Brief Genetics Report

Distinguishing Covariation From Causation in Diabetes

A Lesson From the Protein Disulfide Isomerase mRNA Abundance Trait

Hong Lan,[1] Mary E. Rabaglia,[1] Kathryn L. Schueler,[1] Christine Mata,[1] Brian S. Yandell,[2] and Alan D. Attie[1]

Protein disulfide isomerase (*Pdi*) is reported to be an insulin-regulated gene whose expression level is increased in the livers of rats with streptozotocin-induced diabetes. We found that *Pdi* mRNA is ~20-fold more abundant in the diabetes-susceptible BTBR mouse strain relative to the diabetes-resistant C56BL/6 (B6) strain. A genetic analysis was carried out to determine whether there is a causal relationship between elevated *Pdi* expression and diabetes phenotype in BTBR-*ob/ob* mice. We mapped *Pdi* mRNA abundance as a quantitative trait in 108 (B6 × BTBR)F_2-*ob/ob* mice segregating for diabetes. We detected a single linkage at the telomeric end of chromosome 11, where the *Pdi* gene itself resides (logarithm of odds score >30.0). No linkage was detected for the *Pdi* mRNA trait in the regions where we have previously identified quantitative trait loci for diabetes traits. Sequencing of the *Pdi* promoter and cDNA revealed several single nucleotide polymorphisms between these two mouse strains. We conclude that in our experimental model, elevated *Pdi* expression is *cis* regulated and is not linked to diabetes susceptibility. Genetic analysis is a powerful tool for distinguishing covariation from causation in expression array studies of disease traits. *Diabetes* 53:240–244, 2004

Gene array technology has made it possible to simultaneously interrogate the expression of virtually all expressed genes in a cell. This has prompted numerous studies of gene expression in disease. The chief rationale is that altered patterns of gene expression might shed light on links between gene expression and the disease phenotype. However, a major difficulty in the interpretation of gene expression data is distinguishing covariation from causation. Genes whose expression levels change in a disease state may or may not be related to the etiology or pathology of the underlying disease. This is particularly evident in diabetes research. If tissue samples are collected from diabetic animals, it is difficult to know whether differences in gene expression are a cause or consequence of diabetes. If the samples are collected before the onset of diabetes, it is still difficult to distinguish generic strain differences from strain differences causally related to diabetes. Similarly, when carrying out analyses of genes that covary in comparison groups, it is difficult to infer causation from covariation.

Genetic analysis establishes a one-way chain of causation, from genomic variation to differential gene expression. Gene expression as measured through mRNA abundance is affected by both genetic variations and environmental factors; therefore, it can be a heritable trait (1). Genetic contribution to the expression variation is a reflection of DNA sequence variations in the expression control elements. In eukaryotes, the abundance of an mRNA transcript is controlled either locally by *cis*-regulatory elements, such as sequence polymorphisms in the promoters and in the 3′-untranslated regions (UTRs), or remotely by *trans*-acting elements, such as transcription factors and RNA-binding proteins, or simultaneously by both. In a segregating population, a trait controlled by a *cis*-acting gene will be mapped to a locus where the gene itself resides, whereas a trait controlled by a *trans*-acting gene may be linked to one or more loci. Recent studies in model organisms have shown that one can map mRNA abundance traits as quantitative trait loci (QTL) (2,3).

Our laboratory has been studying gene expression in a diabetes-susceptible congenic mouse strain, BTBR-*ob/ob*. The strain was derived by introgression of the leptin^{ob} allele from the donor C57BL/6 (B6)-*ob/ob* mutant strain. In contrast to B6-*ob/ob* mice, BTBR-*ob/ob* mice are severely diabetic (4,5). In a survey for genes that are differentially expressed in diabetic BTBR-*ob/ob* mice, we observed an ~20-fold increase in the mRNA abundance of the protein disulfide isomerase (*Pdi*) gene. PDI (the mouse gene is also called *P4hb*, *Thbp*, and *Erp59*; locus ID: 18453) is an abundant luminal endoplasmic reticulum protein with diverse functions in the endoplasmic reticulum (6). PDI is also a component of the enzymes prolyl 4-hydroxylase and microsomal triglyceride transfer protein. *Pdi* mRNA is

TABLE 1
Pdi mRNA abundance in pancreatic islets of lean and obese B6 and BTBR mice

