Transcriptome Analysis of the Christianson Syndrome Mouse Model

Jacob Thomas
Senior Honors Thesis
Sc.B Computational Biology
Biological Sciences Track
Advisor: Eric Morrow

April 15th, 2015

Abstract:

Christianson Syndrome is an X-linked neurodevelopmental disorder caused by mutations in SLC9A6. In this study, we present the results of a transcriptome based approach to study a Slc9a6 null mouse model of the disease. In particular, pipelines for the analysis of RNA-Seq data from both mRNA and miRNA are used to identify differentially expressed genes and miRNAs. Overall, 107 differentially expressed genes and 9 differentially expressed miRNAs are identified. Gene set enrichment analyses are performed on the set of 107 differentially expressed genes and a set of 940 predicted target genes of the miRNAs. These sets show enrichment for protein degradation processes and neuronal processes. Additionally, a number of the miRNAs have been shown to affect autophagy and apoptosis.

Introduction

Christianson Syndrome (CS) is an X-linked neurodevelopmental disorder first reported in a large South African pedigree by Christianson *et al* (Christianson et al., 1999; Murtazina, Booth, Bullis, Singh, & Fliegel, 2001; Nakamura, Tanaka, Teko, Mitsui, & Kanazawa, 2005). Male patients with CS have severe intellectual disability, autism, and epilepsy. They exhibit truncal ataxia with an unsteady gait. The majority of patients also have eye movement disorders, sleep problems, and history of microcephaly with delayed trajectories of post-natal brain growth (Pescosolido et al., 2014). Patients with CS are often first diagnosed with Angelman syndrome because of the similar phenotypes. The overlap in phenotypes includes behavioral features such as a happy disposition and unprovoked laughter.

Gilfillan *et al* identified mutations in SLC9A6 as the cause of CS in 2008 (Gilfillan et al., 2008). SLC9A6 is located on the X chromosome and encodes the protein NHE6. NHE6 is a sodium/hydrogen exchanger found on the endosomal membrane. Gilfillan *et al* identified a deletion of six base pairs in SLC9A6 within their pedigree. These bases encode two highly conserved amino acid residues in the exchanger domain of NHE6. These bases have been shown to be essential for ion transport in NHE1 and NHE8, two proteins highly similar to NHE6 (Fafournoux, Noel, & Pouyssegur, 1994; Murtazina et al., 2001; Nakamura et al., 2005). In other families, Gilfillan *et al* were able to identify different mutations in SLC9A6 including a premature stop codon, a splice-site mutation that would cause the skipping of an exon, and a frame shift mutation. Pescosolido *et al* were able to identify additional SLC9A6 mutations in a larger cohort. In particular, they observed 9 SNVs, 2 indels, and 1 CNV deletion. Of these mutations, two were splice-site mutations and the other ten were protein truncations. A

particularly noteworthy feature of the cohort is that seven of the mutations were *de novo* mutations which contrasted the inherited mutations observed in prior literature.

Because of the wide range of SLC9A6 mutations observed in CS patients, many of which are predicted to be loss of function mutations, Slc9a6 null mice are often used to study the disorder. In 2011, Strømme *et al* demonstrated endosomal-lysosomal dysfunction in this model system (Stromme et al., 2011). In particular, they found GM2 ganglioside accumulation in the late endosomes of mutant mouse neurons. These accumulations are particularly significant because they are similar to accumulations observed in lysosomal storage diseases. Lysosomal storage diseases are a large group of genetic diseases which effect the lysosomal degradative pathway. In particular, two thirds of lysosomal storage diseases affect the brain and neurons. Diseases such as Neiman-Pick C and GM2 gangliosidosis cause intellectual disability (Walkley, 2009).

More recently, Ouyang *et al* characterized the effect of the loss of NHE6 in mouse neurons (Ouyang et al., 2013). In this study, they demonstrated that NHE6 is expressed in growing axonal tracts and fiber tracts which include regions of the hippocampus and corpus callosum. Using cultured mouse hippocampal neurons, they then demonstrated that NHE6 null mice had significantly decreased branching and length in both axons and dendrites. They further demonstrated the direct connection to NHE6 by rescuing the arborization deficits with the expression of a vector containing human NHE6. Looking at the cellular function of NHE6, Ouyang *et al* demonstrated that the protein loss leads to an over acidification of the endosomes and a corresponding decrease in TrkB levels. TrkB is the receptor for BDNF, a signaling protein known to play a role in neuronal arborization (Chao & Lee, 2004; Danzer, Crooks, Lo, &

McNamara, 2002). Finally, Ouyang *et al* provided further evidence of the role of BDNF/TrkB signaling by rescuing the NHE6 null mice's arborization problems by exogenously administering BDNF.

