Nat Genet. Author manuscript; available in PMC 2013 September 16.

Published in final edited form as:
Nat Genet. ; 43(10): 984–989. doi:10.1038/ng.921.

Genome-wide association study in people of South Asian ancestry identifies six novel susceptibility loci for type 2 diabetes

Jaspal S Kooner[1,2,3,46], **Danish Saleheen**[4,5,46], **Xueling Sim**[6,46], **Joban Sehmi**[1,2,46], **Weihua Zhang**[7,46], **Philippe Frossard**[4,46], **Latonya F Been**[8], **Kee-Seng Chia**[6,9], **Antigone S Dimas**[10,11], **Neelam Hassanali**[12], **Tazeen Jafar**[13,14], **Jeremy BM Jowett**[15], **Xinzhing Li**[1], **Venkatesan Radha**[16], **Simon D Rees**[17,18], **Fumihiko Takeuchi**[19], **Robin Young**[5], **Tin Aung**[20,21], **Abdul Basit**[22], **Manickam Chidambaram**[16], **Debashish Das**[2], **Elin Grunberg**[23], **Åsa K Hedman**[11], **Zafar I Hydrie**[22], **Muhammed Islam**[13], **Chiea-Chuen Khor**[24,6,21], **Sudhir Kowlessur**[25], **Malene M Kristensen**[15], **Samuel Liju**[16], **Wei-Yen Lim**[6], **David R Matthews**[12], **Jianjun Liu**[24], **Andrew P Morris**[11], **Alexandra C Nica**[10], **Janani M Pinidiyapathirage**[26], **Inga Prokopenko**[11], **Asif Rasheed**[4], **Maria Samuel**[4], **Nabi Shah**[4], **A Samad Shera**[27], **Kerrin S Small**[23,28], **Chen Suo**[6], **Ananda R Wickremasinghe**[26], **Tien Yin Wong**[20,21,29], **Mingyu Yang**[30], **Fan Zhang**[30], **DIAGRAM**[31], **MuTHER**[31], **Goncalo R Abecasis**[32], **Anthony H Barnett**[17,18], **Mark Caulfield**[33], **Panos Deloukas**[34], **Tim Frayling**[35], **Philippe Froguel**[36], **Norihiro Kato**[19], **Prasad Katulanda**[12,37], **M Ann Kelly**[17,18], **Junbin Liang**[30], **Viswanathan Mohan**[16,38], **Dharambir K Sanghera**[8], **James Scott**[1], **Mark Seielstad**[39], **Paul Z Zimmet**[15], **Paul Elliott**[7,40,46], **Yik Ying Teo**[6,9,24,41,42,46], **Mark I McCarthy**[11,12,43,46], **John Danesh**[5,46], **E Shyong Tai**[9,44,45,46], and **John C Chambers**[2,3,7,46]

[1]NHLI, Imperial College London, Hammersmith Hospital, Ducane Road, London, W12 0NN, UK [2]Ealing Hospital NHS Trust, Uxbridge Road, Middlesex, UB1 3HW, UK [3]Imperial College Healthcare NHS Trust, DuCane Road, London, W12 0HS [4]Center for Non-Communicable Diseases Pakistan, Karachi, 75300, Pakistan [5]Department of Public Health and Primary Care, University of Cambridge, worts causeway, Cambridge CB1 8RN, UK [6]Centre for Molecular Epidemiology, National University of Singapore, Singapore [7]Epidemiology and Biostatistics, Imperial College London, Norfolk Place, London, W2 1PG, UK [8]Department of Pediatrics, Section of Genetics, College of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA [9]Department of Epidemiology and Public Health, National University of

Addresses for correspondence Dr John C Chambers, Department of Epidemiology and Public Health, Imperial College London, Norfolk Place, London W2 1PG john.chambers@ic.ac.uk; Professor Jaspal S Kooner, National Heart and Lung Institute, Imperial College London, Hammersmith Hospital Campus, Ducane Road, London W11 0HS, UK. j.kooner@imperial.ac.uk.
[46]These authors contributed equally to the work

The following contributed equally to the study:
JSK, DS, XS, JSehmi, WZ, PFrossard, PE, YYT, MIM, JD, EST and JCC.

Manuscript preparation:
JSK, DS, XS, WZ, JSehmi, WZ, PE, YYT, MIM, JD, EST and JCC wrote the manuscript. All authors read and provided critical comment on the manuscript.

Data collection and analysis in the participating studies:
COBRA study: TJ, MI, TF; Chennai Urban Rural Epidemiology Study: VR, MChidambaram, SL, VM; Diabetes Genetic in Pakistan and UK Asian Diabetes Studies: SDR, AB, ZIH, ASS, AHB, MAK; London Life Sciences Population Study: JSK, WZ, JSehmi, XL, DD, GRA, JS, MCaulfield, PFroguel, PE, MIM, JCC; Mauritius study: JBMJ, SK, MMK, PZZ; Pakistan Risk of Myocardial Infarction Study: DS, PFrossard, RY, AR, MSamuel, NS, PD, JD; Ragama Health Study: NK, FT, ARW, JMP; Singapore Consortium of Cohort Studies: KSC, WL, CK, JLiu, EST; Sikh Diabetes Study: LFB, DKS; Singapore Indian Eye Study: XS, CS, TA, TW, MSeielstad, YYT, EST; Sri Lankan Diabetes Study: NH, IP, DRM, PK, MIM.

Other contributions:
Sequencing of T2D loci: MY, FZ, JLiang, XL, JSK, JCC; Association results amongst Europeans in DIAGRAM: APM, MIM; eQTL analyses in MuTHER: ASD, EG, AKH, ACN, KSS, MIM.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Singapore, Singapore [10]Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland [11]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK [12]Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Churchill Hospital, Oxford, UK [13]Department of Community Health Sciences, Aga Khan University, Stadium Road, P.O. Box 3500, Karachi-74800, Pakistan [14]Department of Medicine, Aga Khan University, Stadium Road, P.O. Box 3500, Karachi-74800, Pakistan [15]Baker IDI Heart and Diabetes Institute, Melbourne, Victoria, 3004, Australia [16]Department of Molecular Genetics, Madras Diabetes Research Foundation-ICMR Advanced Centre for Genomics of Diabetes, Chennai 603 103, India [17]College of Medical and Dental Sciences, University of Birmingham [18]BioMedical Research Centre, Heart of England NHS Foundation Trust, Birmingham [19]Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan, 162-8655 [20]Department of Ophthalomolgy, National University of Singapore, Singapore [21]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore [22]Baqai Institute of Diabetology and Endocrinology, Karachi, Pakistan [23]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK [24]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore [25]Ministry of Health, Port Louis, Mauritius [26]Department of Public Health, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka, 11010 [27]Diabetic Association Pakistan, Karachi, Pakistan [28]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK [29]Center for Eye Research Australia, University of Melbourne, Australia [30]Beijing Genomics Institute, Shenzhen 518083, China [31]Members of the DIAGRAM and MuTHER study are listed in the Supplementary Online Material [32]Center for Statistical Genetics, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan [33]Clinical Pharmacology and Barts and the London Genome Centre, William Harvey Research Institute, Barts and the London School of Medicine, Queen Mary University of London, London, UK [34]Human Genetics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA UK [35]Genetics of Complex Traits, Institute of Biomedical and Clinical Science, Peninsula Medical School, University of Exeter, Exeter, UK [36]Genomics of Common Diseases, School of Public Health, Imperial College London, Hammersmith Hospital, Du Cane Road, London W12 0NN [37]Diabetes Research Unit, Dept of Clinical Medicine, Univ of Colombo, Sri Lanka [38]Dr Mohan's Diabetes Specialties Centre, Chennai 600086, India [39]Institure of Human Genetics, University of California, San Francisco [40]MRC-HPA Centre for Environment and Health, Imperial College London, Norfolk Place, London, W2 1PG, UK [41]Department of Statistics and Applied Probability, National University of Singapore, Singapore [42]NUS Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore [43]Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK [44]Department of Medicine, National University of Singapore, Singapore [45]Duke-National University of Singapore Graduate Medical School, Singapore 169857,Singapore

