Ultrafine Mapping of SNPs From Mouse Strains C57BL/6J, DBA/2J, and C57BLKS/J for Loci Contributing to Diabetes and Atherosclerosis Susceptibility

Richard C. Davis,[1] Eric E. Schadt,[2] Alessandra C.L. Cervino,[2] Miklós Péterfy,[1,3] and Aldons J. Lusis[1,4]

The inbred mouse strain C57BLKS/J (BKS) carrying a mutation of the leptin receptor lepr$^{-/-}$ (BKS-*db*) is a classic mouse model of type 2 diabetes. While BKS was originally presumed to be a substrain of C57BL/6J (B6), it has become apparent that its genome contains introgressed regions from a DBA/2 (DBA)-like strain and perhaps other unidentified sources. It has been hypothesized that the strikingly enhanced diabetes susceptibility of BKS-*db* compared with B6-*db* is conferred by this introgressed DNA. Using high-density single nucleotide polymorphisms, we have mapped the DBA and other contaminating DNA regions present in BKS. Thus, ~70% of its genome appears to derive from B6, with ~20% from DBA and another 9% from an unidentified donor. Comparison with 56 diverse inbred strains suggests that this donor may be a less common inbred strain or an outbred or wild strain. Using expression data from a B6 × DBA cross, we identified differentially regulated genes between these two strains. Those *cis*-regulated genes located on DBA-like blocks in BKS constitute primary candidates for genes contributing to diabetes susceptibility in the BKS-*db* strain. To further prioritize these candidates, we identified those *cis*-acting expression quantitative trait loci whose expression significantly correlates with diabetes-related phenotypes. *Diabetes* 54:1191–1199, 2005

In the 1940s, the inbred mouse strain C57BLKS/J (BKS) was maintained as a substrain of C57BL/6 (B6) by Kaliss. Subsequent analysis has shown that the BKS line carries contaminating DNA regions from a strain similar to DBA/2J (DBA) (1–3). These studies estimated that ~16% of the BKS genome derives from DBA, with the remaining 84% from B6 (2,4). Interestingly, the BKS strain has a greatly increased susceptibility to diabetes and atherosclerosis compared with B6, specifically in animals carrying a mutation in leptin (*ob*) or the leptin receptor (*db*) (5–7). A working hypothesis is that these differences in disease susceptibility derive from shifts in gene expression induced by the introgressed DBA genomic regions. Based on this hypothesis, Mu et al. (4) carried out an F2 cross between BKS and B6-*db/db*. From analysis of female F2 *db/db* offspring, suggestive loci for plasma glucose levels on chromosomes 8 and 17, along with a significant locus for atherosclerotic lesion formation on chromosome 12, were identified. The genes underlying these loci have not been identified. Moreover, there is increasing evidence that the BKS genome carries contributions from strains other than B6 and DBA (2,8).

In this report, we describe the application of ultrafine mapping to characterize the genetic origin of the BKS genome, specifically to identify blocks derived from B6, DBA, and other genetic lines. Further, we combine these data with expression array data of a cross between B6 and DBA to identify those genes within the DBA blocks that are differentially regulated in *cis*. These genes constitute candidates for diabetes and atherosclerosis susceptibility of the BKS strain.

RESEARCH DESIGN AND METHODS

Mouse strains and DNA isolation. DNA for mouse strains B6 (JAX strain 000664), BKS (JAX strain 000662), and DBA (JAX strain 000671) were purchased from The Jackson Laboratory (Bar Harbor, ME). DNA from BKS-*db* (JAX strain 000642) was isolated from tissues using standard phenol/chloroform extraction procedures.

SNP analysis. A total of 15,300 single nucleotide polymorphisms (SNPs), covering the genome at an average spacing of ~167,000 bp (autosomes) and 330,000 bp (X chromosome), were analyzed. SNP analysis was carried out at Illumina (San Diego, CA), using their BeadArray technology (9,10). This involves allele-specific extension from the genomic DNA template followed by PCR amplification of the extended allele. The genotype is determined by the ratio of fluorescent signals from the two allele-specific extension products.

Statistics for distribution of informative SNPs (Runs test). To determine whether SNPs that are informative between B6 and DBA are randomly distributed among all SNPs tested, Runs tests were carried out using the Dataplot program (11) provided by the National Institute of Standards and Technology.

Localization of contributing quantiative trait loci. We previously described (12) an F2 cross between B6 and DBA mice, in which liver mRNAs were analyzed for expression of 23,574 genes (13). Linkage maps were constructed, and quantitative trait locus (QTL) analysis was performed using MapMaker QTL (14,15) and QTL Cartographer (16). Logarithm of odds scores were calculated at 2-cM intervals throughout the genome for each of the

From the [1]Departments of Medicine and Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California; [2]Rosetta Inpharmatics, Merck and Company, Seattle, Washington; the [3]Veterans Administration, Greater Los Angeles Healthcare System, Los Angeles, California; and the [4]Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, California.

Address correspondence and reprint requests to Richard C. Davis, Department of Medicine, Division of Cardiology, David Geffen School of Medicine, University of California at Los Angeles, 47-123 CHS, Los Angeles, CA 90095-1679. E-mail: davisr@ucla.edu.

Received for publication 1 November 2004 and accepted in revised form 18 January 2005.

Additional information for this article can be found in an online appendix at http://diabetes.diabetesjournals.org.

QTL, quantitative trait locus; SNP, single nucleotide polymorphism.

23,574 genes. Using these data, eQTLs controlling expression of these genes were located. Significant (logarithm of odds >4.3) eQTLs were identified for 2,123 genes. Those genes, whose expression levels mapped to within 20 Mb of the location of the gene itself, were identified as *cis*-acting eQTLs.

RESULTS

SNP genotyping was carried out on DNA from B6, DBA, BKS, and BKS-*db*. Of 15,300 SNPs assayed, 12,473 gave genotype data with sufficient quality across a large strain set to be included in the dataset. Based on the rate of heterozygosity, we estimate an error rate of ~0.08%. After removing all SNPs that did not have an identified position in build 32 of the public genome database as well as all those that gave no data or heterozygous genotypes for any of the four strains discussed here, 10,517 SNPs remained. These markers were spaced at an average of 2×10^5 bp on the autosomes and at 5×10^5 bp on the X chromosome. The distributions of total SNPs and informative SNPs for each chromosome are shown in Fig. 2 of the online appendix (available at http://diabetes.diabetesjounrals.org). The SNPs covered the entire length of each chromosome at high density, except for regions within 2–4 Mb of the centromere. Beyond these centromeric regions, there were two notable gaps in SNP coverage on the autosomes. The last 10 Mb of chromosome 7 contained only four SNPs, and, on proximal chromosome X, there was a gap of 5.3 Mb between 22 and 28 Mb. With these exceptions, the maximal interval between SNPs was 2.2 Mb on the autosomes and 2.8 Mb on chromosome X.

