Kikukawa, T., Shimono, K., Tamogami, J., Miyauchi, S., Kim, S., Kimura-Someya, T., Shirouzu, M., Jung, K. H., Yokoyama, S. and Kamo, N	Photochemistry of Acetabularia rhodopsin II from a marine plant, Acetabularia acetabulum.	Biochemistry	50 (41)	8888-8898	2011

Evaluating Japanese Patients With the Marfan Syndrome Using High-Throughput Microarray-Based Mutational Analysis of Fibrillin-1 Gene

Naomi Ogawa, MD, PhD[a,c], Yasushi Imai, MD, PhD[a,c], Yuji Takahashi, MD, PhD[d,e], Kan Nawata, MD, PhD[b], Kazuo Hara, MD, PhD[c,g], Hiroshi Nishimura, MD, PhD[a], Masayoshi Kato, MD, PhD[a], Norifumi Takeda, MD, PhD[a], Takahide Kohro, MD, PhD[c,h], Hiroyuki Morita, MD, PhD[h], Tsuyoshi Taketani, MD, PhD[b], Tetsuro Morota, MD, PhD[b], Tsutomu Yamazaki, MD, PhD[c,f], Jun Goto, MD, PhD[d,e], Shoji Tsuji, MD, PhD[d,e], Shinichi Takamoto, MD, PhD[b], Ryozo Nagai, MD, PhD[a,c], and Yasunobu Hirata, MD, PhD[a,*]

Marfan syndrome (MS) is an inherited connective tissue disorder, and detailed evaluations of multiple organ systems are required for its diagnosis. Genetic testing of the disease-causing fibrillin-1 gene (FBN1) is also important in this diagnostic scheme. The aim of this study was to define the clinical characteristics of Japanese patients with MS and enable the efficient and accurate diagnosis of MS with mutational analysis using a high-throughput microarray-based resequencing system. Fifty-three Japanese probands were recruited, and their clinical characteristics were evaluated using the Ghent criteria. For mutational analysis, an oligonucleotide microarray was designed to interrogate FBN1, and the entire exon and exon-intron boundaries of FBN1 were sequenced. Clinical evaluation revealed more pulmonary phenotypes and fewer skeletal phenotypes in Japanese patients with MS compared to Caucasians. The microarray-based resequencing system detected 35 kinds of mutations, including 23 new mutations. The mutation detection rate for patients who fulfilled the Ghent criteria reached 71%. Of note, splicing mutations accounted for 19% of all mutations, which is more than previously reported. In conclusion, this comprehensive approach successfully detected clinical phenotypes of Japanese patients with MS and demonstrated the usefulness and feasibility of this microarray-based high-throughput resequencing system for mutational analysis of MS. © 2011 Elsevier Inc. All rights reserved. **(Am J Cardiol 2011;108:1801–1807)**

The Marfan syndrome (MS) is an inherited connective tissue disorder with an autosomal dominant inheritance, primarily involving the skeletal, ocular, and cardiovascular systems, caused by mutations in fibrillin-1 gene (FBN1).[1] Diagnosis of the MS has been made using the Ghent criteria[2] on the basis of data from European and American populations, but the Ghent criteria may not be completely suitable for the Japanese population.[3] Therefore, epidemiologic and genetic surveys in the Japanese population are mandatory to establish more Japanese-specific (or Asian-specific) diagnostic criteria for the MS. The Ghent criteria were recently further revised.[4] More weight is now given to FBN1 testing, and a diagnosis can be made if a patient has the FBN1 mutation plus either an aortic phenotype or ectopia lentis. These new criteria are much simpler than the original criteria. Thus, genetic testing of MS is becoming more important. FBN1 spans a 230-kb genomic region and contains 65 exons. More than 1,000 reported mutations are spread throughout the gene and are mostly unique in each affected family.[5,6] Classic genetic analysis methods such as direct sequencing are very time consuming. Thus, the introduction of a more efficient genetic analysis tool is needed. Custom-designed resequencing microarrays enable the analysis of multiple genes spanning 30 to 300 kb on a single array. The microarray identifies individual nucleotides by comparative, high-fidelity hybridization using oligonucleotide probes[7–9] (Figure 1). In the present study, we comprehensively evaluated the clinical characteristics of Japanese patients with suspected MS and also conducted mutational analysis of these patients by adopting a high-throughput genetic diagnosing system to achieve more efficient and accurate diagnoses.

Methods

Fifty-three consecutive probands suspected of having MS who visited the MS clinic at our hospital were enrolled. All patients were assessed using the original Ghent criteria.[2,10] This study was conducted according to the Declara-

[a]Departments of Cardiovascular Medicine, [b]Cardiothoracic Surgery, [c]Clinical and Genetic Informatics, [d]Neurology, [e]Clinical Genomics, [f]Clinical Epidemiology and Systems, [g]Metabolic Diseases, and [h]Translational Research for Healthcare and Clinical Science, Graduate School of Medicine, University of Tokyo, Tokyo, Japan. Manuscript received April 26, 2011; revised manuscript received and accepted July 15, 2011.

This work was supported by Health Labor Science's Research Grants from the Japanese Ministry of Health, Labor, and Welfare (Grant 10103493 to Dr. Hirata), the Human Resources Development Program of the Japanese Ministry of Education, Culture, Sports and Technology and JSPS through its FIRST Program (Drs. Yamazaki and Nagai).

*Corresponding author: Tel: 81-3815-5411; fax: 81-3-5800-9845.
E-mail address: hirata-2im@h.u-tokyo.ac.jp (Y. Hirata).

0002-9149/11/$ – see front matter
doi:10.1016/j.amjcard.2011.07.053

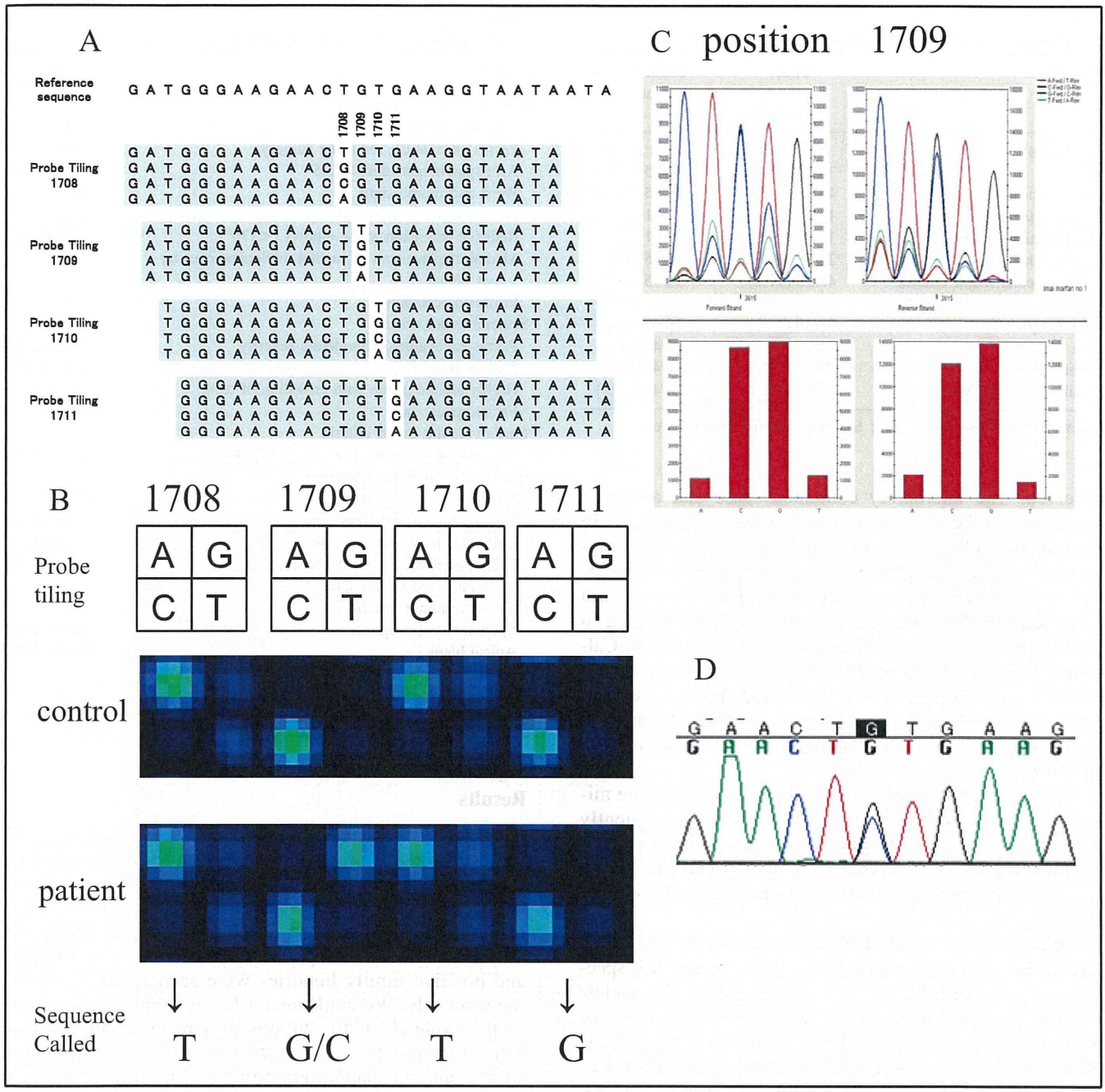

Figure 1. Representative example of mutational analysis using the present microarray-based resequencing system. *(A)* The microarray identifies individual nucleotides by comparative, high-fidelity hybridization using oligonucleotide probes that are synthesized in situ by photolithography and solid-phase DNA synthesis. For each base position, 8 unique 25-mer probes (4 oligonucleotide probes for each strand) are tiled on the array, and each 25-mer probe is varied at the central position to incorporate each possible nucleotide (A, G, C, or T), allowing the detection of all possible nucleotide substitutions. *(B)* Scan images of the probes around the nucleotide position 1709. In patients with the FBN1 c.1709G>C mutation, high signal intensities can be observed in probe G and C at nucleotide 1709 compared to control. *(C,D)* Signal intensity data at nucleotide position 1709. The intensity data for each base position can be also displayed as traces and bar graphs. The missense mutation (c.1709G>C) was successfully detected *(C)* and was verified by direct sequencing *(D)*.

tion of Helsinki and was approved by the institutional ethics committee. Written informed consent was obtained after providing a detailed explanation of the study. Genomic deoxyribonucleic acid (DNA) was extracted from buffy coat using a Genomix DNA extraction kit (Talent, Trieste, Italy). For amplification of the 65 exons of FBN1, polymerase chain reaction (PCR) primers were designed by referring to previous reports.[11–13] After performing the PCRs according to the standard protocol, the PCR products were subjected to hybridization on the microarray.

The resequencing microarray was designed on the basis of the reference sequences from the Ensembl database. Because highly homologous sequences lead to cross-hybridization, FBN1 was checked for possible repetitive sequences using RepeatMasker software (http://repeatmasker.org/chi-bin/webrepeatmasker). No repetitive elements were

Table 1
Background of participants who underwent genetic analysis (n = 53)

Variable	Total (n = 53)	Ghent-Positive Patients (n = 45)
Age (years)	33.1 ± 9.8	33.1 ± 10.4
Men	35/53 (66%)	30/45 (67%)
Ghent positive	45/53 (85%)	45/45 (100%)
Skeletal major criteria	12/49 (25%)	12/41 (29%)
Skeletal minor criteria	19/49 (39%)	17/41 (42%)
Ectopia lentis	25/53 (47%)	25/45 (56%)
Cardiovascular major criteria	48/53 (91%)	44/45 (98%)
Cardiovascular minor criteria	36/48 (75%)	32/41 (78%)
Pulmonary	22/49 (45%)	19/42 (45%)
Skin	26/49 (53%)	23/42 (55%)
Dural ectasia	34/47 (72%)	33/40 (83%)
Family history of MS	31/55 (56%)	28/45 (62%)

Data are expressed as mean ± SD or as number (percentage).

observed. The microarray contained sense and antisense sequences for the 65 exons of FBN1 and ≥12 flanking base pairs of the splice junctions. The PCR product was fragmented, end-labeled with biotin, and hybridized to the array. Washing and staining with streptavidin-phycoerythrin were performed on automated fluidic stations according to the manufacturer's protocol (Affymetrix, Santa Clara, California). Hybridization signals were read by a high-resolution laser scanner, and the data collection and interpretation were carried out using GeneChip Operating Software and GeneChip Sequence Analysis Software (Affymetrix), respectively.

Candidate nucleotide substitutions detected by the microarray-based resequencing system were subsequently validated by fluorescent dideoxy DNA sequencing using BigDye terminator version 3.1 on an ABI PRISM 3100xl genetic analyzer (Applied Biosystems, Foster City, California).

Some patients underwent cardiovascular surgery, and written informed consent for research use of surgical specimens was obtained from each patient. Total ribonucleic acid (RNA) was extracted using an RNeasy Fibrous Tissue Mini Kit (Qiagen, Venlo, The Netherlands). For patients whose aortic tissues were not available, total RNA was extracted from blood using a QIAamp RNA Blood Mini Kit (Qiagen). The RNA was converted to complementary DNA using SuperScript III First-Strand Synthesis SuperMix (Invitrogen, Carlsbad, California). PCR analyses were performed with specific primers designed for the target regions. PCR samples or subcloned plasmids after TA cloning of PCR products using a TOPO-TA vector (Invitrogen) were subjected to fluorescent dideoxy DNA sequencing.

