Der Ida Chen[17], Yuan-Tsong Chen[15], Zhengming Chen[4], Lee-Ming Chuang[39,40], Shufa Du[34], Penny Gordon-Larsen[34], Myron Gross[41], Xiuqing Guo[17], Yu Guo[35], Sohee Han[3], Annie-Green Howard[42], Wei Huang[43], Yi-Jen Hung[44,45], Mi Yeong Hwang[3], Chii-Min Hwu[46,47], Sahoko Ichihara[48], Masato Isono[12], Hye-Mi Jang[3], Guozhi Jiang[8,9], Jost B Jonas[49], Yoichiro Kamatani[5,50], Tomohiro Katsuya[51,52], Takahisa Kawaguchi[10], Chiea-Chuen Khor[53,54], Katsuhiko Kohara[55], Myung-Shik Lee[56], Nannette R Lee[57], Liming Li[58], Jianjun Liu[53,59], Andrea O Luk[8,9], Jun Lv[58], Yukinori Okada[7,60], Mark A Pereira[61], Charumathi Sabanayagam[38,62,63], Jinxiu Shi[43], Dong Mun Shin[3], Wing Yee So[8,36], Atsushi Takahashi[5,64], Brian Tomlinson[8], Fuu-Jen Tsai[65], Rob M van Dam[14,59,66], Yong-Bing Xiang[67], Ken Yamamoto[68], Toshimasa Yamauchi[6], Kyungheon Yoon[3], Canqing Yu[58], Jian-Min Yuan[69,70], Liang Zhang[38], Wei Zheng[13], Michiya Igase[71], Yoon Shin Cho[18], Jerome I Rotter[72], Ya-Xing Wang[73], Wayne HH Sheu[45,47,74], Mitsuhiro Yokota[75], Jer-Yuarn Wu[15], Ching-Yu Cheng[38,62,63], Tien-Yin Wong[38,62,63], Xiao-Ou Shu[13], Norihiro Kato[12], Kyong-Soo Park[11,76,77], E-Shyong Tai[14,59,78], Fumihiko Matsuda[10], Woon-Puay Koh[14,79], Ronald CW Ma[8,9,36,37], Shiro Maeda[2,80,81], Iona Y Millwood[4,82], Juyoung Lee[3], Takashi Kadowaki[6,84*], Robin G Walters[4,82,84*], Bong-Jo Kim[3,84*], Karen L Mohlke[1,84*], Xueling Sim[14,84*]

1 Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
2 Laboratory for Endocrinology, Metabolism and Kidney Diseases, RIKEN Centre for Integrative Medical Sciences, Yokohama, Japan
3 Division of Genome Research, Center for Genome Science, National Institute of Health, Cheongju-si, Korea
4 Nuffield Department of Population Health, University of Oxford, Oxford, UK
5 Laboratory for Statistical and Translational Genetics, RIKEN Centre for Integrative Medical Sciences, Yokohama, Japan
6 Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
7 Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan
8 Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China
9 Chinese University of Hong Kong-Shanghai Jiao Tong University Joint Research Centre in Diabetes Genomics and Precision Medicine, The Chinese University of Hong Kong, Hong Kong, China
10 Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan
11 Department of Internal Medicine, Seoul National University Hospital, Seoul, South Korea
12 Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan
13 Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
14 Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore
15 Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

16 Department of Nursing, Nagoya University Graduate School of Medicine, Nagoya University Hospital, Nagoya, Japan
17 Department of Pediatrics, The Institute for Translational Genomics and Population Sciences, LABioMed at Harbor-UCLA Medical Center, Torrance, CA, USA
18 Biomedical Science, Hallym University, Chuncheon, South Korea
19 Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK
20 Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK
21 Oxford NIHR Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, Churchill Hospital, Oxford, UK
22 Vanderbilt Genetics Institute, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
23 Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA
24 Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA
25 Center for Genomics and Personalized Medicine Research, Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, USA
26 Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA
27 Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
28 Department of Epidemiology and Biostatistics, Imperial College London, London, UK
29 Department of Cardiology, Ealing Hospital, London North West Healthcare NHS Trust, Middlesex, UK
30 Imperial College Healthcare NHS Trust, Imperial College London, London, UK
31 MRC-PHE Centre for Environment and Health, Imperial College London, London, UK
32 Department of Biostatistics, University of Liverpool, Liverpool, UK
33 School of Biological Sciences, University of Manchester, Manchester, UK
34 Department of Nutrition, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
35 Chinese Academy of Medical Sciences, Beijing, China
36 Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong, Hong Kong, China
37 Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China
38 Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore
39 Division of Endocrinology & Metabolism, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan
40 Institute of Preventive Medicine, School of Public Health, National Taiwan University, Taipei, Taiwan
41 Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA
42 Department of Biostatistics, Carolina Population Center, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
43 Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai, China
44 Division of Endocrine and Metabolism, Tri-Service General Hospital Songshan Branch, Taipei, Taiwan
45 School of Medicine, National Defense Medical Center, Taipei, Taiwan
46 Section of Endocrinology and Metabolism, Department of Medicine, Taipei Veterans General Hospital, Taipei, Taiwan

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

47 School of Medicine, National Yang-Ming University, Taipei, Taiwan
48 Department of Environmental and Preventive Medicine, Jichi Medical University School of Medicine, Shimotsuke, Japan
49 Department of Ophthalmology, Medical Faculty Mannheim of the University of Heidelberg, Mannheim, Germany
50 Laboratory of Complex Trait Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan
51 Department of Clinical Gene Therapy, Osaka University Graduate School of Medicine, Osaka, Japan
52 Department of Geriatric and General Medicine, Graduate School of Medicine, Osaka University, Osaka, Japan
53 Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore
54 Department of Biochemistry, National University of Singapore, Singapore, Singapore
55 Department of Regional Resource Management, Ehime University Faculty of Collaborative Regional Innovation, Ehime, Japan
56 Severance Biomedical Science Institute and Department of Internal Medicine, Yonsei University College of Medicine, Seoul, South Korea
57 Department of Anthropology, Sociology and History, University of San Carlos, Cebu City, Philippines
58 Departments of Epidemiology & Biostatistics, Peking University Health Science Centre, Peking University, Beijing, China
59 Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore
60 Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Osaka, Japan
61 Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA
62 Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, Singapore
63 Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore
64 Department of Genomic Medicine, National Cerebral and Cardiovascular Center, Osaka, Japan
65 Department of Medical Genetics and Medical Research, China Medical University Hospital, Taichung, Taiwan
66 Department of Nutrition, Harvard T.H Chan School of Public Health, Boston, MA, USA
67 State Key Laboratory of Oncogene and Related Genes & Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China
68 Department of Medical Biochemistry, Kurume University School of Medicine, Kurume, Japan
69 Division of Cancer Control and Population Sciences, UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA, USA
70 Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA
71 Department of Anti-aging Medicine, Ehime University Graduate School of Medicine, Ehime, Japan
72 Departments of Pediatrics and Medicine, The Institute for Translational Genomics and Population Sciences, LA BioMed at Harbor-UCLA Medical Center, Torrance, CA, USA
73 Beijing Institute of Ophthalmology, Ophthalmology and Visual Sciences Key Laboratory, Beijing Tongren Hospital, Capital Medical University, Beijing, China

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

74 Division of Endocrinology and Metabolism, Department of Medicine, Taichung Veterans General Hospital, Taichung, Taiwan
75 Kurume University School of Medicine, Kurume, Japan
76 Department of Internal Medicine, Seoul National University College of Medicine, Seoul, South Korea
77 Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, South Korea
78 Duke-NUS Medical School, Singapore, Singapore
79 Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore
80 Department of Advanced Genomic and Laboratory Medicine, Graduate School of Medicine, University of the Ryukyus, Okinawa, Japan
81 Division of Clinical Laboratory and Blood Transfusion, University of the Ryukyus Hospital, Okinawa, Japan
82 Medical Research Council Population Health Research Unit, University of Oxford, Oxford, UK
Current address: Genentech, San Francisco, CA, USA

83 These authors contributed equally to this work.
84 These authors jointly supervised this work.
* Corresponding author

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

SUMMARY

Meta-analyses of genome-wide association studies (GWAS) have identified >240 loci associated with type 2 diabetes (T2D), however most loci have been identified in analyses of European-ancestry individuals. To examine T2D risk in East Asian individuals, we meta-analyzed GWAS data in 77,418 cases and 356,122 controls. In the main analysis, we identified 298 distinct association signals at 178 loci, and across T2D association models with and without consideration of body mass index and sex, we identified 56 loci newly implicated in T2D predisposition. Common variants associated with T2D in both East Asian and European populations exhibited strongly correlated effect sizes. New associations include signals in/near *GDAP1*, *PTF1A*, *SIX3, ALDH2,* a microRNA cluster, and genes that affect muscle and adipose differentiation. At another locus, eQTLs at two overlapping T2D signals act through two genes, *NKX6-3* and *ANK1*, in different tissues. Association studies in diverse populations identify additional loci and elucidate disease genes, biology, and pathways.

Type 2 diabetes (T2D) is a common metabolic disease primarily caused by insufficient insulin production and/or secretion by the pancreatic β cells and insulin resistance in peripheral tissues[1]. Most genetic loci associated with T2D have been identified in populations of European (EUR) ancestry, including a recent meta-analysis of genome-wide association studies (GWAS) of nearly 900,000 individuals of European ancestry that identified >240 loci influencing the risk of T2D[2]. Differences in allele frequency between ancestries affect the power to detect associations within a population, particularly among variants rare or monomorphic in one population but more frequent in another[3,4]. Although smaller than studies in European populations, a recent T2D meta-analysis in almost 200,000 Japanese individuals identified 28 additional loci[4]. The relative contributions of different pathways to the pathophysiology of T2D may also differ between ancestry groups. For example, in East Asian (EAS) populations, T2D prevalence is greater than in European populations among people of similar body mass index (BMI) or waist circumference[5]. We performed the largest meta-analysis of East Asian individuals to identify new genetic associations and provide insight into T2D pathogenesis.

RESULTS

We conducted a fixed-effect inverse-variance weighted GWAS meta-analysis combining 23 studies imputed to the 1000 Genomes Phase 3 reference panel from the Asian Genetic Epidemiology Network (AGEN) consortium (Supplementary Tables 1-3). We performed sex-combined T2D association without BMI adjustment in 77,418 T2D cases and 356,122 controls (effective sample size, N_{eff}=211,793) and with BMI adjustment in 54,481 T2D cases and 224,231 controls (N_{eff}= 135,780). In the set of studies with BMI-adjusted analyses, we also tested for T2D association in models stratified by sex (Supplementary Figure 1). We defined "lead" variants as the strongest T2D-associated variants with $P<5x10^{-8}$ and defined the region +/- 500 kb from the lead variant as a locus. A locus was considered novel if the lead variant was located at least 500 kb away from previously reported T2D-associated variants in any ancestry.

Using summary association statistics for ~11.7 million variants without adjustment for BMI (Supplementary Figure 1; Supplementary Tables 1-3), we identified lead variants at 178 loci to be associated with T2D, of which 49 were novel (Table 1; Supplementary Figure 2; Supplementary Table 4). Lead variants at all novel loci were common (MAF≥5%; Supplementary Figure 3), except for two low-frequency lead variants near *GDAP1* (MAF=2.4%), which regulates mitochondrial proteins and metabolic flux in skeletal muscle[6], and *PTF1A* (MAF=4.7%), which encodes a transcription factor required for pancreatic acinar cell development[7]. Lead variants met a stricter *P*-value threshold for significance based on Bonferroni correction for 11.7 million tests ($P<4.3x10^{-9}$) at 147 of the 178 loci, including 31 of the 49 novel loci.