For the analysis presented here, we selected the subset of 4,710 SNPs that are polymorphic between the B6 and DBA. The distribution of such informative SNPs is significantly nonuniform, with obvious long stretches largely devoid of informative SNPs. For example, on chromosome 1, as shown in Figs. 1 and 2 of the online appendix, there is a large gap (8–10 Mb) in informative SNPs from 140 to 150 Mb. Such regions have been previously noted by others (17) and have been hypothesized to represent blocks of DNA inherited from a common ancestral strain. For instance, on chromosome 10, only 10 of 197 SNPs between 21 and 84 Mb were informative. This same region was noted by Wiltshire et al. (17). Strikingly, on the X chromosome, only 12% of the markers were polymorphic between B6 and DBA, suggesting a large portion of this chromosome is derived from a common ancestor.

Even outside these large obvious regions of low informativeness, there is significant clustering of informative and uninformative markers as judged by tests based on the binomial distribution (11). Thus, there is a significant overrepresentation of clusters containing greater than three uninterrupted informative or uninformative markers between B6 and DBA. This is consistent with retention in modern inbred strains of discernable blocks, both large and small, of ancestral genomes. While the presence of large gaps in the set of informative SNPs lowers the resolution with which we may map the blocks of introgressed DBA DNA in BKS, it is also less likely that there

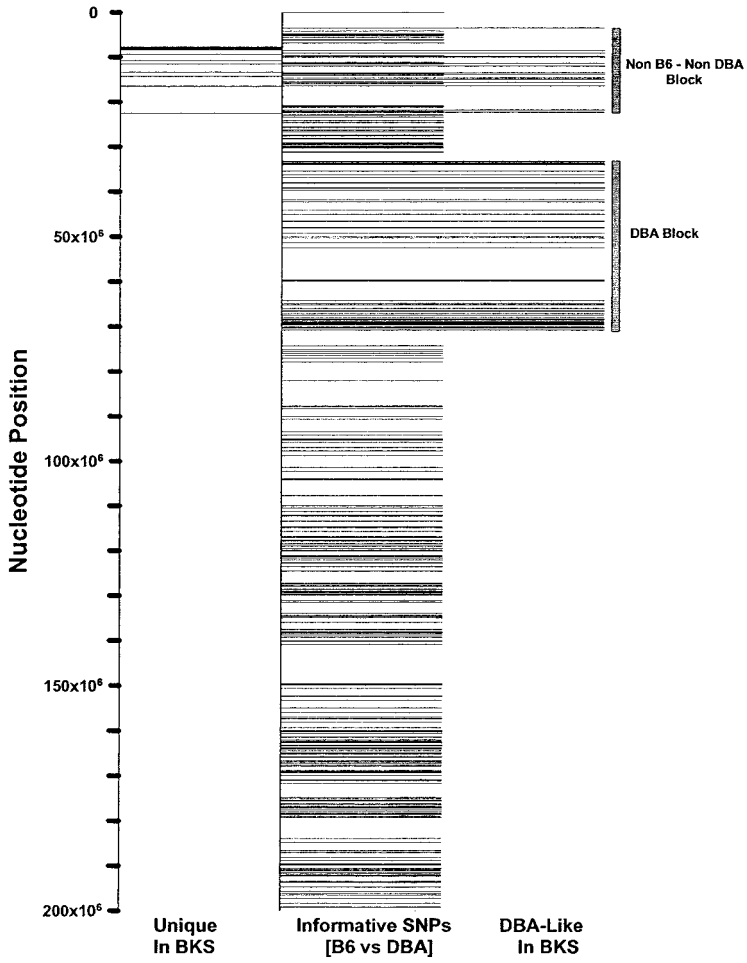

FIG. 1. Genetic origin of DNA blocks on chromosome 1 in BKS. Horizontal lines indicate the location (in bp) of SNPs along the chromosome (centromere at the *top*). The center set of lines indicates the positions of all 376 SNPs tested that were informative between B6 and DBA on chromosome 1. Lines that extend to the *right-hand side* indicate those 86 informative SNPs having the DBA allele. Lines in the *left-hand column* indicate 12 SNPs that are unique to BKS. That is, the SNP has the same allele for B6 and DBA, and yet, for BKS the SNP shows the alternate allele indicating a genetic origin other than B6 or DBA. For maps of the other chromosomes, see Fig. 1 in the online appendix.

TABLE 1
Location of DBA regions in BKS

Chromosome	Proximal SNP Identifier	Proximal SNP location	Distal SNP Identifier	Distal SNP location	Length
1	rs3707642	33,196,011	rs3659932	70,842,743	37,646,732
3	rs3675845	101,223,086	rs6212614	124,871,942	23,648,856
3	rs3679440	133,792,362	rs3676561	140,122,486	6,330,124
4	rs3674982	6,820,703	rs6196764	11,429,647	4,608,944
4	rs6340721	137,737,289	rs3023025	142,099,818	4,362,529*
5	rs3674947	3,948,243	rs3706626	23,176,602	19,228,359†
6	rs3709206	116,095,844	rs3659280	150,258,862	34,163,018
7	rs3726501	103,897,899	rs6216320	137,246,632	33,348,733
8	rs3657344	57,043,519	rs3678433	84,810,857	27,767,338
9	rs3690992	116,472,929	rs8254378	126,510,050	10,037,121
10	rs3664101	9,328,585	rs3667141	19,690,873	10,362,288
11	rs6263230	62,788,840	rs8241202	76,539,840	13,751,000
12	rs3688676	10,414,102	rs4229284	19,543,114	9,129,012
12	rs3023209	32,074,405	rs3699539	84,067,744	51,993,339
13	rs6209128	52,537,356	rs6391937	57,247,520	4,710,164
14	rs3710916	3,304,248	rs6396413	27,667,687	24,363,439
14	rs3707560	45,838,734	rs3700859	56,390,403	10,551,669
14	rs3722416	109,136,607	rs6384035	114,245,366	5,108,759
16	rs4153455	5,150,049	rs4158669	7,343,699	2,193,650‡
17	rs6293022	75,019	rs3703891	51,996,460	51,921,441
18	rs3724798	16,988,430	rs6167189	25,484,137	8,495,707
19	rs3681194	18,417,981	rs3669236	25,953,209	7,535,228
X	rs8276255	39,431,783	rs3693969	155,140,520	115,708,737
				Total length	516,966,187
				Percent of genome (% of 2.6×10^9 bp)	19.9%

Uninterupted blocks of DBA alleles >3 Mb in length in BKS. SNP Identifier uses the standard "rs" (refSNP) prefix for SNP accession numbers from the public database. All SNP locations given are from the National Center for Biotechnology Information build 32 of the mouse genome database. *Interupted by single unique BKS allele: rs224894. †Adjacent SNP with DBA allele (rs6152183) not included (separated from the rest of the block by >10 Mb). ‡DBA block <3 Mb.

is genetic variation in these regions contributing to differences in the BKS phenotype.

