



Published in final edited form as:

*Int Rev Neurobiol.* 2014 ; 116: 1–19. doi:10.1016/B978-0-12-801105-8.00001-1.

## Introduction to Sequencing the Brain Transcriptome

Robert Hitzemann<sup>\*,†,1</sup>, Priscila Darakjian<sup>\*</sup>, Nikki Walter<sup>\*,†</sup>, Ovidiu Dan Iancu<sup>\*</sup>, Robert Searles<sup>‡</sup>, and Shannon McWeeney<sup>§,¶</sup>

<sup>\*</sup>Department of Behavioral Neuroscience, Oregon Health & Science University, Portland, Oregon, USA

<sup>†</sup>Research Service, Veterans Affairs Medical Center, Portland, Oregon, USA

<sup>‡</sup>Integrative Genomics Laboratory, Oregon Health & Science University, Portland, Oregon, USA

<sup>§</sup>Oregon Clinical and Translational Research Institute, Oregon Health & Science University, Portland, Oregon, USA

<sup>¶</sup>Division of Biostatistics, Public Health & Preventative Medicine, Oregon Health & Science University, Portland, Oregon, USA

### Abstract

High-throughput next-generation sequencing is now entering its second decade. However, it was not until 2008 that the first report of sequencing the brain transcriptome appeared (Mortazavi, Williams, Mccue, Schaeffer, & Wold, 2008). These authors compared short-read RNA-Seq data for mouse whole brain with microarray results for the same sample and noted both the advantages and disadvantages of the RNA-Seq approach. While RNA-Seq provided exon level resolution, the majority of the reads were provided by a small proportion of highly expressed genes and the data analysis was exceedingly complex. Over the past 6 years, there have been substantial improvements in both RNA-Seq technology and data analysis. This volume contains 11 chapters that detail various aspects of sequencing the brain transcriptome. Some of the chapters are very methods driven, while others focus on the use of RNA-Seq to study such diverse areas as development, schizophrenia, and drug abuse. This chapter briefly reviews the transition from microarrays to RNA-Seq as the preferred method for analyzing the brain transcriptome. Compared with microarrays, RNA-Seq has a greater dynamic range, detects both coding and noncoding RNAs, is superior for gene network construction, detects alternative spliced transcripts, and can be used to extract genotype information, e.g., nonsynonymous coding single nucleotide polymorphisms. RNA-Seq embraces the complexity of the brain transcriptome and provides a mechanism to understand the underlying regulatory code; the potential to inform the brain–behavior–disease relationships is substantial.

## 1. INTRODUCTION

Next-generation sequencing (NGS) refers to a variety of related technologies, often termed massively parallel sequencing. The first NGS platform (Roche 454) was introduced in 2004.

<sup>1</sup>Corresponding author: hitzeman@ohsu.edu.

Subsequently, other platforms were released by several manufacturers: Illumina (Solexa), Helicos, Pacific Biosciences, and Life Technologies (ABI). Although the instruments differ in the underlying chemistry and technical approach, the platforms are similar in their capability of producing very large numbers of simultaneous reads relative to traditional methods. Thus, it is now possible to sequence whole genomes, exomes, and transcriptomes for a reasonable cost and effort. The technology of transcriptome sequencing, also known as RNA-Seq, has matured to the point that it is reasonable to propose substituting RNA-Seq for microarray-based assessments of global gene expression. Of particular importance to our laboratories are the advantages RNA-Seq has over microarray platforms when analyzing complex rodent crosses, e.g., heterogeneous stocks (HSs). However, the same argument can be made when analyzing any outbred population, including humans. Of particular relevance to the brain transcriptome are the advantages RNA-Seq has over microarrays in analyzing alternative splicing. This chapter provides a starting point for understanding the emergence of RNA-Seq and emphasizes *transcriptome/behavior* relationships.

## 2. FROM MICROARRAYS TO RNA-Seq

Cirelli and Tononi (1999) were among the first to report genomewide brain gene expression profiling associated with a behavioral phenotype; both mRNA differential display and cDNA arrays were used to examine the effects of sleep deprivation on rat prefrontal cortex gene expression. Sandberg et al. (2000) used Affymetrix microarrays to detect differences in brain gene expression between two inbred mouse strains (C57BL/6J [B6] and 129SvEv [129; now 129S6/SvEvTac]). Importantly, these authors observed that some differentially expressed (DE) genes were found in chromosomal regions with known behavioral quantitative trait loci (QTLs). For example, *Kcnj9* that encodes for GIRK3, an inwardly rectifying potassium channel, was DE (higher expression in the 129 strain) and is located on distal chromosome 1 in a region where QTLs had been identified for locomotor activity, alcohol and pentobarbital withdrawal, open-field emotionality, and certain aspects of fear-conditioned behavior (see Sandberg et al., 2000). Subsequently, Buck and colleagues (Buck, Milner, Denmark, Grant, & Kozell, 2012; Kozell, Walter, Milner, Wickman, & Buck, 2009) have shown that *Kcnj9* is a quantitative trait gene (QTG) for the withdrawal phenotypes. Over the past decade, this alignment of global brain gene expression data and behavioral QTLs has been reported in numerous publications and discussed in numerous symposia and reviews (e.g., Bergeson et al., 2005; Farris & Miles, 2012; Hoffman et al., 2003; Matthews et al., 2005; McBride et al., 2005; Saba et al., 2011; Sikela et al., 2006; Tabakoff et al., 2009). The association gained further support as the focus turned to genes whose expression appeared to be regulated by a factor or factors within the behavioral QTL interval. Web tools have been developed to facilitate integrating behavioral and brain microarray data (e.g., [www.genenetwork.org](http://www.genenetwork.org) and <http://phenogen.ucdenver.edu/PhenoGen/index.jsp>; Chapter 8). This integration has been successful in detecting several candidate QTGs for behavioral phenotypes (see, e.g., Hitzemann et al., 2004; Hofstetter et al., 2008; Mulligan et al., 2006; Saba et al., 2011; Tabakoff et al., 2009).

