



A survey of allelic imbalance in F1 mice

Catarina D. Campbell, Andrew Kirby, James Nemesh, et al.

Genome Res. 2008 18: 555-563 originally published online February 6, 2008
Access the most recent version at doi:[10.1101/gr.068692.107](https://doi.org/10.1101/gr.068692.107)

References This article cites 32 articles, 13 of which can be accessed free at:
<http://genome.cshlp.org/content/18/4/555.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center, there is a white box with the words "LEARN MORE" inside. On the right, there is a woman wearing a red superhero mask and cape, and the logo for "CELLECTA" which consists of a green molecular structure.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Copyright © 2008, Cold Spring Harbor Laboratory Press

A survey of allelic imbalance in F1 mice

Catarina D. Campbell,^{1,2} Andrew Kirby,^{3,4} James Nemesh,^{1,3} Mark J. Daly,^{3,4}
and Joel N. Hirschhorn^{1,2,3,5}

¹Program in Genomics and Division of Endocrinology, Children's Hospital, Boston, Massachusetts 02115, USA; ²Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA; ³Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA; ⁴Center for Human Genetic Research and Department of Molecular Biology, Massachusetts General Hospital and Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA

There are widespread, genetically determined differences in gene expression. However, methods that compare transcript levels between individuals are subject to *trans*-acting effects and environmental differences. By looking at allele-specific expression in the F1 progeny of inbred mice, we can directly test for allelic imbalance (AI), which must be due to *cis*-acting variants in the parental strains. We tested over one hundred genes for AI between C57Bl/6J and A/J alleles in F1 mice, including a validation set of 23 genes enriched for *cis*-acting variants and a second set of 92 genes whose orthologs were previously examined for AI in humans. We assayed an average of two transcribed SNPs per gene in liver, spleen, and brain from three male and three female F1 mice. In the set of 92 genes, we observed 33 genes (36%) with significant AI including genes with AI that was specific to certain tissues or transcripts. We also observed extensive tissue-specific AI, with 11 out of 92 genes (12%) having differences in AI between tissues. Interestingly, several genes with alternate transcripts have transcript-specific AI. Finally, we observed that the presence of AI in human genes was correlated to the presence of AI in the mouse orthologs (one-tailed $P = 0.003$), suggesting that certain genes may be more tolerant of *cis*-acting variation across species.

[Supplemental material is available online at www.genome.org.]

Coding variation consisting of polymorphisms or mutations in the amino acid coding sequences of genes has been the traditional focus of genetic studies. Because the genetic code is known, it is straightforward to predict the location of a coding variant. Some examples of strong associations between coding variants and common diseases have been reported (Altshuler et al. 2000; Florez et al. 2004; Klein et al. 2005; Duerr et al. 2006; Sladek et al. 2007). Several noncoding variants have been also strongly associated to complex traits suggesting that they should not be ignored in genetic studies (Ueda et al. 2003; Graham et al. 2006; Grant et al. 2006; McPherson et al. 2007; Saxena et al. 2007).

Gene expression level as determined by microarrays has been shown to be a heritable trait in humans (Cheung et al. 2003). To determine the genetic variants that influence gene expression, the transcript level of each gene is considered a complex genetic trait, and linkage or association analysis is employed to map variants that contribute to transcript level variation (Cheung and Spielman 2002; Schadt et al. 2003). Previous studies in humans and model organisms have reported estimates that 20%–40% of genes with expression level differences are associated (or linked) to a nearby genetic marker (Brem et al. 2002; Yvert et al. 2003; Monks et al. 2004; Morley et al. 2004). However, guidelines based on distance do not definitively prove that a variant is *cis* or *trans* acting.

Measuring allele-specific expression levels allows one to definitively show if a gene is subject to *cis*-acting variants. Allelic imbalance (AI) is determined using the alleles of a polymorphism in the transcribed region of the gene (i.e., variants in the coding and untranslated regions) as markers for the transcripts from each

chromosome. Relative abundance of the alleles of this transcribed polymorphism can be determined in mRNA using many genotyping platforms (Knight 2004). If one allele of this transcribed polymorphism is overrepresented in the mRNA compared to the other allele, then that gene harbors at least one *cis*-acting variant. This *cis*-acting variant is likely not the transcribed polymorphism itself but might be located in a transcription factor binding site or region that determines transcript stability.

Several studies in humans and mice suggest that *cis*-acting variants are common in the mammalian genome. Genes expressed in human lymphoblast cells have been tested for AI, and 18%–53% of the genes tested showed allelic imbalance in at least one individual (Yan et al. 2002; Pastinen et al. 2004; Pant et al. 2006). In a study of 69 genes in F1 mice, four genes had strong evidence of AI (Cowles et al. 2002).

It is currently unknown whether AI is more often tissue-specific or a global phenomenon observed across tissues. In a study of human fetal liver and kidney, high correlation in AI between the two tissues was observed in the AI results, although only seven genes were tested for similarity between tissues (Lo et al. 2003). In contrast, out of four genes with AI in F1 mice, two showed tissue-specific AI (Cowles et al. 2002). These studies of tissue-specific AI have been small and more data are required to determine whether *cis*-acting variants commonly act in a tissue-specific manner.

We performed allelic imbalance assays in F1 mice. Because their parents are inbred, all F1 mice are heterozygous at single nucleotide polymorphisms (SNPs) where the parental strains have different alleles. Therefore, using existing databases, it is straightforward to identify transcribed SNPs that differ between the two parental strains, C57Bl/6J (B6) and A/J in this case. We have tested 115 genes for AI in C57Bl/6J × A/J F1 mice. We have tested genes previously analyzed for AI in mice (Cowles et al.

⁵Corresponding author.

E-mail joelh@broad.mit.edu; fax (617) 730-0253.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.068692.107>.

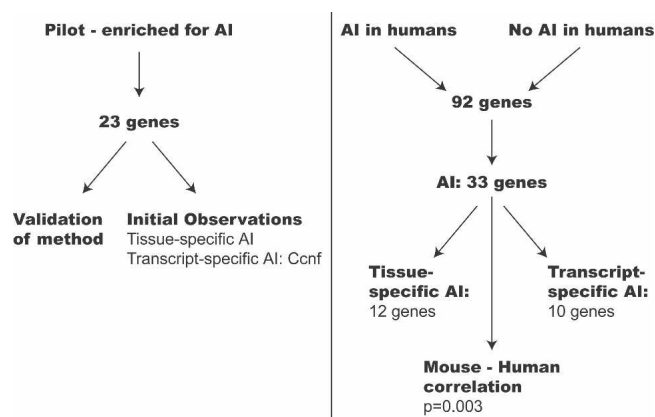


Figure 1. Summary of genes tested and major findings.

2002) and genes with expression differences between A/J and C57Bl/6J determined by microarray analysis. We have also analyzed a set of genes that had been tested for AI in human white blood cells (Pant et al. 2006). We have tested all of the genes in three organs: brain, liver, and spleen, and many genes (12%) show tissue-specific AI. Interestingly, we have observed AI differences between transcribed SNPs in the same genes. Many of these SNPs fall on reported alternate transcripts, confirming that different transcripts can be regulated by different *cis*-acting variants. Finally, we have found that genes that show AI in humans are more likely to show AI in mice.

