PLOS ONE

Joint Effects of Known Type 2 Diabetes Susceptibility Loci in Genome-Wide Association Study of Singapore Chinese: The Singapore Chinese Health Study

Zhanghua Chen[1]*, Mark A. Pereira[2], Mark Seielstad[3], Woon-Puay Koh[4,5], E. Shyong Tai[4,5], Yik-Ying Teo[5], Jianjun Liu[5,6], Chris Hsu[1], Renwei Wang[7], Andrew O. Odegaard[2], Bharat Thyagarajan[2], Revati Koratkar[2], Jian-Min Yuan[7], Myron D. Gross[2]*, Daniel O. Stram[1]*

1 Department of Preventive Medicine, University of Southern California, Los Angeles, California, United States of America, 2 Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, Minnesota, United States of America, 3 Department of Laboratory Medicine, Department of Epidemiology and Biostatistics, and California Institute for Quantitative Biosciences (QB3), University of California San Francisco, San Francisco, California, United States of America, 4 Duke-National University of Singapore Graduate Medical School, Singapore, 5 Saw Swee Hock School of Public Health, National University of Singapore, Singapore, 6 Human Genetics, Genome Institute of Singapore, A*STAR, Singapore, 7 Department of Epidemiology, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania, United States of America

Abstract

Background: Genome-wide association studies (GWAS) have identified genetic factors in type 2 diabetes (T2D), mostly among individuals of European ancestry. We tested whether previously identified T2D-associated single nucleotide polymorphisms (SNPs) replicate and whether SNPs in regions near known T2D SNPs were associated with T2D within the Singapore Chinese Health Study.

Methods: 2338 cases and 2339 T2D controls from the Singapore Chinese Health Study were genotyped for 507,509 SNPs. Imputation extended the genotyped SNPs to 7,514,461 with high estimated certainty ($r^2 > 0.8$). Replication of known index SNP associations in T2D was attempted. Risk scores were computed as the sum of index risk alleles. SNPs in regions ±100 kb around each index were tested for associations with T2D in conditional fine-mapping analysis.

Results: Of 69 index SNPs, 20 were genotyped directly and genotypes at 35 others were well imputed. Among the 55 SNPs with data, disease associations were replicated (at $p < 0.05$) for 15 SNPs, while 32 more were directionally consistent with previous reports. Risk score was a significant predictor with a 2.03 fold higher risk CI (1.69–2.44) of T2D comparing the highest to lowest quintile of risk allele burden ($p = 5.72 \times 10^{-14}$). Two improved SNPs around index rs10923931 and 5 new candidate SNPs around indices rs10965250 and rs1111875 passed simple Bonferroni corrections for significance in conditional analysis. Nonetheless, only a small fraction (2.3% on the disease liability scale) of T2D burden in Singapore is explained by these SNPs.

Conclusions: While diabetes risk in Singapore Chinese involves genetic variants, most disease risk remains unexplained. Further genetic work is ongoing in the Singapore Chinese population to identify unique common variants not already seen in earlier studies. However rapid increases in T2D risk have occurred in recent decades in this population, indicating that dynamic environmental influences and possibly gene by environment interactions complicate the genetic architecture of this disease.

Introduction

T2D remains a very serious health threat in developed countries and is becoming a major health threat in many under-developed countries, particularly those with rapidly growing economies [1–3]. Globally, T2D affected over 360 million people in 2011 [4] and

this number is projected to increase rapidly in upcoming years. This rise in risk is paralleled by a rapidly increasing incidence of obesity in many populations, a major risk factor for diabetes. In addition, incidence may be propelled by an elevated genetic susceptibility in some populations. Other risk factors include

dietary patterns [5,6], sedentary lifestyle [7,8], psychosocial stress [9–11], short sleep hours [12], and smoking [13–16].

Interestingly, the prevalence of T2D is much higher (approximately 2-fold) in several Southeast or East Asian populations than in populations of European-descent, even though most Asians have a much lower average body mass index (BMI) and rates of obesity [17–20]. The prevalence of T2D continues growing rapidly in many Southeast Asian countries, including Singapore [21]. Compared to populations of European ancestry, East Asians, including Chinese and Japanese have been characterized as having a higher proportion of abdominal and visceral fat deposits in the presence of a BMI$\leq$25 kg/m^2 [22,23], which is considered a healthy BMI in populations of European descent. Also, diabetes incidence in young to middle-aged people is disproportionately higher in Southeast Asia than in the West [21]. This apparent difference in susceptibility is recognized by the International Federation of Diabetes, which has established lower BMI cutoffs for overweight and obesity than are used for populations of European-descent [24]. The apparently higher susceptibility persists in individuals migrating from Southeast Asia to other parts of the world and results in even higher levels of diabetes in these populations when living in Western cultures [25–28].

It is well-known that T2D is heritable in many populations [29–31] and has a familial recurrence risk ratio for first degree relatives of approximately two [32,33]. In addition, numerous studies have associated specific genetic variants with the risk of T2D. Several notable associations were identified by linkage analysis and candidate gene studies, and include *PPARγ* [34], *KCNJ11* [35], *WFS1* [36] and *TCF7L2* [37]. The advent of large-scale genetic studies searching the entire genome for common SNPs (frequency>5%) associated with diabetes has significantly increased the number of SNPs associated with diabetes. Since 2007, genome-wide association studies (GWAS) have reported at least 57 additional thoroughly replicated genetic susceptibility loci harboring common variants for T2D [38–57]. Most of these were novel disease loci and contributed to a better understanding of diabetes heritability. However, the effect sizes of these loci were small and only a small proportion of the heritability of T2D was explained [58]. Moreover, most of the SNP associations discovered by GWAS were identified in European populations. However, Asian-specific SNPs have been identified and several loci were first identified by GWAS in Asians including *KCNQ1* [40,59], *UBE2E2* and *C2CD4A-C2CD4B* [48].

We investigated the reproducibility of single SNP associations in a study of T2D among Singapore Chinese using both genotyped and imputed alleles. Beyond investigating associations between single variants and disease risk, it is important to consider the combined effects of various loci on disease risk. In this report, we used the National Human Genome Research Institute (NHGRI) GWAS Catalog [60] to identify 59 single-nucleotide polymorphisms (SNPs) in 46 gene regions that have been associated with T2D. In addition we interrogated regions near GWAS alleles to search for additional or refined associations.

Research Design and Methods

Ethics Statement

This study has been approved by the institutional review boards of the National University of Singapore, the University of Southern California, the University of Minnesota, and the University of Pittsburgh. Informed written consent to participate in biomarker studies was obtained at time of specimen collection. The institutional review boards approved this consent procedure.

Study Population

People of Chinese ancestry comprise the largest ethnic group in Singapore and constitute 74.1% of Singapore's resident population [61]. The design of the Singapore Chinese Health Study has been previously described [62]. Briefly the cohort is drawn from permanent residents or citizens of Singapore aged 45–75 at study entry, who reside in government-built housing estates (~86% of Singapore residents live in such facilities). Migration out of Singapore, especially among housing estates residents is negligible (Department of Statistic, Singapore Ministry of Trade and Industry, 1997). The study subjects are restricted to the two major dialect groups of Chinese in Singapore: The Hokkiens, who originated from southern Fujian Province, and the Cantonese, who came from Guangdong Province (Both provinces are in south eastern China. The gender dialect breakdown of the cohort is as follows, 15,617 (24.7%) Hokkien men, 18,356 (29.0%) Hokkien women, 12,342 (19.5%) Cantonese men, and 16,942 (26.8%) Cantonese women.

