Methods

Study population

We did a GWAS of two study samples: one from IWPC sites across the USA and one from the University of Alabama at Birmingham (UAB; Birmingham, AL, USA). Together, these samples made up the discovery cohort. Patients enrolled in IWPC sites, but who were not used for discovery (enrolled at local site after samples from that site were sent for genotyping or their local site did not join IWPC until after IWPC samples were sent for genotyping), made up an independent replication cohort. All study participants (discovery and replication cohorts) were adults (aged ≥18 years) of self-reported African ancestry who were taking a stable maintenance dose of warfarin, and for whom clinical and demographic data had been obtained as previously reported.[5,13] Definitions of stable dose varied between sites (appendix), but most sites insisted that patients had to have been on a stable dose for at least three consecutive clinic visits, which is indicative of the maintenance phase of warfarin treatment after initial stable dosing has been achieved. Patients provided a DNA sample (whole-blood, saliva, or mouthwash sample). All participants gave written informed consent to participate, and all protocols were approved by local institutional review boards.

Procedures

For the IWPC samples, data curation and quality control procedures were as described previously.[13,14] Data for several variables were obtained for participants from IWPC sites in the discovery cohort (appendix). These individuals were genotyped with the Illumina 610 Quad BeadChip (Illumina, San Diego, CA, USA) at the RIKEN Center for Genomic Medicine (Yokohama, Japan).

Information about demographic and clinical variables was obtained for the UAB study sample as previously described.[5] Participants from UAB were genotyped at the University of Washington's Department of Genome Sciences (Seattle, WA, USA) with the Illumina Human1M-Duo 3.0 (Illumina, San Diego, CA, USA). SNPs previously associated with warfarin dose were genotyped independently (ie, via single SNP genotyping and at local institutions) in both datasets.

In the replication cohort, samples were tested for a genome-wide significant signal ($p<5\times10^{-8}$). To ensure important signals approaching genome-wide significance were not missed, we identified an additional 12 SNPs with $10^{-5}>p>5\times10^{-8}$ as being of interest that were also tested in the replication cohort. IWPC replication samples were genotyped at Vanderbilt University's DNA Resources Core (Nashville, TN, USA) with the Sequenom (San Diego, CA, USA) platform, and UAB replication samples at the University of Washington (Seattle, WA, USA) by Sanger sequencing.

To provide insight into potential functional mechanisms of an identified SNP, we assessed the pharmacokinetics of warfarin in an independent cohort of 60 African Americans on stable warfarin doses, as previously described.[16] Briefly, venous blood was obtained 12–16 h after a warfarin dose. Warfarin enantiomer concentrations were established by a chiral high-performance liquid chromatography-based method, and the oral clearance of both enantiomers, corrected for body surface area, was calculated.[17] Warfarin is almost exclusively cleared by liver metabolism, thus oral clearance approximates the hepatic metabolism of warfarin.

The phenotypic data and a restricted genotypic dataset (eg, *VKORC1*, *CYP2C9*, *CYP4F2*, and rs12777823) are accessible for download at PharmGKB. The IWPC GWAS data are made available under the IWPC Data Usage Agreement and can be accessed at PharmGKB.

Statistical analysis

Conventional GWAS quality control procedures were undertaken, such as checks for sex mismatches, identity by descent testing, deviation from Hardy–Weinberg Equilibrium, and SNP and sample exclusions on the basis of call rates (appendix).

Imputation of additional SNPs was done with hidden Markov models via MACH (version 1.0),[18] on the basis of HapMap Phase 2 Release 22 data (appendix). The quality metric (r^2) was calculated for all imputed SNPs; no SNPs reported here had a value of less than 0·8.

Global estimates of ancestry (appendix) were established by the first and second principal components, with the first principal component used as a global estimate of African ancestry. All covariates obtained in each dataset (including drugs; appendix) and the first ten principal components obtained through EIGENSTRAT (version 4.2)[19] were tested as single covariates in a PLINK (version 1.07) additive model analysis for association with stable warfarin dose. Mean height and weight by sex were used for participants missing these variables in the IWPC cohort.

For all analyses, the primary phenotype (weekly stable warfarin dose) was log transformed to give a normal distribution. We excluded individuals missing data for non-genetic covariates associated with warfarin dose. For the discovery cohort, GWAS analysis was done in PLINK with a linear regression additive genetic model, by use of covariates significantly associated with stable warfarin dose. We combined GWAS results for the IWPC and UAB study samples with fixed effects meta-analysis. We used the same quality-control procedures and analysis approaches in the replication cohort.

Previous warfarin GWAS in individuals of European and Asian ancestry emphasised the importance of conditioning analyses for the well recognised *VKORC1* and *CYP2C9* SNPs.[10,12,20] Therefore, a-priori analysis plans also included a stepwise conditional analysis, conditioning first for *VKORC1* –1639G→A, followed by the composite genotype (presence or absence of a variant allele) of *CYP2C9*2* and *CYP2C9*3* (appendix).

We prespecified a genome-wide significance threshold of $p<5\times10^{-8}$ for the discovery cohort. For the replication

Pharmacotherapy and Translational Research, University of Florida, Gainesville, FL, USA (B M Burkley MS, T Y Langaee PhD, M H Shahin MS, Prof J A Johnson PharmD); **Department of Genetics and Genomics Sciences** (Prof R J Desnick MD, S A Scott PhD) **and The Zena and Michael A Wiener Cardiovascular Institute** (Prof J L Halperin), **Mount Sinai School of Medicine, New York, NY, USA; Pharmaceutical Sciences Section, College of Pharmacy, Qatar University, Doha, Qatar** (Prof S I Khalifa); **Cardiac Arrhythmia Service and Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA** (S A Lubitz MD); **Aurora St Luke's Medical Center, Milwaukee, WI, USA** (M Tector PhD); **Department of Pathology and Laboratory Medicine and Department of Genetics** (Prof K E Weck MD), **and School of Pharmacy** (M J Wagner PhD), **University of North Carolina at Chapel Hill; Chapel Hill, NC, USA; Department of Genome Sciences, University of Washington, Seattle, WA, USA** (M J Rieder PhD); **Department of Laboratory Medicine, University of California, San Francisco, San Francisco, CA, USA** (A H B Wu PhD); **Marshfield Clinic Research Foundation, Marshfield, WI, USA** (J K Burmester PhD); **Department of Medical Sciences, Clinical Pharmacology, Uppsala University, Uppsala, Sweden** (M Wadelius MD); **and RIKEN Center for Genomic Medicine, Yokohama, Japan** (T Mushiroda PhD, M Kubo MD, Prof Y Nakamura MD)

Correspondence to:
Prof Julie A Johnson, Center for Pharmacogenomics, Department of Pharmacotherapy and Translational Research, University of Florida, Box 100486, Gainesville, FL 32610-0486, USA
johnson@cop.ufl.edu

See Online for appendix

To **download phenotypic data from PharmGKB** see http://www.pharmgkb.org/downloads.jsp

	Discovery cohort		Replication cohort (n=432)
	International Warfarin Pharmacogenetics Consortium (n=327)	University of Alabama at Birmingham (n=206)	
Stable warfarin dose (mg/wk)	45·6 (18·6)	42·2 (18·4)	45·7 (23·8)
International normalised ratio	2·46 (0·44)	2·41 (0·40)	2·48 (0·31)
Age (years)	57·1 (14·8)	59·1 (15·3)	58·6 (16·0)
Height (cm)	169·4 (10·1)	170·9 (10·5)	170·4 (27·2)
Weight (kg)	94·4 (10·1)	89·6 (22·9)	93·5 (27·8)
Sex			
Male	121 (37%)	92 (45%)	168 (39%)
Female	206 (63%)	114 (55%)	264 (61%)
Primary warfarin indication*			
Pulmonary embolism or deep vein thrombosis	153 (47%)	97 (47%)	186 (59%)
Atrial fibrillation or flutter	81 (25%)	68 (33%)	67 (21%)
Cardiomyopathy or left ventricle dilation	5 (2%)	0	0
Heart valve replacement	26 (8%)	0	7 (2%)
Stroke	28 (9%)	21 (10%)	29 (9%)
Other	34 (10%)	20 (10%)	27 (9%)
Current smoker	53 (16%)	34 (17%)	NA
Amiodarone use	16 (5%)	7 (3%)	21 (5%)
Aspirin use	92 (28%)	93 (45%)	149 (34%)
Genotype†			
CYP2C9*2	12 (2%)	6 (1%)	19 (2%)
CYP2C9*3	5 (1%)	6 (1%)	13 (2%)
VKORC1 –1639G→A	59 (10%)	38 (10%)	77 (11%)

Data are mean (SD) or n (%). NA=not available. *Primary indication was not available for 116 individuals in the replication cohort. †Reported as number of people carrying one or two variant alleles; percentage is minor allele frequency.

Table 1: **Baseline characteristics**

cohort, we defined the significance level on the basis of the number of samples tested (13 SNPs). Therefore, significant replication was defined as a Bonferroni-corrected p value of less than 0·05/13 or 0·0038.

Additionally, we looked at a previous GWAS of individuals of European and Asian ancestry[9,10,12] to identify any association between the identified SNP and warfarin dose. We tested the SNP in a previously described Egyptian cohort.[21]

We modelled the residual dose (actual dose minus predicted IWPC dose) with linear regression, with genotypes of a replicated SNP as covariates in participants in the discovery cohort for whom complete data were available for calculation of warfarin dose with the IWPC algorithm. We tested the difference between the models (with and without rs12777823 as a covariate) by ANOVA. All statistical analyses were done with R (version 2.10.1).

Pharmacokinetic variables were compared between genotype groups by ANOVA, and between the GG genotype and both the AA and AG genotypes by the one-sided Student's unpaired *t* test, with the expected direction defined by the data from the GWAS.

Role of the funding source

The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. MAP, LHC, NAL, NJC, RBA, TEK, and JAJ had complete access to the data, and JAJ had final responsibility for the decision to submit for publication.

Results

For the IWPC study sample, eight institutions contributed genetic and clinical data for 345 African-American individuals, although 18 were subsequently excluded because of missing data or quality-control issues. The UAB study sample consisted of 206 African Americans; no participant was excluded from analysis because of quality control issues. Therefore, the discovery cohort contained 533 participants after quality control (table 1). The replication cohort consisted of 432 African Americans enrolled from six IWPC sites, including UAB (table 1; appendix).

Participants in both the IWPC and UAB cohorts clustered between the HapMap CEU (northern and western European ancestry) and YRI (African ancestry) samples, as was expected (appendix). Mean proportion of African ancestry in the discovery cohort was 85% (SD 9%) in participants from IWPC sites and 86% (8%) from UAB. The proportion of African ancestry was not associated with warfarin dose in participants from IWPC sites (p=0·08) or from UAB (p=0·944). 557 286 SNPs passed quality-control checks in the IWPC samples (mean call rate 99·96%) and 950 007 in the UAB samples (99·95%).

Age, weight, height, aspirin use, and amiodarone use were associated with warfarin dose in the discovery cohort (p<0·05 for all in both IWPC and UAB samples; appendix). A glomerular filtration rate of less than 30 mL/min in UAB samples was also associated with warfarin dose (p<0·05).

None of the SNPs reported here that were associated with stable warfarin dose significantly deviated from Hardy–Weinberg equilibrium. SNPs on chromosome 16 reached genome-wide significance, including *VKORC1* –1639G→A (rs9923231; $p=2\cdot08\times10^{-9}$; figure 1, appendix). Associations with SNPs in linkage disequilibrium with rs9923231 on chromosome 16 were also significant (appendix).

After the prespecified conditioning on the *VKORC1* locus, *CYP2C9*2*, and *CYP2C9*3*, the most significant SNP association with warfarin dose was rs12777823 on chromosome 10 ($p=1\cdot51\times10^{-8}$; figure 1, appendix). The minor allele frequency was about 25% in the discovery cohort (appendix). Other top signals after conditioning are shown in the appendix.

The SNP that achieved genome-wide significance in the discovery cohort after conditioning (rs12777823) was also significantly associated with stable warfarin dose in the replication cohort after conditioning ($p=5\cdot04\times10^{-5}$; figure 2, appendix). This p value is well below the threshold of 0·0038 that had been defined for significant replication. A combined analysis of the discovery and

replication cohorts for rs12777823 leads to a combined p value of $4{\cdot}5\times10^{-12}$. None of the 12 SNPs of interest were associated with stable warfarin dose (appendix). rs12777823 was also not associated with warfarin dose in individuals of European and Japanese ancestry, or Egyptian individuals (data not shown).

The IWPC warfarin dosing equation is less predictive for African Americans than for other populations.[13] We modelled the residual dose with linear regression in 504 participants from the discovery cohort. Addition of rs12777823 genotype to the IWPC algorithm improved the ability to predict actual therapeutic warfarin dose with the IWPC algorithm by an absolute 5% (relative 21%; table 2). Individuals heterozygous or homozygous for the rs12777823 A allele have reduced stable warfarin doses (table 2).

In the independent cohort of 60 African Americans, the hepatic metabolism of the S-enantiomer of warfarin (as estimated by oral clearance) differed significantly with rs12777823 genotype (table 3). Additionally, the ratio of the R-enantiomer and S-enantiomer differed significantly by rs12777823 genotype, consistent with genotypic differences in metabolism of the S-enantiomer, but not of the R-enantiomer (table 3).

*CYP2C9*2* and *CYP2C9*3* occurred at low frequency (table 1) and were not significantly associated with warfarin dose variability (*CYP2C9*2* nominal p=0·834; *CYP2C9*3* nominal p=0·023). Of the *CYP2C9* non-synonymous polymorphisms specific to individuals of African descent, only *CYP2C9*11* was imputed in the meta-analysis and had a significant association with warfarin dose (p=0·0001). The *CYP2C9*5*, *CYP2C9*6*, and *CYP2C9*8* alleles were not included on the GWAS arrays and could not be imputed. Analyses with additional *CYP2C9* alleles (eg, *CYP2C9*5*, *CYP2C9*6*, and *CYP2C9*8*) had low power, because few participants were genotyped for these alleles independent of the GWAS. Notably, rs12777823 is in low linkage disequilibrium with *CYP2C9*2*, *CYP2C9*3*, and *CYP2C9*11* (all $r^2<0{\cdot}05$), on the basis of the imputed genotypes for these alleles in the GWAS cohorts. Additionally, on the basis of participants independently genotyped for *CYP2C9*5*, *CYP2C9*6*, and *CYP2C9*8*, rs12777823 is also in low linkage disequilibrium with these SNPs (all $r^2<0{\cdot}10$), suggesting that the rs12777823 signal is independent of previously studied *CYP2C9* polymorphisms. The *CYP4F2* Val433Met variant (rs2108622) that has been previously associated with warfarin dose in those of European and Asian descent[10,12] was not associated with warfarin dose in this study (p=0·07).