Results

Validation of an assay for detecting allelic imbalance on a pilot set of genes

We used the Sequenom MassArray genotyping platform to detect allelic imbalance. We designed assays to transcribed SNPs in genes of interest where A/J and C57Bl/6J differed, and genotyped these assays on cDNA samples and genomic DNA for normalization as described in the Methods. An overview of our approach and results is presented in Figure 1. To validate our method for detecting allelic imbalance, we first tested a gene known to be imprinted, *Grb10*. Genetic imprinting can lead to AI, but the transcription imbalance of the two alleles is due to epigenetic influences inherited paternally or maternally and not dependent on genetic variants. In the case of imprinting, one would expect the opposite strain allele to be overrepresented in mice from a reciprocal cross where the strains of the parents were switched. Our control gene, *Grb10*, is expressed from the maternally inherited allele in all tissues except for brain, where it is expressed from the paternally inherited allele only (Hikichi et al. 2003). We observed overrepresentation of the B6 allele in liver and spleen from F1 mice, which have a B6 mother, and overrepresentation of the A/J allele in liver and spleen of RF1 mice which have an A/J mother. The reverse was observed in brain, as expected from the published data (Supplemental Fig. 1A). Therefore, from these data, we concluded that our method could detect allelic imbalance.

As part of our AI assay validation, we tested 23 genes that were likely enriched for the presence of AI. For these 23 genes, previous data existed to suggest the presence or absence of allelic imbalance (Supplemental Table 1). These included genes re-

ported to have *cis*-acting variation, genes reported to not have *cis*-acting differences (Cowles et al. 2002), and genes suggested by microarray data on prefrontal cortex tissue from inbred B6 and A/J mice to be subject to *cis*-acting regulation (A. Kirby and M.J. Daly, unpubl.). We tested for allelic imbalance by comparing the distribution of allelic ratios between F1 genomic DNA replicates and F1 cDNA data points as described in the Methods. For all three positive controls (*Ccnf*, *Il9r*, and *Hmgcr*), our data agree with the published literature (Cowles et al. 2002; Table 1). For the genes suggested by microarrays, six of the genes have AI in brain in the predicted direction (Table 1).

Testing additional genes for allelic imbalance

In addition to the 23 genes described above, we tested 92 more genes for the presence of AI (Supplemental Table 1). These genes are a subset of 1983 genes whose orthologs were studied for AI in 12 human lymphoblast cell lines; 731 genes that showed AI in at least one cell line and 652 genes that showed no AI across all cell lines (Pant et al. 2006). We chose this subset of 92 genes because these genes had at least two transcribed SNPs with different alleles in A/J and C57Bl/6J mice. We assayed an average of two SNPs per gene in these 92 genes (45 with *cis*-acting variants in humans, 47 without).

Ten of these 92 (11%) genes had significant evidence for AI in F1 mice (Table 2). Using standard statistical approaches (Benjamini and Hochberg 1995), we estimate the false discovery rate (FDR) for these 10 genes to be 0.05. We also estimated the rate at which we could replicate these results empirically. The results for nine of these genes were replicated (Bonferroni-corrected $P < 0.05$) or trended in the expected direction (nominal one-tailed $P < 0.05$) in mice from a reciprocal F1 cross (RF1 mice). We therefore estimate a nonreplication rate of 10% (one of 10 genes) for this analysis (see Methods). None of the genes tested showed evidence of genetic imprinting in the RF1 analysis (Supplemental Table 2). We also examined the relationship between statistical significance of AI and magnitude of AI. As expected, we observe correlation ($R^2 = 0.42$ – 0.48) between these metrics for all three tissues tested (Fig. 2).

Gender-specific allelic imbalance

We tested for AI in male and female mice separately. Specifically, we searched for genes where F1 mice of one gender showed significant AI but the other gender did not. None of the genes we examined met the criteria for gender-specific AI (Supplemental Table 3), suggesting that *cis*-acting variants with gender-specific effects are not common.

Tissue-specific allelic imbalance

In the pilot study of 23 genes, we observed that several of the genes had tissue-specific AI (see Methods for criteria used to declare tissue specificity). For *Ccnf*, *Ect2*, *Tap1*, and *Ina*, evidence of AI was observed in one or two tissues, and no evidence of AI was observed in the remaining tissue(s). For *Ugt8a* and *Gfra2*, we observed evidence for AI favoring opposite alleles in different tissues. In spleen, there was significant overexpression of the A/J allele of *Ugt8a*, but in liver the B6 allele of this gene was overexpressed (Table 1). Motivated by the surprisingly high frequency of tissue-specific AI in the pilot gene set, we examined the set of 92 additional genes for tissue-specific AI. We found that 13 genes out of 92 (14%) had significant AI in at least one tissue and either

Table 1. Results from pilot study

Gene	Previous Result	Brain		Liver		Spleen	
		<i>P</i>	Ratio ^a	<i>P</i>	Ratio ^a	<i>P</i>	Ratio ^a
<i>Ccnf</i>	B6 liver^b	1.08×10^{-4}	0.53	6.29×10^{-33}	0.67	0.858	0.50
<i>Ect2</i>	No AI ^b	1.47×10^{-5}	0.58	0.96	0.58	0.147	0.53
<i>Hmgcr</i>	B6 liver^b	1.79×10^{-6}	0.54	5.88×10^{-12}	0.57	5.46×10^{-7}	0.55
<i>Ii9r</i>	B6 spleen^b	ND	ND	ND	ND	2.26×10^{-9}	0.87
<i>Ptger1</i>	No AI ^b	0.849	0.54	0.513	0.53	0.99	0.52
<i>Slc1a2</i>	No AI ^b	0.300	0.45	ND	ND	0.016	0.33
<i>Tap1</i>	No AI ^b	0.842	0.51	3.02×10^{-10}	0.59	0.001	0.54
<i>Ugt8a</i>	No AI ^b	1.54×10^{-14}	0.35	4.49×10^{-9}	0.69	7.95×10^{-17}	0.05
<i>Apex1</i>	AJ PFC^c	0.007	0.47	0.013	0.47	0.974	0.50
<i>Atp8a1</i>	B6 PFC ^c	0.871	0.50	0.673	0.48	0.919	0.50
<i>Dctn6</i>	B6 PFC ^c	0.581	0.49	0.003	0.55	8.42×10^{-7}	0.58
<i>Dnajb1</i>	B6 PFC ^c	0.077	0.52	0.708	0.50	0.673	0.51
<i>Elf2</i>	B6 PFC ^c	0.529	0.49	0.673	0.49	0.081	0.48
<i>Epb4.9</i>	AJ PFC ^c	0.036	0.47	0.353	0.47	0.392	0.51
<i>Epha4</i>	B6 PFC ^c	0.952	0.50	ND	ND	ND	ND
<i>Gfer</i>	AJ PFC ^c	0.156	0.47	0.001	0.57	0.012	0.56
<i>Gfra2</i>	B6 PFC^c	0.001	0.53	1.74×10^{-4}	0.36	0.004	0.49
<i>Impg1</i>	B6 PFC^c	2.68×10^{-5}	0.75	ND	ND	0.002	0.64
<i>Ina</i>	B6 PFC ^c	0.796	0.51	1.55×10^{-4}	0.74	0.041	0.83
<i>Junb</i>	B6 PFC^c	0.003	0.55	0.012	0.53	0.012	0.55
<i>Nme3</i>	AJ PFC ^c	0.117	0.47	ND	ND	0.024	0.47
<i>Scn8a</i>	B6 PFC^c	6.34×10^{-4}	0.53	0.025	0.54	5.14×10^{-4}	0.54
<i>Tsc2</i>	B6 PFC^c	0.001	0.55	2.60×10^{-4}	0.55	0.001	0.55

The results for each gene are given by tissue with the nominal two-tailed *P*-value for the Wilcoxon rank sum test of cDNA data points for that tissue versus genomic DNA data points for the gene. Results in bold represent replication of previous results.

^aRatio = B6 SNR/(B6 SNR + AJ SNR), where SNR is signal-to-noise ratio (see Methods).

