Published in final edited form as:

Int Rev Neurobiol. 2014 ; 116: 195–231. doi:10.1016/B978-0-12-801105-8.00008-4.

Genetics of Gene Expression in CNS

Ashutosh K. Pandey and Robert W. Williams*

Department of Anatomy and Neurobiology, Center for Integrative and Translational Genomics,
University of Tennessee Health Science Center, Memphis, Tennessee, USA

Abstract

Transcriptome studies have revealed a surprisingly high level of variation among individuals in expression of key genes in the CNS under both normal and experimental conditions. Ten-fold variation is common, yet the specific causes and consequences of this variation are largely unknown. By combining classic gene mapping methods—family linkage studies and genome-wide association—with high-throughput genomics it is now possible to define quantitative trait loci (QTLs), single gene variants, and even single SNPs and indels that control gene expression in different brain regions and cells. This review considers some of the major technical and conceptual challenges in analyzing variation in expression in the CNS with a focus on mRNAs, rather than non-coding RNAs or proteins. At one level of analysis this work has been highly successful, and we finally have techniques that can be used to track down small numbers of loci that control expression in the CNS. But at a higher level of analysis, we still do not understand the genetic architecture of gene expression in brain, the consequences of expression QTLs (eQTLs) on protein levels or on cell function, or the combined impact of expression differences on behavior and disease risk. These important gaps are likely to be bridged over the next several decades using 1. much larger sample sizes, 2. more powerful RNA sequencing and proteomic methods, and 3. novel statistical and computational models to predict genome-to-phenome relations.

1. Introduction

For many years, gene mapping studies have focused on the identification of single gene variants and molecular causes of diseases ranging from albinism and phenylketonuria to neurodegenerative diseases such as Huntington's and Alzheimer's disease [1–4]. The same linkage mapping methods that have been used to track down the CAG trinucleotide repeat expansion that causes Huntington's disease [5] can now be used to study the causes of variation in levels of micro-traits, such as RNAs, metabolites, and proteins, in any tissue, organ, or cell. All that is required is a cohort of individuals and matched expression data for a specific brain region or cell type for each subject. A major goal of expression genetics research is to uncover primary and causal sequence variants that modulate expressions levels, but the long term focus is on the complex hierarchical networks that link genetic variation, through mRNA and protein levels, to higher order phenotypes that influence disease risk and progression. If we understand the networks of causal linkages between differences in expression and differences in CNS function, then it may become possible to

*rwilliams@uthsc.edu.

push just the right molecular buttons to prevent and cure many still intractable diseases of the brain.

Compared to a classic genetic analysis of a Mendelian trait such as Huntington's, there are two fundamental differences in mapping RNA or protein expression levels. First, the control of expression is usually genetically complex (polygenic) and large numbers of other genes and sequence variants (polymorphisms) can potentially influence expression of the target transcript or protein. For example a group of cooperating transcription factors may control expression of a key transmitter receptor or an ion channel. These effects give rise to so-called trans eQTLs (Figure 1A) that map far from the target gene itself—usually on different chromosomes. In contrast, expression of mRNAs may also be controlled by sequence variants that are in or very near to the parent gene itself (Figure 1B, C). For example, a polymorphism in a promoter, enhancer, splice acceptor site, or the 3' UTR of a gene may produce differences in transcriptional rates, mRNA stability, or ratios of alternative transcripts. When mapping the expression of mRNAs or proteins, this type of genetic “self-control” produces so-called cis-acting QTLs or cis eQTLs [6]. cis eQTLs are first-order local effects, whereas trans eQTLs are second-order distant effects. In this review we consider both the technical and conceptual utility of cis and trans eQTLs. In short, cis eQTLs can be used to evaluate the quality of expression data sets (more cis eQTLs is always better), and validated cis and trans eQTL can both be used as causal anchor points in genome-to-phenome studies [7].

The history

Since the introduction of proteomic and transcriptome methods in the mid-1990s, gene mapping methods have been applied to study progressively larger molecular data sets generated using segregating populations of F2 intercrosses, backcrosses, sets of recombinant inbred strains, genetic diversity panels, and families and cohorts of humans [8,9]. Damerval, Devienne, and colleagues were the first to apply high throughput methods to map what they called *protein quantity loci* in an F2 intercross of corn in 1994 [10]. Their groundbreaking study is still a model of sophisticated genetic and genomic analysis. In 2002 microarray methods were exploited by Brem and Kruglyak [11] to study gene expression in budding yeast and in 2003 Schadt and colleagues [6] published a remarkable study on the genetic control of mRNA levels in three tissue types from three species: ear leaves of an F3 intercross of corn, livers of an F2 intercross between two strains of mice (C57BL/6J and DBA/2J), and blood cells from four large Mormon families. These landmark studies introduced much of the vocabulary and many of the types of analyses that are still used a decade later. 2005 marked the first publication of a genetic analysis of expression in the CNS by our group. We used a second-generation Affymetrix array—the U74v2—to estimate expression of about 10,000 genes in whole brains of a set of 32 BXD-type recombinant inbred strains of mice. A decade later, these eQTL methods have been applied to study over 20 brain regions in mice, rats, and humans using both arrays and RNA-seq, and most of these large eQTL data sets are accessible on-line for reanalysis and meta-analysis at the GeneNetwork web site (www.genenetwork.org; see Williams and Mulligan [12] for a primer on using GeneNetwork).

How much variation is there in gene expression in brain?

Common and rare gene variants—SNPs, insertions, deletions and inversions—are a major source of phenotypic diversity and of variation in gene expression in wild-type populations, model organisms, and human cohorts [6,10,11,13–18]. Variation in gene expression levels can be high. For example, in hippocampus of normal young strains of mice the coefficient of variation (the standard deviation of strain means divided by the mean of all strains) averages about 7% (Figure 2) but the range is often 2-fold or more. A significant fraction of this variation is under genetic control. Heritability of gene expression data is a function of the genetic diversity of the cases, the genetic complexity of the phenotype, the stability of the environment, and technical error and confounds. Heritability estimates are rough benchmarks that will be depressed by low gene expression (high noise), signal dilution due to cellular heterogeneity, poor technique or specificity, and uncontrolled environmental factors. Conversely, heritability estimates will be inflated by poor experimental design (e.g., processing related individuals in single batches is a well-known statistical mistake), and by allele-specific hybridization or alignment artifacts.

Despite these problems, heritability estimates are used to gauge the likelihood of detecting one or more cis or trans eQTLs that modulate gene expression. It is not uncommon for variation in mRNA expression to have heritabilities in the range of 25 to 50% [19] and to be under relatively strong genetic control in the CNS. This reflects strong modulation by a sizable number of sequence variants in upstream genes, including transcription factors, RNA-binding proteins, transporters and genes and micro RNAs involved in degradation. Even house-keeping genes have not been spared, and genes such as *GAPDH*, *ACTB*, *SNRPD3*, *RAB7*, *PSMB2*, *GPII*, *REEP5*, *RAB7A* are surprisingly variable and heritable among individuals [20].

Brain gene expression studies—a summary

A growing number of eQTL studies have explored the genetics of expression in both whole brain and in specific brain regions of mice, rats, non-human primates, and humans over the last decade [21–28]. Virtually all of this work has been restricted to large and structurally heterogeneous nuclei and cortical regions. In this review we focus on mouse, and even more specifically on the BXD family of mice for which there are remarkably deep and systematic expression data. Large open-access data are available online (Table 1) for neocortex [29], basolateral amygdala [30], striatum [31], nucleus accumbens (Michael Miles and colleagues), hypothalamus [32], hippocampus [33], midbrain [34], ventral tegmental area (Michael Miles and colleagues), cerebellum (Williams and colleagues), and retina [35]. All of these data sets along with tools for gene mapping and eQTL analysis are accessible online at GeneNetwork (reviewed in Williams and Mulligan [12]). Given the high structural heterogeneity of the brain and the logistic difficulties of eQTL studies, there are still no data sets for many other key CNS regions including olfactory bulb, dorsal thalamus, globus pallidus, hindbrain, or spinal cord, or dorsal root ganglia.

What is clear from these initial studies is that the genetic control of gene expression in different CNS regions is highly variable. Even when using the same cases and methods, only a small fraction of cis and trans eQTLs are well conserved across brain regions. In part this

is due to differences in cellular demographics of brain regions, but it would not be surprising if even relatively homogeneous cell types in different regions (e.g., layer 5 projection neurons in different parts of neocortex) had variable eQTL patterns due to cell-extrinsic factors and axonal connectivity differences.

Missing pieces

There are still key missing pieces to the brain's gene expression puzzle. This should not be surprising given the difficulties and costs of eQTL studies of the CNS. Consider the short list that follows as a set of important and still open research areas.