Between April 1993 and December 1998, 63,257 individuals completed an in-person interview that included questions on usual diet, demographics, height and weight, use of tobacco, usual physical activity, menstrual and reproductive history (women only), medical history, and family history of cancer. A follow-up telephone interview took place between 1999 and 2004 for 52,325 cohort members (83% of recruited cohort). Beginning in April 1994, a random 3% sample of cohort participants were asked to provide blood or buccal cells, and spot urine samples. Eligibility for this biospecimen subcohort was extended to all surviving cohort participants starting in January 2000. By April 2005, all surviving cohort subjects had been contacted for biospecimen donation. Samples were obtained from 32,535 subjects, representing a consent rate of about 60%. The institutional review boards at the National University of Singapore, the University of Minnesota, and the University of Pittsburgh approved this study.

Utilizing resources of the Singapore Chinese Health Study, we conducted a genome-wide association study (GWAS) for the risk of developing diabetes that has a two staged design in which approximately 1/2 of all participants in the study are genotyped using a GWAS array with the remaining subjects genotyped as a replication study of the top SNPs found in stage 1. This approach follows the general principles of Satagopan et al [63] and Wang et al [64]. Herein we report results from the first stage of this study focusing on replication and fine-mapping of already-discovered genetic variants.

Ascertainment of Type 2 Diabetes

For each study participant, the history of physician-diagnosed diabetes was asked at a baseline interview administered by a trained interviewer. Diabetes status was assessed again by the following question asked during the first and second follow-up telephone interviews: "Have you been told by a doctor that you have diabetes (high blood sugar)?" If yes: "Please also tell me the age at which you were first diagnosed". The prevalent diabetes cases were those who reported a history of diabetes at the baseline interview whereas the incident diabetes were those reporting the initial diagnosis of diabetes that took place after the baseline interview in either the follow-up I or follow-up II interview (~5.5 years between interviews). A validation study of the incident diabetes mellitus cases used two different methods and was reported in detail previously [65,66]. Based on a hospital-based discharge summary database and a supplementary questionnaire regarding symptoms, diagnostic tests and hyperglycemic therapy during a telephone interview we observed a positive predictive value of 99% [66]. In other words, the self-reported history of

diabetes was a highly reliable measure of diabetes status of the study population.

Eligible Study Subjects

The cohort participants who did not report a history of diabetes at baseline interview and donated blood samples were eligible for the present study. We excluded subjects with prevalent diabetes at the baseline interview (n = 2,080) or did not provide blood samples (n = 36,245). The present study was based on the remaining 24,932 subjects. Among them, we identified 1,284 incident diabetes cases during the follow-up I interview in 2000–2005, and an additional 1,343 incident diabetes cases during the follow-up II interview in 2006–2011. For each incident diabetes case, one control subject was randomly selected among the subjects that provided blood samples but did not have a history of diabetes. Controls were matched to the index cases on gender, dialect group (Cantonese or Hokkien), age at baseline interview (±3 years), year of baseline interview (±2 years), and date of blood draw (±6 months). In addition, the selected controls were screened for the presence of undiagnosed T2D. The criterion for undiagnosed diabetes was the hemoglobin A1c (HbA1c)≥6.0%. All matched controls with HbA1c≥6.0% were ineligible for the study and a replacement control with the same matching criteria was randomly chosen among the remaining eligible subjects. Blood for HbA_{1c} analysis was collected in EDTA (ethylenediaminetet-raacetic acid) tubes. Red-blood cells (RBCs) were isolated from whole blood and frozen until analysis was performed at University of Minnesota, a Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory. HbA1c was measured with a dedicated HPLC instrument in our laboratory which serves as a reference laboratory for this assay. The instrument, a TOSOH HPLC, utilizes ion-exchange chromatography (Tosoh A1c 2.2 Plus HPLC, Tosoh Medics, Inc., Foster City, CA). This instrumentation is also referred to as the Tosoh G7/G8 HPLC Glycohemoglobin Analyzer (Tosoh Medics, Inc., San Francisco, California). A small red blood cell sample was automatically hemolyzed prior to injection onto the column. The labile fraction is separated on-line as a distinct peak and excluded from the calculation of % HbA1c. The hemoglobin fractions (A1a, A1b, F, Labile A1c, Stable A1c, A0 and Hb variants) are separated by a buffer gradient of increasing ionic strength. The Tosoh 2.2+ was calibrated daily using 2 calibrators (2-point calibration) standardized to a reference system and the percentage of HbA1c was calculated based on this system. Using the standards developed in the National Glycohemoglobin Standardization Program, this method was calibrated to the reference range of 4.3%–6.0% and had a laboratory coefficient of variation range 1.4%–1.9% [67].

Genotype Analysis and Quality Control

Peripheral blood samples from 2615 incident diabetes cases and 2615 matched controls were selected for DNA extraction in stage 1. The DNA extraction was conducted at the Molecular Epidemiology and Biomarker Research Laboratory at the University of Minnesota (approximately 2/3rds of the samples) or the Genome Institute of Singapore (approximately 1/3rd of the samples) using the Qiagen method. DNA concentrations were measured by the PicoGreen and Nanodrop methods and prepared for genotype analysis.

Stage 1 genotyping was performed at the Genome Institute of Singapore according to the manufacturer's recommendations using an Affymetrix ASI (Asian) Axiom array. Genotype calling was performed by the Affymetrix Corporation. A standard series of QC steps were followed in order to identify SNPs in case and control samples for genetic association analyses. Starting with

510,584 SNPs provided for 4,918 callable study samples, we excluded samples with SNP call rates of less than 98 percent (n = 22) and SNPs (n = 3,075) with call rates less than 98 percent, leaving 507,509 SNPs.

We estimated relatedness between pairs of samples as the expected number of alleles shared identically by descent, r_{ij}, using PLINK [68]. We dropped two pairs of unintended duplicate samples that were discovered to have r_{ij} close to one; we also dropped samples that appeared to be closely related (r_{ij}>.2) to more than one other sample in the study and one of each remaining pair of samples with r_{ij}>.2 (n = 180 total including the duplicates). We compared reported sex of each sample to sex as inferred on the basis of X chromosome heterozygosity, dropping 29 uncertain or conflicting samples. We computed principal components of the genotype matrix and dropped 9 individuals who were more than 5 standard deviations from the mean on any of the first 4 principal components. One additional sample was dropped because of missing covariate information. A total of 4,677 samples (2338 cases and 2339 controls) remained after QC analysis.

Statistical Analysis

Characteristics of the cohort were compared between diabetes cases and controls. Two sample t-tests were used to compare the mean differences for variables with normal distributions. The Wilcoxon rank sum test was used to compare median differences for variables with skewed distributions. Pearson χ^2 was used to test if the frequency distributions for categorical variables were different between diabetes cases and controls.

