Available online at www.sciencedirect.com

SCIENCE DIRECT

Genomics 85 (2005) 591–599

GENOMICS

www.elsevier.com/locate/ygeno

Identification of candidate genes in the type 2 diabetes modifier locus using expression QTL

Hiroshi Yaguchi[a,b], Katsuhiko Togawa[a,b], Maki Moritani[a], Mitsuo Itakura[a,*]

[a]*Division of Genetic Information, Institute for Genome Research, University of Tokushima, Tokushima 770-8503, Japan*
[b]*Division of Pharmacology, Drug Safety, and Metabolism, Otsuka Pharmaceutical Factory, Inc., Tokushima 772-8601, Japan*

Received 10 December 2004; accepted 20 January 2005

Abstract

To identify new genetic determinants relevant to type 2 diabetes (T2D), diabetic F2 progeny were generated by intercrossing F1 mice obtained from a cross of BKS.Cg-*Lepr*^{db}+/+*m* and DBA/2, and T2D-related phenotypes were measured. In the F2 population, increased susceptibility to diabetes and obesity was observed. We also detected the major quantitative trait loci (QTL) modifying the severity of diabetes on chromosome 9, where peaks of logarithm of odds (LOD) overlapped for three traits. To identify candidate genes in the QTL intervals, we combined "expression QTL" (eQTL), taking mRNA levels as quantitative traits, and "interstrain sequence variations, including cSNPs." As a result, four genes were identified from cosegregation of clinical QTL with eQTL and 13 genes were found from interstrain cSNPs as candidates in the T2D modifier QTL. Our combined approach shows the acceleration of the discovery of candidate genes in the QTL of interest, spanning several megabases.

Keywords: Type 2 diabetes; QTL analysis; Expression QTL; Candidate genes

Quantitative trait loci (QTL) mapping in mice has revealed hundreds of chromosomal regions containing genes affecting polygenic traits. Once QTL are identified, the selection of candidate genes can usually begin. The process of selecting candidate genes relies on a wealth of information gained through molecular approaches and several genomic resources, i.e., the mouse consensus genome sequence, interstrain haplotype information, interstrain polymorphic information, and human–mouse orthologs. Despite the useful bioinformatic tools, it remains difficult to identify suitable candidate genes in the QTL region and few complex-trait genes have been discovered so far [1]. Novel methods to identify the candidate genes in the QTL region need to be developed.

Schadt et al. have performed a genome-wide linkage analysis of gene expression and identified genetic regions (loci) that can account for variation in the levels of gene expression [2]. They term these loci "eQTL" for expression quantitative trait loci against "cQTL" for the classical or clinical trait loci such as height, weight, and blood glucose (BG) concentrations.

In this study, we present two approaches to identify candidate genes responsible for the effect of cQTL: one is based on eQTL analysis and the second on the interstrain coding single-nucleotide polymorphisms (cSNPs). By combining these two approaches, we effectively identified candidate genes in the type 2 diabetes (T2D) modifier locus using F2 progeny by intercrossing F1 mice obtained from a cross of BKS.Cg-*Lepr*^{db}+/+*m* (BKS-*db*/+*m*) and DBA/2 (D2) mice. Seventeen candidates were found from a total of 106 genes in the target QTL region of approximately 10 Mb.

Results

Characterization of D2BKS F2-db/db mice

The diabetes (*db*) mutation is a result of a point mutation in the leptin receptor gene, *Lepr*. The ligand, leptin, is a key

* Corresponding author. Fax: +81 88 633 9455.
E-mail address: itakura@genome.tokushima-u.ac.jp (M. Itakura).

0888-7543/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2005.01.006

Table 1
Values for phenotypes for parental strains and F2 mice

	Mouse strain				
	DBA	BKS	BKS-*db*/+*m*	BKS-*db*/*db*	F2-*db*/*db*
Sex	Female	Female	Female	Female	Female
Number of mice	15	20	12	18	113
Blood glucose at 0 min in ipGTT (mM)	4.2 ± 0.2	3.8 ± 0.2	3.9 ± 0.2	10.5 ± 0.8	18.2 ± 1.0
Blood glucose at 30 min in ipGTT (mM)	10.7 ± 0.5	17.0 ± 0.8	13.0 ± 0.7	29.2 ± 1.6	39.7 ± 0.6
Blood glucose at 60 min in ipGTT (mM)	11.4 ± 0.4	10.2 ± 0.5	11.9 ± 0.8	31.7 ± 1.7	39.9 ± 0.6
Blood glucose at 120 min in ipGTT (mM)	7.8 ± 0.3	5.3 ± 0.2	7.3 ± 0.4	32.6 ± 2.3	39.2 ± 0.7
Body weight at 9 weeks of age (g)	18.6 ± 0.3	18.2 ± 0.2	19.2 ± 0.3	31.6 ± 0.5	36.7 ± 0.4
Parametrial fat pad weight (mg)	350.8 ± 26.5	87.1 ± 6.6	377.9 ± 39.0	1874.1 ± 66.6	2831.9 ± 56.1

Data are shown as means ± SE.

weight control hormone and mice homozygous for the *db* mutation become remarkably obese and hyperglycemic [3]. Compared to the two parental strains and the heterozygotes for the *db* mutation (BKS-*db*/+*m*), homozygotes for the *db* mutation (BKS-*db*/*db* and F2-*db*/*db*) exhibited higher body weights, BG concentration in ipGTT, and parametrial fat pad weight (Table 1 and Fig. 1).