	Pdi expression		Level	Plasma glucose (mg/dl)	
	B6	BTBR	B6	B6	BTBR
Lean mice	1.0	33.0 ± 2.7	ND	ND	
6-week female *ob/ob* mice	2.1 ± 0.1	53.5 ± 5.4	163	166	
6-week male *ob/ob* mice	1.7 ± 0.3	52.8 ± 4.3	185	333	
14-week female *ob/ob* mice	ND	42.4 ± 5.4	ND	289	
14-week male *ob/ob* mice	1.1	65.4 ± 1.8	149	539	

Data are means ± SE for those groups with at least two mice. The mRNA levels were measured by qRT-PCR and were normalized to those of the B6 lean mice, which was set at 1.0. There were 1–5 mice in each of the comparison groups. The average plasma glucose (mg/dl) for each comparison group is also listed for reference.

abundant in a wide spectrum of tissues. The promoter of the human *PDI* gene contains the elements common to constitutively expressed housekeeping genes. Among other elements in the *Pdi* promoter are a TATA box, six CCAAT boxes, and five *Sp1* interacting sites, all located within 600 nucleotides from the transcription start site; all are believed to be responsible for efficient transcription of the *Pdi* gene (7,8).

Of relevance to diabetes, Nieto et al. (9) reported that the mRNA encoding for rat PDI was increased threefold in the livers of rats with streptozotocin-induced diabetes. The higher expression of rat *Pdi* mRNA in diabetes was due to an increase in the transcription rate of the gene, and insulin treatment of diabetic animals reversed the effect (9). *Pdi* has since been regarded as an insulin-regulated gene (6,10). We hypothesized that changes in *Pdi* gene expression might be causally related to the susceptibility of the BTBR strain to diabetes. In this study, we show how genetics can be used to test causation.

Elevated mRNA abundance in BTBR mice. We first observed elevated *Pdi* mRNA abundance in BTBR mice in pancreatic islets, where we searched for differentially expressed genes associated with the diabetes-susceptibility phenotype in the BTBR-*ob/ob* mice (5). The BTBR islets show 20- to 60-fold higher levels of *Pdi* mRNA than the B6 islets, independent of age, sex, obesity, or diabetes (Table 1).

To determine whether the elevated *Pdi* mRNA in BTBR mice is specific for the pancreatic islets, we surveyed the *Pdi* mRNA levels in several tissues from these two mouse strains. Although *Pdi* mRNA levels vary across different tissues, they are consistently higher in all the tissues surveyed in the BTBR mice (Table 2). The difference in

expression ranges from 6- to 40-fold. As in islets, the strain difference in *Pdi* expression was unrelated to body weight; it occurred in both lean and obese mice.

In contrast to the observations in the livers of diabetic rats (9), insulin does not seem to be responsible for the differential *Pdi* expression in our models. The B6-*ob/ob* mice are known to be hyperinsulinemic, with fasting plasma insulin levels >20-fold higher than in the lean mice (4,5). However, in adipose tissue as well as liver and skeletal muscle (all insulin-responsive tissues), *Pdi* mRNA levels are comparable between lean and obese mice. Also, glucose levels do not correlate with *Pdi* expression. The BTBR-*ob/ob* mice have a fasting glucose level >400 mg/dl, but the *Pdi* mRNA levels in the diabetic BTBR-*ob/ob* mice are comparable with those in the lean BTBR mice in all the tissues surveyed.

PDI can be induced by stress-inducing agents, such as heat shock, tunicamycin, cycloheximide, dithiothreitol, etc., but PDI levels seldom change more than three- or fourfold (11). Thus, changes in physiological conditions between these two mouse strains are unlikely to explain the ~20-fold constitutive elevation of *Pdi* mRNA in the BTBR strain. The data in Tables 1 and 2 strongly suggest that the elevated *Pdi* mRNA in the BTBR mice is due to a genetic difference between these two mouse strains. However, the data cannot address whether the strain difference responsible for the *Pdi* expression is causally related to the diabetes susceptibility in the BTBR-*ob/ob* mice. Studying the segregation of *Pdi* mRNA abundance in a population segregating diabetes susceptibility alleles is a powerful tool to distinguish causation from an unrelated strain difference.

Linkage mapping. To identify the expression control elements, we genetically mapped their loci through linkage mapping. We previously mapped gene expression phenotypes in a population of (B6 × BTBR) F_2-*ob/ob* mice (12). We used the same system to map the *Pdi* mRNA determinants in the liver. The mapping panel consists of 108 F_2-*ob/ob* mice. The mice were genotyped for 191 microsatellite markers covering the 19 mouse autosomes with an average spacing of 20 cM.