1. **RNA-seq eQTL studies.** While RNA-seq technology has great promises [36] [37], the method has not been widely exploited yet for eQTL analysis [36,38]. The largest study in mouse that we know of for any CNS tissue is our own modest analysis of whole brains of ~30 genotypes of BXD strains [27,39]. This eQTL RNA-seq data set is accessible in GeneNetwork for analysis of 200,000 exons and 26,400 transcripts. The largest RNA-seq study of humans is the NIH Genotype-Tissue Expression (GTEx) program [40]. GTEx data sets for CNS regions are still small ($n < 30$ cases for most brain regions), but by 2016 there will be excellent data for more than ten regions for several hundred humans.
2. **eQTL studies of alternative splicing.** There are no comprehensive eQTL studies of RNA splice variants in the CNS using array technology or RNA-seq. We know that a majority of genes expressed in brain have multiple isoforms and that eQTL analysis imperfectly combine isoforms into one or two mean "gene level" estimates of expression. The standard protocol used to convert mRNA to complementary DNA relies on a T7 polymerase that is specific to the poly-A tail of the 3' UTR [41]. As a result the great majority of array data only measure expression of the last few coding exons and the 3' UTRs of mRNAs. New RNA amplification methods do not have this 3' bias and the latest generation of arrays—so-called exon arrays and splice-junction arrays—can provide estimates of expression over a 500-fold range for exons and splice-junction sites. It is ironic that just as arrays are reaching full maturity and sophistication, they are being pushed aside (prematurely in our view) by RNA-seq.
3. **Developmental eQTL studies.** There are virtually no developmental studies of genetic control of gene expression during development in any species or tissue type. We know of only two small studies in mouse. Glenn D. Rosen and colleagues analyzed eQTLs in the neocortex of the BXD family at three stages—postnatal days 3, 14, and young adult (unpublished), and Daniel Goldowitz and colleagues (personal communication) are studying expression in cerebellum at eight stages—from embryonic day E12 to young adult—across a subset of BXDs. Given the dynamics of gene expression during development, and the need to understand the coupling between expression, proliferation, differentiation, and cell death in brain, this is a potentially fascinating topic that warrants much more attention.
4. **Experimental eQTL studies.** There are only a handful of studies on changes in eQTLs in brain after experimental perturbations of any type. The reason is that

these studies are doubly hard since they require genetically matched case and control cohorts and exceedingly careful experimental design to avoid statistical confounds. Michael Miles, Lu Lu, and colleagues have studied the impact of ethanol treatment and stress (in isolation and in combination) on expression in prefrontal cortex, nucleus accumbens, ventral tegmental area, and hippocampus [42,43]. What will be required to make these types of experimental eQTL studies more practical is a significant reduction in cost of transcriptome data sets and more sophisticated, accessible, and faster statistical workflows that incorporate linear mixed models.

5. ***Heterogeneity of brain tissue.*** No one has yet attempted an eQTL study of a single type of CNS cell. A genetic dissection of genetic variability of Purkinje cells, dopaminergic neurons, or a subtype of primarily motor neuron would be extremely interesting and could reveal the extent to which cellular heterogeneity obscures eQTL patterns. Use of averaged expression over many cell types may dilute expression variation and obscure genuine eQTLs (more on this topic below). Given the rapid progress in single cell genomic methods [44,45] these critical studies will certainly be accomplished in the next decade, but getting down to this level will probably come at the cost of increased technical noise. Large sample sizes may be a necessity.
6. ***miRNA eQTL studies.*** Finally, no one has yet evaluated the extent, causes, and consequences of micro RNA (miRNA) expression variability at the population level. A large number of miRNAs are expressed in brain and serve as important regulators of gene expression [46,47]. There is a growing body of evidence demonstrating important roles of miRNAs in brain development [46,48,49]. A few eQTL studies have used RNA-seq to sequence small RNA molecules from lymphoblastoid cell lines and adipose tissues [50,51]. Parsons and colleagues used RT-PCR to quantify hippocampal expression of five mature miRNAs in the BXDs ($n = 24$) [52]. The largest study in mouse that we know of is our own RNA-seq analysis of hippocampal miRNA expression differences in 45 genotypes of BXD strains (Lu Lu and colleagues, unpublished). A systematic eQTL study ideally would involve joint analysis of miRNA and mRNA expression from matched biological samples. Such work would be extremely helpful in revealing the shared genetic control of miRNA and mRNA transcripts.

Genetic architecture of expression traits

One of the main surprises of the genetics of gene expression is that it has nearly the same level of complexity as higher order behavioral traits. Cis eQTLs represent one welcome exception to this complexity—they are relatively common, have strong effect sizes, and are easy to validate and interpret [53], albeit with some difficulties related to hybridization artifacts [7]. While the specific SNP or indel that causes a cis eQTL may not be known, there is a very strong prior probability that polymorphisms in or near the parent gene are responsible for almost all cis eQTLs, and this can be proven using allele-specific assays in reciprocal F1 hybrids (more on this below). In contrast, the mapping, analysis, and validation of trans eQTLs is far more interesting and complicated. A single mRNA can be

associated with several trans eQTLs. These associations define the core elements of molecular networks. But a side effect of the polygenic nature of trans eQTLs is that individual effects and the matched LOD scores of trans eQTLs are typically much smaller than those of cis eQTL [53]. Figure 3 plots the abundance of cis and trans eQTLs at different LOD (logarithm of odds) score. Trans eQTLs with smaller effect sizes and lower LOD scores are numerous (left side of Figure 3), but the ratio of cis to trans-eQTLs increases steeply with LOD score.

The naive hope that trans eQTLs would often turn out to be polymorphic transcription factors or RNA metabolism genes has not been borne out by a decade of research. In retrospect, this is perhaps not surprising, since expression of transcription factors, RNA binding proteins, and microRNAs will themselves be under intense genetic control leading to a regression of causality and an increase in complexity. It is also possible that the genetic complexity of trans eQTL effects is an artifact caused by the high cellular heterogeneity of brain regions. The problem may be analogous to trying to follow one conversation in a noisy restaurant with a single microphone placed high above the crowd. If cellular complexity explains the problem then it should be much more effective to dissect and make sense of patterns of eQTLs in relatively homogenous parts of the CNS such as the cerebellum (~90% of all cells are granule cells) or the dorsal striatum (~60% of all cells are medium spiny neurons) than in heterogeneous tissues such as whole brain, neocortex, or hippocampus. We evaluated the impact of cellular heterogeneity and possible signal dilution on the detection of cis eQTLs using the cerebellum as a test case (Figure 4). The cerebellum makes up 12% of the mouse brain—52 mg versus 430 mg—an 8-fold dilution. Cis eQTLs with strong effects in the cerebellum can be easily re-identified using comparable sample sizes and the same array type in whole brain data (compare gray cerebellum bars versus black whole brain bars). However, more than half of small and modest effect eQTLs in cerebellum do not have a large enough signal to be detected in whole brain (left-most bars in Figure 4).

Cellular heterogeneity can also reduce expression correlations between associated genes and transcripts. For example, *Fev* and *Slca64* are two genes with expression in serotonin neurons in midbrain with tightly correlated expression ($r = 0.88$). However, transcripts of these two genes have almost no correlation in whole brain data ($r = 0.15$). To the best of our knowledge there have been no systematic attempts to relate the complexity of cis or trans eQTLs to levels of cellular heterogeneity except in the hematopoietic stem cell lineage [54], but current data for the BXD strains certainly makes this a tractable problem.

This dilution effect does not imply that every cell type has to be isolated for eQTL analysis, but it does mean that the signal-to-noise ratio of mRNA measurements needs to be optimized for mapping. Resampling to reduce noise may often be more effective than the finest laser microdissection. For example in a recent study of the whole midbrain, we measured mRNA levels using four arrays for each genotype. By averaging across these biological replicates we were reduced measurement noise and mapped strong cis eQTLs that originate from the small population of serotonin neurons (Figure 5). Engrailed 1 (*En1*) is a gene with highly selective expression in a few thousand serotonin neurons, and despite ~500-fold dilution in midbrain, *En1* maps as a very strong cis eQTL (LOD ~17).

RNA-seq to the rescue?

Recent progress in high-throughput sequencing has substantially improved the assessment of individual variation at multiple levels including whole genome, transcriptome (RNA-seq), and even the metagenome. RNA-seq uses high-throughput sequencing to profile gene expression and potentially provides more accurate estimates of transcript abundance over a wider dynamic range than arrays. RNA-seq should eventually facilitate detection of eQTLs with small-effects and should also provide insight into the control of alternative splicing and polyadenylation, making it useful to study tissues such as brain [55–60]. Another remarkable feature of RNA-seq is the ability to assess genome-wide allele-specific expression by exploiting isogenic F1 hybrids [61–65].

RNA-seq data generation

Until recently generating high quality RNA-seq data for a large number of samples was not feasible due to technical complexity and cost. However, over the past 6 years RNA-seq is emerging as a viable alternative to exon-arrays for eQTL studies. An RNA-seq sample “library” can now be prepared and sequenced at a depth of 20–30 million reads for the same cost of an exon array (about \$400). Ten million RNA-seq reads have been shown to provide roughly similar dynamic range as arrays [66]. To prevent wastage of sequencing resources due to highly abundant ribosomal RNA (rRNA), RNA libraries are either selectively enriched for mRNAs using poly(A)+ selection methods or depleted of rRNA. Selective enrichment of mRNAs with poly(A) tails is done using poly-T oligo-attached magnetic beads. Alternatively, rRNA can be depleted through a hybridization approach. A comparison of these methods showed higher yield of exon reads from poly(A) enrichment (60% of total reads) compared to ribosomal RNA subtraction (~30% of total reads) [67]. The rRNA depleted libraries generate higher numbers of intron and intergenic reads (~25% and ~45% of total reads) compared to poly(A) methods (~15% and 23%). Although expression estimates from both methods were highly correlated, poly(A) method seems to be a more suitable choice for eQTL studies. Achieving higher numbers of reads in exons is critical to the detection of small expression differences and provides higher power to detect allele-specific expression difference using F1 hybrids. The downside of poly(A) enrichment is that it ignores a small number of mRNAs lacking poly(A) tail and most non-coding RNAs.