For genotype imputation, we first mapped genetic positions of our GWAS data to NCBI build37 using UCSC Genome Browser liftOver [69]. 14,032 (<3%) SNPs failed to be mapped to NCBI build37. The Segmented Haplotype Estimation and Imputation tool (SHAPEIT) [70] was then used to phase the remaining 493,477 SNPs. We applied 1000 Genomes Project Phase I data "version 3" [71] as the reference panel, which contained 1092 individuals of various ethnicities (246 Africans, 181 African Americans, 286 East Asians and 379 Europeans) with 36,648,992 SNPs. IMPUTE2 [72] was run to perform the imputation, which extended our total SNPs to be 36,617,842. After filtering out SNPs imputed to be monomorphic or with estimated r^2<0.8, there were 7,514,461 imputed or genotyped SNPs for association analysis.

For this report, we selected 83 SNPs associated with T2D summarized by NHGRI GWAS Catalog [60] and significantly associated with diabetes risk at a well-recognized criteria for genome-wide significance ($p \leq 5 \times 10^{-8}$). Among these, one SNP was neither genotyped nor imputed in our data, 12 SNPs were poorly imputed with estimated certainty r^2<0.8, and one genotyped SNP had rare minor allele frequency (MAF) less than 0.008. Additionally, 14 of the GWAS SNPs were found to be in LD with 11 other GWAS SNPs with estimated pairwise r^2>0.75 using our genotyped and imputed data. After excluding these 28 SNPs, the logistic regression method was used to analyze the single SNP associations of the remaining 55 GWAS-implicated SNPs with diabetes case-control status after adjusting for age, sex, dialect, and first 10 principal components. The logistic regressions utilized the observed genotyped or expected imputed allele counts as the explanatory variable of interest.

The 55 SNPs from the GWAS catalog are called "index SNPs" in the fine-mapping analysis. Among these 55 SNPs, one SNP had no reported risk allele in the GWAS-catalog and the original papers. Thus, it was not included in the genetic risk score analysis described below. Power calculations were conducted using Quanto

[73] for the 55 SNPs based on the risk allele frequencies in our 4,677 study subjects using a significance level of 0.05 and the odds ratio reported by the GWAS-catalog.

After single SNP association analysis, we constructed genetic risk scores based on genotyped only, imputed only, and both genotyped and imputed known diabetes SNPs combined by adding the observed or expected number of risk alleles for each study participant according to the risk allele reported in GWAS Catalog. The association between the genetic risk score and diabetes mellitus status was assessed using logistic regression adjusting for the same covariates as in the previous single SNP association analysis.

For fine-mapping analysis, regions 100 kb up and down stream of each index SNP were obtained from the combination of genotyped and imputed data. As before, logistic regression was used to test significant associations between the observed or expected allele counts (log additive model) for each SNP and disease status. Additionally, conditional analysis was performed for each SNP in a GWAS-indicated region by adjusting for the index SNP of that region in addition to the other covariates. Such conditional analysis attempts to refine SNP associations and search for stronger signals than index SNPs. Bonferroni adjustment was used to set the significance level for SNP association tests as 0.05/ number of SNPs in each region. Based on fine-mapping results and following the approach of Chen et al [74], we attempted to define two types of SNPs: 1. "Improved SNPs", i.e. SNPs in LD (in the original populations) with index signals ($r^2 \geq 0.5$) but with stronger results in the present GWAS than the index signals; 2. "New SNPs" i.e. SNPs with significant associations, but which were not in close LD with index signals ($r^2 < 0.5$) that may reflect new associations in regions already known to be involved in disease risk. Next, fine-mapping results were used to improve the genetic score by substituting index SNPs with improved SNPs and adding new SNPs into the score. The risk allele and effect sizes of improved SNPs as well as new SNPs were defined based on our fine-mapping results.

Finally, Genome-wide Complex Trait Analysis (GCTA) was performed to estimate the proportion of disease variance (using a liability model) that is explained by GWAS reported diabetes SNPs as well as any newly identified SNPs from the fine-mapping analysis [75].

Results

Characteristics of subjects in this study are presented in Table 1. The mean age and distributions of female gender, dialect group and smoking status or duration of smoking in cases were comparable with those in controls. Compared to controls, cases had a higher BMI ($p < 0.0001$) and lower level of education ($p = 0.004$). More controls had weekly engagement of physical activities than cases ($p = 0.043$).

Among 55 potentially diabetes-related SNPs identified from the NHLBI GWAS Catalog, 20 SNPs were genotyped and 35 SNPs were imputed with $r^2 > 0.8$. Here r^2 was estimated as sample variance (over all individuals in the study) of the expected allele count (i.e. the imputed values) divided by the theoretical value, $2p(1-p)$, of the variance of the count for a SNP in Hardy Weinberg equilibrium where p is the estimated frequency of the allele [76]. Based upon the risk allele frequency seen in our sample from the Singapore Chinese Health Study and on the reported odds ratio and risk allele from the GWAS catalog (for 54 SNPs with this information available) we had an average of 62.8 percent power to replicate true associations at a 5 percent significance level. Of the 54 SNPs with known risk alleles we found that 15 (27.8%) had

significant associations ($p < 0.05$) in the same direction as those reported with diabetes risk after adjusting for age, sex, dialect and 10 principal components (Table S1). Among the remaining 39 non-significant associations a total of 32 (82.1%) of the associations indicated that the same allele was associated with increased risk as listed in the GWAS catalog. Quantile-quantile (QQ) plots (Figure 1) of the p-values for association of the 55 SNPs showed considerable deviation from the distribution expected under the null hypothesis, further indicating that these 55 index SNPs included strong signals for diabetes risk in the Singapore Chinese population.

Non-replication of known or putative disease SNPs may be a result of differing LD patterns in Singapore Chinese relative to the original GWAS populations so that index SNPs might not be sufficiently correlated with the underlying biological causal variant in Singapore Chinese. In order to try to identify better genetic markers of risk in Singapore Chinese, we conducted fine-mapping analysis across all risk regions (± 100 kb of index SNP), using genotyped SNPs on the Affymetrix array and imputed SNPs seen in the 1000 Genomes data (see Methods).

We searched for improved candidate SNPs from among those 1000 genome SNPs that were found to be in high LD ($r^2 \geq 0.5$) in the original GWAS population, as well as for novel SNPs not highly correlated with the index within the reported regions. After applying a Bonferroni correction for the number of SNPs tested in each region, we found two improved signals (both Bonferroni adjusted p-values = 0.033, Figure 2) for rs2453051 and rs2493413 having $r^2 = 1$ (in Europeans based on 1000 genomes pilot data) with index SNP rs10923931. The two improved signals and the index SNP were located in the *NOTCH2* gene on chromosome 1. Additionally, we found five novel independent associations in 2 regions. Four correlated (pairwise $r^2 > 0.97$) novel SNPs (rs10757282, rs7019778, rs10757283, and rs7019437) were found around index rs10965250 (Bonferroni adjusted p-value < 0.044 for all, Figure 2). These SNPs were on chromosome 9 and near the *N2B-AS1* gene. SNP rs10757282 had the most significant association (Bonferroni adjusted p-value = 0.028). Another three significant associations (rs11187139, rs10882102 and rs78216286) were found around index rs1111875 (Bonferroni adjusted p-value < 0.040 for all, Figure 2), however two of these SNPs rs11187139 and rs10882102 were on closer inspection found to be correlated with another nearby index SNP rs5015480 ($r^2 > 0.84$), thus are not included in further analysis. The remaining SNP rs78216286 was on chromosome 10 near the *KIF11* gene. SNP rs78216286 is included in the following risk score analysis. These novel signals may indicate additional causal variants unidentified in the original GWAS.