Also, there were marked mean differences between BKS-*db*/*db* and F2-*db*/*db* for all traits examined. Body weight and BG concentration (0 min in ipGTT) in F2-*db*/*db* were higher than those in BKS-*db*/*db* ($p < 0.01$, 36.7 ± 0.4 versus 31.6 ± 0.5 g, and $p < 0.05$, 18.2 ± 1.0 versus 10.5 ± 0.8 mM, respectively). This tendency was maintained after glucose injection in ipGTT. Parametrial fat pad weight in F2-*db*/*db* was significantly higher than that in BKS-*db*/*db* ($p < 0.01$, 2831.9 ± 56.1 versus 1874.1 ± 66.6 mg). Fig. 1 illustrates the distribution of body weight and BG concentration (0 min in ipGTT). In F2-*db*/*db*, body weight was distributed over a range of 25–48 g in a unimodal fashion. For BG concentrations, F2-*db*/*db* exhibited a much broader bimodal distribution than BKS-*db*/*db*.

cQTL analysis and confidence QTL region

Characterization of D2BKS F2-*db*/*db* mice strongly indicated that the introgression of the D2 background genetically predisposed the F2 population to diabetes and obesity. To identify the chromosomal loci of modifiers determining increased susceptibility to diabetes and obesity, a genome-wide scan was performed to map QTL for six clinical traits, including BG concentrations in ipGTT at 0, 30, 60, and 120 min; body weight; and parametrial fat pad weight. In this analysis, a cQTL was found on chromosome 9 (Chr. 9) with high lod scores for the three traits ipGTT at 120 min, body weight, and parametrial fat pad weight (Fig. 1A). The maximum lod score and genetic variance were 5.67 and 21% (BG concentrations at 120 min in ipGTT), 3.35 and 13% (body weight), and 4.22 and 16% (parametrial fat pad weight), respectively. No significant linkage was observed for BG concentrations at 0 min (LOD 1.97) and 30 min (LOD 0.87) in ipGTT, and weak linkage was found for 60 min (LOD 2.65). The effect on the latter part of ipGTT may suggest this cQTL is not responsible for the glucose-induced initial insulin response but is responsible for insulin resistance in muscle, adipocytes, and liver.

Bootstrap analysis estimated the confidence QTL regions for ipGTT at 120 min (top in Fig. 2B), body weight (middle in Fig. 2B), and parametrial fat pad weight (bottom in Fig. 2B). The highest single peak was not consistently observed at the same location for the three traits, i.e., the high peaks for body weight and parametrial fat pad weight were not reproducible by the peak for BG concentration at 120 min in

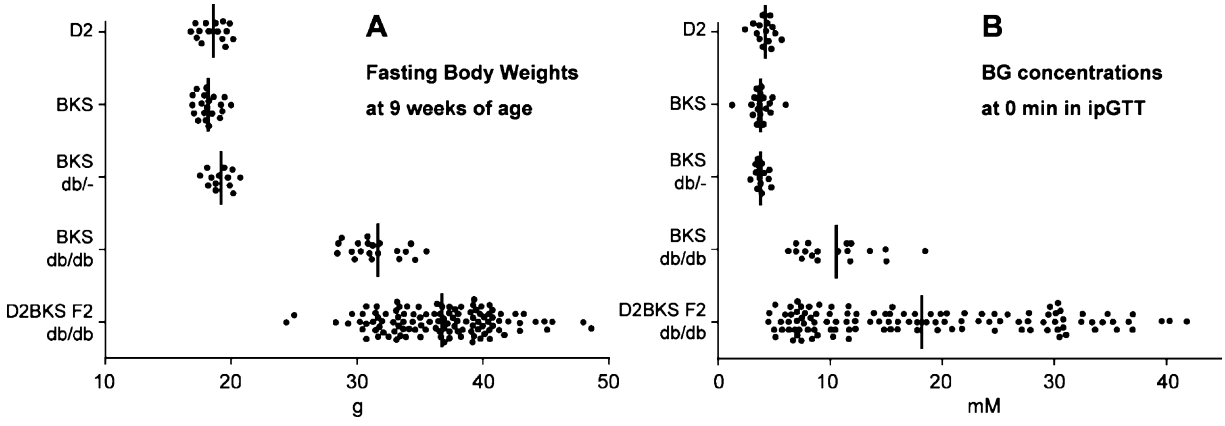

Fig. 1. Scatter plots of (A) fasting body weights at 9 weeks of age and (B) BG concentrations at 0 min in ipGTT. Plots display measurements from D2 ($n = 15$), BKS ($n = 20$), BKS-*db*/− ($n = 12$), BKS-*db*/*db* ($n = 18$), and F2-*db*/*db* ($n = 113$). Bold lines show the mean value.

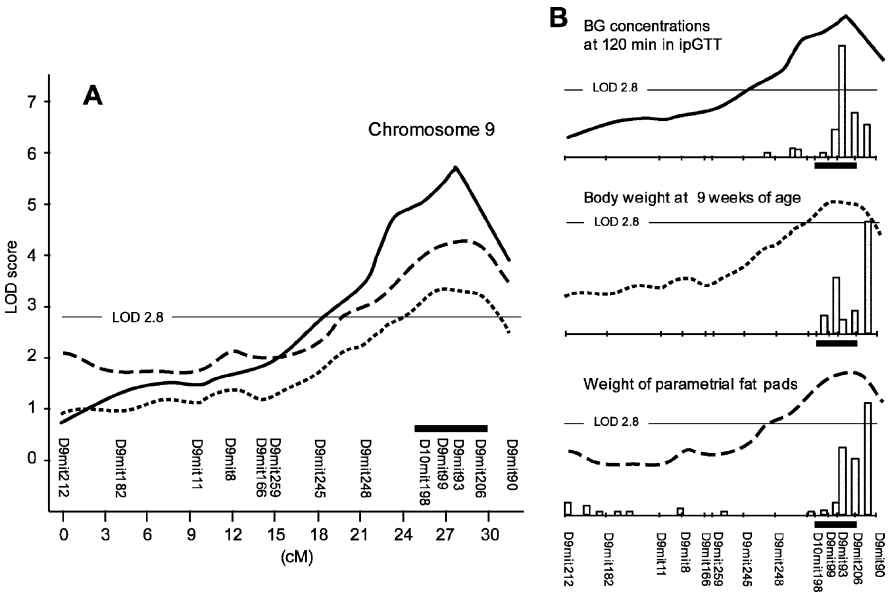

Fig. 2. (A) Lod score curve of three overlapping cQTLs on Chr. 9 and (B) bootstrap analysis. (A) Lod score curves for BG concentrations in ipGTT at 120 min (solid line), body weight at 9 weeks (dashed line), and parametrial fat pad weight (broken line) are shown. The horizontal line indicates the threshold value for a suggestive level of LOD at 2.8. The genetic markers used for typing are illustrated on their genetic location. A short bold line shows the target region between *D10mit198* and *D9mit206*. (B) Results of bootstrap analysis for BG concentrations in ipGTT at 120 min (top), body weight at 9 weeks of age (middle), and parametrial fat pad weight (bottom) are shown.

ipGTT. In contrast, the single peak for BG concentrations at 120 min in ipGTT was associated with the second highest peaks for body weight and parametrial fat pad weight. Based on the shared peaks in the lod score curve in Fig. 2A and bootstrap results in Fig. 2B, we assumed that the interval between *D10mit198* and *D9mit206* (~10 Mb physical distance) is the most likely cQTL target region.