2. Genetic Resources for eQTL Analysis in Mice

Mapping eQTLs involves linkage analysis between variation in expression and genetic polymorphisms (markers) that segregate in a family, cohort, or population of individuals. F2 intercrosses, sets of recombinant inbred strains, and heterogeneous stock have been used to map eQTLs [68–73]. Cis eQTLs are relatively easily identifiable compared to trans eQTLs due to their strong effects and prior knowledge about their location. Mapping and narrowing down trans eQTLs to single genes remains challenging because trans eQTLs have small effects and require a large sample size for detection, but also because mapping precision of the currently available mouse crosses is poor (particularly in F2s), generally in the range of tens to hundreds of genes. In the next section, we consider advantages and disadvantages of different crosses currently available for eQTL mapping. Data sets generated from different crosses can now be easily combined using a number of statistical methods, and there are

good reasons to combine the best of each of the mapping resources described below in eQTL mapping.

Intercross progeny

Test cross progeny—either F2 intercrosses or backcrosses—are the traditional mapping population described in the experiments of Gregor Mendel. Generating an F2 intercross is a simple two generation affair. Two distinct strains (often inbred strains) are bred to produce the first filial (F1) generation. F1s are mated to generate a cohort of F2s—usually several hundred individuals. Alternatively, F1s can be backcrossed to either parent to generate a backcross. Meiotic recombination in the F1s produces genetically diverse F2 individuals that segregate for gene variants and heritable phenotypes. Lusi, Schadt, and colleagues have successfully exploited large mouse F2 populations to study eQTLs in the brain and other tissues [74,75] and many of their data sets are available in GeneNetwork. Each member of an intercross needs to be genotyped at 100 to 200 markers. Large sample sizes ($n > 100$) are often required to map eQTLs in F2s because recombination density per animal is low. For the same reason, F2 crosses often lack adequate positional precision. This makes narrowing down a trans eQTL to a single gene almost impossible.

Recombinant inbred strains

Recombinant inbred (RI) strains have been used widely for mapping of both Mendelian and quantitative traits and, for reasons described below, are advantageous for eQTL studies. RI strains are families of fully inbred strains that are produced by intercrossing two parental strains, followed by repeated sibling matings for at least 20 generations. Each RI strain represents a unique and fixed chromosomal mosaic of the parental genomes. Once an RI strain is fully inbred and genotyped, it can be used as an immortal and genetically defined resource. RI strains are ideal for developmental and experimental eQTL studies since the same genotype can be studied at many time points and under many conditions. They are also ideal for studies of gene-by-environmental interactions because cases and controls can be matched. Finally, in the context of noisy eQTL experiments, one can resample a given strain and brain regions (multiple independent biological replicates) to reduce technical and unintended experimental variability (as in Figure 5). But the most important advantage of RI strains is that legacy phenotypes and eQTL studies can be combined to assemble massive phenomes. The current champion in terms of phenome depth is the BXD family of RI strains—a group of individuals for whom there are now close to 100 independent eQTL studies, all of which are assembled within the GeneNetwork web site.

The historical disadvantage of RI strains was their limited numbers and modest power and precision of associated QTL studies. Throughout most of the 1990s there were fewer than 30 strains per family. Now however, three mouse RI families—BXD, LXS, and the Collaborative Cross, each have well over 60 strains [76,77]. The main disadvantage of RI strains is not QTL power or precision but steadily rising costs of acquisition and maintenance of large numbers of strains. It is now possible to achieve average eQTL precision of ± 1 –4 Mb across most of the genome with a set of only 60–80 RI strains (Figure 6), even without any replication within strain. Thus, RI strains together with whole genome sequence data of parental strains can be used to achieve single gene resolution.

The BXD family

The BXDs—a RI set made by crossing C57BL/6J (B) with DBA/2J (D)—is the largest and oldest RI set. They have been used to study complex traits since the mid 1970s and the genetics of gene expression since the early 2000s (www.genenetwork.org). In addition to the remarkably deep phenome data sets available for the BXDs a further advantage is that both parents have been fully sequenced [78,79]. A complete compendium of *B* versus *D* sequence variants is available online and can be used to track down causal SNPs, indels, and CNVs. It is possible to use reverse genetic methods with the BXDs and to look up those phenotypes that map to the location of a particular sequence variant [80]. The current BXD panel contains around 120 lines that are almost fully inbred and available from the Jackson Laboratory, and another set of 30–40 that are being inbred by Williams, Lu, and colleagues at UTHSC.

Heterogeneous stock

Heterogeneous stock (HS) mice [71] and rats [81] were created by repeated random mating of stock animals. Unlike RI strains that descend from two parental strains, some HS crosses have incorporated as many as eight founder strains. This adds a high level of genetic diversity to HS progeny. The high recombinant density of HS increases the resolution of QTL and eQTL mapping. Huang and colleagues used an HS cross to map eQTLs in several tissues with an average precision of 2.45 and 3.75 Mb for cis and trans eQTLs respectively [82]. However, eQTL mapping of HS progeny is not straightforward [37]. The family structure causes genotype correlations that can produce spurious eQTLs. As a result, sophisticated statistical approaches such as mixed model associations have been designed for QTL analysis in heterogeneous populations [15,82–84]. High level of genetic diversity may also cause spurious cis eQTLs due to hybridization artifacts (microarray) or allelic bias in aligning RNA-seq reads. The eQTL study mentioned above found a significant enrichment of SNPs in probes corresponding to large-effect cis eQTLs. HS also requires high-density genotyping and large sample sizes to map moderate effect eQTLs.

The Collaborative Cross

The Collaborative Cross (CC) is multi-parental RI set derived from eight genetically diverse strains [85] that combine features of an HS with those of a conventional RI family. The parents of the CC include five common inbred strains (A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, and NZO/HILtJ) and three wild-derived strains (CAST/EiJ, PWK/PhJ, and WSB/EiJ), and consequently has much higher genetic complexity than a normal RI panel. Like an HS panel, the CC can have high potential eQTL precision at a given sample size [86,87]. Aylor and colleagues used the CC to map cis eQTLs at a resolution of <1 Mb [87]. A 1 Mb interval in the CC will contain nearly 10 times as many sequence variants as the BXDs. In some cases this will be a major advantage, but in other cases, it will make it hard to find the causative SNPs. As of 2014, 70 CC are ready for distribution. This is a resource that is now ready for prime time.

3. Genetic Mapping Methods

Several statistical approaches have been developed for genome-wide linkage analysis of traditional phenotypes. The same approaches can be used to map eQTLs. These approaches range from single marker tests (*t*-test, ANOVA, and simple regression analysis) to multiple locus mapping methods. The only major difference is that eQTL studies involve tens of thousands of expression traits and require fast algorithms. Since an eQTL study tests for thousands of markers and many thousands of molecular traits, associations must be statistically adjusted to account for multiple testing at two levels of analysis [21].

Single marker test

As the name suggests, the single marker test considers an individual marker or SNP without regard to information about adjacent markers. Single marker tests can be as simple as *t* test between two sets of expression values, where each set represents expression values for a distinct genotype. Analysis of variance (ANOVA) can also be used. The advantage of ANOVA is that it can incorporate covariates such as sex and non-genetic variables such as environmental effects and technical error.

Interval mapping

Interval mapping is widely used for QTL mapping of F2 and RI crosses and human linkage analysis. Interval mapping interrogates region between the two adjacent marker loci to precisely determine the location of the QTL. It imputes the genotypes at intervals (for e.g., every 1 cM) between each pair of adjacent markers and tests for presence of QTL. The results are expressed as logarithm of the odds (LOD) scores. This score represents the ratio of likelihoods of a statistical model that includes a genetic effect at a particular locus versus a model that does not include that effect (the null expectation). A downside of interval mapping is that it is computationally intensive. Haley and Knott [88] devised accurate and computationally tractable regression-based methods to compute interval maps. Their method is now widely used in eQTL analysis because of its speed. It is now possible to map an entire transcriptome in short time, and single mRNAs and proteins can be mapped with up to 10,000 permutation tests to compute genome-wide significance in less than a minute.

Composite interval mapping

The mapping methods discussed so far assume that a QTL acts independently. However a QTL can be linked to or interact epistatically with other QTLs. Composite interval mapping [89,90] combines interval mapping with multiple-marker regression analysis, which controls for the effect of a known QTL. In short, composite interval mapping uses a subset of significantly associated markers as covariates and controls for variation produced at these controlled markers. Thus, composite interval mapping helps to detect weaker but biologically relevant QTLs. This is particularly important for mapping small effect eQTLs, as strong *cis* eQTLs often mask secondary eQTLs.