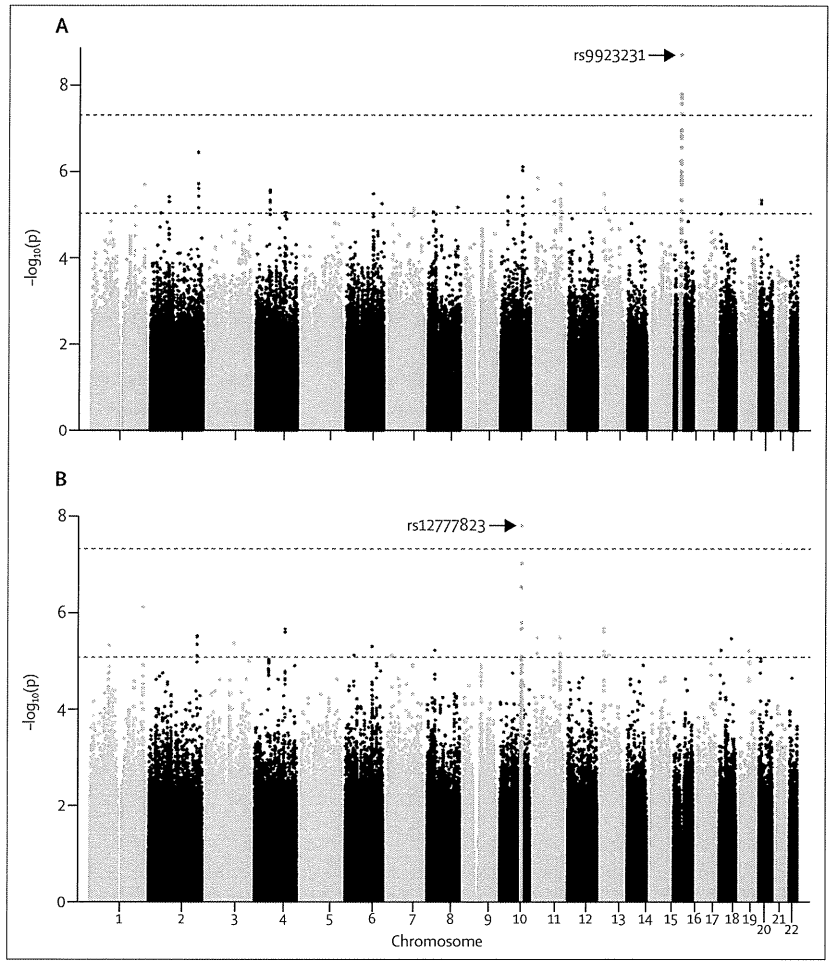

Figure 1: **Manhattan plots showing associations between single nucleotide polymorphisms and stable warfarin dose in the discovery cohort**
(A) Without and (B) with prespecified conditioning. The top line shows $p=5\times10^{-8}$. Green points indicate genome-wide significant regions.

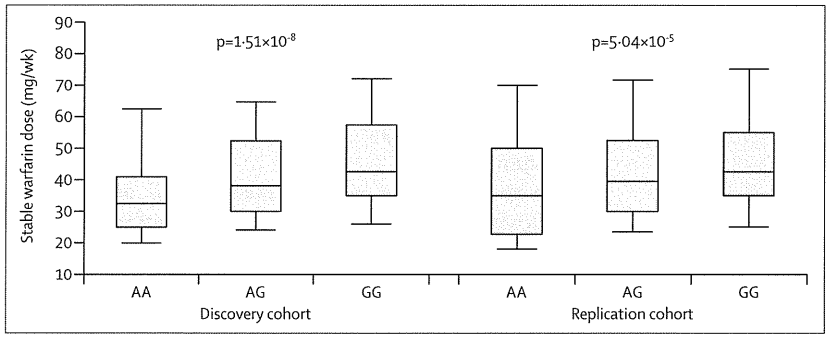

Figure 2: **Stable warfarin dose by rs12777823 genotype**
Lines within boxes represent medians; lower and upper boundaries represent IQRs; and whiskers represent tenth and 90th percentiles. p values calculated after conditioning on the *VKORC1* locus, *CYP2C9*2*, *CYP2C9*3*, and non-genetic covariates associated with warfarin dose.

Discussion

Our results show that *VKORC1* genotype is the primary genetic determinant of variability in warfarin dose requirements in African Americans, which is consistent with previous GWAS data from European and Japanese populations.[9,10,12] After adjustment for *VKORC1* −1639G→A, *CYP2C9*2*, and *CYP2C9*3*, we identified a novel association between rs12777823 and warfarin dose variability; we replicated this finding in an independent African American cohort.

	Coefficient*	p value	Adjusted model R^2†
IWPC dosing equation Predicted Dose	..	$<2\times10^{-16}$	0·2193
rs12777823 AG	−6·92	6.76×10^{-6}	..
rs12777823 AA	−9·34	0·00050	..
IWPC dosing equation and rs1277823	..	1.7×10^{-6}	0·2666

IWPC=International Warfarin Pharmacogenetics Consortium. *Represents change in warfarin dose (mg/week). †Increase indicates increase in amount of variability explained.

Table 2: **Effects of novel *CYP2C* variation on warfarin dose prediction with the IWPC dosing algorithm**

	AA (n=6)	AG (n=24)	GG (n=30)
Oral clearance of S-enantiomer (mL/min/m²)*	1·20 (0·69)	1·38 (0·48)	1·80 (0·83)
Plasma concentration ratio of R-enantiomer to S-enantiomer†	1·29 (0·50)	1·42 (0·43)	1·76 (0·68)

Data are mean (SD). Clearance values are normalised for body surface area. *p=0·042 for association between the three genotype groups (one-way ANOVA); p=0·049 for comparison between AA and GG genotypes, and p=0·016 for comparison between AG and GG genotypes (one-sided Student's unpaired *t* test). †p=0·049 for association between the three genotype groups (one-way ANOVA); p=0·039 for comparison between AA and GG genotypes, and p=0·018 for comparison between AG and GG genotypes (one-sided Student's unpaired *t* test).

Table 3: **Pharmacokinetic parameters of warfarin R-enantiomer and S-enantiomer by rs12777823 genotype**

Panel: **Research in context**

Systematic review

We searched PubMed for reports of studies and meta-analyses published in any language before Dec 1, 2012. We used the search terms "warfarin" and "genome wide association study". We identified three genome-wide association studies of warfarin dose requirement[9,10,12] and one of acenocoumarol.[20] All previous studies were done in either European or Asian populations. They all showed that that *VKORC1* −1639G→A (rs9923231), *CYP2C9**2 (rs1799853), *CYP2C9**3 (rs1057910), and *CYP4F2* (rs2108622) are the main genetic determinants of drug dose. The study of acenocoumarol[20] identified a single nucleotide polymorphism (SNP) upstream of *CYP2C18*, although it did not reach genome-wide significance after conditioning for *CYP2C9**2 and *CYP2C9**3. To identify previous studies testing genetic variation associated with warfarin dose requirements in African Americans, we did another search of PubMed with the same date criteria and no language restrictions, but used the search terms "warfarin" and "African American". We identified five studies,[22–26] all of which used a candidate-gene approach.

Interpretation

As far as we are aware, ours is the first genome-wide association study to investigate warfarin dose requirements in African Americans. Our results identified a SNP upstream of *CYP2C18*, rs12777823, that is significantly associated with warfarin dose requirement in African Americans and is independent of previous associations with *VKORC1* and *CYP2C9*. Patients carrying this SNP have a significantly lower stable dose of warfarin than do those without this variant, and the addition of this SNP improved the International Warfarin Pharmacogenetics Consortium algorithm by 21%. We have also shown that this SNP significantly alters warfarin clearance. To our knowledge, no other studies of warfarin dose requirement have identified this SNP.

In the context of other genotypes and clinical factors in the IWPC model, the rs12777823 genotype was associated with a reduction in dose of almost 7 mg/week in heterozygotes and of 9 mg/week in homozygotes—an effect similar to that of the *CYP2C9**2* allele in the IWPC dosing algorithm.[13] Thus, the effect is consistent with a polymorphism that is well accepted to be clinically relevant for individuals. The effect is also important at a population level, because of both the effect size and the frequency of the polymorphism. Specifically, rs12777823 explains 5% of the variability in warfarin dose in African Americans, which is consistent with the variability explained by *CYP2C9**2* and *CYP2C9**3* in individuals of European ancestry,[8] and greatly exceeds the variability explained by *CYP2C9**2* and *CYP2C9**3* in African Americans (1–2%).[5,22]

Our findings have important implications for prediction of warfarin maintenance dose in individuals of African ancestry, for whom current warfarin pharmacogenetic algorithms are less predictive.[13,14] On the basis of our data, the IWPC algorithm could be used to estimate warfarin dose, and then the rs12777823 variant in African Americans could be accounted for with dose reductions (panel). Further improvement in dose estimation might eventually be possible for African Americans when additional SNPs specific to individuals with African ancestry are identified.

rs12777823 is located within the *CYP2C* gene cluster, upstream from *CYP2C18* on chromosome 10q23, and includes the *CYP2C9*, *CYP2C8*, *CYP2C18*, and *CYP2C19* genes. In a GWAS of clopidogrel antiplatelet effect in Amish individuals,[27] rs12777823 was the strongest signal, although it was in strong linkage disequilibrium (r^2=0·87) with *CYP2C19**2* (rs4244285), which produces a non-functional truncated CYP2C19 protein and is believed to be the causative variant. We reported low linkage disequilibrium (r^2=0·5) between rs12777823 and *CYP2C19**2*, and the association between warfarin dose requirements and *CYP2C19**2* was not significant at the genome-wide level and was absent when conditioned on rs12777823. Therefore, the effect of rs12777823 in African Americans is probably independent of *CYP2C19**2*.

The functional basis for the reported association between warfarin dose and rs12777823 is unknown, but the pharmacokinetic data indicate an effect on metabolism of the S-enantiomer. The S-enantiomer is metabolised primarily by CYP2C9, and the stereoselective nature of the genetic association suggests rs12777823 affects CYP2C9 activity. Importantly, the low linkage disequilibrium between this SNP and other known functional SNPs in *CYP2C9* suggest that it has an independent and novel genetic effect on CYP2C9 metabolism, which could have clinically important implications for many other drugs that are CYP2C9 substrates.

rs12777823 is common in African Americans: in our study, the minor allele frequency was 25% in both the discovery and replication cohorts. Although common in European (minor allele frequency 14%) and Japanese (32%) populations, rs12777823 did not have a significant

association with warfarin dose in previous GWAS,[9,10,12] or when we interrogated the European and Japanese GWAS datasets for this SNP. Similarly, this SNP did not have an association with warfarin dose in about 200 individuals from Egypt. Collectively, these data suggest that the association in African Americans might not be due to rs12777823 itself, but a causal SNP in linkage disequilibrium with rs12777823 in African Americans but not in other populations.

Although the *CYP2C9*2* and *CYP2C9*3* alleles contribute to warfarin dose variability in populations of European descent,[9,10] we recorded no genome-wide significant association in the GWAS of African Americans. The low frequency of these alleles in this population might have contributed to this finding.

GWAS has limitations because not all variants are interrogated—eg, a novel *VKORC1* variant (rs61162043) identified through resequencing and associated with increased warfarin dose requirements in African Americans[22] is not captured on the GWAS genotyping platforms, nor is it in HapMap.[28] *CYP2C9*5*, *CYP2C9*6*, *CYP2C9*8*, and *CYP2C9*11* have been previously associated with warfarin dose variability,[5,22,29] but were not included on the GWAS genotyping platforms; only *CYP2C9*11* was available through imputation. Because few sites independently genotyped for these variants, we could not reliably estimate their contribution to warfarin dose requirements in our study. However, we could rule out the *CYP2C9*2*, *CYP2C9*3*, and *CYP2C9*11* alleles as causes of the recorded association with the rs12777823 variant on the basis of conditional analyses and the computed pairwise linkage disequilibrium between these SNPs. Additionally, linkage disequilibrium between rs12777823 and *CYP2C9*5*, *CYP2C9*6*, and *CYP2C9*8* was low in participants for whom these data were available.

Although use of new oral anticoagulants is expected to increase with time, warfarin will probably be widely used for the foreseeable future. A study reported in 2012[1] suggested that dabigatran use has risen consistently since its approval; 8·1% of oral anticoagulant prescriptions in the USA in late 2011 were for this drug. In Sweden, the proportion of prescriptions for oral anticoagulants that were for new drugs rose from 3·0% in 2011, to 4·7% in 2012.[30] Thus, although new oral anticoagulant use is increasing, warfarin remains the predominant oral anticoagulant. Additionally, in December, 2012, the US Food and Drugs Administration issued a safety notice that outlined that dabigatran is contraindicated in patients with mechanical heart valves, meaning new oral anticoagulants will not be used for all the indications for which warfarin is presently used.

Importantly, because of cost issues, warfarin will probably remain the mainstay oral anticoagulant for the uninsured or underinsured in the USA and individuals in developing countries—particularly in Africa—for many years to come. Therefore, identification of a way to achieve therapeutic anticoagulation efficiently with warfarin is a priority. Pharmacogenetic warfarin dosing algorithms, such as the IWPC algorithm, could reduce the time needed to reach stable therapeutic dosing;[8,13,31] genotype-guided warfarin dosing better predicts warfarin maintenance dose than does standard dosing, and reduces out-of-range international normalised ratio values.[31] Secondary analyses in a previous study[31] also suggested the potential for improved outcomes via a pharmacogenetic approach. Randomised controlled trials are assessing the clinical utility of genotype-guided warfarin dosing.[32,33] If the trials show benefits for pharmacogenetic approaches to warfarin, then future consensus guidelines will probably support warfarin pharmacogenetics in the clinical setting.

In conclusion, we have shown that *VKORC1* is the major gene affecting warfarin dose variability in African Americans, with the *CYP2C* locus exerting influence independent of well known *CYP2C9*2*, *CYP2C9*3*, and *CYP2C19*2* polymorphisms. Our findings suggest that warfarin dose variability is affected by variants other than the well established *VKORC1* and *CYP2C9* ones in African Americans; these new variants could improve dose prediction in these individuals.

Contributors

MAP, LHC, NAL, PD, MJW, DMR, RBA, TEK, and JAJ did the literature search. MAP, LHC, NAL, DCC, RJD, SIK, SAL, MJR, SAS, AHBW, JKB, PD, MJW, TM, DMR, NJC, RBA, TEK, YN, and JAJ designed the study. MAP, LHC, NAL, DCC, NT, HT, YB, BMB, JLH, TYL, EAN, MO, MHS, SRP, HS, MT, KEW, MJR, SAS, MJW, TM, and MK gathered data and generated laboratory data. MAP, LHC, NAL, ERG, AK, and AP prepared figures. MAP, LHC, NAL, ERG, AK, RD, AP, DCC, JW, NL, NT, SB, HT, MJR, MK, and NJC analysed data. MAP, LHC, NAL, ERG, AK, RD, AP, DCC, JL, NL, NT, SB, HT, SAS, AHBW, JKB, PD, MJW, TM, DMR, NJC, RBA, TEK, YN, and JAJ interpreted data. MAP, LHC, NAL, RBA, TEK, and JAJ drafted the report. All authors contributed to revision of the report.