Application of these informative SNPs to identify regions of BKS derived from DBA is shown in Fig. 1 and Table 1. The middle of Fig. 1 shows the locations for all SNPs on chromosome 1 that are informative between B6 and DBA. Horizontal lines on the right side show regions of BKS that carry the DBA alleles for these informative SNPs. Thus, between 33.2 and 70.8 Mb on chromosome 1, all 70 SNPs showed a genotype consistent with this region of BKS deriving from DBA. By contrast, all informative SNPs between 70 Mb and the telomere at ~200 Mb carried the B6 allele. Table 1 shows the location of all regions in BKS carrying uninterrupted blocks of DBA alleles. These large DBA-derived regions are present on all chromosomes except 2 and 15 and in total represent ~20% of the genome. Interestingly, the largest DBA-derived block appears on the X chromosome. This chromosome appears to be >80% from DBA.

As previously shown, the BKS strain carries some genomic regions derived from strains other than B6 or DBA (2,8). We have termed these regions as "non-B6–non-DBA." The fine-scale SNP genotyping described here allows us to closely map the boundaries of these regions. For example in Fig. 1, on proximal chromosome 1 between 3.5 and 22.5 Mb, the SNP genotype rapidly alternated between B6 and DBA alleles. This could derive from a high degree of recombination between B6 and DBA alleles but more likely results from introgression of a DNA segment derived from an unrelated strain. Consistent with this, we observed a high frequency of SNPs interspersed in the same interval that, while not informative between B6 and DBA, nevertheless shows the alternate allele for the SNP in question. These SNPs are labeled "Unique in BKS" on the left side of Fig. 1. The distributions of DBA-like SNPs and BKS-unique SNPs for each chromosome are shown in Fig. 1 of the online appendix.

There are a total of 144 BKS-unique SNPs in the BKS genome. Throughout, the BKS-unique SNPs are most frequently associated with regions where the informative SNPs rapidly alternate between B6 and DBA alleles. In some cases, such as on proximal chromosome 9, the BKS-unique SNPs are concentrated in regions with a low density of SNPs informative between B6 and DBA. Both of these circumstances are consistent with introgression of genomic fragments from an unrelated strain. We have termed these general genomic regions as non-B6–non-DBA to indicate their unknown genetic origin. We use the term "BKS-unique" for individual SNPs whose origin is clearly from a strain other than B6 or DBA.

In no case do we see significant runs of BKS-unique SNPs that are interspersed with purely B6-like or DBA-like regions and well separated (>2–3 Mb) from a B6-DBA transition. The non-B6–non-DBA regions are summarized in Table 2. In total, these regions constitute 8–9% of the BKS genome.

In an attempt to identify the origin of the non-B6–non-DBA regions, we compared alleles for BKS-unique SNPs

TABLE 2
Location of non-B6–non-DBA regions in BKS

Chromosome	Proximal SNP Identifier	Proximal SNP location	Distal SNP Identifier	Distal SNP location	Length
1	rs6269442	3,493,529	rs3708897	22,515,307	19,021,778*
3	mCV24984125	87,702,042	rs3720007	98,222,780	10,520,738
3	rs3671858	141,318,437	rs3688244	149,350,137	8,031,700
4	rs3701432	29,303,790	rs3714457	39,214,252	9,910,462
4	rs3655671	85,993,583	rs6370137	129,311,429	43,317,846*
5	mCV23798232	110,754,738	rs3663155	126,368,833	15,614,095*
6	rs6239446	110,235,439	mCV23042866	115,587,647	5,352,208
8	rs3694208	15,420,149	rs3695658	19,267,741	3,847,592†
8	rs3720150	48,951,381	rs3659366	56,385,791	7,434,410
9	rs3722569	19,342,565	rs6385855	30,308,187	10,965,622
9	rs3655717	67,932,226	rs3676124	85,876,377	17,944,151*
11	rs3659787	4,406,639	rs3088673	8,400,103	3,993,464
11	rs3705582	17,502,025	rs3708339	21,756,038	4,254,013
11	rs8243055	51,879,197	rs3657266	58,827,272	6,948,075
11	rs3688710	80,124,992	rs3697441	89,124,377	8,999,385
13	rs3711004	111,614,693	mCV24397011	121,083,641	9,468,948*
15	rs6354330	77,377,173	rs3686133	104,469,414	27,092,241
16	mCV23083514	92,259,824	rs6254145	99,700,954	7,441,130
19	rs6254371	26,463,268	rs6224900	30,231,403	3,768,135
				Total length	223,925,993
				Percent of genome (% of 2.6×10^9 bp)	8.6%

Blocks of non-B6–non-DBA DNA >3 Mb in length and containing at least one BKS-unique SNP. Blocks generally contain no uninterrupted internal runs of B6 or DBA alleles >3 Mb. *Exceptions occur for some blocks where terminal SNP encloses a slightly larger block or in the case of a large gap in informative markers (*). †Contains no BKS-unique SNPs. SNP Identifier uses the standard "rs" (refSNP) prefix for SNP accession numbers from the public database and "mCV" (mouse Celera Variant) for SNPs developed by Celera.

among a panel of DNAs from 56 other inbred strains (Cervino et al., unpublished data). There was no strong match for this set of SNP alleles other than with BKS-*db* (Fig. 3 in the online appendix). Of the strains tested, the best match was with 129P3J and other strains of the 129 series. For these strains, 98 of 144 BKS-unique SNPs match BKS. However, while strain 129 matches this specific subset of SNPs in some regions, the overall match between 129 and BKS in the same regions is not high. Similarly, there is no strong match in this set of strains across the complete set of SNP alleles from the non-B6–non-DBA regions listed in Table 2 (data not shown). Petkov et al. (8) suggested that a mouse strain related to BTBR may have contributed these DNA regions, but we observe a close match to BTBR only on part of chromosome 4 (data not shown). Thus, the origin of the non-B6–non-DBA regions in BKS is undetermined. A match may be found in a larger panel of inbred strains, or the contaminant may derive from an outbred or wild strain.