^bGenes with previously published AI results (Cowles et al. 2002).

^cGenes suggested by microarrays on the prefrontal cortex of B6 and AJ mice.

no AI (10 genes) or AI favoring the opposite allele (seen in three genes: *Emilin2*, *Klhdc5*, and *Synj2*) in at least one of the remaining two tissues (Fig. 3; Table 2). We estimated the FDR for this analysis to be 0.05 (Benjamini and Hochberg 1995), and also empirically determined the rate of nonreplication in RF1 mice. We strongly replicated the tissue-specific AI results for 11 out of the 13 genes (Table 2), with an estimated nonreplication rate of 19% (see Methods).

There was partial overlap in the genes that show AI in each of the three tissues (Fig. 4). In total, 15 out of 92 genes had evidence of AI in brain (16%), 22 out of 92 genes in liver (24%), and 26 out of 92 in spleen (28%). Each tissue had at least one gene where AI was specific to that tissue, with one, six, and 11 genes having AI specific to brain, liver, and spleen, respectively.

For the remaining genes, the allelic imbalance results were generally consistent across at least two tissues (Fig. 4).

Transcript-specific allelic imbalance

For the original set of 23 genes, we attempted to test as many reported SNPs in each gene as possible. For most genes, the expression results are consistent across the gene for all SNPs tested. However, for *Ccnf*, three of the eight SNPs assayed showed overrepresentation of the A/J allele in spleen, while the remainder showed no AI (Supplemental Table 5). These three SNPs were located near the end of the 3' untranslated region (UTR) (Fig. 5). This result suggested that *Ccnf* transcripts of different length might be present and regulated differentially by a *cis*-acting vari-

Table 2. Genes with significant global AI

Gene	F1 <i>P</i>	Ratio ^a	RF1 <i>P</i>	Ratio ^a	Result
<i>Trpv1</i>	2.06×10^{-10}	0.76	2.54×10^{-7}	0.83	Replicated
<i>Slc7a7</i>	1.15×10^{-10}	0.41	3.11×10^{-7}	0.44	Replicated
<i>6720456H20Rik</i>	2.04×10^{-9}	0.60	5.33×10^{-8}	0.57	Replicated
<i>Nav1</i>	3.39×10^{-9}	0.57	2.58×10^{-7}	0.55	Replicated
<i>Stch</i>	1.04×10^{-8}	0.43	5.69×10^{-5}	0.45	Replicated
<i>Adrbk2</i>	5.79×10^{-7}	0.42	0.004	0.45	Trend
<i>Tchp</i>	5.67×10^{-6}	0.55	0.079	0.53	Trend
<i>Slc22a15</i>	8.31×10^{-5}	0.42	0.115	0.46	Not replicated
<i>Snx13</i>	1.00×10^{-4}	0.43	2.67×10^{-4}	0.43	Replicated
<i>Rab2b</i>	2.30×10^{-4}	0.44	1.05×10^{-4}	0.44	Replicated

Results from Wilcoxon rank sum tests between cDNA data points and genomic DNA replicates are presented for each gene. Nominal *P*-values are given for F1 and RF1 mice. The threshold of significance in the RF1 data was $P < 0.005$, which is corrected for 10 tests. One-tailed $P < 0.1$ in the RF1 mice was considered a nonsignificant trend.

^aRatio = B6 SNR/(B6 SNR + AJ SNR).

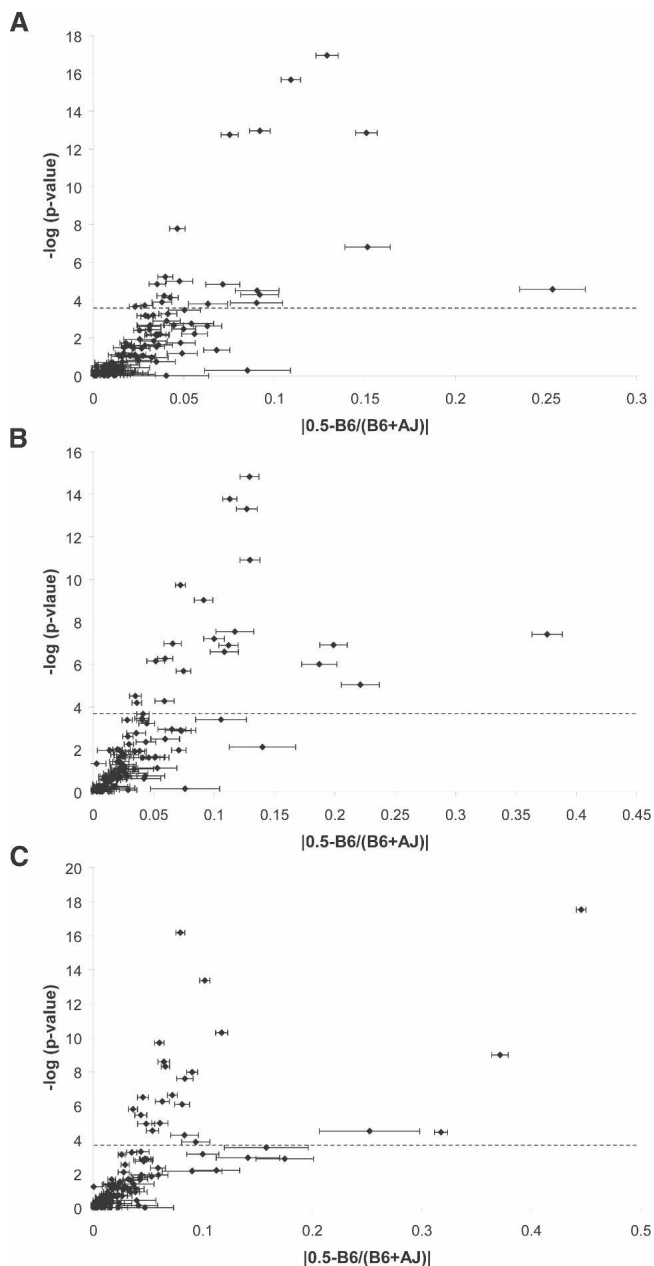


Figure 2. Magnitude of allelic imbalance. The statistical significance from Wilcoxon rank sum tests ($-\log$ of the P -value) is plotted against the magnitude of AI (absolute value of the difference between 0.5 [no AI] and the average ratio $B6/[A + B6]$) for each of the 115 genes in each tissue. The dotted line indicates an empirical $P = 0.05$. Error bars are the standard error of the mean; genes with large standard errors have SNP differences in AI. (A) brain; (B) liver; (C) spleen.

ant, similar to the case of *IRF5* in humans (Graham et al. 2006, 2007a,b). Data from the UCSC Genome Browser (genome.ucsc.edu) confirmed that this gene has multiple reported transcripts. In fact, the three SNPs in *Ccnf* with evidence of AI are located on the 3' UTR of one transcript but not the other (Fig. 5). We confirmed these results using quantitative PCR in A/J and B6 animals (data not shown).

Based on the results in *Ccnf*, we tested for transcript-specific AI in the set 92 additional genes (see Methods). AI was deter-

mined for each SNP tested in a given gene (Supplemental Table 5). Ten genes out of 92 (11%) show significant differences in AI between SNPs (Table 4, see below). The estimated FDR for this analysis is 0.05 (Benjamini and Hochberg 1995). All of these differences were observed in the RF1 mice, and we estimated a non-replication rate of 7% for these results (see Methods). Two genes had especially strong evidence of transcript-specific allelic imbalance. Specifically, *Brd8* has two reported transcripts with different 5' UTR lengths, and a SNP that lay only on the longer transcript showed a different pattern of AI from the remaining SNPs (Fig. 5; Supplemental Fig. 2). *Snx13* has a shorter alternate transcript with a different last exon and 3' UTR. A SNP in this alternate last exon showed a different pattern of AI from other SNPs in the gene (Fig. 5). We confirmed these transcript-specific observations by retesting the critical SNPs, by confirming these results in an analysis of mixed parental cDNA (Supplemental Table 5), and with qPCR on cDNA from the parental strains (data not shown).