Conflicts of interest

NAL, RJD, JLH, JAJ are principal investigators; TYL, SAS, and JKB are members of the genotyping committee; and JAJ is a member of the executive committee for the National Institutes of Health Clarification of Optimal Anticoagulation through Genetics trial. KEW is an unpaid consultant for Roche Diagnostics and a paid consultant for the Gentris Corporation. JKB is a site principal investigator for the Iverson Warfarin Trial and holds a patent pending for CYP4F2 use for warfarin dosing. RBA is founder, equity holder, and consultant for Personalis. The other authors declare that they have no conflicts of interest.

Acknowledgments

MAP (K23 HL089808), NAL (K23 NS45598; RO1 HL092173; UL1 TR000165), NL (R01GM081488), MT and JKB (UL1 TR000427), SAS (KL2 TR000069), DMR (U19 HL 065962; RC2 GM092618), NJC (U19 GM61393; U19 HL065962), RBA and TEK (R24 GM061374), and JAJ (U01 GM074492) were supported by National Institutes of Health grants. LHC was supported by an American Heart Association Grant-In-Aid (10GRNT3750024). RD was supported by the Howard Hughes Medical Institute. MT and JKB were also supported by the Wisconsin Network for Health Research. PD was supported by the Wellcome Trust (grant 098051).

References

1 Kirley K, Qato DM, Kornfield R, Stafford RS, Alexander GC. National trends in oral anticoagulant use in the United States, 2007 to 2011. *Circ Cardiovasc Qual Outcomes* 2012; **5**: 615–21.

2 Budnitz DS, Lovegrove MC, Shehab N, Richards CL. Emergency hospitalizations for adverse drug events in older Americans. *N Engl J Med* 2011; **365**: 2002–12.

3 Hylek EM, Go AS, Chang Y, et al. Effect of intensity of oral anticoagulation on stroke severity and mortality in atrial fibrillation. *N Engl J Med* 2003; **349:** 1019–26.

4 Hylek EM, Evans-Molina C, Shea C, Henault LE, Regan S. Major hemorrhage and tolerability of warfarin in the first year of therapy among elderly patients with atrial fibrillation. *Circulation* 2007; **115:** 2689–96.

5 Limdi NA, Arnett DK, Goldstein JA, et al. Influence of *CYP2C9* and *VKORC1* on warfarin dose, anticoagulation attainment and maintenance among European-Americans and African-Americans. *Pharmacogenomics* 2008; **9:** 511–26.

6 Ferder N, Eby CS, Deych E, et al. Ability of VKORC1 and CYP2C9 to predict therapeutic warfarin dose during the initial weeks of therapy. *J Thromb Haemost* 2010; **8:** 95–100.

7 Rieder MJ, Reiner AP, Gage BF, et al. Effect of *VKORC1* haplotypes on transcriptional regulation and warfarin dose. *N Engl J Med* 2005; **352:** 2285–93.

8 Gage B, Eby C, Johnson J, et al. Use of pharmacogenetic and clinical factors to predict the therapeutic dose of warfarin. *Clin Pharmacol Ther* 2008; **84:** 326–31.

9 Cooper GM, Johnson JA, Langaee TY, et al. A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose. *Blood* 2008; **112:** 1022–27.

10 Takeuchi F, McGinnis R, Bourgeois S, et al. A genome-wide association study confirms *VKORC1, CYP2C9,* and *CYP4F2* as principal genetic determinants of warfarin dose. *PLoS Genet* 2009; **5:** e1000433.

11 Limdi NA, Beasley TM, Crowley MR, et al. *VKORC1* polymorphisms, haplotypes and haplotype groups on warfarin dose among African-Americans and European-Americans. *Pharmacogenomics* 2008; **9:** 1445–58.

12 Cha PC, Mushiroda T, Takahashi A, et al. Genome-wide association study identifies genetic determinants of warfarin responsiveness for Japanese. *Hum Mol Genet* 2011; **19:** 4735–44.

13 Klein TE, Altman RB, Eriksson N, et al. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N Engl J Med* 2009; **360:** 753–64.

14 Limdi NA, Wadelius M, Cavallari L, et al. Warfarin pharmacogenetics: a single *VKORC1* polymorphism is predictive of dose across 3 racial groups. *Blood* 2010; **115:** 3827–34.

15 Tishkoff SA, Dietzsch E, Speed W, et al. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* 1996; **271:** 1380–87.

16 Liu Y, Jeong H, Takahashi H, et al. Decreased warfarin clearance associated with the CYP2C9 R150H (*8) polymorphism. *Clin Pharmacol Ther* 2012; **91:** 660–65.

17 Takahashi H, Kashima T, Nomizo Y, et al. Metabolism of warfarin enantiomers in Japanese patients with heart disease having different *CYP2C9* and *CYP2C19* genotypes. *Clin Pharmacol Ther* 1998; **63:** 519–28.

18 Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 2010; **34:** 816–34.

19 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:** 904–09.

20 Teichert M, Eijgelsheim M, Rivadeneira F, et al. A genome-wide association study of acenocoumarol maintenance dosage. *Hum Mol Genet* 2009; **18:** 3758–68.

21 Shahin MH, Khalifa SI, Gong Y, et al. Genetic and nongenetic factors associated with warfarin dose requirements in Egyptian patients. *Pharmacogenet Genomics* 2011; **21:** 130–35.

22 Perera MA, Gamazon E, Cavallari LH, et al. The missing association: sequencing-based discovery of novel SNPs in VKORC1 and CYP2C9 that affect warfarin dose in African Americans. *Clin Pharmacol Ther* 2011; **89:** 408–15.

23 Ramirez AH, Shi Y, Schildcrout JS, et al. Predicting warfarin dosage in European-Americans and African-Americans using DNA samples linked to an electronic health record. *Pharmacogenomics* 2012; **13:** 407–18.

24 Cavallari LH, Perera M, Wadelius M, et al. Association of the GGCX (CAA)16/17 repeat polymorphism with higher warfarin dose requirements in African Americans. *Pharmacogenet Genomics* 2012; **22:** 152–58.

25 Schelleman H, Brensinger CM, Chen J, Finkelman BS, Rieder MJ, Kimmel SE. New genetic variant that might improve warfarin dose prediction in African Americans. *Br J Clin Pharmacol* 2010; **70:** 393–99.

26 Voora D, Koboldt DC, King CR, et al. A polymorphism in the VKORC1 regulator calumenin predicts higher warfarin dose requirements in African Americans. *Clin Pharmacol Ther* 2010; **87:** 445–51.

27 Shuldiner AR, O'Connell JR, Bliden KP, et al. Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *JAMA* 2009; **302:** 849–57.

28 Gamazon ER, Ziliak D, Im HK, et al. Genetic architecture of microRNA expression: implications for the transcriptome and complex traits. *Am J Hum Genet* 2012; **90:** 1046–63.

29 Cavallari LH, Langaee TY, Momary KM, et al. Genetic and clinical predictors of warfarin dose requirements in African Americans. *Clin Pharmacol Ther* 2010; **87:** 459–64.

30 The National Board of Health and Welfare. Läkemedelsstatistik. http://192.137.163.49/sdb/lak/val.aspx (accessed May 10, 2013).

31 Anderson JL, Horne BD, Stevens SM, et al. A randomized and clinical effectiveness trial comparing two pharmacogenetic algorithms and standard care for individualizing warfarin dosing (CoumaGen-II). *Circulation* 2012; **125:** 1997–2005.

32 French B, Joo J, Geller NL, et al. Statistical design of personalized medicine interventions: the Clarification of Optimal Anticoagulation through Genetics (COAG) trial. *Trials* 2010; **11:** 108.

33 van Schie RM, Wadelius MI, Kamali F, et al. Genotype-guided dosing of coumarin derivatives: the European pharmacogenetics of anticoagulant therapy (EU-PACT) trial design. *Pharmacogenomics* 2009; **10:** 1687–95.

Human Molecular Genetics, 2014, Vol. 23, No. 1 **239–246**
doi:10.1093/hmg/ddt399
Advance Access published on August 14, 2013

Genome-wide association study identifies three novel loci for type 2 diabetes

Kazuo Hara[1,2,*,†], Hayato Fujita[1,†], Todd A. Johnson[3,†], Toshimasa Yamauchi[1,6], Kazuki Yasuda[8], Momoko Horikoshi[1], Chen Peng[9], Cheng Hu[10], Ronald C. W. Ma[11,12,13], Minako Imamura[4], Minoru Iwata[15], Tatsuhiko Tsunoda[3], Takashi Morizono[3], Nobuhiro Shojima[1], Wing Yee So[11,12,13], Ting Fan Leung[14], Patrick Kwan[11], Rong Zhang[10], Jie Wang[10], Weihui Yu[10], Hiroshi Maegawa[16], Hiroshi Hirose[17], DIAGRAM consortium[18], Kohei Kaku[19], Chikako Ito[20], Hirotaka Watada[7], Yasushi Tanaka[21], Kazuyuki Tobe[15], Atsunori Kashiwagi[16], Ryuzo Kawamori[7], Weiping Jia[10], Juliana C. N. Chan[11,12,13], Yik Ying Teo[9,22,23,24,26], Tai E. Shyong[9,25,27], Naoyuki Kamatani[28], Michiaki Kubo[5], Shiro Maeda[4] and Takashi Kadowaki[1]

[1]Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo 113-8655, Japan, [2]Department of Integrated Molecular Science on Metabolic Diseases, 22nd Century Medical and Research Center, the University of Tokyo, Tokyo 113-8655, Japan, [3]Laboratory for Medical Science Mathematics, [4]Laboratory for Endocrinology, Metabolism, and Kidney Diseases and [5]Research Group for Genotyping, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa 230-0045, Japan, [6]Sportology Center, Graduate School of Medicine and [7]Department of Medicine, Metabolism and Endocrinology, School of Medicine, Juntendo University, Tokyo 113-8421, Japan, [8]Department of Metabolic Disorder, Diabetes Research Center, National Center for Global Health and Medicine, Research Institute, Tokyo 162-8655, Japan, [9]Saw Swee Hock School of Public Health, National University of Singapore, MD3, 16 Medical Drive, Singapore 117597, Singapore, Singapore, [10]Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai Diabetes Institute, Shanghai Key Laboratory of Diabetes Mellitus, 600 Yishan Road, Shanghai 200233, China, [11]Department of Medicine and Therapeutics, Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, China, [12]Li Ka Shing Institute of Health Sciences, [13]Hong Kong Institute of Diabetes and Obesity and [14]Department of Paediatrics, Chinese University of Hong Kong, Hong Kong, China, [15]First Department of Internal Medicine, University of Toyama, Toyama 930-0194, Japan, [16]Department of Medicine, Shiga University of Medical Science, Otsu, Shiga 520-2192, Japan, [17]Health Center, Keio University School of Medicine, Tokyo 160-8582, Japan, [18]For a Full List of Authors, see Supplementary Material and [19]Division of Diabetes, Endocrinology and Metabolism, Department of Internal Medicine, Kawasaki Medical School, Kurashiki, Okayama 701-0192, Japan, [20]Medical Court Life Care Clinic, Hiroshima 730-0012, Japan, [21]Department of Internal Medicine, Division of Metabolism and Endocrinology, St. Marianna University School of Medicine, Kawasaki, Kanagawa 216-8511, Japan, [22]Life Sciences Institute, [23]NUS Graduate School for Integrative Science and Engineering, [24]Department of Statistics and Applied Probability and [25]Department of Medicine, National University of Singapore, Singapore, Singapore, [26]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore, [27]Duke-National University of Singapore Graduate Medical School, Singapore, Singapore and [28]StaGen, Tokyo 111-0051, Japan

Received March 20, 2013; Revised and Accepted August 9, 2013

Although over 60 loci for type 2 diabetes (T2D) have been identified, there still remains a large genetic component to be clarified. To explore unidentified loci for T2D, we performed a genome-wide association study (GWAS) of 6 209 637 single-nucleotide polymorphisms (SNPs), which were directly genotyped or imputed using East Asian

*To whom correspondence should be addressed at: Department of Diabetes and Metabolic Diseases, University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo 115-0055, Japan. Tel: 81 358009161; Fax: 81 358009161; Email: thara-tky@umin.ac.jp
†These authors contributed equally to this work.

references from the 1000 Genomes Project (June 2011 release) in 5976 Japanese patients with T2D and 20 829 nondiabetic individuals. Nineteen unreported loci were selected and taken forward to follow-up analyses. Combined discovery and follow-up analyses (30 392 cases and 34 814 controls) identified three new loci with genome-wide significance, which were *MIR129-LEP* [rs791595; risk allele = A; risk allele frequency (RAF) = 0.080; $P = 2.55 \times 10^{-13}$; odds ratio (OR) = 1.17], *GPSM1* [rs11787792; risk allele = A; RAF = 0.874; $P = 1.74 \times 10^{-10}$; OR = 1.15] and *SLC16A13* (rs312457; risk allele = G; RAF = 0.078; $P = 7.69 \times 10^{-13}$; OR = 1.20). This study demonstrates that GWASs based on the imputation of genotypes using modern reference haplotypes such as that from the 1000 Genomes Project data can assist in identification of new loci for common diseases.

INTRODUCTION

T2D is a complex disease characterized by hyperglycemia resulting from impaired pancreatic β-cell function and a decreased action of insulin on target tissues (1,2). Familial aggregation and twin studies have shown that a genetic component plays a major role in the onset of T2D. Although there has been a marked increase in the identification of genetic loci for T2D, with over 60 of the discoveries made through genome-wide association studies (GWASs), it is estimated that at best 10% of the genetic component of T2D can be explained by the loci identified so far (3–5). The 1000 Genomes Project was founded with the aim of characterizing over 95% of variants with a minor allele frequency (MAF) >1% to provide a more extensive catalog of the genetic variations in major ethnic groups (6,7). We reasoned that we could take advantage of this recent advance to explore the remaining unidentified part of the genetic component of T2D by extending our search. This idea is supported by a recent article that reported that imputing the Wellcome Trust Case Control Consortium Phase I genotype data by using references from the 1000 Genomes Project detected loci for diabetes that were not identified in the original case–control study (8). In the meantime, genetic studies have failed to explain the reason why the prevalence of T2D differs across different ethnic groups. In light of the high and rising prevalence of T2D in East Asia (9), there is a huge need for exploring genetic variations vigorously in East Asian populations, although GWASs for T2D have been performed not only in European (10–15), but also in Asian (16–22) populations.