Historically, the *db* mutation arose on the BKS genetic background, and most early phenotypic data for the diabetes mutation were derived from this strain. The original BKS-*db* strain is no longer available. The BKS-*db* breeding stock currently available from The Jackson Laboratory is heterozygous for the *db* mutation (because homozygotes are infertile), and as a convenience for breeding, the heterozygous *db* mutation at 46.7 cM is maintained in opposition to a coat color mutation in the closely linked misty (*m*) locus at 46.1 cM (Fig. 2). Moreover, the mutant *misty* gene derives from DBA and was first transferred to B6 (18) and subsequently bred into the BKS-*db* line (E. Leiter, personal communication). Thus, the current heterozygous breeding stock for BKS-*db* represents a second-generation congenic with the potential for contamination with genomic regions originating in both B6 and DBA. In congenic strains, the introgressed segment always carries additional sequence beyond markers used to select the congenic segment. In addition, because congenic construction begins with outcrossing between donor and recipient strains, additional regions of the donor genome beyond that of the intended congenic segment frequently become introgressed in the final congenic strain. We used the high-density SNP genotype data to address these issues in the homozygous BKS-*db* genome.

The leptin receptor gene occurs at ~100.6 Mb on chromosome 4 in a region where there are a number of non-B6–non-DBA SNPs in the BKS genome (Fig. 2). In fact, for BKS chromosome 4, a large block from ~85 to ~130 Mb shows the characteristic SNP pattern of non-B6–non-DBA DNA. This suggests that the original *db* mutation occurred on a region of the genome derived from a strain other than B6 or DBA. The SNPs proximal to and in the immediate vicinity of the leptin receptor gene are identical between BKS and BKS-*db*. However, distal to the leptin receptor gene, from ~102 to ~110 Mb, the original BKS sequences have been replaced with DNA carrying DBA alleles (indicated by the bar in Fig. 2). This block of DBA alleles in BKS-*db* is consistent with residual DBA contamination carried along with the *misty* congenic region derived from DBA. Sequence differences in this congenic contamination may impact the phenotype driven by the *db* mutation. Distal to this block, only two SNPs, at 111.0 and 111.6 Mb on chromosome 4, distinguish BKS and BKS-*db*. SNP analysis also reveals a short block of five SNPs on chromosome 15 from 93.9 to 95.8 Mb that were DBA-like in BKS but have become B6-like in BKS-*db* (Fig. 3). Thus,

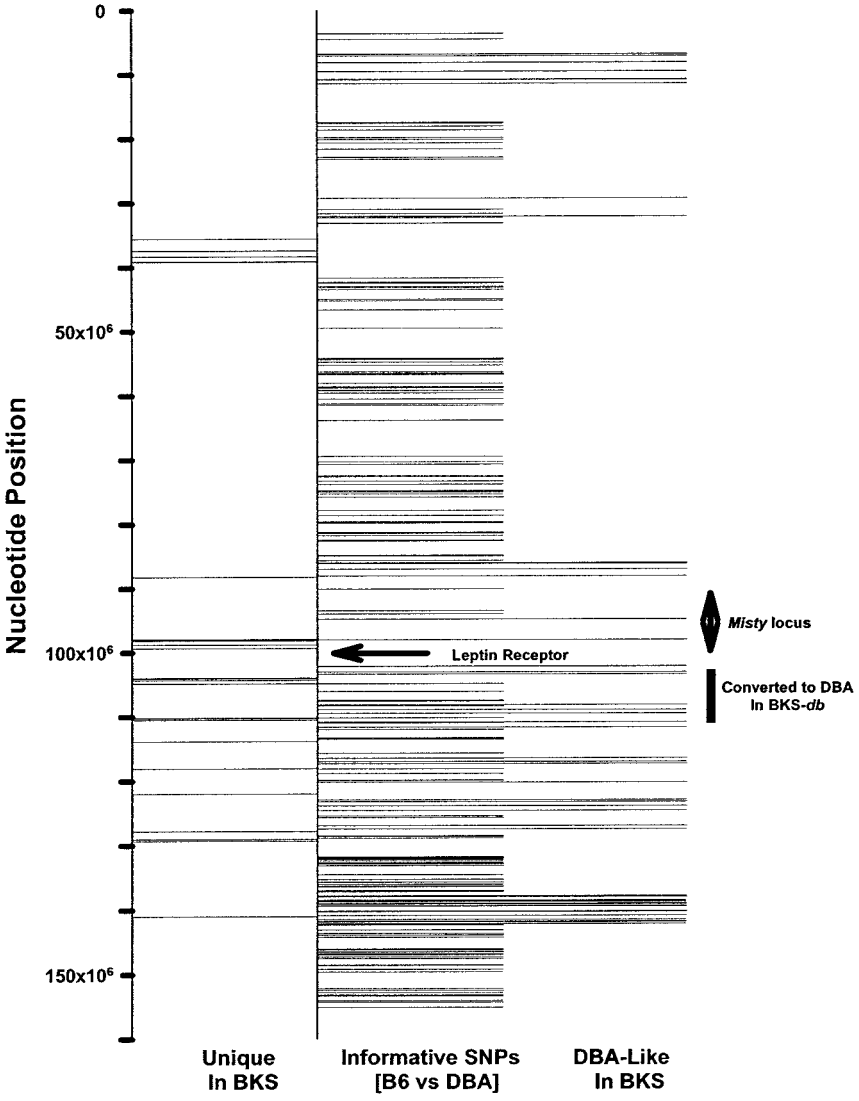

FIG. 2. Comparison of chromosome 4 in BKS and BKS-*db*. BKS-*db* arose from a spontaneous mutation in the leptin receptor gene at 100 Mb (←) in a large block of non-B6–non-DBA DNA that apparently extends from ~87 to 130 Mb. In BKS-*db*, a DNA block extending from 102 to 110 Mb has been replaced with DNA showing the DBA allele for all SNPs (indicated by a vertical bar on the *right*). This replacement is consistent with DNA carried along with and transferred from the *misty* congenic locus that is used in the BKS-*db* breeding stock to help identify offspring that do not carry the *db* mutation. (The *misty* gene is unidentified but based on its locus cM position relative to the leptin receptor, it should be located at ~99 Mb, as indicated by ↕.)

both possible contaminants from the *misty* congenic are now fixed in the BKS-*db* genome. Strikingly, no other contaminating regions associated with construction of the BKS-*db* congenic were detected by the fine-scale SNP analysis, although additional segments may have been transferred in either regions where the SNPs are not informative between B6 and DBA or smaller insertions between informative SNPs.