DNA from patients whose mutations were not found by the aforementioned methods was screened by multiplex ligation-dependent probe amplification using a SALSA MLPA kit P065/P066 (MRC-Holland, Amsterdam, The Netherlands)[14] for large deletions and duplications.

All quantitative data are expressed as mean ± SD. Statistical comparisons of distributions between groups were made using the chi-square test. Significance was taken as $p < 0.05$.

Table 2
Detailed clinical findings of Ghent-positive patients (n = 45)

Criterion	n (%)
Skeletal major criteria	
Pectus carinatum	9/42 (21%)
Pectus excavatum, requiring surgery	7/44 (16%)
Arm span/height ratio >1.05	8/41 (20%)
Wrist and thumb signs	32/43 (74%)
Scoliosis of >20% or spondylolisthesis	21/44 (48%)
Reduced extension at the elbows (<170°)	2/41 (5%)
Medial displacement of medial malleolus, causing pes planus	16/41 (39%)
Protrusio acetabuli	8/39 (21%)
Skeletal minor criteria	
Pectus excavatum of moderate severity	10/44 (23%)
Joint hypermobility	7/41 (17%)
Highly arched palate with crowding of teeth	31/40 (78%)
Facial appearance	15/40 (38%)
Cardiovascular major criteria	
Dilatation/dissection of the ascending aorta	44/45 (98%)
Cardiovascular minor criteria	
Mitral valve prolapse	23/42 (55%)
Dilatation of main pulmonary artery	9/20 (45%)
Calcification of mitral annulus	0/34 (0%)
Dilatation/dissection of descending thoracic/abdominal aorta	12/43 (28%)
Pulmonary minor criteria	
Spontaneous pneumothorax	13/43 (30%)
Apical blebs	15/44 (34%)
Skin minor criteria	
Striae atrophicae	24/42 (57%)
Recurrent or incisional herniae	0/41 (0%)

Results

Of the 53 probands enrolled, 45 were diagnosed with MS according to the original Ghent criteria. Because our Marfan clinic offers cardiac surgery and some patients were referred for aortic surgery from other hospitals, most of the patients had aortic phenotypes (Table 1). Dural ectasia and ectopia lentis were common findings, and positive family histories were seen in about half of the probands. We confirmed a lower frequency for some of the skeletal manifestations in Japanese patients with MS compared to that reported in a Western database, such as an arm span/height ratio >1.05 (20% in our study vs 55% in Western populations) and reduced extension at the elbows (<170°) (5% vs 15%), findings that were similar to the report of Akutsu et al[3,6] (Table 2). However, the frequency of major skeletal criteria (29%) was higher than a previous Japanese report (15%), which is partially due to a lack of evaluation of protrusio acetabuli in the earlier study. We found a higher frequency of spontaneous pneumothorax (30% vs 7%) in our Japanese population compared to a previous study conducted in Western patients. Calcification of the mitral annulus and frequency of dilatation of the main pulmonary artery were rarely reported. Actually, mitral annular calcification was not detected at all. However, pulmonary artery dilatation was relatively frequent (45% [9 of 20]) in our study, after excluding those patients whose main pulmonary artery diameters were difficult to evaluate.

Table 3
Mutations found in this study

Exon	Complementary DNA	Protein
Missense mutations		
4	c.386G>A	p.Cys 129 Tyr
13	c.1709G>C*	p.Cys 570 Ser
14	c.1786T>G*	p.Cys 596 Gly
15	c.1911T>G*	p.Cys 637 Trp
18	c.2171T>G*	p.Ile 724 Arg
18	c.2201G>T	p.Cys 734 Phe
21	c.2638G>A	p.Gly 880 Ser
24	c.3043G>A	p.Ala 1015 Thr
26	c.3263A>G*	p.Asn 1088 Ser
28	c.3503A>G	p.Asn 1168 Ser
34	c.4280A>G*	p.Tyr 1427 Cys
43	c.5371T>C*	p.Cys 1791 Arg
47	c.5873G>A*	p.Cys 1958 Tyr
50	c.6296G>T	p.Cys 2099 Phe
53	c.6518G>A*	p.Gly 2173 Ser
57	c.7015T>G*	p.Cys 2339 Gly
60	c.7466G>A*	p.Cys 2489 Tyr
62	c.7754T>C	p.Ile 2585 Thr (2 probands)
Nonsense mutations		
8	c.945T>A*	p.Cys 315 X
12	c.1585C>T	p.Arg 529 X
29	c.3603C>A*	p.Cys 1201 X
37	c.4709G>A*	p.Trp 1570 X
38	c.4777G>T*	p.Glu 1593 X
38	c.4786C>T	p.Arg 1596 X
54	c.6658C>T	p.Arg 2220 X
58	c.7240C>T	p.Arg 2414 X
65	c.8521G>T*	p.Glu 2841 X
Splicing mutations		
11–12	c.IVS11+5G>A	p.Cys474Tyr Glu475_Asp490del
15–16	c.IVS15-3T>G*	
16–17	c.IVS16+3A>C*	
18–19	c.IVS18+1G>C*	
34–35	c.IVS34-1G>A*	p.Asp1446ValfsX21
40–41	c.IVS40+1G>A*	
52–53	c.6453C>T*	p.Cys2151Tyr, Glu2152_Asp2166del
56–57	c.IVS56+5G>A*	
Deletion mutations		
54	c.6665delT*	p.Val2222GlyFsX69
54	c.6703-6704delGG*	p.Gly2235IlefsX7
55	c.6837delG*	p.Tyr2280IlefsX10
57	c. 7071_7079delCGTCACCAA*	p.Val2358SerfsX511
65	c. 8532_8delTACAACT*	p.Thr2785X
3	Exon 3 deletion*	

* Newly found mutation.

In our mutational analysis, the base call rate of this system for FBN1 was >96% when examining 5 representative cases, and resequencing as many as 12,688 bp per patient was easily accomplished in 3 working days, demonstrating the high fidelity and high throughput of this system.

In the 53 probands, 35 kinds of FBN1 mutations were found in 36 probands using this system (Table 3). There were 18 missense and 9 nonsense mutations. Eight other mutations located near the exon-intron boundaries were thought to alter the splicing patterns. Supplemental direct sequencing in probands with no mutation detected by the microarray-based method revealed 5 deletion mutations in FBN1 (Table 3). Furthermore, multiplex ligation-dependent probe amplification assay revealed a large deletion mutation (exon 3) in 1 proband. Finally, novel mutations were found in 23 probands using microarray and in 29 probands in total. All possible mutations found by the microarray-based resequencing system were verified by direct sequencing, and thus the microarray detected point mutations with 100% accuracy. A representative example of genetic analysis using the microarray-based resequencing system is shown in Figure 1. Of 18 missense mutations, 11 were either affecting or creating cysteine residues. For other novel missense mutations, none of the mutations were found in ≥200 ethnically matched control subjects. The mutation detection rate

Table 4
Number of mutations detected

Mutation Detection Method	Total (n = 53)	Ghent Positive (n = 45)	Other (n = 8)
Microarray	36 (68%)	32 (71%)	4 (50%)
Direct sequencing	5 (9%)	5 (11%)	0
Multiplex ligation-dependent probe amplification	1 (2%)	1 (2%)	0
Total of all 3 modalities	42 (79%)	38 (84%)	4 (50%)

of the microarray-based resequencing system for the Ghent-positive patients was 71%. The overall mutation detection rate after additional analysis by fluorescent dideoxy DNA sequencing and multiplex ligation-dependent probe amplification reached 84% (Table 4).

Eight possible splicing mutations were identified, and these mutations constituted 19% of all mutations, which was more than the 11% currently reported in the UMD-FBN1 mutation database.[5] One patient and his 2 relatives with MS had the same silent mutation in FBN1 exon 52 (c.6453C>T, p.Cys2151Cys; Figure 2). Therefore, we re-sequenced complementary DNA from his aortic tissue and verified an alternation of the splicing pattern between FBN1 exon 52 and 53. The C at nucleotide position 6453 of FBN1 complementary DNA was substituted with a T, which resulted in the creation of a new splicing donor site, causing abnormal shorter messenger RNA. Another patient had a mutation at the fifth nucleotide of the beginning of intron 11 (c.IVS11+5G>A), although it is well known that the first 2 nucleotides at the beginning of the intron are very important as a splice donor site. We found by sequencing the complementary DNA that the latent splice donor site within exon 11 was activated and created the frame-shift mutation (Figure 2).

Six additional mutations possibly causing a splicing aberration were also found (Table 3). Although aortic tissue was unavailable for these patients, splicing aberrations were successfully confirmed in 2 whose complementary DNA was clinically available by resequencing FBN1 complementary DNA obtained from peripheral blood (Figure 2).

In published research, it has been suggested that mutations causing the in-frame loss or gain of the central coding sequence through deletions, insertions, or splicing errors are thought to be associated with more severe disease phenotypes. In contrast, nonsense mutations that result in rapid degradation of mutant transcripts are reported to be potentially associated with milder conditions. However, we could not find any associations between mutation types and clinical severity in our study subjects. A higher incidence of ectopia lentis in patients who carried a missense mutation involving a cysteine substitution or splicing mutation has been reported.[15] However, these correlations were not observed in our study. Among 4 patients who had mutations located between FBN1 exons 24 and 32, the so-called "neonatal region," none had the neonatal or early-onset form of MS.

Discussion

The Ghent criteria for MS diagnosis are based on data obtained mainly from European and American populations. Our clinical evaluations revealed that there were more pulmonary phenotypes and fewer skeletal phenotypes in Japanese patients with MS compared to Western patients. Therefore, the criteria for systemic and orthopedic features in the Ghent nosology may not be entirely suitable for application to Japanese and perhaps other Asian populations. Further epidemiologic and genetic studies in the Japanese population should be conducted to establish Asian- or Japanese-specific diagnostic criteria for MS.

The present microarray-based resequencing system is an efficient method for rapid and affordable mutation analysis of heterogenous disorders such as MS. The mutation detection rate is influenced by the accuracy of the clinical diagnosis of MS, the type of mutation, and the ability of the testing method. It ranged from 55% to 91% in previous reports.[11,16–18] The mutation detection rate of our system was concordant with previous reports. Its greatest advantages are high throughput and digitalized sequencing data. The digitally retrieved sequencing data are easily computable and can be displayed in various ways. In most of the cases, we could identify the mutations within a few minutes of data collection. Several other causative genes, such as transforming growth factor receptor types 1 and 2 (TGFBR1 and TGFBR2),[19] smooth muscle α-actin (ACTA2),[20] myosin heavy chain 11 (MYH11),[21] and SMAD3,[22] have been identified for syndromic or nonsyndromic aortic aneurysms and dissection. Such additional candidate genes can also be included on the same array because 1 array can resequence up to 300 kb.

Our system can detect point mutations with 100% accuracy and thus is a reliable first screening method for detecting single nucleotide substitutions. In contrast, it is difficult to detect heterozygous deletion or insertion mutations, because an abnormal allele containing a deletion or an insertion mutation is difficult to hybridize to probes. For patients with MS with no mutation detected by the microarray system, conventional direct sequencing and multiplex ligation-dependent probe amplification was helpful for searching for possible deletion or insertion mutations. Because there is a certain number of patients with MS without mutations in FBN1,[12,19] the 7 probands without any mutations may have possessed mutations in undiscovered disease-causing genes.

Eight splicing mutations that accounted for 19% of all the mutations were found. Because this type of mutation represented a greater proportion than that of previous reports, every exon-intron boundary should be resequenced. It is also advisable to obtain messenger RNA in addition to DNA for analyzing the splicing pattern. We successfully demonstrated altered splicing patterns using FBN1 messenger RNA extracted from peripheral leukocytes. Thus, we also recommend the extraction of RNA as well as genomic DNA from peripheral blood, if a surgically retrieved specimen is not available.