Effects of the D9mit93 genotype on phenotypes

Fig. 3 shows the characteristics of the 113 female F2-*db*/*db* mice separated by their genotypes at *D9mit93*. In 113 F2-*db*/*db* mice, 35 were homozygous for the D2 allele (DD), 62 were heterozygous (DB), and 16 were homozygous for the BKS allele (BB).

DD mice had higher BG concentrations at 120 min in ipGTT than BB mice (Fig. 3A). In contrast, the body weight and parametrial fat pad weight in DD mice were lower than those in BB mice (Figs. 3B and 3C). In heterozygous DB mice, body weight and parametrial fat pad weight were intermediate between those of the two homozygotes, and BG concentrations at 120 min in ipGTT were close to those of DD mice. Given that there were marked differences in the effects of the *D9mit93* genotype, we considered that *D9mit93* was a cQTL predisposing for hyperglycemia and affecting body weight and parametrial fat pad weight.

We deduce that the decreased body weight and fat pad weight are secondary to the higher prevalence of diabetes in DD mice and not to the obesity gene contributed from D2 mice. Insulin resistance is associated with an increased flux of free fatty acids (FFA) from intra-abdominal fat pads to the liver because of increased hormone (insulin) sensitive lipase

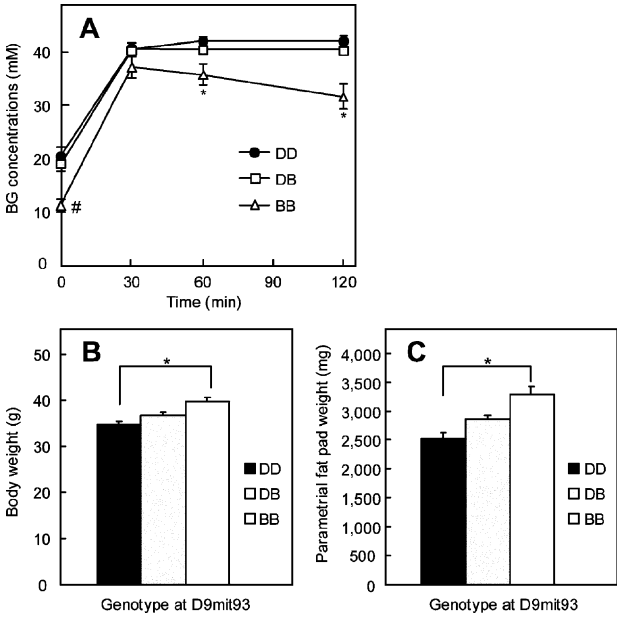

Fig. 3. (A) BG concentrations in ipGTT, (B) body weight, and (C) parametrial fat pad weight in female F2-*db*/*db* mice according to the *D9mit93* genotype. (A) BG concentrations in ipGTT homozygous for the D2 allele (DD, $n = 35$) and heterozygous (DB, $n = 62$) and homozygous mice for the BKS allele (BB, $n = 16$) are shown. (B and C) Black columns denote homozygous mice for the D2 allele ($n = 35$). Gray columns indicate heterozygous mice ($n = 62$). White columns show homozygous mice for the BKS allele ($n = 16$). Data in A, B, and C are the means ± SE. *Significant difference at $p < 0.01$ between the BB genotype and the DD genotype. #Significant difference at $p < 0.05$ between the BB genotype and the DD genotype.

activity [4]. The increased delivery of FFA to the liver leads to decreased fat weight and increased hepatic very low density lipoprotein–triglyceride secretion. In our F2-*db*/*db* mice, the DD genotype exhibited higher serum triglyceride concentrations (159.91 ± 11.9 mg/dl, $n = 35$) than BB mice (125.81 ± 15.3 mg/dl, $n = 16$). This supports our interpretation that the effect of *D9mit93* on body weight and parametrial fat pad weight is secondary to diabetes.

Identification of causative candidates by the eQTL analysis

To detect candidate genes responsible for the effects of the cQTL at *D9mit93*, we selected 76 genes in the target region (*D10mit198–D9mit206*) using public and Celera databases. First, we measured the mRNA levels in liver, parametrial fat pads, and soleus of BKS-*db*/*db* and D2 mice. Table 2 shows 15 genes encoded in the target region, which were differentially expressed between BKS-*db*/*db* and D2 mice. Their transcriptional levels were not rare in relation to those of the reference genes, e.g., glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) or β-actin, except for *ApoA1*. In comparing BKS-*db*/*db* and D2 mice, we were looking at both the effects of the interstrain differences (i.e., BKS versus D2) and the *db* mutation. To assess the effects of only the interstrain differences, we compared mRNA levels of these 15 genes between BKS-*db*/+*m* and D2 mice, assuming *db* was fully recessive. Thirteen genes were found, excluding *9030425E11Rik* (GenBank Accession No. BC026447) and *Hyou1* (GenBank Accession No. BC050107), as differentially expressed between BKS-*db*/+*m* and D2 mice.

In addition, we found that *Tagln* (GenBank Accession No. BC003795) and *Nnmt* (GenBank Accession No. AH006989) had no change in their transcriptional levels by the *db* mutation and that *9030425E11Rik*, *Abcg4* (GenBank Accession No. BC016200), *ApoA1* (GenBank Accession No. BC012253), and *Cryab* (GenBank Accession No. M73741) were up-regulated, while other genes were down-regulated, comparing the expression ratios of BKS-*db*/+*m* vs D2 to BKS-*db*/*db* vs D2 in Table 2.