The cumulative effect of all T2D risk variants was tested using unweighted counts of all diabetes risk SNPs. We did association analysis using a risk score comprised of four sets of risk alleles: 1) 19 genotyped SNPs; 2) 35 imputed SNPs; 3) 54 SNPs (genotyped and imputed); 4) original 54 SNPs with rs10923931 replaced by rs2453051, and including 2 new independent SNPs identified from fine-mapping analysis (rs10757282, and rs78216286) (Table 2). Using the 54 index SNPs from the GWAS catalog, the risk per allele was 1.049 (95% confidence interval (CI) 1.036–1.062; $p = 2.93 \times 10^{-14}$). Individuals in the highest quintile of the risk allele distribution were at 2.0-fold greater risk ($p = 5.72 \times 10^{-14}$) of T2D compared to individuals in the lowest quintile (Table 2). In single SNP analysis for the genotyped SNPs the mean odds ratio in the Singapore data was 1.100 while for the imputed SNPs the mean odds ratio was 1.058. In the risk score using genotyped SNPs the estimated OR per allele was 1.073 (1.049–1.097; $p = 4.30 \times 10^{-10}$). For the risk score with only imputed SNPs the

Table 1. Characteristics of subjects comparing T2D cases and controls.

	Cases (n = 2338)	Controls (n = 2339)	p
Anthropometrics			
Age [yrs]*	55.3 (7.2)	55.2 (7.1)	0.575
Females‡	1227 (52.5)	1247 (53.3)	0.568
BMI [kg/m²]*	24.6 (3.4)	22.7 (3.1)	<0.0001
Physical activity weekly (%)‡	32.7	35.5	0.043
Food energy intake [kcal/day]*	1619 (589)	1578 (547)	0.014
Smoking History			
non-smokers‡§	1579 (69.3)	1643 (71.9)	
Former smokers‡§	285 (12.5)	258 (11.3)	0.143
Current smokers‡§	415 (18.2)	383 (16.8)	
Years of smoking (among current and former smokers) † §	25 (15, 35)	25 (15, 35)	0.808
Education levels			
No education‡	543 (23.2)	483 (20.7)	
Primary school education‡	1074 (45.9)	1033 (44.2)	0.004
Secondary school education or more‡	721 (30.8)	823 (35.2)	
Dialect group			
Cantonese‡	1166 (49.9)	1165 (49.8)	0.965
Hokkien‡	1172 (50.1)	1174 (50.2)	

*Variables are presented as mean (standard deviation). Two-sample independent t-test is used to test the mean differences between cases and controls.
†Variables are not normally distributed and are presented as median (25th, 75th percentiles). Wilcoxon rank-sum test is used to compare the median differences between cases and controls.
‡Categorical variables are presented as frequencies (%).χ^2 test is used to test whether the distribution between cases and controls is different.
§Variables have 59 cases and 55 controls missing.
doi:10.1371/journal.pone.0087762.t001

odds ratio per allele was OR = 1.048 (95% CI: 1.031–1.065; p = 3.19×10^{-8}). When the three new or improved SNPs were included in the risk score the association with T2D was slightly strengthened (per allele OR = 1.053; 95% CI 1.040–1.066; p = 6.68×10^{-16}). Compared to individuals in the lowest quintile
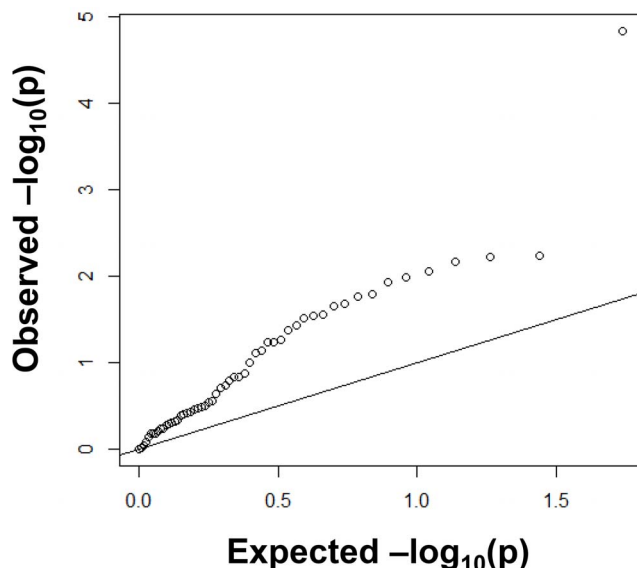

Figure 1. QQ-plot of 55 T2D SNPs reported by GWAS-catalog. Observed distribution of −log P-values were compared to the expected (null) distribution.
doi:10.1371/journal.pone.0087762.g001

of this risk score, those in the highest quintile had a 2.1 times greater risk of the disease (p = 2.09×10^{-16}). Interestingly we noted no evidence that the per allele odds ratios were different depending upon whether the index SNP was reported in GWAS of either a European or Asian population (mean OR in the Singapore sample was 1.063 for the 19 SNPs reported from GWAS in Asian populations versus 1.078 for the 35 SNPs reported from GWAS in European populations, Table S1).

Finally, we estimated the proportion of variance of diabetes risk (on the liability scale) explained by these SNPs using the GCTA program [77]. We assumed the prevalence of diabetes among the population to be 0.08 based on International Diabetes Federation report [4] and found that the 55 GWAS-reported diabetes SNPs explained 2.3% of disease liability variance after adjusting for age, sex, dialect and first 10 principal components (p = 0.007). After adding two novel SNPs from our fine-mapping analysis, the entire 57 SNPs were again estimated to explain 2.3% variance of the liability of diabetes in the sample (p = 0.007).

Discussion

Replication and fine mapping of GWAS index disease associations in additional populations is useful for defining the relevancy of associations discovered in one population to other ethnic groups. In addition, studies of ethnically diverse groups contribute to the localization of associations and the discovery of new disease risk alleles in previously identified regions [74].