The genome-wide scan revealed a single linkage peak on chromosome 11 (Fig. 1) with a maximum logarithm of odds (LOD) score >30.0 (Fig. 1*A*). This locus explains 88% of the phenotypic variance as estimated by MAPMAKER/QTL. The strikingly high LOD score and the lack of linkage elsewhere suggest a single-gene control of the liver *Pdi* mRNA phenotype. The linkage peak is located near the telomeric end of mouse chromosome 11, where the *Pdi*

TABLE 2
Pdi mRNA abundance in various tissues of B6 and BTBR mice

	Lean mice		*ob/ob* mice	
	B6	BTBR	B6	BTBR
Muscle	2.7 ± 0.7	101.7 ± 6.7	8.2 ± 4.7	53.8 ± 4.3
Adipose tissue	1.8 ± 0.4	40.0 ± 3.7	1.0 ± 0.2	41.9 ± 20.3
Liver	8.9 ± 1.3	273.6 ± 24.6	14.6 ± 1.4	241.8 ± 54.4
Pancreatic acinar	ND	ND	8.8 ± 3.1	223.5 ± 11.9
Pancreatic islets	6.1 ± 1.3	52.1 ± 10.5	1.9 ± 0.2	58.8 ± 3.3

Data are means ± SE. The mRNA levels were measured by qRT-PCR and were normalized to those in the adipose tissue of B6-*ob/ob* mice. There were 2–4 mice in each of the comparison groups.
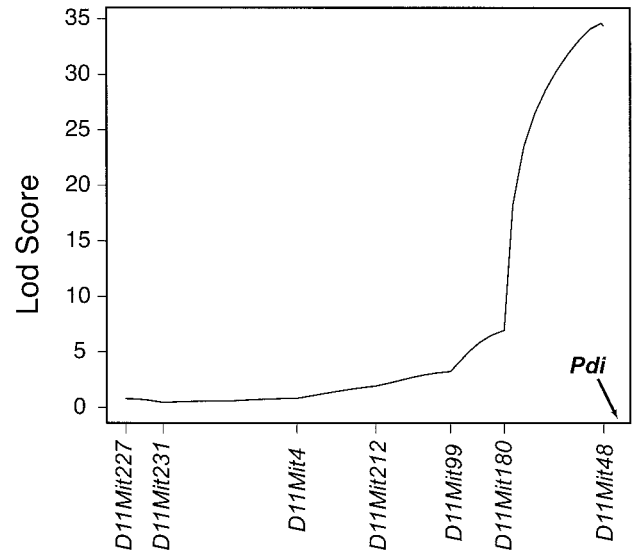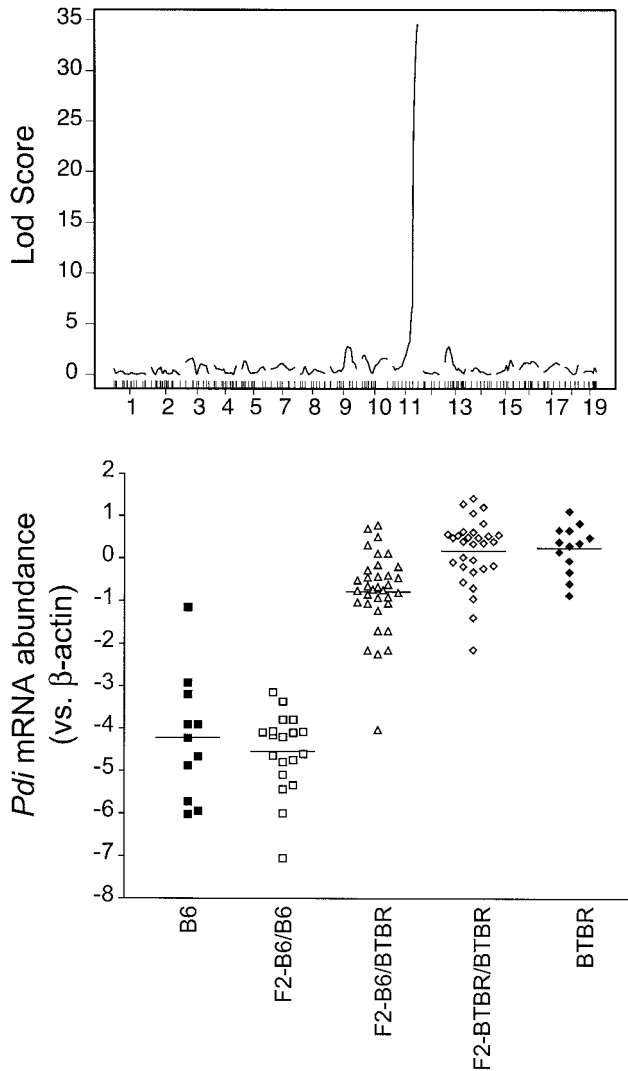
FIG. 1. *A*: Genome-wide LOD profile of liver *Pdi* mRNA abundance. The LOD scores were computed with interval mapping in the MAPMAKER/QTL program. The ticks on the *x*-axis are locations of genetic markers. The numbers below the *x*-axis represent chromosomes (chromosomes 12, 14, 16, and 18 are not annotated). *B*: Expansion of the *Pdi* mRNA LOD peak on chromosome 11. The genetic markers and their mapping locations are listed on the *x*-axis. The arrow in the lower right-hand corner marks the physical location of the *Pdi* gene. *C*: Dot plot of *Pdi* mRNA levels in the parental strains and in the F_2 mice. Each dot represents the *Pdi* mRNA abundance of an individual mouse as fold difference ($\log_2$) relative to the level of the housekeeping gene β-actin. F_2 mice with different genotypes at marker *D11Mit48* are represented by different symbols.