Classically, QTL analysis found genes responsible for complex traits such as body weight or diabetes. Similarly, when the expression levels of genes are defined as a quantitative trait, QTL analysis can map the genetic determinants responsible for their transcriptional levels (eQTL analysis). Often, the eQTL are mapped to the gene itself, indicating that *cis* changes are responsible for the different expression levels. However, the presence of many genes that are coordinately regulated by a single unlinked QTL suggests that *trans*-acting factors are controlling expression [2]. If the cQTL were caused by the interstrain differences of gene expression, the genetic determinants responsible for the cQTL would be restricted to the genes that were encoded inside the cQTL region and provide variations of gene expression under *cis*-acting transcriptional fashion in the F2 population. In this case, their eQTL were found to reside at the same chromosomal positions at which they were encoded and the lod score curves with the peak of eQTL should coincide with those of the cQTL.

We examined mRNA levels in the F2-*db*/*db* population for 15 differentially expressed genes (Table 2) and carried them through an eQTL analysis. *9030425E11Rik* and

Table 2
Genes encoded in the target region that were differentially expressed between BKS-*db*/*db* and D2 mice

Symbol	Tissue measured for gene expression	Expression ratio (between original strains)		eQTL results (in F2 *db*/*db*)	
		BKS-*db*/+*m* vs D2	BKS-*db*/*db* vs D2	*cis*/*trans*	Lod score in target region
9030425E11Rik	Liver	1.07 (1.86/1.74)	2.99 (5.22/1.74)	*trans*	0.48
Abcg4	FAT	0.39 (0.35/0.89)	0.47 (0.42/0.89)	*trans*	1.04
Hyou1	Liver	1.05 (0.59/0.57)	0.30 (0.17/0.57)	*trans*	0.18
Eva	FAT	0.08 (0.31/3.96)	0.03 (0.12/3.96)	*cis*/*trans*	2.50
Tagln	Soleus	0.48 (0.96/2.01)	0.48 (0.96/2.01)	*trans*	0.90
BC033915	FAT	1.30 (1.16/0.89)	0.51 (0.46/0.89)	*trans*	0.26
ApoA1	Liver	0.66 (0.87/1.32)	0.21 (0.28/1.32)	*trans*	0.76
ApoA4	Liver	0.15 (0.47/3.05)	0.10 (0.31/3.05)	*trans*	0.90
ApoA5	Liver	1.36 (0.07/0.05)	5.18 (0.28/0.05)	*trans*	0.96
D030060M11Rik	Liver	0.54 (1.51/2.78)	0.44 (1.23/2.78)	*trans*	1.36
C130036J11	FAT	8.59 (ND)[a]	4.30 (ND)[a]	*cis*	7.29
	Soleus	18.39 (ND)[a]	8.76 (ND)[a]	*cis*	35.21
4432416J03Rik	Liver	0.54 (0.91/1.68)	0.33 (0.56/1.68)	*cis*	4.72
	Soleus	1.25 (0.54/0.43)	0.61 (0.26/0.43)	*cis*	7.14
Nnmt	FAT	3.17 (1.94/0.61)	3.25 (1.99/0.61)	*cis*	4.46
2310030G06Rik	FAT	0.56 (3.91/6.96)	0.45 (3.14/6.96)	*cis*/*trans*	2.12
Cryab	FAT	2.00 (1.39/0.69)	5.09 (3.53/0.69)	*cis*	12.26

FAT stands for parametrial fat pads. The ratios of gene expression between BKS-*db*/+*m* and D2 and also between BKS-*db*/*db* and D2 are shown in the Expression ratio column. The measurement values of the gene of interest in BKS-*db*/+*m* and DBA and in BKS-*db*/*db* and DBA are shown in parentheses. All measurement values represent arbitrary units normalized to the reference gene β-actin or *GAPDH*. In the eQTL column, the classification of transcription (between *cis* and *trans*) and the peak lod scores in the target region are shown. The four underlined genes represent those with peak eQTL lod scores greater than 2.8 in at least one examined tissue.

[a] The comparative C_t method was used to assess relative changes in mRNA levels of *C130036J11*.

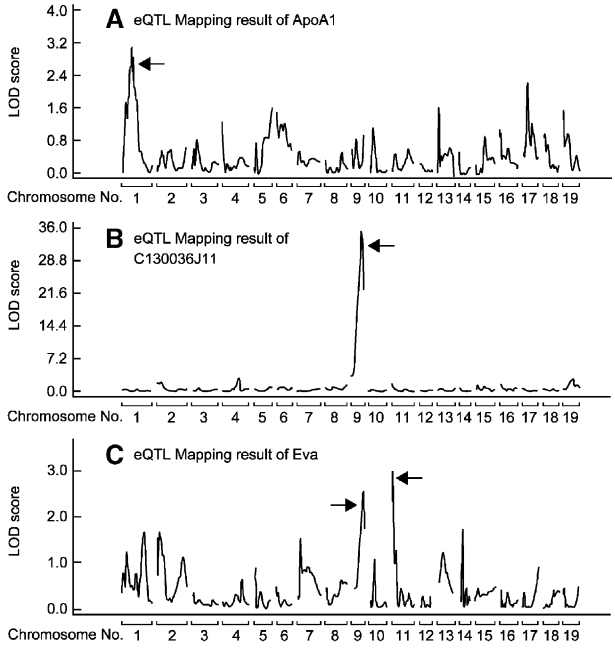

Fig. 4. Genome-wide eQTL mapping results for the expression levels of (A) *ApoA1*, (B) *C130036J11*, and (C) *Eva*, summarized by chromosome. The arrows indicate chromosomal regions that vary in the level of gene expression. (A) Example of *trans*-acting transcription. (B) Example of *cis*-acting transcription. (C) Example of both *cis*- and *trans*-acting transcription.

Hyou1 were included as the subjects of eQTL analysis despite no difference in their expression levels between BKS-*db*/+*m* and D2, because the cQTL analysis was done for the F2-*db*/*db* population, which had effects of both the interstrain differences and the *db* mutation.