RESULTS

GWAS of typed and imputed SNPs using 1000 Genomes Project data

This is a three-stage study comprised of (i) a discovery GWAS, (ii) follow-up analysis and (iii) validation analysis. During the discovery GWAS, we obtained directly genotyped data with a 610K SNP array in 5976 Japanese patients with T2D and 20 829 subjects without diabetes. To increase the genome-wide coverage of genetic variations in this population, we imputed 10 811 164 SNPs of 572 East Asian haplotypes [194 Han Chinese in Beijing (CHB), 200 Chinese in Metropolitan Denver (CHS) and 178 Japanese in Tokyo (JPT)], with 194 generated by the 1000 Genomes Project (June 2011 release). We found that concordance between the directly genotyped data and the imputed genotypes was generally good (94.2%). We then tested the association with T2D for 6 209 637 typed and imputed SNPs that passed our quality control criteria. An outline of the present study is shown in Supplementary Material, Figure S1, and details of the methods used in each stage are given in Supplementary Material, Table S1. Based on the principal component analysis (PCA), there was no apparent heterogeneity in genetic background among the subjects of the discovery stage, and we observed no indication of a population stratification influencing the result of the discovery analysis; the genomic control inflation factor (λ) was 1.074 (adjusted for 1000 cases and 1000 controls: $\lambda_{1000} = 1.008$) (Supplementary Material, Fig. S2). As shown in Figure 1, 9 known loci (in black) reached genome-wide significance ($P < 5 \times 10^{-8}$), and we observed 39 loci with a P-value for association with T2D of $<1 \times 10^{-4}$, consisting of 19 previously reported and 20 unreported loci. We selected a top SNP at each of the 20 unreported loci to take forward to follow-up analyses.

Replication and validation of top SNPs selected in the discovery GWAS

A top SNP at each of the two unreported loci was taken forward to follow-up analysis by *in silico* (2799 cases and 3793 controls) and *de novo* (10 319 cases and 6795 controls) genotyping in East Asian populations (Supplementary Material, Table S2). By combining the results obtained in the first discovery GWAS and the second follow-up analyses (in total 19 094 cases and 31 417 controls), four novel SNPs reached genome-wide significance ($P < 5 \times 10^{-8}$), which were near *MIR129-LEP* (rs791595; $P = 5.46 \times 10^{-11}$; OR = 1.19), *GPSM1* (rs11787792; $P = 7.26 \times 10^{-11}$; OR = 1.18), *MRPS35* (rs7316898; $P = 7.36 \times 10^{-9}$, OR = 1.10) and *SLC16A13* (rs312457; $P = 2.15 \times 10^{-8}$; OR = 1.18) (Supplementary Material, Table S2). To validate these associations, the four top SNPs were subsequently genotyped further in another 11 298 cases and 3397 controls. We confirmed that *MIR129-LEP* (rs791595; risk allele = A; RAF = 0.08; $P = 2.55 \times 10^{-13}$; OR = 1.17), *GPSM1* (rs11787792; risk allele = A; RAF = 0.874 ; $P = 1.74 \times 10^{-10}$; OR = 1.15) and *SLC16A13* (rs312457; risk allele = G ; RAF = 0.078; $P = 7.69 \times 10^{-13}$; OR = 1.20) were associated with T2D, reaching a Bonferroni-adjusted P-value for significance of 8.05×10^{-9}, which is more stringent than the conventional cut-off of 5×10^{-8} (Table 1) (Fig. 2). In contrast, rs7316898 in *MRPS35* did not reach genome-wide significance ($P = 5.42 \times 10^{-5}$; OR = 0.94) after combining three-stage analyses. Rs147689733 in *TNKS2* also reached genome-wide significance ($P = 2.43 \times 10^{-9}$) (Supplementary Material, Table S2). However, conditioning for rs12219514 in the previously

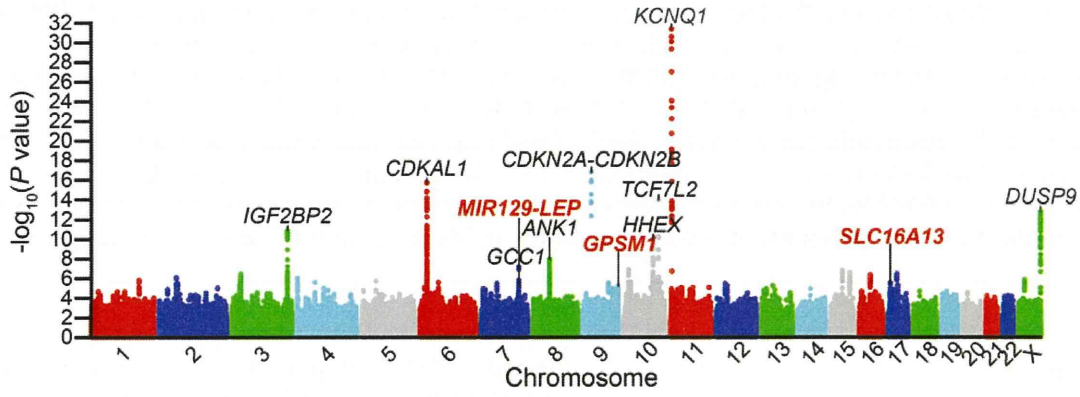

Figure 1. Manhattan plot for the discovery analysis of directly genotyped and imputed SNPs in 5976 T2D cases and 20 829 controls. Known loci that reached genome-wide significance ($P < 5 \times 10^{-8}$) in the discovery analysis (5976 T2D cases and 20 829 controls) are indicated in black and the three loci that reached genome-wide significance in the combined analysis of discovery and follow-up analyses 30 392 cases and 34 814 controls) are indicated in red.

reported locus *HHEX* (13,14) abrogated the signal at rs147689733, showing that it was a proxy for the stronger *HHEX* signal. Therefore, this SNP was not taken forward for validation analysis. To further confirm our imputation-based results, we performed additional direct genotyping of the three novel SNPs in a subset of the discovery GWAS samples that consisted of ~2600 samples and observed almost perfect concordance with the imputed data (Supplementary Material, Table S3). Combined with the direct genotyping performed in the other two analysis stages, the three novel SNPs were directly genotyped in over 34 000 samples.

Similarities and differences between East Asians and Europeans

To briefly assess the similarity of genetic architecture between East Asians and European populations, we compared the reported effect sizes of loci reported first in Europeans and then those obtained in the present study (Supplementary Material, Table S4). A high concordance in the direction of the effects and correlation of ORs between the two populations was observed ($r = 0.49$, $P = 0.0018$), showing that the two populations shared the same susceptibility genes when the data are restricted to common variants identified by previous GWASs. As for the novel loci identified in the present study, there is no evidence for an association between SNPs in and near *MIR129-LEP* (Supplementary Material, Fig. S3A) and *SLC16A13* (Supplementary Material, Fig. S3C) and only modest evidence for association around *GPSM1* (Supplementary Material, Fig. S3B) in European populations (15). Rs4731420, being in perfect linkage disequilibrium (LD) with rs791595 near *MIR129-LEP* ($r^2 = 1.00$), was not associated with T2D in the DIAGRAMv3 ($P = 0.760$). There are no available data on rs312457 in *SLC16A13* and its proxy in European populations. However, rs11652868, which was in modest LD with another SNP that was associated with T2D in the present study (rs312458), was not associated with T2D in the European data. The *GPSM1* locus has been reported to be linked to diabetes-related traits (23,24): rs3829109, located in the adjacent gene *DNLZ*, was previously associated with the fasting glucose level (23), and rs60980157, a nonsynonymous SNP (Ser391Leu of *GPSM1*), was previously reported to be associated with an insulin secretion measure, the insulinogenic index (24). These two SNPs were in modest LD with rs11787792 ($r^2 = 0.127$ for rs3829109 and 0.344 for rs60980157) and were associated with T2D ($P = 6.10 \times 10^{-6}$ and 9.03×10^{-6}, respectively) in this study, but we did not detect any association between rs11787792 and fasting glucose level or an insulin secretion index, HOMA-β (Supplementary Material, Table S5).

Comparison between association statistics of typed and imputed SNPs

Supplementary Material, Figure S4, demonstrates the usefulness of using imputation to explore previously unknown loci; imputed SNPs (grey circles) were more significantly associated with T2D than typed (red circles) SNPs. The signal at *GPSM1* would not have been taken forward to follow-up analysis, and therefore, it would have been missed (Supplementary Material, Fig. S4). Moreover, rs11787792 in *GPSM1* is included only in 1000 Genomes Project data and not in HapMap2 data. We also sought to define the most relevant SNPs for susceptibility to T2D in previously identified loci. We found rs7656416 in *CTBP1* ($P = 1.29 \times 10^{-8}$) to be more significantly associated with T2D than the previously reported SNP, rs6815464 in *MAEA* ($P = 1.32 \times 10^{-5}$) (20). Rs7656416 and rs6815464 were in LD ($r^2 = 0.58$), and conditioning for rs7656416 in the logistic regression abrogated the signal at rs6815464.

BMI-stratifying analysis

It has been reported that the genetic predisposition to T2D is different in lean subjects compared with findings for obese subjects (25,26). We tested the association between the top SNP in each of the eight autosomal known loci with $P < 5 \times 10^{-8}$ (adjusted for sex, age, and the first four principal components from PCA in the present discovery analysis) after dichotomizing the T2D subjects into lean (body mass index, BMI < 25 kg/m^2) and overweight (BMI ≥ 25 kg/m^2) groups. We found that the top SNPs in the known loci were more significantly associated with T2D

Table 1. Three new T2D loci reaching genome-wide significance from combined analysis

SNP	Chr	Chromosome position	Nearby gene	Risk allele	Other allele	RAF	First stage OR (95%CI)	*P*-value	Second + third stage OR (95%CI)	*P*-value	1st + 2nd + 3rd Stage OR (95%CI)	*P*-value
							5976 cases and 20 829 controls		24 416 cases and 13 985 controls		30 392 cases and 34 814 controls	
rs791595	7	127 862 802	*MIR129-LEP*	A	G	0.080	1.19 (1.11–1.28)	4.69E−06	1.16 (1.06–1.18)	5.64E−07	1.17 (1.12–1.22)	2.55E−13
rs11787792	9	13 92 52 148	*GPSM1*	A	G	0.874	1.17 (1.09–1.25)	7.12E−06	1.14 (1.06–1.18)	3.94E−06	1.15 (1.10–1.20)	1.74E−10
rs312457	17	69 40 393	*SLC16A13*	G	A	0.078	1.19 (1.10–1.29)	9.40E−06	1.20 (1.13–1.28)	1.04E−08	1.20 (1.14–1.26)	7.69E−13

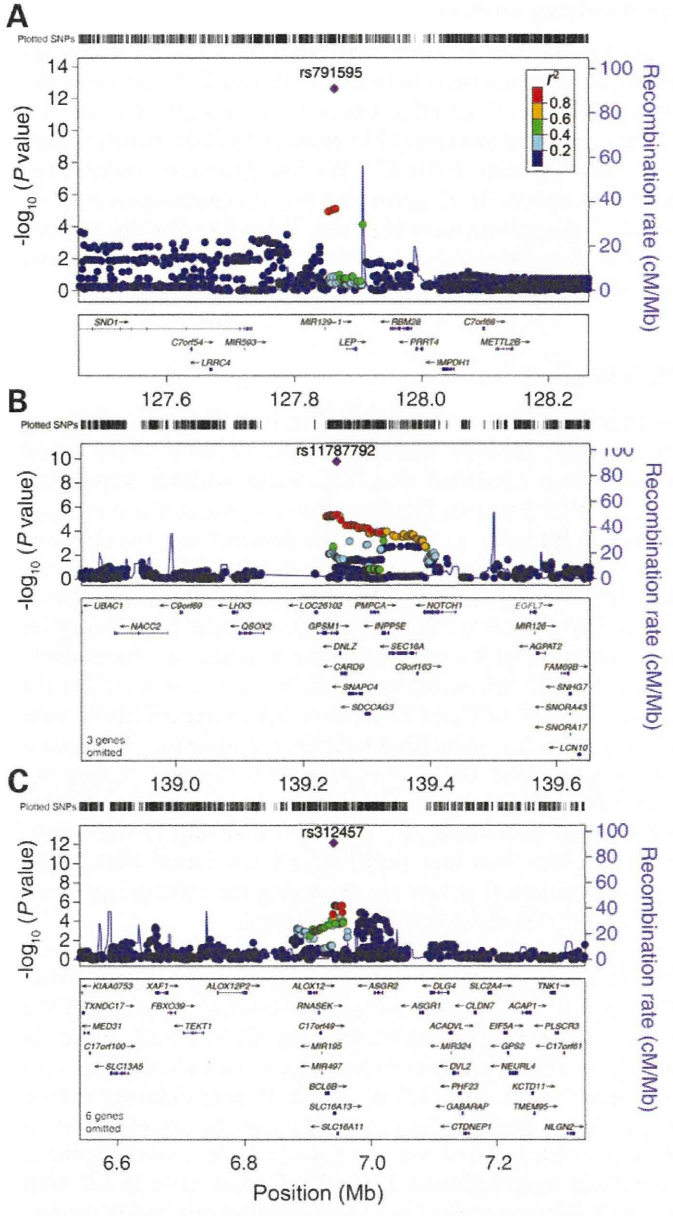

Figure 2. Regional plots of the three newly discovered T2D loci: (**A**) *MIR129-LEP,* (**B**) *GPSM1* and (**C**) *SLC16A13*. Regional associations were plotted for the three novel loci that showed genome-wide significance after combing the results from Stages 1, 2 and 3. Genotyped and imputed SNPs are plotted with the *P*-values (as −log10 values) from discovery analysis versus their physical position (NCBI Build 37). In each panel, the top SNP is represented by a purple diamond and *P*-values derived by combining first + second + third stage results are shown. *P*-values of other SNPs were derived from Stage 1 result alone and are color coded according to their pairwise LD with the top SNP based on 1000 Genomes Project East Asian reference data. Estimated recombination rates are plotted to reflect the local LD structure.

in lean subjects than in overweight subjects (Supplementary Material, Fig. S5). Several loci (*KCNQ1*, *CDC123, IGF2BP2* and *CDKAL1*) had large enough heterogeneity *z*-scores to suggest that a substantial difference may exist in the association statistics between lean and overweight groups (Supplementary Material, Table S6).

Sex-stratifying analysis

We performed a sex-differentiated analysis to test for sexual dimorphism of associations between SNPs and T2D, allowing for heterogeneity in allelic effects between males (3938 cases and 9553 controls) and females (1831 cases and 9220 controls) (Supplementary Material, Table S7). We found modest evidence for sexual dimorphism in 23 genes and regions (heterogeneity $P < 1 \times 10^{-4}$) (Supplementary Material, Table S7). We did not find evidence of sex heterogeneity at previously reported loci such as *KCNQ1, DGKB,* and *GRB14*, nor present novel loci.