Correlation between allelic imbalance in mice and humans

Because this set of 92 genes was drawn from genes that had been tested for AI in 12 human white blood cell samples (Pant et al. 2006), we were able to test whether the presence of AI in mice was correlated with the presence of AI in humans. Out of the 33 genes with strong evidence for AI in mice, 19 showed AI in humans and 14 did not, while of the 27 genes with no evidence of AI in mice, six showed AI in humans and 21 did not (one-tailed $P = 0.003$). These data suggest that genes that show AI in humans are more likely to have evidence of AI in mice.

Discussion

We have tested 115 genes for AI in C57Bl/6J \times A/J F1 mice. Each of these genes harbors at least one transcribed SNP that differs

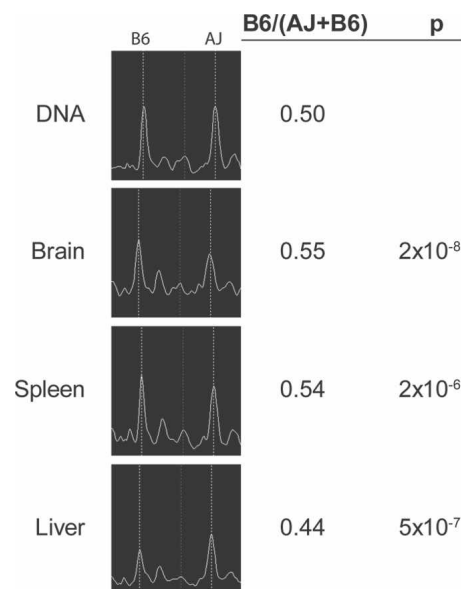


Figure 3. Example of tissue-specific allelic imbalance in *Emilin2*. Representative spectra of a transcribed SNP assayed in F1 genomic DNA and cDNA from three tissues. The peak locations for the A/J and B6 alleles are marked. For this gene, the B6 allele peak is higher than the A/J allele peak in brain and spleen, but the A/J allele peak is higher than the B6 peak in liver.

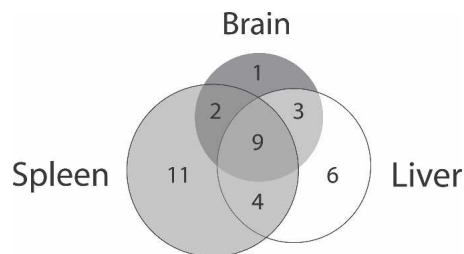


Figure 4. Overlap of genes with allelic imbalance between tissues. For this analysis, only the 33 genes with any evidence of AI were considered. Three genes are represented twice in the diagram because two of the tissues had evidence for allelic imbalance favoring alleles from opposite strains.

between C57Bl/6J (B6) and A/J, which we used as a marker for transcripts from the chromosomes inherited from the two strains. We did not observe substantial gender-specific AI, but tissue-specific AI was prevalent. For some genes with several transcripts, certain transcripts had different levels of AI than others. Interestingly, genes with AI in humans were more likely to show AI in mice, and genes with no evidence of AI in humans were less likely to show AI in mice. These results constitute one of the largest surveys of *cis*-acting regulatory variation in mice to date.

We initially studied two groups of genes that were likely enriched for AI: genes from Cowles et al. (2002) and genes suggested by microarray analysis of prefrontal cortex tissue of A/J and B6 mice (A. Kirby and M. Daly, unpubl.). In this pilot study, we attempted to replicate some of these earlier results. We were able to replicate tissue-specific AI observed in three genes from Cowles et al. (2002) (Table 1). However, for many of the genes reported to not have *cis*-acting differences in Cowles et al. (2002) we observed strong AI. The reason for this discrepancy is unclear, although it might be explained by technical differences between the two assays or that we tested multiple SNPs in most genes whereas the original study only tested a single SNP per gene. It is unlikely that the AI we observed in these genes are false positive results, since we reproduced the same result in RF1 mice (Supplemental Tables 2, 4). We observed significant AI in the direction predicted from microarray experiments for six out of 15 genes. We did not observe AI in all of these 15 genes, which may be explained by *trans*-acting factors in the microarray data, or that the microarrays were performed on prefrontal cortex tissue whereas we tested whole brain.

The 115 genes that we examined for the presence of allelic imbalance were not randomly selected. Twenty-three of these genes were highly enriched for genes with AI, and these genes were not included in our estimates of the prevalence of AI or tissue-specific AI. For about half of the other 92 genes in our study, there was prior evidence of AI in human white blood cells, and we acknowledge that inclusion of these genes could introduce a small bias in our estimates of the extent of AI in mice, given the correlation that we discovered between genes showing AI in mice and human. However, we have no reason to believe that this set of genes would be biased in estimating the prevalence of gender- or tissue-specific AI within those genes that show AI.

In the 92 genes, we do not observe any evidence for strong gender-specific AI. Gender-specific strain expression differences have been observed with microarrays (Yang et al. 2006), and we observe gender-specific expression differences between strains in seven genes using mixed A/J and B6 cDNA (Supplemental Table

3). These data suggest that gender-specific *trans*-acting variants and environmental influences are more common than *cis*-acting variants with effects in only one gender.

This study is the largest study to date of AI across different tissues. Data from smaller studies of AI in mice and humans have been inconclusive on the extent of tissue-specific AI (Cowles et al. 2002; Lo et al. 2003), although genes with tissue-specific expression patterns have been mapped to expression quantitative trait loci (eQTL) in mice (Yang et al. 2006). Interestingly, three genes in our study (*Synj2*, *Emilin2*, and *Klhdc5*) have highly significant AI favoring alleles from opposite strains in different tissues (Fig. 3; Table 3). One possible explanation for these data is the presence of multiple *cis*-acting variants including one or more variant(s) in tissue-specific regulatory elements that influence expression in only one tissue.

Interestingly, several genes in our study show evidence of *cis*-acting variation that only affects one transcript. Recently, a