We were able to replicate disease associations (p<0.05) for 15 of 54 SNPs considered validated by prior studies. Of the 39 SNPs that were not replicated at p<0.05, the average power based on the GWAS-reported OR and Singapore Chinese risk allele
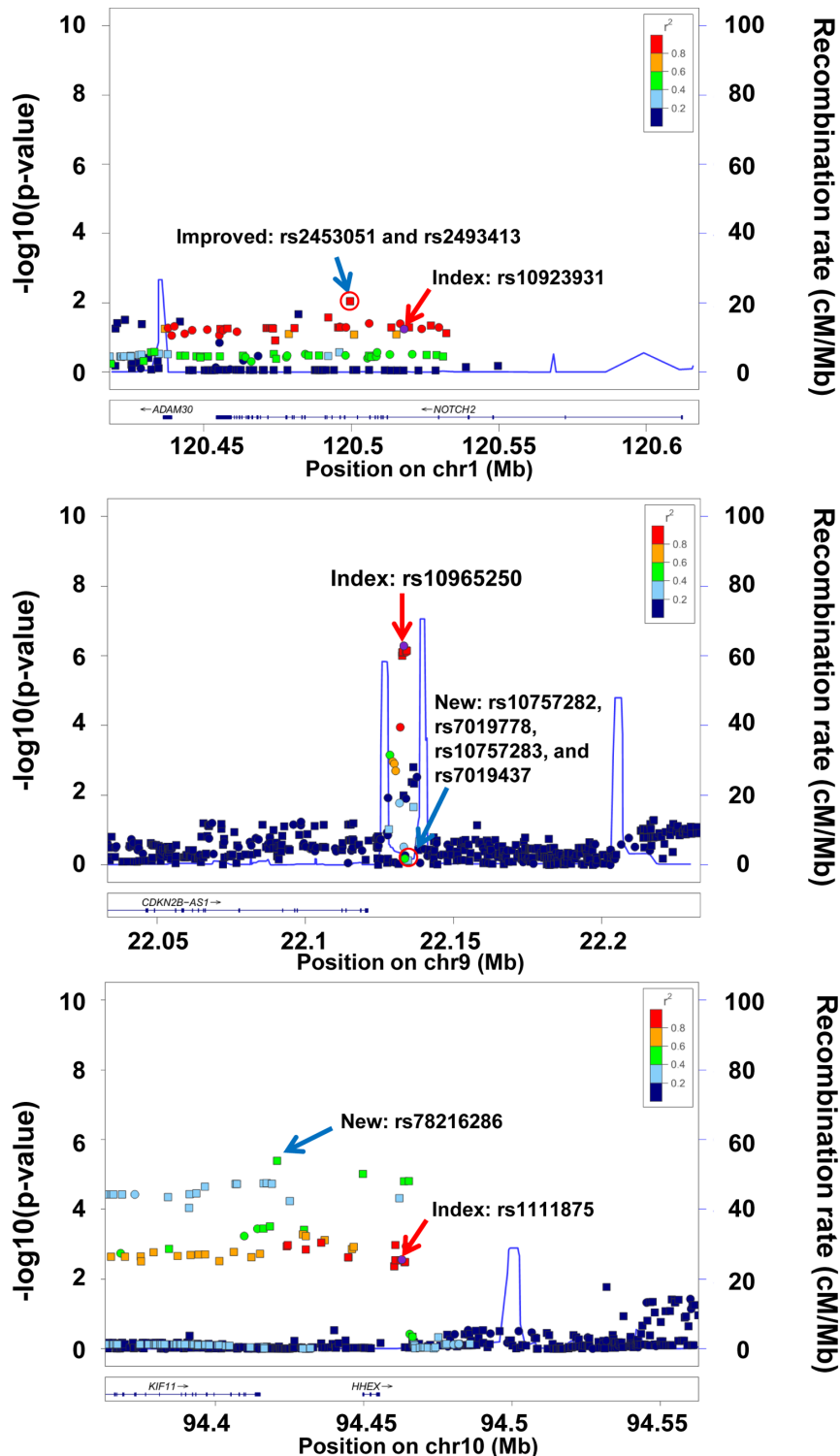
Figure 2. −log *P*-plots of the improved and novel T2D signals around index SNPs: rs10923931, rs10965250 and rs1111875 found by fine-mapping analysis. −Log *P*-value for risk-associated allele from the logistic regression model adjusted for age, sex, dialect and global ancestry (the first 10 principal components). Pairwise correlations (r²) in the 1000 Genomes Asian population are shown in relation to markers identified through fine-mapping in our sample. Squares denote genotyped SNPs; circles, imputed SNPs. Gray squares and circles denote that r² cannot be estimated (not in 1000 Genomes). Red arrows and diamond denote the index SNP. Blue arrows denote the novel signal. The plots were generated using LocusZoom [86].

doi:10.1371/journal.pone.0087762.g002

Table 2. Summary risk scores in association with T2D.

		2,338 cases, 2,339 controls
Summary score of genotyped index markers (19 markers)		
Mean risk score of cases (range)/controls (range)		14.8 (7.0–25.0)/14.3 (6.0–25.0)
Per allele OR		1.073 (1.049–1.097)
P_{trend}		4.301×10^{-10}
Quintile OR (95% CI)	Q1	1.00 (ref)
	Q2	1.26 (1.03–1.53)
	Q3	1.40 (1.19–1.65)
	Q4	1.28 (1.04–1.56)
	Q5	1.68 (1.41–1.99)
Summary score of well imputed index markers (35 markers)		
Mean risk score of cases (range)/controls (range)		37.1 (24.3–48.8)/36.5 (21.5–47.8)
Per allele OR		1.048 (1.031–1.065)
P_{trend}		3.193×10^{-8}
Quintile OR (95% CI)	Q1	1.00 (ref)
	Q2	1.19 (0.99–1.43)
	Q3	1.36 (1.14–1.64)
	Q4	1.32 (1.10–1.59)
	Q5	1.66 (1.38–1.99)
Summary score of all imputed index markers (44 markers)		
Mean risk score of cases (range)/controls (range)		49.4 (36.3–61.7)/48.7 (33.9–60.9)
Per allele OR		1.048 (1.032–1.064)
P_{trend}		1.172×10^{-9}
Quintile OR (95% CI)	Q1	1.00 (ref)
	Q2	1.04 (0.87–1.25)
	Q3	1.27 (1.06–1.52)
	Q4	1.37 (1.14–1.64)
	Q5	1.57 (1.31–1.89)
Summary score of genotyped and well imputed index markers (54 markers)		
Mean risk score of cases (range)/controls (range)		51.9 (36.8–66.9)/50.8 (33.5–70.5)
Per allele OR		1.049 (1.036–1.062)
P_{trend}		2.931×10^{-14}
Quintile OR (95% CI)	Q1	1.00 (ref)
	Q2	1.39 (1.16–1.67)
	Q3	1.34 (1.12–1.61)
	Q4	1.57 (1.31–1.89)
	Q5	2.03 (1.69–2.44)
Summary score of all index markers (63 markers)		
Mean risk score of cases (range)/controls (range)		64.2 (47.3–80.4)/63.0 (45.9–80.1)
Per allele OR		1.050 (1.037–1.062)
P_{trend}		1.147×10^{-15}
Quintile OR (95% CI)	Q1	1.00 (ref)
	Q2	1.26 (1.05–1.52)
	Q3	1.23 (1.02–1.48)
	Q4	1.57 (1.31–1.89)
	Q5	2.03 (1.69–2.44)
Summary score of genotyped, well imputed index markers and 2 novel marker (56 markers)		
Mean risk score of cases (range)/controls (range)		53.3 (38.3–68.2)/52.2 (36.1–70.5)

Table 2. Cont.