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Using GCTA[8], we identified 298 distinct signals that met a locus-wide significance threshold of $P<1x10^{-5}$ (Supplementary Table 5), 204 of which were genome-wide significant ($P<5x10^{-8}$). Overall, we observed 2-4 signals at 50 loci and ≥5 signals at 11 loci. Among the 49 novel loci, 9 loci had two signals and the locus at *WNT7B* had three signals. Among the ten loci with the most significant meta-analysis *P*-values of association, seven contained ≥5 distinct signals (16 signals at *INS*/*IGF2*/*KCNQ1*; 7 signals at *CDKN2A/B* and *GRM8*/*PAX4*/*LEP*; 5 signals at *CDKAL1*, *HHEX*/*IDE, CDC123*/*CAMK1D,* and *TCF7L2;* Supplementary Figure 4; Supplementary Tables 5). The seven distinct association signals at the *GRM8*/*PAX4*/*LEP* locus span 1.4 Mb, and no evidence of T2D association at this locus has yet been reported in non-East Asian ancestry groups[2,9] (Supplementary Figure 4C). Joint analyses confirmed independent associations (LD r^2=0.0025) at two previously reported *PAX4* missense variants[10], rs2233580 [Arg192His: risk allele frequency (RAF)=8.6%, OR=1.31, 95% CI 1.27 – 1.34, $P_{GCTA}=3.0x10^{-89}$] and rs3824004 (Arg192Ser: RAF=3.4%, OR=1.23, 95% CI 1.19-1.28, $P_{GCTA}=4.3x10^{-29}$). The association signals at this locus also include variants near *LEP*, which encodes leptin, a hormone that regulates appetite[11]; increased leptin levels are associated with obesity and T2D, with greater increase in leptin levels per unit of BMI in Chinese individuals compared to those of African-American, European, and Hispanic ancestries[12].

At the previously reported *ANK1/NKX6-3* locus[2,13,14], we observed three distinct T2D association signals, two of which overlap and consist of variants spanning only ~25 kb (Figure 1). Given conflicting interpretation of candidate genes[2,15,16], we compared the T2D-association signals identified in East Asian individuals to eQTLs reported at this locus in islets[2,16-18], subcutaneous adipose[19], and skeletal muscle[15]. At the strongest signal, the lead T2D-associated variant, rs33981001, is in high LD with the lead *cis*-eQTL variant for *NKX6-3* in pancreatic islets (rs12549902; EAS LD r^2=0.79, EUR r^2=0.83)[16], and the T2D risk allele is associated with decreased expression of *NKX6-3* (β=-0.36, $P=6.1x10^{-7}$; Figure 1)[20]. *NKX6-3*, or NK6 homeobox 3, encodes a pancreatic islet transcription factor required for the development of alpha and β cells in the pancreas[21] and has been shown to influence insulin secretion[16]. At the second T2D-association signal, rs62508166 is in high LD with the lead *cis*-eQTL variant for *ANK1* in subcutaneous adipose tissue[19] and skeletal muscle[15] (rs516946; EAS LD r^2=0.96, EUR r^2=0.80), and the T2D risk allele is associated with increased expression of *ANK1* (subcutaneous adipose: β=0.20, $P=1.8x10^{-7}$; skeletal muscle: β=1.01, $P=2.8x10^{-22}$). *ANK1* belongs to the ankyrin family of integral membrane proteins that has been shown to affect glucose uptake in skeletal muscle, and changes in expression levels may lead to insulin resistance[22]. Together, these GWAS and eQTL signals suggest that variants within this ~25 kb region act to increase or decrease expression levels of two different genes in different tissues to increase T2D risk.

In T2D association analyses adjusted for BMI, we identified an additional ten loci, four of which were not reported previously for T2D, including loci near *MYOM3/SRSF10*, *TSN*, *GRB10*, and *NID2* (Supplementary Figure 5A; Supplementary Table 4). At the *NID2* locus, the T2D risk allele is associated with lower BMI in East Asian individuals, consistent with a lipodystrophy phenotype[23,24]. Among the combined 188 loci identified in models with and without adjustment for BMI, effect sizes were highly correlated (Pearson correlation r=0.99; Supplementary Figures 1 and 6). The locus with the strongest heterogeneity between the two models was *FTO* ($P_{het}=5.6x10^{-3}$), although even this locus failed to surpass a Bonferroni-corrected threshold for significant heterogeneity ($P_{het}<0.05/188=2.7x10^{-4}$).

In sex-stratified analyses of males (28,027 cases and 89,312 controls) and females (27,370 cases and 135,055 controls), we identified one additional novel male-specific locus near *IFT81* and two additional novel female-specific loci near *CPS1* and *LMTK2* (Supplementary Figure 5B and 5C; Supplementary Table 6). The lead *CPS1* variant rs1047891 (Thr1412Asn) has been reported previously to have a stronger effect in females than in males for cardiovascular disease and several blood metabolites[25]. Taken

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

together, we identified a total of 56 novel loci across BMI-unadjusted, BMI-adjusted, and sex-stratified models.

Among all T2D-associated loci, a region spanning almost 2 Mb on chromosome 12 near *ALDH2* exhibited the strongest differences between sexes (rs12231737, P_{het}=2.6x10^{-19}), with compelling evidence of association in males (P=5.5x10^{-27}) and no evidence for association in females (P=0.19) (Supplementary Figure 7; Supplementary Table 6). Further, joint conditional analyses revealed two conditionally distinct signals (rs12231737, P_{GCTA}=1.7x10^{-21}; rs557597782, P_{GCTA}=4.7x10^{-7}) in males only. *ALDH2* encodes aldehyde dehydrogenase 2 family member, a key enzyme in alcohol metabolism that converts acetaldehyde into acetic acid. This stretch of T2D associations in males reflects a long LD block that arose due to a recent selective sweep in East Asian individuals and results in flushing, nausea, and headache following alcohol consumption[26]. The most significantly associated missense variant in moderate LD with rs12231737 (r^2=0.68) was common functional variant rs671 (*ALDH2* Glu504Lys: RAF=77.4%, OR=1.16, 95% CI 1.15 – 1.18, P_{males}=4.2x10^{-24}), which leads to reduced ALDH2 activity and reduced alcohol metabolism, and have previously been reported to be associated with cardiometabolic traits in East Asian populations; the T2D risk allele is associated with better tolerance for alcohol and increased BMI, increased systolic and diastolic blood pressure, and increased triglycerides, but increased high-density lipoprotein, decreased low-density lipoprotein, and decreased cardiovascular risk[27-31]. The strong sexual dimorphism observed at this locus may be due to differences in alcohol consumption patterns between males and females[27,29] and/or differences in the effect of alcohol on insulin sensitivity[32].

With an effective sample size comparable to the largest study of T2D in European individuals (East Asian N_{eff}=211,793; European N_{eff} = 231,436)[2] and imputation to a dense 1000 Genomes reference panel, our results provide the most comprehensive and precise catalogue of East Asian T2D effects to date for comparisons across ancestries (Figure 2; Supplementary Table 7). For 178 EAS T2D loci and 231 EUR T2D loci (unadjusted for BMI) identified in a European meta-analysis[2], we compared the per-allele effect sizes for the 343 variants available in both datasets (i.e. polymorphic and passed quality control), including lead variants from both ancestries at shared signals. Overall, the per-allele effect sizes between the two ancestries were moderately correlated (r=0.54; Figure 2A). When the comparison was restricted to the 290 variants that are common (MAF≥5%) in both ancestries, the effect size correlation increased to r=0.71 (Figure 2B; Supplementary Figure 8). This effect size correlation further increased to r=0.88 for 116 variants significantly associated with T2D (P<5x10^{-8}) in both ancestries. While the overall effect sizes for all 343 variants appear, on average, to be stronger in East Asian individuals than European, this trend is reduced when each locus is represented only by the lead variant from one population (Supplementary Figure 9). Specifically, many variants identified with larger effect sizes in the European meta-analysis are missing from the comparison because they were rare/monomorphic or poorly imputed in the East Asian meta-analysis, for which imputation reference panels are less comprehensive compared to the European-centric Haplotype Reference Consortium panel.

Variants exhibiting the largest differences in effect sizes across ancestries are generally rare (MAF ≤0.1%) in European populations but common (e.g. *PAX4*, *RANBP3L*) or low frequency (e.g. *ZNF257*, *DGKD*) in East Asian populations. For example, rs142395395 near *ZNF257* (RAF=96.9%, OR=1.24, 95% CI 1.19-1.29, P=7.0x10^{-23}) has been reported only twice in 15,414 individuals of non-Finnish European ancestry from the gnomAD database[33]. This variant tags a previously described inversion of 415 kb observed only in East Asian individuals that disrupts the coding sequence and expression of *ZNF257,* as well as lymphoblastoid expression of 81 downstream genes and transcripts[34]. These data suggest that *ZNF257* and/or downstream target genes influence T2D susceptibility (Supplementary Figure 10).

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We identified many loci for which the lead variants exhibited similar allele frequencies and effect sizes in both the East Asian and European meta-analyses, but only reached genome-wide significance in the East Asian meta-analysis. Given shared susceptibility across ancestry groups, these loci may be detected in non-East Asian populations when sample sizes increase. Among these variants is rs117624659, located near *NKX6-1* (P_{EAS} = 2.0x10^{-16}, P_{EUR} =2.2x10^{-4}). This lead variant overlaps a highly conserved region that shows open chromatin specific to pancreatic islets. We conducted transcriptional reporter assays in MIN6 mouse insulinoma cells and observed that rs117624659 exhibited significant allelic differences in enhancer activity (Figure 3; Supplementary Figure 11). In the pancreas, NK6 homeobox 1 (NKX6.1) is required for the development of insulin-producing β cells and is a potent bifunctional transcriptional regulator[35]. Further, inactivation of *Nkx6.1* in mice demonstrated rapid-onset diabetes due to defects in insulin biosynthesis and secretion[36]. Unexpectedly, the T2D risk allele showed increased transcriptional activity, suggesting that the variant does not act in isolation or that *NXK6-1* is not the target gene.

At one of the novel T2D-associated loci near *SIX3*, the risk allele of East Asian lead variant rs12712928 (RAF=40.2%, OR=1.06, 95% CI 1.04 – 1.07, *P*=3.2x10^{-14}) is common across non-East Asian ancestries, ranging from 16.0% in Europeans to 26.4% in South Asians; however, there was no evidence of association in the other ancestry groups (meta-analysis: OR=0.98, 95% CI 0.96 – 0.99, *P*=2.9x10^{-3}) (Figure 4A; Supplementary Figure 12; Supplementary Table 8). Within the East Asian meta-analysis, the direction of effect is consistent across East Asian countries (Figure 4B) and within the contributing cohorts (Supplementary Figure 13). The T2D risk allele rs12712928-C is associated with higher fasting glucose levels in East Asian populations[37,38], has the strongest association with lower expression levels of both *SIX3* and *SIX2* in pancreatic islets[17], and demonstrated allele-specific binding to the transcription factor GABPA and significantly lower levels of transcriptional activity[38]. While larger studies in other ancestry groups could improve the accuracy of the effect estimate, current evidence suggests that the T2D association near *SIX3* is specific to East Asian populations.