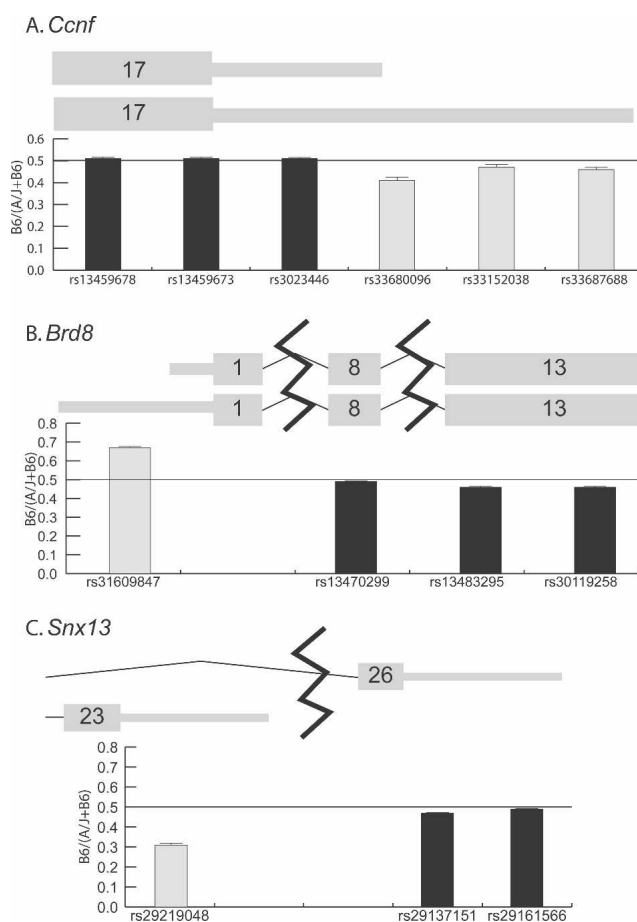


Figure 5. Examples of transcript-specific allelic imbalance. Representations of the exon and UTR structures are plotted with the thicker bars symbolizing coding sequence and the thinner bars symbolizing UTR sequence. Exon numbers are given in the bars for each exon. These diagrams are based on information from the UCSC Genome Browser (genome.ucsc.edu). Below the representation of each gene, a graph of $B6/(A/J + B6)$ is plotted for SNPs that fall within the exons shown. The horizontal line across each graph marks a 50:50 ratio of B6 to A/J alleles. (A) *Ccnf* has two transcripts that differ in 3' UTR length; data shown are from spleen. (B) *Brd8* has two transcripts that differ in 5' UTR length; data shown are the average of all tissues. (C) *Snx13* has two transcripts with different last exons and 3' UTRs; data shown are the average of all tissues.

Table 3. Genes with tissue-specific AI

Gene	Brain		Liver		Spleen		RF1 result
	Ratio ^a	P	Ratio ^a	P	Ratio ^a	P	
<i>Coro2a</i>	0.50	0.541	<u>0.39</u>	4.00×10^{-4}	0.82	3.69×10^{-5}	Replicated
<i>Elf3</i>	0.55	0.002	<u>0.48</u>	<u>0.011</u>	0.66	2.93×10^{-4}	Replicated
<i>Emilin2</i>	0.55	1.63×10^{-8}	<u>0.44</u>	5.37×10^{-7}	0.54	1.62×10^{-6}	Replicated
<i>Gnpat</i>	<u>0.50</u>	<u>1.000</u>	0.56	5.36×10^{-5}	0.52	0.03	Not replicated
<i>Klhdc5</i>	0.53	8.09×10^{-4}	<u>0.53</u>	5.19×10^{-5}	0.42	1.14×10^{-12}	Replicated
<i>Lnx2</i>	0.53	0.014	0.60	6.27×10^{-8}	0.50	<u>0.518</u>	Replicated
<i>Ly9</i>	0.59	3.16×10^{-5}	0.48	0.259	<u>0.47</u>	<u>0.021</u>	Replicated
<i>Mical2</i>	0.41	5.16×10^{-5}	0.46	0.178	0.41	1.39×10^{-4}	Not replicated
<i>Mmp9</i>	<u>0.48</u>	<u>0.578</u>	0.61	2.58×10^{-7}	0.57	2.41×10^{-7}	Replicated
<i>Rnf122</i>	<u>0.53</u>	<u>0.002</u>	0.48	0.038	0.41	1.07×10^{-8}	Replicated
<i>Sipa1l2</i>	<u>0.48</u>	<u>0.583</u>	0.37	1.49×10^{-15}	0.48	0.561	Replicated
<i>Slc22a15</i>	<u>0.49</u>	<u>0.292</u>	0.39	1.28×10^{-7}	0.38	5.14×10^{-11}	Replicated
<i>Synj2</i>	0.56	0.002	0.28	9.12×10^{-6}	<u>0.60</u>	<u>7.19×10^{-4}</u>	Replicated

Genes with at least one tissue with a Bonferroni-corrected $P < 0.05$ for AI (the most significant tissue result is in bold and italics) and another tissue with a nominal $P > 0.1$ or another tissue favoring the opposite allele (the nonsignificant result or the result favoring the opposite allele is underlined) in F1 mice. RF1 mice were used to validate these results, as described in the text. Nominal two-tailed P -values are given.

^aRatio = B6 SNR/(B6 SNR + AJ SNR).

similar result was described for *IRF5* in humans. In this gene, a SNP in the 3' UTR creates a new polyadenylation signal, leading to a more stable shorter transcript (Graham et al. 2006, 2007a,b). We have studied three mouse genes in detail that show differences in AI between SNPs along their transcripts (Fig. 5), and these genes all have multiple reported transcripts (genome.ucsc.edu). In these three genes, we observe different patterns of transcript-specific AI (Fig. 5). More generally, we observed significant, reproducible differences in AI between SNPs in the same gene for 10 out of 60 genes for which we assayed more than one SNP. These data suggest that this phenomenon is relatively common in the mouse genome. We cannot suggest a mechanism for any of these genes, but differences in transcript stability due to different polyadenylation site usage and differences in microRNA binding efficiency are among the possible explanations; SNPs could also affect the efficiency of transcript export from the nucleus, the binding of transcript-specific regulators, or the overall rates of transcription of each transcript. Our results emphasize the potential importance of measuring the level of each transcript for a given gene when trying to understand the impact of *cis*-acting variants.

This is the first study to compare the presence of allelic imbalance across species. There are several caveats to this experiment: first, the tissues tested in mouse and human were different; second, the number of tested genes was small, so the power to detect an association is low; and third, the number of human samples was small with limited power to detect AI. Despite these caveats, we did observe a nominally significant association between genes with AI in both species ($P = 0.003$). Purifying selection has recently been described for *cis*-eQTL in yeast (Ronald and Akey 2007), which lends support to our findings that certain genes may be more tolerant of regulatory variation.

In our survey of allelic imbalance in mice, we observed many genes with tissue-specific AI, and our data suggest that the choice of tissue for studies of the genetics of gene expression should be considered carefully. We particularly recommend the creation of a human tissue bank, so that studies of the genetics of gene expression could be performed across a wide range of tissue types. We have also observed that different transcripts in the same gene can have differing AI, suggesting that expression should ideally be assessed at the transcript level. Determining the

causal variants for the AI phenotypes observed would lead to new insights into the genetic control of gene expression.

Methods

Mice

Mice were bred and reared and organs were extracted by the Jackson Laboratory. Brains, livers, and spleens were obtained for three male and three female C57Bl/6J \times A/J F1 mice, three male and three female A/J \times C57Bl/6J F1 mice, three male and three female C57Bl/6J (B6) mice, and three male and three female A/J mice. Organs were flash frozen following extraction. All mice were 4–6 wk of age and were fed a standard chow diet.

DNA and RNA isolation

Genomic DNA was isolated from the F1, C57Bl/6J, and A/J mice using the DNeasy kit (Qiagen). To isolate RNA, frozen tissue was homogenized in Qiazol (Qiagen) using a mechanical rotor/stator homogenizer. Total RNA was purified using the RNeasy kit (Qiagen). mRNA was isolated from total RNA using the Oligotex kit (Qiagen). mRNA samples were treated with 10 units of DNaseI at 37°C for 30 min. Then, 2.5 mM EDTA was added to stabilize the RNA and the sample was incubated at 75°C for 10 min to denature the DNaseI.

cDNA synthesis

To synthesize cDNA, 100 ng of mRNA, 150 ng random hexamers, and 1 mM dNTPs were incubated at 65°C for 5 min and 4°C for 1 min. Next, 1 \times RT buffer, 5 mM MgCl₂, 10 mM DTT, 120 units of RNaseOut, and 600 units of Superscript III (Invitrogen) were added to make a 60 μ L reaction volume and the sample was incubated at 25°C for 10 min, 50°C for 50 min, and 85°C for 5 min. Controls lacking reverse transcriptase were included to indicate the presence of genomic DNA contamination. Synthesized cDNA samples were treated with 5 units of RNaseH at 37°C for 20 min.