Summary score of genotyped, well imputed index markers and 2 novel marker (56 markers)		
Per allele OR		1.053 (1.040–1.066)
P_{trend}		6.679×10^{-16}
Quintile OR (95% CI)	Q1	1.00 (ref)
	Q2	1.27 (1.06–1.52)
	Q3	1.25 (1.04–1.50)
	Q4	1.60 (1.33–1.92)
	Q5	2.10 (1.74–2.52)
Summary score of all index markers and 2 novel marker (65 markers)		
Mean risk score of cases (range)/controls (range)		65.6 (49.9–81.4)/64.4 (48.9–80.1)
Per allele OR		1.053 (1.040–1.065)
P_{trend}		2.684×10^{-17}
Quintile OR (95% CI)	Q1	1.00 (ref)
	Q2	1.30 (1.09–1.57)
	Q3	1.36 (1.13–1.63)
	Q4	1.58 (1.31–1.90)
	Q5	2.18 (1.81–2.62)

doi:10.1371/journal.pone.0087762.t002

frequency was 58.5 percent (Table S1, Figure S1) compared to 73.8 percent for the replicated SNPs. Our failure to replicate more known associations, despite reasonable power to do so, may be due to several reasons; it is possible that the odds ratios estimated for the reported risk alleles were biased upwards by a "winner's curse" phenomenon [78] thus causing an overestimation of statistical power for replication. Our risk score analysis using the sum of all 54 (both genotyped and well imputed) GWAS-significant risk alleles as a predictor of T2D risk in the Singapore Chinese Health Study population, while highly significant statistically ($p < 10^{-13}$), showed per-allele ORs that are smaller on average (1.05) than the mean (1.16) of the published ORs for these alleles or of the mean (1.07) of the single SNP ORs estimated in this study. This appears to be indicating either a sub-multiplicative effect of the SNPs in aggregate and/or reflecting a slight negative correlation ($r = -0.25$) between risk allele frequency and OR evident in Table S1. Additionally including nine poorly imputed SNPs into the risk score did not significantly influence previous results (per-allele ORs = 1.05, 95% CI: 1.04–1.06, p = 1.147×10^{-16}, Table S1).

The attenuation of effect between the reported ORs and the ORs estimated here may also be due to differences in LD between the initial GWAS populations and the Singapore Chinese so that the correlation between index SNP and underlying causal variant is lower. In fine mapping analysis we found an improved signal for two SNPs (rs2453051 and rs2493413) that were in high LD with the index SNP rs10923931 in the original (European) reporting population but not in our study ($r^2 = .426$). We also found five novel candidate SNPs (rs10757282, rs7019778, rs10757283, rs7019437, rs78216286) near two index SNPs, rs10965250 and rs1111875, which passed our criteria for significance but were not among the ones in LD with the original index SNPs in the reporting populations; SNP rs10965250 was reported in European population [47], and rs1111875 was reported in both European and Japanese populations [38,79–81]. While these results may be novel associations, i.e. new signals in a region already implicated in GWAS studies, further replication (as in stage 2 of this two-stage

GWAS study) will be needed before these will be well-accepted risk alleles.

It appears that our efforts to impute ungenotyped SNPs implemented by the programs SHAPEIT [70] and IMPUTE2 [72] were largely successful; as shown in the Results section we were able to impute with a high degree of estimated certainty for the large majority of ungenotyped risk alleles. We do note that the fraction of replicated risk alleles among imputed SNPs (6 of 35, 17.1%, Table S1) were smaller compared to the directly genotyped ones (9 of 19, 47.4%). This is partly explained by allele frequency and odds ratio differences which lead to somewhat decreased power (60.6% versus vs. 66.7%, Table S1) for imputed and genotyped SNPs respectively. In addition imputation involves some loss of power, governed by the r^2 between the imputed and true genotypes [82]. Nevertheless the score involving only imputed SNPs was a highly significant predictor of diabetes risk (p = 3.19×10^{-8}).

More generally our findings indicate that only a very small fraction of T2D in Singapore Chinese can be explained by the SNPs in the risk regions examined to date. The rapid increase in T2D in Singapore and in other Asian and South East Asian communities [4,21,22] strongly indicate environmental factors are at play, yet susceptibility to these factors (notably BMI) appears to differ greatly by racial/ethnic group [83–85]. Understanding the interplay between genes and lifestyle-related risk factors that could produce such notable racial/ethnic disparities would seem to be among the most important needs in diabetes epidemiology. A separate report on genetic interactions between individual risk SNPs, genetic scores, and lifestyle or other "environmental" variables is under development using these data. It is clear also that very large sample sizes are needed to establish new T2D risk alleles since it is evident that each one plays a small role by itself even when strongly significantly predictive in composite (as in our risk score analyses). Our ability to extend through imputation the set of SNPs used in the present study (based on the Affymetrix Axiom ASI array) to over 7 million SNPs with good reliability and demonstrated predictive ability means that this study can

contribute to the very large scale highly collaborative studies that may be needed to make further progress in understanding the genetics of T2D. Alternatively, significant differences may exist between ethnic groups, such that, the effect size of specific SNPs may differ between the ethnic groups as a result of differences due to early development and/or environment. In addition, the identification of less common SNPs ($<5\%$) may be important and studies of T2D in ethnic groups would benefit from sequencing studies.

Supporting Information

Figure S1 Observed $-\log P$ compared to the corresponding power for each of the 54 reported T2D SNPs. The reference solid line indicates observed $P = 0.05$.
(TIF)