As a result of eQTL analysis, we could distinguish 15 differentially expressed genes by their transcriptional behavior. We defined *cis*-acting transcription as eQTL that corresponded to genes with physical locations within 5 Mb of the eQTL peak, and all other eQTL were considered as *trans*-acting transcription. By these definitions of *cis* and *trans*, the following distribution of 15 differentially expressed genes was found: 4 genes had only *cis*-acting transcription, 9 genes had only *trans*-acting transcription, and 2 genes had both *cis*- and *trans*-acting transcription (Table 2). Examples of the eQTL results of three genes, *ApoA1* (*trans*-acting transcription), *C130036J11* (GenBank Accession No. AK048132) (*cis*-acting transcription), and *Eva* (GenBank Accession No. AF030454) (*cis*- and *trans*-acting transcription) are shown in Fig. 4. The trend observed here was that eQTL with high lod scores were *cis*-acting, whereas moderately significant QTL were *trans*-acting in most cases.

We found the eQTL of four genes, *C130036J11*, *4432416J03Rik* (GenBank Accession No. AK019469), *Nnmt*, and *Cryab*, were comapped with the cQTL at *D9mit93*. *4432416J03Rik* in liver showed a peak lod score of 4.72 (Fig. 5A, dashed line) and in soleus 7.14 (Fig. 5A, solid line). *C130036J11* showed a peak lod score of 7.29 in parametrial fat pads (Fig. 5B, dashed line) and 35.21 in the soleus (Fig. 5B, solid line). *Nnmt* in parametrial fat pads showed the peak lod score of 4.46 (Fig. 5C). *Cryab* showed the peak lod score of 12.26 in parametrial fat pads (Fig. 5D). Bootstrap analysis also supported that these eQTL were colocalized in the target region spanning *D10mit198* and *D9mit206* (Figs. 5A–5D). These results indicated that these four genes were apparently under *cis*-acting transcriptional regulation in the target region. We considered that they

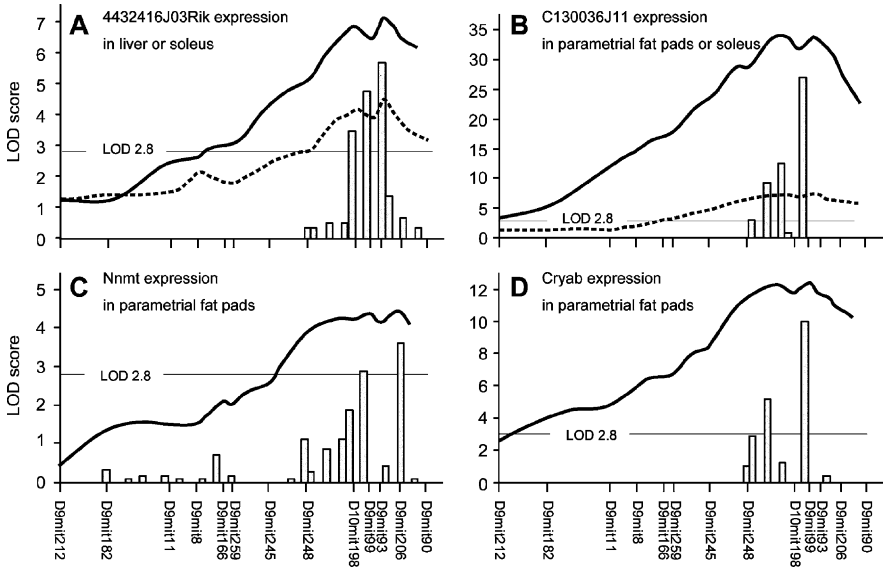

Fig. 5. Lod score curves for (A) *4432416J03Rik*, (B) *C130036J11*, (C) *Nnmt*, and (D) *Cryab*, which were comapped with the cQTL at *D9mit93*. (A) *4432416J03Rik* in the liver showed a peak lod score of 4.72 (dashed line) and in the soleus 7.14 (solid line). (B) *C130036J11* showed a peak lod score of 7.29 in parametrial fat pads (dashed line) and 35.21 in the soleus (solid line). (C) *Nnmt* in parametrial fat pads showed the peak lod score of 4.46. (D) *Cryab* showed the peak lod score of 12.26 in parametrial fat pads. Gray columns represent the results of bootstrap analysis.

were causative candidates for the effects of the cQTL at *D9mit93*.

Identification of causative candidates by the interstrain cSNPs

As well as the differences in gene expression, the amino acid substitutions are also important in providing functional variation. Genes that have cSNPs accompanied by amino acid variations afford a group of candidate genes in the target region. We searched the Celera RefSNP database based on the interstrain sequence differences and listed 14 polymorphic genes between C57BL/6 (B6) and D2: 13 genes contained the missense cSNPs and 1 gene had a SNP in a splice acceptor site (Table 3). The BKS strain is a mixture of B6 and D2 genetic background, respectively, at 84 and 16% [5]; therefore, SNPs between B6 and D2 include all SNPs between BKS and D2. These cSNPs were confirmed to be polymorphic between BKS and D2 by sequencing the BKS genome directly. We did not sequence F2-*db*/*db* mice because the microsatellite genotyping results provided us the information of these cSNPs in the F2-*db*/*db* population.

From sequence analysis, *Hyou1* contains no polymorphisms in the splice acceptor site between BKS and D2 (Table 3). The absence of polymorphisms in *Hyou1* may be caused by genetic contamination of the D2 genome into a B6 background. The rest of the candidates were really polymorphic and showed amino acid variations. We found that *4432416J03Rik*, near *D9mit93*, contained the largest number of amino acid variations (5). In these 13 candidates, no significant difference was observed in the expression levels between BKS-*db*/+*m* and D2 or between BKS-*db*/*db* and D2 mice. Therefore, it may be possible that these sequence variations affect enzymatic activities or gene interactions and ultimately lead to phenotypic variations.