To identify potential candidate genes underlying the T2D-association signals identified in East Asian individuals, we further characterized 88 loci, including known and novel loci, for which the lead variant at the primary East Asian association signal is located >500 kb from the lead variant of any primary European T2D association signal[2] (Supplementary Table 9). We characterized loci using prior trait associations, *cis*-regulatory effects on expression (colocalized eQTL), predicted effects on protein sequence, and a literature search (Supplementary Tables 10-13). Based on association results from cardiometabolic trait consortia[39], Biobank Japan[40], and the UK Biobank[41], the lead T2D-associated variant at 19 of the 88 loci was associated (*P*<5x10^{-8}) with at least one additional cardiometabolic trait, most frequently BMI or a fat mass-related trait (16 loci; Supplementary Tables 10 and 12). At 12 of the examined loci, T2D signals were colocalized with *cis*-eQTLs for transcripts in subcutaneous adipose tissue (n=5), skeletal muscle (n=3), pancreas (n=2), pancreatic islets (n=3), or whole blood (n=5; Supplementary Tables 11-12). At 19 loci, the lead T2D-associated variant or a variant in high LD with it (East Asian r^2>0.80) alter the protein sequence (Supplementary Tables 12). These variants affect mesenchymal stem cell differentiation and adipogenesis (*GIT2*, *STEAP2* and *JMJD1C*), muscle stem cell biology (*CALCR*), glucose metabolism (*PGM1* and *SCTR*), and insulin secretion (*FGFR4*; Supplementary Table 13). While mechanistic inference is required, these potential molecular mechanisms suggest new T2D susceptibility genes primarily detected by analyses in East Asians.

T2D loci were also identified at clusters of noncoding RNAs with roles in islet β cell function. One locus includes a set of microRNAs specifically expressed in islet β cells, the maternally expressed noncoding RNA *MEG3,* and the paternally expressed gene *DLK1*. Targets of the microRNAs at this locus increase β

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

cell apoptosis[42], and reduced *Meg3* impairs insulin secretion[43]. *DLK1* inhibits adipocyte differentiation, protecting from obesity[44], and promotes pancreatic ductal cell differentiation into β cells, increasing insulin secretion[45,46]. Other variants near *MEG3* have been associated with type 1 diabetes (EAS and EUR LD r^2=0 with EAS lead variant)[47]. The other noncoding RNA locus is the *MIR17HG* cluster of miRNAs that regulate glucose-stimulated insulin secretion and pancreatic β cell proliferation stress[48]; one of these microRNAs, miR-19a, affects hepatic gluconeogenesis[49]. Yet another independent T2D association locus is located near *TRAF3*, which is a direct target of the *MIR17HG* microRNA cluster and promotes hyperglycemia by increasing hepatic glucose production[50,51]. The T2D association results suggest that these noncoding RNAs influence disease susceptibility.

DISCUSSION

These T2D GWAS meta-analyses in the largest number of East Asian individuals analyzed to date identified 56 novel loci, providing additional insight into the biological basis of T2D. The results emphasize substantial shared T2D susceptibility with European individuals, as shown by the strong correlation of effect sizes among T2D-associated genetic variants with common allele frequencies in both East Asian and European ancestry populations. The results also detect novel associations in East Asian individuals, several of which are identified because they have higher allele frequencies in East Asian populations, exhibit larger effect sizes, and/or are influenced by other environmental risk factors or lifestyle behaviors such as alcohol consumption.

The identified loci point to multiple plausible molecular mechanisms and many new candidate genes linking T2D susceptibility to diverse biological processes. Annotation of loci identified in the East Asian meta-analysis suggests a substantial role for insulin resistance in T2D pathogenesis among East Asian individuals through skeletal muscle, adipose, and liver development and function. We also provide evidence that multiple distinct association signals in the same region do not necessarily act through the same gene. Conditionally distinct association signals in close proximity can affect different genes that may act in different tissues by different mechanisms, emphasizing the value of identifying functional variants that enable variant-to-gene links to be examined directly. Our results provide a foundation for future biological research in T2D pathogenesis and offer potential targets for mechanisms for interventions in disease risk.

METHODS

Ethics statement

All human research was approved by the relevant institutional review boards for each study at their respective sites and conducted according to the Declaration of Helsinki. All participants provided written informed consent.

Study cohorts and quality control

The East Asian type 2 diabetes (T2D) meta-analyses were performed with studies participating in the Asian Genetic Epidemiology Network (AGEN), a consortium of genetic epidemiology studies of T2D and related traits conducted in individuals of East Asian ancestry, and the Diabetes Meta-analysis of Trans-ethnic Association Studies (DIAMANTE), a consortium examining the genetic contribution to T2D across diverse ancestry populations including African-American, East Asian, European, Hispanic, and South Asian. The East Asian meta-analysis included 77,418 T2D cases and 356,122 controls from 23 GWAS, including three biobanks, CKB, KBA[52,53], and BBJ[4] [effective sample size (N_{eff}) = 211,793; Supplementary Figure 1]. A subset of studies was analyzed in BMI-adjusted and sex-specific models (54,481 cases, 224,231 cases; N_{eff} = 135,780). For each study, T2D case control ascertainment is described in Supplementary Table 1 and summary statistics are provided in Supplementary Table 2. Included studies

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

were genotyped on either commercially available or customized Affymetrix or Illumina genome-wide genotyping arrays. Array quality control criteria implemented within each study, including variant call rate and Hardy-Weinberg equilibrium, are summarized in Supplementary Table 3. To harmonize study-level genotype scaffold for imputation to 1000 Genomes (1000G) reference panels, each study adopted a uniform protocol for pre-imputation quality checks. Each study applied the protocol to exclude variants with: i) mismatched chromosomal positions or alleles not present in the reference panel; ii) ambiguous alleles (AT/CG) with minor allele frequency (MAF) >40% in the reference panel; or iii) absolute allele frequency differences > 20% compared to East Asian-specific allele frequencies. The genotype scaffold for each study was then imputed to the 1000G Phase 1 or 3 reference panel[54] using minimac3[55] or IMPUTEv2[56]. In BMI-unadjusted analyses, all studies were imputed to 1000G Phase 3. In BMI-adjusted and sex-stratified analyses, all studies were imputed to 1000G Phase 3 except for Biobank Japan[14], which was imputed to the 1000G Phase 1 reference panel.

Study-level association analyses

Within each study, all variants were tested for association with T2D assuming an additive model of inheritance within a regression framework, including age, sex, and other study-specific covariates (Supplementary Table 3). To account for population structure and relatedness, association analyses were either performed using FIRTH[57] or mach2dat with additional adjustment for principal components in unrelated individuals or a linear mixed model with kinship matrix implemented in BOLT-LMM[58]. In studies analyzed with the linear mixed model, allelic effects and standard errors were converted to the log-odds scale that accounts for case-control imbalance[59]. Within each study, variants were removed if the: i) imputation quality score was poor (minimac3 r^2<0.30; IMPUTE2 info score <0.40); ii) combined case control minor allele count <5; or iii) standard error of the log-OR>10. For a subset of the studies, BMI was added as an additional covariate, and association analyses were also performed separately in males and females. For each study and model, association statistics were corrected with genomic control inflation factor[60] calculated from common variants (MAF≥5%) (Supplementary Table 3). For BBJ, we applied the genomic control inflation factor 1.21 as reported[4].

Sex-combined meta-analysis

We combined study-level association statistics using fixed effects meta-analysis with inverse-variance weighting of log-ORs implemented in METAL[61]. Variants with allele frequency differences >20% between 1000 Genomes Phase 1 and 3 panels were excluded from the meta-analysis. To assess excess inflation arising from cryptic relatedness and population structure, we applied LD score regression to the meta-analysis summary statistics to estimate residual inflation of summary statistics, using a set of 1,889 unrelated Chinese individuals from the Singapore Chinese Eye Study[62]. The LD score regression intercepts were 0.991 for BMI-unadjusted, and 1.0148 for BMI-adjusted models. As the LD score regression intercepts indicated absence of excess inflation, the meta-analysis results were corrected for inflation using these LD score regression intercepts. For subsequent analyses, we considered only variants that were present in at least 50% of the effective sample size N_{eff} [computed as $4/(1/N_{cases} + 1/N_{controls})$][61]. Heterogeneity in allelic effect sizes between studies were assessed with fixed effects inverse variance weighted meta-analysis P_{het}. We further compared the genetic effects from BMI-unadjusted and BMI-adjusted models using fixed effects inverse variance weighted meta-analysis P_{het}. Loci were defined as novel if the lead variant is: (1) at least 500 kb away and confirmed by GCTA to be conditionally independent from previously reported T2D-associated variants in any ancestry, and (2) assessed using LocusZoom plots and detailed literature review to be away from known loci with extended LD.

Sex-differentiated meta-analysis

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The meta-analyses described above were repeated for males and females separately. The male-specific meta-analyses included up to 28,027 cases and 89,312 controls (N_{eff} = 65,660) and the female-specific analyses included up to 27,370 cases and 135,055 controls (N_{eff} = 70,332). LD score regression intercepts were 1.0035 for BMI-unadjusted and 1.0034 for BMI-adjusted models in males and 1.0035 for BMI-unadjusted and 1.0034 for BMI-adjusted models in females. We further performed a test for heterogeneity in allelic effects between males and females as implemented in GWAMA[63,64].

Detection of distinct association signals

To detect multiple distinct association signals at each associated locus, we combined overlapping loci when the distance between any pair of lead variants was <1 Mb. We then performed approximate conditional analyses using GCTA[8] with genome-wide meta-analysis summary statistics and LD estimated from 78,000 samples from the Korean Biobank Array[53].

Comparing loci effects between East Asian and European populations

We compared the per-allele effect sizes of lead variants identified from the East Asian BMI-unadjusted sex-combined meta-analysis (178 lead variants) and European BMI-unadjusted sex-combined meta-analysis[2] (231 lead variants; Supplementary Table 7). Across the 409 associated variants from the two ancestries, 11 lead variants overlapped, resulting in 398 unique variants. As the variants in the European analysis were imputed using the Haplotype Reference Consortium reference panel and did not include indel variants, a variant in strong LD (East Asian r^2>0.90) with the lead East Asian variant was used when the lead variant was an indel, when possible. If the lead East Asian variant or a variant in strong LD (East Asian r^2>0.90) was not available in the European data from DIAMANTE, we used results from a previous European type 2 diabetes meta-analysis[65]. The effect size comparison plot was restricted to 343 variants where data was available for both ancestries (Figure 2A). For loci that were significant in both the East Asian and European meta-analyses, if the lead variants were different, both lead variants were plotted (see Supplementary Table 7). Effect size plots were further restricted to: i) 290 lead variants with MAF≥5% in both East Asian and European meta-analyses (Supplementary Figure 7); ii) 162 lead variants significant in the East Asian meta-analysis (Supplementary Figure 8A); and iii) 192 lead variants significant in the European meta-analysis (Supplementary Figure 8B).

Associations with other metabolic traits and outcomes

We used the Type 2 Diabetes Knowledge Portal[39] to explore associations of the newly identified loci with other metabolic traits and outcomes. Association statistics from the following consortia were available for query on the portal (last accessed March 18, 2019): coronary artery disease from CARDIoGRAM[66], BMI and waist-hip-ratio from GIANT[67,68], lipid traits from GLGC[69], and glycemic traits from MAGIC[70,71]. Additionally, we used available data from AGEN East Asian meta-analyses for lipids[72] and adiponectin[73], along with the phenotypic data from the UK Biobank[74]. Effect sizes were obtained from publicly available summary statistic files.