In the presence of the leptin receptor mutation, BKS exhibits a progression into full type 2 diabetes accompanied by extensive atherosclerosis. Strikingly, the diabetic and atherosclerotic phenotypes in BKS-*db* are significantly more severe than in B6-*db* with, for instance, an ~100-fold increase in lesion area. While rates of lesion progression and underlying differences in plaque histology and stability have not been widely explored, the basic effect is assumed to result from susceptibility loci conferred by the introgressed segments of DBA genome (4,19). To map the

		BKS Genotype	BKS-*db* Genotype
SNP Id	SNP Location		
		Proximal Flanking Marker*	
rs3717268	92,739,852	DBA	DBA
rs3722990	93,935,308	DBA	B6
rs3658059	94,404,332	DBA	B6
rs3666934	94,481,050	B6	B6
rs3667785	95,387,208	DBA	B6
rs4137261	95,771,810	DBA	B6
rs3685284	96,081,940	DBA	DBA
		Distal Flanking Marker*	

FIG. 3. B6 congenic region on BKS-*db* chromosome 15. *No difference between strains at flanking markers or any other proximal or distal marker.

TABLE 3
Cis-regulated genes in DBA blocks of BKS (chromosome 12)

Locus link/ Genbank ID	Gene symbol	Full gene name	Phenotype code*	Gene location	Morgan location	eQTL location	Logarithm of odds score	r^2
Chromosome 12:10.4–84.7Mb								
70619	5730510P18Rik	RIKEN cDNA 5730510P18 gene		13,254,130	0.0001	3,520,027	23.806	0.5169669
14245	Lpin1	lipin 1		16,930,916	0.0601	15,536,538	22.330	0.6045347
50496	E2f6	E2F transcription factor 6		17,204,840	0.0201	7,525,531	5.861	0.2192264
71853	Txndc7	thioredoxin domain containing 7		17,659,137	0.0401	11,531,035	12.708	0.4389293
BE992406		UI-M-BZ1-beg-g-08-0-UI.s1 NIH_BMAP_MHI2_S1 Mus musculus cDNA clone		19,125,145	0.0401	11,531,035	7.399	0.3078787
20135	**Rrm2**	**ribonucleotide reductase M2**	hfffa, hfldlv, hftcho, loghfffa, loghftg	**19,155,671**	**0.0801**	**19,542,042**	**13.928**	**0.4947893**
77480	C330002I19Rik	RIKEN cDNA C330002I19 gene		19,506,771	0.0401	11,531,035	13.122	0.4488543
72877	2900003A17Rik	RIKEN cDNA 2900003A17 gene		19,509,495	0.0401	11,531,035	32.468	0.7620641
15902	Idb2	inhibitor of DNA binding 2		19,542,044	0.0401	11,531,035	5.075	0.206266
246196	Zfp277	zinc finger protein 277		35,174,591	0.1401	31,558,554	28.458	0.6095555
BB347868	**Dock4**	**RIKEN cDNA clone B930067K06**	hfffa, loghfffa	**35,351,646**	**0.1652**	**36,577,818**	**8.753**	**0.2576814**
12632	**Cfl2**	**cofilin 2, muscle**	loghfhdl	**50,275,555**	**0.3652**	**61,347,573**	**10.339**	**0.4205977**
73385	1700047I17Rik	RIKEN cDNA 1700047I17 gene		50,466,305	0.2652	48,962,695	6.123	0.2173526
24067	Srp54	signal recognition particle 54		50,505,057	0.2252	44,008,745	7.549	0.2703286
15375	Foxa1	forkhead box A1		52,888,408	0.2452	46,485,720	9.522	0.4084019

*Phenotype code definition: hfffa, Plasma free fatty acid (high-fat diet); hfldlv, Plasma LDL and VLDL cholesterol level (high-fat diet); hftcho, Plasma total cholesterol level (high-fat-diet); logfffa, Log of plasma free fatty acid (high-fat diet); loghfhdl, Log of plasma HDL level (high-fat diet); loghftg, Log of plasma triglyceride level (high-fat diet). A complete Table 3 covering all chromosomes is found in the online appendix.

underlying loci, Mu et al. (4) performed a genetic cross between B6-*db* and BKS and then analyzed female F2 offspring carrying the homozygous leptin receptor mutation. They identified a significant QTL for atherosclerosis on chromosome 12 and suggestive QTLs for plasma glucose on chromosomes 8 and 17 and for plasma HDL cholesterol on chromosomes 5 and 17.

The atherosclerosis locus has been further localized to a subcentimorgan interval on proximal chromosome 12 (19), spanning 12.4–13.2 Mb. Comparison of this locus with the SNP genotype data presented here for BKS (Table 1) shows that it is contained within a short block of DBA-derived genomic sequence that extends from 10.4 to 19.5 Mb.

To further narrow the list of candidate genes, we have attempted to identify genes that are differentially regulated between B6 and DBA and whose expression is regulated in *cis*. We previously identified *cis*-acting eQTLs in an F2 intercross between strains DBA and B6 (12). The genes listed in Table 3 were selected on the following criteria: *1*) Logarithm of odds score for the eQTLs was greater than 4.3, *2*) physical map location of the gene is within 20 Mb of the eQTL peak, and *3*) physical map location of the gene is within a DBA block introgressed in BKS. The r^2 value indicates the percentage of expression variation for the gene attributable to the eQTL at this locus. (Online appendix Table 3 is a complete version of Table 3, which only shows chromosome 12.) That these genes are regulated in *cis* strongly suggests that local sequence variation within or near the gene affects expression differences observed among animals in the B6 × DBA cross.

Thus, the genes listed in Table 3 are likely candidates for any phenotypic differences introduced by the DBA blocks within BKS. This greatly focuses the attempt to identify candidate genes. For instance, there are two DBA-like regions on chromosome 12 in BKS covering a total of ~55 Mb. The National Center for Biotechnology Information MapViewer identifies ~430 genes in these combined regions. For the same regions, we identify only 18 *cis*-acting eQTLs meeting our criteria. To further focus on those genes specifically related to diabetes susceptibility, we measured correlation of each gene's expression with diabetes-related traits in the same animals. These traits included plasma levels of insulin and lipids (triglycerides, free fatty acids, and total, unesterified, and HDL cholesterol), measures of obesity (body weight, total fat pad weight, and weights of individual fat depots), and atherosclerosis (fatty streak lesions in the aortic arch). Genes whose expression showed significant correlation ($P < 0.05$) with one of these traits are shown in bold in Table 3, and the specific correlated phenotype is noted.