Evaluation of mapping precision

Mapping precision for a genetic population can be empirically determined by measuring the offset distance between cis eQTL peaks and the locations of the parent genes. Cis eQTLs are essentially used as positional “gold standards” and any errors in this assumption will tend to be conservative and degrade apparent precision. In other words, cis eQTL offsets will be conservatively biased estimates. We evaluated the mapping precision of the BXD family using a hippocampus exon array data set generated using 67 BXD strains. At a crude level of analysis the offset or errors of QTL mapping in this sample were almost always less than ± 2 Mb—a 4 Mb error. But the error varies as a function of the underlying quality of the genetic map. Some regions of the genome provide much better resolution. We divided the genome into 5 Mb bins and counted the number of markers used to compute QTL maps within each bin. Strong cis eQTLs with LOD scores greater than 22 (genome-wide p value $<10^{-6}$) have a mean gene-to-QTL error of 1.44, 0.70, 0.61 and 0.53 Mb for regions with progressively higher marker densities (Figure 6). Similar results are seen when an offset distance of less than 10 Mb is used to define cis eQTL.

4. RNA-seq eQTL Studies

Only a handful of eQTL studies have yet exploited RNA-seq [66,91]. Most have unfortunately involved immortalized human lymphoblastoid cells—a cell type that is riddled with chromosomal abnormalities that make analysis problematic. As far as we know, the only accessible RNA-seq eQTL study of the brain is our own modest study of the BXD strains, in which we generated an average of 30 million 50 nt reads for each of 28 BXD strains and both parents [27,39] but without any biological replication. Each fragment library was generated using a pool of RNAs from three or more cases using an rRNA depletion method and a protocol that preserved strand polarity. We mapped 350 transcripts with LOD scores above 5, of which 225 were cis eQTLs with a median LOD of 6.4 and an average gene-to-marker offset of ± 3 Mb, whereas 125 were trans eQTLs with a median LOD score of 5.3. The precision of these QTLs is impressive given the small sample size and the lack of biological replication. The precision of trans eQTLs can be expected to be closely matched to those of cis eQTLs—perhaps ± 3.5 Mb—a small enough interval to begin candidate gene analysis. Among the most interesting trans eQTLs relevant to CNS function are *Atf4*, *Atp2b1*, *Atp13a2*, *Atrx*, *Cacnb4*, *Foxa1*, *Foxc1*, *Gap43*, *Kcnk10*, *Lifr*, *Ntsr2*, *Per3*, *Pdyn*, *Pou5f1*, and *Ptprz1*. Those with strong cis eQTLs are *Mrps5* [92], *Alad*, *Ckb*, *Glo1* [93], *Ntn4*, *Prdx2*, and *Sae1*.

5. Pros and Cons of Arrays and RNA-seq for eQTL Studies

Advantages of arrays

For fully sequenced model organisms, commercial arrays now have essentially comprehensive coverage of protein-coding RNAs. The latest arrays from Affymetrix also include a nearly complete set of probes for miRNAs, exons, and even splice junctions. This intense focus on the core subset of RNAs can be an advantage in many situations—particularly in eQTL studies in which investigators are interested on the impact of mRNA variation on phenotypes. The modest dynamic range of arrays relative to RNA-seq is rarely

an issue in eQTL studies, since the key variable is variation across individuals rather than across transcripts. In fact, many eQTL studies discard mean expression values and retain only the offset from the mean (the z score). A second advantage of arrays is that every transcript has its own “real estate” on the array. Even those genes with low expression have an opportunity to produce a hybridization signal. In contrast, RNA-seq count data have a highly skewed distribution, and a small number of genes account for a large fraction of the reads, and many transcripts have no, or very low read counts (<5). Consequently, the power to detect differential expression among shorter and more modestly expressed genes is poor due to high Poisson noise [94]. A final pragmatic advantage of arrays is that the analysis workflow is far less computationally intensive and can be performed on a desktop computer. A strong case can still be made in favor of arrays for large eQTL studies.

Advantage of RNA-seq

RNA-seq offers advantages relative to arrays and can provide more accurate estimates of isoform abundance over a wider dynamic range. Dynamic range is only limited by the RNA complexity of samples (library complexity) and the depth of sequencing. In a small study, Fu and colleagues compared RNA-seq and array data with protein levels in cerebellar cortex and found a slightly better relation between RNA-seq and protein [95]. The higher dynamic range of RNA-seq could potentially facilitate detection of eQTLs associated with transcripts that have either low or very high expression. Saturation of signal at the high end or at the low end of the expression spectrum could obscure genetic expression differences. However, to the best of our knowledge this has never been tested. In Figure 7, we compare cis eQTL effect sizes from the RNA-seq data discussed above with a matched Affymetrix M430 array data set. We extracted all cis eQTLs in both data sets—around 2,000 and 3,000, respectively—and compared their LOD scores. For those transcripts with cis eQTLs in both data sets, there was no advantage to the RNA-seq in terms of effect size or LOD scores of the eQTLs. In this particular case, four arrays per strain outperformed 30 million reads per strain.

The second factor is more compelling—RNA-seq enables expression quantification of novel transcripts and transcripts not represented on arrays. RNA-seq analysis tools such as Cufflinks [96] can utilize reads to annotate novel transcripts by performing reference-based *de novo* assembly of transcripts. This is particularly important in analyzing a complex tissue such as brain known to have high frequency of alternative splicing events [97]. However, most RNA-seq analysis ignores reads mapped to unannotated regions in the reference genome, somewhat reducing the significance of this advantage.

Thirdly, RNA-seq is a hybridization-free approach and does not suffer from confounds such as cross hybridization and artifacts due to variants in probe sequences. Probes with variants—SNPs and small indels—influence hybridization kinetics and cause incorrect detection of the expression level of genes. Ciobanu and colleagues found 25% of apparent cis-modulation detected in the hippocampus was caused by probe variants rather than genuine mRNA quantitative differences [7]. RNA-seq suffers from a similar problem of allele bias inherent when aligning reads to a single reference genome. However, alignment methods allow for mismatches and are less sensitive to sequence differences (See below).

Finally, RNA-seq makes it possible to study complex transcriptional events including alternative splicing and polyadenylation. Hybridization-based approaches use probes targeting small regions of the transcript, mostly single exons. As a result, they have not been used to study splicing extensively. High density exon arrays consisting of both exon and splice junction probes are now available; although they still suffer from systematic errors that lead to over-estimation of alternative splicing [98,99]. Splicing array also require *a priori* information of isoforms for probe design. In contrast, RNA-seq reads mapping to splice junctions provide direct evidence of splicing. These reads can be quantified to roughly estimate the relative abundance of alternative isoforms. Additionally, they can be combined with read distribution across different exons to precisely quantify alternatively spliced isoforms of a gene [100–102]. Of course, the near term technical goal is to sequence and count entire mRNA molecules—so-called single molecule sequencing. Once this goal has been reached it will be possible to use quantitative genetic methods to study splice isoform usage in brain regions and even single cells.

6. RNA-seq Read Alignment and Normalization

Accurate estimation of transcript abundance is critical for the success of eQTL studies. Unbiased alignment of short sequences and correct normalization of RNA-seq counts are important for accurate estimates. There are now accurate RNA-seq aligners that can align reads in a splice-aware manner, but normalization methods for RNA-seq counts are still evolving. Details of RNA-seq data analysis are covered in Chapter 2. Here we will only discuss data analysis issues that are important for conducting a more error-free eQTL study.

Allelic bias in read mapping

Though RNA-seq is a hybridization free approach, it still requires read mapping—essentially digital hybridization—against the reference genome. Unlike hybridization, RNA-seq alignment is not particularly sensitive to the presence of variants because the algorithms allow for specified numbers of mismatches and gaps. Nevertheless, a significant allelic bias can still occur when aligning reads with large numbers of genetic variants relative to a single reference genome [103]. This is because reads with reference alleles will match the reference genome precisely, whereas reads that contain non-reference alleles will not. For example, reads originating from *B* haplotypes in BXD strains will show better alignment rate to the reference genome (*B* genome) than the *D* haplotypes. This bias will appear as a strong *cis* eQTL, similar to ones produced by hybridization artifacts in arrays. This bias can be more of a problem in complicated crosses with multiple and highly divergent haplotypes such as HS and CC.

Degner and colleagues [103] proposed using a masked reference genome in which each polymorphic position (SNP) in the reference genome is masked with a third allele that is neither the reference allele nor the allele from the non-reference haplotype. Their approach cannot reduce allelic bias completely; moreover, it increases the number of unmapped reads because SNP masking this way adds an obligatory mismatch in the alignment. As a result, expression values of genes with large number of variants are underestimated. The advantage of their method is that it only requires read alignment against a single genome. A more widely-used method [56,65] to reduce allelic bias is to create strain-specific genomes for

alignment by substituting reference alleles with known variants. Reads are aligned to both the reference genome and substituted strain-specific genomes. Aligned reads are then combined in a non-redundant manner to estimate transcript abundance. If sequence variants between a non-reference genome and the reference genome are not known, RNA-seq can also be used to detect variants. In this case, RNA-seq reads are as usual aligned against the reference genome and sequence variants are generated for the non-reference case. The locations of these variants can be used to identify genomic regions corresponding to different haplotypes. RNA-seq reads can then be realigned in a haplotype-sensitive manner to reduce bias.