References

1. Wild S, Roglic G, Green A, Sicree R, King H (2004) Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care 27: 1047–1053.
2. Yoon KH, Lee JH, Kim JW, Cho JH, Choi YH, et al. (2006) Epidemic obesity and type 2 diabetes in Asia. Lancet 368: 1681–1688.
3. Hossain P, Kawar B, El Nahas M (2007) Obesity and diabetes in the developing world–a growing challenge. N Engl J Med 356: 213–215.
4. International Diabetes Federation website. Available: http://www.idf.org/global-diabetes-plan-2011-2021. Accessed 2012 Oct 17.
5. Haag M, Dippenaar NG (2005) Dietary fats, fatty acids and insulin resistance: short review of a multifaceted connection. Med Sci Monit 11: RA359–367.
6. Villegas R, Liu S, Gao YT, Yang G, Li H, et al. (2007) Prospective study of dietary carbohydrates, glycemic index, glycemic load, and incidence of type 2 diabetes mellitus in middle-aged Chinese women. Arch Intern Med 167: 2310–2316.
7. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, et al. (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 346: 393–403.
8. Lindstrom J, Louheranta A, Mannelin M, Rastas M, Salminen V, et al. (2003) The Finnish Diabetes Prevention Study (DPS): Lifestyle intervention and 3-year results on diet and physical activity. Diabetes Care 26: 3230–3236.
9. Ko GT, Chan JC, Yeung VT, Chow CC, Tsang LW, et al. (2001) A low socioeconomic status is an additional risk factor for glucose intolerance in high risk Hong Kong Chinese. Eur J Epidemiol 17: 289–295.
10. Takeuchi T, Nakao M, Nomura K, Yano E (2009) Association of metabolic syndrome with depression and anxiety in Japanese men. Diabetes Metab 35: 32–36.
11. Mezuk B, Eaton WW, Albrecht S, Golden SH (2008) Depression and type 2 diabetes over the lifespan: a meta-analysis. Diabetes Care 31: 2383–2390.
12. Ko GT, Chan JC, Chan AW, Wong PT, Hui SS, et al. (2007) Association between sleeping hours, working hours and obesity in Hong Kong Chinese: the 'better health for better Hong Kong' health promotion campaign. Int J Obes (Lond) 31: 254–260.
13. Willi C, Bodenmann P, Ghali WA, Faris PD, Cornuz J (2007) Active smoking and the risk of type 2 diabetes: a systematic review and meta-analysis. Jama 298: 2654–2664.
14. Chen CC, Li TC, Chang PC, Liu CS, Lin WY, et al. (2008) Association among cigarette smoking, metabolic syndrome, and its individual components: the metabolic syndrome study in Taiwan. Metabolism 57: 544–548.
15. Yeh HC, Duncan BB, Schmidt MI, Wang NY, Brancati FL (2010) Smoking, smoking cessation, and risk for type 2 diabetes mellitus: a cohort study. Ann Intern Med 152: 10–17.
16. Ko GT, Chan JC, Tsang LW, Critchley JA, Cockram CS (2001) Smoking and diabetes in Chinese men. Postgrad Med J 77: 240–243.
17. Deurenberg-Yap M, Chew SK, Deurenberg P (2002) Elevated body fat percentage and cardiovascular risks at low body mass index levels among Singaporean Chinese, Malays and Indians. Obes Rev 3: 209–215.
18. Deurenberg-Yap M, Chew SK, Lin VF, Tan BY, van Staveren WA, et al. (2001) Relationships between indices of obesity and its co-morbidities in multi-ethnic Singapore. Int J Obes Relat Metab Disord 25: 1554–1562.
19. Deurenberg P, Deurenberg-Yap M, Guricci S (2002) Asians are different from Caucasians and from each other in their body mass index/body fat per cent relationship. Obes Rev 3: 141–146.
20. Deurenberg P, Yap M, van Staveren WA (1998) Body mass index and percent body fat: a meta analysis among different ethnic groups. Int J Obes Relat Metab Disord 22: 1164–1171.
21. Chan JN, Malik V, Jia W, Kadowaki T, Yajnik C, et al. (2009) Diabetes in asia: Epidemiology, risk factors, and pathophysiology. JAMA: The Journal of the American Medical Association 301: 2129–2140.
22. Deurenberg P, Deurenberg-Yap M, Guricci S (2002) Asians are different from Caucasians and from each other in their body mass index/body fat per cent relationship. Obesity Reviews 3: 141–146.
23. Huxley R, James WPT, Barzi F, Patel JV, Lear SA, et al. (2008) Ethnic comparisons of the cross-sectional relationships between measures of body size with diabetes and hypertension. Obesity Reviews 9: 53–61.
24. Alberti KGMM, Zimmet P, Shaw J (2007) International Diabetes Federation: a consensus on Type 2 diabetes prevention. Diabet Med 24: 451–463.
25. Fujimoto WY (1996) Overview of non-insulin-dependent diabetes mellitus (NIDDM) in different population groups. Diabet Med 13: S7–10.
26. Misra A, Ganda OP (2007) Migration and its impact on adiposity and type 2 diabetes. Nutrition 23: 696–708.
27. Zheng Y, Lamoureux EL, Ikram MK, Mitchell P, Wang JJ, et al. (2012) Impact of migration and acculturation on prevalence of type 2 diabetes and related eye complications in Indians living in a newly urbanised society. PLoS One 7: e34829.
28. WHO Expert Consultation (2004) Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 363: 157–163.
29. Permutt MA, Wasson J, Cox N (2005) Genetic epidemiology of diabetes. The journal of Clinical Investigation 115: 1431–1439.
30. Jowett JB, Diego VP, Kotea N, Kowlessur S, Chitson P, et al. (2009) Genetic Influences on Type 2 Diabetes and Metabolic Syndrome Related Quantitative Traits in Mauritius. Twin Research and Human Genetics 12: 44–52.
31. Almgren P, Lehtovirta M, Isomaa B, Sarelin L, Taskinen M, et al. (2011) Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. Diabetologia 54: 2811–2819.
32. Weijnen CF, Rich SS, Meigs JB, Krolewski AS, Warram JH (2002) Risk of diabetes in siblings of index cases with Type 2 diabetes: implications for genetic studies. Diabetic Medicine 19: 41–50.
33. Chege MP (2010) Risk factors for type 2 diabetes mellitus among patients attending a rural Kenyan hospital.
34. Altshuler D, Hirschhorn J, Klannemark M, Lindgren C, Vohl V, et al. (2000) The common PPARG Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet 26: 76–80.
35. Gloyn AL, Weedon MN, Owen KR, Turner MJ, Knight BA, et al. (2003) Large-Scale Association Studies of Variants in Genes Encoding the Pancreatic β-Cell KATP Channel Subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) Confirm That the KCNJ11 E23K Variant Is Associated With Type 2 Diabetes. Diabetes 52: 568–572.
36. Sandhu MS, Weedon MN, Fawcett KA, Wasson J, Debenham SL, et al. (2007) Common variants in WFS1 confer risk of type 2 diabetes. Nat Genet 39: 951–953.
37. Grant SFA, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, et al. (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet 38: 320–323.
38. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, et al. (2007) A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants. Science 316: 1341–1345.
39. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, et al. (2007) Replication of Genome-Wide Association Signals in UK Samples Reveals Risk Loci for Type 2 Diabetes. Science 316: 1336–1341.
40. Yasuda K, Miyake K, Horikawa Y, Hara K, Osawa H, et al. (2008) Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. Nat Genet 40: 1092–1097.
41. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, et al. (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 40: 638–645.

Table S1 Association results of Singapore GWAS for known T2D loci.
(DOCX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: MAP MDG JMY DOS. Performed the experiments: WPK EST YYT JL RW AOO MDG BT RK. Analyzed the data: ZC CH. Wrote the paper: ZC DOS MAP MS MDG.

42. Rung J, Cauchi S, Albrechtsen A, Shen L, Rocheleau G, et al. (2009) Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. Nat Genet 41: 1110–1115.

43. Takeuchi F, Serizawa M, Yamamoto K, Fujisawa T, Nakashima E, et al. (2009) Confirmation of Multiple Risk Loci and Genetic Impacts by a Genome-Wide Association Study of Type 2 Diabetes in the Japanese Population. Diabetes 58: 1690–1699.

44. Qi L, Cornelis MC, Kraft P, Stanya KJ, Linda Kao WH, et al. (2010) Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. Human Molecular Genetics 19: 2706–2715.