DISCUSSION

We identified three novel T2D loci in East Asian populations: *MIR129-LEP*, *GPSM1* and *SLC16A13*. *GPSM1* locus could not have been identified as a T2D locus without imputation using the 1000 Genomes Project reference data, as shown in Supplementary Material, Figure S4, This demonstrates the utility of using up-to-date reference haplotype data such as that from the 1000 Genomes Project to perform imputation to identify novel loci for T2D. Given the fact that whole-genome sequencing for thousands of samples is still highly cost intensive, an imputation-based GWAS is still useful to search for novel common disease loci. At *MIR129-LEP* and *SLC16A13* loci, imputed SNPs were more significantly associated with type 2 diabetes (T2D) than directly genotyped SNPs, but imputation was not mandatory for the identification of these loci, because a directly genotyped SNP reached the *P*-value of 1×10^{-4}, the cut-off for follow-up analysis. Those two loci could have been found because of improved statistical power by increasing the sample size from ~7000 to 27 000 in the first-stage analysis.

We found no association between the three novel genes identified in this study and T2D in European populations (DIAGRAMv3) (15). We think there could be three sources for the lack of evidence for associations in DIAGRAM. First, in general, the difference in allele frequencies would cause discrepancy between the association results among different ethnic groups. A lower allele frequency reduces statistical power to detect association. However, we found no apparent differences in the allele frequencies of the top SNPs and those in LD with the top SNP between the East Asian populations and European populations. Taking into account that the DIAGRAM conducted the largest GWAS for T2D in European populations covering ~2.5 million SNPs, it is unlikely that the DIAGRAM did not find our loci due to the lack of statistical power. Secondly and most likely is dissimilarities in the patterns of LD among populations, which could lead to a substantial difference in the strength of LD between a causal variant and its proxy SNP between our population and European populations. While in *GPSM1* locus, which showed modest association with T2D in DIAGRAMv3 (Supplementary Material, Fig. S3B), CEU and JPT + CHB show quite similar haplotype structure using HapMap Project data (Supplementary Material, Fig. S6B), while the haplotype structure around the other two loci appears to be notably different between the two population samples (Supplementary Material, Fig. S6A and S6C). In particular, the region around *SLC16A13* shows a large block in which many SNPs are in complete LD ($D' = 1$). Further examination using the HapMap project's phased haplotype browser shows a single high-frequency haplotype extending roughly 20 kb in each direction from rs321457 in the JPT or CHB haplotypes that is not present in the CEU haplotypes (data not shown). Lastly and least likely, it is possible that a difference in effect size could lead to a situation where a disease locus would be discovered in a particular group with a strong effect. But no such locus, at least for T2D, has been reported so far, and effect sizes generally appear to be comparable among different ethnic groups. Further study will be needed to clarify whether three loci detected in this study are specific to East Asians.

Of the three newly identified loci, rs791595 is located between *MIR129-1* and *LEP*. The coding product of *LEP*, leptin, plays a critical role in the regulation of body weight by inhibiting food intake and stimulating energy expenditure, and its deficiency in mice and humans causes morbid obesity and diabetes (27,28). Leptin is a hormone produced and secreted by white adipose tissue, and its circulating levels are closely related to body fat mass (29). Thus, *LEP* is one of the most plausible candidates for a susceptibility gene of obesity and T2D, although we could not find any association with BMI in the controls, and adjustment for BMI did not influence the strength of the association with T2D ($P = 9.42 \times 10^{-9}$). Instead, we did find rs791595 to be significantly associated with the homeostasis model assessment of insulin resistance (HOMA-IR) index (30) ($P = 0.005$) (Supplementary Material, Table S5). As for *MIR129–1*, miR129, the mature product of *MIR129–1*, has been reported to be up-regulated in bladder cancers in accordance with the down regulation of its targeted genes *SOX4* and *GALNT1*, which are involved in cell death processes (31). However, whether miR129 has any role in tissues relevant to T2D is unknown. The Genomic Evolutionary Rate Profiling scores for the position of rs791595 near *MIR129-LEP* is 2.96, indicating that this site may be under evolutionary constraint.

The coding product of G-protein signaling modulator 1 isoform *(GPSM1)* influences the basal activity of G-protein signaling systems through interaction with G-protein subunits (32). *Gpsm1* null mice have a lean phenotype with reduced fat mass and increased nocturnal energy expenditure (33), suggesting that *GPSM1* is a biologically plausible obesity gene. We could not find any association with BMI ($P = 0.11$) and adjustment for BMI did not influence the strength of the association with T2D.

SLC16A13 encodes solute carrier family 16, member 13, which is one of the monocarboxylate transporters (MCTs) (34). The first four MCTs (MCT1-4) catalyze the transport of monocarboxylates, such as lactate and pyruvate, but the functions of MCT13 (encoded by *SLC16A13*) and MCT11 (encoded by *SLC16A11,* which is located adjacent to *SLC16A13*) are unknown, except for one report that showed that intestinal expression of *SLC16A13* was up-regulated by peroxisome proliferator-activated receptor-α agonists (35). Rs312457 in *SLC16A13* is in perfect LD with rs17203120 (RAF = 0.079; $P = 6.71 \times 10^{-6}$, OR = 1.19), which is located between *SLC16A13* and *SLC16A11*. There is evidence that rs17203120 is associated with the expression level of *SLC16A11* in lymphoblastoid cell lines (36).

There was no low-frequency variant with MAF ranging from 1 to 5% that reached genome-wide significance among 967 419 variants tested for association with T2D in this study. The present study yielded ~80% power to detect variants with an

MAF of 1% and OR of 1.6 and retained 80% power to detect association for variants with an MAF of 1% and OR of 2.0 even when imputed with a minimum quality metric r^2 of 0.3, which was sufficient for detecting low-frequency variants with a relatively large effect. Further studies that use newer data from the 1000 Genomes Project or large-scale meta-analyses across populations will be required to build on these results and to further elucidate the global genetic architecture of T2D.

MATERIALS AND METHODS

Study design

The present study was comprised of a three-stage analysis: a discovery stage (first stage), *de novo* genotyping and *in silico* follow-up (second stage), and *de novo* genotyping (third stage) analysis. The discovery stage was a GWAS that tested directly genotyped and imputed variants for association with T2D. We phased SNPs from the Illumina 610K SNP array and performed imputation using the 1000 Genomes East Asian reference panel. We selected 39 loci that had a top SNP with a $P < 1 \times 10^{-4}$ (adjusted for age, sex, BMI and the first four principal components from PCA), and filtered out 19 loci that were located in or near genes that were found to be associated with T2D in previous GWAS. The remaining 20 top SNPs were subsequently investigated for association with T2D by *de novo* genotyping analysis or *in silico* imputation analysis. We combined the results of the discovery and follow-up analyses to identify new T2D loci. An outline of the study design is shown in Supplementary Material, Figure S1.

Subjects

Subjects for the discovery analysis [BioBank Japan (BBJ) 1; 5976 cases with T2D and 20 829 controls] were recruited at several medical institutions in Japan (37,38). The follow-up analysis consisted of five studies, which were BBJ2, Shanghai Jiao Tong University (SJTU), Singapore Diabetes Cohort Study (SDCS)/Singapore Prospective Study Program (SP2), Singapore Chinese Eye Study (SCES) and Chinese University of Hong Kong (CUHK). We obtained a total of 2799 cases and 3793 controls for *in silico* follow-up analysis and a total of 10 319 cases and 6795 controls for the follow-up analysis by *de novo* genotyping. Samples in the validation analysis included an independent case–control sample (11 298 cases and 3397 controls) from subjects enrolled in the BioBank Japan (BBJ3). T2D cases were from individuals registered as having T2D. Diabetes was diagnosed according to the WHO criteria. The exclusion criteria for cases were individuals positive for antibody to glutamic acid decarboxylase (GAD) or those with diabetes due to liver dysfunction, steroids and other drugs that might raise glucose levels, malignancy or monogenic disorder known to cause diabetes. Control subjects were healthy volunteers confirmed at annual health check-up or individuals not having T2D but with diseases other than T2D: bronchial asthma, myocardial infarction, breast cancer, Basedow's disease, cerebral infarction, cerebral aneurism, osteoporosis, heart failure, unstable angina, pollinosis, arteriosclerosis obliterance, emphysema or atopic dermatitis. Note that BBJ1, BBJ2 and BBJ3 did not overlap for samples. Ethnicity was self-reported by the enrolled individuals. For each study, approval was obtained from the appropriate institutional review boards of the participating institutions, and a written informed consent was obtained from all participants.

Genotyping and imputation: in the discovery (BBJ1) and *in silico* follow-up analyses (SDCS/SP2, SCES, and CUHK), genotyping was done with genome-wide SNP arrays. In the *de novo* follow-up analysis, genotyping was carried out by using a multiplex polymerase chain reaction invader assay (BBJ2) and Mass ARRAY (SJTU). The typing platforms and quality control methods for each study are described in Supplementary Material, Table S1. We included SNPs from the SNP array for imputation and the association analysis with a call rate of $\geq$0.99 and a Hardy–Weinberg equilibrium (HWE) $P \geq 1 \times 10^{-6}$, selecting 6 209 637 imputed SNPs with an MAF of ≥ 0.01 and r^2 higher than a set of MAF specific thresholds as described in the previous study: MAF 0–0.1 = 0.75, MAF 0.1–0.2 = 0.70, MAF 0.2–0.3 = 0.66, MAF 0.3–0.4 = 0.60, MAF 0.4–0.5 = 0.55 for the discovery analysis. In the discovery and *in silico* analyses, SNP imputation was done using 572 East Asian haplotypes (194 CHB, 200 CHS and 178 JPT) from the 1000 Genomes Project data (June 2011 release).

eQTL

Potential candidates for association with T2D were pursued with eQTL studies in available datasets in lymphoblastoid cell lines (36) and a liver tissue gene expression database (39).

Statistical analysis

Associations between SNPs and T2D were tested by logistic regression analysis using an additive model with or without adjustment for age, sex, BMI and the first four principal components from PCA. A quantile–quantile plot was constructed by plotting the distribution of the observed *P*-values for the SNPs against the theoretical distribution of the expected *P*-values for T2D. In the discovery analysis, the genomic control inflation factor (λ) was calculated as the median χ^2 statistic divided by 0.456. Meta-analysis was performed by an inverse variance method assuming fixed effects using *R* software. Quantitative trait analyses were done for BMI, FPG, HbA1c, log-transformed HOMA-β, log-transformed HOMA-IR, total cholesterol, HDL cholesterol and triglycerides by multiple linear regression analysis, employing an additive association model with or without adjustment for the relevant covariates. The power of detecting previously reported loci in the present study was estimated by using QUANTO, employing the RAF and the sample sizes in the discovery stage, the reported ORs, an assumed T2D prevalence of 10%, and $\alpha = 0.05$. For the analysis of lean versus obese subjects, samples were dichotomized into lean (BMI < 25) and obese (BMI ≥ 25) groups and SNPs analyzed for each group using a logistic regression base model adjusted for gender, age and four principal components. Heterogeneity between lean and obese samples was calculated as a *z*-score using the beta and standard error (s.e.) from the logistic regression estimates as: $z = (\text{beta.lean} - \text{beta.obese})/\text{sqrt}\ (\text{s.e.lean}^2 + \text{s.e.obese}^2)$. To test whether the reduced sample size of the obese sample group could affect its observed decrease in significance, we performed a re-sampling-based analysis, whereby samples with

sizes matching that of the obese case and control sizes were drawn without replacement 10 000 times from the complete set of samples. We estimated the probability of observing the obese group's *P*-value due to random sample size differences using R's empirical distribution function.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

B.B.J.: We thank all the participants and the staff of the BioBank Japan project. We thank all participating doctors and staff from collaborating institutes for providing DNA samples. We also thank the technical staff of the Laboratory for Endocrinology and Metabolism at the RIKEN Center for Genomic Medicine for providing the technical assistance. Likewise, we thank the technical staff of the Laboratory for Genotyping Development at RIKEN Center for Genomic Medicine for performing SNP genotyping.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by a grant from the Leading Project of Ministry of Education, Culture, Sports, Science and Technology, Japan. S.J.T.U.: We thank all medical staff of the Shanghai Clinical Center for Diabetes. This work was supported by grants from the National 973 Program (2011CB504001), 863 Program (2006AA02A409), National Science Foundation of China (30800617, 81170735), Excellent Young Medical Expert of Shanghai (XYQ2011041) and the major program of the Shanghai Municipality for Basic Research (08dj1400601), China. SP2: The Singapore BioBank and the Genome Institute of Singapore, Agency for Science, Technology, and Research provided services for tissue archival and genotyping, respectively. This work was supported by grants from the Biomedical Research Council of Singapore (BMRC 05/1/36/19/413 and 03/1/27/18/216) and the National Medical Research Council of Singapore (NMRC/1174/2008).C.U.H.K.: We thank all medical and nursing staff of the Prince of Wales Hospital Diabetes Mellitus Education Centre for their commitment and professionalism. We would like to thank the Genome Institution at Quebec for help with replication genotyping, and the Chinese University of Hong Kong Information Technology Services Centre for their support with computing resources. This work was supported by the Hong Kong Foundation for Research and Development in Diabetes established under the auspices of the Chinese University of Hong Kong, the Hong Kong Government Research Grant Committee Central Allocation Scheme (CUHK 1/04C), Research Grants Council Earmarked Research Grant (CUHK4727/0M), the Innovation and Technology Fund (ITS/088/08 and ITS/487/09FP), a Chinese University Direct Grant, the Research Fund of the Department of Medicine and Therapeutics and the Diabetes and Endocrine Research Fund of the Chinese University of Hong Kong. The summary results of DIAGRAMv3 study were downloaded via the website at http://diagram-consortium.org/index.html.