Correct normalization of RNA-seq counts

Choice of normalization procedure can significantly affect the outcomes of gene expression studies. Similar to arrays, normalized estimates of expression are necessary to evaluate whether differences among samples used for eQTL mapping are genuine. Systematic biases in RNA-seq experiments can result from differences in the sequencing depth of the sample libraries, and differences in the length of genes. Deeply sequenced libraries generate more reads per gene than less deeply sequenced libraries. Similarly, longer genes will have more aligned reads compared to shorter genes. Mortazavi and colleagues introduced a normalization method called the *reads per kilobase per million* mapped reads (RPKM) that rescales counts to correct for differences in library size and gene length [104]. RPKM values enables between- and within-library sample comparisons. Sandberg and colleagues found that RPKM values that take only exonic reads into account correlate better with qRT-PCR data [105]. Recent, work has criticized RPKM normalization when detecting for differentially expressed transcripts [94,106,107]. Dudoit and colleagues found a bias that favors longer transcripts with small differences over shorter transcripts with large differences [94].

The total number of RNA-seq counts generated for a gene depends not only on its expression level, length, and depth of sequencing, but also on the library composition and complexity that is used for sequencing. Differences in library composition between samples can contribute to high levels of variability and spuriously high variance in expression. To address this problem, normalization methods such as *Trimmed Mean of M-values* (implemented in edgeR Bioconductor package) [108] and DESeq [109] have been proposed. Both model RNA-seq counts using a negative binomial distribution and have outperformed several other methods in studies focused on differential expression [94,106,107]. No comparison of normalization methods with respect to eQTL mapping has yet been performed. Traditional eQTL mapping methods based on linear regression are highly sensitive to outliers and work best when the expression data are roughly normally distributed. Unfortunately raw RNA-seq expression counts are strongly right skewed, with a few transcripts having extraordinarily high expression. Thus, RNA-seq counts must be transformed in order to apply linear regression or equivalent approaches for eQTL mapping. A log₂ transform is often suitable. Pickerell and colleagues used a normal quantile transformation of RNA-seq counts in their eQTL study [91]. Another study showed that modelling RNA-seq counts using a discrete distribution, such as a negative binomial or a beta binomial distribution, results in higher statistical power in eQTL mapping [110].

7. eQTL Mapping of Alternative Splicing and Polyadenylation

Alternative splicing and alternative polyadenylation increase the complexity of the transcriptome and diversity protein isoforms. Xu and colleagues showed that brain is highly enriched for alternative splice forms [60]. Similarly, highly expressed genes in mammalian brain are known to have unusually long 3' UTRs [111]. Alternative splicing plays an important role in neuronal differentiation, synaptic transmission, and plasticity. Splicing differences and mutations have been linked with several disorders [112,113]. Alternative polyadenylation also plays a role in the stability and localization of the mRNA through interactions with RNA-binding proteins, ribosomes, miRNA etc. Several association studies have utilized exon-arrays and RNA-seq to study variability and heritability of splicing in human lymphoblastoid cell lines [66,91,114]. These studies have confirmed genetic control of splicing variation in humans. However, eQTLs regulating alternative splicing in brain have not been studied extensively. To the best of our knowledge-the only eQTL study that has systematically investigated splicing eQTLs genome-wide is by Heinzen and colleagues. They studied expression in human cortex and peripheral blood using the Affymetrix Human Exon 1.0 ST array [28] and identified 23 and 84 associations at the transcript and the exon level, respectively. This is likely to be a massive underestimate of the actual number of splice isoforms that are under genetic control.

8. RNA-seq for Allele Specific Expression

A remarkable feature of RNA-seq is its ability to assay genome-wide allele-specific expression (ASE) using isogenic F1 hybrids [61–65] made by crossing inbred parents. RNA-seq can reliably distinguish mRNA representing the alternative alleles and can be used to detect unequal production of alleles. An advantage of using F1 animal for ASE analysis is that the two alleles in these animals share all environmental and trans-acting influences. As a result, any genetic expression differences in heterozygotes must be attributed to the local allele-specific endogenous effect.

Key factors in design of genome-wide ASE

A key factor to consider for measuring cis eQTLs on a genomic scale is the presence of appropriate coding variants—usually SNPs—to assay allelic imbalance. Another factor is the sequencing depth needed to detect differences with good statistical power. Fontanillas and colleagues showed that the read depth required to detect an allelic imbalance depends on the size of difference in the allelic expression [115]. They determined that 50 reads per SNP is enough to provide 60% statistical power for larger than 2-fold differences in expression. Small allelic expression differences of less than 25% fold will require more than 500 reads to reach the same power.

ASE can be used to identify imprinted genes by comparing ratios of expression in reciprocal F1 crosses. The reciprocal F1 females are genetically identical but of the polarity of parents differ (e.g., B mother to D father, or D mother to B father). An initial RNA-seq study of this type reported an implausibly high number of imprinted genes in the CNS [116,117]. Correct modeling of biological and technical variation brought this estimate down to less than 100 genes [118]. Several other factors can contribute to error in estimating ASE. Allelic bias in

read mapping to a single reference genome has already been discussed. Other mapping artifacts can also produce false positives include using non-unique reads (reads that can be mapped to multiple locations) and reads that map to low complexity genomic regions. PCR amplification bias during library preparation can also cause false positive allelic imbalances.

Advantages and disadvantages of ASE

An advantage of using ASE based approach to identify cis eQTLs on a genomic scale is that it requires relatively few samples. Additionally, it does not depend on arbitrary window cut-off as used in eQTL mapping. A disadvantage of ASE analysis over eQTL mapping is its complete inability to locate trans eQTLs. Babak and colleagues compared array-based eQTL mapping with RNA-seq ASE to detect cis eQTLs [119]. They found an extensive agreement between cis eQTL results. For genes showing discrepancies between methods, RNA-seq more frequently matched subsequent validation using conventional qRT-PCR protocols.

9. Conclusions

The last decade has seen a rapid growth in the number of eQTL studies of the CNS and large efforts to accumulate massive gene expression data sets across multiple brain regions and cell types. There are two very general findings. First, cis eQTLs have large effects, are often replicable across different data sets and even regions, and are comparatively easy to validate and interpret. However, these first-order cis effects usually do not expose critical gene-gene interactions that define molecular networks. They can, however, be used as seeds to define downstream effects on protein levels and higher order behavioral traits [7]. Second, trans eQTLs usually have smaller effects, are harder to validate, and often don't replicate well across different data sets. But they are also most interesting because they can define gene-gene interactions (e.g. Ciobanu et al. [7] figure 4). Although trans eQTLs are common in expression data sets, they are hard to pin down to single causative genes. But this problem is being resolved. For example, eQTL mapping resolution in the CC will soon provide 1 Mb resolution (5–10 genes in mouse). The BXD family can already routinely achieve a resolution of 2–5 Mb (around 10–50 genes, Figure 6) with high power. Once they are more fully developed, these RI families will each contain 150 strains and they should routinely achieve single gene resolution suitable for high power trans eQTL studies of any part of the CNS.

RNA-seq offers the high dynamic range and resolution essential for capturing small expression differences. Additionally, this method can be used for isoform-specific eQTL mapping. While RNA-seq offers great promises, it has not yet been widely exploited for eQTL mapping studies. This is primarily because of the cost and complexity of library preparation and the high bioinformatics overhead required to process and analyze data. Rapid technical advances have dramatically reduced both types of costs, and the interpretation of RNA-seq is now becoming much more tractable.

Gene expression studies of the brain are particularly challenging due to the extreme cellular heterogeneity. There are probably well over 7,000 statistically distinct cell types with unique mRNA and protein expression profiles in brain. This estimate is based on the well-known cellular complexity of retina—a CNS tissue in which there are ~70 cell types in mouse and

human [120]—and a conservative multiplier of 100 for the effective number of equally complex CNS regions. The high level of still undefined cellular and molecular heterogeneity in the brain is a major issue that still confounds neuroscience. For eQTL studies, use of averaged expression over diverse cell types dilutes, but does not eliminate the important genetic signals. The ultimate genetic studies of gene expression will require extremely efficient workflows to quantify mRNAs, proteins, and metabolites for hundreds of cells belonging to thousands of unique CNS cell types. This may now seem daunting, but rapid progress in single-cell genomics methods [44,45] will make this just as practical in a decade as whole genome sequencing is today.

Acknowledgments

We thank Drs. Megan K. Mulligan, Snigdha Roy, and Evan G. Williams, for their critical reading of drafts of this review. We also thank a group of highly supportive colleagues for generating many massive eQTL data sets over the past decade that are now easily accessible to the neurogenetics research community. Our thanks to Drs. Lu Lu, Glenn D. Rosen, Michael F. Miles, Eldon E. Geisert, Randy D. Blakely, Daniel Goldowitz, Gerd Kempermann, Rupert W. Overall, Boris Tabakoff, Divyen Patel, Kristin M. Hamre, and Robert J. Rooney. Much of the work reviewed here has been generously supported by NIAAA's Integrative Neuroscience Initiative on Alcoholism program (U01 AA016662, U01 AA013499, U24 AA013513, U01 AA014425) and by the UT Center for Integrative and Translational Genomics.