The alignment of DE genes with a behavioral phenotype can be further examined using a variety of secondary analyses, e.g., examining if the DE genes cluster within known gene ontology categories (Pavlidis, Qin, Arango, Mann, & Sibille, 2004) or are part of a known

protein–protein interaction network (Bebek & Yang, 2007; Feng, Shaw, Rosen, Lin, & Kibbe, 2012). DE genes can also be grouped on the basis of common transcription factors and other regulatory elements (e.g., Vadigepalli, Chakravarthula, Zak, Schwaber, & Gonye, 2003). In addition to DE genes, microarrays have also facilitated gene coexpression-based analyses, such as the Weighted Gene Coexpression Network Analysis (WGCNA; Horvath et al., 2006; Zhang & Horvath, 2005). The rationale behind these approaches is that coexpressed genes frequently code for interacting proteins, which in turn leads to new insights into protein function(s) and in some cases leads to discovery of protein function (Zhao et al., 2010). Coexpression analysis has been used to analyze differences in functional brain organization between nonhuman primates and humans (Oldham, Horvath, & Geschwind, 2006), regional differences in the functional organization of the human brain (Oldham et al., 2008), and the molecular pathology of autism (Voineagu et al., 2011) and alcoholism (see Chapter 11).

Despite these successes, microarray-based approaches are not without problems. First, differences in brain gene expression among genetically unique individuals or lines selected for behavioral traits are generally small; reported differences of 15–25% are not uncommon. To some extent, these small variations occur because hybridization isotherms for oligonucleotide arrays are frequently not linear due to probe saturation (Pozhitkov, Boube, Brouwer, & Noble, 2010).

A second problem with oligonucleotide arrays is the effect of single nucleotide polymorphisms (SNPs; Duan, Pauley, Spindel, Zhang, & Norgren, 2010; Peirce et al., 2006; Sliwerska et al., 2007; Walter et al., 2009, 2007). Rodent oligonucleotide arrays are based upon the sequence of the B6 mouse or Brown-Norway (BN) rat. Even inbred strains closely related to the B6 or BN strains may differ by several million SNPs (see, e.g., Keane et al., 2011), which in turn can cause significant hybridization artifacts (Walter et al., 2009, 2007). Masking for SNPs can improve this situation but results in deleting probes or even an entire probe set from the analysis. Walter et al. (2009) used NGS to address the SNP problem, building upon the repeated observation that, when comparing gene expression in the B6 and DBA/2J (D2) inbred mouse strains (or crosses and selected lines formed from these strains) and after masking for known SNPs in the D2 strain, there remained an excess of genes showing higher expression in the B6 strain. Similarly, this was also observed in the case of cis-eQTLs showing higher expression associated with the B6 allele (see Mulligan et al., 2006; Peirce et al., 2006; Walter et al., 2007). The two possible explanations for these observations were the following: (a) gene expression was actually higher in the B6 strain or (b) there were many uncharacterized D2 SNPs, which led to decreased binding of D2-derived target on probes containing the SNP locale. Preliminary direct sequencing and quantitative PCR data pointed to missing SNPs. NGS was used to analyze a 3-Mbp region of Chr 1 (171.5–174.5 Mbp) that was enriched in a number of behavioral QTLs and transcripts DE between the B6 and D2 strains. B6 and D2 BAC clones tiled across the region were sequenced using the short-read Illumina Ix and ABI SOLiD 2 platforms. The results obtained (30–100 × coverage) illustrated that there were 160% more SNPs in the region than previously reported (Walter et al., 2009); these data have been confirmed (Keane et al., 2011; R. Williams, unpublished observations). The integration of these SNPs to the mask markedly reduced the disparity in DE genes between the B6 and D2 strains.

A third problem with oligonucleotide arrays is the annotation and summarization issues associated with predefined reporters/probes (e.g., Allison, Cui, Page, & Sabripour, 2006; Lu, Lee, Salit, & Cam, 2007). Interestingly, on some arrays, a significant number of the represented transcripts are actually long noncoding RNAs (ncRNAs) (see, e.g., Liao et al., 2011). But tens of thousands of ncRNAs, many of which have important regulatory functions (Mattick, 2011; also see Chapter 7), are not represented on the arrays.

A fourth problem with oligonucleotide arrays is that 3'UTR-orientated microarrays provide relatively little information about alternative splicing, which is particularly high in brain (Johnson et al., 2009; Li, Lee, & Black, 2007; Licatalosi & Darnell, 2006; Mortazavi, Williams, Mccue, Schaeffer, & Wold, 2008). The Affymetrix Mouse 1.0 Exon ST array collects data on alternative splicing, but when used to detect differential alternative splicing, it is particularly sensitive to the “SNP effect” due to the smaller number of probes per probe set (Laderas et al., 2011).

### 3. NGS PLATFORMS

There are several excellent reviews of the various NGS platforms (e.g., Mardis, 2008, 2011; Martin & Wang, 2011; Metzker, 2010; Ozsolak & Milos, 2011; Rothberg et al., 2011). Understanding in some depth how the platforms work is critical to understanding where errors develop and are propagated from sample preparation to alignment to data analysis. The differences in platforms will not be discussed here. We simply note that for RNA-Seq experiments, the majority have used the Illumina platform (see, e.g., Costa, Angelini, De Feis, & Ciccodicola, 2010). The promise of a high-throughput, high read instrument with minimal library preparation remains a promise. Such an instrument would be particularly welcome for sequencing the brain transcriptome given the diversity of cell types present and the numerous comparisons that could be made.

### 4. RNA-Seq OVERVIEW

The first and perhaps the most important step of an RNA-Seq experiment is the same as that for a microarray experiment, the isolation of high-quality RNA. Although both RNA-Seq and microarrays can be used on fragmented RNA such as that found in formalin-fixed-paraffin-embedded samples, the biases present in such samples for genome-wide sequencing are difficult to assess. RNA quality is routinely examined on the Agilent BioAnalyzer or a similar instrument; an RNA integrity number (RIN) of 8 is generally considered high quality. Unfortunately for brain samples, the amount of beginning tissue may be very small, and obtaining a reliable RIN or even accurately measuring the amount of RNA may be difficult. Even within very discrete brain regions, there are multiple cell types, and some experiments need to focus on a specific subset of cells or even a single cell. Eberwine and colleagues at the University of Pennsylvania have pioneered techniques for the linear amplification of small amounts of RNA; an online audio describing the procedures when beginning with only femtograms of material is available (Morris, Singh, & Eberwine, 2011). Many RNA-Seq experiments begin with postmortem material that has been stored, often under variable conditions, including differences in the postmortem interval (PMI). Depending on the length of the PMI, the RNA in a sample may be moderately to