Colocalization with expression quantitative trait loci (eQTL)

We searched publicly available eQTL databases such as GTEx[75] and the Parker lab Islet Browser[17], to identify *cis*-eQTLs at the novel loci in adipose (subcutaneous and visceral), blood, pancreas, pancreatic islet, and skeletal muscle tissue. We also searched for *cis*-eQTLs in subcutaneous adipose tissue data from the METSIM study[19]. Colocalized eQTLs were identified if the lead expression level-associated variant and the GWAS lead variant were in high LD (r^2>0.80) in Europeans to accommodate the predominantly European eQTL data. Reciprocal conditional analyses were also performed using the METSIM data to determine if the GWAS lead variant and the lead eSNP were part of the same eQTL signal.

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Literature review

We conducted a traditional literature review to identify candidate genes at each novel locus using NCBI Entrez Gene, PubMed and OMIM. We included gene symbols and the following keywords as search terms in PubMed: diabetes, glucose, insulin, islet, adipose, muscle, liver, obesity. A gene was considered a potential candidate if an apparent link to T2D biology existed based on prior studies of gene function.

Functional annotation and experimentation at *NKX6-1*

We used ENCODE[76], ChromHMM[77], and Human Epigenome Atlas[78] data available through the UCSC Genome Browser to identify candidate variants at the association signal near *NKX6-1* that overlapped open-chromatin peaks, ChromHMM chromatin states, and chromatin-immunoprecipitation sequencing (ChIP-seq) peaks of histone modifications H4K4me1, H3K4me3, and H3K27ac, and transcription factors in the pancreas and pancreatic islets. MIN6 mouse insulinoma cells[79] and 823/13 rat insulinoma cells[80] were cultured in DMEM (Sigma) supplemented with 10% FBS, 1mM sodium pyruvate, and 0.1 mM beta-mercaptoethanol. The cell cultures were maintained at 37° C with 5% CO_2. To measure variant allelic differences in enhancer activity at the *NKX6*-1 locus, we designed oligonucleotide primers (forward: CCCTAGTAATGCCCTTTTTGTT; reverse: TCAGCCTGAGAAGTCTGTGA) with KpnI and XhoI restriction sites, and amplified the 400-bp DNA region (GRCh37/hg19 -chr4: 85,339,430-85,339,829) around rs117624659. As previously described[80], we ligated amplified DNA from individuals homozygous for each allele into the multiple cloning site of the pGL4.23 (Promega) minimal promoter luciferase reporter vector in both the forward and reverse orientations with respect to the genome. Clones were isolated and sequenced for genotype and fidelity. $2.1x10^5$ MIN6 or $3.0x10^5$ 823/13 cells were seeded per well and grown to 90% confluence in 24-well plates. We co-transfected five independent luciferase constructs and *Renilla* control reporter vector (phRL-TK, Promega) using Lipofectamine 2000 (Life Technologies) and incubated. 48-hours post-transfection, the cells were lysed with Passive Lysis Buffer (Promega). Luciferase activity was measured using the Dual-luciferase Reporter Assay System (Promega) per manufacturer instructions and as previously described[81].

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

REFERENCES

1 Stumvoll, M., Goldstein, B. J. & van Haeften, T. W. Type 2 diabetes: principles of pathogenesis and therapy. *Lancet (London, England)* **365**, 1333-1346, doi:10.1016/s0140-6736(05)61032-x (2005).

2 Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* **50**, 1505-1513, doi:10.1038/s41588-018-0241-6 (2018).

3 Cho, Y. S. *et al.* Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat Genet* **44**, 67-72, doi:10.1038/ng.1019 (2011).

4 Suzuki, K. *et al.* Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. *Nat Genet* **51**, 379-386, doi:10.1038/s41588-018-0332-4 (2019).

5 Huxley, R. *et al.* Ethnic comparisons of the cross-sectional relationships between measures of body size with diabetes and hypertension. *Obesity reviews : an official journal of the International Association for the Study of Obesity* **9 Suppl 1**, 53-61, doi:10.1111/j.1467-789X.2007.00439.x (2008).

6 Lassiter, D. G., Sjogren, R. J. O., Gabriel, B. M., Krook, A. & Zierath, J. R. AMPK activation negatively regulates GDAP1, which influences metabolic processes and circadian gene expression in skeletal muscle. *Mol Metab* **16**, 12-23, doi:10.1016/j.molmet.2018.07.004 (2018).

7 Hoang, C. Q. *et al.* Transcriptional Maintenance of Pancreatic Acinar Identity, Differentiation, and Homeostasis by PTF1A. *Molecular and cellular biology* **36**, 3033-3047, doi:10.1128/MCB.00358-16 (2016).

8 Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-375, S361-363, doi:10.1038/ng.2213 (2012).

9 Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41-47, doi:10.1038/nature18642 (2016).

10 Kwak, S. H. *et al.* Nonsynonymous Variants in PAX4 and GLP1R Are Associated With Type 2 Diabetes in an East Asian Population. *Diabetes* **67**, 1892-1902, doi:10.2337/db18-0361 (2018).

11 Klok, M. D., Jakobsdottir, S. & Drent, M. L. The role of leptin and ghrelin in the regulation of food intake and body weight in humans: a review. *Obesity reviews : an official journal of the International Association for the Study of Obesity* **8**, 21-34, doi:10.1111/j.1467-789X.2006.00270.x (2007).

12 Rasmussen-Torvik, L. 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* **22**, 705-709, doi:10.1016/j.annepidem.2012.07.011 (2012).

13 Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature genetics* **44**, 981-990, doi:10.1038/ng.2383 (2012).

14 Imamura, M. *et al.* Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes. *Nat Commun* **7**, 10531, doi:10.1038/ncomms10531 (2016).

15 Scott, L. J. *et al.* The genetic regulatory signature of type 2 diabetes in human skeletal muscle. *Nat Commun* **7**, 11764, doi:10.1038/ncomms11764 (2016).

16 van de Bunt, M. *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* **11**, e1005694, doi:10.1371/journal.pgen.1005694 (2015).

17 Varshney, A. *et al.* Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc Natl Acad Sci U S A* **114**, 2301-2306, doi:10.1073/pnas.1621192114 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

18 Thurner, M. *et al.* Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci. *eLife* **7**, doi:10.7554/eLife.31977 (2018).
19 Civelek, M. *et al.* Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic Traits. *Am J Hum Genet* **100**, 428-443, doi:10.1016/j.ajhg.2017.01.027 (2017).
20 van de Bunt, M. *et al.* The miRNA profile of human pancreatic islets and beta-cells and relationship to type 2 diabetes pathogenesis. *PLoS One* **8**, e55272, doi:10.1371/journal.pone.0055272 (2013).
21 Henseleit, K. D. *et al.* NKX6 transcription factor activity is required for alpha- and beta-cell development in the pancreas. *Development* **132**, 3139-3149, doi:10.1242/dev.01875 (2005).
22 Yan, R. *et al.* A novel type 2 diabetes risk allele increases the promoter activity of the muscle-specific small ankyrin 1 gene. *Sci Rep* **6**, 25105, doi:10.1038/srep25105 (2016).
23 Wen, W. *et al.* Genome-wide association studies in East Asians identify new loci for waist-hip ratio and waist circumference. *Sci Rep* **6**, 17958, doi:10.1038/srep17958 (2016).
24 Akiyama, M. *et al.* Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat Genet* **49**, 1458-1467, doi:10.1038/ng.3951 (2017).
25 Hartiala, J. A. *et al.* Genome-wide association study and targeted metabolomics identifies sex-specific association of CPS1 with coronary artery disease. *Nat Commun* **7**, 10558, doi:10.1038/ncomms10558 (2016).
26 Okada, Y. *et al.* Deep whole-genome sequencing reveals recent selection signatures linked to evolution and disease risk of Japanese. *Nat Commun* **9**, 1631, doi:10.1038/s41467-018-03274-0 (2018).
27 Xu, F. *et al.* ALDH2 genetic polymorphism and the risk of type II diabetes mellitus in CAD patients. *Hypertens Res* **33**, 49-55, doi:10.1038/hr.2009.178 (2010).
28 Kato, N. *et al.* Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat Genet* **43**, 531-538, doi:10.1038/ng.834 (2011).
29 Takeuchi, F. *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* **75**, 911-918 (2011).
30 Kim, Y. K. *et al.* Evaluation of pleiotropic effects among common genetic loci identified for cardio-metabolic traits in a Korean population. *Cardiovascular diabetology* **15**, 20, doi:10.1186/s12933-016-0337-1 (2016).
31 Ma, C. *et al.* Associations between aldehyde dehydrogenase 2 (ALDH2) rs671 genetic polymorphisms, lifestyles and hypertension risk in Chinese Han people. *Sci Rep* **7**, 11136, doi:10.1038/s41598-017-11071-w (2017).
32 Schrieks, I. C., Heil, A. L., Hendriks, H. F., Mukamal, K. J. & Beulens, J. W. The effect of alcohol consumption on insulin sensitivity and glycemic status: a systematic review and meta-analysis of intervention studies. *Diabetes Care* **38**, 723-732, doi:10.2337/dc14-1556 (2015).
33 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291, doi:10.1038/nature19057 (2016).
34 Puig, M. *et al.* Functional Impact and Evolution of a Novel Human Polymorphic Inversion That Disrupts a Gene and Creates a Fusion Transcript. *PLoS Genet* **11**, e1005495, doi:10.1371/journal.pgen.1005495 (2015).
35 Iype, T. *et al.* The transcriptional repressor Nkx6.1 also functions as a deoxyribonucleic acid context-dependent transcriptional activator during pancreatic beta-cell differentiation: evidence for feedback activation of the nkx6.1 gene by Nkx6.1. *Molecular endocrinology (Baltimore, Md.)* **18**, 1363-1375, doi:10.1210/me.2004-0006 (2004).

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

36 Taylor, B. L., Liu, F. F. & Sander, M. Nkx6.1 is essential for maintaining the functional state of pancreatic beta cells. *Cell Rep* **4**, 1262-1275, doi:10.1016/j.celrep.2013.08.010 (2013).

37 Kim, Y. J. *et al.* Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. *Nat Genet* **43**, 990-995, doi:10.1038/ng.939 (2011).

38 Spracklen, C. N. *et al.* Identification and functional analysis of glycemic trait loci in the China Health and Nutrition Survey. *PLoS Genet* **14**, e1007275, doi:10.1371/journal.pgen.1007275 (2018).

39 *Type 2 Diabetes Knowledge Portal*, <http://www.type2diabetesgenetics.org/home/portalHome> (2019).

40 Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* **50**, 390-400, doi:10.1038/s41588-018-0047-6 (2018).

41 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209, doi:10.1038/s41586-018-0579-z (2018).

42 Kameswaran, V. *et al.* Epigenetic regulation of the DLK1-MEG3 microRNA cluster in human type 2 diabetic islets. *Cell metabolism* **19**, 135-145, doi:10.1016/j.cmet.2013.11.016 (2014).

43 You, L. *et al.* Downregulation of Long Noncoding RNA Meg3 Affects Insulin Synthesis and Secretion in Mouse Pancreatic Beta Cells. *J Cell Physiol* **231**, 852-862, doi:10.1002/jcp.25175 (2016).

44 Moon, Y. S. *et al.* Mice lacking paternally expressed Pref-1/Dlk1 display growth retardation and accelerated adiposity. *Molecular and cellular biology* **22**, 5585-5592, doi:10.1128/mcb.22.15.5585-5592.2002 (2002).

45 Wang, Y. *et al.* Overexpression of Pref-1 in pancreatic islet beta-cells in mice causes hyperinsulinemia with increased islet mass and insulin secretion. *Biochem Biophys Res Commun* **461**, 630-635, doi:10.1016/j.bbrc.2015.04.078 (2015).