gene itself resides (Fig. 1*B*). The linkage was confirmed with multiple interval mapping (data not shown). Figure 1*C* shows the values of the F_2 mice according to their genotypes at the closest marker, *D11Mit48*, which is ~3 cM upstream of the *Pdi* locus. Mice with homozygous B6/B6 or homozygous BTBR/BTBR alleles have *Pdi* mRNA levels similar to the respective parental strains ($P > 0.05$). Mice with heterozygous genotypes are closer to the BTBR strain than to the B6 strain, implying a dominant effect of the BTBR allele. The degree of dominance (dominance/additivity) is ~60%.

Polymorphisms in promoter and cDNA sequences between B6 and BTBR mice. As a *cis*-regulated gene, the higher *Pdi* mRNA abundance in BTBR mice could be due to either a higher rate of transcription or higher mRNA stability. In an attempt to determine the molecular basis of this phenomenon, we searched the promoter regions of both strains for sequence variations. No difference was found in the region within 1.0 kb of the transcription site. There was one single nucleotide polymorphism (SNP) in the promoter region between B6 and BTBR at nucleotide 1341 (Mouse Genome Browser, February 2003 Assembly, coordinate at chromosome 11:121454350), a T→G transition (B6→BTBR). We used the TFSEARCH (13) program, which uses the TRANSFAC databases (14), to search for

transcription factor binding sites. The results show that this SNP may affect one of the several binding sites for the GATA transcription factor (15). *Pdi* gene transcription is largely controlled by elements in the promoter within 600 nucleotides of the transcription site (8). Deletion of all the six CCAAT boxes in the *Pdi* promoter still retains 10% transcriptional activity (7). Since the mutation in the BTBR strain only affects one of the several potential GATA sites, it is unlikely this single point mutation would be responsible for an ~20-fold constitutional elevation of *Pdi* gene transcription.

Sequence changes in the 3′-UTR can affect RNA stability. We sequenced the full-length *Pdi* cDNA sequences in B6 and BTBR strains. There are seven SNPs, four of which are in the coding region (ORF) and the other three in 3′-UTR. None of the SNPs in the ORF would change the encoded amino acids. To assess the possible impacts of these SNPs on mRNA stability, we checked whether these changes would be predicted to affect RNA secondary structure. We used the Zuker minimum free energy algorithm (16), implemented in the software mFOLD (version 3.1), to predict the mRNA secondary structures of the B6 and BTBR transcripts. Among the seven SNPs, six are predicted to have no or a minor effect on RNA secondary structure. A C→U substitution at position 1509 in the

TABLE 3
SNPs in the B6 and BTBR *Pdi* cDNA sequences and their predicted impact on mRNA secondary structure

Position (B6-BTBR)	Region	B6			BTBR			Impact
		Base	Paired with	Structure	Base	Paired with	Structure	
831(U-C)	ORF	831-U	952-A	start of a 9-bp stem	831-C	none	next to a 8-bp stem	Decrease stem size by 1
1077(U-C)	ORF	1077-U	none	loop	1077-C	none	loop	No change
1209(A-G)	ORF	1209-A	2487-U	start of a 15-bp stem	1209-G	2487-U	start of a 15-bp stem	No change
1509(C-U)	ORF	1509-C	none	loop	1509-U	2170-A	10-bp stem	Structure change
2049(C-U)	3′-UTR	2049-C	none	loop	2049-U	none	loop	No change
2098(C-U)	3′-UTR	2098-C	2127-G	5-bp stem	2098-U	2127-G	5-bp stem	No change
2240(A-G)	3′-UTR	2240-A	2203-U	14-bp stem	2204-G	2203-U	14-bp stem	No change

The table shows the most probable structure predicted by mFOLD. The positions in column 1 are indexed with coordinates of the *Pdi* mRNA sequence in the Genebank (*J05185*).