significantly degraded as assessed by the RIN and other  $Q/C$  measures. For samples with integrity numbers  $<6$ , one should consider ribosome depletion as opposed to a polyA +preparation. Ribosome-depleted samples also have the advantage of including coding and ncRNAs which are not polyadenylated; tiling array data suggest that more than 40% of transcripts are not polyadenylated (Cheng et al., 2005). Cui et al. (2010) have compared RNA-Seq of RiboMinus (rmRNA) and poly(A)-selected (mRNA) samples; the starting total RNA was extracted from BALB/c mouse whole brain. The authors found (on a percent basis) that there were marked read distribution differences between samples. The percentage of known exon reads was twice as high in the mRNA sample (60%), while the percentage of both intronic and intergenic reads was twice as high in the rmRNA sample (25% and 44%, respectively). Both samples detected reads in essentially the same population of RefSeq-defined genes, i.e., there was not a substantial read bias. So the use of rRNA-depleted or poly(A)-selected RNA depends on the questions being asked and the estimated read density per sample. Data collected in our laboratory and elsewhere (Bottomly et al., 2011; Marionni, Mason, Mane, Stephens, & Gilad, 2008; Mortazavi et al., 2008) have found that 20–40 million reads are generally adequate for most estimates of gene expression. If the goal is to quantitatively measure expression at the exon level, then the read density must be increased significantly, perhaps by an order of magnitude (see Labaj et al., 2011; Lee, Mayfield, & Harris, 2014). Such exon level measurements are obviously best suited for poly(A)-selected samples, especially when one is dealing with multiple biological replicates and assuming resources are reasonably limited; i.e., it is very likely that it will be necessary to multiplex samples. But if one is only interested in gene expression and can maintain total exonic read density at 20–40 million, then rmRNA could be used, and significant information on ncRNAs and mRNAs without a poly(A) tail can be obtained. Cui et al. (2010) also used a procedure that facilitates both the quantification of transcripts derived from opposite strands and determining the directionality of transcription (Costa et al., 2010; Martin & Wang, 2011; see also Chapter 2). Using the strand-specific data, Cui et al. (2010) made several salient observations: (a) 99.9% of the junction reads are in the sense orientation; (b) nearly all expressed genes have natural antisense transcripts (the proportion may be as high as 70% of expressed genes [Katayama et al., 2005]); (c) poorly expressed genes tend to have more pronounced antisense transcription; and (d) the antisense transcripts are enriched in the promoter and terminal transcript regions. This enrichment is likely the result of divergent transcription initiation of RNA polymerase II (Core, Waterfall, & Lis, 2008; Preker et al., 2008).

Samples from very discrete brain regions are often prepared by laser capture microdissection (LCM). Given the steps involved in preparing the LCM samples, including staining and dehydration, care needs to be taken to maintain RNA quality. Chen et al. (2011) appear to be the first to couple LCM and RNA-Seq to examine brain gene expression. They examined rat GABAergic neurons projecting from the nucleus accumbens to the ventral pallidum. Cells were labeled using the retrograde tracer, Fluorogold. Approximately 1500 cells were labeled and isolated by LCM in each of four animals; this in turn produced ~4 ng of RNA per animal, and the average RIN was 8.1. Samples were independently amplified for microarray and RNA-Seq; for genes detected on both platforms, the correlation for gene expression was ~0.7. Not surprisingly, the correlation was better for the highly expressed genes. We have

used LCM to examine gene expression in discrete regions of the mouse brain (prelimbic cortex, nucleus accumbens shell, and central nucleus of the amygdala; Colville, AM & Hitzemann, RJ unpublished observations). Sufficient high-quality RNA was obtained from each sample (>100 ng) that amplification was not necessary. Although the samples were only used for RNA-Seq, the data obtained for the nucleus accumbens shell appear at the gene level to be very similar to data previously obtained for the ventral striatum when using microarrays (e.g., Iancu et al., 2010).

In addition to examining gene expression in discrete brain regions and discrete cell types, for some applications, it is desirable to assess the synaptic transcriptome (see, e.g., Eipper-Mains, Eipper, & Mains, 2012). A key mechanism of synaptic plasticity is the local synthesis of proteins from synaptic mRNA. Techniques for isolating synaptosomes from adult brains and growth cones from developing brains are well established using gradient centrifugation (e.g., Hitzemann & Loh, 1978). Synaptoneuroosomes are prepared by filtration of tissue homogenate through a series of filters to obtain a fraction that is enriched in pinched-off dendritic spines (Lugli & Smalheiser, 2013). Regardless of preparation, once isolated, these fractions can be subjected to sequencing as outlined earlier (e.g., Eipper-Mains et al., 2011). A key to the use of these fractions will be assessments of subcellular contamination.

The next step in an RNA-Seq experiment involves the synthesis of high-quality double-stranded (ds) cDNA. The most widely used procedure fragments the RNA before reverse transcription, followed by second-strand synthesis. This approach has the advantage of minimizing the effects of secondary RNA structure on first-strand synthesis. If the adapters needed for the sequencing are added after the ds cDNA is formed, information on strandedness is lost. There are several procedures, including ligating adapters to the fragmented RNA, that will maintain strand information (Ingolia, Ghaemmaghami, Newman, & Weissman, 2009; Li et al., 2008; Parkhomchuk et al., 2009). The alternative to using fragmented RNA is to synthesize the cDNA from intact RNA and then fragment. This approach has a clear advantage for platforms that are capable of long to very long reads. For the Illumina, SOLiD, and 454 platforms, the final step prior to the actual sequencing is the clonal amplification of the fragmented cDNA. Both 454 and SOLiD use emulsion PCR on a bead surface, while Illumina uses enzymatic amplification on a glass surface (flow cell). The sequencing and detection methods differ among the three platforms (see Mardis, 2011 and Metzker, 2010 for details). The 454 sequencer use a polymerase-mediated incorporation of unlabeled nucleotides; detection is via light emitted by secondary reactions with the released PPi. Illumina also uses a polymerase-mediated sequencing but uses end-blocked fluorescent nucleotides in a protocol similar to traditional Big Dye sequencing; detection comes from following the incorporation of the nucleotide attached fluorescent tags. SOLiD sequencing uses the ligase-mediated addition of 2-base encoded fluorescent oligonucleotides; detection is from fluorescent emission of the incorporated oligonucleotides. The SOLiD system differs from Illumina and 454 in that each base is determined twice. The quality of the base calls for all three platforms is very good. Quality is measured in terms of a Phred Score ( $Q$ ), which was originally developed to assess base calls for the human genome project (Ewing, Hillier, Wendl, & Green, 1998). A  $Q$  score of 20 indicates a 99% accuracy rate, and a score of 30

indicates a 99.9% accuracy rate.  $Q_{30}$  values are routinely obtained for NGS platforms. Typically, the  $Q$  value decreases with increasing read length.