46 Rhee, M. *et al.* Preadipocyte factor 1 induces pancreatic ductal cell differentiation into insulin-producing cells. *Sci Rep* **6**, 23960, doi:10.1038/srep23960 (2016).

47 Onengut-Gumuscu, S. *et al.* Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet* **47**, 381-386, doi:10.1038/ng.3245 (2015).

48 Chen, Y. *et al.* MicroRNA-17-92 cluster regulates pancreatic beta-cell proliferation and adaptation. *Mol Cell Endocrinol* **437**, 213-223, doi:10.1016/j.mce.2016.08.037 (2016).

49 Dou, L. *et al.* MiR-19a mediates gluconeogenesis by targeting PTEN in hepatocytes. *Mol Med Rep* **17**, 3967-3971, doi:10.3892/mmr.2017.8312 (2018).

50 Chen, Z. *et al.* Hepatocyte TRAF3 promotes insulin resistance and type 2 diabetes in mice with obesity. *Molecular metabolism* **4**, 951-960, doi:10.1016/j.molmet.2015.09.013 (2015).

51 Liu, F., Cheng, L., Xu, J., Guo, F. & Chen, W. miR-17-92 functions as an oncogene and modulates NF-kappaB signaling by targeting TRAF3 in MGC-803 human gastric cancer cells. *Int J Oncol* **53**, 2241-2257, doi:10.3892/ijo.2018.4543 (2018).

52 Kim, Y., Han, B. G. & Ko, G. E. S. g. Cohort Profile: The Korean Genome and Epidemiology Study (KoGES) Consortium. *Int J Epidemiol* **46**, e20, doi:10.1093/ije/dyv316 (2017).

53 Moon, S. *et al.* The Korea Biobank Array: Design and Identification of Coding Variants Associated with Blood Biochemical Traits. *Sci Rep* **9**, 1382, doi:10.1038/s41598-018-37832-9 (2019).

54 The 1000 Genomes Project Consortium. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).

55 Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287, doi:10.1038/ng.3656 (2016).

56 Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3 (Bethesda)* **1**, 457-470, doi:10.1534/g3.111.001198 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

57 Ma, C., Blackwell, T., Boehnke, M. & Scott, L. J. Recommended joint and meta-analysis strategies for case-control association testing of single low-count variants. *Genet Epidemiol* **37**, 539-550, doi:10.1002/gepi.21742 (2013).

58 Loh, P. R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-290, doi:10.1038/ng.3190 (2015).

59 Cook, J. P., Mahajan, A. & Morris, A. P. Guidance for the utility of linear models in meta-analysis of genetic association studies of binary phenotypes. *Eur J Hum Genet* **25**, 240-245, doi:10.1038/ejhg.2016.150 (2017).

60 Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).

61 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191, doi:10.1093/bioinformatics/btq340 (2010).

62 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295, doi:10.1038/ng.3211 (2015).

63 Magi, R., Lindgren, C. M. & Morris, A. P. Meta-analysis of sex-specific genome-wide association studies. *Genet Epidemiol* **34**, 846-853, doi:10.1002/gepi.20540 (2010).

64 Magi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288, doi:10.1186/1471-2105-11-288 (2010).

65 Scott, R. A. *et al.* An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes* **66**, 2888-2902, doi:10.2337/db16-1253 (2017).

66 Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* **43**, 333-338, doi:10.1038/ng.784 (2011).

67 Turcot, V. *et al.* Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat Genet* **50**, 26-41, doi:10.1038/s41588-017-0011-x (2018).

68 Justice, A. E. *et al.* Protein-coding variants implicate novel genes related to lipid homeostasis contributing to body-fat distribution. *Nat Genet* **51**, 452-469, doi:10.1038/s41588-018-0334-2 (2019).

69 Willer, C. J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274-1283, doi:10.1038/ng.2797 (2013).

70 Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* **42**, 105-116, doi:10.1038/ng.520 (2010).

71 Soranzo, N. *et al.* Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemic and nonglycemic pathways. *Diabetes* **59**, 3229-3239, doi:10.2337/db10-0502 (2010).

72 Spracklen, C. N. *et al.* Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. *Hum Mol Genet* **27**, 1122, doi:10.1093/hmg/ddx439 (2018).

73 Wu, Y. *et al.* A meta-analysis of genome-wide association studies for adiponectin levels in East Asians identifies a novel locus near WDR11-FGFR2. *Hum Mol Genet* **23**, 1108-1119, doi:10.1093/hmg/ddt488 (2014).

74 Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine* **12**, e1001779, doi:10.1371/journal.pmed.1001779 (2015).

75 Gamazon, E. R. *et al.* Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat Genet* **50**, 956-967, doi:10.1038/s41588-018-0154-4 (2018).

76 An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

77 Ezzat, S. *et al.* The cancer-associated FGFR4-G388R polymorphism enhances pancreatic insulin secretion and modifies the risk of diabetes. *Cell Metab* **17**, 929-940, doi:10.1016/j.cmet.2013.05.002 (2013).

78 Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).

79 Miyazaki, J. *et al.* Establishment of a pancreatic beta cell line that retains glucose-inducible insulin secretion: special reference to expression of glucose transporter isoforms. *Endocrinology* **127**, 126-132, doi:10.1210/endo-127-1-126 (1990).

80 Hohmeier, H. E. *et al.* Isolation of INS-1-derived cell lines with robust ATP-sensitive K+ channel-dependent and -independent glucose-stimulated insulin secretion. *Diabetes* **49**, 424-430 (2000).

81 Fogarty, M. P., Cannon, M. E., Vadlamudi, S., Gaulton, K. J. & Mohlke, K. L. Identification of a regulatory variant that binds FOXA1 and FOXA2 at the CDC123/CAMK1D type 2 diabetes GWAS locus. *PLoS Genet* **10**, e1004633, doi:10.1371/journal.pgen.1004633 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ACKNOWLEDGEMENTS

This work was supported by subawards (to X.S and Y.S.C.) from NIDDK U01DK105554 (Jose C. Florez). The authors thank all investigators, staff members and study participants for their contributions to all participating studies. A full list of funding, and individual and study acknowledgements are available in Supplementary Materials.

AUTHOR CONTRIBUTIONS

Project coordination: K.L.M, X.S. Writing: C.N.S., E.S.T, M.B., M.H., Y.J.K., K.L., K.L.M, X.S. Core analyses: C.N.S., M.H., Y.J.K., K.L., A.K.I., H.J.P., S.M.B., X.S. eQTL lookups: M.vdB, A.L.G. DIAMANTE analysis group: J.E.B., M.B., D.W.B., J.C.C., A.M., M.I.M., M.C.Y.N., L.E.P., W.Zhang., A.P.M. Statistical analysis in individual studies: Y.T., M.H., K.S., X.S., F.T., M.N., C.N.S., K.L., F.B., Y.J.K., S.Moon., C.H.T.T., J.Y., X.G., J.Long., J.F.C., V.J.Y.L., S.H.K., H.S.C., C.H.C. Individual study design and principal investigators: M.Igase., T.Kadowaki., Y.X.W., N.K., M.Y., K.L.M., R.G.W., E.S.T., B.J.K., R.C.W.M., J.I.R., F.M., X.O.S., C.Y.C., W.P.K., T.Y.W., K.S.P., Y.S.C., W.H.H.S., J.Y.W. Genotyping and phenotyping: K.K., A.T., Y.K., T.Y., Y.O., J.B.J., T.Katsuya., M.Isono., S.I., K.Yamamoto., A.H., S.D., W.H., J.S., P.G.L., C.Y., Y.G., Z.B., J.Lv., L.L., Z.C., N.R.L., L.S.A., J.Liu., R.M.vD., S.H., K.Yoon., H.M.J., D.M.S., G.J., A.O.L., B.T., W.Y.S., J.C.N.C., M.Y.H., Y.D.I.C., T.Kawaguchi., Y.B.X., W.Zheng., L.Z., C.C.K., M.A.P., M.G., J.M.Y., C.S., M.L.C., M.S.L., C.M.H., L.M.C., Y.J.H., L.C.C., Y.T.C., F.J.T., J.I.R.

AUTHOR INFORMATION

Summary-level statistics will be available at the AGEN consortium website https://blog.nus.edu.sg/agen/summary-statistics/ and the Accelerating Medicines Partnership T2D portal http://www.type2diabetesgenetics.org/.

The authors declare no competing interest.

Correspondence and requests for materials should be addressed to mohlke@med.unc.edu and/or ephsx@nus.edu.sg.

WEB RESOURCES

Pre-imputation preparation and quality control, http://www.well.ox.ac.uk/~wrayner/tools/
Michigan Imputation Server, https://imputationserver.sph.umich.edu/index.html
EPACTS, https://github.com/statgen/EPACTS
RVTESTS, https://github.com/zhanxw/rvtests
METAL, http://csg.sph.umich.edu/abecasis/metal/
LDSC, https://github.com/bulik/ldsc
GWAMA, https://www.geenivaramu.ee/en/tools/gwama
Type 2 Diabetes Knowledge Portal, http://www.type2diabetesgenetics.org
GTEx Portal, https://gtexportal.org/home/
Parker lab Islet Browser, http://theparkerlab.org/tools/isleteqtl/

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

814 **Table 1. Novel lead variants associated with type 2 diabetes in East Asians**