Discussion

The severity of diabetes caused by the *db* mutation in mice is markedly strain dependent. When the *db* mutation is maintained in the BKS and D2 strains, diabetes is severe, whereas diabetes is mild in the B6 genetic background [6]. We also presented that the D2 genetic background carries genetic factors as modifiers to aggravate *db* mutation-related phenotypes, such as obesity and diabetes, in Table 1. This report was only about the target region between *D10mit198* and *D9mit206* on Chr. 9, where cQTL of BG concentrations in ipGTT, body weight, and parametrial fat pad weight cosegregated. We found many other cQTL on other

Table 3
Genes with missense cSNPs and one gene without a splice acceptor site polymorphism between BKS and DBA in the target region

Symbol	Celera gene ID	Celera missense SNP ID	Protein variation (B6 vs DBA)	Sequencing results		
				B6	BKS	D2
D9Mit206						
Bsx	mCG51682	mCV24919748	Glu(GAA)13Gly(GGA)	G	G	A
		mCV24919749	Met(ATG)19Leu(TTG)	A	A	T
		mCV24919760	Asn(AAT)131Tyr(TAT)	A	A	T
Crtam	mCG5069	mCV24022091	Gln(CAG)360Arg(CGG)	G	G	A
Tecta (LOC270149)	mCG10666	mCV24933599	Ile(ATC)2148Val(GTC)	A	A	G
		mCV24313489	Thr(ACT)1032Pro(CCT)	C	C	A
		mCV24314178	Asp(GAC)559Gly(GGC)	A	A	G
Pvrl1	mCG142054	mCV22858958	Ser(AGC)428Gly(GGC)	G	G	A
Hyou1[a]	mCG10711	mCV24941980	Splice acceptor site	A	G	G
Mll	mCG1547	mCV24950707	Ser(TCA)2030Leu(TTA)	C	C	T
		mCV24951774	Pro(CCG)561Leu(CTG)	T	T	C
		mCV24951785	Glu(GAA)546Asp(GAT)	A	A	T
Ube4a	mCG50277	mCV24043625	Thr(ACG)518Met(ATG)	C	C	T
Fxyd2	mCG4536	mCV25163361	Gln(CAA)80Arg(CGA)	A	A	G
4930435C18Rik	mCG141917	mCV23102204	Ala(GCG)75Val(GTG)	C	C	T
		mCV25167864	Ser(AGT)61Ile(ATT)	T	T	G
4432416J03Rik	mCG9661	mCV25187606	Tyr(TAT)384Phe(TTT)	T	T	A
		mCV25187920	Ile(ATC)176Val(GTC)	A	A	G
		mCV25187951	Thr(ACT)162Ala(GCT)	G	G	A
		mCV25187952	Met(ATG)158Leu(CTG)	C	C	A
		mCV25187953	Ile(ATA)153Met(ATG)	G	G	A
9930020N01	mCG3832	mCV25201504	Leu(CTC)133Phe(TTC)	C	C	T
Dlat	mCG4170	mCV22305679	Glu(GAA)181Asp(GAT)	T	T	A
Hspb2	mCG4168	mCV24348544	Arg(CGC)28Cys(TGC)	C	C	T
1110032A03Rik	mCG4164	mCV25206353	Asn(AAC)8Ser(AGC)	A	A	G
D10Mit198						

The Celera database was used to compare B6 with D2. Fourteen genes with similar expression levels between BKS and D2 are listed. The sequence was confirmed by sequencing the coding regions of the BKS genome.

[a] *Hyou 1* does not contain a splice acceptor site polymorphism between BKS and D2, although this polymorphism is present between B6 and D2.

chromosomes (data not shown) that will be reported independently.

In the F2-*db*/*db* population, BB mice at *D9Mit93* exhibited higher body weight and parametrial fat pad weight (Figs. 3B and 3C). However, in Table 1, BKS-*db*/*db* (BB at *D9Mit93*) had lower scores for these traits than F2-*db*/*db*. This finding led us to suspect some interaction existed between the cQTL on Chr. 9 and other loci. We searched for interaction effects between the cQTL on Chr. 9 and other loci by testing all pairs of marker loci in Map Manager QTX [7], but significant interactions were not detected. Some QTL found on the other chromosomes clearly presented the DD genotype at their loci with higher body weight and parametrial fat pad weight (data not shown). Thus, it may be possible that the phenotype of F2-*db*/*db* was the simple sum of contributions from many other independent QTL that affected body weight and parametrial fat pad weight.

As described under Effects of the *D9mit93* genotype on phenotypes, the D allele appears to be significantly over-represented in the F2-*db*/*db* population on Chr. 9. There are 132 D and 94 B alleles. We observed some bias of the genotype ratio at *D9mit93* in all F2 progenies generated (i.e., 15 BB:35 DB:18 DD) and this generation bias increased in the female F2-*db*/*db* population. Currently, the reason behind this bias is unknown. To clarify the effect of this bias on our mapping results, we made 10 extra datasets including equal alleles (94 D and 94 B) by random sampling from the complete dataset of F2-*db*/*db*. As a result of all QTL analyses using the 10 extra datasets, no change in the statistical significance of cQTL or eQTL on Chr. 9 was found.

As described under Identification of causative candidates by the eQTL analysis, four genes, *C130036J11, 4432416JRik, Nnmt*, and *Cryab*, were considered as causative candidates. These genes were under *cis*-acting transcriptional regulation in the target region. Additionally, they were coordinately regulated; all four were up-regulated by the B allele (Supplementary Fig. 1). This finding made us speculate the existence of a common promoter region, but the molecular basis for the expression of these genes remains unknown because we did not determine the sequences of their promoters or introns. Moreover, they did not contain the missense cSNPs between BKS and D2.

The target region on Chr. 9 reported in this study is located in almost the same region as *Obq5* (gender-influenced obesity QTL) reported by Taylor et al. [8]. Gender influence was also observed in our study in the sense that cQTL was seen only in female F2-*db*/*db* mice (data not shown). Although Taylor et al. regarded the apolipoprotein gene cluster as responsible for the change of *Obq5*, we did not observe eQTL or any amino acid change in these apolipoproteins.