## 5. RNA-Seq AND DATA ANALYSIS

Before commenting on the analysis of RNA-Seq data, it is useful to recount the analysis controversies that arose with the introduction of microarrays. In 1999, *Nature Genetics* devoted an entire issue (volume 21—January) to microarrays. Cautionary concerns were raised around issues of data analysis (Lander, 1999). Microarray experiments, at the time, were generally expensive, limiting sample sizes. Small sample sizes and thousands of independent observations per sample were seen as a prescription for statistical disaster. Initial attempts to deal with this problem frequently involved using a nonstatistical threshold for a meaningful difference, e.g., a twofold difference in expression. This approach frequently worked well in some applications, e.g., when comparing cancerous and noncancerous tissue; however, this approach was destined not to work well in brain, where differences in expression among experimental groups were much smaller. Initially, journal reviewers, editors, and study sections panned microarray experiments as being “fishing expeditions,” with no clear hypothesis. The idea of discovery science as a valuable strategy was a minority opinion.

Despite the obstacles, microarray experiments eventually flourished; technology and analysis methods improved. One might have predicted that the microarray experience would have laid the groundwork for the acceptance of NGS. However, the introduction of the 454 sequencer (Margulies et al., 2005) was met with a similar resistance; the argument was made that the data sets were so large that only one of the established genome centers would have the necessary bioinformatics expertise. But as NGS technology improved so did the analytic approaches, such that by 2007/2008, RNA-Seq data appeared from several different laboratories (Marioni et al., 2008; Mortazavi et al., 2008; Sugarbaker et al., 2008; Torres, Metta, Ottenwalder, & Schlotterer, 2008; Weber, Weber, Carr, Wilkerson, & Ohlrogge, 2007). Workflows emerged that addressed the measurement of not only DE genes but also differential alternative splicing and the detection of novel transcripts (Marioni et al., 2008). Bullard, Purdom, Hansen, and Dudoit (2010) examined a number of statistical issues associated with using RNA-Seq to detect DE genes. Similar to Marioni et al. (2008), they found that most sources of technical variation had only small effects on detecting DE transcripts. The most significant effect on DE transcripts was data normalization. Bullard et al. (2010) concluded that their “main novel finding is the extent to which normalization affects differential expression results: sensitivity varies more between normalization procedures than between test statistics...we propose scaling gene counts by a quantile of the gene count distribution (the upper-quartile).”

This volume contains several chapters that address in some detail the analysis of RNA-Seq data (see Chapters 2, 3, and 11); these chapters especially emphasize the evolution of RNA-Seq analysis over the past 3–4 years. In addition to improvements in analysis strategy, sample power has in general improved with decreasing costs and the ability to multiplex samples with adequate read depth (at least at a level sufficient for gene summarization

statistics). If one is interested in quantifying alternative splicing, then substantially greater read depth is required (see, e.g., Lee et al., 2014).

RNA-Seq data have some unique properties that affect the strategies for data analysis (Garber, Grabherr, Guttman, & Trapnell, 2011). First, unlike microarray data where the output is fluorescence intensity (more or less a continuous measure), the output from an RNA-Seq experiment is digital in the form of read counts. For the microarray experiment, familiar statistics such as a *t*-test or ANOVA are appropriate (assuming variances are equal); for RNA-Seq data, these statistics are not directly applicable. Robinson, Mccarthy, and Smyth (2009) proposed the use of the empirical analysis of digital gene expression in R (edgeR), a variant of a procedure used to analyze SAGE data. edgeR models count data using an over-dispersed Poisson model and use an empirical Bayes' procedure to moderate the degree of overdispersion across genes; the overdispersion reflects the biological variation among samples (Robinson et al., 2009). An implementation of edgeR to mouse brain RNA-Seq data is found in Bottomly et al. (2011).

Second, RNA-Seq data are biased in several important ways. First, the majority of the counts are produced by a small number (<10% of the total) of very highly expressed genes. Thus, many genes of interest may have only moderate to low counts. Also, for genes with equal levels of expression, the long genes will be overrepresented, distorting the relative expression among genes. Similarly, within a given gene, long exons are overrepresented. Normalization and weighting algorithms can be used to address these issues, but they in turn may introduce new biases (Bullard et al., 2010).

Third, RNA-Seq provides a substantial amount of data with very low read counts, which will be quite variable (see, e.g., Cui et al., 2010), and thus, regardless of the analytic strategy, makes detecting DE genes difficult.

Fourth, RNA-Seq data includes multireads, i.e., reads that map equally well to multiple genomic locations. The multireads arise predominantly from conserved domains in paralogous genes and from repeats (Costa et al., 2010). Mortazavi et al. (2008) found that, in the mouse brain, 76% of the 25-bp transcriptome sequence segments uniquely mapped; 6% mapped 2–10 times in the genome; and the remainder mapped more than 10 times. Depending on the gene model used and assuming a high-read density, ignoring the multireads may only have a minimal effect on detecting DE genes. But one can easily contrive a situation involving alternative splicing and multireads where this would not be the case.

Fifth, RNA-Seq collects data across splice junctions that (a) are ignored by many alignment tools and (b) may be unknown. While there are <25,000 known protein-coding genes in the mammalian genome, the number of gene-related transcripts may well be 10–20 times higher (Pan, Shai, Lee, Frey, & Blencowe, 2008; Johnson et al., 2009). Given the heterogeneous nature of brain tissue, the complexity problem is significantly amplified. Tools are available that detect splice junctions and will estimate the minimum number of gene isoforms that account for the observed data (Guttman et al., 2010; Katz, Wang, Airolidi, & Burge, 2010; Trapnell, Pachter, & Salzberg, 2009; Trapnell et al., 2012, 2010). Roberts, Pimentel,



Trapnell, and Pachter (2011) illustrate a procedure that makes use of annotated model organism genomes, such as those available for the laboratory mouse and rat. For both correctly aligning multireads and splice junctions, paired-end sequencing is a useful approach. The downside is the added expense of sequencing the cDNA fragment from both ends.