We also assessed patients using the recently published revised Ghent criteria. Forty-two of the 45 original Ghent-positive patients were also diagnosed with MS using the revised criteria. One patient, who was positive according to the original Ghent criteria, did not satisfy the revised criteria and was diagnosed with ectopia lentis syndrome. Two patients (aged 20 and 30 years) failed to meet the revised Ghent criteria because their z scores of aortic diameter were not

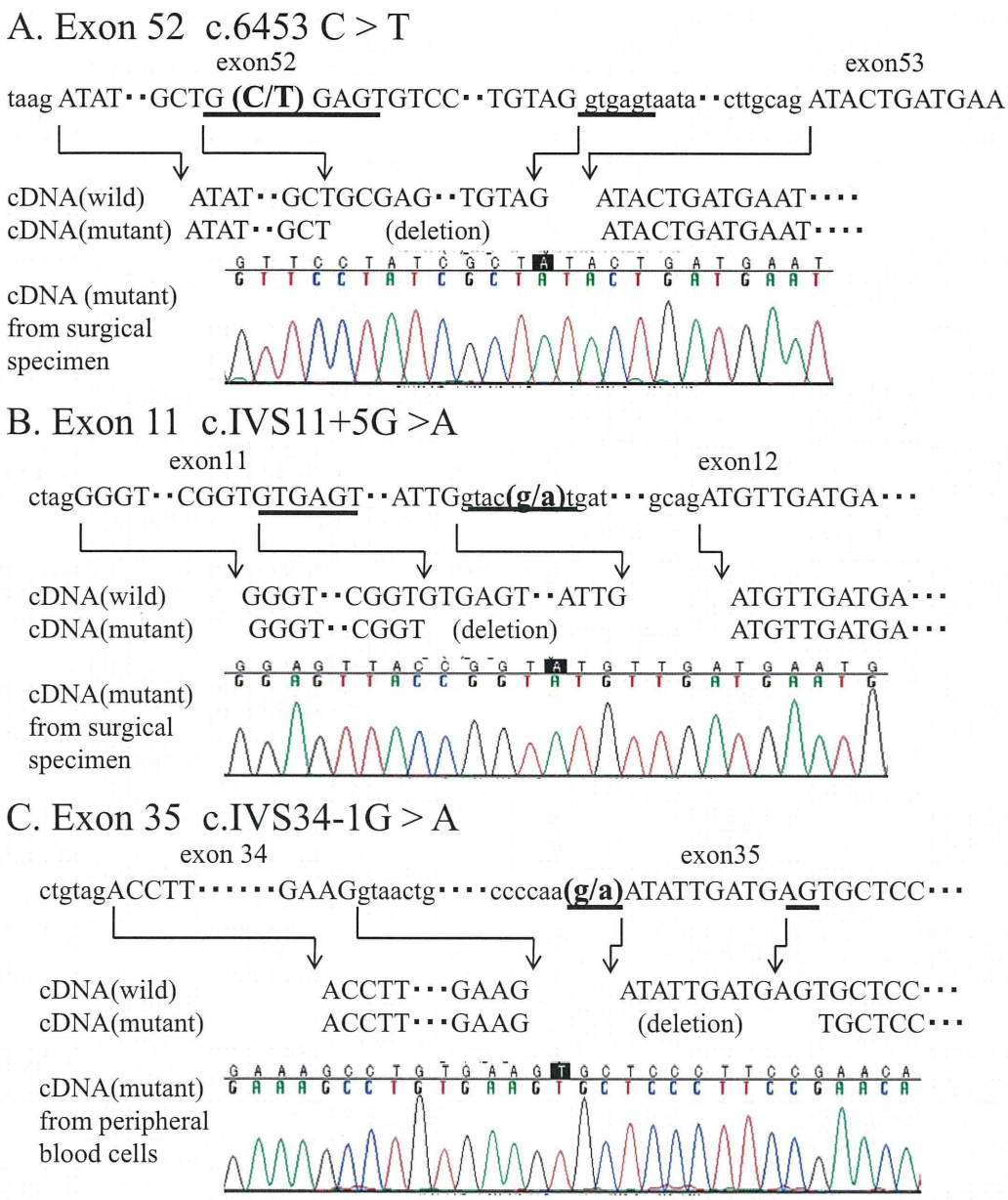

Figure 2. Representative splicing mutations in our study. *(A)* C-to-T substitution at nucleotide position 6453 produced a 6–nucleic acid sequence (c.6452-6457) *(underlined)* identical to that of IVS52+1 to +6 (gtgagt) *(underlined)*, and this was recognized as the splicing donor site. The following nucleotides of exon 52 from this point were deleted in the mutant allele. *(B)* The fifth nucleotide at the beginning of intron 11 was substituted from G to A (c.IVS11+5G>A). Resequencing of complementary DNA revealed that the latent splice donor site within exon 11 *(underlined)* was activated and became a new splice donor site and created the frame-shift mutation. *(C)* The last nucleotide of intron 34 was substituted from G to A (c.IVS34-1G>A). Resequencing FBN1 complementary DNA obtained from the peripheral blood demonstrated the splicing aberration, resulting in the deletion of 11 nucleotides of exon 35.

satisfactory, although their aortas dilated gradually later on. One patient met the revised Ghent criteria after genetic analysis of FBN1. Although further studies are warranted to clarify the properties and usefulness of the novel Ghent criteria, genetic testing is more important, although it is not mandatory. In such a setting, a high-throughput resequencing method such as the present microarray-based resequencing system can be a powerful tool for making an accurate diagnosis of MS.

Acknowledgments: We would like to express our sincere gratitude to the patients and their families for agreeing to participate in this work and to all the doctors in the MS clinic for their valuable advice. We also thank Wu Zhenghong, Eri Kasukawa, Ryohei Kataoka, Keiko Hori, Yumiko Fujimaki, and Takako Kawamura for their excellent technical assistance.

1. Dietz HC, Pyeritz RE. Mutations in the human gene for fibrillin-1 (FBN1) in the Marfan syndrome and related disorders. *Hum Mol Genet* 1995;4:1799–1809.
2. De Paepe A, Devereux RB, Dietz HC, Hennekam RC, Pyeritz RE. Revised diagnostic criteria for the Marfan syndrome. *Am J Med Genet* 1996;62:417–426.

3. Akutsu K, Morisaki H, Takeshita S, Ogino H, Higashi M, Okajima T, Yoshimuta T, Tsutsumi Y, Nonogi H, Morisaki T. Characteristics in phenotypic manifestations of genetically proved Marfan syndrome in a Japanese population. *Am J Cardiol* 2009;103:1146–1148.
4. Loeys BL, Dietz HC, Braverman AC, Callewaert BL, De Backer J, Devereux RB, Hilhorst-Hofstee Y, Jondeau G, Faivre L, Milewicz DM, Pyeritz RE, Sponseller PD, Wordsworth P, De Paepe AM. The revised Ghent nosology for the Marfan syndrome. *J Med Genet* 2010; 47:476–485.
5. Collod-Béroud G, Le Bourdelles S, Ades L, Ala-Kokko L, Booms P, Boxer M, Child A, Comeglio P, De Paepe A, Hyland JC, Holman K, Kaitila I, Loeys B, Matyas G, Nuytinck L, Peltonen L, Rantamaki T, Robinson P, Steinmann B, Junien C, Béroud C, Boileau C. Update of the UMD-FBN1 mutation database and creation of an FBN1 polymorphism database. *Hum Mutat* 2003;22:199–208.
6. Faivre L, Collod-Beroud G, Loeys BL, Child A, Binquet C, Gautier E, Callewaert B, Arbustini E, Mayer K, Arslan-Kirchner M, Kiotsekoglou A, Comeglio P, Marziliano N, Dietz HC, Halliday D, Beroud C, Bonithon-Kopp C, Claustres M, Muti C, Plauchu H, Robinson PN, Adès LC, Biggin A, Benetts B, Brett M, Holman KJ, De Backer J, Coucke P, Francke U, De Paepe A, Jondeau G, Boileau C. Effect of mutation type and location on clinical outcome in 1,013 probands with Marfan syndrome or related phenotypes and FBN1 mutations: an international study. *Am J Hum Genet* 2007;81:454–466.
7. Cutler DJ, Zwick ME, Carrasquillo MM, Yohn CT, Tobin KP, Kashuk C, Mathews DJ, Shah NA, Eichler EE, Warrington JA, Chakravarti A. High-throughput variation detection and genotyping using microarrays. *Genome Res* 2001;11:1913–1925.
8. Warrington JA, Shah NA, Chen X, Janis M, Liu C, Kondapalli S, Reyes V, Savage MP, Zhang Z, Watts R, DeGuzman M, Berno A, Snyder J, Baid J. New developments in high throughput resequencing and variation detection using high-density microarrays. *Hum Mutat* 2002;19:402–409.
9. Takahashi Y, Seki N, Ishiura H, Mitsui J, Matsukawa T, Kishino A, Onodera O, Aoki M, Shimozawa N, Murayama S, Itoyama Y, Suzuki Y, Sobue G, Nishizawa M, Goto J, Tsuji S. Development of a high-throughput microarray-based resequencing system for neurological disorders and its application to molecular genetics of amyotrophic lateral sclerosis. *Arch Neurol* 2008;65:1326–1332.
10. Roman MJ, Devereux RB, Kramer-Fox R, O'Loughlin J. Two-dimensional echocardiographic aortic root dimensions in normal children and adults. *Am J Cardiol* 1989;64:507–512.
11. Sakai H, Visser R, Ikegawa S, Ito E, Numabe H, Watanabe Y, Mikami H, Kondoh T, Kitoh H, Sugiyama R, Okamoto N, Ogata T, Fodde R, Mizuno S, Takamura K, Egashira M, Sasaki N, Watanabe S, Nishimaki S, Takada F, Nagai T, Okada Y, Aoka Y, Yasuda K, Iwasa M, Kogaki S, Harada N, Mizuguchi T, Matsumoto N. Comprehensive genetic analysis of relevant four genes in 49 patients with Marfan syndrome or Marfan-related phenotypes. *Am J Med Genet* 2006;140: 1719–1725.
12. Mizuguchi T, Collod-Beroud G, Akiyama T, Abifadel M, Harada N, Morisaki T, Allard D, Varret M, Claustres M, Morisaki H, Ihara M, Kinoshita A, Yoshiura K, Junien C, Kajii T, Jondeau G, Ohta T, Kishino T, Furukawa Y, Nakamura Y, Niikawa N, Boileau C, Matsumoto N. Heterozygous TGFBR2 mutations in Marfan syndrome. *Nat Genet* 2004;36:855–860.
13. Nijbroek G, Sood S, McIntosh I, Francomano CA, Bull E, Pereira L, Ramirez F, Pyeritz RE, Dietz HC. Fifteen novel FBN1 mutations causing marfan syndrome detected by heteroduplex analysis of genomic amplicons. *Am J Hum Genet* 1995;57:8–21.
14. Mátyás G, Alonso S, Patrignani A, Marti M, Arnold E, Magyar I, Henggeler C, Carrel T, Steinmann B, Berger W. Large genomic fibrillin-1 (FBN1) gene deletions provide evidence for true haploinsufficiency in Marfan syndrome. *Hum Genet* 2007;122:23–32.
15. Rommel K, Karck M, Haverich A, von Kodolitsch Y, Rybczynski M, Müller G, Singh KK, Schmidtke J, Arslan-Kirchner M. Identification of 29 novel and nine recurrent fibrillin-1 (FBN1) mutations and genotype-phenotype correlations in 76 patients with Marfan syndrome. *Hum Mutat* 2005;26:529–539.
16. Liu WO, Oefner PJ, Qian C, Odom RS, Francke U. Denaturing HPLC-identified novel FBN1 mutations, polymorphisms, and sequence variants in Marfan syndrome and related connective tissue disorders. *Genet Test* 1997;1:237–242.
17. Pepe G, Giusti B, Evangelisti L, Porciani MC, Brunelli T, Giurlani L, Attanasio M, Fattori R, Bagni C, Comeglio P, Abbate R, Gensini GF. Fibrillin-1 (FBN1) gene frameshift mutations in Marfan patients: genotype-phenotype correlation. *Clin Genet* 2001;59:444–450.
18. Loeys B, De Backer J, Van Acker P, Wettinck K, Pals G, Nuytinck L, Coucke P, De Paepe A. Comprehensive molecular screening of the FBN1 gene favors locus homogeneity of classical Marfan syndrome. *Hum Mutat* 2004;24:140–146.
19. Loeys BL, Schwarze U, Holm T, Callewaert BL, Thomas GH, Pannu H, De Backer JF, Oswald GL, Symoens S, Monouvrier S, Roberts AE, Faravelli F, Greco MA, Pyeritz RE, Milewicz DM, Coucke PJ, Cameron DE, Braverman AC, Byers PH, De Paepe AM, Dietz HC. Aneurysm syndromes caused by mutations in the TGFb receptor. *N Engl J Med* 2006;355:788–798.
20. Guo DC, Pannu H, Tran-Fadulu V, Papke CL, Yu RK, Avidan N, Bourgeois S, Estrera AL, Safi HJ, Sparks E, Amor D, Ades L, McConnell V, Willoughby CE, Abuelo D, Willing M, Lewis RA, Kim DH, Scherer S, Tung PP, Ahn C, Buja LM, Raman CS, Shete SS, Milewicz DM. Mutations in smooth muscle alpha-actin (ACTA2) lead to thoracic aortic aneurysms and dissections. *Nat Genet* 2007;39: 1488–1493.
21. Zhu L, Vranckx R, Khau Van Kien P, Lalande A, Boisset N, Mathieu F, Wegman M, Glancy L, Gasc JM, Brunotte F, Bruneval P, Wolf JE, Michel JB, Jeunemaitre X. Mutations in myosin heavy chain 11 cause a syndrome associating thoracic aortic aneurysm/aortic dissection and patent ductus arteriosus. *Nat Genet* 2006;38:343–349.
22. van de Laar IM, Oldenburg RA, Pals G, Roos-Hesselink JW, de Graaf BM, Verhagen JM, Hoedemaekers YM, Willemsen R, Severijnen LA, Venselaar H, Vriend G, Pattynama PM, Collée M, Majoor-Krakauer D, Poldermans D, Frohn-Mulder IM, Micha D, Timmermans J, Hilhorst-Hofstee Y, Bierma-Zeinstra SM, Willems PJ, Kros JM, Oei EH, Oostra BA, Wessels MW, Bertoli-Avella AM. Mutations in SMAD3 cause a syndromic form of aortic aneurysms and dissections with early-onset osteoarthritis. *Nat Genet* 2011;43:121–126.