Allelic Imbalance genotyping

All genotyping was performed using the mass spectrometry-based MassArray platform (Sequenom) (Gabriel et al. 2002; Altshuler et al. 2005). Primers and probes were designed using Spec-

troDesigner (Sequenom). Assays were multiplexed up to a maximum of five using the homologous Mass Extension (hME) protocol (Gabriel et al. 2002; Campbell et al. 2007). The iPLEX protocol was used to multiplex assays up to a maximum of 25 as follows. PCR was performed in 6 μ L with 10 ng of genomic or cDNA, 0.6 pmol of each primer, 0.5 mM dNTPs, and 0.1 units of Taq DNA polymerase (Qiagen) in 1.25 \times PCR buffer (Qiagen) and 1.625 mM MgCl₂. PCR conditions were 94°C for 15 min; 45 cycles of 94°C for 20 sec, 56°C for 20 sec, 72°C for 1 min; 72°C for 3 min. Extra dNTPs were inactivated using 0.3 U of shrimp alkaline phosphatase at 37°C for 20 min, and the enzyme was denatured at 85°C for 5 min. Single base extension was performed with 7.15–12.51 pmol of probe based on probe mass for each assay, 0.2 mM termination mix, 0.22 \times extension buffer, and 0.041 μ L extension enzyme (Sequenom). Single base extension conditions were 94°C for 30 sec; 40 cycles of 94°C for 5 sec, 5 cycles of 52°C for 5 sec and 80°C for 5 sec; 72°C for 3 min. The iPLEX and hME assays yielded similar AI results when tested on the same set of genes.

Mouse homologs of human genes in Pant et al. (2006)

Using the human gene symbols provided in Pant et al. (2006), we performed database searches to identify the mouse homologs. We used the homolog table from Mouse Genome Informatics (MGI; www.informatics.jax.org). For human genes that were not listed in this table, we used Gene and Homologene from the National Center for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov) and the Ensembl database (www.ensembl.org) to determine the mouse homologs (Supplemental Table 1).

SNPs

SNPs were obtained from Ensembl, which has curated all mouse SNPs in dbSNP as well as SNPs discovered by the Sanger Institute. SNPs were also obtained from a database at the Broad Institute maintained by the mouse HapMap project. To design the AI genotyping assays, sequences for transcribed SNPs were trimmed so that both PCR primers would be in the same exon as the SNP. These design constraints allow the same assay to be run on genomic DNA and cDNA. SNP sites were avoided in the design of PCR primers and single base extension probes.

Data normalization and quality control

Each AI assay was run on three cDNA replicates for each F1 mouse and each organ. Six replicates of F1 genomic DNA were also run for each assay. Signal-to-noise (SNR) estimation was done automatically in the software provided by Sequenom. We chose SNR (as opposed to other measures, such as peak heights) because in pilot experiments this was the most reliable indicator of the relative proportion of alleles. For each SNP, the two alleles are present in equal amounts in the genomic DNA sample, and the SNRs for the two alleles should also be equal. Therefore, we used the genomic DNA replicates to calculate a correction factor to account for any technical biases or unknown copy number variants leading to unequal SNRs between the C57Bl/6J and A/J alleles. Specifically, the difference in SNR between the C57Bl/6J allele and A/J allele in the genomic DNA data was divided by the A/J allele SNR, and the mean of this correction factor was calculated across the six DNA replicates. The A/J SNRs were multiplied by the correction factor, and this value was added to the original A/J SNR value to obtain a normalized A/J SNR. The final AI ratio was calculated as the B6 SNR divided by the sum of the B6 and normalized A/J SNRs. Equal representation of the two alleles yielded a ratio of 0.5. Values between 0.5 and 1 indicate that the

B6 allele is overrepresented and values between 0 and 0.5 indicate that the A/J allele is overrepresented.

Assays were required to pass several quality control (QC) measurements in order to be analyzed further. At least four of the six genomic DNA replicates had to be correctly called heterozygous by the Sequenom software indicating a high accuracy of genotyping, and the standard deviation of the AI ratio for the working genomic DNA replicates had to be <0.06 . For each gender and tissue group (e.g., livers of male mice or spleens of female mice), there were a maximum of nine data points (three mice times three replicates). The group was only analyzed for AI if at least five data points had data (raw B6 SNR + raw A/J SNR >15) and the standard deviation of the replicates was <0.1 . Mouse-to-mouse variability was tested for each group, and groups with significant mouse-to-mouse variability ($P < 0.005$) were excluded from further analyses. Each group was considered for QC individually, so that genes that were not expressed in one group would not lead to the failure of AI assays in the other groups.

It is unlikely that our AI results are due to random single-copy gene inactivation. We only analyzed data that were consistent across three F1 mice. In addition, for all genes with AI, we required that this result was reproduced in three RF1 mice from the reciprocal cross. Also, because we did not observe any reproducible gender-specific differences in AI, all AI results are consistent across six F1 mice (three male and three female) and six RF1 mice.

Statistical analysis

All statistical tests were coded and performed in R (www.r-project.org). For all tests, genomic DNA replicate data points were compared to cDNA data points using Wilcoxon rank sum tests. We observed that most of the variability in our assay was from experimental replicates and not mouse-to-mouse variability for data from the same tissue, gender, and SNP. Therefore, we used the experimental replicates as the cDNA data points. Genes were first evaluated for overall AI by performing Wilcoxon rank sum tests between all cDNA data points for the gene and all genomic DNA data points. A Bonferroni correction was applied to this data set for 92 tests. Genes were considered to show AI if the corrected $P < 0.05$.

There were three classes of analysis with multiple groups in each: gender analysis with male and female groups; tissue analysis with brain, liver, and spleen groups; and SNP analysis with groups for each transcribed SNP tested in a gene. The data in each class for a gene were evaluated for normality, and more deviation from normality was observed than expected by chance. Therefore, we have chosen to use nonparametric Wilcoxon rank sum tests in the analysis. We have performed the analysis using t -tests as well and observed similar results suggesting there is not a substantial loss of power when using the Wilcoxon rank sum tests (data not shown). In each class, two-tailed Wilcoxon rank sum tests were performed between the data points of each group and the genomic DNA data points to evaluate the presence of AI in that group. For example, to test for AI in the males for a gene, Wilcoxon tests between all cDNA data points from male mice for that gene and all genomic DNA data points for that gene were performed. In each class, the P -values obtained from the Wilcoxon tests were corrected for the number of tests performed using the Bonferroni method. For gender, the number of tests was two (one for each gender) times the number of genes, or 184 tests. For tissue, the number of tests was three (one for each organ) times the number of genes, or 276 tests. For SNPs, the number of tests was the number of SNPs in genes containing more than one SNP, which was equal to 169 tests. A gene was considered to have gender, organ, or SNP specific differences if

one of the groups showed AI (corrected $P < 0.05$) and another group did not (nominal $P > 0.1$) or two of the groups showed overexpression of opposite alleles. We used these criteria to avoid declaring differences in AI for genes where all groups in a class (e.g., all tissues) show a trend favoring the same allele, since results like these are likely driven by statistical fluctuation and not true biological differences.

An additional minimum threshold for the magnitude of AI was applied to the genes with statistically significant results as defined by the corrected P -values above. Genes were considered to show AI if the ratio of SNR for the B6 allele to the sum of SNRs for both alleles was ≤ 0.44 or ≥ 0.55 . These values were determined based on an empirical null distribution as follows. For each working assay with six genomic DNA data points (226 SNPs), we assigned three data points as genomic DNA control and the remaining three genomic DNA data points as “experimental” values. We then calculated the mean ratio for the “experimental” data points in each assay and examined the distribution of means across all the assays; 95% of the assays had means between 0.44 and 0.55, so we used these values as our thresholds for the magnitude of AI.