Sixth, RNA-Seq data can be used to detect allele-specific expression and both synonymous and nonsynonymous SNPs within gene-coding sequences. This application may be particularly useful in complex crosses such as the HS-CC (Iancu et al., 2010) where RNA-Seq can provide detailed genotype information. In the RNA-Seq context, the advantages of using model organisms with a well-annotated genome cannot be underestimated (Martin & Wang, 2011). Reference genome alignment is computationally simpler and faster as the problem is reduced from assembling millions of reads to assembling a much smaller number of reads to known loci. For both the mouse and rat, the reference genomic sequence was obtained using tiled BAC clones, and thus, there are essentially no gaps. But if one believes that there are a substantial number of missed exons, then some combination of reference-based and *de novo* alignment may be the most effective approach (Martin & Wang, 2011). The Mouse Genomes Project (Keane et al., 2011) released genomic sequence data for 17 inbred strains; the data are aligned to the B6 reference strain. It is important to note that these data are not equivalent to the reference genome. The data were acquired using a short-read NGS platform (Illumina), which naturally means that in regions of high repeats/low genetic complexity, it is not possible to correctly align the sequence data. For the standard laboratory strains, this effect is most notable on the proximal aspect of chromosome 7 (Keane et al., 2011). RNA-Seq data are also available for six tissues from a B6D2 F<sub>1</sub> hybrid and for whole brain transcriptome data from 15 strains. These data sets can be freely downloaded and provide an excellent training set for RNA-Seq analysis.

## 6. SEQUENCING THE BRAIN TRANSCRIPTOME

PubMed lists 2702 RNA-Seq publications (6/1/14) with the first appearing in June 2008 (Nagalakshmi et al., 2008); the number has steadily increased from 11 in 2008, to 34 in 2009, to 127 in 2010, to 339 in 2011, to 639 in 2012, and to 1123 in 2013. Of these publications, 162 are also coded as “RNA-Seq and Brain” (~6% of total). However, this number most certainly represents a low estimate of the number of publications where RNA-Seq is used to assess the brain transcriptome or brain surrogates such as induced pluripotent stem cells. Nonetheless, sequencing the brain transcriptome is still an emerging area. The first publication using RNA-Seq to compare brain gene expression between two inbred mouse strains appeared in 2011 (Bottomly et al., 2011). The first application of RNA-Seq to brain WGCNA appeared in 2012 (Iancu et al., 2012). Iancu and colleagues extend this network approach to cosplicing in Chapter 4 building upon the earlier work of Dai, Li, Liu, & Zhou (2012) and Aschoff et al. (2013). Mudge et al. (2008) is an early example of using RNA-Seq in a neuropsychiatric context (schizophrenia) but as noted by Wang and Cairns in Chapter 6 most of the work in this area has appeared within the last 2 years. Chapter 7 details just how quickly our understanding of the functional roles of the ncRNAs has changed due to the introduction of RNA-Seq; further, Guennewig and Copper make compelling arguments for the roles of the ncRNAs in both normal brain function and disease

states. Alternative splicing is higher in the brain as compared to other tissues (Johnson et al., 2009); RNA-Seq facilitates a genome-wide assessment of alternative splicing which is key to understanding both brain development (Dillman and Cookson—Chapter 9) and normal brain function (Zaghlool et al.—Chapter 5). Lewohl et al. (2000) were among the first to use microarrays to study the human brain transcriptome, comparing alcoholics and matched controls. Zhou et al. (Chapter 10) and Farris and Mayfield (Chapter 11) illustrate how readily investigators in the fields of alcoholism and drug abuse research have adopted RNA-Seq to examine human samples. Although still in the preliminary data stage, RNA-Seq is being extensively used to examine the brain transcriptome in nonhuman primates chronically exposed to alcohol (Grant KA, Hitzemann RJ, Darakjian P, & Iancu OD, unpublished observations).

For many investigators, the interest in RNA-Seq and the brain transcriptome is not matched by available funding. Williams and Pandey (Chapter 8) describe a number of freely available mouse resources that allows one to interrogate the relationship(s) between phenotypes and RNA-Seq data. A key element to these resources has been the use of mouse reference populations such as the BXD recombinant inbred series and the Collaborative Cross (Churchill et al., 2004).

RNA-Seq has many applications outside of those mentioned in this volume. One area where it proven to have particular value has been in the examination of the brain transcriptome in nonmodel organisms. Frequently, these organisms have a significant behavioral and/or evolutionary value. A *de novo* assembly of the data can be used in the absence of high-quality genomic sequence data by aligning the reads to conserved protein sequence and/or the annotated genomes of closely related organisms. Four examples are described. Fraser, Weadick, Janowitz, Rodd, and Hughes (2011) assembled brain transcriptome data from the guppy (*Poecilia reticulata*) and were able to detect both sex-specific expression and the effect of predator (*Rivulus hartii*) exposure. Malik et al. (2011) examined the brain transcriptome of blind subterranean mole rat (*Spalax galili*); some modest differences in brain gene expression were found after prolonged exposure to low oxygen concentrations (a normally occurring condition in the underground tunnels). Tzika, Helaers, Schramm, and Milinkovitch (2011) used RNA-Seq in an evolutionary context to compare brain transcriptomes of four divergent reptilian and one reference avian species: the Nile crocodile, the corn snake, the bearded dragon, the red-eared turtle, and the chicken. Somewhat surprisingly, the data suggest that the turtle was evolutionarily closer to the crocodile than was expected. All three of these examples used the Roche 454 platform for sequencing; the longer reads compared with other instruments facilitated the *de novo* transcriptome assembly. Balakrishnan et al. (2014) used RNA-Seq to examine the relationships among the brain transcriptome, avian vocal communication, and social behavior. Brain transcriptomes were sequenced for three emberizid model systems, song sparrow *Melospiza melodia*, white-throated sparrow *Zonotrichia albicollis*, and Gambel's white-crowned sparrow *Zonotrichia leucophrys gambelii*. Each of the assemblies covered fully or in part, over 89% of the previously annotated protein-coding genes in the zebra finch *Taeniopygia guttata*, with 16,846, 15,805, and 16,646 unique BLAST hits in song, white-throated and white-crowned sparrows, respectively. As in previous studies, these

authors found tissue of origin (auditory forebrain versus hypothalamus and whole brain) as an important determinant of the expression profile.