Locus	Lead variant	Chr	Pos	RA	NRA	RAF	Neff	Cases (N)	Controls (N)	OR (95% CI)	*P*
VWA5B1	rs60573766	1	20,688,352	C	T	0.635	210,985	77,181	354,521	1.04 (1.03-1.06)	4.30E-10
MAST2	rs562138031	1	46,244,900	C	CT	0.726	210,231	76,984	350,185	1.06 (1.04-1.07)	8.66E-12
PGM1	rs2269239	1	64,109,359	C	G	0.813	211,039	77,221	351,786	1.06 (1.04-1.07)	5.81E-10
TSEN15	rs1327123	1	184,014,593	C	G	0.460	210,985	77,181	354,521	1.04 (1.03-1.05)	6.40E-09
MDM4	rs201297151	1	204,474,581	CAAAAAAAA	C	0.440	210,231	76,984	350,185	1.04 (1.03-1.05)	4.64E-08
SIX3	rs12712928	2	45,192,080	C	G	0.402	211,321	77,275	355,448	1.06 (1.04-1.07)	3.16E-14
IKZF2	rs75179644	2	213,687,103	T	C	0.899	210,985	77,181	354,521	1.08 (1.05-1.10)	6.01E-10
ZBTB20	rs6768463	3	114,964,264	T	A	0.614	211,793	77,418	356,122	1.05 (1.03-1.06)	3.19E-11
TFRC	rs9866168	3	195,830,310	T	A	0.640	210,985	77,181	354,521	1.05 (1.03-1.06)	1.31E-09
RANBP3L	rs16902871	5	36,257,018	G	A	0.149	211,793	77,418	356,122	1.06 (1.04-1.08)	3.34E-09
PCSK1	rs11960326	5	95,850,807	C	T	0.416	211,680	77,388	355,625	1.04 (1.03-1.05)	3.74E-09
REPS1	rs556058292	6	139,259,142	C	CT	0.651	210,985	77,181	354,521	1.04 (1.03-1.06)	1.42E-08
HIVEP2	rs9390022	6	143,056,556	T	C	0.800	211,793	77,418	356,122	1.05 (1.03-1.07)	6.35E-09
ZNF713	rs565050730	7	55,984,953	GA	G	0.334	210,985	77,181	354,521	1.04 (1.03-1.06)	4.78E-08
STEAP1	rs62469016	7	89,752,238	C	G	0.223	211,793	77,418	356,122	1.07 (1.05-1.08)	1.52E-15
CALCR	rs2074120	7	93,107,093	A	C	0.323	211,793	77,418	356,122	1.04 (1.03-1.06)	8.38E-09
GRM8/PAX4	rs117737118	7	126,526,991	G	A	0.093	209,652	76,825	348,561	1.18 (1.15-1.21)	1.01E-31
ASAH1	rs34642578	8	17,927,609	T	C	0.053	210,985	77,181	354,521	1.09 (1.06-1.13)	1.67E-09
ZNF703	rs12680217	8	37,397,803	T	C	0.553	211,457	77,324	355,195	1.05 (1.03-1.06)	1.64E-11
GDAP1	rs149265787	8	75,214,398	G	A	0.024	211,694	77,392	355,608	1.14 (1.10-1.19)	5.68E-10
TRIB1	rs10103720	8	126,471,770	C	T	0.378	211,793	77,418	356,122	1.04 (1.03-1.06)	3.33E-09
EFR3A	rs73708055	8	132,879,882	G	A	0.252	211,793	77,418	356,122	1.04 (1.03-1.06)	4.41E-08
DMRT2	rs1016565	9	1,032,567	A	G	0.421	211,793	77,418	356,122	1.04 (1.02-1.05)	2.18E-08
PTCH1	rs113154802	9	98,278,413	C	T	0.888	210,985	77,181	354,521	1.06 (1.04-1.08)	4.91E-08
ABCA1	rs201375651	9	107,597,527	CA	C	0.395	211,793	77,418	356,122	1.04 (1.03-1.06)	2.56E-08
PTF1A	rs77065181	10	23,487,778	A	G	0.047	211,098	77,211	355,018	1.09 (1.06-1.13)	1.63E-08
ARID5B	rs141583966	10	63,712,602	G	GGTGT	0.909	210,066	76,938	349,783	1.08 (1.05-1.11)	6.58E-10
JMJD1C	rs148928116	10	64,976,133	T	TA	0.795	210,985	77,181	354,521	1.06 (1.05-1.08)	2.31E-13
ARHGAP19	rs10736116	10	99,056,921	C	G	0.306	211,793	77,418	356,122	1.05 (1.03-1.06)	9.21E-11
BBIP1	rs7895872	10	112,678,657	T	G	0.579	210,985	77,181	354,521	1.05 (1.03-1.06)	2.29E-11
BDNF	rs988748	11	27,724,745	G	C	0.565	211,793	77,418	356,122	1.04 (1.03-1.06)	1.62E-10
FAIM2	rs77978149	12	50,269,863	T	C	0.090	210,406	77,022	352,897	1.08 (1.05-1.10)	3.89E-09
ALDH2	rs149212747	12	111,836,771	A	AC	0.795	210,231	76,984	350,185	1.07 (1.05-1.09)	2.07E-11
RBM19	rs7307263	12	114,123,722	G	C	0.427	210,985	77,181	354,521	1.04 (1.02-1.05)	3.96E-08
FGF9	rs9316706	13	22,589,883	A	G	0.351	211,793	77,418	356,122	1.04 (1.03-1.06)	3.33E-09
NYNRIN	rs12437434	14	24,878,370	C	T	0.713	211,039	77,221	351,786	1.05 (1.03-1.06)	1.02E-09
LRRC74A	rs140431144	14	77,382,093	TA	T	0.327	207,126	75,090	353,783	1.05 (1.03-1.06)	9.21E-11
DLK1/MEG3/ miRNA cluster	rs73347525	14	101,255,172	A	G	0.756	205,739	74,694	350,558	1.06 (1.04-1.08)	5.54E-11
TRAF3	rs55700915	14	103,237,952	A	G	0.434	211,039	77,221	351,786	1.04 (1.03-1.06)	1.50E-08
HERC2	rs76704029	15	28,546,173	T	C	0.732	173,598	66,475	264,668	1.06 (1.04-1.08)	3.03E-08
MYO5C	rs149336329	15	52,587,740	G	T	0.949	210,231	76,984	350,185	1.11 (1.07-1.14)	1.31E-09
RGMA	rs61021634	15	93,825,384	A	G	0.438	210,985	77,181	354,521	1.05 (1.03-1.06)	9.47E-12
IGF1R	rs79826452	15	99,366,409	A	G	0.890	210,985	77,181	354,521	1.07 (1.04-1.10)	3.80E-08
PKD1L3	rs12600132	16	72,022,534	T	C	0.432	211,793	77,418	356,122	1.04 (1.03-1.05)	5.95E-09
ZFHX3	rs6416749	16	73,100,308	C	T	0.375	210,460	77,062	350,162	1.05 (1.04-1.07)	3.40E-12
SUMO2	rs35559984	17	73,187,031	CA	C	0.652	206,547	74,931	352,159	1.05 (1.03-1.07)	7.88E-09
ZNF799	rs4604181	19	12,509,536	A	C	0.514	209,652	76,825	348,561	1.04 (1.03-1.06)	1.80E-08
ZNRF3	rs147413364	22	29,380,119	T	TA	0.357	210,874	77,175	351,384	1.04 (1.03-1.06)	3.38E-08
WNT7B	rs28637892	22	46,313,618	T	G	0.215	175,881	63,772	319,376	1.05 (1.04-1.07)	3.66E-09

Single variant association results from East Asian fixed-effect inverse-variance meta-analysis (BMI-unadjusted sex-combined model adjusted for age, sex, and study-specific covariates) using METAL. Loci were defined as novel if the lead variant is (1) at least 500 kb away and confirmed by GCTA to be conditionally independent from previously reported T2D-associated variants in any ancestry, and (2) assessed using locuszoom plots and biology lookups to be away from known loci with extended LD. Four additional variants met the definition for a novel locus but are located within the previously reported major histocompatibility complex (MHC) region; see Supplementary Table S5 for the full list of distinct association signals at the MHC region. rs4804181 was >500kb from primary signal rs3111316 in European meta-analysis, but <500kb from their secondary signal rs755734872. Genome-wide significant association is defined as P<5.0E-08. Physical position based on hg19. Effect alleles are assoicated with increased risk for T2D. Odds ratios reflect per allele effects of variants on T2D risk.

Abbreviations: Chr, chromosome; Pos, position; RA, risk allele; NRA, non-risk allele; RAF, risk allele frequency; Neff, effective sample size; OR, odds ratio; CI, confidence interval; P, P-value

815
816
817
818
819

Figure 1: Two distinct T2D-association signals at the *ANK1-NKX6-3* locus associated with expression levels of two transcripts in two tissues. (A) Regional association plot for East Asian sex-combined BMI-unadjusted meta-analysis at *ANK1-NKX6-3* locus. Approximate conditional analysis using GTCA identified three distinct T2D-association signals at this locus (signal 1, rs33981001; signal 2, rs62508166; signal 3, rs144239281, in order of strength of association). Using 1000G Phase3 East Asian LD, variants are colored in red and blue with the first and second distinct signals respectively (lead variants represented as diamonds). (B) Variant rs12549902, in high LD (EAS LD r^2=0.80, EUR r^2=0.83) with T2D signal 1, shows the strongest association with expression levels of *NXK6-3* in pancreatic islets in 118 individuals16. (C) Variant rs516946, in high LD (EAS LD r^2=0.96, EUR r^2=0.80) with T2D signal 2, shows the strongest association with expression levels of *ANK1* in subcutaneous adipose tissue in 770 individuals[19]. As rs62508166 is not available in the subcutaneous adipose tissue data set, a variant in perfect LD (rs28591316) was used and is represented by the blue diamond variant.

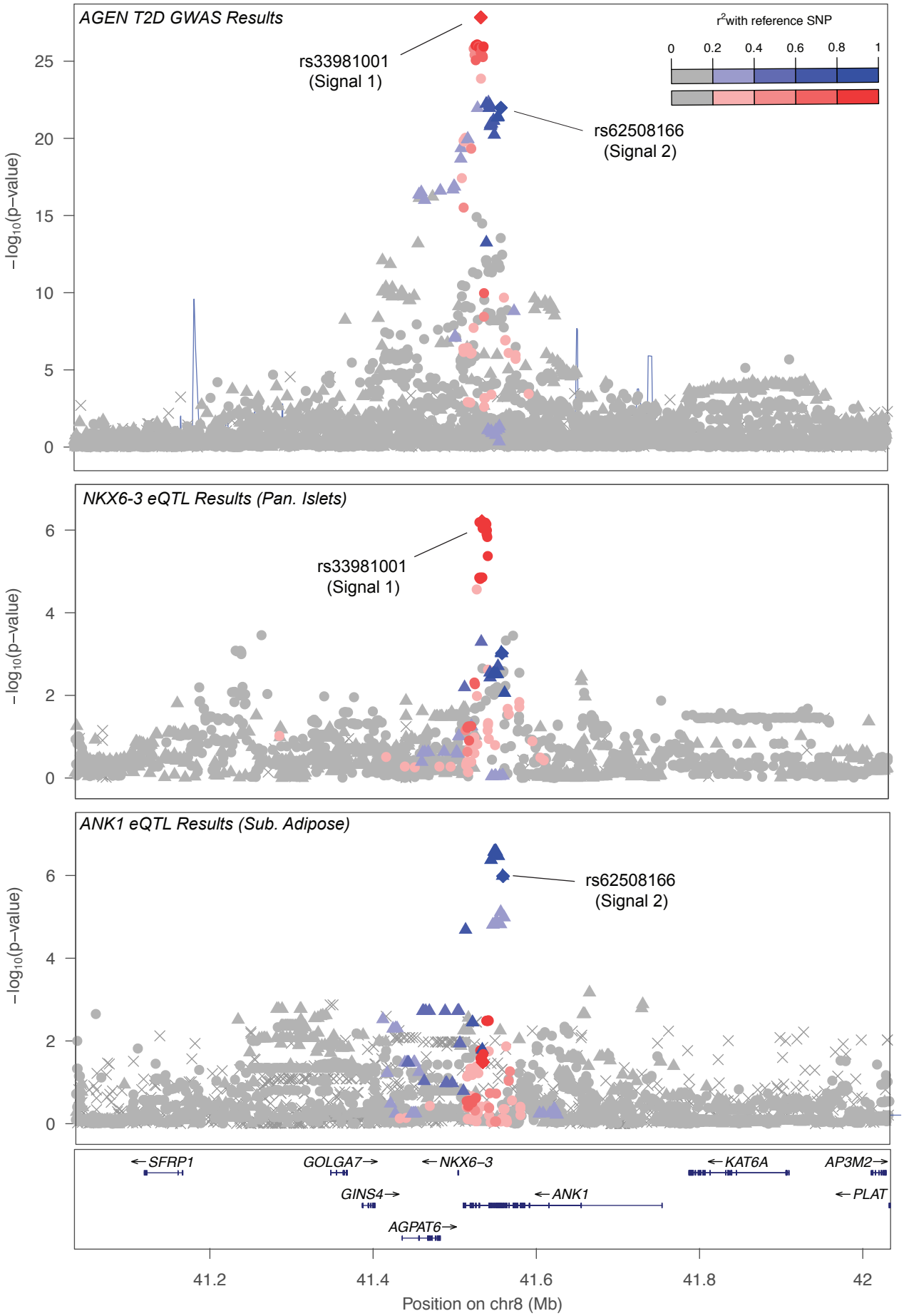
AGEN T2D GWAS Results
r2with reference SNP
0 0.2 0.4 0.6 0.8 1
rs33981001
(Signal 1)
rs62508166
(Signal 2)
−log10(p−value)
NKX6-3 eQTL Results (Pan. Islets)
rs33981001
(Signal 1)
−log10(p−value)
ANK1 eQTL Results (Sub. Adipose)
rs62508166
(Signal 2)
−log10(p−value)
SFRP1
GOLGA7
NKX6−3
KAT6A
AP3M2
GINS4
ANK1
PLAT
AGPAT6
41.2
41.4
41.6
41.8
42
Position on chr8 (Mb)