REFERENCES

1. Ashcroft, F.M. and Rorsman, P. (2012) Diabetes mellitus and the β cell: the last ten years. *Cell*, **148**, 1160–1171.
2. Samuel, V.T. and Shulman, G.I. (2012) Mechanisms for insulin resistance: common threads and missing links. *Cell*, **148**, 852–871.
3. McCarthy, M.I. (2010) Genomics, type 2 diabetes, and obesity. *N. Engl. J. Med.*, **363**, 2339–2350.
4. Prokopenko, I., McCarthy, M.I. and Lindgren, C.M. (2008) Type 2 diabetes: new genes, new understanding. *Trends Genet.*, **24**, 613–621.
5. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A. *et al.* (2009) Finding the missing heritability of complex diseases. *Nature*, **461**, 747–753.
6. 1000 Genomes Project ConsortiumAbecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E. and McVean, G.A. The 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061–1073.
7. 1000 Genomes Project ConsortiumAbecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. and McVean, G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.
8. Huang, J., Ellinghaus, D., Franke, A., Howie, B. and 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.
9. Yang, W., Lu, J., Weng, J., Jia, W., Ji, L., Xiao, J., Shan, Z., Liu, J., Tian, H., Ji, Q. *et al.* (2010) Prevalence of diabetes among men and women in China. *N. Engl. J. Med.*, **362**, 1090–1101.
10. Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., Boutin, P., Vincent, D., Belisle, A., Hadjadj, S. *et al.* (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*, **445**, 881–885.
11. Scott, L.J., Mohlke, K.L., Bonnycastle, L.L., Willer, C.J., Li, Y., Duren, W.L., Erdos, M.R., Stringham, H.M., Chines, P.S., Jackson, A.U. *et al.* (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*, **316**, 1341–1345.
12. Wellcome Trust Case Control Consortium. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661–678.
13. Zeggini, E., Scott, L.J., Saxena, R., Voight, B.F., Marchini, J.L., Hu, T., de Bakker, P.I., Abecasis, G.R., Almgren, P., Andersen, G. *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.
14. Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., Zeggini, E., Huth, C., Aulchenko, Y.S., Thorleifsson, G. *et al.* (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.*, **42**, 579–589.
15. Morris, A.P., Voight, B.F., Teslovich, T.M., Ferreira, T., Segrè, A.V., Steinthorsdottir, V., Strawbridge, R.J., Khan, H., Grallert, H., Mahajan, A. *et al.* (2012) Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.*, **44**, 981–990.
16. Yasuda, K., Miyake, K., Horikawa, Y., Hara, K., Osawa, H., Furuta, H., Hirota, Y., Mori, H., Jonsson, A., Sato, Y. *et al.* (2008) Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat. Genet.*, **40**, 1092–1097.
17. Unoki, H., Takahashi, A., Kawaguchi, T., Hara, K., Horikoshi, M., Andersen, G., Ng, D.P., Holmkvist, J., Borch-Johnsen, K., Jørgensen, T. *et al.* (2008) SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asians and European populations. *Nat. Genet.*, **40**, 1098–1102.
18. Yamauchi, T., Hara, K., Maeda, S., Yasuda, K., Takahashi, A., Horikoshi, M., Nakamura, M., Fujita, H., Grarup, N., Cauchi, S. *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.
19. Kooner, J.S., Saleheen, D., Sim, X., Sehmi, J., Zhang, W., Frossard, P., Been, L.F., Chia, K.S., Dimas, A.S., Hassanali, N. *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.
20. Cho, Y.S., Chen, C.H., Hu, C., Long, J., Ong, R.T., Sim, X., Takeuchi, F., Wu, Y., Go, M.J., Yamauchi, T. *et al.* (2011) Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in East Asians. *Nat. Genet.*, **44**, 67–72.
21. Imamura, M., Maeda, S., Yamauchi, T., Hara, K., Yasuda, K., Morizono, T., Takahashi, A., Horikoshi, M., Nakamura, M., Fujita, H. *et al.* (2012) A single-nucleotide polymorphism in ANK1 is associated with susceptibility to type 2 diabetes in Japanese populations. *Hum. Mol. Genet.*, **21**, 3042–3049.
22. Saxena, R., Saleheen, D., Been, L.F., Garavito, M.L., Braun, T., Bjonnes, A., Young, R., Ho, W.K., Rasheed, A., Frossard, P. *et al.* (2013) Genome-wide association study identifies a novel locus contributing to type 2 diabetes susceptibility in sikhs of punjabi origin from India. *Diabetes*, **62**, 1746–1755.
23. Scott, R.A., Lagou, V., Welch, R.P., Wheeler, E., Montasser, M.E., Luan, J., Mägi, R., Strawbridge, R.J., Rehnberg, E., Gustafsson, S. *et al.* (2012) Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet.*, **44**, 991–1005.
24. Huyghe, J.R., Jackson, A.U., Fogarty, M.P., Buchkovich, M.L., Stančáková, A., Stringham, H.M., Sim, X., Yang, L., Fuchsberger, C., Cederberg, H. *et al.* (2013) Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat. Genet.*, **45**, 197–201.
25. Perry, J.R., Voight, B.F., Yengo, L., Amin, N., Dupuis, J., Ganser, M., Grallert, H., Navarro, P., Li, M., Qi, L. *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.
26. Timpson, N.J., Lindgren, C.M., Weedon, M.N., Randall, J., Ouwehand, W.H., Strachan, D.P., Rayner, N.W., Walker, M., Hitman, G.A., Doney, A.S. *et al.* (2009) Adiposity-related heterogeneity in patterns of type 2 diabetes susceptibility observed in genome-wide association data. *Diabetes*, **58**, 505–510.
27. Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L. and Friedman, J.M. (1994) Positional cloning of the mouse obese gene and its human homologue. *Nature*, **372**, 425–432.
28. Montague, C.T., Farooqi, I.S., Whitehead, J.P., Soos, M.A., Rau, H., Wareham, N.J., Sewter, C.P., Digby, J.E., Mohammed, S.N., Hurst, J.A. *et al.* (1997) Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature*, **387**, 903–908.
29. Maffei, M., Halaas, J., Ravussin, E., Pratley, R.E., Lee, G.H., Zhang, Y., Fei, H., Kim, S., Lallone, R., Ranganathan, S. *et al.* (1995) Leptin levels in human and rodent: measurement of plasma leptin and ob RNA in obese and weight-reduced subjects. *Nat. Med.*, **1**, 1155–1161.
30. Matthews, D.R., Hosker, J.P., Rudenski, A.S., Naylor, B.A., Treacher, D.F. and Turner, R.C. (1985) Homeostasis model assessment: insulin resistance and B cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia*, **28**, 412–419.
31. Dyrskjøt, L., Ostenfeld, M.S., Bramsen, J.B., Silahtaroglu, A.N., Lamy, P., Ramanathan, R., Fristrup, N., Jensen, J.L., Andersen, C.L., Zieger, K. *et al.* (2009) Genomic profiling of microRNAs in bladder cancer: miR-129 is associated with poor outcome and promotes cell death in vitro. *Cancer Res.*, **69**, 4851–4860.
32. Pizzinat, N., Takesono, A. and Lanier, S.M. (2001) Identification of a truncated form of the G-protein regulator AGS3 in heart that lacks the tetratricopeptide repeat domains. *J. Biol. Chem.*, **276**, 16601–16610.
33. Blumer, J.B., Lord, K., Saunders, T.L., Pacchioni, A., Black, C., Lazartigues, E., Varner, K.J., Gettys, T.W. and Lanier, S.M. (2008) Activator of G protein signaling 3 null mice: unexpected alterations in metabolic and cardiovascular function. *Endocrinology*, **149**, 3842–3849.
34. Halestrap, A.P. and Meredith, D. (2004) The SLC16 gene family-from monocarboxylate transporters (MCTs) to aromatic amino acid transporters and beyond. *Pflugers Arch.*, **447**, 619–628.
35. Hirai, T., Fukui, Y. and Motojima, K. (2007) PPAR Alpha agonists positively and negatively regulate the expression of several nutrient/drug transporters in mouse small intestine. *Biol. Pharm. Bull.*, **30**, 2185–2190.
36. Veyrieras, J.B., Kudaravalli, S., Kim, S.Y., Dermitzakis, E.T., Gilad, Y., Stephens, M. and Pritchard, J.K. (2008) High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.*, **4**, e1000214.
37. Okada, Y., Sim, X., Go, M.J., Wu, J.Y., Gu, D., Takeuchi, F., Takahashi, A., Maeda, S., Tsunoda, T., Chen, P. *et al.* (2012) Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nat. Genet.*, **44**, 904–909.
38. Okada, Y., Kubo, M., Ohmiya, H., Takahashi, A., Kumasaka, N., Hosono, N., Maeda, S., Wen, W., Dorajoo, R., Go, M.J. *et al.* (2012) Common variants at CDKAL1 and KLF9 are associated with body mass index in east Asian populations. *Nat. Genet.*, **44**, 302–306.
39. Schadt, E.E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P.Y., Kasarskis, A., Zhang, B., Wang, S., Suver, C. *et al.* (2008) Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.*, **6**, e107.

nature genetics

Genetic variants in *GPR126* are associated with adolescent idiopathic scoliosis

Ikuyo Kou[1], Yohei Takahashi[2], Todd A Johnson[3], Atsushi Takahashi[4], Long Guo[5], Jin Dai[6], Xusheng Qiu[7], Swarkar Sharma[8,21], Aki Takimoto[5], Yoji Ogura[1,2], Hua Jiang[7], Huang Yan[7], Katsuki Kono[9], Noriaki Kawakami[10], Koki Uno[11], Manabu Ito[12], Shohei Minami[13], Haruhisa Yanagida[14], Hiroshi Taneichi[15], Naoya Hosono[16], Taichi Tsuji[10], Teppei Suzuki[11], Hideki Sudo[12], Toshiaki Kotani[13], Ikuho Yonezawa[17], Douglas Londono[18], Derek Gordon[18], John A Herring[19,20], Kota Watanabe[2], Kazuhiro Chiba[2], Naoyuki Kamatani[4], Qing Jiang[6], Yuji Hiraki[5], Michiaki Kubo[16], Yoshiaki Toyama[2], Tatsuhiko Tsunoda[3], Carol A Wise[8,20], Yong Qiu[7], Chisa Shukunami[5], Morio Matsumoto[2] & Shiro Ikegawa[1]

Adolescent idiopathic scoliosis (AIS) is the most common pediatric skeletal disease[1]. We previously reported a locus on chromosome 10q24.31 associated with AIS susceptibility in Japanese using a genome-wide association study (GWAS) consisting of 1,033 cases and 1,473 controls[2]. To identify additional AIS-associated loci, we expanded the study by adding X-chromosome SNPs in the GWAS and increasing the size of the replication cohorts. Through a stepwise association study including 1,819 cases and 25,939 controls, we identified a new susceptibility locus on chromosome 6q24.1 in Japanese ($P = 2.25 \times 10^{-10}$; odds ratio (OR) = 1.28). The most significantly associated SNP, rs6570507, was in *GPR126* (encoding G protein–coupled receptor 126). Its association was replicated in Han Chinese and European-ancestry populations (combined $P = 1.27 \times 10^{-14}$; OR = 1.27). *GPR126* was highly expressed in cartilage, and the knockdown of *gpr126* in zebrafish caused delayed ossification of the developing spine. Our results should provide insights into the etiology and pathogenesis of AIS.

AIS is a complex rotational spinal deformity that occurs during the pubertal growth spurt. The etiology of AIS remains largely unknown; however, many clinical and genetic studies suggest a contribution of genetic factors to the development of AIS[3,4]. We previously conducted a GWAS of Japanese females with AIS, identifying a single locus on chromosome 10q24.31 with an association that surpassed the genome-wide significance level of $P < 5 \times 10^{-8}$ (ref. 2). The association of this locus was subsequently replicated in two independent Chinese populations with $P = 9.1 \times 10^{-10}$ and 1.8×10^{-9} for rs11190870 (refs. 5,6). Thus, the chromosome 10q24.31 locus has provided the most compelling evidence of association with AIS so far.

To identify additional AIS susceptibility loci, we extended our previous GWAS (**Supplementary Fig. 1**) by including 10,641 X-chromosome SNPs. We evaluated the quality of genotyping data as previously described[2]. In addition to three SNPs located on chromosome 10q24.31 with associations that reached a genome-wide significance threshold as previously described[2], six SNPs yielded suggestive evidence of association ($P < 1 \times 10^{-5}$; **Supplementary Fig. 2**). By excluding SNPs within the same locus ($r^2 > 0.8$) from the six SNPs, we selected three SNPs for a replication study. We genotyped the SNPs in an independent set of Japanese subjects (786 cases and 24,466 controls) and found evidence for association with one SNP, rs6570507 on chromosome 6q24.1, with $P = 3.02 \times 10^{-5}$ (Bonferroni-corrected $P < 1.67 \times 10^{-2}$; **Table 1** and **Supplementary Table 1**). Combining the results from the GWAS and replication study, rs6570507 reached a genome-wide significance threshold of $P < 5 \times 10^{-8}$ (combined $P = 2.25 \times 10^{-10}$; OR = 1.28, 95% confidence interval (CI) = 1.18–1.38; **Table 1**).

[1]Laboratory for Bone and Joint Diseases, Center for Integrative Medical Sciences, RIKEN, Tokyo, Japan. [2]Department of Orthopaedic Surgery, School of Medicine, Keio University, Tokyo, Japan. [3]Laboratory for Medical Science Mathematics, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan. [4]Laboratory for Statistical Analysis, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan. [5]Department of Cellular Differentiation, Institute for Frontier Medical Sciences, Kyoto University, Kyoto, Japan. [6]The Center of Diagnosis and Treatment for Joint Disease, the Affiliated Drum Tower Hospital of Nanjing University Medical School, Nanjing, China. [7]Department of Spine Surgery, the Affiliated Drum Tower Hospital of Nanjing University Medical School, Nanjing, China. [8]Sarah M. and Charles E. Seay Center for Musculoskeletal Research, Texas Scottish Rite Hospital for Children, Dallas, Texas, USA. [9]Scoliosis Center, Saiseikai Central Hospital, Tokyo, Japan. [10]Department of Orthopaedic Surgery, Meijo Hospital, Nagoya, Japan. [11]Department of Orthopaedic Surgery, National Hospital Organization, Kobe Medical Center, Kobe, Japan. [12]Department of Advanced Medicine for Spine and Spinal Cord Disorders, Hokkaido University Graduate School of Medicine, Sapporo, Japan. [13]Department of Orthopaedic Surgery, Seirei Sakura Citizen Hospital, Sakura, Japan. [14]Department of Orthopaedic Surgery, Fukuoka Children's Hospital, Fukuoka, Japan. [15]Department of Orthopaedic Surgery, Dokkyo Medical University School of Medicine, Mibu, Japan. [16]Laboratory for Genotyping Development, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan. [17]Department of Orthopaedic Surgery, Juntendo University School of Medicine, Tokyo, Japan. [18]Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, New Jersey, USA. [19]Department of Orthopedics, Texas Scottish Rite Hospital for Children, Dallas, Texas, USA. [20]Department of Orthopedic Surgery, University of Texas Southwestern Medical Center at Dallas, Dallas, Texas, USA. [21]Present address: School of Biology and Chemistry, Shri Mata Vaishno Devi University, Katra, India. Correspondence should be addressed to M.M. (morio@a5.keio.jp) or S.I. (sikegawa@ims.u-tokyo.ac.jp).

Received 20 November 2012; accepted 19 April 2013; published online 12 May 2013; doi:10.1038/ng.2639

Table 1 Association of rs6570507 with AIS

Population	Study	Number of samples		RAF				
		Case	Control	Case	Control	*P* value[a]	OR (95% CI)[b]	P^{c}_{het}
Japanese	GWAS	1,033	1,473	0.49	0.42	1.37×10^{-6}	1.32 (1.18–1.48)	
	Replication	786	24,466	0.48	0.43	3.02×10^{-5}	1.24 (1.12–1.37)	
	Japanese combined[d]	1,819	25,939			2.25×10^{-10}	1.28 (1.18–1.38)	0.40
Chinese	Replication	743	1,209	0.39	0.35	3.36×10^{-3}	1.22 (1.06–1.39)	
	East Asian combined[d]	2,562	27,148			3.90×10^{-12}	1.26 (1.18–1.35)	0.58
European ancestry	Replication	447	737	0.36	0.29	4.83×10^{-4}	1.37 (1.14–1.63)	
	All combined[d]	3,009	27,885			1.27×10^{-14}	1.27 (1.20–1.35)	0.62

RAF, risk allele frequency.
[a]*P* values from the Cochran-Armitage trend test. [b]OR estimates for the risk allele from a two-by-two allele frequency table. [c]Results of the Breslow-Day test; *P* values for heterogeneity. [d]*P* values were calculated by the Mantel-Haenszel method.