To validate the AI results and to test whether any of these results were due to genetic imprinting, we tested the same genes in the brain, liver, and spleen of reciprocal A/J \times C57Bl/6J F1 mice (RF1). These mice are genetically identical to the original F1 mice with the strains of the mother and father reversed. For global tests of AI, we compared all cDNA and genomic DNA values as was done in the F1 mice. To validate potential results from F1 mice that showed specificity in AI for gender, tissue, or SNPs, we created a list of the two most divergent groups for each potential instance of AI, and performed Wilcoxon rank sum tests between these same two groups in the RF1 data set. The P -values were Bonferroni corrected for the number of potential positive results in each class: 13 for tissue-specific AI, 10 for SNP-specific AI, and 10 for overall AI. RF1 data were considered to have replicated the F1 result if the Bonferroni-corrected P -value was < 0.05 in RF1 mice and the allele overrepresented in the RF1 data was the same as in the F1 mice.

To determine empirically a nonreplication rate, we not only calculated the rate of negative findings in the RF1 mice but also estimated the number of apparent replications that we expected to see by chance in the RF1 mice. To estimate this latter quantity for genes with global AI, the 10 genes with the most nonsignificant P -values in the F1 data set for the Wilcoxon rank sum test between all cDNA data points and genomic DNA data points were analyzed in the RF1 mice, to mimic our analysis of the 10 genes with putative global AI in the F1 mice. None of these 10

genes was a false positive in the RF1 data set (Bonferroni-corrected P -value < 0.05). Because we observed that one out of 10 genes with significant global AI in F1 mice did not replicate in RF1 mice, we estimated an empirical rate of nonreplication of 10%.

To determine a rate of experimental nonreplication for tissue- and SNP-specific AI, we performed a similar analysis, this time using genes that had no evidence for tissue or SNP-specific effects in the F1 mice. Genes were considered negative for tissue- or SNP-specific effects in F1 mice if the nominal P -values for AI were > 0.1 for all tissues (or all SNPs), suggesting no AI in the gene, or if all tissues (or SNPs) had a Bonferroni-corrected P -value < 0.05 favoring the same allele, suggesting consistent AI. Wilcoxon rank sum tests were performed for these genes in the RF1 data between the two most similar groups in the F1 data set, in other words, those showing the least evidence of specificity, and P -values were Bonferroni corrected. Genes with corrected P -values < 0.05 and magnitude of AI > 0.55 or < 0.44 in the RF1 data set were considered false-positive results. For tissue-specific AI, we tested 24 genes that had no evidence of tissue-specific AI in F1 mice, and one (4.2%) of these genes showed apparent evidence of tissue-specific AI in RF1 mice. From this result, we considered that 4.2% (0.42 genes) of the 11 genes with reproduced tissue-specific AI (Table 4) could represent false positives, despite the apparent replication. Including the two genes with apparent evidence of tissue-specific AI in F1 but no replication in the RF1 mice, we estimate that the rate of nonreplication is 2.42 genes (0.42 + 2) out of 13, or 19%. For SNP-specific AI, there were 14 genes with no evidence of SNP differences in AI in the F1 mice. Of these 14 genes, one (7.1%) had significant apparent SNP-specific differences in RF1 mice despite no evidence in the F1 mice. All 10 genes that were positive of SNP differences in AI were replicated in RF1 mice, so we estimate the rate of nonreplication for SNP-specific AI differences in F1 mice as 7.1%.

To compare genes between mice and humans, we considered any genes with reproduced tissue- or SNP-specific AI as well as genes with AI across all data points as showing AI in mice. Genes were considered to not show AI in mice if the nominal $P > 0.05$ for AI in both F1 and RF1 mice. Thirty-two genes met neither criteria and were not included in the analysis. Results were similar using a χ^2 test with one degree of freedom or a Fisher's exact test.

Mixed cDNA studies

To assess gene expression in the parental mouse strains using Sequenom AI assays, we assayed an equal mix of C57Bl/6J and A/J cDNA for each gender and organ. cDNA from three B6 and three

Table 4. Genes with significant SNP-specific AI

Gene	SNPs	Ratio 1 ^a	P_1	Ratio 2 ^a	P_2	RF1 result
<i>Adrbk2</i>	3	0.39	7.33×10^{-5}	0.47	0.152	Replicated
<i>Akap9</i>	5	0.43	8.62×10^{-6}	0.50	0.989	Replicated
<i>Brd8</i>	2	0.67	2.07×10^{-6}	0.40	8.4×10^{-5}	Replicated
<i>Dusp12</i>	2	0.45	2.30×10^{-4}	0.57	0.003	Replicated
<i>Elf3</i>	4	0.62	1.09×10^{-5}	0.49	0.490	Replicated
<i>Lasp1</i>	4	0.63	1.38×10^{-7}	0.49	0.149	Replicated
<i>Nav1</i>	4	0.60	8.71×10^{-8}	0.49	0.591	Replicated
<i>Nolc1</i>	2	0.31	8.81×10^{-7}	0.54	8.34×10^{-4}	Replicated
<i>Sipa1l2</i>	3	0.37	6.9×10^{-8}	0.50	0.679	Replicated
<i>Snx13</i>	3	0.31	1.51×10^{-6}	0.49	0.409	Replicated

Data for the most statistically significant SNP is presented as P_1 (all are Bonferroni-corrected $P < 0.05$). Data for the least significant SNP (or SNP favoring the allele from the opposite strain) are presented as P_2 . RF1 mice were used to validate these results as described in the text. Nominal two-tailed P -values are given.

^aRatio = B6 SNR/(B6 SNR + A/J SNR).

A/J mice was mixed in equal proportion to approximate a F1 mouse. The mixed cDNA for each gender and tissue was tested in the same AI assays described above with six replicates. To account for small differences in the amounts of B6 and A/J cDNA, all data points for the mixed cDNA from each group were averaged. Operating under the assumption that the mean should equal 0.5, we calculated the difference between 0.5 and the actual mean. Then, we added this value to all the data points in the group so the mean would be equal to 0.5. Analysis of the mixed cDNA sample then proceeded as described above.

Quantitative PCR

Quantitative PCR was performed using SYBR green (Applied Biosystems). Three to four replicates of cDNA from each mouse were tested, and three mice each of B6 and A/J were tested. In a 15 μ L reaction volume, 7.5 μ L of SYBR green master mix, 10 pmol of each PCR primer, and 5 ng of cDNA were added. Cycling conditions were 95°C for 10 min, 40 cycles of 95°C for 15 sec and 60°C for 1 min. PCR primers were the same as used in the Sequenom AI assays.

Acknowledgments

We thank members of the Hirschhorn laboratory for advice on this project, especially Guillaume Lettre for helpful discussions. We thank Tracey Petryshen for the prefrontal cortex gene expression data. All mouse work and organ extraction were done at the Jackson Laboratories in Bar Harbor, Maine. This work has been supported by the Smith Family Pinnacle Program Project grant from the American Diabetes Association.