## 7. CONCLUSIONS

Historically, the main arguments against using RNA-Seq (as opposed to using microarrays) have been cost and difficulties with data analysis. Over the past 6 years, technical improvements have and will continue to reduce costs; if the primary goal is gene-wide summarization, transcriptome samples can now be multiplexed and sequenced at adequate depth for less than \$200/sample (not including the cost of library preparation). RNA-Seq data analysis remains substantially more complex than a comparable microarray analysis. The data sets are much larger and are generally not suitable for analysis on a personal computer. While the analysis of RNA-Seq data could still be described as not for the “faint of heart,” a rapidly improving data analysis trajectory is clear as indicated by the numerous reports described in this volume. RNA-Seq has several distinct advantages over microarray-based approaches to transcriptome analysis. RNA-Seq data have a significantly greater dynamic range (there are no probe saturation effects); the gene expression data are not biased to the 3'UTR (although there is a bias to the most highly expressed and longest genes) and data are collected on both alternative splicing and inter- and intragenic ncRNAs. Overall, RNA-Seq embraces the complexity of the transcriptome and provides a mechanism to understand the underlying regulatory code.

## ACKNOWLEDGMENTS

This study was supported in part by grants MH 51372, AA 11034, AA 13484, and a grant from the Veterans Affairs Research Service. The authors want to thank Dr. Kristin Demarest and Kris Thomason for editorial assistance in preparing the review.

## REFERENCES

- Allison DB, Cui X, Page GP, Sabripour M. Microarray data analysis: From disarray to consolidation and consensus. *Nature Reviews. Genetics*. 2006; 7:55–65.
- Aschoff M, Hotz-Wagenblatt A, Glatting KH, Fischer M, Eils R, König R. SplicingCompass: differential splicing detection using RNA-seq data. *Bioinformatics*. 2013; 29:1141–1148. [PubMed: 23449093]
- Balakrishnan CN, Mukai M, Gonser RA, Wingfield JC, London SE, Tuttle EM, et al. Brain transcriptome sequencing and assembly of three songbird model systems for the study of social behavior. *PeerJ*. 2014; 2:e396. [PubMed: 24883256]
- Bebek G, Yang J. Pathfinder: Mining signal transduction pathway segments from protein-protein interaction networks. *BMC Bioinformatics*. 2007; 8:335. [PubMed: 17854489]
- Bergeson SE, Berman AE, Dodd PR, Edenberg HJ, Hitzemann RJ, Lewohl JM, et al. Expression profiling in alcoholism research. *Alcoholism: Clinical and Experimental Research*. 2005; 29:1066–1073.
- Bottomly D, Walter NA, Hunter JE, Darakjian P, Kawane S, Buck KJ, et al. Evaluating gene expression in C57BL/6J and DBA/2J mouse striatum using RNA-Seq and microarrays. *PLoS One*. 2011; 6:e17820. [PubMed: 21455293]
- Buck KJ, Milner LC, Denmark DL, Grant SG, Kozell LB. Discovering genes involved in alcohol dependence and other alcohol responses: Role of animal models. *Alcohol Research: Current Reviews*. 2012; 34(3):367–374. [PubMed: 23134054]

- 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]
- Chen H, Liu Z, Gong S, Wu X, Taylor WL, Williams RW, et al. Genome-wide gene expression profiling of nucleus accumbens neurons projecting to ventral pallidum using both microarray and transcriptome sequencing. *Frontiers in Neuroscience*. 2011; 5:98. [PubMed: 21886604]
- Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, et al. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*. 2005; 308:1149–1154. [PubMed: 15790807]
- Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J. The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nature Genetics*. 2004; 36:1133–1137. [PubMed: 15514660]
- Cirelli C, Tononi G. Differences in brain gene expression between sleep and waking as revealed by mRNA differential display and cDNA microarray technology. *Journal of Sleep Research*. 1999; 8(Suppl. 1):44–52. [PubMed: 10389106]
- Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*. 2008; 322:1845–1848. [PubMed: 19056941]
- Costa V, Angelini C, De Feis I, Ciccodicola A. Uncovering the complexity of transcriptomes with RNA-Seq. *Journal of Biomedicine & Biotechnology*. 2010; 2010:853916. [PubMed: 20625424]
- Cui P, Lin Q, Ding F, Xin C, Gong W, Zhang L, et al. A comparison between ribo-minus RNA-sequencing and poly-a-selected RNA-sequencing. *Genomics*. 2010; 96:259–265. [PubMed: 20688152]
- Dai C, Li W, Liu J, Zhou XJ. Integrating many co-splicing networks to reconstruct splicing regulatory models. *BMC Systems Biology*. 2012; 16:S1–S17.
- Duan F, Pauley MA, Spindel ER, Zhang L, Norgren RB Jr. Large scale analysis of positional effects of single-base mismatches on microarray gene expression data. *BioData Mining*. 2010; 3:2. [PubMed: 20429935]
- Eipper-Mains JE, Eipper BA, Mains RE. Global approaches to the role of miRNAs in drug-induced changes in gene expression. *Frontiers in Genetics*. 2012; 3:109. [PubMed: 22707957]
- Eipper-Mains JE, Kiraly DD, Palakodeti D, Mains RE, Eipper BA, Graveley BR. MicroRNA-Seq reveals cocaine-regulated expression of striatal microRNAs. *RNA*. 2011; 17(8):1529–1543. [PubMed: 21708909]
- Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*. 1998; 8:175–185. [PubMed: 9521921]
- Farris SP, Miles MF. Ethanol modulation of gene networks: Implications for alcoholism. *Neurobiology of Disease*. 2012; 45:115–121. [PubMed: 21536129]
- Feng G, Shaw P, Rosen ST, Lin SM, Kibbe WA. Using the bioconductor gene answers package to interpret gene lists. *Methods in Molecular Biology*. 2012; 802:101–112. [PubMed: 22130876]
- Fraser BA, Weadick CJ, Janowitz I, Rodd FH, Hughes KA. Sequencing and characterization of the guppy (*Poecilia reticulata*) transcriptome. *BMC Genomics*. 2011; 12:202. [PubMed: 21507250]
- Garber M, Grabherr MG, Guttman M, Trapnell C. Computational methods for transcriptome annotation and quantification using RNA-Seq. *Nature Methods*. 2011; 8:469–477. [PubMed: 21623353]
- Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nature Biotechnology*. 2010; 28:503–510.
- Hitzemann RJ, Loh HH. High-affinity GABA and glutamate transport in developing nerve ending particles. *Brain Research*. 1978; 159(1):29–40. [PubMed: 728799]
- Hitzemann R, Reed C, Malmanger B, Lawler M, Hitzemann B, Cunningham B, et al. On the integration of alcohol-related quantitative trait loci and gene expression analyses. *Alcoholism: Clinical and Experimental Research*. 2004; 28:1437–1448.
- Hoffman PL, Miles M, Edenberg HJ, Sommer W, Tabakoff B, Wehner JM, et al. Gene expression in brain: A window on ethanol dependence, neuroadaptation, and preference. *Alcoholism: Clinical and Experimental Research*. 2003; 27:155–168.