45. Shu XO, Long J, Cai Q, Qi L, Xiang Y-B, et al. (2010) Identification of New Genetic Risk Variants for Type 2 Diabetes. PLoS Genet 6: e1001127.

46. Tsai F-J, Yang C-F, Chen C-C, Chuang L-M, Lu C-H, et al. (2010) A Genome-Wide Association Study Identifies Susceptibility Variants for Type 2 Diabetes in Han Chinese. PLoS Genet 6: e1000847.

47. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, et al. (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 42: 579–589.

48. Yamauchi T, Hara K, Maeda S, Yasuda K, Takahashi A, et al. (2010) A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. Nat Genet 42: 864–868.

49. Cui B, Zhu X, Xu M, Guo T, Zhu D, et al. (2011) A Genome-Wide Association Study Confirms Previously Reported Loci for Type 2 Diabetes in Han Chinese. PLoS ONE 6: e22353.

50. Kooner JS, Saleheen D, Sim X, Sehmi J, Zhang W, et al. (2011) Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat Genet 43: 984–989.

51. Sim X, Ong RT-H, Suo C, Tay W-T, Liu J, et al. (2011) Transferability of Type 2 Diabetes Implicated Loci in Multi-Ethnic Cohorts from Southeast Asia. PLoS Genet 7: e1001363.

52. Cho YS, Chen C-H, Hu C, Long J, Hee Ong RT, et al. (2012) Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat Genet 44: 67–72.

53. Huang J, Ellinghaus D, Franke A, Howie B, Li Y (2012) 1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data. Eur J Hum Genet 20: 801–805.

54. Imamura M, Maeda S, Yamauchi T, Hara K, Yasuda K, et al. (2012) A single-nucleotide polymorphism in ANK1 is associated with susceptibility to type 2 diabetes in Japanese populations. Human Molecular Genetics 21: 3042–3049.

55. Palmer ND, McDonough CW, Hicks PJ, Roh BH, Wing MR, et al. (2012) A Genome-Wide Association Search for Type 2 Diabetes Genes in African Americans. PLoS ONE 7: e29202.

56. Perry JRB, Voight BF, Yengo L, Amin N, Dupuis J, et al. (2012) Stratifying Type 2 Diabetes Cases by BMI Identifies Genetic Risk Variants in LAMA1 and Enrichment for Risk Variants in Lean Compared to Obese Cases. PLoS Genet 8: e1002741.

57. Wellcome Trust Case Control C (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661–678.

58. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. (2009) Finding the missing heritability of complex diseases. Nature 461: 747–753.

59. Unoki H, Takahashi A, Kawaguchi T, Hara K, Horikoshi M, et al. (2008) SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. Nat Genet 40: 1098–1102.

60. Hindorff LA MJ, Wise A, Junkins HA, Hall PN, et al. A Catalog of Published Genome-Wide Association Studies. Available: www.genome.gov/gwastudies. Accessed 2013 May 17.

61. Department of Statistics S (2013) Population Trends 2013. Available: http://www.singstat.gov.sg/publications/publications_and_papers/population_and_population_structure/population2013.pdf. Accessed 2014 Jan 15.

62. Hankin JH, Stram DO, Arakawa K, Park S, Low S-H, et al. (2001) Singapore Chinese Health Study: Development, Validation, and Calibration of the Quantitative Food Frequency Questionnaire. Nutrition and Cancer 39: 187–195.

63. Satagopan JM, Elston RC (2003) Optimal two-stage genotyping in population-based association studies. Genet Epidemiol 25: 149–157.

64. Wang H, Stram DO (2006) Optimal two-stage genome-wide association designs based on False Discovery Rate. Statistical and Computational Data Analysis 51(2): 457–465.

65. Odegaard AO, Pereira MA, Koh W-P, Arakawa K, Lee H-P, et al. (2008) Coffee, tea, and incident type 2 diabetes: the Singapore Chinese Health Study. The American Journal of Clinical Nutrition 88: 979–985.

66. Odegaard AO, Koh W-P, Arakawa K, Yu MC, Pereira MA (2010) Soft Drink and Juice Consumption and Risk of Physician-diagnosed Incident Type 2 Diabetes: The Singapore Chinese Health Study. American Journal of Epidemiology 171: 701–708.

67. Steffes M, Cleary P, Goldstein D, Little R, Wiedmeyer HM, et al. (2005) Hemoglobin A1c measurements over nearly two decades: sustaining comparable values throughout the Diabetes Control and Complications Trial and the Epidemiology of Diabetes Interventions and Complications study. Clin Chem 51: 753–758.

68. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.

69. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, et al. (2006) The UCSC Genome Browser Database: update 2006. Nucleic Acids Res 34: D590–598.

70. Delaneau O, Marchini J, Zagury JF (2012) A linear complexity phasing method for thousands of genomes. Nature methods 9: 179–181.

71. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, et al. (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061–1073.

72. Howie BN, Donnelly P, Marchini J (2009) A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. PLoS Genet 5: e1000529.

73. Gauderman WJ, Morrison JM (2006) QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies. Available: http://hydra.usc.edu/gxe/. Accessed 2014 Jan 15.

74. Chen F, Chen GK, Millikan RC, John EM, Ambrosone CB, et al. (2011) Fine-mapping of breast cancer susceptibility loci characterizes genetic risk in African Americans. Human Molecular Genetics 20: 4491–4503.

75. Lee Sang H, Wray Naomi R, Goddard Michael E, Visscher Peter M (2011) Estimating Missing Heritability for Disease from Genome-wide Association Studies. American journal of human genetics 88: 294–305.

76. Stram DO (2005) Software for tag single nucleotide polymorphism selection. Hum Genomics 2: 144–151.

77. Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: A Tool for Genome-wide Complex Trait Analysis. The American Journal of Human Genetics 88: 76–82.

78. Zollner S, Pritchard JK (2007) Overcoming the winner's curse: estimating penetrance parameters from case-control data. Am J Hum Genet 80: 605–615.

79. Takeuchi F, Serizawa M, Yamamoto K, Fujisawa T, Nakashima E, et al. (2009) Confirmation of multiple risk Loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population. Diabetes 58: 1690–1699.

80. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, et al. (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445: 881–885.

81. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, Novartis Institutes of BioMedical Research, Saxena R, Voight BF, et al. (2007) Genome-Wide Association Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels. SCIENCE 316: 1331–1336.

82. Stram DO (2004) Tag SNP selection for association studies. Genet Epidemiol 27: 365–374.

83. Maskarinec G, Grandinetti A, Matsuura G, Sharma S, Mau M, et al. (2009) Diabetes prevalence and body mass index differ by ethnicity: the Multiethnic Cohort. Ethn Dis 19: 49–55.

84. Abate N, Chandalia M (2003) The impact of ethnicity on type 2 diabetes. J Diabetes Complications 17: 39–58.

85. Lear SA, Humphries KH, Kohli S, Chockalingam A, Frohlich JJ, et al. (2007) Visceral adipose tissue accumulation differs according to ethnic background: results of the Multicultural Community Health Assessment Trial (M-CHAT). Am J Clin Nutr 86: 353–359.

86. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, et al. (2010) LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26: 2336–2337.