Figure 1

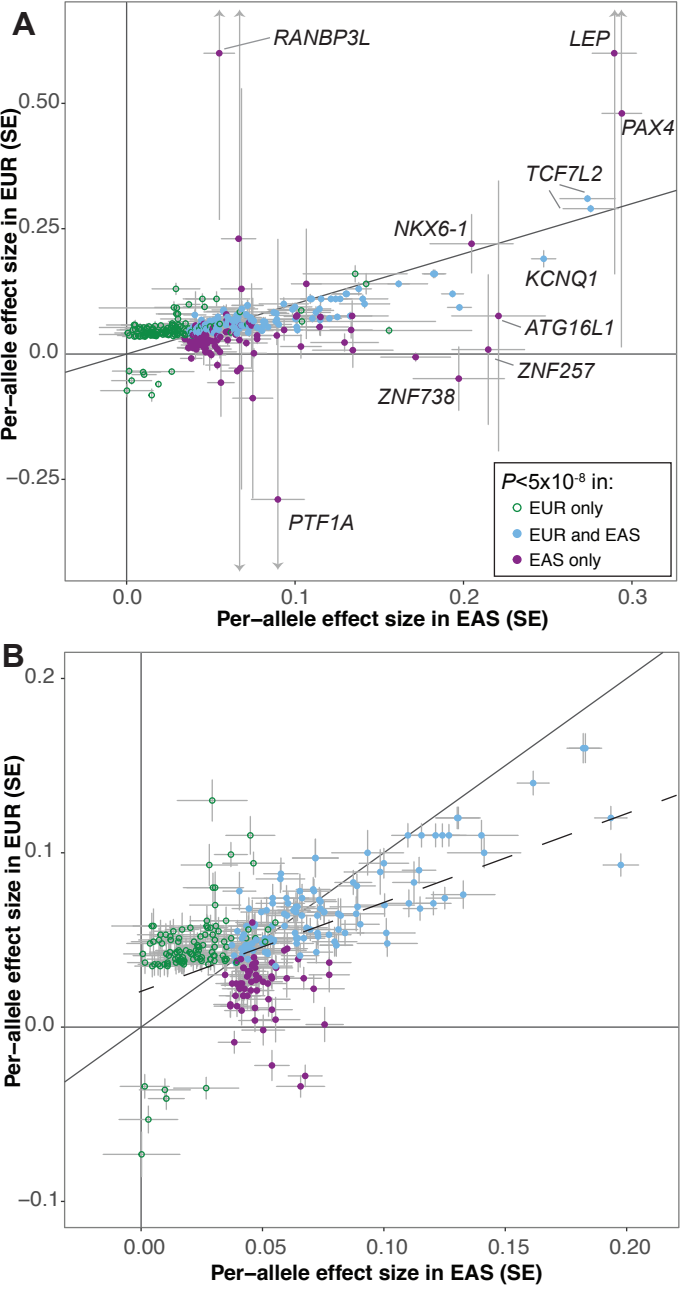

Figure 2: Effect size comparison of lead variants identified in this T2D GWAS BMI-unadjusted meta-analysis in East Asians and previous T2D GWAS meta-analysis in Europeans[2]. For 343 lead variants identified from the two BMI unadjusted meta-analyses, per-allele effect sizes (beta) from Europeans meta-analysis (y-axis) were plotted against per-allele effect sizes from this East Asians meta-analysis (x-axis). (A) All 343 lead variants; (B) 290 lead variants with minor allele frequency at least 5% in both ancestries. Variants are colored purple if they were significant (P<5x10-8) in East Asians only, green if they were significant in Europeans only, and blue if they were significant in both East Asians and Europeans (see Methods and Supplementary Table 7).

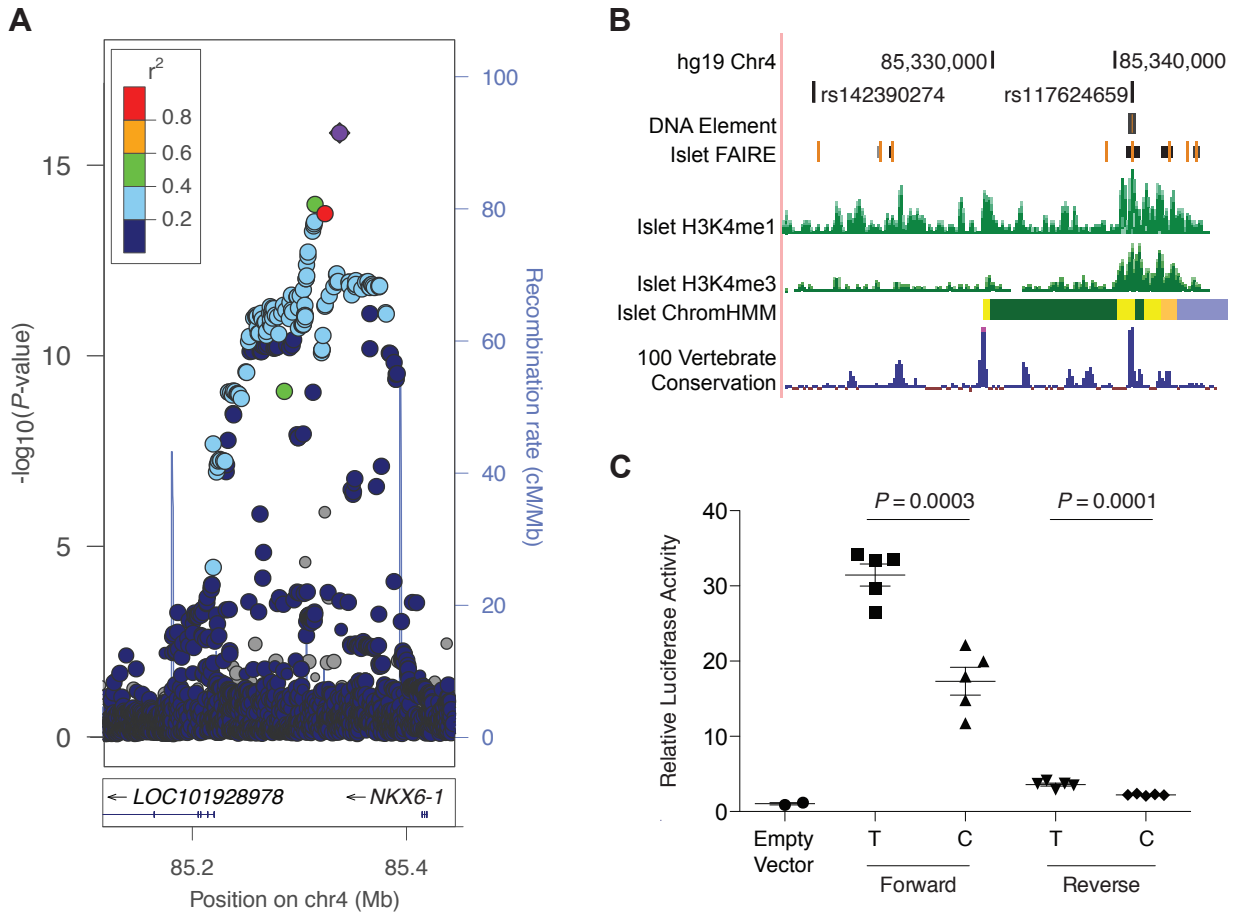

Figure 3: rs117624659 at *NKX6-1* locus exhibits allelic differences in transcriptional activity. (A) rs117624659 (purple diamond) shows the strongest association with type 2 diabetes in the region. Variants are colored based on 1000G Phase 3 East Asian LD with rs117624659. (B) rs117624659 and an additional candidate variant rs142390274 in high pairwise LD (r^2>0.80) span a 22 kb region approximately 75 kb upstream of *NKX6-1*. rs117624659 overlaps a region of open chromatin in pancreatic islets and lies within a region conserved across vertebrates. (C) rs117624659-T, associated with increased risk of T2D, showed greater transcriptional activity in an element cloned in both forward and reverse orientations with respect to *NKX6-1* in MIN6 cells compared to rs117624659-C and an "empty vector" containing a minimal promoter.

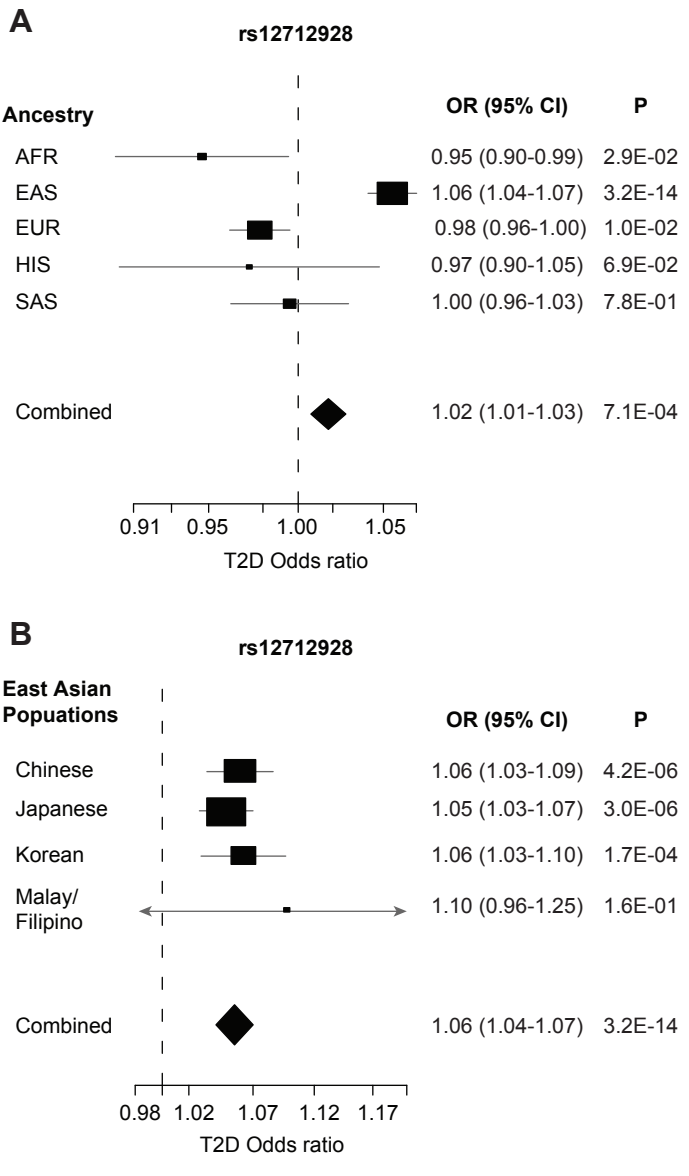

Figure 4: Forest plots of BMI-unadjusted meta-analysis association results at *SIX3-SIX2* locus. Odds ratios (black boxes) and 95% confidence intervals (horizontal lines) for T2D associations at the lead East Asian variant (rs12721928) are presented (A) across ancestries of African-American (AFR), East Asian (EAS), European (EUR)[2], Hispanic (HIS), and South Asian (SAS) individuals, and (B) within East Asians by four major East Asian populations (Chinese, Japanese, Korean, and Malay/Filipino combined due to small sample sizes). The size of the box is proportional to the sample size of each contributing population.