HMG Advance Access published April 6, 2012

Human Molecular Genetics, 2012 **1–8**
doi:10.1093/hmg/dds113

A single-nucleotide polymorphism in *ANK1* is associated with susceptibility to type 2 diabetes in Japanese populations

Minako Imamura[1,†], Shiro Maeda[1,5,*,†], Toshimasa Yamauchi[5,6,†], Kazuo Hara[6,7,†], Kazuki Yasuda[9], Takashi Morizono[2], Atsushi Takahashi[3], Momoko Horikoshi[6], Masahiro Nakamura[6], Hayato Fujita[6], Tatsuhiko Tsunoda[2], Michiaki Kubo[4], Hirotaka Watada[5,10], Hiroshi Maegawa[11], Miki Okada-Iwabu[6,8], Masato Iwabu[6,7], Nobuhiro Shojima[6], Toshihiko Ohshige[12], Shintaro Omori[12], Minoru Iwata[13], Hiroshi Hirose[14], Kohei Kaku[15], Chikako Ito[16], Yasushi Tanaka[12], Kazuyuki Tobe[13], Atsunori Kashiwagi[11], Ryuzo Kawamori[5], Masato Kasuga[17], Naoyuki Kamatani[3], Diabetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium, Yusuke Nakamura[18] and Takashi Kadowaki[6,*]

[1]Laboratory for Endocrinology and Metabolism, [2]Laboratory for Medical Informatics, [3]Laboratory for Statistical Analysis, [4]Laboratory for Genotyping Development, RIKEN Center for Genomic Medicine, Yokohama Kanagawa 230-0045, Japan, [5]Sportology Center, Graduate School of Medicine, Juntendo University, Tokyo 113-8421, Japan, [6]Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, [7]Department of Integrated Molecular Science on Metabolic Diseases, 22nd Century Medical and Research Center, [8]Department of Molecular Medicinal Sciences on Metabolic Regulation, 22nd Century Medical and Research Center, University of Tokyo, Tokyo 113-8655 Japan, [9]Department of Metabolic Disorder, Diabetes Research Center, Research Institute, National Center for Global Health and Medicine, Tokyo 162-8655, Japan, [10]Department of Medicine, Metabolism and Endocrinology, School of Medicine, Juntendo University, Tokyo 113-8421, Japan, [11]Department of Medicine, Shiga University of Medical Science, Otsu, Shiga 520-2192, Japan, [12]Department of Internal Medicine, Division of Metabolism and Endocrinology, St. Marianna University School of Medicine, Kawasaki, Kanagawa 216-8511, Japan, [13]First Department of Internal Medicine, University of Toyama, Toyama 930-0194, Japan, [14]Health Center, Keio University School of Medicine, Tokyo 160-8582, Japan, [15]Division of Diabetes, Endocrinology and Metabolism, Department of Internal Medicine, Kawasaki Medical School, Kurashiki, Okayama 701-0192, Japan, [16]Medical Court Life Care Clinic, Hiroshima 730-0012, Japan, [17]Diabetes Research Center, Research Institute, National Center for Global Health and Medicine, Tokyo 162-8655, Japan, and [18]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan

Received November 15, 2011; Revised March 15, 2012; Accepted March 21, 2012

To identify a novel susceptibility locus for type 2 diabetes, we performed an imputation-based, genome-wide association study (GWAS) in a Japanese population using newly obtained imputed-genotype data for 2 229 890 single-nucleotide polymorphisms (SNPs) estimated from previously reported, directly genotyped GWAS data in the same samples (stage 1: 4470 type 2 diabetes versus 3071 controls). We directly genotyped 43 new SNPs with *P*-values of $<10^{-4}$ in a part of stage-1 samples (2692 type 2 diabetes versus 3071 controls), and the associations of validated SNPs were evaluated in another 11 139 Japanese individuals (stage 2: 7605

*To whom correspondence should be addressed at: Laboratory for Endocrinology and Metabolism, RIKEN Center for Genomic Medicine, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. Tel: +81 455039595; Fax: +81 455039567; Email: smaeda@src.riken.jp (S.M.); Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. Tel: +81 358008815; Fax: +81 358009797; Email: kadowaki-3im@h.u-tokyo.ac.jp (T.K.)
†The authors wish it to be known that, in their opinion, the first four authors (M.I., S.M., T.Y. and K.H.) should be regarded as joint First Authors.

type 2 diabetes versus 3534 controls). Combined meta-analysis using directly genotyped data for stages 1 and 2 revealed that rs515071 in *ANK1* and rs7656416 near *MGC21675* were associated with type 2 diabetes in the Japanese population at the genome-wide significant level ($P < 5 \times 10^{-8}$). The association of rs515071 was also observed in European GWAS data (combined P for all populations = 6.14×10^{-10}). Rs7656416 was in linkage disequilibrium to rs6815464, which had recently been identified as a top signal in a meta-analysis of East Asian GWAS for type 2 diabetes ($r^2 = 0.76$ in stage 2). The association of rs7656416 with type 2 diabetes disappeared after conditioning on rs6815464. These results indicate that the *ANK1* locus is a new, common susceptibility locus for type 2 diabetes across different ethnic groups. The signal of association was weaker in the directly genotyped data, so the improvement in signal indicates the importance of imputation in this particular case.

INTRODUCTION

Genome-wide association studies (GWAS) for type 2 diabetes have been conducted extensively and have successfully identified over 40 susceptibility loci, mostly in European populations (1,2). The first round of GWAS for type 2 diabetes reported in 2007 confirmed five new loci—*HHEX*, *SLC30A8*, *CDKAL1*, *CDKN2A-CDKN2B* and *IGF2BP2* (3–7)—in addition to three previously reported loci—*TCF7L2* (8), *PPARG* (9) and *KCNJ11* (10). The Wellcome Trust Case Control Consortium/United Kingdom Type 2 Diabetes Genetics Consortium (WTCCC/UKT2D) study also identified a strong association between *FTO* variants and type 2 diabetes, although the effect of the *FTO* variants was mostly mediated through an increased body weight (11). After the first round of European GWAS, additional studies combined individual GWAS data to increase the sample size and make common variants with lower effect sizes detectable. These studies have so far identified more than 30 additional susceptibility loci for type 2 diabetes (12–14). Additionally, some of them have been shown to confer similar susceptibility to type 2 diabetes in non-European populations (15–19). However, the integration of all this information can explain only ~10% of type 2 diabetes heritability (1,2,13), suggesting that most of the genetic factors for the condition remain to be identified, especially in non-European populations. Cumulative evidence suggests that Asians may be more genetically susceptible to type 2 diabetes than populations of European ancestry (20). Also, there are significant interethnic differences in the risk allele frequency or in effect sizes at several loci, which may affect the power to detect the associations in these populations (2,20). Therefore, it is considered to be relevant to perform GWAS for type 2 diabetes using non-European populations as well as European populations to uncover the missing heritability of type 2 diabetes.

In 2008, two Japanese GWAS simultaneously identified the *KCNQ1* locus as a strong susceptibility locus for type 2 diabetes (21,22). Recently, we performed a larger scale Japanese GWAS, identifying the additional loci *UBE2E2* and *C2CD4A-C2CD4B* (23). The associations of *KCNQ1* and *C2CD4A-C2CD4B* with type 2 diabetes were consistently observed among European populations, underlining the importance of examining non-European populations through GWAS. This will help us to identify not only ethnicity-specific loci, but also common-susceptibility loci among different ethnic groups.

Here, we show the results of an imputation-based GWAS as an extended analysis of our previous Japanese GWAS for 459 359 directly genotyped single-nucleotide polymorphisms (SNPs). We obtained imputed-genotype data for 2 229 890 SNPs estimated from 459 359 directly genotyped SNPs in our previous report (23). The analysis using over 2 million of newly obtained imputed SNPs data and subsequent *in silico* replication study in European GWAS data provide the evidence that the *ANK1* locus is a novel common-susceptibility locus for type 2 diabetes across different ethnic groups.

RESULTS

We successfully obtained the new information of 2 229 890 imputed SNPs with a quality score (proper_info) of >0.40, minor allele frequency (MAF) of >0.01, and Hardy–Weinberg equilibrium (HWE) P-value of $>1 \times 10^{-6}$ by using IMPUTE with previously reported GWAS data (459 359 directly genotyped SNPs) (stage 1) (23) and from 89 HapMap samples (44 JPT and 45 CHB in HapMap phase 2).

Among 2 229 890 imputed SNPs, we found that the *KCNQ1* locus appeared as the top signal, although the association did not attain genome-wide significance levels (rs2237896; $P = 5.9 \times 10^{-8}$, rs2283228; $P = 8.7 \times 10^{-8}$). We did not observe evidence for population stratification in stage-1 samples [Supplementary Material, Fig. S1, genomic inflation score (λ_{GC}) = 1.07347]. We also identified 330 SNPs with P-values between 1×10^{-7} and 1×10^{-4} that were derived from 70 distinct loci, including 7 already confirmed loci and 27 loci evaluated in the prior analysis (23; Supplementary Material, Table S1). Two hundred and sixty-six SNPs within these 34 (7 + 27) loci and 23 proxies ($r^2 > 0.8$) out of the remaining 66 SNPs in 36 loci were excluded from further analysis in order to focus on identifying novel T2D susceptible loci. Therefore, we selected 43 SNPs within the 36 loci and directly genotyped these SNPs using a part of stage-1 samples (2692 type 2 diabetes versus 3071 controls); the remaining 1778 type 2 diabetes samples were not available for the direct genotyping. In this analysis, we successfully obtained information for 40 SNPs. Among them, 10 were excluded from stage-2 analysis because an association study using directly genotyped data showed that they were not associated with type 2 diabetes ($P \geq 0.01$, Supplementary Material, Table S2).

In stage-2 analysis (7605 type 2 diabetes versus 3534 controls), four SNPs were associated with type 2 diabetes ($P < 0.01$), although none of the SNPs showed an association with a genome-wide significance level.

Next, we performed the combined meta-analysis by using directly genotyped data for stages 1 and 2 using the Mantel–Haenszel procedure. In this combined analysis, we found that two SNPs—rs515071 in *ANK1* and rs7656416 near *MGC21675*—were significantly associated with type 2 diabetes in the Japanese population [rs515071: $P = 1.37 \times 10^{-8}$, odds ratio (OR) = 1.18, 95% CI 1.12–1.25, rs7656416: $P = 1.37 \times 10^{-8}$, OR = 1.15, 95% CI 1.10–1.21, Table 1, Supplementary Material, Table S3]. We identified additional two SNPs associated with type 2 diabetes, but the association did not attain genome-wide significance level (Supplementary Material, Table S3, rs1327796: $P = 3.17 \times 10^{-6}$, rs10993738: $P = 4.61 \times 10^{-6}$). The association of these SNPs with type 2 diabetes was not affected by adjusting for age, sex or body mass index (BMI; Supplementary Material, Table S4). We then searched for data on the top SNP—rs515071 in *ANK1*—in publicly available, European GWAS studies. Analysis of these data showed that rs515071 was also associated with type 2 diabetes in European populations [$P = 0.0129$, OR = 1.1, 95 % CI 1.02–1.19, combined data for WTCCC/UKT2D and the Diabetes Genetics Initiative (DGI), Table 2]. The association of rs515071 with type 2 diabetes was further strengthened in a larger European GWAS meta-analysis data [13; $P = 8.54 \times 10^{-4}$, OR = 1.09, 95% CI 1.03–1.14, DIAGRAM Diabetes Genetics Replication and Meta-analysis (DIAGRAM) Table 2]. The effect direction of rs515071 was consistent throughout all studies. In the previously reported Japanese GWAS data using directly genotyped SNPs, the best *P*-value for directly genotyped SNPs in this region was $>1 \times 10^{-4}$ (rs6989203, $P = 4.65 \times 10^{-4}$), whereas the *P*-value of the top imputed signal for this region (rs515071) in the present study (stage 1) was 2.69×10^{-5} (Supplementary Material, Tables S1 and S2).

Recently, the SNP rs6815464, located 54 kb downstream of rs7656416, was shown to be associated with type 2 diabetes in East Asian genome-wide association meta-analysis (24). Therefore, we also genotyped rs6815464 in stage-2 samples, which did not overlap with the samples in the study for East Asian meta-analysis. We found that this SNP was also associated with type 2 diabetes and that these two SNPs were in modest linkage disequilibrium (LD; $r^2 = 0.76$ in our stage-2 samples, Supplementary Material, Fig. S2 and Table S5). Subsequent conditional analysis that included both rs7656416 and rs6815464 in the same logistic regression model revealed that conditioning the SNPs on each other removed their significance (before conditioning: rs7656416, $P = 1.01 \times 10^{-5}$; rs6815464; $P = 1.26 \times 10^{-5}$; after conditioning: rs7656416, $P = 0.22$; rs6815464, $P = 0.29$, Supplementary Material, Table S5).

We further examined the association of rs515071 and rs7656416 with quantitative metabolic traits among control participants. Rs515071-C, the risk allele for type 2 diabetes, was modestly associated with a decrease in BMI (beta = −0.012, S.E. 0.005, $P = 0.016$, adjusting age and sex, Table 3). Participants without a risk allele for diabetes (TT, $n = 115$) showed higher BMI than those with a homozygote

Table 1. Two SNPs significantly associated with type 2 diabetes in the Japanese population

Nearest gene	Stage 1[a] RAF		Stage 2 RAF		Stage 1 + 2[b]		Heterogeneity		Stage 1 + 2[c]
	Cases	Controls	Cases	Controls	Combined *P*	OR (95% CI)	Q	I^2	Combined *P*
Rs515071 **C** > T Ch 8 *ANK1*	0.826	0.804	0.822	0.794	1.37×10^{-8}	1.18 (1.12–1.25)	0.59	0	6.45×10^{-10}
Rs7656416 **C** > T Ch 4 *MGC21675*	0.706	0.675	0.705	0.675	1.37×10^{-8}	1.15 (1.10–1.21)	0.96	0	1.32×10^{-9}

Combined *P*-values with a fixed effect model are presented.
Ch, chromosome; RAF, risk allele frequency; OR, odds ratio;
Q, *P*-values for Cochran Q statistics; $I^2 = [Q \text{ statistics} - (K - 1)]/Q$ statistics, *K* is the number of the study.
[a]Direct genotyped data in 2692 type 2 diabetes and 3071 controls.
[b]Direct genotyped stage 1 + stage 2 data.
[c]Imputed data after GC correction in 4470 type 2 diabetes and 3071 controls (stage 1) + stage 2 data.