References

1. Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, et al. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature*. 1983; 306:234–238. [PubMed: 6316146]
2. Charles SJ, Moore AT, Yates JR. Genetic mapping of X linked ocular albinism: linkage analysis in British families. *J Med Genet*. 1992; 29:552–554. [PubMed: 1355560]
3. Woo SL, Lidsky AS, Guttler F, Thirumalachary C, Robson KJ. Prenatal diagnosis of classical phenylketonuria by gene mapping. *JAMA*. 1984; 251:1998–2002. [PubMed: 6700105]
4. St George-Hyslop PH, Tanzi RE, Polinsky RJ, Haines JL, Nee L, et al. The genetic defect causing familial Alzheimer's disease maps on chromosome 21. *Science*. 1987; 235:885–890. [PubMed: 2880399]
5. MacDonald Marcy E, Ambrose Christine M, Duyao Mabel P, Myers Richard H, et al. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell*. 1993; 72:971–983. [PubMed: 8458085]
6. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, et al. Genetics of gene expression surveyed in maize, mouse and man. *Nature*. 2003; 422:297–302. [PubMed: 12646919]
7. Ciobanu DC, Lu L, Mozhui K, Wang X, Jagalur M, et al. Detection, validation, and downstream analysis of allelic variation in gene expression. *Genetics*. 2010; 184:119–128. [PubMed: 19884314]
8. Jansen RC, Nap JP. Genetical genomics: the added value from segregation. *Trends Genet*. 2001; 17:388–391. [PubMed: 11418218]
9. Li J, Burmeister M. Genetical genomics: combining genetics with gene expression analysis. *Hum Mol Genet*. 2005; 14(Spec No 2):R163–169. [PubMed: 16244315]
10. Damerval C, Maurice A, Josse JM, de Vienne D. Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. *Genetics*. 1994; 137:289–301. [PubMed: 7914503]
11. Brem RB, Yvert G, Clinton R, Kruglyak L. Genetic dissection of transcriptional regulation in budding yeast. *Science*. 2002; 296:752–755. [PubMed: 11923494]
12. Williams RW, Mulligan MK. Genetic and molecular network analysis of behavior. *Int Rev Neurobiol*. 2012; 104:135–157. [PubMed: 23195314]
13. Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, et al. Genetic analysis of genome-wide variation in human gene expression. *Nature*. 2004; 430:743–747. [PubMed: 15269782]

14. Turk R, t Hoen PA, Sterrenburg E, de Menezes RX, de Meijer EJ, et al. Gene expression variation between mouse inbred strains. *BMC Genomics*. 2004; 5:57. [PubMed: 15317656]
15. Brem RB, Kruglyak L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc Natl Acad Sci U S A*. 2005; 102:1572–1577. [PubMed: 15659551]
16. Hubner N, Wallace CA, Zimdahl H, Petretto E, Schulz H, et al. Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat Genet*. 2005; 37:243–253. [PubMed: 15711544]
17. Storey JD, Madeoy J, Strout JL, Wurfel M, Ronald J, et al. Gene-expression variation within and among human populations. *Am J Hum Genet*. 2007; 80:502–509. [PubMed: 17273971]
18. Massouras A, Waszak SM, Albarca-Aguilera M, Hens K, Holcombe W, et al. Genomic variation and its impact on gene expression in *Drosophila melanogaster*. *PLoS Genet*. 2012; 8:e1003055. [PubMed: 23189034]
19. Geisert EE, Lu L, Freeman-Anderson NE, Templeton JP, Nassr M, et al. Gene expression in the mouse eye: an online resource for genetics using 103 strains of mice. *Mol Vis*. 2009; 15:1730–1763. [PubMed: 19727342]
20. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. *Trends Genet*. 2013; 29:569–574. [PubMed: 23810203]
21. Chesler EJ, Lu L, Shou S, Qu Y, Gu J, et al. Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat Genet*. 2005; 37:233–242. [PubMed: 15711545]
22. Hovatta I, Zapala MA, Broide RS, Schadt EE, Libiger O, et al. DNA variation and brain region-specific expression profiles exhibit different relationships between inbred mouse strains: implications for eQTL mapping studies. *Genome Biol*. 2007; 8:R25. [PubMed: 17324278]
23. Zou F, Chai HS, Younkin CS, Allen M, Crook J, et al. Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet*. 2012; 8:e1002707. [PubMed: 22685416]
24. MacLaren EJ, Sikela JM. Cerebellar gene expression profiling and eQTL analysis in inbred mouse strains selected for ethanol sensitivity. *Alcohol Clin Exp Res*. 2005; 29:1568–1579. [PubMed: 16205357]
25. Myers AJ, Gibbs JR, Webster JA, Rohrer K, Zhao A, et al. A survey of genetic human cortical gene expression. *Nat Genet*. 2007; 39:1494–1499. [PubMed: 17982457]
26. Webster JA, Gibbs JR, Clarke J, Ray M, Zhang W, et al. Genetic control of human brain transcript expression in Alzheimer disease. *Am J Hum Genet*. 2009; 84:445–458. [PubMed: 19361613]
27. Mulligan MK, Wang X, Adler AL, Mozhui K, Lu L, et al. Complex control of GABA(A) receptor subunit mRNA expression: variation, covariation, and genetic regulation. *PLoS One*. 2012; 7:e34586. [PubMed: 22506031]
28. Heinzen EL, Ge D, Cronin KD, Maia JM, Shianna KV, et al. Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol*. 2008; 6:e1. [PubMed: 19222302]
29. Gaglani SM, Lu L, Williams RW, Rosen GD. The genetic control of neocortex volume and covariation with neocortical gene expression in mice. *BMC Neurosci*. 2009; 10:44. [PubMed: 19426526]
30. Mozhui K, Hamre KM, Holmes A, Lu L, Williams RW. Genetic and structural analysis of the basolateral amygdala complex in BXD recombinant inbred mice. *Behav Genet*. 2007; 37:223–243. [PubMed: 17131200]
31. Rosen GD, Pung CJ, Owens CB, Caplow J, Kim H, et al. Genetic modulation of striatal volume by loci on Chrs 6 and 17 in BXD recombinant inbred mice. *Genes Brain Behav*. 2009; 8:296–308. [PubMed: 19191878]
32. Mozhui K, Lu L, Armstrong WE, Williams RW. Sex-specific modulation of gene expression networks in murine hypothalamus. *Front Neurosci*. 2012; 6:63. [PubMed: 22593731]
33. Overall RW, Kempermann G, Peirce J, Lu L, Goldowitz D, et al. Genetics of the hippocampal transcriptome in mouse: a systematic survey and online neurogenomics resource. *Front Neurosci*. 2009; 3:55. [PubMed: 20582282]

34. Ye R, Carneiro AM, Airey D, Sanders-Bush E, Williams RW, et al. Evaluation of heritable determinants of blood and brain serotonin homeostasis using recombinant inbred mice. *Genes Brain Behav.* 2014; 13:247–260. [PubMed: 24102824]
35. Templeton JP, Wang X, Freeman NE, Ma Z, Lu A, et al. A crystallin gene network in the mouse retina. *Exp Eye Res.* 2013; 116:129–140. [PubMed: 23978599]
36. Hitzemann R, Bottomly D, Darakjian P, Walter N, Iancu O, et al. Genes, behavior and next-generation RNA sequencing. *Genes Brain Behav.* 2013; 12:1–12. [PubMed: 23194347]
37. Hitzemann R, Bottomly D, Iancu O, Buck K, Wilmot B, et al. The genetics of gene expression in complex mouse crosses as a tool to study the molecular underpinnings of behavior traits. *Mamm Genome.* 2014; 25:12–22. [PubMed: 24374554]
38. Sun W, Hu Y. eQTL Mapping Using RNA-seq Data. *Stat Biosci.* 2013; 5:198–219. [PubMed: 23667399]
39. Li Z, Mulligan MK, Wang X, Miles MF, Lu L, et al. A transposon in *Comt* generates mRNA variants and causes widespread expression and behavioral differences among mice. *PLoS One.* 2010; 5:e12181. [PubMed: 20808911]
40. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013; 45:580–585. [PubMed: 23715323]
41. Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, et al. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci U S A.* 1990; 87:1663–1667. [PubMed: 1689846]
42. Vanderlinden LA, Saba LM, Kechris K, Miles MF, Hoffman PL, et al. Whole brain and brain regional coexpression network interactions associated with predisposition to alcohol consumption. *PLoS One.* 2013; 8:e68878. [PubMed: 23894363]
43. Wolen AR, Phillips CA, Langston MA, Putman AH, Vorster PJ, et al. Genetic dissection of acute ethanol responsive gene networks in prefrontal cortex: functional and mechanistic implications. *PLoS One.* 2012; 7:e33575. [PubMed: 22511924]
44. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science.* 2014; 343:776–779. [PubMed: 24531970]
45. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, et al. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods.* 2014; 11:163–166. [PubMed: 24363023]
46. Shao NY, Hu HY, Yan Z, Xu Y, Hu H, et al. Comprehensive survey of human brain microRNA by deep sequencing. *BMC Genomics.* 2010; 11:409. [PubMed: 20591156]
47. Bak M, Silahtaroglu A, Moller M, Christensen M, Rath MF, et al. MicroRNA expression in the adult mouse central nervous system. *RNA.* 2008; 14:432–444. [PubMed: 18230762]
48. Somel M, Liu X, Tang L, Yan Z, Hu H, et al. MicroRNA-driven developmental remodeling in the brain distinguishes humans from other primates. *PLoS Biol.* 2011; 9:e1001214. [PubMed: 22162950]
49. Ziats MN, Rennert OM. Identification of differentially expressed microRNAs across the developing human brain. *Mol Psychiatry.* 2013
50. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature.* 2013; 501:506–511. [PubMed: 24037378]
51. Parts L, Hedman AK, Keildson S, Knights AJ, Abreu-Goodger C, et al. Extent, causes, and consequences of small RNA expression variation in human adipose tissue. *PLoS Genet.* 2012; 8:e1002704. [PubMed: 22589741]
52. Parsons MJ, Grimm C, Paya-Cano JL, Fernandes C, Liu L, et al. Genetic variation in hippocampal microRNA expression differences in C57BL/6 J X DBA/2 J (BXD) recombinant inbred mouse strains. *BMC Genomics.* 2012; 13:476. [PubMed: 22974136]
53. Peirce JL, Li H, Wang J, Manly KF, Hitzemann RJ, et al. How replicable are mRNA expression QTL? *Mamm Genome.* 2006; 17:643–656. [PubMed: 16783644]
54. Gerrits A, Li Y, Tesson BM, Bystrykh LV, Weersing E, et al. Expression quantitative trait loci are highly sensitive to cellular differentiation state. *PLoS Genet.* 2009; 5:e1000692. [PubMed: 19834560]