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

820 **FIGURE LEGENDS**
821
822 **Figure 1: Two distinct T2D-association signals at the *ANK1-NKX6-3* locus associated with expression**
823 **levels of two transcripts in two tissues.** (A) Regional association plot for East Asian sex-combined BMI-
824 unadjusted meta-analysis at *ANK1-NKX6-3* locus. Approximate conditional analysis using GTCA identified
825 three distinct T2D-association signals at this locus (signal 1, rs33981001; signal 2, rs62508166; signal 3,
826 rs144239281, in order of strength of association). Using 1000G Phase3 East Asian LD, variants are
827 colored in red and blue with the first and second distinct signals respectively (lead variants represented
828 as diamonds). (B) Variant rs12549902, in high LD (EAS LD r^2=0.80, EUR r^2=0.83) with T2D signal 1, shows
829 the strongest association with expression levels of *NXK6-3* in pancreatic islets in 118 individuals[16]. (C)
830 Variant rs516946, in high LD (EAS LD r^2=0.96, EUR r^2=0.80) with T2D signal 2, shows the strongest
831 association with expression levels of *ANK1* in subcutaneous adipose tissue in 770 individuals[19]. As
832 rs62508166 is not available in the subcutaneous adipose tissue data set, a variant in perfect LD
833 (rs28591316) was used and is represented by the blue diamond variant.
834
835 **Figure 2: Effect size comparison of lead variants identified in this East Asian T2D GWAS BMI-**
836 **unadjusted meta-analysis and previous European T2D GWAS meta-analysis[2].**
837 For 343 lead variants identified from the two BMI unadjusted meta-analyses, per-allele effect sizes (β)
838 from the European meta-analysis (*y*-axis) were plotted against per-allele effect sizes from this East Asian
839 meta-analysis (*x*-axis). (A) All 343 lead variants; (B) 290 lead variants with minor allele frequency ≥5% in
840 both ancestries. Variants are colored purple if they were significant (P<5x10^{-8}) in the East Asian analysis
841 only, green if they were significant in European analysis only, and blue if they were significant in both
842 the East Asian and European analysis (see Methods and Supplementary Table 7). The dashed diagonal
843 line represents the trend line across all plotted variants.
844
845 **Figure 3: rs117624659 at *NKX6-1* locus exhibits allelic differences in transcriptional activity.** (A)
846 rs117624659 (purple diamond) shows the strongest association with T2D in the region. Variants are
847 colored based on 1000G Phase 3 East Asian LD with rs117624659. (B) rs117624659 and an additional
848 candidate variant rs142390274 in high pairwise LD (r^2>0.80) span a 22 kb region approximately 75 kb
849 upstream of *NKX6-1*. rs117624659 overlaps a region of open chromatin in pancreatic islets and lies
850 within a region conserved across vertebrates. (C) rs117624659-T, associated with increased risk of T2D,
851 showed greater transcriptional activity in an element cloned in both forward and reverse orientations
852 with respect to *NKX6-1* in MIN6 cells compared to rs117624659-C and an "empty vector" containing a
853 minimal promoter.
854
855 **Figure 4: Forest plots of BMI-unadjusted meta-analysis association results at *SIX3-SIX2* locus.** Odds
856 ratios (black boxes) and 95% confidence intervals (horizontal lines) for T2D associations at the lead East
857 Asian variant (rs12721928) are presented (A) across ancestries of African-American (AFR), East Asian
858 (EAS), European (EUR)[2], Hispanic (HIS), and South Asian (SAS) individuals, and (B) within four major East
859 Asian populations (Chinese, Japanese, Korean, and Malay/Filipino combined due to small sample sizes).
860 The size of the box is proportional to the sample size of each contributing population.
861
862 **Supplementary Figure 1: Flow chart of study design, depicting the different data analyses performed.**
863
864 **Supplementary Figure 2: Manhattan plot for East Asian T2D meta-analysis association results in model**
865 **unadjusted for BMI.** -log$_{10}$(P) from fixed effects inverse variance weighted genome-wide meta-analysis
866 association results for each variant (*y*-axis) was plotted against the genomic position (hg19; *x*-axis).
867 Known T2D loci achieving genome-wide significance (P<5.0x10^{-8}) meta-analysis are shown in blue. Loci

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

achieving genome-wide significance that are previously unreported for T2D association are shown in red.

Supplementary Figure 3: The relationship between effect size and minor allele frequency. Odds ratios (*y*-axis) and minor allele frequencies (*x*-axis) for 178 primary association signals from the T2D BMI-unadjusted models.

Supplementary Figure 4: Regional association plots at seven T2D associated loci with the strongest association *P*-value and more than five distinct association signals in East Asians. (A) *INS/IGF2/KCNQ1*, (B) *CDKN2A/B*, (C) *PAX4/LEP*, (D) *CDKAL1*, (E) *HHEX/IDE*, (F) *CDC123/CAMK1D*, and (G) *TCF7L2*. Variants are colored based on East Asian 1000G Phase 3 LD with the lead variants for each association signal, shown as diamonds.

Supplementary Figure 5: Miami plots of East Asian T2D meta-analysis association results adjusted for BMI. For each plot, -$\log_{10}$(*P*) from fixed effects inverse-variance weighted genome-wide meta-analysis association results for each variant (*y*-axis) was plotted against the genomic position (hg19; *x*-axis). (A) Sex-combined meta-analyses in models unadjusted for BMI (top) and adjusted for BMI (bottom). Both sex-combined models include the same set of studies for comparable sample size. Novel T2D-associated loci are shown in blue (models unadjusted for BMI), purple (models adjusted for BMI), or green (both); (B) sex-specific meta-analyses for males (top) and females (bottom) without BMI adjustment; and (C) sex-specific meta-analyses for males (top) and females (bottom) with BMI adjustment. For (B) and (C), loci significantly associated with T2D in females only are shown in purple and loci significantly associated with T2D in males only are shown in blue.

Supplementary Figure 6: Effect size comparison of lead variants in sex-combined models unadjusted and adjusted for BMI. At 178 lead variants identified in the East Asian BMI-unadjusted sex-combined T2D meta-analysis, per-allele effect sizes (β) from BMI-adjusted sex-combined models were plotted against BMI-unadjusted sex-combined model. Both sex-combined models include the same set of studies for comparable sample size. Error bars indicate 95% confidence intervals. Effect sizes between the two models are highly correlated with a Pearson correlation coefficient r=0.99 (Supplementary Table 4).

Supplementary Figure 7: Regional plots of male-specific T2D-associated locus, *ALDH2*. (A) Males only, (B) sex-combined, and (C) females only. For each plot, -$\log_{10}$(*P*) from association results for each variant (*y*-axis) was plotted against the genomic position (hg19; *x*-axis). The lead variant rs12231737 plotted is the lead variant from BMI-unadjusted male-specific meta-analysis, and also the sex-combined meta-analysis from the same subset of individuals included in the sex-stratified analyses. This lead variant rs12231737 is in high LD with rs77768175, identified from the larger BMI-unadjusted sex-combined meta-analysis (East Asian r^2=0.80). Variants are shaded based on East Asian 1000G Phase 3 LD with the lead variant, shown as a purple diamond.

Supplementary Figure 8: Effect size comparison of common lead variants (MAF≥5%) identified in East Asian meta-analysis and previously published European T2D GWAS meta-analysis[2].
For 290 lead variants with MAF≥5% in both East Asian and European BMI-unadjusted meta-analyses, per-allele effect sizes (β) from Mahajan et al.[2] (*y*-axis) were plotted against per-allele effect sizes from this East Asian meta-analysis (*x*-axis). Variants are colored purple if they were significant in the East Asian meta-analysis only, green if they were significant in European meta-analysis only, and blue if they

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

were significant in both the East Asian and European meta-analyses. Error bars indicate 95% confidence intervals. (see Methods and Supplementary Table 7).

Supplementary Figure 9: Effect size comparison of lead variants identified in East Asian BMI-unadjusted meta-analysis and previously published European T2D GWAS meta-analysis[2].
For 343 lead variants identified from the two BMI-unadjusted meta-analyses, per-allele effect sizes (β) from a European meta-analysis (*y*-axis) were plotted against per-allele effect sizes from this East Asian meta-analysis (*x*-axis). (A) 162 lead variants significant in the East Asian meta-analysis (purple) or both the East Asian and European meta-analysis (blue) and (B) 192 lead variants significant in the European meta-analysis (green) or both the East Asian and European meta-analysis (blue). These plots include only one variant per locus, in contrast to Figure 2 and Supplementary Figure 8.

Supplementary Figure 10: T2D-association near *ZNF257* and its relationship with a previously reported *ZNF257* inversion. The lead T2D-associated variant rs142395395 near *ZNF257* tags an inversion observed almost exclusively in East Asian individuals. We observed an association between the variants tagging the inversion and a decreased risk for T2D. In the reference haplotype ("Reference"), there is no disruption to *ZNF257* and its promoter, thus there is normal ZNF257 function and normal expression of downstream transcripts. When the alternate alleles are present ("Inversion"), an inversion is observed, marked by the separation of the promoter and first two exons of *ZNF257* from the rest of the gene and moving them over 400 kb upstream. While this has not been observed to affect expression of *ZNF208*, *ZNF43*, or *ZNF100,* the inversion results in a loss of ZNF257 function and altered expression of downstream targets[33].

Supplementary Figure 11: rs117624659 at *NKX6-1* exhibits allelic differences in transcriptional activity. (A) Chromatin marks in pancreatic islets in the intergenic region near *NKX6-1* from the UCSC Genome Browser (hg19) spanning the lead T2D associated variant, rs117624659, and the only other candidate variant rs142390274 in high LD (East Asian r^2>0.80) with rs117624659. *NKX6-1* is transcribed from right to left. rs117624659 overlaps regions of accessible chromatin in human pancreatic islets detected by islet FAIRE-seq along with H3K4me1 and H3K4me3 ChIP-seq. It also overlaps a conserved region in vertebrates. The tested candidate regulatory DNA element is represented by a horizontal black rectangle in the upper portion of the figure. (B) rs117624659 at the *NKX6-1* locus exhibited significant allelic differences in transcriptional activity in MIN6 mouse insulinoma cells on a second day. (C-D) rs11762465 at the *NKX6-1* locus did not exhibit significant allelic differences in transcriptional activity in 832/13 rat insulinoma cells on experimental day 1 (C) or day 2 (D).

Supplementary Figure 12: Forest plots of 49 novel T2D-associated variants using data from five ancestries in the DIAMANTE consortium. T2D meta-analysis results for the 49 novel T2D-associated variants identified in this East Asian meta-analysis were obtained from the other four ancestry groups within the DIAMANTE consortium (African-American, European, Hispanic, and South Asian) for comparison. Odds ratios (black boxes) and 95% confidence intervals (horizontal lines) for the association between the lead East Asian variants and T2D from each ancestry are presented, along with a combined odds ratio and *P*-value. The size of the box is proportional to the sample size of each contributing ancestry group.

Supplementary Figure 13: Forest plot of BMI-unadjusted East Asian meta-analysis association results at *SIX3-SIX2* locus from each contributing cohort. Odds ratios (black boxes) and 95% confidence intervals (horizontal lines) for T2D associations at the lead East Asian variant (rs12721928) are presented

bioRxiv preprint doi: https://doi.org/10.1101/685172; this version posted June 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

962 for each East Asian contributing cohort. The size of the box is proportional to the sample size of each
963 dataset. Full study names can be found in Supplementary Table 1.