Table 2. Association of rs515071 in the *ANK1* locus with type 2 diabetes

	Sample size		CC/CT/TT (RAF)		*P*	OR (95% CI)	Q	I^2
	Cases	Controls	Cases	Controls				
1. Stage 1 (GWAS)	2692	3071	1804/757/80, −0.826	1905/950/104, −0.804	2.57×10^{-3}	1.16 (1.05–1.27)		
2. Stage 2	7605	3534	4806/2053/239, −0.822	2148/1116/145, −0.794	1.33×10^{-6}	1.20 (1.11–1.29)		
1 + 2, all Japanese	10 297	6605			1.37×10^{-8}	1.18 (1.12–1.25)	0.59	0
WTCCC/UKT2D	1924	2938	1167/669/86, −0.781	1696/1058/177, −0.759	0.012	1.13 (1.03–1.25)		
DGI	1464	1467	949/442/40, −0.818	933/450/47, −0.81	0.441	1.05 (0.92–1.20)		
3. WTCCC/UKT2D + DGI	3388	4405			0.0129	1.10 (1.02–1.19)	0.38	0
1 + 2 + 3	13 685	11 010			1.59×10^{-9}	1.15 (1.10–1.21)	0.4	0
4. DIAGRAM v2	22 570		N/A	N/A	8.54×10^{-4}	1.09 (1.03–1.14)		
1 + 2 + 4	39 472				6.14×10^{-10}	1.12 (1.08–1.17)	0.07	62.33

Combined *P*-values with a fixed effect model are presented.
RAF, risk allele frequency; OR, odds ratio; *P*; *P*-value is calculated with additive model;
Q, *P*-values for Cochran Q statistics, I^2 = [Q statistics − $(K - 1)$]/Q statistics, K = number of the study.

Table 3. Association of rs515071-C and rs7656416-C with BMI[a]

	β (S.E.)	*P*-value	BMI (mean ± SD)		
rs515071					
Unadjusted	−0.013 (0.005)	0.008	CC ($n = 1527$)	CT ($n = 813$)	TT ($n = 115$)
Adjusted[b]	−0.012 (0.005)	0.016	23.0 ± 3.3	23.2 ± 3.4	24.0 ± 3.3[c]
rs7656416					
Unadjusted	−0.004 (0.004)	0.39	CC ($n = 1129$)	CT ($n = 1107$)	TT ($n = 265$)
Adjusted[b]	−0.004 (0.004)	0.37	23.0 ± 3.4	23.2 ± 3.3	23.1 ± 3.2

β, regression coefficient; *P*, *P*-values are calculated on linear regression analysis with additive model.
[a]Log-transformed BMI was used for the linear regression analysis.
[b]Adjusted for age and sex.
[c]$P < 0.05$ TT versus CC.

Table 4. Association of rs515071-C and rs7656416-C with quantitative glycemic traits

Adjustment	FPG (mmol/l)		HOMA-IR[a]		HOMA-β[a]	
	β (S.E.)	*P*-value	β (S.E.)	*P*-value	β (S.E.)	*P*-value
rs515071						
–	−0.022 (0.019)	0.26	−0.038 (0.024)	0.11	−0.026 (0.024)	0.29
Sex, age	−0.020 (0.019)	0.31	−0.034 (0.024)	0.15	−0.024 (0.024)	0.33
Sex, age, BMI[a]	−0.011 (0.019)	0.57	0.003 (0.02)	0.88	0.005 (0.022)	0.81
rs7656416						
–	0.015 (0.017)	0.38	0.006 (0.021)	0.79	−0.007 (0.021)	0.74
Sex, age	0.013 (0.017)	0.43	0.005 (0.021)	0.8	−0.006 (0.021)	0.77
Sex, age, BMI[a]	0.016 (0.017)	0.35	0.01 (0.018)	0.6	−0.001 (0.019)	0.95

β, regression coefficient; *P*, *P*-values are calculated on linear regression analysis with additive model.
[a]Values were log-transformed and used for the analysis.
Sample size: FPG, $n = 2347$; HOMA-IR, $n = 1623$; HOMA-β, $n = 1623$.

of the risk allele (CC, $n = 1527$; $P = 0.01$, 24.0 ± 3.3, compared with 23.0 ± 3.3). Meanwhile, no association was observed between rs7656416 and BMI.

We did not observe an association of rs515071 or rs7656416 with any glycemic traits, such as fasting plasma glucose (FPG), homeostasis model assessment of insulin resistance (HOMA-IR) or homeostasis model assessment of beta-cell function (HOMA-β) (Table 4).

DISCUSSION

By using imputation-based GWAS, we identified a novel susceptibility variant for type 2 diabetes at the *ANK1* locus. We also identified a strong signal at *MGC21675*, located in the same LD block as the *MAEA*, recently reported as a top signal in a meta-analysis of GWAS for East Asian type 2 diabetes.

Currently performed GWAS have examined ~1 000 000 directly genotyped SNPs and additional ~2 000 000 imputed SNPs. Those genotypes are estimated based on the degree of LD in directly genotyped alleles. The accuracy of imputed SNPs that pass the quality-control standards (proper_info for IMPUTE, r^2 for MACH) can widely be accepted in European populations. This bioinformatics technology significantly contributes to the identification of additional novel loci or SNPs more strongly associated with the disease. Approximately 40 loci have been identified and confirmed by examining more than 2 million directly genotyped and imputed SNPs in European populations. Recently, an East Asian study group identified eight novel loci by using the same strategy as the European GWAS meta-analysis (24). However, integration of all these data is still not sufficient to completely explain type 2 diabetes heritability. Thus, more efforts are necessary to identify additional susceptibility variants for the disease, especially among non-European populations.

In the present study, we identified the SNP rs515071 located at an intron of the Ankyrin1 gene as a novel susceptibility variant for type 2 diabetes. Because the analysis of European GWAS data has also shown a significant association between rs515071 and type 2 diabetes, rs515071 is likely a common locus for type 2 diabetes across multiethnic populations. This confirms the importance of extended analyses in multiethnic groups.

ANK1 is located on chromosome 8p11.1 and encodes a member of the ankyrin family. The ankyrins act as adaptors among a variety of integral membrane proteins and the spectrin skeleton (25). Ankyrin1, the prototype of this family, was first discovered in the erythrocytes, but it has also been found in the brain and muscle cells. In humans, mutations to *ANK1* cause hereditary spherocytosis; therefore, *ANK1* has been considered pivotal in stabilizing the membrane structure of erythrocytes (25). Recently, variants in *ANK1*—rs4737009 and rs6474359—were shown to influence HbA1c levels in European, non-diabetic adults (rs4737009, $\beta = 0.027$, S.E. $= 0.004$, $P = 6.11 \times 10^{-12}$; rs6474359, $\beta = 0.058$, S.E. $= 0.011$, $P = 1.18 \times 10^{-8}$) (26). In the same report, the effect size of the *ANK1* variants has been shown to remain essentially unchanged after conditioning by either fasting or 2 h plasma glucose levels, and neither SNPs were associated with type 2 diabetes ($P = 0.069$, OR $= 1.05$, 95% CI 1.00–1.10). Therefore, the association of the *ANK1* variants with HbA1c was likely mediated by non-glycemic factors. Furthermore, *ANK1* variants may influence erythrocyte lifespan and lower HbA1c levels without affecting plasma-glucose levels.

In the present study, however, rs515071 in *ANK1* was significantly associated with susceptibility to type 2 diabetes and was found to be in weak LD with rs4737009 ($r^2 = 0$ in JPT, CEU) and rs6474359 ($r^2 = 0.22$ in JPT and 0 in CEU; Supplementary Material, Fig. S3). We further performed conditional analysis of rs515071 for type 2 diabetes susceptibility including two reported variants rs4737009 and rs6474359 into the same logistic model as co-valuables, and the results indicated that the association of rs515071 with type 2 diabetes was independent of these two SNPs (Supplementary Material, Table S8). In addition, we could not observe any significant association of these three SNPs with HbA1c levels in our stage-2 controls (Supplementary Material, Table S9). Taken together, we concluded that the association of rs515071 with type 2 diabetes is independent of already reported association signals for affecting HbA1c levels. The mechanisms by which the SNP in the *ANK1* contributes to susceptibility to type 2 diabetes are unknown.

We also examined the expression profile of *ANK1* in various tissues, clearly observing its expression in human islet, pancreas, skeletal muscle, adipose and liver tissues, along with the mouse pancreatic β-cell line, all of which are important organs for glucose metabolism (Supplementary Material, Fig. S4). The physiological or pathological role of ankyrin1 in pancreatic β-cell has not yet been reported. However, another member of the ankyrin family, ankyrin B (also termed ankyrin 2), is known to regulate K-ATP channel membrane trafficking and gating in excitable cells. In pancreatic β-cells, ankyrin B directly interacts with potassium inward rectifier 6.2 (Kir6.2)/sulphonylurea receptor ATP-sensitive potassium (K_{ATP}) channel and plays a key role in regulating ATP sensitivity (27). On the other hand, SNPs in the *ANK1* promoter have been reported to be associated with intramuscular fat in bovine or porcine tissues (28,29), suggesting that *ANK1* also contributes to the development of muscular insulin resistance (30). Therefore, *ANK1* may be one of the genes conferring susceptibility to type 2 diabetes, although a possibility still exists that other nearby genes confer the true causal effects. Although we did not observe a significant association between the *ANK1* SNP and glycemic traits (FPG, HOMA-IR or HOMA-β) in our limited, non-diabetic controls, the susceptibility allele for type 2 diabetes (rs515071-C) modestly reduced BMI. This might operate via the effects on lean body mass, because ankyrin1 appears to have a role in the organization of myofibrils during assembly and seems to cooperate with obscurin in mediating interactions between the sarcoplasmic reticulum and myofibrils (31,32). Because the association between rs515071 and type 2 diabetes was not affected by adjustment of BMI (Supplementary Material, Table S4), the effects of the variant on type 2 diabetes susceptibility and reducing BMI are probably independent of each other.

We also found another significant signal at the rs7656416 locus, which was not identified in the previous Japanese GWAS. This site is located on chromosome 4p16.3 near *MGC21675* (alternatively, *C4orf42*), which encodes the hypothetical protein *LOC92070*. Recently, eight novel loci for type 2 diabetes were identified in an East Asian GWAS meta-analysis for type 2 diabetes, including rs6815464 at the *MAEA* locus. Rs6815464, located at 54 kb downstream of rs7656416, was identified as a top signal in the East Asian meta-analysis. Both SNPs are located in the same LD block. The association of these SNPs with type 2 diabetes disappeared after conditioning on one another in the present stage-2 samples, indicating that our results for rs7656416 simply replicated the previously identified association of this locus with type 2 diabetes. There are several genes in this locus, including *MGC21675*, macrophage erythroblast attacher (*MAEA*) and C-terminal binding protein-1 (*CTBP1*). Studies have not yet elucidated the roles of the proteins encoded by these genes in pancreatic β-cell or peripheral tissues involved in glucose metabolism. Further studies,

including fine mapping and functional analysis, are needed to clarify the mechanisms by which these variants confer susceptibility to type 2 diabetes.

Our present findings, identification of two loci for type 2 diabetes susceptibility with genome-wide significant levels, indicate that imputation-based genotype data are also useful in Japanese populations, because both loci, *ANK1* and *MGC21675-MAEA*, did not show strong evidence being associated with type 2 diabetes in a previously reported direct-genotyped GWAS data ($P > 10^{-4}$). However, in some cases, there are significant discrepancies between the imputation-based data and directly genotyped data (Supplementary Material, Table S10), probably resulting from insufficient sample number or no information for trio in the reference panel we used. Therefore, in contrast to the European data, association studies using imputation-based genotyped data should be performed with some caution in Japanese populations. Currently, the efforts to improve the quantity and quality of data for reference panels, and to improve the imputation programs themselves, are in progress and will further contribute to advances in genetic studies.

In summary, using imputation-based GWAS, we identified *ANK1* as a novel locus associated with type 2 diabetes at genome-wide significance levels in Japanese populations. The risk allele rs515071-C in *ANK1* was associated with type 2 diabetes susceptibility in European populations as well. Additionally, we showed that the association of the *MGC21675-MAEA* locus with type 2 diabetes also occurs among Japanese populations.

MATERIALS AND METHODS

Participants, DNA preparation and SNP genotyping

Stage-1 samples

For the GWAS, we selected case–control samples (4470 cases and 3071 controls) from subjects enrolled in the BioBank Japan, as previously reported (23). We selected type 2 diabetes cases from individuals registered as having type 2 diabetes. Control groups were healthy volunteers or individuals registered as individuals not having type 2 diabetes but with diseases other than type 2 diabetes, comprised of 13 distinct diseases.