These *cis*-regulated genes will be useful in identifying candidates for genes influencing the BKS phenotypes. For instance, genetic crosses between B6 and BKS were analyzed by Coleman and Hummel (7) revealing non-Mendelian inheritance of diabetes susceptibility and suggesting that multiple loci contribute to the phenotype (7). Coleman and colleagues (20,21) identified variations in cytosolic malic enzyme activity between BKS and B6 and speculated that they may play a role in diabetes susceptibility. Malic enzyme catalyzes the $NADP^+$-dependent oxidative decarboxylation of malate to pyruvate and CO_2,

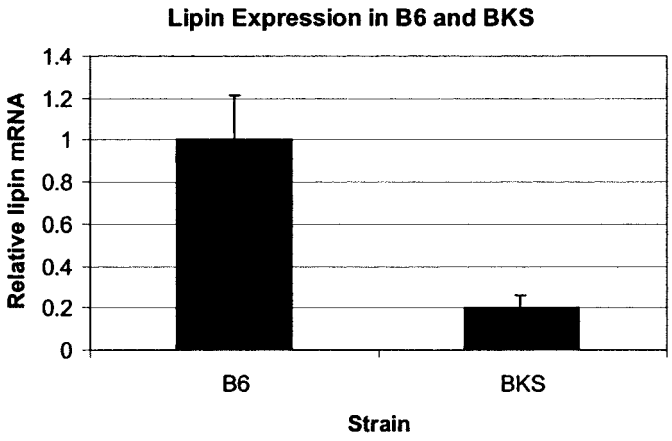

FIG. 4. Comparison of lipin expression in B6 and BKS mice. Lipin mRNA levels are significantly different ($P < 0.05$) in livers of male B6 and BKS mice at 1 month of age, consistent with differential regulation in *cis* deduced from expression analysis of the B6 × DBA cross (Table 3). Levels were measured by real-time quantitative RT-PCR as previously described (37) ($n = 5$ for each strain).

producing the NADPH required for de novo fatty acid synthesis. Coleman and Kuzava's (20) studies showed that the B6 strain carries the high-activity allele of the malic enzyme regulator (*Mod1r*), whereas BKS carries the low allele and that the *Mod1r* locus segregates in part with diabetes susceptibility. Coleman and collegues (20,21) mapped the *Mod1r* locus to proximal chromosome 12. Examination of the *cis*-regulated genes on proximal chromosome 12 (Table 3) shows a number of possible candidates for this susceptibility, most of which are uncharacterized genes. One particularly interesting candidate in this region is lipin, a gene known to modulate adiposity, insulin resistance, and atherosclerosis (22,23). The *cis* regulation of lipin, as detected in expression array data the B6 × DBA cross, is consistent with the fivefold differential expression of the mRNA that we observe between B6 and BKS by quantitative PCR (Fig. 4). Moreover, the lipodystrophy observed in mice carrying a natural mutation in lipin and the increased adiposity seen in muscle- and adipose-specific transgenics (24) strongly suggest that variation in mRNA levels has a corresponding impact on lipin activity. While lipin expression was not correlated with diabetes-related traits in the B6 × DBA cross, it is quite possible that the observed *cis*-regulated variation in lipin expression only becomes relevant in the context of obesity driven by leptin receptor deficiency or that the insulin resistance/diabetes phenotype results from interaction of lipin expression differences with other loci within the non-B6–non-DBA regions of BKS.

Of course, this use of expression array data will only detect candidate genes whose expression is regulated in *cis* in liver and not genes whose function is altered in the absence of a corresponding shift in expression. Moreover, these expression data are derived from a B6 × DBA cross and are only appropriate for identifying candidates for BKS phenotypic variation in regions of the BKS genome that derive from DBA.

DISCUSSION

The identification of genetic factors contributing to the diabetes phenotype in strain BKS is an important goal that may reveal new pathways involved in the disease. We now report the characterization of genomic differences between BKS and B6. We also survey *cis*-acting genetic variations between B6 and DBA at these loci. These results should be useful for the further mapping and, ultimately, positional cloning of genes underlying susceptibility to diabetes- and atherosclerosis-related traits.

The identification of *cis*-acting genetic variants in liver is an exciting strategy for narrowing the list of candidate genes in the BKS-*db* phenotype. However, while liver is likely to be a major organ impacting susceptibility to insulin resistance and diabetes in BKS, it is important to pursue the impact of genetic variations in other tissues as likely contributors to these phenotypes. Also, it will be important to verify the involvement of specific genomic regions of BKS using congenic strains. Such congenics will provide an ideal platform for subsequently investigating the effects of specific candidates and pursuing mechanistic and protein expression studies.

Numerous regions of BKS appear to have been derived from strain DBA (Table 1), and a smaller number are of unknown origin (Table 2). Lueders (3) provided the first evidence for DBA-like regions in BKS by studying strain-specific proviral loci. Naggert et al. (2) subsequently used 321 microsatellite markers polymorphic between B6 and DBA to genotype the BKS strain. They observed that several DBA clusters were observed in BKS, but because the markers were widely and unevenly spaced, some clusters were not identified and the remainder poorly defined (2). Recently, Petkov et al. (8) reported an analysis of the BKS genome using a panel of 1,636 SNPs. From this data, they identified 6 of the 19 non-B6–non-DBA blocks that we report here. Furthermore, the finding that DBA-*db* mice exhibit a diabetes-susceptible phenotype similar to BKS-*db* suggests that most or all diabetes susceptibility loci in BKS are of DBA origin (25).

The density of our SNP mapping is not sufficient to detect very short non-B6 regions in the BKS strain. For instance, although all SNPs tested on chromosome 2 showed the B6 allele, previous microsatellite mapping has shown a DBA allele for D2Mit1 at 3.80 Mb in BKS (2). This is consistent with the presence of a small fragment of DBA genome (<0.44 Mb) inserted between flanking SNP markers rs422932 at 3.45 Mb and rs3680350 at 3.88 Mb, both of which show the B6 allele in BKS. It is also possible that the D2Mit1 genotype in BKS recently evolved independently of DBA.

More recently, Mu et al. (4) examined the relationship between diabetes and atherosclerosis in the BKS-*db* strain. They constructed a cross between BKS on the *db* background and, following the feeding of an atherogenic cholic acid–containing diet, observed striking differences in lesion size, glucose levels, and HDL levels. Interestingly, no relationship was observed between glucose levels and lesion size. These studies (4) led to the identification of a major gene locus for atherosclerosis designated *Ath6* located on chromosome 12 in a DBA-like cluster. Interestingly, the peak of this QTL occurred at ~17 Mb, coincident with the *Lipn1* gene described above.