Abstract

We carried out a genome wide association study of type-2 diabetes (T2D) amongst 20,119 people of South Asian ancestry (5,561 with T2D); we identified 20 independent SNPs associated with T2D at $P<10^{-4}$ for testing amongst a further 38,568 South Asians (13,170 with T2D). In combined analysis, common genetic variants at six novel loci (*GRB14*, *ST6GAL1*, *VPS26A*, *HMG20A*, *AP3S2* and *HNF4A*) were associated with T2D ($P=4.1\times10^{-8}$ to $P=1.9\times10^{-11}$); SNPs at *GRB14* were also associated with insulin sensitivity, and at *ST6GAL1* and *HNF4A* with pancreatic beta-cell function respectively. Our findings provide additional insight into mechanisms underlying T2D, and demonstrate the potential for new discovery from genetic association studies in South Asians who have increased susceptibility to T2D.

People of South Asian ancestry are at up to 4 fold higher risk of type-2 diabetes (T2D) compared to European populations[1, 2]. T2D currently affects ~55 million South Asians worldwide, and is projected to affect ~80 million South Asians by 2030[3], which will comprise one-quarter of all people with T2D worldwide. Though unhealthy diet, obesity and physical inactivity contribute, T2D also has an important genetic contribution[4]. T2D is heritable in South Asians, as in other populations[5].

To date, genome-wide association (GWA) studies have identified variants in and around 42 genes as determinants of T2D risk[4]. However, these studies have been predominantly performed in populations of European ancestry. We therefore used genome-wide association to identify common genetic variants underlying risk of T2D in South Asians. The study design is shown in Figure 1.

Genome-wide association and replication testing

In stage one, GWA scans were done in 5,561 South AsianT2D cases and 14,458 South Asian controls from the London Life Sciences Population (LOLIPOP) study[6], the Pakistan Risk of Myocardial Infarction Study (PROMIS)[7] and Singapore Indian Eye (SINDI) Study[8], using Illumina genotyping arrays. South Asians were identified as people originating from the Indian subcontinent (India, Pakistan, Sri Lanka and Bangladesh). Characteristics of participants and genotyping arrays used are summarised (Supplementary note and Supplementary Table 1). Samples with <95% call rate were excluded as were SNPs with call rate < 97%, Hardy-Weinburg $P<10^{-6}$ or minor allele frequency (MAF) <1%. Principal components analysis was used to assess for population substructure[9]; results confirmed that the GWA samples were representative of South Asian ancestry, with no evidence for significant stratification between cases and controls, or between the three studies (Supplementary Figures 1 and 2).

The primary analysis tested the association with T2D of the 568,976 autosomal SNPs that had been directly genotyped and passed QC, amongst the South Asian GWA participants. We chose to limit the primary analysis to directly genotyped, rather than imputed SNPs, since a South Asian specific dense haplotype map has not need described. This approach was enabled by use of Illumina microarrays for all samples (17,880 samples on Illumina 610/660; 2,139 samples on Illumina 317). SNP associations with T2D were tested separately amongst men and women in each study, using logistic regression, and an additive genetic model. Principal components were included as covariates to adjust for population substructure[9]; no other covariates were used in the regression analyses. Results from the separate studies were combined by fixed effects inverse variance meta-analysis implemented in METAL, and with output association test results adjusted for genomic control inflation factor. There was no evidence for inflation of test statistics (λ = 1.00 to 1.03, Supplementary Table 2, Supplementary Figure 3).

In the primary analysis, one locus reached genome-wide significance ($P<5\times10^{-8}$); the lead SNP was rs7903146 in *TCF7L2*, a well described T2D risk variant (Figure 2)[10]. There were a further 59 independent loci associated with T2D at $P\leq1\times10^{-4}$, of which 50 have not previously been described in GWA studies of T2D (Supplementary Table 3). To help prioritise these novel loci for replication testing in stage two, we carried out a combined analysis of the South Asian discovery data with results from the DIAGRAM+ GWA meta-analysis (8,130 T2D cases and 38,987 controls of European ancestry)[11]. We then took forward for replication testing: i. all loci associated with T2D in South Asians alone at $P\leq1\times10^{-5}$ (N=7); and ii. amongst the loci associated with T2D in South Asians at $P>1\times10^{-5}$ and $P\leq1\times10^{-4}$, the 12 independent SNPs with lowest P value in combined analysis with results from DIAGRAM+ (corresponding to ~$P<10^{-3}$). This strategy was designed to

maximise discovery of both genetic loci specific to South Asians, as well as loci shared with European populations.

Replication testing was carried out amongst 13,170 T2D cases and 25,398 controls of South Asian ancestry (Supplementary note and Supplementary Table 4). SNP associations with T2D were tested in each cohort separately, then combined by inverse variance meta-analysis. Six SNPs were associated with T2D in the South Asian replication samples at $P<2.5\times10^{-3}$ ($P<0.05$ after correction for multiple testing): rs3923113 near *GRB14*, rs16861329 in *ST6GAL1*, rs1802295 in *VPS26A*, rs2028299 near *AP3S2*, rs7178572 in *HMG20A*, and rs4812829 in *HNF4A* (Table 1, Supplementary Table 5 and Supplementary Figure 4). The 6 SNPs reached genome-wide significance ($P<5\times10^{-8}$) in combined analysis of GWA data from South Asians, and results for South Asians from the replication samples (Table 1).

Of the 12 SNPs with modest statistical evidence in the South Asian GWA ($P>1\times10^{-5}$ and $P\leq1\times10^{-4}$) that were carried forward for replication testing based on combined analysis with DIAGRAM+, 3 showed replication and reached genome-wide significance in South Asians. This lends support to the view that selection of SNPs based on combined analysis with Europeans is likely to enrich for true associations.

As secondary analyses we carried out the following five GWA meta-analyses: i. male gender specific; ii. female gender specific; iii. BMI adjusted; iv. lean T2D cases ($BMI<25kg/m^2$) vs overweight controls ($BMI>25kg/m^2$); and v. analysis of GWA data imputed with untyped SNPs from the HapMap2 reference panel (total 2,646,472 SNPs). One locus reached $P<5\times10^{-8}$ for association with T2D amongst women (Supplementary Figure 5). However, the sentinel SNP (rs17052370, near *UBBP4*, OR=1.37 [1.23-1.52], $P=8.9\times10^{-9}$) did not replicate amongst South Asian women, either in the replication samples (OR=1.03 [0.97-1.09], P=0.32) or in combined analysis of GWA and replication data (OR=1.10 [1.04-1.15], $P=3.4\times10^{-4}$). There were no additional loci identified from the other secondary analysis (Supplementary Figure 5).