55. Grabowski PJ, Black DL. Alternative RNA splicing in the nervous system. *Prog Neurobiol.* 2001; 65:289–308. [PubMed: 11473790]
56. Blencowe BJ. Alternative splicing: new insights from global analyses. *Cell.* 2006; 126:37–47. [PubMed: 16839875]
57. Dredge BK, Polydorides AD, Darnell RB. The splice of life: alternative splicing and neurological disease. *Nat Rev Neurosci.* 2001; 2:43–50. [PubMed: 11253358]
58. Ule J, Ule A, Spencer J, Williams A, Hu JS, et al. Nova regulates brain-specific splicing to shape the synapse. *Nat Genet.* 2005; 37:844–852. [PubMed: 16041372]
59. Johnson MB, Kawasaki YI, Mason CE, Krsnik Z, Coppola G, et al. Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron.* 2009; 62:494–509. [PubMed: 19477152]
60. Xu Q, Modrek B, Lee C. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res.* 2002; 30:3754–3766. [PubMed: 12202761]
61. Skelly DA, Johansson M, Madeoy J, Wakefield J, Akey JM. A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Res.* 2011; 21:1728–1737. [PubMed: 21873452]
62. Bell GD, Kane NC, Rieseberg LH, Adams KL. RNA-seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. *Genome Biol Evol.* 2013; 5:1309–1323. [PubMed: 23677938]
63. Korir PK, Seoighe C. Inference of allele-specific expression from RNA-seq data. *Methods Mol Biol.* 2014; 1112:49–69. [PubMed: 24478007]
64. McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, et al. Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res.* 2010; 20:816–825. [PubMed: 20354124]
65. Rozowsky J, Abyzov A, Wang J, Alves P, Raha D, et al. AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol Syst Biol.* 2011; 7:522. [PubMed: 21811232]
66. Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, et al. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature.* 2010; 464:773–777. [PubMed: 20220756]
67. Cui P, Lin Q, Ding F, Xin C, Gong W, et al. A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. *Genomics.* 2010; 96:259–265. [PubMed: 20688152]
68. Taylor BA, Wnek C, Kotlus BS, Roemer N, MacTaggart T, et al. Genotyping new BXD recombinant inbred mouse strains and comparison of BXD and consensus maps. *Mamm Genome.* 1999; 10:335–348. [PubMed: 10087289]
69. Threadgill DW, Miller DR, Churchill GA, de Villena FP. The collaborative cross: a recombinant inbred mouse population for the systems genetic era. *ILAR J.* 2011; 52:24–31. [PubMed: 21411855]
70. Valdar W, Solberg LC, Gauguier D, Burnett S, Klennerman P, et al. Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat Genet.* 2006; 38:879–887. [PubMed: 16832355]
71. Hitzemann R, Belknap JK, McWeeney SK. Quantitative trait locus analysis: multiple cross and heterogeneous stock mapping. *Alcohol Res Health.* 2008; 31:261–265. [PubMed: 23584873]
72. Churchill GA, Gatti DM, Munger SC, Svenson KL. The Diversity Outbred mouse population. *Mamm Genome.* 2012; 23:713–718. [PubMed: 22892839]
73. Svenson KL, Gatti DM, Valdar W, Welsh CE, Cheng R, et al. High-resolution genetic mapping using the Mouse Diversity outbred population. *Genetics.* 2012; 190:437–447. [PubMed: 22345611]
74. Lum PY, Chen Y, Zhu J, Lamb J, Melmed S, et al. Elucidating the murine brain transcriptional network in a segregating mouse population to identify core functional modules for obesity and diabetes. *J Neurochem.* 2006; 97(Suppl 1):50–62. [PubMed: 16635250]
75. Yang X, Schadt EE, Wang S, Wang H, Arnold AP, et al. Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.* 2006; 16:995–1004. [PubMed: 16825664]
76. Williams RW, Bennett B, Lu L, Gu J, DeFries JC, et al. Genetic structure of the LXS panel of recombinant inbred mouse strains: a powerful resource for complex trait analysis. *Mamm Genome.* 2004; 15:637–647. [PubMed: 15457343]

77. Williams RW, Gu J, Qi S, Lu L. The genetic structure of recombinant inbred mice: high-resolution consensus maps for complex trait analysis. *Genome Biol.* 2001; 2:RESEARCH0046. [PubMed: 11737945]
78. Keane TM, Goodstadt L, Danecek P, White MA, Wong K, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature.* 2011; 477:289–294. [PubMed: 21921910]
79. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature.* 2002; 420:520–562. [PubMed: 12466850]
80. Carneiro AM, Airey DC, Thompson B, Zhu CB, Lu L, et al. Functional coding variation in recombinant inbred mouse lines reveals multiple serotonin transporter-associated phenotypes. *Proc Natl Acad Sci U S A.* 2009; 106:2047–2052. [PubMed: 19179283]
81. Baud A, Hermesen R, Guryev V, Stridh P, Graham D, et al. Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. *Nat Genet.* 2013; 45:767–775. [PubMed: 23708188]
82. Huang GJ, Shifman S, Valdar W, Johannesson M, Yalcin B, et al. High resolution mapping of expression QTLs in heterogeneous stock mice in multiple tissues. *Genome Res.* 2009; 19:1133–1140. [PubMed: 19376938]
83. Valdar WS, Flint J, Mott R. QTL fine-mapping with recombinant-inbred heterogeneous stocks and in vitro heterogeneous stocks. *Mamm Genome.* 2003; 14:830–838. [PubMed: 14724734]
84. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, et al. Efficient control of population structure in model organism association mapping. *Genetics.* 2008; 178:1709–1723. [PubMed: 18385116]
85. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, et al. The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet.* 2004; 36:1133–1137. [PubMed: 15514660]
86. Philip VM, Sokoloff G, Ackert-Bicknell CL, Striz M, Branstetter L, et al. Genetic analysis in the Collaborative Cross breeding population. *Genome Res.* 2011; 21:1223–1238. [PubMed: 21734011]
87. Aylor DL, Valdar W, Foulds-Mathes W, Buus RJ, Verdugo RA, et al. Genetic analysis of complex traits in the emerging Collaborative Cross. *Genome Res.* 2011; 21:1213–1222. [PubMed: 21406540]
88. Haley CS, Knott SA. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity (Edinb).* 1992; 69:315–324. [PubMed: 16718932]
89. Zeng ZB. Precision mapping of quantitative trait loci. *Genetics.* 1994; 136:1457–1468. [PubMed: 8013918]
90. Jansen RC, Stam P. High resolution of quantitative traits into multiple loci via interval mapping. *Genetics.* 1994; 136:1447–1455. [PubMed: 8013917]
91. Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature.* 2010; 464:768–772. [PubMed: 20220758]
92. Houtkooper RH, Mouchiroud L, Ryu D, Moullan N, Katsyuba E, et al. Mitonuclear protein imbalance as a conserved longevity mechanism. *Nature.* 2013; 497:451–457. [PubMed: 23698443]
93. Williams, Rt; Lim, JE.; Harr, B.; Wing, C.; Walters, R., et al. A common and unstable copy number variant is associated with differences in *Glo1* expression and anxiety-like behavior. *PLoS One.* 2009; 4:e4649. [PubMed: 19266052]
94. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics.* 2010; 11:94. [PubMed: 20167110]
95. Fu X, Fu N, Guo S, Yan Z, Xu Y, et al. Estimating accuracy of RNA-Seq and microarrays with proteomics. *BMC Genomics.* 2009; 10:161. [PubMed: 19371429]
96. Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics.* 2011; 27:2325–2329. [PubMed: 21697122]
97. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet.* 2008; 40:1413–1415. [PubMed: 18978789]