As a further test of this association, we genotyped rs6570507 in a Han Chinese population (743 cases and 1,209 controls) and a population of European ancestry (447 cases and 737 controls). The association was replicated in these distinct populations. The Breslow-Day test identified no significant heterogeneity ($P > 0.05$) between studies for this SNP. The overall meta-analysis combined *P* value was 1.27×10^{-14} (**Table 1**).

To investigate additional susceptibility loci, we imputed genotypes across the genome using reference haplotypes from the 1000 Genomes Project East Asian population sample (EAS) and tested association with AIS. We demarcated local peaks in the association statistics, thereby grouping SNPs into putative loci, identified the top SNP in each locus and examined linkage disequilibrium (LD) to confirm the absence of secondary signals. We found suggestive associations ($P < 1 \times 10^{-5}$) in 12 loci, but only rs11190870 on chromosome 10q24.31 reached the genome-wide significance threshold ($P < 5 \times 10^{-8}$). We excluded 4 investigated loci and genotyped the top SNPs in each of the remaining 8 loci in the Japanese replication cohort (786 cases and 3,422 controls); however, no SNP associations were replicated after Bonferroni correction (**Supplementary Table 2**).

To further characterize the chromosome 6q24.1 locus, we imputed additional genotypes and tested their associations with AIS (**Fig. 1a**). SNPs rs7774095 and rs6570507 yielded the strongest evidence for association; genotyping rs7774095 using GWAS samples provided a less significant association than that of rs6570507 ($P = 1.23 \times 10^{-5}$; OR = 1.29). Seventeen additional SNPs yielded evidence for association and were strongly correlated ($r^2 > 0.9$) with rs6570507 (**Fig. 1a**). All of these SNPs were located in introns of the *GPR126* gene encoding G protein–coupled receptor 126. Given that a previous analysis of height in individuals of European ancestry identified both primary and secondary association signals around *GPR126* (ref. 7), we performed an additional analysis of SNPs surrounding *GRP126* while conditioning on the top imputed SNP (rs7774095). There was no apparent secondary association signal for AIS within the region (**Fig. 1b**).

We investigated possible functional effects of associated SNPs by cross-referencing expression quantitative trait locus (eQTL) data for our chromosome 6q24.1 locus using the Genevar (GENe Expression VARiation)[8] and SNPExpress[9] databases (see URLs). None of the SNPs were significantly associated with *GPR126* expression levels. We also evaluated the overlap of associated SNPs with Encyclopedia of DNA Elements (ENCODE)-annotated genomic elements[10], using the UCSC Genome Browser (see URLs). Two SNPs (rs7774095 and rs9403380) intersected regions bound by multiple transcription factors in the cell lines investigated by the ENCODE Project. SNP rs9403380 also resided within DNase I–hypersensitive regions across multiple tissues. These findings suggest that these variants have the potential to regulate the transcriptional activity of *GPR126*.

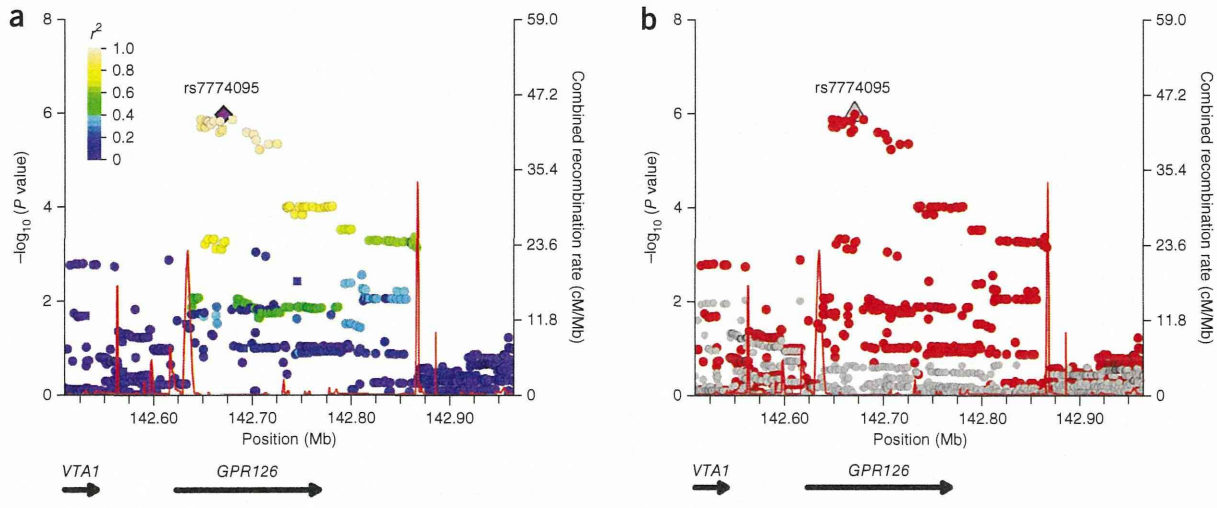

Figure 1 Regional association plots and recombination rates at the AIS susceptibility locus on chromosome 6q24.1. The $-\log_{10}$ (*P* values) from a logistic regression analysis plotted against the chromosome positions (NCBI Build 37) of SNPs are shown together with the positions and orientations of genes. The recombination rate estimate (red line) is based on 1000 Genomes Project EAS genotype data downloaded from the IMPUTE2 website. (**a**) Analysis without conditioning. The SNP with the highest association signal (rs7774095) is represented by a purple diamond. Imputed SNPs (circles) and genotyped SNPs (squares) are colored according to their LD (r^2) with rs7774095. (**b**) Analysis with conditioning. Red circles represent unconditioned associations and gray circles represent associations conditioned on the signal at rs7774095 (gray triangle).

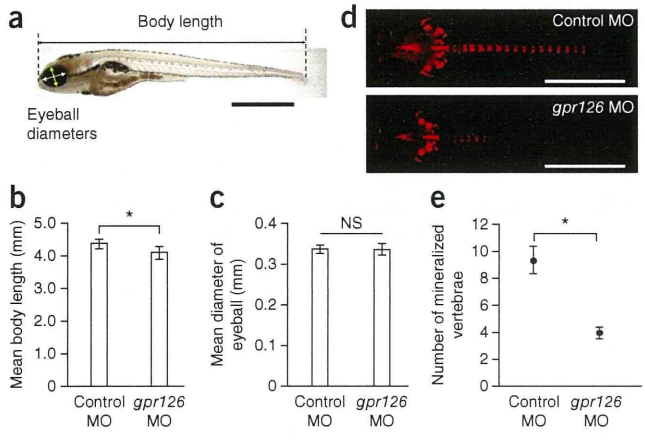

Figure 2 Phenotypes of *gpr126* zebrafish morphants. (**a**) Lateral view of a zebrafish with measurements of body length and eyeball diameter. Scale bar, 1 mm. (**b**,**c**) Body lengths (**b**) and eyeball diameters (**c**) of embryos with *gpr126* or control morpholino (MO) at 14 d post-fertilization (d.p.f.) (*gpr126* morphants and control embryos, $n = 15$ each; mean ± s.d.). (**d**,**e**) Alizarin red staining for the skeleton (**d**) and the number of mineralized vertebrae (**e**) at 14 d.p.f. (*gpr126* morphants and control embryos, $n = 10$ each; mean ± s.e.m.). Scale bars in **d**, 1 mm. Vertebral development is significantly delayed in *gpr126* morphants; $^*P < 0.01$; NS, not significant.

The GPR126 protein is an orphan receptor of the adhesion–G protein–coupled receptor (GPCR) family[11]. Previous studies in mice indicated that loss of *Sox9*, a key regulator of chondrogenesis, decreased the expression of *Gpr126* in intervertebral disc[12]. Because *GPR126* expression in humans has not been examined in skeletal tissues related to scoliosis, we examined the expression of *GPR126* mRNA using quantitative RT-PCR in various human tissues, including bone, cartilage and intervertebral disc. We found that *GPR126* was highly expressed in cartilage (**Supplementary Fig. 3**). We then analyzed the expression of *Gpr126* in mouse spine using *in situ* hybridization and found that *Gpr126* was expressed in the proliferating chondrocytes of the vertebral body at embryonic day (E) 16.5 (**Supplementary Fig. 4**).

We next assessed the effects of *GPR126* on skeletal development using morpholino-based knockdown experiments in zebrafish. Compared with controls, *gpr126* morphants had shorter body lengths and delayed ossification of the vertebrae (**Fig. 2**). In addition, *gpr126* morphants showed slower escape responses, suggesting possible neurological defects (data not shown). We found that the phenotype of *gpr126* morphants with a splice-blocking morpholino was indistinguishable from that of morphants with a translation-blocking morpholino (data not shown).

It is widely hypothesized that abnormal skeletal growth could induce scoliosis[13]. Several morphological studies have reported that individuals with AIS have imbalanced and faster longitudinal growth of the vertebral bodies, as well as extraspinal skeletal length asymmetries[14–16]. A recent study reported that *Gpr126*-null mice have limb posture abnormalities and growth failure[17]. GWAS of human height yielded associations with common SNPs in *GPR126* in both children and adults[18–21]. Notably, our most significantly associated SNP rs6570507 was also associated with trunk length in a GWAS meta-analysis of height in European populations[19]. We observed association in the same direction; the susceptibility allele is the same for AIS and for shortened trunk. These observations suggest that *GPR126* may affect both AIS susceptibility and height through abnormal spinal development and/or growth.

A neuropathic etiology is also widely hypothesized for AIS[22]. GPR126 also functions in nervous system control, and this process may contribute to AIS susceptibility. Several studies in zebrafish and mice have indicated that GPR126 is essential for myelination, a process necessary for proper nerve conduction velocity[17,23–25]. *Gpr126*-null mice show severe congenital hypomyelinating peripheral neuropathy[17]. The slower escape response we observed in *gpr126* zebrafish morphants might be related to such a myelination defect. Scoliosis is secondary to various neurological disorders in humans, including demyelinating peripheral neuropathies such as Charcot-Marie-Tooth syndrome[13,26]. It is noteworthy that *Gpr126*-null mice also present with joint contractures, a frequent comorbidity with scoliosis seen in humans with various disorders, including Marfan syndrome, congenital contractural arachnodactyly and arthrogryposis multiplex congenita[27,28].

Through meta-analyses of GWAS, it was reported that *GPR126* is associated with obstructive pulmonary dysfunction, indicated by the ratio of forced expiratory volume in the first second to forced vital capacity (FEV1/FVC)[29]. Pulmonary dysfunction is a clinical concern in individuals with AIS. Although the pulmonary dysfunction observed in these individuals is mostly restrictive dysfunction, it was reported that individuals with AIS also have obstructive dysfunction[30]. Thus, further studies are necessary to evaluate the relationship between *GPR126* and obstructive pulmonary dysfunction in AIS.

Through a stepwise association study, starting from an extended GWAS and followed by replication studies using Japanese, Chinese and European-ancestry populations, we identified a new AIS susceptibility locus on chromosome 6q24.1. *GPR126* in the locus is a promising candidate for AIS susceptibility, as zebrafish studies demonstrate its clear involvement in both the growth and ossification of the developing spine, as well as in neurological development. Further functional studies are necessary to elucidate how alterations in *GPR126* increase the risk of AIS in humans. We estimate that the two loci thus far associated with AIS susceptibility, on chromosomes 10q24 and 6q24, explain approximately 1% of the trait variance in AIS. Clearly, additional AIS risk factors await discovery. A global GWAS meta-analysis (consortium type) of existing AIS should be the next step.

URLs. The Leading Project for Personalized Medicine (in Japanese), http://biobankjp.org/; 1000 Genomes Project, http://www.1000genomes.org/; MACH, http://www.sph.umich.edu/csg/yli/mach/index.html; Minimac, http://genome.sph.umich.edu/wiki/Minimac; R v2.13.0, http://www.r-project.org/; splitPed.pl, http://www.sph.umich.edu/csg/yli/splitPed/ and http://genome.sph.umich.edu/wiki/SplitPed; Genevar, http://www.sanger.ac.uk/resources/software/genevar/; SNPExpress, http://compute1.lsrc.duke.edu/softwares/SNPExpress/; UCSC Genome Browser (ENCODE data), http://genome.ucsc.edu/ENCODE/.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank all participating individuals with AIS and the doctors and staff of the collaborating institutes. We especially thank N. Suzuki, M. Saito, M. Kamata, N. Hosogane, E. Okada and H. Hase for their help in collecting samples. We also thank S. Tominaga and T. Isono for technical assistance. This work was supported by a grant-in-aid to M.M. from the Japanese Orthopaedic Association (JOA–Subsidized Science Project Research 2009-1) and by funding from the US National Institutes of Health (R01052973), the Crystal Charity Ball, the Scoliosis Research Society and the Texas Scottish Rite Hospital Research Fund to C.A.W.