Adiposity and the novel associations with T2D

Since overweight and obesity are major risk factors for T2D[12], we investigated whether the relationship of the sentinel SNPs with T2D might be mediated through adiposity. We found that the sentinel SNPs are not associated with body mass index or waist-hip ratio in South Asians (Supplementary Table 6). Furthermore, the association of the sentinel SNPs with T2D is not materially influenced by additional adjustment for measures of obesity amongst South Asians in the replication samples (Supplementary Table 7). Adiposity therefore does not mediate the relationships between these SNPs and T2D.

Coding and transcriptional mechanisms at the six novel loci

To provide insight into the genetic mechanisms underlying the observed associations with T2D, we sequenced the six regions (1MB either side of sentinel SNPs) in 109 South Asians. We identified 49,145 genetic variants at these loci amongst South Asians, of which 24,902 are not present in the HapMap (www.hapmap.org) or 1000G[13] datasets (Supplementary Table 8). However, there were no pathogenic SNPs (coding, splice site or regulatory) in LD at $r^2>0.5$ with the sentinel SNPs; results were the same in searches of HapMap and 1000G datasets.

To investigate possible transcriptional mechanisms underlying our associations with T2D, we investigated the relationships of sentinel SNPs with cis-eQTLs in adipose tissue, liver, peripheral blood leucocytes and other tissues (Supplementary Table 9). At 2q24, rs3923113

is in high LD (r^2=0.85) with rs10195252, recently reported to be associated with central adiposity and increased expression of *GRB14* in adipose tissue ($P<10^{-10}$)[14]. At 15q26, rs2028299 is closely associated with expression of *C15orf38* in skin and fat ($P<10^{-16}$). At 15q24, rs7178572 is weakly associated with expression of *PSTPIP1*, but is in low LD (r^2<0.3) with the peak SNP associated with this eQTL; *PSTPIP1* is unlikely to be the causal gene at this locus.

Heterogeneity, LD structure and comparisons to Europeans

To test for heterogeneity between the participating South Asian GWA cohorts, we compared effect sizes at i. the loci identified in the present study, and ii. the 42 loci previously reported to be associated with T2D in GWA studies. After correction for multiple testing there was also no evidence for heterogeneity of effect at any of these loci (Supplementary Table 10). Findings were similar amongst South Asians in the replication samples, although there was some evidence for heterogeneity at the *GRB14* locus (Supplementary Figure 4). There was also no evidence for heterogeneity of effect between South Asian men and women, except at rs17052370, which was suggested through the women only gender-specific secondary analysis, but which did not replicate (Supplementary Table 11).

We then investigated whether there was heterogeneity of effect between South Asians and Europeans. Amongst the 42 loci reported to be associated with T2D, 37 show consistent direction of effect amongst South Asians and Europeans ($P=2.8\times10^{-8}$), and 27 are associated with T2D at P<0.05 amongst South Asians (Supplementary Table 12). Only three loci showed statistical evidence for heterogeneity of effect at P<0.001 (P<0.05 after correction for multiple testing). These findings support the view that the effects of common variants are largely shared between populations,[6, 15-17] and lend further support to the strategy of prioritising SNPs from the South Asian GWA through combined analysis with results from Europeans. There was also no evidence for heterogeneity of effect between South Asians and Europeans at the six novel loci discovered in the present study, including those carried forward for further testing based on GWA results from South Asians only (Table 1).

Finally we compared LD structure of the six loci amongst South Asians and Europeans (Supplementary Table 13). There was some evidence at the *VPS26A* locus for different pairwise LD between the two populations, however the differences were small (Supplementary Figure 6). At the *GRB14, ST6GAL1, AP3S2, HMG20A* and *HNF4A* loci haplotype structure were similar, with no evidence for differences in pairwise LD, between South Asians and Europeans.

Six loci associated with T2D

We identify common variants at six novel genetic loci associated with T2D in people of South Asian ancestry. Regional plots of directly genotyped SNPs are shown in Figure 3. Results for genotyped and imputed SNPs are shown in Supplementary Figure 7 and 8, and the top 10 SNPs (genotyped or imputed) at each locus are listed in Supplementary Table 14. At the *GRB14, VPS26A* and *HMG20A* loci, the genotyped SNP is the most strongly associated with T2D. At the *ST6GAL1, AP3S2* and *HNF4A* loci, there were imputed SNPs with stronger P values. For the *STGAL1* and *AP3S2* loci, the sentinel genotyped SNP is in perfect LD with the lead imputed SNP. At *HNF4A*, the sentinel genotyped SNP (rs4812829) and the lead imputed SNP (rs4812831) are in partial LD (r^2=0.43). However, these LD estimates are calculated from HapMap CEU data, so may not be applicable to South Asians.

At 2q24, rs3923113 is nearest to *GRB14*, a strong candidate for the observed association. GRB14 is an adapter protein which binds to insulin receptors and insulin-like growth-factor

receptors, to inhibit tyrosine kinase signalling[18, 19]. *Grb14*−/− mice have higher lean mass, better glucose homeostasis despite lower insulin, and improved insulin sensitivity[20]. Risk allele of rs3923113 is associated with T2D and with reduced insulin sensitivity in South Asians suggesting gain of function (Supplementary Table 6). SNP rs3923113 is also in high LD (r^2=0.85) with rs10195252 (Supplementary Table 9), recently reported to be associated with central adiposity and expression of *GRB14* in adipose tissue[14].

At 3q27, rs16861329 is intronic in *ST6GAL1*, encoding an enzyme predominantly located in the golgi apparatus. ST6GAL1 is involved in post-translational modification of cell-surface components by glycosylation. Although *ST6GAL1* has not previously been linked with glucose metabolism or T2D, glycosylation through addition of sialic acid residues is reported to influence both insulin action and cell surface trafficking[21]. SNP rs16861329 is also near *ADIPOQ* encoding adiponectin, a hormone secreted by adipocytes which promotes insulin sensitivity. Adiponectin knockout-mice show severe insulin resistance[22] and previous candidate gene studies have suggested an association of genetic variation at *ADIPOQ* with adiponectin levels, obesity and T2D[23]. However, these associations have not been consistently found, and rs16861329 (identified in the present study) is not in LD (r^2<0.1) with reported *ADIPOQ* variants.

At 10q22, rs1802295 is in *VPS26A*, encoding a component of the retromer complex, a multimeric protein involved in transport of proteins from endosomes to the trans-Golgi network[24, 25]. *VPS26A* is expressed in pancreatic, adipose and other tissues[26] but a relationship with glucose metabolism or T2D has not been described. At 15q26, rs2028299 is nearest *AP3S2*, encoding a clathrin associated adaptor complex expressed in adipocytes, pancreatic islets and other tissues, which may be involved in vesicle transport and sorting[27]. SNP rs2028299 is associated with expression of *C15orf38*, encoding a member of an uncharacterised family of proteins. Amongst the other genes at this locus, *PLIN1* is also a possible candidate for the association with T2D. *PLIN1* encodes Perilipin-1, a phosphoprotein which coats fat droplets in adipocytes and regulates lipolysis by hormone sensitive lipase[28]. Genetic variation at *PLIN1* has been associated with obesity in man and in experimental animal models[29, 30]. SNP rs2028299 is 1.2MB away from rs8042680 in *PRC1*, which is associated with T2D in Europeans[11]. Although these two SNPs are not in LD (r^2=0), we cannot exclude the possibility of a shared mechanism through remote regulatory effects. At 15q24, known biology does not identify any compelling candidates. SNP rs7178572 is intronic in *HMG20A*, a non-histone chromosomal protein that is widely expressed, which may influence histone methylation and be involved in neuronal development[31, 32].