The human syntenic interval of our target region on Chr. 9 corresponds to 11q23. This genomic interval was previously linked to T2D or obesity in Pima Indians, for which *ORP150* (oxygen-regulated protein 150 kDa) was the presumed candidate gene [9]. However, our study foreclosed the possibility that *Hyou1*, a mouse homolog of *OPR150*, was certified as a candidate. Recently, it was reported that *Cryab*, with the second highest eQTL lod score of 12.26, was related to diabetic neuropathy [10]. It was reported that *Cryab* might be expressed in oligodendrocytes after long-term hyperglycemic stress.

In this report, we presented one practical method for providing the list of candidate genes responsible for the effects of cQTL. Actually, 17 candidates were found from 106 genes in the target region of approximately 10 Mb. There are many genes in the QTL region that showed differences in their expression levels between parental strains of the F2 population. However, they are not helpful for reducing the number of candidate genes because most of them are expressed under *trans*-acting transcriptional regulation in the F2 population (Table 2). The genetic determinants responsible for cQTL surely exist inside the cQTL region but their causes of differential expression are derived from other loci outside the target QTL region. How do we exclude them from the candidates? One way is the construction of congenic strains. Compared with the levels of gene expression between congenic and background (host) strains, the differences in gene expression are usually the result of *cis*-acting transcription because congenic strains are identical with the background strains, except for the QTL region. Nonetheless, constructing congenic strains still requires a great deal of effort and congenic strains do not always reproduce the phenotypes assigned to the cQTL. In contrast, eQTL analysis of the F2 population enabled effective identification of candidates that were expressed under *cis*-acting regulation without constructing congenic strains.

To cope with the extremely laborious task of identifying the candidate genes for a complex trait variation, here we showed that our combined approach based on “the overlapping of eQTL with cQTL” and “interstrain sequence variations including cSNPs” accelerates the discovery of candidate genes in the QTL of interest spanning several megabases. Our combined approach should circumvent or at least alleviate the need for the generation of a series of congenic mice with a small interval assignment. The identification of novel candidate genes and the clarification of their mechanisms of influence should contribute to the eventual understanding of functional variability not only in mice, but also in humans. The candidate genes found here in mice can also serve as candidate modifier genes of T2D in the human association study.

Materials and methods

Animals and genetic crosses for QTL analysis

Parental inbred male BKS.Cg-*Lepr*^{d}+/+*m* and female D2 mice were purchased from CLEA Japan, Inc. (Tokyo,

Japan). BKS-*db*/+*m* is maintained with *db* and *m* in repulsion. The recessive misty (*m*) mutation causes a mild dilution of coat color. Because of the sterility of *db* homozygotes, *m* has been incorporated into stocks for maintenance of the *db* mutation.

Mice were fed ad libitum and maintained in a temperature- and humidity-controlled environment with a 12-h light–dark cycle. Female DBA mice were crossed with male BKS-*db*/+*m* to produce F1 hybrid mice. F1 progeny heterozygous for *db* were intercrossed to produce the F2 generation. We selected 113 female progeny homozygous for *db* from 805 F2 intercrosses for this study.

Genetic diagnosis of the db gene in F2 mice

Genomic DNA was extracted from tail tips at 4 weeks of age by an automatic DNA Isolation System NA-1000 (Kurabo, Osaka, Japan). The genotype at the *db* locus in F2 mice was determined using the TaqMan PCR method (Applied Biosystems (ABI), Foster City, CA, USA). The minor groove binder (MGB) TaqMan probe was created using a G/T point mutation of the leptin receptor [3]. The sense primer was 5′-CAACTTCCCAACAGTCCATACAA-TATTA-3′ and the antisense primer was 5′-AAACTGAAC-TACATCAAACCTACATTGTG-3′. A TaqMan MGB probe for the wild type was 5′-FAM-TGGAGGGAAACAAA-MGB-3′. A TaqMan MGB probe for the mutation was 5′-VIC-TGGAGGTAAACAAAC-MGB-3′. The PCR assay mixture contained 20–200 ng of genomic DNA, 900 nM each primer, 200 nM each TaqMan MGB probe, and Platinum qPCR Supermix UDG (Invitrogen Corp., Carlsbad, CA, USA). PCR conditions were based on the manufacturer's recommendations.

Measurement of quantitative traits including BG concentrations, body weight, and parametrial fat pad weight

The ipGTT were performed at 8 weeks of age (56 ± 1 days). After the mice were fasted for 16 h, 2 mg/g body wt of glucose in physiological saline was intraperitoneally administered, and blood was drawn from a tail vein at 0, 30, 60, and 120 min to measure BG concentrations. BG concentrations were measured by the glucose oxidase method (Antisense II; Bayer Sankyo, Inc., Japan). At 9 weeks of age, mice were weighed and sacrificed after fasting for 16 h and parametrial fat pads, liver, and soleus were sampled and weighed for RNA preparation.

Microsatellite genotyping and QTL analysis

A total of 227 well-amplified microsatellite markers with a difference of 2 bp or more that are polymorphic between BKS and DBA were selected. PCR was performed to genotype all markers using fluorescently labeled primers and Platinum *Taq* DNA polymerase (Invitrogen Corp.). PCR conditions were based on the manufacturer's instructions. Amplicons were electrophoretically separated on a 3700 capillary DNA sequencer (ABI), and two independent researchers performed the size calling. Analysis and positional assignment of the markers were performed with Genotyper version 3.7 (ABI), in each of the two strains independently for each marker. QTL analysis, bootstrap analysis, and calculation of genetic variance or interaction effects were performed with Map Manager QTXb17 [7] and QTL Cartographer version 2.0 (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm). We adopted a lod score of 2.8 as the lower limit for the suggestive level in the analysis [11].

RNA preparation

For total RNA preparation, cubic tissue 5 mm in size was obtained from parametrial fat pads or liver of the mice, and total soleus was removed at 9 weeks of age. All tissue samples were flash-frozen in liquid nitrogen and stored at −80°C. Total RNA was purified using a RNeasy Mini kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions and stored at −80°C until used.

Measurement of gene expression

To measure gene expression, only established sets of primers and probes from Mouse Assays-on-Demand Gene Expression Products (ABI) were utilized. A total of 76 predesigned primer and probe sets were commercially available in our target region between *D9mit206* and *D10mit198*. These 76 genes correspond to about 70% of the 106 genes with reference sequences assigned to this target region.