The same study identified suggestive QTLs for plasma glucose levels on chromosomes 8 and 17 and for plasma HDL cholesterol levels on chromosomes 5 and 17. The

plasma glucose QTLs were identified near D8Mit195 (83.4 Mb) and D17Mit24 (36.0 Mb), both corresponding to DBA-like regions in BKS. In both cases, the DBA-like region contains a number of strongly *cis*-regulated genes including likely regulatory factors such as a zinc finger protein on chromosome 8 (Zfp617) and the E2F transcription factor on chromosome 17, in addition to several uncharacterized genes that may underlie the QTLs.

These studies (4,21) did not reveal evidence for sex-linked inheritance of diabetes or atherosclerosis susceptibility, and therefore the extensive DBA-like regions on the X chromosome are unlikely to be involved in the diabetes susceptibility.

Attie and colleagues (26–30) carried out detailed studies of diabetes susceptibility, using genetic segregation, congenic strains, and expression array analysis, in a cross between B6 and BTBR on an obese (*ob*) background. They identified loci contributing to adiposity, lipid metabolism, and insulin/glucose levels differing between BTBR (diabetes susceptible) and B6 (diabetes resistant). However, none of these loci, with the exception of an obesity locus on chromosome 13, map within the DBA- or non-B6–non-DBA–derived chromosomal regions of BKS. Thus, although the phenotypic differences observed in the BTBR × B6 and BKS × B6 crosses are similar, separate genes likely determine these traits. Even so, there may be significant overlap in the overall pathways and mechanisms leading to diabetes susceptibility in these two models. In particular, Attie and colleagues have suggested that suppressed hepatic lipogenesis may be central to the relative diabetes susceptibility of BTBR-*ob*, and our ongoing studies show that BKS-*db* has strikingly suppressed hepatic lipogenesis relative to B6-*db* (data not shown).

As described above, in a cross between strains DBA and B6 (12), we used expression array data to identify eQTLs for several thousand genes. As shown in Table 3, about one-third of these eQTLs mapped over the corresponding structural gene and are therefore likely to be *cis* acting. We propose that this type of data will be very valuable in prioritizing candidate genes underlying the increased diabetes susceptibility observed in BKS-*db*. However, there are some hypotheses implicit in this proposal that merit validation. First, we assume that mRNA variation measured by array reflects real differences in mRNA levels, as might be determined by more quantitative methodology. The validity of mRNA measurements by array is usually checked only on a small scale by investigators pursuing specific candidate genes. Recently however, systematic assessment of this issue showed 70% correlation between Agilent array data and rigorous quantitative PCR for a panel of 100 mRNAs selected from a wide set of abundance classes (31,32). Moreover, when the same arrays are used in fluor-reversed pairs, as was done for the B6 × DBA cross, the correlation greatly improves (33), in our experience reaching 90%.

A second hypothesis is that mRNA variation mapping to the gene location (i.e., a *cis*-regulated gene) reflects a local sequence variation that differentially modulates the mRNA level in an allele-specific manner. Such *cis* regulation is what we would expect for genetic variation underlying a phenotypic QTL. To test this hypothesis, we used a direct test of *cis* regulation. Among the genes whose expression was predicted to be *cis* regulated in the B6 × DBA cross, we selected a set of 29 genes containing SNPs within the coding region. We then isolated liver mRNA from (B6 × DBA) F1 animals. Such animals are heterozygous at all loci, and their cells will contain a mixture of B6- and DBA-specific transcripts. Genes that are regulated in *trans* should have equal transcript levels from each allele. By contrast, genes that are differentially regulated in *cis* will show a corresponding enrichment in transcripts from one of the two alleles. Taking advantage of the SNPs in these genes, we used direct sequencing of RT-PCR products to determine the relative transcript levels from each allele. *Cis* regulation was confirmed in 19 of 29 genes (34).

To provide investigators with potential candidates underlying the DBA regions in BKS, we have catalogued the strongest *cis*-acting variations in Table 3. As discussed above, there is high correlation of this expression array data with more quantitative measures of mRNA levels, and, furthermore, independent *cis-trans* tests confirm that genes in Table 3 show strong differences in allele-specific transcript levels, strongly suggesting true *cis* regulation. However, investigators pursuing specific QTLs will need to validate their most promising candidate genes at the level of protein mass or activity. Because of significant posttranscriptional effects for many genes, overall mRNA levels from array data show only 40–75% correlation with the corresponding protein (35,36). Also, it should be noted that the eQTLs were not identified on *db* background, and thus there are probably differences in gene regulation between BKS-*db* and B6-*db* that were not observed in our cross. Nevertheless, most such *cis*-acting variations are likely to be at least partly independent of the genetic background or the state of obesity and diabetes of the mice.

ACKNOWLEDGMENTS

National Institutes of Health Grants HL28481, HL60030, and HL70526 supported this work.

REFERENCES

1. Slingsby JH, Hogarth MB, Simpson E, Walport MJ, Morley BJ: New microsatellite polymorphisms identified between C57BL/6, C57BL/10, and C57BL/KsJ inbred mouse strains. *Immunogenetics* 43:72–75, 1996
2. Naggert JK, Mu JL, Frankel W, Bailey DW, Paigen B: Genomic analysis of the C57BL/Ks mouse strain. *Mamm Genome* 6:131–133, 1995
3. Lueders KK: Differences in intracisternal A-particle and GLN proviral loci suggest a genetic contribution from a DBA/2-like strain in generation of the C57BL/Ks strain. *Mamm Genome* 6:134–136, 1995
4. Mu JL, Naggert JK, Svenson KL, Collin GB, Kim JH, McFarland C, Nishina PM, Levine DM, Williams KJ, Paigen B: Quantitative trait loci analysis for the differences in susceptibility to atherosclerosis and diabetes between inbred mouse strains C57BL/6J and C57BLKS/J. *J Lipid Res* 40:1328–1335, 1999
5. Hummel KP, Coleman DL, Lane PW: 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* 7:1–13, 1972
6. Hummel KP, Dickie MM, Coleman DL: Diabetes: a new mutation in the mouse. *Science* 153:1127–1128, 1966
7. Coleman DL, Hummel KP: Influence of genetic background on the expression of mutations at the diabetes locus in the mouse. II. Studies on background modifiers. In *Contemporary Topics in the Study of Diabetes and Metabolic Endocrinology.* Shafrir E, Ed. New York, Academic Press, 1975, p. 182–187
8. Petkov PM, Ding Y, Cassell MA, Zhang W, Wagner G, Sargent EE, Asquith S, Crew V, Johnson KA, Robinson P, Scott VE, Wiles MV: An efficient SNP system for mouse genome scanning and elucidating strain relationships. *Genome Res* 14:1806–1811, 2004
9. Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen

M, Steemers F, Butler SL, Deloukas P, Galver L, Hunt S, McBride C, Bibikova M, Rubano T, Chen J, Wickham E, Doucet D, Chang W, Campbell D, Zhang B, Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelpnagel J, Chee MS: Highly parallel SNP genotyping. *Cold Spr Harb Symp Quant Biol* 68:69–78, 2003
10. Gunderson KL, Kruglyak S, Graige MS, Garcia F, Kermani BG, Zhao C, Che D, Dickinson T, Wickham E, Bierle J, Doucet D, Milewski M, Yang R, Siegmund C, Haas J, Zhou L, Oliphant A, Fan JB, Barnard S, Chee MS: Decoding randomly ordered DNA arrays. *Genome Res* 14:870–877, 2004
11. Filliben JJ, Heckert A: Dataplot [Internet], 2004. National Institute of Standards and Technology. Available at http://www.itl.nist.gov/div898/software/dataplot.html/. Accessed 30 July 2004
12. 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
13. Colinayo VV, Qiao JH, Wang X, Krass KL, Schadt E, Lusis AJ, Drake TA: Genetic loci for diet-induced atherosclerotic lesions and plasma lipids in mice. *Mamm Genome* 14:464–471, 2003
14. Lincoln SE, Daly M, Lander ES: Mapping genes controlling quantitative traits with Mapmaker/QTL 1.1 [article online], 1993. Available at http://www.broad.mit.edu/genome_software/other/qtl.htm. Accessed 16 February 2005
15. Lincoln SE, Daly M, Lander ES: Constructing genetic maps with Mapmaker/Exp 3.0 [article online], 1993. Available at http://linkage.rockefeller.edu/soft/mapmaker/. Accessed 16 February 2005
16. Wang S, Basten C, Gaffney P, Zeng Z-B. QTL cartographer [article online], 2004. Available at http://statgen.ncsu.edu/qtlcart/index.php. Accessed 16 February 2005
17. Wiltshire T, Pletcher MT, Batalov S, Barnes SW, Tarantino LM, Cooke MP, Wu H, Smylie K, Santrosyan A, Copeland NG, Jenkins NA, Kalush F, Mural RJ, Glynne RJ, Kay SA, Adams MD, Fletcher CF: Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse. *Proc Natl Acad Sci U S A* 100:3380–3385, 2003
18. Chick WL, Lavine RL, Like AA: Studies in the diabetic mutant mouse. V. Glucose tolerance in mice homozygous and heterozygous for the diabetes (db) gene. *Diabetologia* 6:257–262, 1970
19. Purcell MK, Mu JL, Higgins DC, Elango R, Whitmore H, Harris S, Paigen B: Fine mapping of Ath6, a quantitative trait locus for atherosclerosis in mice. *Mamm Genome* 12:495–500, 2001
20. Coleman DL, Kuzava JE: Genetic regulation of malic enzyme activity in the mouse. *J Biol Chem* 266:21997–22002, 1991
21. Coleman DL: The influence of genetic background on the expression of mutations at the diabetes (db) locus in the mouse. VI: Hepatic malic enzyme activity is associated with diabetes severity. *Metabolism* 41:1134–1136, 1992
22. Peterfy M, Phan J, Xu P, Reue K: Lipodystrophy in the fld mouse results from mutation of a new gene encoding a nuclear protein, lipin. *Nat Genet* 27:121–124, 2001
23. Reue K, Xu P, Wang XP, Slavin BG: Adipose tissue deficiency, glucose intolerance, and increased atherosclerosis result from mutation in the mouse fatty liver dystrophy (fld) gene. *J Lipid Res* 41:1067–1076, 2000
24. Phan J, Reue K: Lipin, a lipodystrohy and obesity gene. *Cell Metabolism* 1:73–83, 2005
25. Leiter EH, Coleman DL, Hummel KP: The influence of genetic background on the expression of mutations at the diabetes locus in the mouse. III. Effect of H-2 haplotype and sex. *Diabetes* 30:1029–1034, 1981
26. 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
27. 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
28. Stoehr JP, Byers JE, Clee SM, Lan H, Boronenkov IV, Schueler KL, Yandell BS, Attie AD: Identification of major quantitative trait loci controlling body weight variation in *ob/ob* mice. *Diabetes* 53:245–249, 2003
29. Nadler ST, Stoehr JP, Schueler KL, Tanimoto G, Yandell BS, Attie AD: The expression of adipogenic genes is decreased in obesity and diabetes mellitus. *Proc Natl Acad Sci U S A* 97:11371–11376, 2000
30. Ranheim T, Dumke C, Schueler KL, Cartee GD, Attie AD: Interaction between BTBR and C57BL/6J genomes produces an insulin resistance syndrome in (BTBR x C57BL/6J) F1 mice. *Arterioscler Thromb Vasc Biol* 17:3286–3293, 1997
31. Lu J, Lal A, Merriman B, Nelson S, Riggins G: A comparison of gene expression profiles produced by SAGE, long SAGE, and oligonucleotide chips. *Genomics* 84:631–636, 2004
32. Nelson SF, Merriman B: A comparison of multiple microarray platforms for gene expression [article online], 2004. Available at http://www.ashg.org/cgi-bin/ashg03s/ashg03. Accessed 16 February 2005
33. He YD, Dai H, Schadt EE, Cavet G, Edwards SW, Stepaniants SB, Duenwald S, Kleinhanz R, Jones AR, Shoemaker DD, Stoughton RB: Microarray standard data set and figures of merit for comparing data processing methods and experiment designs. *Bioinformatics* 19:956–965, 2003
34. Doss S, Schadt EE, Drake TA, Lusis AJ: *Cis*-acting expression quantitative trait loci in mice. *Genome Res.* In press
35. Klausner RD, Harford JB: Cis-trans models for post-transcriptional gene regulation. *Science* 246:870–872, 1989
36. Tian Q, Stepaniants SB, Mao M, Weng L, Feetham MC, Doyle MJ, Yi EC, Dai H, Thorsson V, Eng J, Goodlett D, Berger JP, Gunter B, Linseley PS, Stoughton RB, Aebersold R, Collins SJ, Hanlon WA, Hood LE: Integrated genomic and proteomic analyses of gene expression in mammalian cells. *Mol Cell Proteomics* 3:960–969, 2004
37. Phan J, Peterfy M, Reue K: Lipin expression preceding peroxisome proliferator-activated receptor-gamma is critical for adipogenesis in vivo and in vitro. *J Biol Chem* 279:29558–29564, 2004