At 20q13, the lead genotyped SNP rs4812829 is intronic in *HNF4A*, a strong candidate for the observed association. HNF4A is a nuclear transcription factor strongly expressed in liver[33], which regulates transcription of a number of genes including *HNF1A*[34]. Mutations in *HNF4A* are known to cause maturity-onset diabetes of the young (MODY) type 1, characterised by defective pancreatic beta cell function and impaired insulin secretion[35]. In keeping with this, risk allele of rs4812829 is associated with reduced pancreatic beta cell function in South Asians (Supplementary Table 9). At the *HNF4A* locus, there were four genotyped SNPs associated with T2D at $P<10^{-5}$; in conditional analysis the effect sizes are substantially reduced indicating that these are unlikely to be independent signals (Supplementary Table 15). Analysis of imputed data identifies rs4812831 as the strongest signal at this locus (Supplementary Table 14), suggesting that rs4812831 may either be the causal variant, or in high LD with it. The regional plots also reveal a separate cluster of SNPs associated with T2D, that are not in LD with rs4812829 (lead SNP rs12625067), raising the possibility of two separate causal variants at this locus.

Europe PMC Funders Author Manuscripts

This is the first GWA study to investigate genetic factors underlying T2D amongst people of South Asian ancestry who have increased susceptibility to T2D. We identify common genetic variants at six novel loci (*GRB14, ST6GAL1, VPS26A, AP3S2, HMG20A*, and *HNF4A*) associated with T2D. Our findings provide insight into the genetic mechanisms underlying T2D, and demonstrate the potential for new discovery from genetic association studies in populations of non-European ancestry.

Online Methods

Participants

South Asian T2D cases and controls—Genome-wide association was carried out amongst 5,561 South Asian T2D cases and 14,458 South Asian controls from the London Life Sciences Population (LOLIPOP) study, the Pakistan Risk of Myocardial Infarction Study (PROMIS) and the Singapore Indian Eye (SINDI) study. Replication testing amongst South Asians was carried out amongst 13,170 T2D cases and 25,398 controls participants from the following studies: the Chennai Urban Rural Epidemiology Study (CURES)[36] the COBRA study[37] the Diabetes Genetics in Pakistan (DGP) and UK Asian Diabetes Study (UKADS),[38, 39] the Mauritius study[40], the Ragama Health Study (RHS)[41], the Sikh Diabetes Study (SDS)[42], the Singapore Consortium of Cohort Studies (SCCS), the Sri Lankan Diabetes Study (SLDS)[43], and the LOLIPOP and PROMIS studies. For LOLIPOP and PROMIS, there was no overlap of participants between the genome-wide association and replication testing stages. Full details of the contributing cohorts are provided in Supplementary Methods, along with characteristics of participants (Supplementary note and Supplementary Tables 1 and 4).

European T2D cases and controls—Associations of SNPs with T2D amongst Europeans were tested *in silico* using results from the GWA phase of the DIAGRAM+ study, which comprises 8,130 T2D cases and 38,987 controls of European ancestry[11]. T2D case-control status were defined using study specific criteria. SNP associations were tested using an additive genetic model, and combined across studies by inverse variance meta-analysis using a fixed effects model.

Genotyping, quality control and statistical methods

Genome-wide association—Genome-wide association scans were performed using Illumina Infinium Beadchips, genotypes were called using GenCall or Illuminus algorithms (Supplementary Table 1). Samples with a SNP call rate of <95% were removed, as were SNPs with call rate <97%, minor allele frequency<1%, or Hardy–Weinberg equilibrium $P<1.0\times10^{-6}$. Hidden relatedness or duplicate samples were sought using identity-by-descent methods implemented in PLINK; for individuals with evidence for relatedness were excluded (pi_hat≥0.5 in LOLIPOP and SINDI, or pi_hat≥0.37 in PROMIS to allow for the higher prevalence of consanguinuity in Pakistanis). In the absence of an South Asian specific haplotype map, the primary analysis tested the association with T2D of the 568,976 autosomal SNPs that had been directly genotyped and passed QC, amongst the South Asian GWA participants. As secondary analyses, and to help fine-map the loci identified, we repeated the GWA analyses with after imputation of untyped SNPs from HapMap2. Imputation of genotypes not directly measured was performed using pooled haplotypes from the CEU, YRI and CHB/JPT HapMap2 reference panel, and the IMPUTE2 software package, as previously described.[44]

Principal components analysis (PCA) was used to identify population outliers by comparison to reference samples from the Hapmap YRI, CHB, JPT and CEU panels and the Indian samples collected by Reich and colleagues[45]; samples with Eigenvalues inconsistent

with South Asian ancestry were removed. PCA was performed in Eigensoft v3.0[9], using a set of 100,864 SNPs common to all three studies, and pruned to reduce pairwise LD – this SNP set was selected using the 'indep' option of PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/) with window size 50, step 3, VIF=1.3.

Associations of SNPs with T2D were tested using logistic regression and an additive genetic model. Analyses were carried out in men and women separately. Principal components were included as covariates to adjust for residual population stratification; no other covariate adjustments were made. The number of principal component included was study specific: LOLIPOP 10; PROMIS 5 and SINDI 3. A fixed-effects inverse variance meta-analysis was used to combine the results for individual studies. P values were adjusted for study-level genomic control inflation factor before the meta-analysis, and then again for the meta-analysis genomic control inflation factor (double GC correction)[46].

The primary analysis comprised meta-analysis of association results for men and women combined, at the 568,976 SNPs that had been directly genotyped. Secondary analyses included the following five GWA meta-analyses: i. male gender specific; ii. female gender specific; iii. BMI adjusted; iv. lean T2D cases (BMI<25kg/m^2) vs overweight controls (BMI>25kg/m^2); and v. analysis of GWA data imputed with missing genotypes from the HapMap2 reference panel (total 2,646,472 SNPs).

Selection for replication testing—SNPs that were located within a locus that had been previously reported to be associated with T2D were excluded, as were SNPs in LD (r^2>0.5) with another SNP that had a more significant association with T2D.

Twenty SNPs were carried forward for replication testing, comprising: i. all SNPs associated with T2D at P≤10^{-5} in the primary South Asian specific genome-wide association analysis of directly genotyped SNPs (N=7); ii. all SNPs associated with T2D at P<5×10^{-8} in one of the five secondary GWA analyses (N=1, identified from the female gender specific analysis); iii. 12 SNPs prioritised from amongst the 43 SNPs associated with T2D at P>10^{-5} and P≤10^{-4} in the primary analysis, based on lowest P value in fixed-effects inverse variance meta-analysis with results from the GWA stage of the DIAGRAM+ study (corresponding to ~P<10^{-3} in combined analysis)[11].