AUTHOR CONTRIBUTIONS

I.K., Y. Takahashi, M.M., K.C. and S.I. designed the study. M.K. and N. Kamatani contributed to overall GWAS design. I.K., M.M. and S.I. drafted the manuscript. T.A.J., A. Takahashi and T. Tsunoda analyzed the GWAS data. I.K., Y. Takahashi, N.H. and M.K. performed the genotyping for the GWAS and replication study. M.M., K.K., K.W., N. Kawakami, T. Tsuji, K.U., T.S., M.I., H.S., S.M., T.K., H. Yanagida, I.Y., Y.O., H.T. and Y. Toyama managed DNA samples from individuals with AIS and clinical data. M.K. managed DNA samples from control individuals. J.D., X.Q., H.J., H. Yan, Y.Q. and Q.J. designed and carried out the Han Chinese replication study. S.S., D.L., D.G., J.A.H. and C.A.W. designed and carried out the replication study in individuals of European ancestry. I.K., L.G., A. Takimoto, C.S. and Y.H. designed and performed the functional studies. I.K. and S.I. summarized the results.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Weinstein, S.L. Natural history. *Spine (Phila Pa 1976)* **24**, 2592–2600 (1999).
2. Takahashi, Y. *et al.* A genome-wide association study identifies common variants near *LBX1* associated with adolescent idiopathic scoliosis. *Nat. Genet.* **43**, 1237–1240 (2011).
3. Wynne-Davies, R. Genetic aspects of idiopathic scoliosis. *Dev. Med. Child Neurol.* **15**, 809–811 (1973).
4. Ward, K. *et al.* Polygenic inheritance of adolescent idiopathic scoliosis: a study of extended families in Utah. *Am. J. Med. Genet. A.* **152A**, 1178–1188 (2010).
5. Fan, Y.H. *et al.* SNP rs11190870 near *LBX1* is associated with adolescent idiopathic scoliosis in southern Chinese. *J. Hum. Genet.* **57**, 244–246 (2012).
6. Jiang, H. *et al.* Association of rs11190870 near *LBX1* with adolescent idiopathic scoliosis susceptibility in a Han Chinese population. *Eur. Spine J.* **22**, 282–286 (2013).
7. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
8. Yang, T.P. *et al.* Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* **26**, 2474–2476 (2010).
9. Heinzen, E.L. *et al.* Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol.* **6**, e1 (2008).
10. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
11. Stehlik, C., Kroismayr, R., Dorfleutner, A., Binder, B.R. & Lipp, J. VIGR—a novel inducible adhesion family G-protein coupled receptor in endothelial cells. *FEBS Lett.* **569**, 149–155 (2004).
12. Henry, S.P., Liang, S., Akdemir, K.C. & de Crombrugghe, B. The postnatal role of Sox9 in cartilage. *J. Bone Miner. Res.* **27**, 2511–2525 (2012).
13. Kouwenhoven, J.W. & Castelein, R.M. The pathogenesis of adolescent idiopathic scoliosis: review of the literature. *Spine (Phila Pa 1976)* **33**, 2898–2908 (2008).
14. Guo, X., Chau, W.W., Chan, Y.L. & Cheng, J.C. Relative anterior spinal overgrowth in adolescent idiopathic scoliosis. Results of disproportionate endochondral-membranous bone growth. *J. Bone Joint Surg. Br.* **85**, 1026–1031 (2003).
15. Burwell, R.G., Dangerfield, P.H. & Freeman, B.J. Concepts on the pathogenesis of adolescent idiopathic scoliosis. Bone growth and mass, vertebral column, spinal cord, brain, skull, extra-spinal left-right skeletal length asymmetries, disproportions and molecular pathogenesis. *Stud. Health Technol. Inform.* **135**, 3–52 (2008).
16. Burwell, R.G. *et al.* Pathogenesis of adolescent idiopathic scoliosis in girls—a double neuro-osseous theory involving disharmony between two nervous systems, somatic and autonomic expressed in the spine and trunk: possible dependency on sympathetic nervous system and hormones with implications for medical therapy. *Scoliosis* **4**, 24 (2009).
17. Monk, K.R., Oshima, K., Jors, S., Heller, S. & Talbot, W.S. Gpr126 is essential for peripheral nerve development and myelination in mammals. *Development* **138**, 2673–2680 (2011).
18. Zhao, J. *et al.* The role of height-associated loci identified in genome wide association studies in the determination of pediatric stature. *BMC Med. Genet.* **11**, 96 (2010).
19. Soranzo, N. *et al.* Meta-analysis of genome-wide scans for human adult stature identifies novel loci and associations with measures of skeletal frame size. *PLoS Genet.* **5**, e1000445 (2009).
20. Sovio, U. *et al.* Genetic determinants of height growth assessed longitudinally from infancy to adulthood in the northern Finland birth cohort 1966. *PLoS Genet.* **5**, e1000409 (2009).
21. Lettre, G. *et al.* Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat. Genet.* **40**, 584–591 (2008).
22. Sharma, S. & Wise, C. Current understanding of genetic factors in idiopathic scoliosis. in *The Genetics and Development of Scoliosis* (eds. Katsumi, K. & Dunwoodie, S.L.) 167–190 (Springer, New York, 2010).
23. Nave, K.A. Myelination and support of axonal integrity by glia. *Nature* **468**, 244–252 (2010).
24. Monk, K.R. *et al.* A G protein–coupled receptor is essential for Schwann cells to initiate myelination. *Science* **325**, 1402–1405 (2009).
25. Pogoda, H.M. *et al.* A genetic screen identifies genes essential for development of myelinated axons in zebrafish. *Dev. Biol.* **298**, 118–131 (2006).
26. Carter, G.T. *et al.* Profiles of neuromuscular diseases. Hereditary motor and sensory neuropathy, types I and II. *Am. J. Phys. Med. Rehabil.* **74**, S140–S149 (1995).
27. Martin, A.G., Foguet, P.R., Marks, D.S., Thompson, A.G. & Child, A.H. Infantile scoliosis in Beals syndrome: the use of a non-fusion technique for surgical correction. *Eur. Spine J.* **15**, 433–439 (2006).
28. Yingsakmongkol, W. & Kumar, S.J. Scoliosis in arthrogryposis multiplex congenita: results after nonsurgical and surgical treatment. *J. Pediatr. Orthop.* **20**, 656–661 (2000).
29. Hancock, D.B. *et al.* Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat. Genet.* **42**, 45–52 (2010).
30. Boyer, J., Amin, N., Taddonio, R. & Dozor, A.J. Evidence of airway obstruction in children with idiopathic scoliosis. *Chest* **109**, 1532–1535 (1996).

ONLINE METHODS

Subjects and genotyping. GWAS samples, quality control measures and genotyping platforms have been described previously[2]. The Japanese replication samples have also been described previously[2], with 460 cases and 14,643 controls being added for the present replication study. To validate the imputation analysis, we genotyped the subjects using a multiplex PCR–based Invader assay (Third Wave Technologies)[31]. Genotypes were called by visual inspection for all SNPs. The 743 Han Chinese subjects with AIS were recruited from the Affiliated Drum Tower Hospital of Nanjing University Medical School between 2005 and 2011. They underwent clinical and radiological examinations by surgeons specializing in scoliosis treatment, according to previously described criteria[32]. The control subjects consisted of 1,209 healthy volunteers. We genotyped the Han Chinese subjects using a multiplex PCR–based Invader assay. The subjects of European ancestry were recruited under a protocol approved by the University of Texas Southwestern Medical Center Institutional Review Board. The 457 cases were ascertained sequentially in orthopedic clinics at Texas Scottish Rite Hospital for Children and underwent clinical and radiological examinations by surgeons specializing in AIS treatment, as previously described[33]. The 744 controls were ascertained from within the local Texas population or from non-orthopedic clinics at Texas Scottish Rite Hospital for Children, and a diagnosis of scoliosis was excluded by questionnaire. All subjects were of self-reported white, non-Hispanic ancestry. Subjects of European ancestry were genotyped at 730,498 markers, including rs6570507, using Illumina Human OmniExpress 12v1 BeadChips. Quality control was performed using algorithms in PLINK v1.07 (ref. 34). A total of 17 individuals were removed from the analyses, including 15 individuals with genotyping missing rates greater or equal to 5% and 2 individuals with ambiguous sex according to X-chromosome inbreeding coefficient[35]. Confounding by population stratification was assessed by applying the genomic control method, confirming the absence of stratification in the samples ($\lambda_{GC} = 1.0992$). Informed consent was obtained from all the subjects and from the parents of subjects who were minors. The ethics committees of the participating institutions and of RIKEN approved the study.

Whole-genome imputation. Genotype data for cases (Illumina Human610 Genotyping BeadChip) and controls (Illumina HumanHap550v3 Genotyping BeadChip) were separately processed to exclude SNPs with call rate of <0.99, minor allele frequency (MAF) of <0.01 or Hardy-Weinberg equilibrium P value of $<1 \times 10^{-6}$; samples were excluded if they had a call rate of <0.98. We oriented the alleles of the genotyped SNPs to both the plus strand of the UCSC hg19 genome build and the 1000 Genomes Project reference (REF) and alternate (ALT) alleles, merging data from the two platforms to create a consensus data set (number of SNPs passing quality control = 438,242). We divided genotype data for each chromosome arm into groups using the splitPed.pl script (see URLs; parameters: -windowSize = 1,000, -overlapSize = 100) and phased each group using MACH version 1.0.16.c. (rounds = 20, states = 200). We imputed untyped loci using Minimac software (rounds = 5, states = 200; Minimac source code dated 1 March 2011, obtained from C. Fuchsberger; for the latest Minimac version, see URLs) with reference haplotypes from the 1000 Genomes Project June 2011 EAS population after excluding monomorphic SNPs. As previously described[2], we excluded SNPs from statistical analyses if MAF was <0.01 or if the Minimac Rsq value fell below thresholds specific to five MAF ranges (cutoff values: MAF of 0.0–0.1 = 0.75, MAF of 0.1–0.2 = 0.70, MAF of 0.2–0.3 = 0.66, MAF of 0.3–0.4 = 0.60, and MAF of 0.4–0.5 = 0.55). In total, 6,140,112 genotyped and imputed SNPs remained after quality control filtering.

Statistical analysis. In the GWAS and the replication studies, we assessed the association of each SNP with the Cochrane-Armitage trend test. We calculated OR values and their CIs using the risk allele as the reference. We used the Mantel-Haenszel method for meta-analyses and examined heterogeneity between studies using the Breslow-Day test. In the imputation analysis, pairwise r^2 values were calculated using the R Bioconductor package snpMatrix (version 1.16.2), and the LD map was created using in-house programs. We performed association analysis of imputed data using the single.snp.tests function in the R package snpStats version 1.3.4 after converting Minimac output to the uncertain genotype data format for snpStats. Regional association plots were generated using R statistical environment version 2.13.0. For general statistical analyses, we used R statistical environment version 2.13.0 or Microsoft Excel. Explained trait variance was calculated as the square of the correlation coefficient, r^2.

Quantitative RT-PCR. We synthesized cDNA from the total RNA of intervertebral disc, bone, articular cartilage and spinal cord samples using Multiscribe reverse transcriptase and a random hexamer primer (Applied Biosystems). We used this cDNA, multiple-tissue cDNA panels (Clontech) and *GPR126*-specific primers (**Supplementary Table 3**) for PCR experiments to examine the tissue-specific expression of *GPR126*. We carried out quantitative RT-PCR using an ABI PRISM 7700 sequence detector with the QuantiTect SYBR Green PCR kit (Qiagen) in accordance with the manufacturers' instructions.

***In situ* hybridization.** Antisense and sense RNA probes for each gene were transcribed from plasmids with a digoxigenin (DIG) RNA labeling kit (Roche). To generate RNA probes, *Gpr126* cDNA was amplified by RT-PCR on the basis of sequence information in GenBank (NM_001002268). *In situ* hybridization was performed essentially as previously described[36].

Morpholino injections and bone visualization. Antisense and control morpholinos were ordered from Gene Tools. The sequences of the *gpr126* translation-blocking and splice-blocking morpholinos and of the control morpholino were 5′-ACCGACCACTGATGAACGAAATCAT-3′, 5′-TGCCACTG CAAGCAAACAGGGCACA-3′ and 5′-CCTCTTACCTCAGTTACAATTTAT A-3, respectively. We injected 15 ng of each morpholino into the embryo yolk at the one-cell stage. The major and minor eyeball diameters were measured, and we give the means of these values. We assessed skeletal development of zebrafish using Alizarin red staining[37].

31. Ohnishi, Y. *et al.* A high-throughput SNP typing system for genome-wide association studies. *J. Hum. Genet.* **46**, 471–477 (2001).
32. Takahashi, Y. *et al.* Lack of association between adolescent idiopathic scoliosis and previously reported single nucleotide polymorphisms in *MATN1*, *MTNR1B*, *TPH1*, and *IGF1* in a Japanese population. *J. Orthop. Res.* **29**, 1055–1058 (2011).
33. Sharma, S. *et al.* Genome-wide association studies of adolescent idiopathic scoliosis suggest candidate susceptibility genes. *Hum. Mol. Genet.* **20**, 1456–1466 (2011).
34. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
35. Falconer, D.S. & Mackay, T.F.C. *Introduction to Quantitative Genetics*, 4th edn (Longman, Harlow, UK, 1996).
36. Shukunami, C., Takimoto, A., Miura, S., Nishizaki, Y. & Hiraki, Y. Chondromodulin-I and tenomodulin are differentially expressed in the avascular mesenchyme during mouse and chick development. *Cell Tissue Res.* **332**, 111–122 (2008).
37. Fleming, A., Keynes, R. & Tannahill, D. A central role for the notochord in vertebral patterning. *Development* **131**, 873–880 (2004).

Fujii et al. Nutrition Journal 2013, 12:159
http://www.nutritionj.com/content/12/1/159

NUTRITION JOURNAL

RESEARCH Open Access

Impact of dietary fiber intake on glycemic control, cardiovascular risk factors and chronic kidney disease in Japanese patients with type 2 diabetes mellitus: the Fukuoka Diabetes Registry

Hiroki Fujii[1], Masanori Iwase[1,2*], Toshiaki Ohkuma[1], Shinako Ogata-Kaizu[1], Hitoshi Ide[1], Yohei Kikuchi[1], Yasuhiro Idewaki[1], Tamaki Joudai[1], Yoichiro Hirakawa[3], Kazuhiro Uchida[4], Satoshi Sasaki[5], Udai Nakamura[1] and Takanari Kitazono[1]

Abstract

Background: Dietary fiber is beneficial for the treatment of type 2 diabetes mellitus, although it is consumed differently in ethnic foods around the world. We investigated the association between dietary fiber intake and obesity, glycemic control, cardiovascular risk factors and chronic kidney disease in Japanese type 2 diabetic patients.

Methods: A total of 4,399 patients were assessed for dietary fiber intake using a brief self-administered diet history questionnaire. The associations between dietary fiber intake and various cardiovascular risk factors were investigated cross-sectionally.

Results: Body mass index, fasting plasma glucose, HbA1c, triglyceride and high-sensitivity C-reactive protein negatively associated with dietary fiber intake after adjusting for age, sex, duration of diabetes, current smoking, current drinking, total energy intake, fat intake, saturated fatty acid intake, leisure-time physical activity and use of oral hypoglycemic agents or insulin. The homeostasis model assessment insulin sensitivity and HDL cholesterol positively associated with dietary fiber intake. Dietary fiber intake was associated with reduced prevalence of abdominal obesity, hypertension and metabolic syndrome after multivariate adjustments including obesity. Furthermore, dietary fiber intake was associated with lower prevalence of albuminuria, low estimated glomerular filtration rate and chronic kidney disease after multivariate adjustments including protein intake. Additional adjustments for obesity, hypertension or metabolic syndrome did not change these associations.

Conclusion: We demonstrated that increased dietary fiber intake was associated with better glycemic control and more favorable cardiovascular disease risk factors including chronic kidney disease in Japanese type 2 diabetic patients. Diabetic patients should be encouraged to consume more dietary fiber in daily life.

Keywords: Adiponectin, Albuminuria, Chronic kidney disease, Diabetes mellitus, Dietary fiber, Homeostasis model assessment, Hypertension, Inflammation, Insulin resistance, Metabolic syndrome

* Correspondence: iwase@intmed2.med.kyushu-u.ac.jp
[1]Department of Medicine and Clinical Science, Graduate School of Medical Sciences, Kyushu University, Maidashi 3-1-1, Higashi-ku, Fukuoka 812-8582, Japan
[2]Diabetes Centre, Hakujyuji Hospital, Fukuoka, Japan
Full list of author information is available at the end of the article