98. Gaidatzis D, Jacobeit K, Oakeley EJ, Stadler MB. Overestimation of alternative splicing caused by variable probe characteristics in exon arrays. *Nucleic Acids Res.* 2009; 37:e107. [PubMed: 19528075]
99. Laderas TG, Walter NA, Mooney M, Vartanian K, Darakjian P, et al. Computational detection of alternative exon usage. *Front Neurosci.* 2011; 5:69. [PubMed: 21625610]
100. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010; 28:511–515. [PubMed: 20436464]
101. Jiang H, Wong WH. Statistical inferences for isoform expression in RNA-Seq. *Bioinformatics.* 2009; 25:1026–1032. [PubMed: 19244387]
102. Griffith M, Griffith OL, Mwenifumbo J, Goya R, Morrissy AS, et al. Alternative expression analysis by RNA sequencing. *Nat Methods.* 2010; 7:843–847. [PubMed: 20835245]
103. Degner JF, Marioni JC, Pai AA, Pickrell JK, Nkadori E, et al. Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics.* 2009; 25:3207–3212. [PubMed: 19808877]
104. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008; 5:621–628. [PubMed: 18516045]
105. Ramskold D, Wang ET, Burge CB, Sandberg R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput Biol.* 2009; 5:e1000598. [PubMed: 20011106]
106. Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, et al. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform.* 2013; 14:671–683. [PubMed: 22988256]
107. Rapaport F, Khanin R, Liang Y, Pirun M, Krek A, et al. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol.* 2013; 14:R95. [PubMed: 24020486]
108. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010; 26:139–140. [PubMed: 19910308]
109. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010; 11:R106. [PubMed: 20979621]
110. Sun W. A statistical framework for eQTL mapping using RNA-seq data. *Biometrics.* 2012; 68:1–11. [PubMed: 21838806]
111. Miura P, Shenker S, Andreu-Agullo C, Westholm JO, Lai EC. Widespread and extensive lengthening of 3' UTRs in the mammalian brain. *Genome Res.* 2013; 23:812–825. [PubMed: 23520388]
112. Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. *Genes Dev.* 2003; 17:419–437. [PubMed: 12600935]
113. Nissim-Rafinia M, Kerem B. The splicing machinery is a genetic modifier of disease severity. *Trends Genet.* 2005; 21:480–483. [PubMed: 16039004]
114. Nembaware V, Lupindo B, Schouest K, Spillane C, Scheffler K, et al. Genome-wide survey of allele-specific splicing in humans. *BMC Genomics.* 2008; 9:265. [PubMed: 18518984]
115. Fontanillas P, Landry CR, Wittkopp PJ, Russ C, Gruber JD, et al. Key considerations for measuring allelic expression on a genomic scale using high-throughput sequencing. *Mol Ecol.* 2010; 19(Suppl 1):212–227. [PubMed: 20331781]
116. Gregg C, Zhang J, Butler JE, Haig D, Dulac C. Sex-specific parent-of-origin allelic expression in the mouse brain. *Science.* 2010; 329:682–685. [PubMed: 20616234]
117. Gregg C, Zhang J, Weissbourd B, Luo S, Schroth GP, et al. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science.* 2010; 329:643–648. [PubMed: 20616232]
118. DeVeale B, van der Kooy D, Babak T. Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. *PLoS Genet.* 2012; 8:e1002600. [PubMed: 22479196]
119. Babak T, Garrett-Engle P, Armour CD, Raymond CK, Keller MP, et al. Genetic validation of whole-transcriptome sequencing for mapping expression affected by cis-regulatory variation. *BMC Genomics.* 2010; 11:473. [PubMed: 20707912]

120. Marc RE, Jones BW, Lauritzen JS, Watt CB, Anderson JR. Building retinal connectomes. *Curr Opin Neurobiol.* 2012; 22:568–574. [PubMed: 22498714]

Figure 1.

Linkage maps of cis and trans eQTLs in mouse hippocampus. (A) *Gabrg2* expression is controlled by a trans eQTL on Chr 5 at 138 Mb (LRS = 18.2 on the Y axis, equivalent to a LOD score of 3.94). The *Gabrg2* gene itself is located on Chr 11 at 41 Mb (triangle on X axis). (B) In contrast, *Grin2b* expression is controlled by a cis eQTL with a peak LRS score of 77 located on Chr 6 at 135 Mb. This location corresponds precisely to the location of the *Grin2b* gene (triangle). (C) Magnified view of the *Grin2b* cis eQTL that provides much more detail on the QTL map and its chromosomal context. The small shaded or colored blocks along the top represent genes on mouse Chr 6. Shading is used to encode the density of polymorphic SNPs within each gene. The horizontal lines provide genome-wide significance thresholds for the QTL determined by permutation analysis (upper <.05 and lower <.63). The hash on the X axis summarizes the density of SNPs along the chromosome. Regions of the genome that are identical by descent (i.e., not variable in the BXD family) have almost no X axis hash. Finally, the so-called additive genetic effect (see Ref 12) is marked by the thinner line and the right-side Y axis. All data here were generated in GeneNetwork (www.genenetwork.org) using the BXD mouse *Hippocampus Consortium*

M430v2 (Jun06) PDNN array data set (GeneNetwork.org, accession number GN112, $n = 67$, probe sets 1418177_at and 1457003_at).

Figure 2.

Expression variation in hippocampal mRNA expression. Log transformed gene expression values were used to calculate the coefficient of variation (X axis) across 99 genetically diverse strains. The Y axis represents $\log_2$ of numbers of assays. A total of ~44,500 probe sets in the Hippocampus Consortium data (GN112) were used to generate this plot.

Figure 3.

Comparison of LOD scores of cis and trans eQTLs. Numbers of cis eQTLs (solid) and trans eQTLs (dashed) are plotted on left y-axis as a function of LOD score (X axis). The fraction of cis eQTLs (dotted) are plotted on right Y axis. Cis eQTLs are defined as those eQTLs within 5 Mb of the parent gene. Trans eQTL are usually on different chromosome or more than 10 Mb from the parent gene. A total of 44,500 probe sets in the Hippocampus Consortium data were analyzed.

Figure 4.

Impact of cellular heterogeneity on detection of cis eQTLs. Numbers of cis eQTLs identified in the cerebellum (grey bars) are plotted as a function of LOD score. A subset of these cerebellar cis eQTLs were also identified in matched whole brain data (black bars). A total of ~44,500 probe sets in the GE-NIAAA cerebellum (GN72, $n = 28$) and whole brain (GN123, $n = 30$) set were used for this analysis.

Figure 5. Highly selective but diluted expression of *En1* (*Engrailed 1*) in midbrain. (A) Sagittal section of the brainstem and ventral tegmental area (VTA) from the Allen Brain Atlas (www.brain-map.org). (B) Matched *in situ* hybridization image of *En1* expression in serotonin neurons from the Allen Brain Atlas. (C) *En1* expression in midbrain is controlled by a strong cis eQTL on Chr 1 with a LOD of 16.7 (Y axis). This location corresponds to the location of *En1* itself (triangle on the X axis). Dilution is clearly not a factor in this instance.

Figure 6.

QTL mapping precision of the BXD family. The precision of QTL mapping was estimated empirically by measuring the distance between the marker closest to the *cis* QTL peak and the location of the parent gene (specifically, the position of the proximal-most nucleotide in each probe set associated with a *cis* eQTL). A total of ~27,500 *cis* eQTLs were used for this analysis (GN206, $n = 67$ BXD strains). Average precision in megabases (Y axis) is plotted as a function of LOD score (X axis). Each bar includes a count of *cis* eQTLs and the SEM. Four levels of shadings were used to evaluate the effects of marker density (per 5 Mb bin) on precision. Precision varies from ± 1.2 – 1.8 Mb for *cis* eQTLs with modest LOD scores to ± 0.5 Mb for *cis* eQTL with high LOD scores in regions with high marker density. Precision would be improved by a factor of two by including all strains

Figure 7.

Comparison of cis eQTLs identified by arrays and RNA-seq. (A) The box plots show LOD score distributions (Y axis) for RNA-seq ($n = 1,779$) versus arrays ($n = 2,839$). Cis eQTLs are defined as those eQTLs having a LOD score of greater than 2.0 and within 5 Mb of the parent gene. A total of ~44,500 assays in a whole brain array data set (GN123, $n = 30$) were compared to ~200,000 exons in a whole brain RNA-seq data set (GN394, $n = 28$). (B) Scatter plot of LOD scores for 105 cis eQTLs shared by array and RNA-seq data sets.

Table 1

CNS eQTL data sets for BXD strains (see www.genenetwork.org for a complete list)

GN accession	CNS region	BXD <i>n</i>	Platform	mRNA assays	cis* eQTLs	trans* eQTLs
GN323	Amygdala (BLA)	54	Affy Mouse Gene 1.0 ST	34,760	1,824	3,058
GN394	Whole brain	28	SOLiD RNA-seq. (transcript level)	26,408	418	1,075
GN123	Whole brain	39	Affy Mouse 430 2.0	45,101	2,162	2,771
GN56	Cerebellum	30	Affy Mouse 430 2.0	45,101	2,559	6,631
GN112	Hippocampus	67	Affy Mouse 430 2.0	45,101	4,927	5,520
GN281	Hypothalamus	46	Affy Mouse Gene 1.0 ST	34,760	1,759	4,230
GN375	Neocortex	43	Illumina Mouse WG-6 v2	45,281	2,614	4,120
GN156	Nucleus accumbens	34	Affy Mouse 430 2.0	45,101	3,648	4,624
GN135	Prefrontal cortex	27	Affy Mouse 430 2.0	45,101	2,256	4,392
GN399	Striatum	32	Affy Mouse 430 2.0	45,101	2,115	3,652
GN228	Ventral tegmental area	35	Affy Mouse 430 2.0	45,101	3,156	3,629
GN381	Midbrain	37	Agilent SurePrint G3 GE	55,681	6,526	7,258
GN302	Retina	73	Illumina Mouse WG-6 v2	45,281	3,833	6,941

* Cis eQTLs defined as LOD > 3 and within ± 5 Mb of parent gene. Trans eQTLs defined as LOD > 3 and more than 10 Mb from parent gene.