Replication testing—Genotyping of the replication samples was performed by KASPAR (K-Bioscience Ltd, UK), Sequenom MassArray or TaqMan assays (Supplementary Table 4). Samples with <90% call rate were excluded, as were SNPs with call rate <95% or that deviated from Hardy–Weinberg equilibrium at P<2.5×10^{-3}. The associations of SNPs with T2D were tested in each cohort separately; heterogeneity between studies was assessed using Cochran's Q statistic. A fixed-effects meta-analysis was then used to combine the results for each SNP across all replication studies with available data, and then in combined analysis with results from the genome-wide association stage. Statistical significance was inferred at P<2.5×10^{-3} in the replication stage (ie P<0.05 after Bonferroni correction for 20 SNPs). For the combined analysis of genome-wide and replication data, genome-wide significance was inferred at P<5×10^{-8}.

Power—For SNPs with minor allele frequency >20% the study had >80% power to identify SNPs with odds ratios for T2D of: >1.11 per allele copy at P<10^{-4} in the genome-wide association phase, 1.06 per allele copy at P<2.5×10^{-3} in the replication testing stage, and 1.08 per allele copy at P<5×10^{-8} in combined analysis.

Europe PMC Funders Author Manuscripts

Sequencing

Sequencing was performed using a Genome Analyser-2 platform (Illumina) at the Beijing Genomics Institute, with library preparation and a 91bp paired-end sequencing strategy, according to manufacturer's instructions. Reads were aligned to the human reference genome (NCBI Build 36) using BWA[47], duplicates removed using SAM tools, and genotype likelihood calculated by SOAPsnp[48]. 109 South Asian samples (34 with T2D) were sequenced; average sequencing depth was ~4x.

Comparison of regional LD patterns

We use the varLD algorithm to compare the regional pattern of LD surrounding the index SNPs for each of the six regions[49]. The use of the targeted varLD algorithm tests the null hypothesis that the regional pattern of correlation between every pair of SNPs in the window is identical across two populations, and yields a Monte Carlo *P*-value that effectively quantifies the statistical evidence of a deviation from this identity. In our comparisons, we implemented 1,000 iterations for the Monte Carlo procedure across a 300kb window around the associated SNPs in each region. Each analysis compares between two of the following populations: (i) the 60 Europeans in phase 2 of the HapMap (CEU); (ii) the 83 South Asians from the Singapore Genome Variation Project[50] and (iii) 60 randomly chosen control samples from the LOLIPOP study.

Expression QTLs

To determine whether the T2D-risk variants detected in this study influenced expression of nearby genes, we accessed a variety of sources, including (a) publicly available cis eQTL data for brain[51, 52], lymphoblastoid cell lines[53-55], fibroblasts[55], liver[56], and T-cells[55], (b) expression data from HapMap3 (Stranger et al., under review) and the MuTHER consortium [57]; and (c) cis-eQTL data for *GRB14* reported in a recent genome wide association study of fat distribution[14].

The HapMap3 resource (Stranger et al, under review) comprises LCLs from 726 HapMap3 individuals (CEU: 109, CHB: 80, GIH: 82, JPT: 82, LWK: 82, MEX: 45, MKK: 138, and YRI: 108), with mRNA transcript levels measured using Illumina's whole genome expression array Sentrix Human-6 Expression BeadChip version 2. Log2 transformed expression signals were normalized as follows: quantile normalization across replicates of a single individual, followed by median normalization across all individuals of the eight populations. Expression data from GIH, LWK, MEX, and MKK (populations with admixture) were subjected to correction for genetic structure. For each gene we tested for association between SNP genotype and normalized expression values using Spearman rank correlation (SRC), testing all SNPs mapping within a 2MB window centred on the gene's transcription start site (TSS). Statistical significance was evaluated through permutations of expression phenotypes relative to genotypes defining a significance threshold of 0.01.

The MuTHER resource (www.muther.org) includes LCLs, skin and adipose tissue derived simultaneously from a subset of well-phenotyped healthy female twins[57]. Whole-genome expression profiling of the samples, each with either two or three technical replicates, were performed using the Illumina Human HT-12 V3 BeadChips (Illumina Inc) according to the protocol supplied by the manufacturer. Log2 transformed expression signals were normalized separately per tissue as follows: quantile normalization was performed across technical replicates of each individual followed by quantile normalization across all individuals. Genotyping was done with a combination of Illumina arrays (HumanHap300, HumanHap610Q, 1MDuo and 1.2MDuo). Untyped HapMap2 SNPs were imputed using the IMPUTE software package (v2). The number of samples with genotypes and expression values per tissue is 778 LCL, 667 skin and 776 adipose, respectively. Association between

all SNPs (MAF>5%, IMPUTE info >0.8) within a gene or within 1MB of the gene transcription start or end site and normalized expression values were performed with the GenABEL/ProbABEL packages using the polygenic linear model incorporating a kinship matrix in GenABEL followed by the ProbABEL mmscore score test with imputed genotypes. Age and experimental batch were included as cofactors in the adipose and LCL analysis, while age, experimental batch and concentration were included as cofactors in the skin analysis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank the many colleagues who contributed to collection and phenotypic characterization of the clinical samples, as well as genotyping and analysis of the GWA data. They would also like to acknowledge those who agreed to participate in these studies. Major funding for the work described in this paper comes from Wellcome Trust awards (070854/Z/03/Z, 080747/Z/06/Z, 083270/Z/07/Z, 084723/Z/08/Z); Chennai Wellingdon Corporate Foundation; Diabetes UK (07/0003512); National Institute for Health Research Comprehensive Biomedical Research Centre at Imperial College Healthcare NHS Trust; British Heart Foundation (SP/04/002); Medical Research Council (G0700931); National Institute for Health Research (RP-PG-0407-10371); US National Institutes of Health (DK-25446); KAKENHI (Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan; National Center for Global Health and Medicine; National Institute of Health (KO1TW006087); National Institute of Diabetes and Digestive and Kidney Diseases (R01DK082766); A*STAR Biomedical Research Council (05/1/21/19/425); Biomedical Research Council Singapore (09/1/35/19/616, 08/1/35/19/550); National Medical Research Council Singapore (NMRC/STaR/0003/2008, 1174/2008); National Science Foundation of Sri Lanka; and Oxford NIHR Biomedical Research Centre. A full list of acknowledgments is provided in the Supplementary Online Material.

References

1. Chambers JC, Obeid OA, Refsum H, et al. Plasma homocysteine concentrations and risk of coronary heart disease in UK Indian Asian and European men. Lancet. 2000; 355(9203):523–527. [PubMed: 10683001]
2. Ramachandran A, Ma RC, Snehalatha C. Diabetes in Asia. Lancet. 2010; 375(9712):408–418. [PubMed: 19875164]
3. Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract. 2010; 87(1):4–14. [PubMed: 19896746]
4. McCarthy MI. Genomics, type 2 diabetes, and obesity. N Engl J Med. 2010; 363:2339–50. [PubMed: 21142536]
5. Jowett JB, Diego VP, Kotea N, et al. Genetic influences on type 2 diabetes and metabolic syndrome related quantitative traits in Mauritius. Twin Res Hum Genet. 2009; 12(1):44–52. [PubMed: 19210179]
6. Chambers JC, Zhao J, Terracciano CM, et al. Genetic variation in SCN10A influences cardiac conduction. Nat Genet. 2010; 42(2):149–152. [PubMed: 20062061]
7. Saleheen D, Zaidi M, Rasheed A, et al. The Pakistan Risk of Myocardial Infarction Study: a resource for the study of genetic, lifestyle and other determinants of myocardial infarction in South Asia. Eur J Epidemiol. 2009; 24(6):329–338. [PubMed: 19404752]
8. Lavanya R, Jeganathan VS, Zheng Y, et al. Methodology of the Singapore Indian Chinese Cohort (SICC) eye study: quantifying ethnic variations in the epidemiology of eye diseases in Asians. Ophthalmic Epidemiol. 2009; 16(6):325–336. [PubMed: 19995197]
9. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38(8):904–909. [PubMed: 16862161]
10. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature. 2007; 445(7130):881–885. [PubMed: 17293876]