- Hofstetter JR, Hitzemann RJ, Belknap JK, Walter NA, Mcweeney SK, Mayeda AR. Characterization of the quantitative trait locus for haloperidol-induced catalepsy on distal mouse chromosome 1. *Genes, Brain, and Behavior*. 2008; 7:214–223.
- Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, Felciano RM, et al. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:17402–17407. [PubMed: 17090670]
- Iancu OD, Darakjian P, Walter NA, Malmanger B, Oberbeck D, Belknap J, et al. Genetic diversity and striatal gene networks: Focus on the heterogeneous stock-collaborative cross (hs-cc) mouse. *BMC Genomics*. 2010; 11:585. [PubMed: 20959017]
- Iancu OD, Kawane S, Bottomly D, Searles R, Hitzemann R, McWeeney S. Utilizing RNA-Seq data for de novo coexpression network inference. *Bioinformatics*. 2012; 28(12):1592–1597. [PubMed: 22556371]
- Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genomewide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*. 2009; 324:218–223. [PubMed: 19213877]
- Johnson MB, Kawasawa YI, Mason CE, Krsnik Z, Coppola G, Bogdanovic D, et al. Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron*. 2009; 62:494–509. [PubMed: 19477152]
- Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, et al. Antisense transcription in the mammalian transcriptome. *Science*. 2005; 309:1564–1566. [PubMed: 16141073]
- Katz Y, Wang ET, Airoidi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods*. 2010; 7:1009–1015. [PubMed: 21057496]
- Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*. 2011; 477:289–294. [PubMed: 21921910]
- Kozell LB, Walter NA, Milner LC, Wickman K, Buck KJ. Mapping a barbiturate withdrawal locus to a 0.44 Mb interval and analysis of a novel null mutant identify a role for Kcnj9 (GIRK3) in withdrawal from pentobarbital, zolpidem, and ethanol. *Journal of Neuroscience*. 2009; 29(37): 11662–11673. [PubMed: 19759313]
- Labaj PP, Leparc GG, Linggi BE, Markillie LM, Wiley HS, Kreil DP. Characterization and improvement of RNA-Seq precision in quantitative transcript expression profiling. *Bioinformatics*. 2011; 27:i383–i391. [PubMed: 21685096]
- Laderas TG, Walter NA, Mooney M, Vartanian K, Darakjian P, Buck K, et al. Computational detection of alternative exon usage. *Frontiers in Neuroscience*. 2011; 5:69. [PubMed: 21625610]
- Lander ES. Array of hope. *Nature Genetics*. 1999; 21(Suppl. 1):3–4. [PubMed: 9915492]
- Lee C, Mayfield RD, Harris RA. Altered gamma-aminobutyric acid type B receptor subunit 1 splicing in alcoholics. *Biological Psychiatry*. 2014; 75:765–773. [PubMed: 24209778]
- Lewohl JM, Wang J, Miles MF, Zhang L, Dodd PR, Harris RA. Gene expression in human alcoholism: microarray analysis of frontal cortex. *Alcohol, Clinical and Experimental Research*. 2000; 24:1873–1882.
- Li Q, Lee JA, Black DL. Neuronal regulation of alternative pre-mRNA splicing. *Nature Reviews. Neuroscience*. 2007; 8:819–831. [PubMed: 17895907]
- Li H, Lovci MT, Kwon YS, Rosenfeld MG, Fu XD, Yeo GW. Determination of tag density required for digital transcriptome analysis: Application to an androgen-sensitive prostate cancer model. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105:20179–20184. [PubMed: 19088194]
- Liao Q, Liu C, Yuan X, Kang S, Miao R, Xiao H, et al. Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. *Nucleic Acids Research*. 2011; 39(9):3864–3878. [PubMed: 21247874]
- Licatalosi DD, Darnell RB. Splicing regulation in neurologic disease. *Neuron*. 2006; 52:93–101. [PubMed: 17015229]

- Lu J, Lee JC, Salit ML, Cam MC. Transcript-based redefinition of grouped oligonucleotide probe sets using AceView: High-resolution annotation for microarrays. *BMC Bioinformatics*. 2007; 8:108. [PubMed: 17394657]
- Lugli G, Smalheiser NR. Preparing synaptoneurosomes from adult mouse forebrain. *Methods in Molecular Biology*. 2013; 936:173–179. [PubMed: 23007508]
- Malik A, Korol A, Hubner S, Hernandez AG, Thimmapuram J, Ali S, et al. Transcriptome sequencing of the blind subterranean mole rat, *Spalax galili*: Utility and potential for the discovery of novel evolutionary patterns. *PLoS One*. 2011; 6:e21227. [PubMed: 21857902]
- Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends in Genetics*. 2008; 24:133–141. [PubMed: 18262675]
- Mardis ER. A decade's perspective on DNA sequencing technology. *Nature*. 2011; 470:198–203. [PubMed: 21307932]
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005; 437:376–380. [PubMed: 16056220]
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-Seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*. 2008; 18:1509–1517. [PubMed: 18550803]
- Martin JA, Wang Z. Next-generation transcriptome assembly. *Nature Reviews. Genetics*. 2011; 12:671–682.
- Matthews DB, Bhawe SV, Belknap JK, Brittingham C, Chesler EJ, Hitzemann RJ, et al. Complex genetics of interactions of alcohol and CNS function and behavior. *Alcoholism: Clinical and Experimental Research*. 2005; 29:1706–1719.
- Mattick JS. The central role of RNA in human development and cognition. *FEBS Letters*. 2011; 585:1600–1616. [PubMed: 21557942]
- Mcbride WJ, Kerns RT, Rodd ZA, Strother WN, Edenberg HJ, Hashimoto JG, et al. Alcohol effects on central nervous system gene expression in genetic animal models. *Alcoholism: Clinical and Experimental Research*. 2005; 29:167–175.
- Metzker ML. Sequencing technologies—The next generation. *Nature Reviews. Genetics*. 2010; 11:31–46.
- Morris J, Singh JM, Eberwine JH. Transcriptome analysis of single cells. *Journal of Visualized Experiments*. 2011; 50:2634. [PubMed: 21540826]
- Mortazavi A, Williams BA, Mccue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*. 2008; 5:621–628. [PubMed: 18516045]
- Mudge J, Miller NA, Khrebtukova I, Lindquist IE, May GD, Huntley JJ, et al. Genomic convergence analysis of schizophrenia: mRNA sequencing reveals altered synaptic vesicular transport in post-mortem cerebellum. *PLoS One*. 2008; 3:e3625. [PubMed: 18985160]
- Mulligan MK, Ponomarev I, Hitzemann RJ, Belknap JK, Tabakoff B, Harris RA, et al. Toward understanding the genetics of alcohol drinking through transcriptome meta-analysis. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:6368–6373. [PubMed: 16618939]
- Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*. 2008; 320:1344–1349. [PubMed: 18451266]
- Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:17973–17978. [PubMed: 17101986]
- Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, et al. Functional organization of the transcriptome in human brain. *Nature Neuroscience*. 2008; 11:1271–1282. [PubMed: 18849986]
- Ozsolak F, Milos PM. RNA sequencing: Advances, challenges and opportunities. *Nature Reviews. Genetics*. 2011; 12:87–98.
- Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*. 2008; 40:1413–1415. [PubMed: 18978789]

- Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitch S, et al. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Research*. 2009; 37:e123. [PubMed: 19620212]
- Pavlidis P, Qin J, Arango V, Mann JJ, Sibille E. Using the gene ontology for microarray data mining: A comparison of methods and application to age effects in human prefrontal cortex. *Neurochemical Research*. 2004; 29:1213–1222. [PubMed: 15176478]
- Peirce JL, Li H, Wang J, Manly KF, Hitzemann RJ, Belknap JK, et al. How replicable are mRNA expression QTL? *Mammalian Genome*. 2006; 17:643–656. [PubMed: 16783644]
- Pozhitkov AE, Boube I, Brouwer MH, Noble PA. Beyond Affymetrix arrays: Expanding the set of known hybridization isotherms and observing pre-wash signal intensities. *Nucleic Acids Research*. 2010; 38:e28. [PubMed: 19969547]
- Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, et al. RNA exosome depletion reveals transcription upstream of active human promoters. *Science*. 2008; 322:1851–1854. [PubMed: 19056938]
- 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]
- Robinson MD, McCarthy DJ, Smyth GK. EdgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2009; 26:139–140. [PubMed: 19910308]
- Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*. 2011; 475:348–352. [PubMed: 21776081]
- Saba LM, Bennett B, Hoffman PL, Barcomb K, Ishii T, Kechris K, et al. A systems genetic analysis of alcohol drinking by mice, rats and men: Influence of brain GABAergic transmission. *Neuropharmacology*. 2011; 60:1269–1280. [PubMed: 21185315]
- Sandberg R, Yasuda R, Pankratz DG, Carter TA, Del Rio JA, Wodicka L, et al. Regional and strain-specific gene expression mapping in the adult mouse brain. *Proceedings of the National Academy of Sciences of the United States of America*. 2000; 97:11038–11043. [PubMed: 11005875]
- Sikela JM, Maclaren EJ, Kim Y, Karimpour-Fard A, Cai WW, Pollack J, et al. DNA microarray and proteomic strategies for understanding alcohol action. *Alcoholism: Clinical and Experimental Research*. 2006; 30:700–708.
- Sliwerska E, Meng F, Speed TP, Jones EG, Bunney WE, Akil H, et al. SNPs on chips: The hidden genetic code in expression arrays. *Biological Psychiatry*. 2007; 61:13–16. [PubMed: 16690034]
- Sugarbaker DJ, Richards WG, Gordon GJ, Dong L, De Rienzo A, Maulik G, et al. Transcriptome sequencing of malignant pleural mesothelioma tumors. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105:3521–3526. [PubMed: 18303113]
- Tabakoff B, Saba L, Printz M, Flodman P, Hodgkinson C, Goldman D, et al. Genetical genomic determinants of alcohol consumption in rats and humans. *BMC Biology*. 2009; 7:70. [PubMed: 19874574]
- Torres TT, Metta M, Ottenwalder B, Schlotterer C. Gene expression profiling by massively parallel sequencing. *Genome Research*. 2008; 18:172–177. [PubMed: 18032722]
- Trapnell C, Pachter L, Salzberg SL. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009; 25:1105–1111. [PubMed: 19289445]
- Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-Seq experiments with TopHat and Cufflinks. *Nature Protocols*. 2012; 7:562–578. [PubMed: 22383036]
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*. 2010; 28:511–515.
- Tzika AC, Helaers R, Schramm G, Milinkovitch MC. Reptilian-transcriptome v1.0, a glimpse in the brain transcriptome of five divergent Sauropsida lineages and the phylogenetic position of turtles. *EvoDevo*. 2011; 2:19. [PubMed: 21943375]
- Vadigepalli R, Chakravarthula P, Zak DE, Schwaber JS, Gonye GE. Paint: A promoter analysis and interaction network generation tool for gene regulatory network identification. *OMICS: A Journal of Integrative Biology*. 2003; 7:235–252. [PubMed: 14583114]

- Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ, Geschwind DH. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*. 2011; 474:380–384. [PubMed: 21614001]
- Walter NA, Bottomly D, Laderas T, Mooney MA, Darakjian P, Searles RP, et al. High throughput sequencing in mice: A platform comparison identifies a preponderance of cryptic SNPs. *BMC Genomics*. 2009; 10:379. [PubMed: 19686600]
- Walter NA, Mcweeney SK, Peters ST, Belknap JK, Hitzemann R, Buck KJ. SNPs matter: Impact on detection of differential expression. *Nature Methods*. 2007; 4:679–680. [PubMed: 17762873]
- Weber AP, Weber KL, Carr K, Wilkerson C, Ohlrogge JB. Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. *Plant Physiology*. 2007; 144:32–42. [PubMed: 17351049]
- Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*. 2005; 4 Article17.
- Zhao W, Langfelder P, Fuller T, Dong J, Li A, Hovarth S. Weighted gene coexpression network analysis: State of the art. *Journal of Biopharmaceutical Statistics*. 2010; 20:281–300. [PubMed: 20309759]