References

- Altshuler, D., Hirschhorn, J.N., Klannemark, M., Lindgren, C.M., Vohl, M.C., Nemesh, J., Lane, C.R., Schaffner, S.F., Bolk, S., Brewer, C., et al. 2000. The common PPAR γ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.* **26**: 76–80.
- Altshuler, D., Brooks, L.D., Chakravarti, A., Collins, F.S., Daly, M.J., and Donnelly, P. 2005. A haplotype map of the human genome. *Nature* **437**: 1299–1320.
- Benjamini, Y. and Hochberg, Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**: 289–300.
- Brem, R.B., Yvert, G., Clinton, R., and Kruglyak, L. 2002. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**: 752–755.
- Campbell, C.D., Lyon, H.N., Nemesh, J., Drake, J.A., Tuomi, T., Gaudet, D., Zhu, X., Cooper, R.S., Ardlie, K.G., Groop, L.C., et al. 2007. Association studies of BMI and type 2 diabetes in the neuropeptide Y pathway: A possible role for NPY2R as a candidate gene for type 2 diabetes in men. *Diabetes* **56**: 1460–1467.
- Cheung, V.G. and Spielman, R.S. 2002. The genetics of variation in gene expression. *Nat. Genet. (Suppl.)* **32**: S22–S25.
- Cheung, V.G., Conlin, L.K., Weber, T.M., Arcaro, M., Jen, K.Y., Morley, M., and Spielman, R.S. 2003. Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat. Genet.* **33**: 422–425.
- Cowles, C.R., Hirschhorn, J.N., Altshuler, D., and Lander, E.S. 2002. Detection of regulatory variation in mouse genes. *Nat. Genet.* **32**: 432–437.
- Duerr, R.H., Taylor, K.D., Brant, S.R., Rioux, J.D., Silverberg, M.S., Daly, M.J., Steinhardt, A.H., Abraham, C., Regueiro, M., Griffiths, A., et al. 2006. A genome-wide association study identifies *IL23R* as an inflammatory bowel disease gene. *Science* **314**: 1461–1463.
- Florez, J.C., Burt, N., de Bakker, P.I., Almgren, P., Tuomi, T., Holmkvist, J., Gaudet, D., Hudson, T.J., Schaffner, S.F., Daly, M.J., et al. 2004. Haplotype structure and genotype-phenotype correlations of the sulfonyleurea receptor and the islet ATP-sensitive potassium channel gene region. *Diabetes* **53**: 1360–1368.
- Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., Defelice, M., Lochner, A., Faggart, M., et al. 2002. The structure of haplotype blocks in the human genome. *Science* **296**: 2225–2229.
- Graham, R.R., Kozyrev, S.V., Baechler, E.C., Reddy, M.V., Plenge, R.M., Bauer, J.W., Ortmann, W.A., Koeuth, T., Gonzalez Escribano, M.F., Pons-Estel, B., et al. 2006. A common haplotype of interferon regulatory factor 5 (*IRF5*) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nat. Genet.* **38**: 550–555.
- Graham, R.R., Kyogoku, C., Sigurdsson, S., Vlasova, I.A., Davies, L.R., Baechler, E.C., Plenge, R.M., Koeuth, T., Ortmann, W.A., Hom, G., et al. 2007a. Three functional variants of IFN regulatory factor 5 (*IRF5*) define risk and protective haplotypes for human lupus. *Proc. Natl. Acad. Sci.* **104**: 6758–6763.
- Graham, D.S., Manku, H., Wagner, S., Reid, J., Timms, K., Gutin, A., Lanchbury, J.S., and Vyse, T.J. 2007b. Association of *IRF5* in UK SLE families identifies a variant involved in polyadenylation. *Hum. Mol. Genet.* **16**: 579–591.
- Grant, S.F., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Manolescu, A., Sainz, J., Helgason, A., Stefansson, H., Emilsson, V., Helgadóttir, A., et al. 2006. Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nat. Genet.* **38**: 320–323.
- Hikichi, T., Kohda, T., Kaneko-Ishino, T., and Ishino, F. 2003. Imprinting regulation of the murine *Meg1/Grb10* and human *GRB10* genes; roles of brain-specific promoters and mouse-specific CTCF-binding sites. *Nucleic Acids Res.* **31**: 1398–1406.
- Klein, R.J., Zeiss, C., Chew, E.Y., Tsai, J.Y., Sackler, R.S., Haynes, C., Henning, A.K., SanGiovanni, J.P., Mane, S.M., Mayne, S.T., et al. 2005. Complement factor H polymorphism in age-related macular degeneration. *Science* **308**: 385–389.
- Knight, J.C. 2004. Allele-specific gene expression uncovered. *Trends Genet.* **20**: 113–116.
- Lo, H.S., Wang, Z., Hu, Y., Yang, H.H., Gere, S., Buetow, K.H., and Lee, M.P. 2003. Allelic variation in gene expression is common in the human genome. *Genome Res.* **13**: 1855–1862.
- McPherson, R., Pertsemidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D.R., Hinds, D.A., Pennacchio, L.A., Tybjaerg-Hansen, A., Folsom, A.R., et al. 2007. A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**: 1488–1491.
- Monks, S.A., Leonardson, A., Zhu, H., Cundiff, P., Pietrusiak, P., Edwards, S., Phillips, J.W., Sachs, A., and Schadt, E.E. 2004. Genetic inheritance of gene expression in human cell lines. *Am. J. Hum. Genet.* **75**: 1094–1105.
- Morley, M., Molony, C.M., Weber, T.M., Devlin, J.L., Ewens, K.G., Spielman, R.S., and Cheung, V.G. 2004. Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**: 743–747.
- Pant, P.V., Tao, H., Beilharz, E.J., Ballinger, D.G., Cox, D.R., and Frazer, K.A. 2006. Analysis of allelic differential expression in human white blood cells. *Genome Res.* **16**: 331–339.
- Pastinen, T., Sladek, R., Gurd, S., Sammak, A., Ge, B., Lepage, P., Lavergne, K., Villeneuve, A., Gaudin, T., Brandstrom, H., et al. 2004. A survey of genetic and epigenetic variation affecting human gene expression. *Physiol. Genomics* **16**: 184–193.
- Ronald, J. and Akey, J.M. 2007. The evolution of gene expression QTL in *Saccharomyces cerevisiae*. *PLoS One* **2**: e678. doi: 10.1371/journal.pone.0000678.
- Saxena, R., Voight, B.F., Lyssenko, V., Burt, N.P., de Bakker, P.I., Chen, H., Roix, J.J., Kathiresan, S., Hirschhorn, J.N., Daly, M.J., et al. 2007. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**: 1331–1336.
- Schadt, E.E., Monks, S.A., Drake, T.A., Lusk, A.J., Che, N., Colinaco, V., Ruff, T.G., Milligan, S.B., Lamb, J.R., Cavet, G., et al. 2003. Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**: 297–302.
- Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., Boutin, P., Vincent, D., Belisle, A., Hadjadj, S., et al. 2007. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**: 881–885.
- Ueda, H., Howson, J.M., Esposito, L., Heward, J., Snook, H., Chamberlain, G., Rainbow, D.B., Hunter, K.M., Smith, A.N., Di Genova, G., et al. 2003. Association of the T-cell regulatory gene *CTLA4* with susceptibility to autoimmune disease. *Nature* **423**: 506–511.
- Yan, H., Yuan, W., Velculescu, V.E., Vogelstein, B., and Kinzler, K.W. 2002. Allelic variation in human gene expression. *Science* **297**: 1143.
- Yang, X., Schadt, E.E., Wang, S., Wang, H., Arnold, A.P., Ingram-Drake, L., Drake, T.A., and Lusk, A.J. 2006. Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.* **16**: 995–1004.
- Yvert, G., Brem, R.B., Whittle, J., Akey, J.M., Foss, E., Smith, E.N., Mackelprang, R., and Kruglyak, L. 2003. *Trans*-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat. Genet.* **35**: 57–64.

Received June 29, 2007; accepted in revised form January 28, 2008.