11. Voight BF, Scott LJ, Steinthorsdottir V, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010; 42(7):579–589. [PubMed: 20581827]
12. Kahn SE, Hull RL, Utzschneider KM. Mechanisms linking obesity to insulin resistance and type 2 diabetes. Nature. 2006; 444(7121):840–846. [PubMed: 17167471]
13. Durbin RM, Abecasis GR, Altshuler DL, et al. A map of human genome variation from population-scale sequencing. Nature. 2010; 467(7319):1061–1073. [PubMed: 20981092]
14. Heid IM, Jackson AU, Randall JC, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet. 2010; 42(11):949–960. [PubMed: 20935629]
15. Chambers JC, Elliott P, Zabaneh D, et al. Common genetic variation near MC4R is associated with waist circumference and insulin resistance. Nat Genet. 2008; 40(6):716–718. [PubMed: 18454146]
16. Kooner JS, Chambers JC, Aguilar-Salinas CA, et al. Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. Nat Genet. 2008; 40(2):149–151. [PubMed: 18193046]
17. Peden JF, Hopewell JC, Saleheen D, et al. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nat Genet. 2011
18. Dufresne AM, Smith RJ. The adapter protein GRB10 is an endogenous negative regulator of insulin-like growth factor signaling. Endocrinology. 2005; 146(10):4399–4409. [PubMed: 16037382]
19. Depetris RS, Wu J, Hubbard SR. Structural and functional studies of the Ras-associating and pleckstrin-homology domains of Grb10 and Grb14. Nat Struct Mol Biol. 2009; 16(8):833–839. [PubMed: 19648926]
20. Holt LJ, Lyons RJ, Ryan AS, et al. Dual ablation of Grb10 and Grb14 in mice reveals their combined role in regulation of insulin signaling and glucose homeostasis. Mol Endocrinol. 2009; 23(9):1406–1414. [PubMed: 19541746]
21. Woodard-Grice AV, McBrayer AC, Wakefield JK, Zhuo Y, Bellis SL. Proteolytic shedding of ST6Gal-I by BACE1 regulates the glycosylation and function of alpha4beta1 integrins. J Biol Chem. 2008; 283(39):26364–26373. [PubMed: 18650447]
22. Maeda N, Shimomura I, Kishida K, et al. Diet-induced insulin resistance in mice lacking adiponectin/ACRP30. Nat Med. 2002; 8(7):731–737. [PubMed: 12068289]
23. Siitonen N, Pulkkinen L, Lindstrom J, et al. Association of ADIPOQ gene variants with body weight, type 2 diabetes and serum adiponectin concentrations: the Finnish Diabetes Prevention Study. BMC Med Genet. 2011; 12(1):5. [PubMed: 21219602]
24. Seaman MN, Marcusson EG, Cereghino JL, Emr SD. Endosome to Golgi retrieval of the vacuolar protein sorting receptor, Vps10p, requires the function of the VPS29, VPS30, and VPS35 gene products. J Cell Biol. 1997; 137(1):79–92. [PubMed: 9105038]
25. Seaman MN, Harbour ME, Tattersall D, Read E, Bright N. Membrane recruitment of the cargo-selective retromer subcomplex is catalysed by the small GTPase Rab7 and inhibited by the Rab-GAP TBC1D5. J Cell Sci. 2009; 122(Pt 14):2371–2382. [PubMed: 19531583]
26. Kim E, Lee JW, Baek DC, et al. Identification of novel retromer complexes in the mouse testis. Biochem Biophys Res Commun. 2008; 375(1):16–21. [PubMed: 18656452]
27. Dell'Angelica EC, Ohno H, Ooi CE, Rabinovich E, Roche KW, Bonifacino JS. AP-3: an adaptor-like protein complex with ubiquitous expression. EMBO J. 1997; 16(5):917–928. [PubMed: 9118953]
28. Brasaemle DL, Rubin B, Harten IA, Gruia-Gray J, Kimmel AR, Londos C. Perilipin A increases triacylglycerol storage by decreasing the rate of triacylglycerol hydrolysis. J Biol Chem. 2000; 275(49):38486–38493. [PubMed: 10948207]
29. Qi L, Corella D, Sorli JV, et al. Genetic variation at the perilipin (PLIN) locus is associated with obesity-related phenotypes in White women. Clin Genet. 2004; 66(4):299–310. [PubMed: 15355432]
30. Beller M, Bulankina AV, Hsiao HH, Urlaub H, Jackle H, Kuhnlein RP. PERILIPIN-dependent control of lipid droplet structure and fat storage in Drosophila. Cell Metab. 2010; 12(5):521–532. [PubMed: 21035762]