BTBR strain is predicted to form a new stem, which is absent in the B6 strain (Table 3). The predicted secondary structures of *Pdi* mRNA for B6 and BTBR strain look very similar (data not shown). It is not known whether this mutation would cause an ~20-fold reduction of mRNA degradation in the BTBR mice. It is puzzling that the BTBR phenotype is dominant over the B6 phenotype, even though it is a *cis*-acting difference.

In conclusion, we have used genetic analysis to distinguish covariation from causation in a gene expression study. Previously, we mapped QTLs of the insulin trait to chromosomes 2, 16, and 19 and the glucose trait to chromosomes 16 and 19 (4). *Pdi* mRNA abundance is independent of insulin or glucose traits in the same F_2 population. Here, we show that *Pdi* mRNA maps to chromosome 11. Therefore, we conclude that the *Pdi* mRNA abundance trait of the BTBR strain is not associated with insulin or glucose and is unrelated to the strain's diabetes susceptibility.

RESEARCH DESIGN AND METHODS

Quantitation of *Pdi* mRNA. The mRNA abundance in B6 and BTBR mice was estimated using the quantitative real time RT-PCR (qRT-PCR) assay. The primers were designed against the mouse full-length *Pdi* mRNA sequence (J05185) in the GeneBank. The primer sequences are forward 5′-tttcaccatg-gcagacctcc and reverse 5′-ccatggcaacactaggacaagg. The primers and the amplicon do not harbor polymorphisms between the B6 and BTBR strains, so that any differential expression observed cannot be due to a primer-annealing artifact. The housekeeping gene, *β-actin* (M12481), was used as a normalization control. The primer sequences are forward 5′-ccatcctgcgtctggacttg and reverse 5′-ttccctctcagctgtggtgg. qRT-PCR was performed as described previously (12).

Linkage mapping. The 108 F_2-*ob/ob* mice were a subset of the F_2 population derived from the B6 and BTBR strains that we previously used to study QTLs associated with obesity and diabetes (4). There were 191 microsatellite markers spanning the 19 mouse autosomes that were genotyped and assembled into a framework map using MAPMAKER/EXP (17). Interval mapping of liver *Pdi* mRNA traits in MAPMAKER/QTL was used to reveal the statistical significance of the linkages. Multiple interval mapping (18) implemented in WinQTLCart v2.0 was used to confirm the results from the interval mapping.

Sequencing.
Promoter. A genomic DNA segment of 2,000 bp, indexed at chromosome 11:121453010–121455009 in the February 2003 Assembly, was retrieved with the Mouse Genome Browser (19) and used as a reference sequence for the *Pdi* promoter region. The 5′ transcription start site of the mouse *Pdi* gene is unknown. The published human *PDI* gene promoter (7), M22803, may in fact contain some mRNA sequences. BLAST search and pairwise sequence alignment showed that the first 61 nucleotides of the mouse *Pdi* mRNA (J05185) are mapped into the human "promoter" sequence, including the proposed TATA box, suggesting some of the promoter sequence is actually exon 1 of the human *PDI* gene. Among the available *Pdi* mRNA and EST sequences, J05185 has the most extended 5′ sequence, so we treated the first base of the J05185 sequence as the transcription start site. The *Pdi* mRNA was transcribed from the reverse genomic DNA strain. Our template sequence starts two nucleotides before the published sequence J05185, whose first nucleotide aligned to chromosome 11:121453008 in the genome assembly. The *Pdi* mRNA sequence XM_126743.3 has more 5′ sequence than J05185, in which the first 397 nucleotides match the 5′ end of our promoter template sequence. However, this 397-nucleotide "mRNA" sequence was based on automatic computer prediction. The most 5′ sequence was not supported by any experimental mRNA or EST data. We therefore treat it as nontranscriptional sequence. We designed four pairs of PCR primers to amplify overlapping fragments in the *Pdi* promoter region in both the B6 and the BTBR genomic DNA. PCR products were purified using Qiagen columns before they were subject to sequencing. Sequencing was performed using Big-Dye reagents from Applied Biosystems. Sequences are assembled and analyzed using Sequencher 4.1.4 (Gene Codes).
cDNA. We designed four pairs of PCR primers based on the *Pdi* mRNA sequence J05185 to amplify overlapping fragments from cDNA made from total RNA isolated from both B6 and BTBR livers. The fragments cover ~2.5 kb *Pdi* transcript of 5′-UTR, coding region, and 3′-UTR. PCR sequencing was performed and analyzed as with promoter sequencing.
RNA secondary structure analysis. The RNA secondary structure prediction and the affects of mutations were performed using the Zuker minimum free energy algorithm implemented in mFOLD version 3.1 (16).