Stage-2 samples

We selected another 7605 cases from the BioBank Japan or from subjects with type 2 diabetes who visited outpatient clinics at one of the nine different institutions: The University of Tokyo, Juntendo University, National Center for Global Health and Medicine, Hiranuma Clinic, St. Marianna University School of Medicine, the Hiroshima Atomic Bomb Casualty Council Health Management Center, Kawasaki Medical School, Toyama University Hospital or the Shiga University of Medical Science. We also examined 3534 controls enrolled during an annual health check-up at six institutions: The Hiroshima Atomic Bomb Casualty Council Health Management Center, The National Center for Global Health and Medicine, Keio University, Hiranuma Clinic, St. Marianna University School of Medicine or Toyama University Hospital. Diabetes was diagnosed according to World Health Organization (WHO) criteria. We excluded individuals who were positive for antibody to glutamic acid decarboxylase or those with diabetes due to (i) liver dysfunction, (ii) steroids and other drugs that might raise glucose levels, (iii) malignancy or (iv) a monogenic disorder known to cause diabetes. Clinical characteristics of stage-1 and -2 participants are shown in Supplementary Material, Table S6.

Genomic DNA was extracted from peripheral leukocytes using the standard phenol–chloroform procedure. Genotyping in stage-1 validation and stage 2 was performed using the multiplex-polymerase chain reaction (PCR)-invader assay (21,23).

The protocol was approved by the ethics committee of the Institute of Physical and Chemical Research (RIKEN), the University of Tokyo and each participating institution (Juntendo University, National Center for Global Health and Medicine, Keio University, St. Marianna University School of Medicine, Kawasaki Medical School, Toyama University and the Shiga University of Medical Science).

Genome-wide imputation by the IMPUTE

We performed genome-wide imputation by using IMPUTE (https://mathgen.stats.ox.ac.uk/impute/impute.html) with previously reported GWAS data (459 359 directly genotyped SNPs; stage 1; 23) and from 89 HapMap samples (44 JPT and 45 CHB in HapMap phase 2). We successfully obtained the new information of 2 229 890 imputed SNPs with a quality score (proper_info) of >0.40, MAF of >0.01 and HWE P-value of $>1 \times 10^{-6}$.

Cell culture

Hepa1-6 and C2C12 were purchased from ATCC (Manassas, VA, USA). 3T3-L1 was purchased from Health Science Research Resources Bank (Sennan, Japan). MIN6-m9 cells were kindly provided by Prof. Susumu Seino (Kobe University, Kobe, Japan).

Hepa1-6 and C2C12 cells were maintained in Dulbecco's Modified Eagle Medium, containing 10% fetal bovine serum (FBS) and 1% penicillin–streptomycin. The differentiation in C2C12 was induced by depleting FBS from 10 to 1% and subsequently culturing for 8 days. MIN6-m9 cells were cultured as previously described (33). 3T3-L1 cells were maintained and their differentiation was induced as described previously (34). Differentiated 3T3-L1 cells were harvested 8 days after initiating induction.

Quantitative reverse transcription-PCR

Each cell was harvested at the indicated time and the total RNA was extracted using the RNeasy Kit (Qiagen, Germantown, MD, USA). First-strand cDNAs were synthesized using the PrimeScript® II 1st Strand cDNA Synthesis Kit (Takara Bio, Inc., Otsu, Japan) following the manufacturer's protocol. We obtained human cDNAs from multiple tissues from CLONTECH, Inc. (Palo Alto, CA, USA). Human islet cDNA were kindly provided by Primary Cell Co., Ltd. (Sapporo, Japan).

The amount of first-strand cDNAs was quantified using SYBR premix Ex Taq II (Takara Bio Inc., Otsu, Japan) for amplification and Mx3000P multiplex quantitative PCR system (Stratagene, La Jolla, CA, USA) for detection. The thermal profile was 95°C for 30 s, followed by 40 cycles of 95°C for 5 s and 60°C for 30 s. Relative expressions of human ankyrin1 isoforms were normalized with GAPDH, and the expressions of mouse ankyrin1 were normalized with normalization factor derived from mouse Eef1g, Hmbs and Ppia, calculated using GeNorm software (http://medgen.ugent.be/~jvdesomp/genorm/).

The primers for quantitative PCR are described in Supplementary Material, Table S7.

Statistical analysis

Statistical methods for determining the associations and calculating the LD coefficients (r^2) have previously been described (23). We performed the HWE test according to a previously described method (35). The cut-off value for the HWE test in the control groups was 0.000001 for the first stage. The SNPs with *P*-values less than this were excluded from the analysis. We performed the imputation GWAS by SNPTEST (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html) and used gene dosages for the analysis. As for directly genotyped data in Stage 1 validation and Stage 2, we analyzed the differences between the case and control groups in genotype distribution by using the Armitage test for trends, based on an additive model, as previously described (21,23). Combined meta-analysis was performed using the Mantel–Haenszel procedure with a fixed-effect model after testing for heterogeneity. We performed quantitative trait analyses for BMI, FPG, HbA1c, HOMA-β and HOMA-IR by using multiple linear regression analysis in an additive association model with or without adjusting for age, sex and log-transformed BMI. Because the Japanese samples studied here show the skewed distribution values for BMI, HOMA-IR and HOMA-β, we have analyzed the quantitative traits by using log-transformed BMI, HOMA-IR and HOMA-β. Tests for multiple comparisons were performed by analysis of variance followed by Scheffe's *post hoc* procedure.

AUTHORS' ROLE

M.Im., T.K., S.M., T.Y. and K.H. planned and coordinated the study; Y.N. and M.K. managed BioBank Japan; K.H., S.M., K.Y., M.H., M.N., H.F., H.W., H.M., M.O-I., M.Iw.. N.S., T.O., S.O., M.Iw., H.H., K.K., C.I., Y.T., K.T., A.K., R.K., M.K. and T.K. recruited and phenotyped or genotyped the patient cohorts; M.Im., T.M., A.T., T.T., N.K., S.M., K.H. and T.K. analyzed the genotyping data; M.Im. and S.M. wrote the manuscript; All authors contributed to the final version of the manuscript.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

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. Finally, we thank all the participants and the staff of the BioBank Japan project.

Conflict of Interest statement. None declared.

FUNDING

This work was partly supported by a grant from the Leading Project of Ministry of Education, Culture, Sports, Science and Technology, Japan.

REFERENCES

1. McCarthy, M.I. (2010) Genomics, type 2 diabetes, and obesity. *N. Engl. J. Med.*, **363**, 2339–2350.
2. Imamura, M. and Maeda, S. (2011) Genetics of type 2 diabetes: the GWAS era and future perspectives. *Endocr. J.*, **58**, 723–739.
3. 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.
4. Steinthorsdottir, V., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Jonsdottir, T., Walters, G.B., Styrkarsdottir, U., Gretarsdottir, S., Emilsson, V., Ghosh, S. *et al.* (2007) A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat. Genet.*, **39**, 770–775.
5. Saxena, R., Voight, B.F., Lyssenko, V., Burtt, N.P., de Bakker, P.I., Chen, H., Roix, J.J., Kathiresan, S., Hirschhorn, J.N., Daly, M.J. *et al.* (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*, **316**, 1331–1336.
6. Zeggini, E., Weedon, M.N., Lindgren, C.M., Frayling, T.M., Elliott, K.S., Lango, H., Timpson, N.J., Perry, J.R., Rayner, N.W. and Freathy, R.M. (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*, **316**, 1336–1341.
7. 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.
8. Grant, S.F., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Manolescu, A., Sainz, J., Helgason, A., Stefansson, H., Emilsson, V., Helgadottir, A. *et al.* (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat. Genet.*, **38**, 320–323.
9. Altshuler, D., Hirschhorn, J.N., Klannemark, M., Lindgren, C.M., Vohl, M.C., Nemesh, J., Lane, C.R., Schaffner, S.F., Bolk, S., Brewer, C. *et al.* (2000) The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.*, **26**, 76–80.
10. Gloyn, A.L., Weedon, M.N., Owen, K.R., Turner, M.J., Knight, B.A., Hitman, G., Walker, M., Levy, J.C., Sampson, M., Halford, S. *et al.* (2003) Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes*, **52**, 568–572.
11. Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R., Elliott, K.S., Lango, H., Rayner, N.W. *et al.* (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, **316**, 889–894.
12. 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.
13. 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.
14. Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A.U., Wheeler, E., Glazer, N.L., Bouatia-Naji, N., Gloyn, A.L. *et al.* (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.*, **42**, 105–116.
15. Hayashi, T., Iwamoto, Y., Kaku, K., Hirose, H. and Maeda, S. (2007) Replication study for the association of TCF7L2 with susceptibility to type 2 diabetes in a Japanese population. *Diabetologia*, **50**, 980–984.
16. Horikoshi, M., Hara, K., Ito, C., Nagai, R., Froguel, P. and Kadowaki, T. (2007) A genetic variation of the transcription factor 7-like 2 gene is associated with risk of type 2 diabetes in the Japanese population. *Diabetologia*, **50**, 747–751.
17. Omori, S., Tanaka, Y., Takahashi, A., Hirose, H., Kashiwagi, A., Kaku, K., Kawamori, R., Nakamura, Y. and Maeda, S. (2008) Association of CDKAL1, IGF2BP2, CDKN2A/B, HHEX, SLC30A8, and KCNJ11 with susceptibility to type 2 diabetes in a Japanese population. *Diabetes*, **57**, 791–795.
18. Horikawa, Y., Miyake, K., Yasuda, K., Enya, M., Hirota, Y., Yamagata, K., Hinokio, Y., Oka, Y., Iwasaki, N., Iwamoto, Y. *et al.* (2008) Replication of genome-wide association studies of type 2 diabetes susceptibility in Japan. *J. Clin. Endocrinol. Metab.*, **93**, 3136–3141.
19. Horikoshi, M., Hara, K., Ito, C., Shojima, N., Nagai, R., Ueki, K., Froguel, P. and Kadowaki, T. (2007) Variations in the HHEX gene are associated with increased risk of type 2 diabetes in the Japanese population. *Diabetologia*, **50**, 2461–2466.
20. Chan, J.C.N., Malik, V., Jia, W., Kadowaki, T., Yajnik, C.S., Yoon, K.H. and Hu, F.B. (2009) Diabetes in Asia: epidemiology, risk factors, and pathophysiology. *J. Am. Med. Assoc.*, **301**, 2129–2140.
21. 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 Asian and European populations. *Nat. Gene.*, **40**, 1098–1102.
22. 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.
23. 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.
24. Cho, Y.S., Chen, C.H., Hu, C., Long, J., Ong, R.T.H., Sim, X., Takeuchi, F., Wu, Y., Go, M.J., Yamauchi, T. *et al.* (2012) East Asian genome-wide association meta-analysis identifies 8 new loci for type 2 diabetes. *Nat. Genet.*, **44**, 67–72.
25. Bennett, V. and Baines, A.J. (2001) Spectrin and ankyrin-based pathways: metazoan inventions for integrating cells into tissues. *Physiol. Rev.*, **81**, 1353–1392.
26. Soranzo, N., Sanna, S., Wheeler, E., Gieger, C., Radke, D., Dupuis, J., Bouatia-Naji, N., Langenberg, C., Prokopenko, I., Stolerman, E. *et al.* (2010) Common variants at 10 genomic loci influence hemoglobin A1 (C) levels via glycemic and nonglycemic pathways. *Diabetes*, **59**, 3229–3239.
27. Kline, C.F., Kurata, H.T., Hund, T.J., Cunha, S.R., Koval, O.M., Wright, P.J., Christensen, M., Anderson, M.E., Nichols, C.G. and Mohler, P.J. (2009) Dual role of K ATP channel C-terminal motif in membrane targeting and metabolic regulation. *Proc. Natl Acad. Sci. USA*, **106**, 16669–16674.
28. Aslan, O., Sweeney, T., Mullen, A.M. and Hamill, R.M. (2010) Regulatory polymorphisms in the bovine Ankyrin 1 gene promoter are associated with tenderness and intramuscular fat content. *BMC. Genet.*, **11**, 111.
29. Aslan, O., Hamill, R.M., Mullen, A.M., Davey, G.C., Gil, M., Gladney, C.D. and Sweeney, T. (2012) Association between promoter polymorphisms in a key cytoskeletal gene (Ankyrin 1) and intramuscular fat and water-holding capacity in porcine muscle. *Mol. Biol. Rep.*, **39**, 3903–3914.
30. Hegarty, B.D., Furler, S.M., Ye, J., Cooney, G.J. and Kraegen, E.W. (2003) The role of intramuscular lipid in insulin resistance. *Acta. Physiol. Scand.*, **178**, 373–383.
31. Busby, B., Oashi, T., Willis, C.D., Ackermann, M.A., Kontrogianni-Konstantopoulos, A., Mackerell, A.D. Jr and Bloch, R.J. (2011) Electrostatic interactions mediate binding of obscurin to small ankyrin 1: biochemical and molecular modeling studies. *J. Mol. Biol.*, **408**, 321–334.
32. Kontrogianni-Konstantopoulos, A., Jones, E.M., Van Rossum, D.B. and Bloch, R.J. (2003) Obscurin is a ligand for small ankyrin 1 in skeletal muscle. *Mol. Biol. Cell.*, **14**, 1138–1148.
33. Minami, K., Yano, H., Miki, T., Nagashima, K., Wang, C.Z., Tanaka, H., Miyazaki, J.I. and Seino, S. (2000) Insulin secretion and differential gene expression in glucose-responsive and -unresponsive MIN6 sublines. *Am. J. Physiol. Endocrinol. Metab.*, **279**, E773–E781.
34. Imamura, M., Inoguchi, T., Ikuyama, S., Taniguchi, S., Kobayashi, K., Nakashima, N. and Nawata, H. (2002) ADRP stimulates lipid accumulation and lipid droplet formation in murine fibroblasts. *Am. J. Physiol. Endocrinol. Metab.*, **283**, E775–E783.
35. Nielsen, D.M., Ehm, M.G. and Weir, B.S. (1998) Detecting marker-disease association by testing for Hardy–Weinberg disequilibrium at a marker locus. *Am. J. Hum. Genet.*, **63**, 1531–1540.