31. Sumoy L, Carim L, Escarceller M, et al. HMG20A and HMG20B map to human chromosomes 15q24 and 19p13.3 and constitute a distinct class of HMG-box genes with ubiquitous expression. Cytogenet Cell Genet. 2000; 88(1-2):62–67. [PubMed: 10773667]
32. Artegiani B, Labbaye C, Sferra A, et al. The interaction with HMG20a/b proteins suggests a potential role for beta-dystrobrevin in neuronal differentiation. J Biol Chem. 2010; 285(32): 24740–24750. [PubMed: 20530487]
33. Battle MA, Konopka G, Parviz F, et al. Hepatocyte nuclear factor 4alpha orchestrates expression of cell adhesion proteins during the epithelial transformation of the developing liver. Proc Natl Acad Sci U S A. 2006; 103(22):8419–8424. [PubMed: 16714383]
34. Harries LW, Brown JE, Gloyn AL. Species-specific differences in the expression of the HNF1A, HNF1B and HNF4A genes. PLoS One. 2009; 4(11):e7855. [PubMed: 19924231]
35. Yamagata K, Furuta H, Oda N, et al. Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). Nature. 1996; 384(6608):458–460. [PubMed: 8945471]
36. Chidambaram M, Radha V, Mohan V. Replication of recently described type 2 diabetes gene variants in a South Indian population. Metabolism. 2010; 59(12):1760–1766. [PubMed: 20580033]
37. Jafar TH, Hatcher J, Poulter N, et al. Community-based interventions to promote blood pressure control in a developing country: a cluster randomized trial. Ann Intern Med. 2009; 151(9):593–601. [PubMed: 19884620]
38. Bellary S, O'Hare JP, Raymond NT, et al. Enhanced diabetes care to patients of south Asian ethnic origin (the United Kingdom Asian Diabetes Study): a cluster randomised controlled trial. Lancet. 2008; 371(9626):1769–1776. [PubMed: 18502301]
39. Rees SD, Islam M, Hydrie MZ, et al. An FTO variant is associated with Type 2 diabetes in South Asian populations after accounting for body mass index and waist circumference. Diabet Med. 2011
40. Soderberg S, Zimmet P, Tuomilehto J, et al. High incidence of type 2 diabetes and increasing conversion rates from impaired fasting glucose and impaired glucose tolerance to diabetes in Mauritius. J Intern Med. 2004; 256(1):37–47. [PubMed: 15189364]
41. Takeuchi F, Katsuya T, Chakrewarthy S, et al. Common variants at the GCK, GCKR, G6PC2-ABCB11 and MTNR1B loci are associated with fasting glucose in two Asian populations. Diabetologia. 2010; 53(2):299–308. [PubMed: 19937311]
42. Sanghera DK, Bhatti JS, Bhatti GK, et al. The Khatri Sikh Diabetes Study (SDS): study design, methodology, sample collection, and initial results. Hum Biol. 2006; 78(1):43–63. [PubMed: 16900881]
43. Katulanda P, Constantine GR, Mahesh JG, et al. Prevalence and projections of diabetes and pre-diabetes in adults in Sri Lanka--Sri Lanka Diabetes, Cardiovascular Study (SLDCS). Diabet Med. 2008; 25(9):1062–1069. [PubMed: 19183311]
44. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet. 2007; 39(7):906–913. [PubMed: 17572673]
45. Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. Nature. 2009; 461(7263):489–494. [PubMed: 19779445]
46. Bacanu SA, Devlin B, Roeder K. The power of genomic control. Am J Hum Genet. 2000; 66(6): 1933–1944. [PubMed: 10801388]
47. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010; 26(5):589–595. [PubMed: 20080505]
48. Li R, Li Y, Fang X, et al. SNP detection for massively parallel whole-genome resequencing. Genome Res. 2009; 19(6):1124–1132. [PubMed: 19420381]
49. Ong RT, Teo YY. varLD: a program for quantifying variation in linkage disequilibrium patterns between populations. Bioinformatics. 2010; 26(9):1269–1270. [PubMed: 20308177]
50. Teo YY, Sim X, Ong RT, et al. Singapore Genome Variation Project: a haplotype map of three Southeast Asian populations. Genome Res. 2009; 19(11):2154–2162. [PubMed: 19700652]
51. Myers AJ, Gibbs JR, Webster JA, et al. A survey of genetic human cortical gene expression. Nat Genet. 2007; 39(12):1494–1499. [PubMed: 17982457]

52. Webster JA, Gibbs JR, Clarke J, et al. Genetic control of human brain transcript expression in Alzheimer disease. Am J Hum Genet. 2009; 84(4):445–458. [PubMed: 19361613]
53. Dixon AL, Liang L, Moffatt MF, et al. A genome-wide association study of global gene expression. Nat Genet. 2007; 39(10):1202–1207. [PubMed: 17873877]
54. Ge B, Pokholok DK, Kwan T, et al. Global patterns of cis variation in human cells revealed by high-density allelic expression analysis. Nat Genet. 2009; 41(11):1216–1222. [PubMed: 19838192]
55. Dimas AS, Deutsch S, Stranger BE, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science. 2009; 325(5945):1246–1250. [PubMed: 19644074]
56. Schadt EE, Molony C, Chudin E, et al. Mapping the genetic architecture of gene expression in human liver. PLoS Biol. 2008; 6(5):e107. [PubMed: 18462017]
57. Nica AC, Parts L, Glass D, et al. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet. 2011; 7(2):e1002003. [PubMed: 21304890]

Europe PMC Funders Author Manuscripts

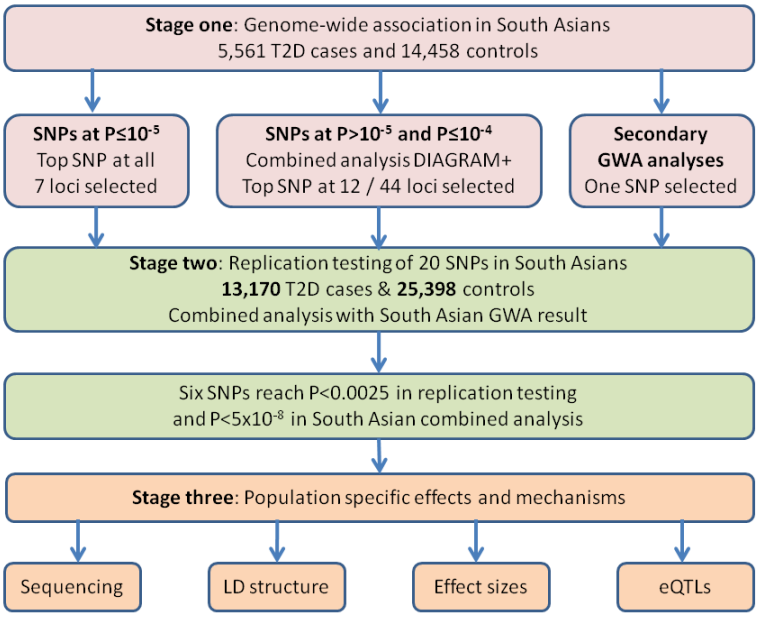

Figure 1.
Summary of study design

Europe PMC Funders Author Manuscripts

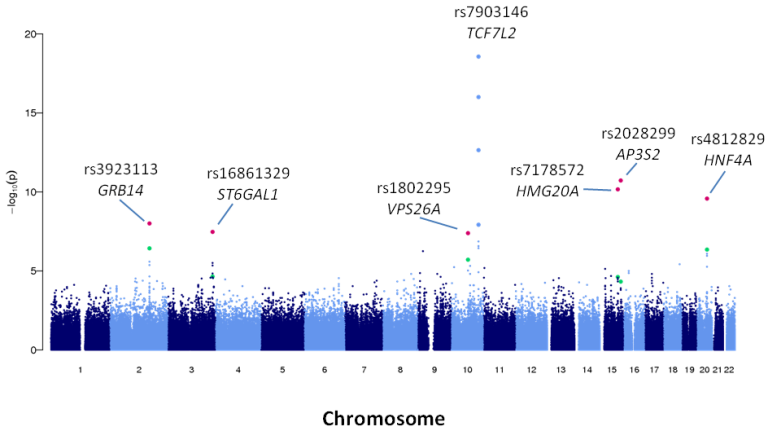

Figure 2.