REFERENCES

1. Cheung VG, Conlin LK, Weber TM, Arcaro M, Jen KY, Morley M, Spielman RS: Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat Genet* 33:422–425, 2003
2. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH: Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422:297–302, 2003
3. Brem RB, Yvert G, Clinton R, Kruglyak L: Genetic dissection of transcriptional regulation in budding yeast. *Science* 296:752–755, 2002
4. Stoehr JP, Nadler ST, Schueler KL, Rabaglia ME, Yandell BS, Metz SA, Attie AD: Genetic obesity unmasks nonlinear interactions between murine type 2 diabetes susceptibility loci. *Diabetes* 49:1946–1954, 2000
5. Lan H, Rabaglia ME, Stoehr JP, Nadler ST, Schueler KL, Zou F, Yandell BS, Attie AD: Gene expression profiles of nondiabetic and diabetic obese mice suggest a role of hepatic lipogenic capacity in diabetes susceptibility. *Diabetes* 52:688–700, 2003
6. Noiva R: Protein disulfide isomerase: the multifunctional redox chaperone of the endoplasmic reticulum. *Semin Cell Dev Biol* 10:481–493, 1999
7. Tasanen K, Oikarinen J, Kivirikko KI, Pihlajaniemi T: Promoter of the gene for the multifunctional protein disulfide isomerase polypeptide: functional significance of the six CCAAT boxes and other promoter elements. *J Biol Chem* 267:11513–11519, 1992

8. Tasanen K, Oikarinen J, Kivirikko KI, Pihlajaniemi T: Interaction of transcription factor Sp1 with the promoter of the gene for the multifunctional protein disulphide isomerase polypeptide. *Biochem J* 292:41–45, 1993

9. Nieto A, Mira E, Castano JG: Transcriptional regulation of rat liver protein disulphide-isomerase gene by insulin and in diabetes. *Biochem J* 267:317–323, 1990

10. O'Brien RM, Granner DK: Regulation of gene expression by insulin. *Physiol Rev* 76:1109–1161, 1996

11. Odani N, Negishi M, Takahashi S, Ichikawa A: Induction of protein disulfide isomerase mRNA by delta 12-prostaglandin J2. *Biochem Biophys Res Commun* 220:264–268, 1996

12. Lan H, Stoehr JP, Nadler ST, Schueler KL, Yandell BS, Attie AD: Dimension reduction for mapping mRNA abundance as quantitative traits. *Genetics* 164:1607–1614, 2003

13. Akiyama Y: TFSEARCH: Searching transcription factor binding sites [article online]. 1.3 ed. Tokyo, Japan, Real World Computing Partnership, 1998. Available from http://www.crbc.jp/research/db/TFSEARCH-notice.html

14. Heinemeyer T, Wingender E, Reuter I, Hermjakob H, Kel AE, Kel OV, Ignatieva EV, Ananko EA, Podkolodnaya OA, Kolpakov FA, Podkolodny NL, Kolchanov NA: Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. *Nucleic Acid Res* 26:362–367, 1998

15. Merika M, Orkin SH: DNA-binding specificity of GATA family transcription factors. *Mol Cell Biol* 13:3999–4010, 1993

16. Jaeger JA, Turner DH, Zuker M: Improved predictions of secondary structures for RNA. *Proc Natl Acad Sci U S A* 86:7706–7710, 1989

17. Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE: MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1:174–181, 1987

18. Kao CH, Zeng ZB, Teasdale RD: Multiple interval mapping for quantitative trait loci. *Genetics* 152:1203–1216, 1999

19. UCSC Genome Browser [article online], 2002 (updated Feb 2003). Available from http://genome.ucsc.edu/. Accessed 7 October 2002.