nature genetics

Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians

Yoon Shin Cho[1,46], Chien-Hsiun Chen[2,3,46], Cheng Hu[4,46], Jirong Long[5,46], Rick Twee Hee Ong[6,46], Xueling Sim[7,46], Fumihiko Takeuchi[8,46], Ying Wu[9,46], Min Jin Go[1,46], Toshimasa Yamauchi[10,46], Yi-Cheng Chang[11,46], Soo Heon Kwak[12,46], Ronald C W Ma[13,46], Ken Yamamoto[14,46], Linda S Adair[15], Tin Aung[16,17], Qiuyin Cai[5], Li-Ching Chang[2], Yuan-Tsong Chen[2], Yutang Gao[18], Frank B Hu[19], Hyung-Lae Kim[1,20], Sangsoo Kim[21], Young Jin Kim[1], Jeannette Jen-Mai Lee[22], Nanette R Lee[23], Yun Li[9,24], Jian Jun Liu[25], Wei Lu[26], Jiro Nakamura[27], Eitaro Nakashima[27,28], Daniel Peng-Keat Ng[22], Wan Ting Tay[16], Fuu-Jen Tsai[3], Tien Yin Wong[16,17,29], Mitsuhiro Yokota[30], Wei Zheng[5], Rong Zhang[4], Congrong Wang[4], Wing Yee So[13], Keizo Ohnaka[31], Hiroshi Ikegami[32], Kazuo Hara[10], Young Min Cho[12], Nam H Cho[33], Tien-Jyun Chang[11], Yuqian Bao[4], Åsa K Hedman[34], Andrew P Morris[34], Mark I McCarthy[34,35], DIAGRAM Consortium[36], MuTHER Consortium[36], Ryoichi Takayanagi[37,47], Kyong Soo Park[12,38,47], Weiping Jia[4,47], Lee-Ming Chuang[11,39,47], Juliana C N Chan[13,47], Shiro Maeda[39,47], Takashi Kadowaki[10,47], Jong-Young Lee[1,47], Jer-Yuarn Wu[2,3,47], Yik Ying Teo[6,7,22,25,41,47], E Shyong Tai[22,42,43,47], Xiao Ou Shu[5,47], Karen L Mohlke[9,47], Norihiro Kato[8,47], Bok-Ghee Han[1,47] & Mark Seielstad[25,44,45,47]

We conducted a three-stage genetic study to identify susceptibility loci for type 2 diabetes (T2D) in east Asian populations. We followed our stage 1 meta-analysis of eight T2D genome-wide association studies (6,952 cases with T2D and 11,865 controls) with a stage 2 *in silico* replication analysis (5,843 cases and 4,574 controls) and a stage 3 *de novo* replication analysis (12,284 cases and 13,172 controls). The combined analysis identified eight new T2D loci reaching genome-wide significance, which mapped in or near *GLIS3*, *PEPD*, *FITM2-R3HDML-HNF4A*, *KCNK16*, *MAEA*, *GCC1-PAX4*, *PSMD6* and *ZFAND3*. *GLIS3*, which is involved in pancreatic beta cell development and insulin gene expression[1,2], is known for its association with fasting glucose levels[3,4]. The evidence of an association with T2D for *PEPD*[5] and *HNF4A*[6,7] has been shown in previous studies. *KCNK16* may regulate glucose-dependent insulin secretion in the pancreas. These findings, derived from an east Asian population, provide new perspectives on the etiology of T2D.

T2D is a major public health problem with a rapidly rising global prevalence[8]. The development of T2D is influenced by diverse factors, and decades of epidemiological studies have linked obesity, hypertension and dyslipidemia with the risk of T2D[9]. It is also known that T2D has considerable heritability. Within only the last 3 years, genetic studies have produced a rapidly lengthening list of loci harboring disease-predisposing variations[10]. To date, genetic variants at 45 loci have been identified for T2D[10,11]. Despite these advances toward a better understanding of the genetic basis of T2D, its heritability has not been fully explained[12]. In addition, most of the T2D loci were detected initially in population samples of European origin, with the exceptions of *KCNQ1*, *UBE2E2* and *C2CD4A-C2CD4B*, which were first identified in studies of east Asian populations[13–15]. Additional efforts involving east Asian populations identified variants associated with T2D at the *SPRY2*, *PTPRD* and *SRR* loci[5,16,17]. However, these associations need more validation from additional studies of east Asians as well as in studies of other populations. A large meta-analysis in east Asians would be expected to identify new genetic associations and provide insights into T2D pathogenesis. In addition to differences in the allele frequencies between east Asians and Europeans, which may affect the power to detect associations in these populations, T2D epidemiology also differs considerably between European populations and east Asian populations. In east Asians, the rates of diabetes are often higher at lower average body mass indices (BMIs)[18], suggesting that some different pathways may be involved in pathogenesis of T2D in east Asians and Europeans.

To discover new T2D loci, we conducted a three-stage association study in individuals of east Asian descent (**Supplementary Fig. 1**). We performed the stage 1 meta-analysis by combining eight T2D genome-wide association studies (GWAS) participating in the Asian Genetic Epidemiology Network (AGEN) consortium (6,952 cases and 11,865 controls) with association data for 2,626,356 imputed and genotyped autosomal SNPs, and we used the inverse-variance method

A full list of affiliations are at the end of article.

Received 12 April; accepted 2 November; published online 11 December 2011; doi:10.1038/ng.1019

for fixed effects for the statistical analyses (**Supplementary Table 1**). All imputed and genotyped SNPs (minor allele frequency (MAF) > 0.01) passed quality control filters in each of the eight stage 1 datasets prior to conducting the meta-analysis (**Supplementary Table 2**). The genomic control inflation factor (λ) for the meta-analysis was 1.046 (and was less than 1.062 for each of the individual studies), indicating that the results seen in stage 1 were probably not the result of population stratification (**Supplementary Fig. 2**). Individuals from each component study that participated in stage 1 mainly clustered together with the samples from the CHB/JPT HapMap population in the principal component analysis plot (**Supplementary Fig. 3**), further showing the similarity in ethnicity between the stage 1 samples. Most signals showing strong evidence for T2D associations were in known T2D genes (**Fig. 1**). Stage 1 *P* values, odds ratios (ORs) and average risk allele frequencies for 45 previously reported T2D-associated SNPs are listed in **Supplementary Table 3**.

After removing known T2D variants, we selected 297 SNPs from independent loci from the stage 1 meta-analysis based on our arbitrary inclusion criteria for *in silico* follow-up replication: meta-analysis $P < 5 \times 10^{-4}$ (based on the divergence between the observed and expected *P* values on the quantile-quantile plot; **Supplementary Fig. 2**), heterogeneity $P > 0.01$ and at least seven studies having been included in the meta-analysis (**Supplementary Table 4**). We took a total of 3,756 SNPs, including the 297 selected SNPs and their proxies ($r^2 > 0.8$ based on phase 2 CHB/JPT HapMap data), forward to stage 2 (*in silico* replication) in three independent GWAS (5,843 cases and 4,574 controls). After a meta-analysis that combined stage 1 and 2 data for 3,756 SNPs, we selected the 19 SNPs that showed the most compelling evidence for association (stage 1 and 2 combined $P < 10^{-5}$) (**Supplementary Table 5**) for stage 3 *de novo* genotyping in up to 12,284 cases and 13,172 controls recruited from five independent studies (**Supplementary Tables 1** and **2**). This resulted in eight new T2D loci that reached genome-wide significance in the combined meta-analysis across all three stages (**Table 1** and **Fig. 2**).

Three of these T2D-associated loci were previously associated with metabolic traits or related diseases or were suggestively associated with T2D. We detected one such locus within an intron of *GLIS3*, a gene that is highly expressed in islet beta cells. The coding product of this gene, a Krüppel-like zinc finger transcription factor, has been proposed as a key player in the regulation of pancreatic beta cell development and insulin gene expression[1,2]. SNPs in high linkage disequilibrium (LD) with this locus have been implicated in association with type 1 diabetes (T1D)[19] and fasting plasma glucose[3]. The second such locus, on 19q13, is located in an intron of *PEPD*. Several SNPs (lead SNP: rs10425678) in this gene were previously associated with T2D in a Japanese population[5]. However, the SNP in *PEPD* identified in our study (rs3786897) is not in LD with those identified in the Japanese population ($r^2 = 0.008$, $D' = 0.143$ between rs3786897 and rs10425678 based on phase 2 CHB/JPT HapMap data), and our GWAS data do not support an association for T2D with rs10425678 ($P = 0.528$). The third such signal is near *FITM2-R3HDML-HNF4A*. *FITM2* may be involved in lipid droplet accumulation[20], and the function of *R3HDML* is not known. Mutations in *HNF4A* cause maturity onset diabetes of the young type 1 (ref. 21). Common variants in the P2 promoter region of this gene (rs1884613 and rs2144908) have been associated with T2D in a population-specific manner[6,22]. The SNP near *FITM2-R3HDML-HNF4A* identified in our study (rs6017317) is not in strong LD with the *HNF4A* P2 promoter SNPs ($r^2 = 0.23$–0.25, $D' = 0.50$–0.54 between rs6017317 and rs1884613 or rs2144908 based on phase 2 CHB/JPT HapMap data), indicating that rs6017317 is a new T2D signal in the 20q13.12 region where *HNF4A* resides.

The other five loci reaching genome-wide significance in our study have not previously been reported in the context of any metabolic traits, including the loci mapped in or near *KCNK16*, *MAEA*, *GCC1-PAX4*, *PSMD6* and *ZFAND3*. *KCNK16*, which is expressed predominantly in the pancreas, encodes a potassium channel protein containing two pore-forming P domains[23]. In pancreatic β cells, potassium channels that are inhibited by ATP regulate glucose-dependent insulin secretion. Among the variants in strong LD with the signal reaching genome-wide significance in *KCNK16* (rs1535500) is rs11756091 ($r^2 = 0.977$, $D' = 1.0$ based on phase 2 CHB/JPT HapMap data), which encodes a substitution of proline to histidine in two isoforms of KCNK16. This variant or others influencing *KCNK16* may result in the defective regulation of potassium channel activity that contributes to the etiology of T2D[24]. *MAEA* encodes a protein that has a role in erythroblast enucleation and in the development of mature macrophages[25]. A gene-set analysis of the stage 1 *P* values using GSA-SNP[26] indicated that *MAEA* belongs to a group of genes that previously showed significant association with T2D and includes *IDE*, which is located at a known T2D susceptibility locus[27] (stage 1 $P = 1.41 \times 10^{-7}$ for rs6583826 at the *IDE* locus in this study). The GRIP-domain–containing protein that is encoded by *GCC1* might have a role in the organization of the trans-Golgi network, which is involved in membrane transport[28]. *PAX4*, which is only 30 kb away from *GCC1*, is a good candidate for T2D given its involvement in pancreatic islet development. *PAX4* was recently implicated in a Japanese individual with maturity onset diabetes of the young[29]. The expression product of *PSMD6*, which acts as a regulatory subunit of the 26S proteasome, is probably involved in the ATP-dependent degradation of ubiquitinated proteins[30]. Although the function of *ZFAND3* has not been fully elucidated, it is noteworthy that a member of the same gene family, *ZFAND6*, is present along with *FAH* at a previously detected T2D locus[31]. We examined whether eight new loci are potentially associated with T2D through an effect on obesity, as is the case with *FTO*[32]. All of the T2D association signals we initially detected remained after adjustment for BMI (**Supplementary Table 6**), indicating that the associations with T2D of these eight loci are not mediated through an effect on obesity.

In addition to the eight loci reaching genome-wide significance, we identified two loci showing moderate evidence (combined $P < 10^{-6}$) of association with T2D, including *WWOX* and *CMIP* loci (**Table 1**). We obtained the association results for these ten loci in GWAS data from up to 47,117 European samples generated by the DIAGRAM consortium (DIAGRAM+ is the current version of dataset)[31].

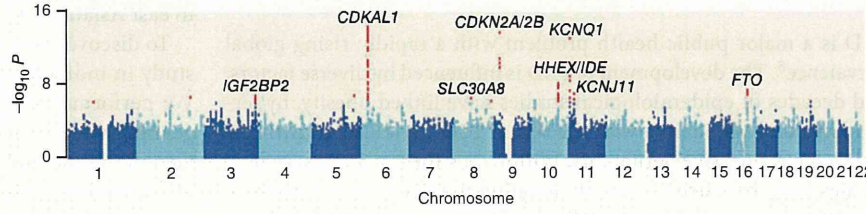

Figure 1 Genome-wide Manhattan plot for the east Asian T2D stage 1 meta-analysis. Shown are the $-\log_{10} P$ values using the trend test for SNPs distributed across the entire autosomal genome. The red dots at each locus indicate the signals with $P < 10^{-6}$ detected in the genome-wide meta-analysis. A total of 1,934,619 SNPs that were present in at least five stage 1 studies were used to generate the plot.