Quantitative real-time RT-PCR was performed with TaqMan One-Step RT-PCR Master Mix Reagents Kit (ABI) and mRNA levels were quantitated with an ABI Prism 7900HT sequence detector (ABI). We used two quantitative methods to assess the amount of a particular transcript: the relative standard curve and the comparative cycle time (C_t) methods [12]. The relative standard curve method requires that a template dilution series of a cDNA standard be included for each gene. For relative quantitation, the template concentration values are arbitrary. Quantities interpolated from the resulting standard curve are used to calculate relative mRNA levels in each unknown sample. For each sample, we determined the quantity of the gene of interest and the reference gene in triplicate. Additionally, the quantity of the gene of interest was normalized to that of a reference gene, e.g., *GAPDH* for fat samples and β-actin for liver and soleus samples. The comparative C_t method was used to assess relative changes in mRNA levels between two samples and does not require the use of a template dilution series. It does, however, require the same amplification efficiencies of the compared genes. Cycle times rather than interpolated template quantities were used for the calculations.

First, for 76 genes encoded in the target region (*D10mit198–D9mit206*), their mRNA levels were measured in triplicate in parametrial fat pads, liver, and soleus of BKS-*db*/*db* (n = 18) and D2 (n = 15) mice. Fifteen genes that are differentially expressed between BKS-*db*/*db* and D2 were identified. The mRNA levels of these 15 genes were also confirmed for BKS-*db*/+*m* (n = 12) and D2 (n = 15) mice in triplicate. For the 15 differentially expressed genes, mRNA levels in the F2-*db*/*db* population were examined through an eQTL analysis. In this case, the mRNA levels were examined only in the same tissues in which differences in expression between BKS-*db*/*db* and D2 were detected.

Statistical analysis

Statistical analysis of measured traits was carried out using StatView for Windows version 5.0 (SAS Institute, Inc., Cary, NC, USA). Phenotypic comparison for different genotype groups was performed by the classical Kruskal–Wallis test with a post hoc test using Scheffe's test for multiple comparisons.

Sequencing analysis of the BKS genome to confirm cSNPs

We selected cSNPs accompanied by amino acid variation between B6 and D2 mice in our target region between *D10mit198* and *D9mit93* based on the Celera database. These cSNPs were confirmed to be polymorphic between BKS and D2 by sequencing the BKS genome. PCR products were purified with a QIAquick PCR purification Kit (Qiagen GmbH) and were used as template in a sequencing reaction with a BigDye Terminator version 1.1 cycle sequencing kit (ABI). Cycle sequencing conditions were based on the manufacturer's instructions. Sequencing was performed on a 3100 automated fluorescence sequencer (ABI).

Detection of interstrain polymorphisms in silico

All consensus transcript sequences within the region between *D10mit198* and *D9mit206* were downloaded from Celera's Mouse Genome Assembly under the Celera Discovery System (http://cds.celera.com/biolib/cdsTopLibrary.jsp). To search for coding variations in these transcripts, the Celera Mouse RefSNP database (release 1.0) containing 2,566,706 SNPs between 129X1/SvJ, 129S1/SvImJ, D2, A/J, and B6 mouse strains was utilized. This database was utilized to screen for SNPs that cause missense mutations between B6 and DBA within the target region.

Acknowledgments

This study was supported by a grant from Otsuka Pharmaceutical Factory, Inc., a grant from the Japan Society for the Promotion of Science (Grant for Genome Research in the Research for the Future Program), and the Cooperative Link of Unique Science and Technology for Economy Revitalization (CLUSTER).

Appendix A. Supplementary data

Supplementary data for this article may be found on ScienceDirect.

References

[1] R. Korstanje, B. Paigen, From QTL to gene: the harvest begins, Nat. Genet. 31 (2002) 235–236.
[2] E.E. Schadt, et al., Genetics of gene expression surveyed in maize, mouse and man, Nature 422 (2003) 297–301.
[3] H. Chen, et al., Evidence that the diabetes gene encodes the leptin receptor: identification of a mutation in the leptin receptor gene in db/db mice, Cell 84 (1996) 491–495.
[4] P. Bjorntorp, Fatty acids, hyperinsulinemia, and insulin resistance: which comes first? Curr. Opin. Lipidol. 5 (3) (1994) 166–174.
[5] J.K. Naggert, J.L. Mu, W. Frankel, D.W. Bailey, B. Paigen, Genomic analysis of the C57BL/Ks mouse strain, Mamm. Genome 6 (1995) 131–133.
[6] K.P. Hummel, D.L. Coleman, P.W. Lane, The influence of genetic background on expression of mutations at the diabetes locus in the mouse. I. C57BL/KsJ and C57BL/6J strains, Biochem. Genet. 7 (1972) 1–13.
[7] K.F. Manly, J.M. Olson, Overview of QTL mapping software and introduction to MapManager QT, Mamm. Genome 10 (1999) 327–334.
[8] B.A. Taylor, L.M. Tarantino, S.J. Phillips, Gender-influenced obesity QTLs identified in a cross involving the KK type II diabetes-prone mouse strain, Mamm. Genome 10 (1999) 963–968.
[9] P. Kovacs, X. Yang, P.A. Permana, C. Bogardus, L.J. Baier, Polymorphisms in the oxygen-regulated protein 150 gene (ORP150) are associated with insulin resistance in Pima Indians, Diabetes 51 (2002) 1618–1621.
[10] M. Yaguchi, K. Nagashima, T. Izumi, K. Okamoto, Neuropathological study of C57BL/6 Akita mouse, type 2 diabetic model: enhanced expression of alpha B-crystalline in oligodendrocytes, Neuropathology 23 (2003) 44–50.
[11] E.S. Lander, L. Kruglyak, Genetic dissection of complex traits: guidelines for interpreting and reporting linkage result, Nat. Genet. 11 (1995) 241–247.
[12] A.L. Bookout, D.J. Mangelsdorf, A quantitative real-time PCR protocol for analysis of nuclear receptor signaling pathways, NURSA e-J 1 (6) (2003) (ID 1.11082003.1).