Manhattan plot for the primary South Asian genome-wide association analysis. of men and women, using directly genotyped SNPs. At the six novel loci reaching genome wide-significance, the sentinel SNPs are indicated: green dot for GWA result, and red dot for combined analysis of GWA and replication data in South Asians.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

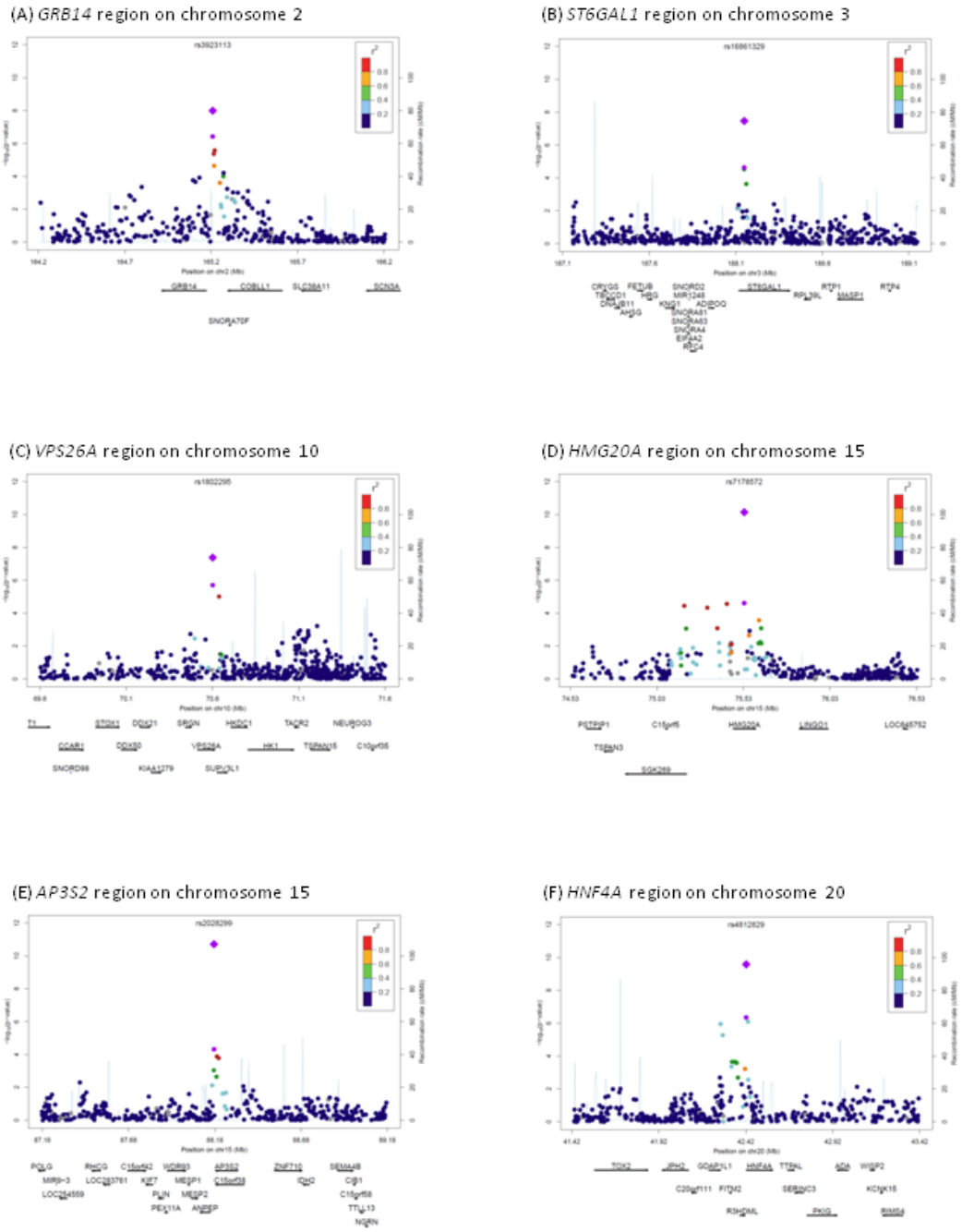

Figure 3.

Regional plots for the six loci associated with type-2 diabetes in South Asians. (A) *GRB14* region, (B) *ST6GAL1* region, (C) *VPS26A* region, (D) *HMG20A* region, (E) *AP3S2* region and (F) *HNF4A* region.

Sentinel SNP: purple circle for genome-wide result; purple diamond for combined analysis with replication studies. Other genotyped SNPs colour coded according to pairwise LD with sentinel SNP, calculated in a representative sample of 83 South Asians from the Singapore Genome Variation Project.[50] Recombination rates estimated from Hapmap Phase II combined panels. Regional plots also incorporating SNPs imputed from HapMap2 are shown in Supplementary Figures 7 and 8.

Europe PMC Funders Author Manuscripts

Table 1

Genomic location and association test results for the sentinel SNPs from the six loci reaching $P<5\times10^{-8}$ amongst South Asians. Results are provided as odds ratio [OR] (95% confidence interval) per copy of risk allele.

				RAF		GWA South Asians			Replication South Asians		Combined analysis South Asians		Europeans (DIAGRAM+)			Global analysis	
SNP	**Chr (position)**	**Nearest Gene**	**Alleles (A/R)**	**SA**	**EW**	**N**	**OR**	**P**	**OR**	**P**	**OR**	**P**	**OR**	**P**	**P_{hetero}**	**OR**	**P**
rs3923113	2 (165210095)	*GRB14*	C / A	0.74	0.64	18,174	1.15 (1.09-1.21)	3.7×10^{-7}	1.07 (1.03-1.11)	6.7×10^{-4}	1.09 (1.06-1.13)	1.0×10^{-8}	1.05 (1.01-1.10)	2.0×10^{-2}	0.64	1.08 (1.05-1.11)	1.6×10^{-9}
rs16861329	3 (188149155)	*ST6GAL1*	A / G	0.75	0.86	18,186	1.12 (1.07-1.19)	2.3×10^{-5}	1.07 (1.03-1.11)	1.6×10^{-4}	1.09 (1.06-1.12)	3.4×10^{-8}	1.02 (0.95-1.09)	0.62	0.21	1.08 (1.05-1.11)	1.3×10^{-7}
rs1802295	10 (70601480)	*VPS26A*	G / A	0.26	0.31	15,506 *	1.14 (1.08-1.20)	1.9×10^{-6}	1.06 (1.03-1.10)	6.6×10^{-4}	1.08 (1.05-1.12)	4.1×10^{-8}	1.04 (1.00-1.09)	6.0×10^{-2}	0.50	1.07 (1.05-1.10)	2.1×10^{-8}
rs7178572	15 (75534245)	*HMG20A*	A / G	0.52	0.71	18,193	1.10 (1.05-1.15)	2.4×10^{-5}	1.08 (1.05-1.12)	7.0×10^{-7}	1.09 (1.06-1.12)	7.1×10^{-11}	1.07 (1.02-1.12)	2.6×10^{-3}	0.59	1.08 (1.06-1.11)	9.2×10^{-13}
rs2028299	15 (88175261)	*AP3S2*	A / C	0.31	0.31	18,076	1.11 (1.05-1.16)	4.8×10^{-5}	1.09 (1.06-1.13)	1.1×10^{-7}	1.10 (1.07-1.13)	1.9×10^{-11}	1.05 (1.00-1.09)	4.0×10^{-2}	0.12	1.08 (1.06-1.11)	1.2×10^{-11}
rs4812829	20 (42422681)	*HNF4A*	G / A	0.29	0.19	18,186	1.14 (1.08-1.19)	4.5×10^{-7}	1.07 (1.04-1.11)	2.8×10^{-5}	1.09 (1.06-1.12)	2.6×10^{-10}	1.08 (1.02-1.14)	1.0×10^{-2}	0.90	1.09 (1.06-1.12)	8.2×10^{-12}

Alleles: R=risk, A=alternate. RAF: risk allele frequency in South Asians (SA) and Europeans (EW). Results for Europeans are from the GWA stage of DIAGRAM+. P_{hetero} is for the comparison between the odds ratios for T2D amongst South Asians in the replication sample, and Europeans in DIAGRAM+.

* rs1802295 not available on the Illumina 317K array.