Table 1 Eight new T2D loci reaching genome-wide significance from a combined meta-analysis of stages 1, 2 and 3

SNP	Chr.	Position (bp)	Nearby gene	Risk allele	Other allele	Stage 1 (discovery)[a] OR (95% CI)	Stage 1 (discovery)[a] *P*	Stage 2 (*in silico* replication)[b] OR (95% CI)	Stage 2 (*in silico* replication)[b] *P*	Stage 3 (*de novo* replication)[c] OR (95% CI)	Stage 3 (*de novo* replication)[c] *P*	Combined (stages 1, 2 and 3)[d] OR (95% CI)	Combined (stages 1, 2 and 3)[d] *P*
Loci showing strong evidence of association with T2D													
rs6815464	4	1,299,901	*MAEA*	C	G	1.09 (1.04–1.14)	8.21×10^{-4}	1.13 (1.07–1.20)	3.67×10^{-5}	1.16 (1.11–1.20)	4.15×10^{-15}	1.13 (1.10–1.16)	1.57×10^{-20}
rs7041847	9	4,277,466	*GLIS3*	A	G	1.09 (1.04–1.14)	1.29×10^{-4}	1.09 (1.03–1.15)	2.20×10^{-3}	1.11 (1.07–1.15)	2.89×10^{-9}	1.10 (1.07–1.13)	1.99×10^{-14}
rs6017317	20	42,380,380	*FITM2-R3HDML-HNF4A*	G	T	1.10 (1.05–1.15)	2.43×10^{-5}	1.07 (0.99–1.15)	8.42×10^{-2}	1.10 (1.06–1.14)	3.96×10^{-7}	1.09 (1.07–1.12)	1.12×10^{-11}
rs6467136	7	126,952,194	*GCC1-PAX4*	G	A	1.12 (1.06–1.18)	6.47×10^{-5}	1.11 (1.04–1.18)	2.09×10^{-3}	1.10 (1.05–1.15)	2.31×10^{-5}	1.11 (1.07–1.14)	4.96×10^{-11}
rs831571	3	64,023,337	*PSMD6*	C	T	1.11 (1.06–1.17)	4.85×10^{-6}	1.06 (1.00–1.13)	4.46×10^{-2}	1.08 (1.05–1.12)	1.41×10^{-5}	1.09 (1.06–1.12)	8.41×10^{-11}
rs9470794	6	38,214,822	*ZFAND3*	C	T	1.11 (1.05–1.17)	1.45×10^{-4}	1.09 (1.02–1.17)	1.48×10^{-2}	1.16 (1.09–1.23)	3.20×10^{-6}	1.12 (1.08–1.16)	2.06×10^{-10}
rs3786897	19	38,584,848	*PEPD*	A	G	1.14 (1.08–1.20)	3.74×10^{-6}	1.05 (0.99–1.12)	1.28×10^{-1}	1.11 (1.04–1.17)	5.46×10^{-4}	1.10 (1.07–1.14)	1.30×10^{-8}
rs1535500	6	39,392,028	*KCNK16*	T	G	1.11 (1.06–1.16)	5.34×10^{-6}	1.07 (1.01–1.15)	3.33×10^{-2}	1.06 (1.02–1.10)	3.50×10^{-3}	1.08 (1.05–1.11)	2.30×10^{-8}
Loci showing moderate evidence of association with T2D													
rs16955379[e]	16	80,046,874	*CMIP*	C	T	1.13 (1.07–1.20)	2.20×10^{-5}	1.10 (1.03–1.17)	6.59×10^{-3}	1.05 (1.01–1.10)	2.19×10^{-2}	1.08 (1.05–1.12)	2.84×10^{-7}
rs17797882	16	77,964,419	*WWOX*	T	C	1.12 (1.05–1.18)	1.76×10^{-4}	1.09 (1.02–1.16)	1.21×10^{-2}	1.06 (1.01–1.11)	1.61×10^{-2}	1.08 (1.05–1.12)	9.49×10^{-7}

[a]Up to 6,952 cases and 11,865 controls. [b]Up to 5,843 cases and 4,574 controls. [c]Up to 12,284 cases and 13,172 controls. [d]Up to 25,079 cases and 29,611 controls. [e]The proxy SNP rs9930117 ($r^2 = 1$) was genotyped in the stage 3 CAGE study.

The DIAGRAM-generated results for these loci indicated that three loci, including the *FITM2-R3HDML-HNF4A* (rs6017317: $P = 1.47 \times 10^{-2}$, OR = 1.07), *CMIP* (rs16955379: $P = 3.33 \times 10^{-2}$, OR = 1.20) and *MAEA* (using the proxy SNP for rs6815464, rs11247991 ($r^2 = 0.96$): $P = 6.56 \times 10^{-3}$, OR = 1.19) loci, were modestly associated with T2D, whereas a locus in *GLIS3* (rs7041847: $P = 6.43 \times 10^{-2}$, OR = 1.04) was nominally associated with T2D. The direction of effect was consistent in four (*PSMD6*, *PEPD*, *WWOX* and *KCNK16*) of the six loci that were not replicated in DIAGRAM+ (**Supplementary Table 5**).

We analyzed the functional connections among the 10 new T2D genes and the 28 known T2D genes that we replicated in this study (**Supplementary Table 3**) using GRAIL[33]. The connection results highlighted notable biological functions for sets of genes within T2D-associated regions (**Supplementary Fig. 4** and **Supplementary Tables 7** and **8**). For example, *KCNK16* has strong connections with previously known T2D genes encoding potassium channels (*KCNJ11* and *KCNQ1*), implying that is has a physiological role in the regulation of potassium transport in pancreatic cells.

We examined the association between each new T2D SNP and the expression of genes within 1 Mb of these SNPs by an expression quantitative trait locus (eQTL) analysis using the data from the MuTHER consortium. One SNP (rs3786897) in an intron of *PEPD* was highly associated with the mRNA expression of *PEPD* in the adipose tissue of 776 individuals of European ancestry ($P_{eQTL} = 2.14 \times 10^{-8}$) (**Supplementary Table 9**). However, this SNP did not show an association with T2D in populations of European ancestry, thus the importance of this finding is unclear.

We considered the possibility that autoimmune diabetes (rather than T2D) may be driving some of the signals that we observed. First, the cases from all the studies we examined predominantly had adult-onset diabetes (age of disease onset ≥30 years), and none of the clinically diagnosed subjects had T1D, which is defined by the presence of acute ketosis and the continuous requirement of insulin beginning within 1 year after diagnosis. Second, we researched the associations for all known T1D-associated variants in our dataset. Only a small number of loci showed association after this analysis (**Supplementary Table 10**). These results are in distinct contrast to those for known T2D-associated variants, many of which replicated in our study (**Supplementary Table 3**), further suggesting that our findings are most relevant to T2D. Third, as variants close to the *GLIS3* locus have been shown to be associated with T1D[19], we examined the association between rs7041847 and diabetes in four studies ($n = 8,383$) in which individulas with positive glutamic acid decarboxylase (GAD) antibodies had been excluded (as individuals with T1D are positive for GAD antibodies, whereas individuals with T2D are not) (data not shown). In each study, the associations between this SNP and diabetes were the same as the association found when we included all the samples (meta-analysis $P = 3.4 \times 10^{-4}$, OR = 1.12). This finding, along with the fact that SNPs at the *GLIS3* locus also show associations with fasting plasma glucose in nondiabetic adults[3] and in healthy children and adolescents[4], is consistent with the hypothesis that SNPs at this locus may affect fasting glucose homeostasis rather than the immune system. Taken together, it is unlikely that a substantial proportion of the positive associations observed in our study were driven by autoimmune diabetes.

This study is the largest GWAS meta-analysis, to our knowledge, conducted for T2D in east Asians. Findings from this study highlight not only previously unknown biological pathways but also population-specific loci for T2D. The association of rs9470794 in *ZFAND3* with T2D seems to be highly specific to east Asian populations (**Supplementary Table 5**), whereas the association of

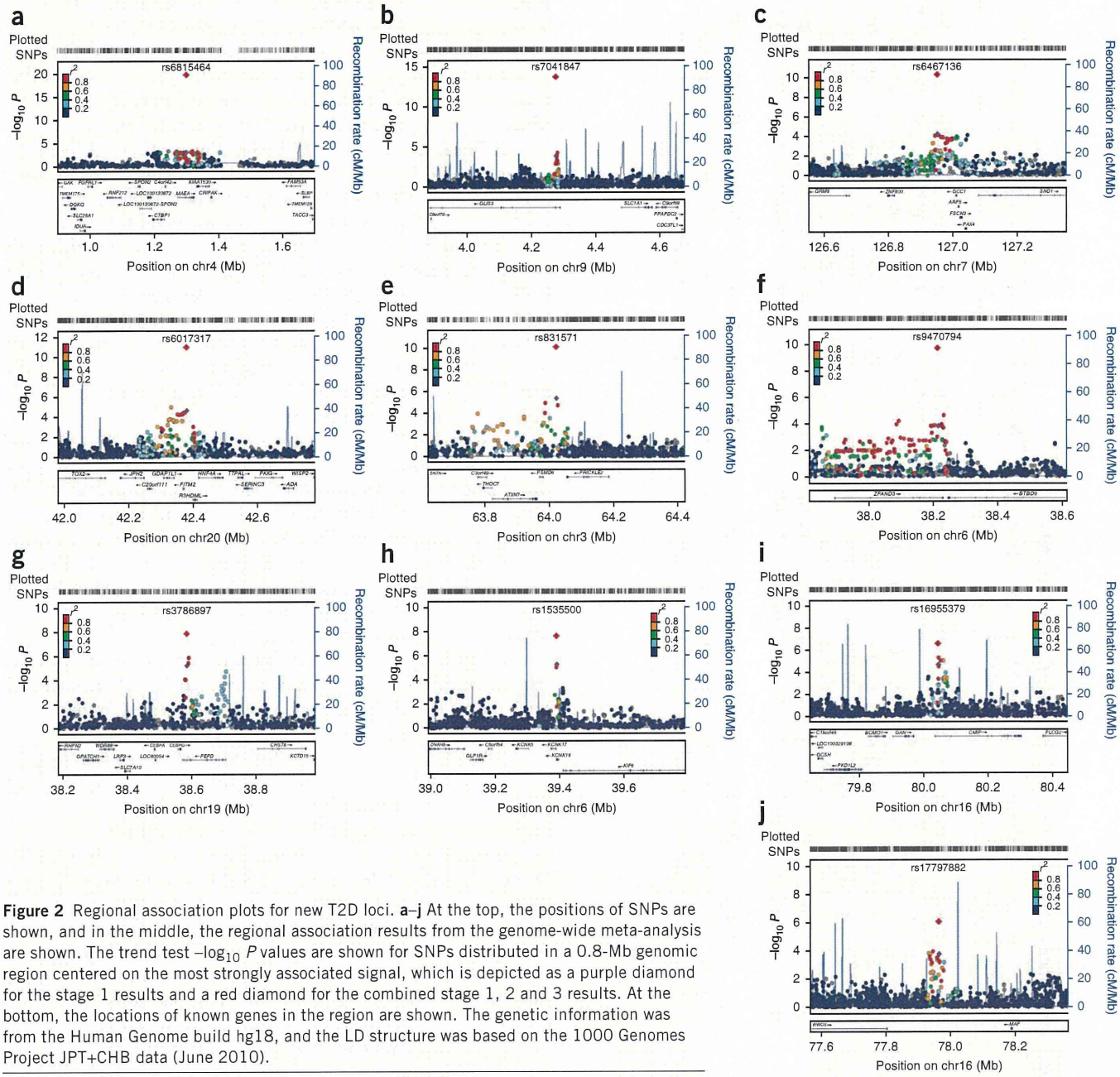

Figure 2 Regional association plots for new T2D loci. **a–j** At the top, the positions of SNPs are shown, and in the middle, the regional association results from the genome-wide meta-analysis are shown. The trend test $-\log_{10}$ *P* values are shown for SNPs distributed in a 0.8-Mb genomic region centered on the most strongly associated signal, which is depicted as a purple diamond for the stage 1 results and a red diamond for the combined stage 1, 2 and 3 results. At the bottom, the locations of known genes in the region are shown. The genetic information was from the Human Genome build hg18, and the LD structure was based on the 1000 Genomes Project JPT+CHB data (June 2010).

rs11634397 near *ZFAND6* seems to be specific to European populations (**Supplementary Table 3**). We observed a substantial difference in the risk allele frequencies of both loci between the two continental (Asian and European) populations (rs9470794: risk allele frequency (RAF) = 0.32 for the Asian CHB/JPT HapMap population compared to RAF = 0.12 for the European CEU HapMap population; rs11634397: RAF = 0.07 for CHB/JPT compared to RAF = 0.64 for CEU). Although these loci are related to T2D differently in the two populations (the *ZFAND3* locus is specific to Asians, whereas the *ZFAND6* locus to Europeans), these results lead to speculation that the broader A20 domain-containing zinc finger protein family has a role in the etiology of T2D. Additional population-specific T2D loci were also suggested by our analysis, for example, *WWOX* (rs17797882) (**Supplementary Table 5**) in east Asians and *ZBED3* (rs4457053) (**Supplementary Table 3**) in Europeans. Despite the lack of clear physiological evidence on T2D pathogenesis, these findings may provide clues to help understand T2D phenotypes characteristic of each population, for example, the high rates of diabetes seen at lower average BMIs in east Asians.

URLs. IMPUTE, http://mathgen.stats.ox.ac.uk/impute/impute.html; MACH, http://www.sph.umich.edu/csg/abecasis/MACH/; BEAGLE, http://faculty.washington.edu/browning/beagle/beagle.html; METAL, http://www.sph.umich.edu/csg/abecasis/Metal; WGAViewer, http://compute1.lsrc.duke.edu/softwares/WGAViewer/; SNAP, http://www.broadinstitute.org/mpg/snap/; **LocusZoom**, http://csg.sph.umich.edu/locuszoom/; GenABEL, http://www.genabel.org/; ProbABEL, http://www.genabel.org/packages/ProbABEL.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.