In order to provide a broader molecular understanding of CS and have a larger cellular perspective of the disease, we present a transcriptome based approach to study the disease. The transcriptome includes the set of all transcribed genes. With the advances in sequencing technologies over the last ten years, new methods have been developed that can produce data on a whole transcriptome level. In particular, next generation sequencing has enabled scientists to sequence all the transcribed RNA in a sample in a technique called RNA-seq (Z. Wang, Gerstein, & Snyder, 2009). First, RNA is extracted from all samples. Then, the total RNA is converted to cDNA with a reverse transcriptase. Once a cDNA library is made, it is sequenced using a next-generation platform such as the Illumina NextSeq500. Rather than produce long reads like the traditional Sanger sequencing, these platforms make much shorter reads that are mapped onto a reference genome in order to determine which genes are expressed. By quantifying the number of reads that map to each gene, scientists can not only measure which genes are expressed but also their relative expression levels. These expression levels are used to identify differentially expressed genes from varying experimental conditions. More recently, this technique has been adapted to allow for the study of small RNAs like microRNA(miRNA) (Eminaga, Christodoulou, Vigneault, Church, & Seidman, 2013).

miRNAs are a group of small RNAs that are between 21 and 25 base pairs (He & Hannon, 2004). They were first reported in *C. elegans* but have been found in a large range of species including both plants and mammals (Chalfie, Horvitz, & Sulston, 1981; He & Hannon,

2004). In particular, miRNAs have been reported in both mice and humans (Pasquinelli et al., 2000). miRNAs have been linked to post-transcriptional gene repression and bind to sites in the 3' untranslated regions of mRNAs (Chen & Rajewsky, 2007). The 21-25 base pair mature miRNA are derived from longer precursor miRNAs. These precursors form a hairpin RNA loop which is bound by a RNase protein like Dicer. This protein cleaves the precursor into two mature miRNAs with one miRNA from each side of the loop (He & Hannon, 2004). When naming a miRNA, the precursor miRNA is assigned a number, and then based on which side of hairpin loop the mature miRNA is derived, the miRNA is called either 5p (5 prime) or 3p (3 prime) (Ambros et al., 2003).

In this study, we present the results from using RNA-seq to analyze both mRNA and miRNA expression in Nhe6 null mouse brains. Differential gene expression is also determined relative to wild-type mice.

Materials

This study used an NHE6 null mouse model with a LacZ-neo cassette inserted into exon 6 of Slc9a6. This mouse will be referred to as the mutant line. The RNA for sequencing was extracted from the hippocampus of post-natal day 19 mice. Five litters of mice were used. The litters were labeled A-E. For each litter, one male mutant mouse and a wild-type brother were selected. Within each litter, the wild-type was labeled sample 1 and the mutant was labeled sample 2. For example, from the first litter, A1 is the wild-type mouse and A2 is the mutant mouse. The mRNA-seq and miRNA-seq were performed on Illumina NextSeq500. The reads were 75 base pair single end reads. The raw reads from the sequencing were returned in FASTAQ format. A more detailed description of the RNA extraction and sequencing library

generation is described in the methods.

For all analyses performed, the *Mus musculus* reference genome used was mmu10. This was downloaded from UCSC genome browser with the data provided by the Genome Research Consortium (Waterston et al., 2002).

All computations were performed using the resources at the Center for Computation and Visualization, Brown University. Specifically, they were conducted using 4 cores and 16 gigabytes of memory.

Methods

Poly-A Selection and Sequencing Library Generation

Total RNA was processed for library construction according to the following procedure. Oligo-dT beads were used to select poly-adenylated RNA (poly-A). Poly-A selected RNA was sheared to appropriate size for cDNA synthesis. Double-stranded cDNA was end-repaired and a-tailed to prepare for adapter ligation. Indexed adapters were ligated to sample cDNA, and the adapter-ligated cDNA was then size-selected on a 2% SizeSelect™ E-Gel (Invitrogen, Carlsbad, CA) and amplified by PCR. Library size and quality was assessed on an Agilent Bioanalyzer and library yield was quantified by qPCR using the Kapa Biosystems library quantification kit (Wilmington, MA) prior to sequencing on the Illumina NextSeq500 (San Diego, CA) as 1x75 base reads following manufacturer's protocols.

microRNA Extraction

Total RNA was processed for library construction according to the following procedure. Briefly, a 3’ adapter was ligated to the RNA molecules. After ligation, a RT primer was hybridized to the adapter. The 5’ adapter was then ligated to total RNA, followed by reverse

transcription of RNA in a 1st strand synthesis reaction. After 1st strand cDNA synthesis, adapter ligated molecules were amplified with adapter specific primers to enrich for the miRNA fraction of total RNA. Amplified material was then purified and size selected on a 6% polyacrylamide gel. Library quality was assessed by measuring concentration with Qubit (ThermoFisher Scientific, Waltham, MA), and average library size was measured using a Bioanalyzer (Agilent Technologies, Santa Clara, CA).

mRNA analysis

To analyze the mRNA data, an approach similar to that taken in Gamsiz *et al* was used (Gamsiz, Ouyang, Schmidt, Nagpal, & Morrow, 2012). The actual pipeline used for analysis in this study is outlined in Figure 1.

Before beginning the actual expression analysis, the quality of the reads was checked using FastQC (version 0.11.4) (Andrews, 2010). In particular, the percentage of reads with Phred quality scores greater than 30 was checked for each sample. Contamination from primer sequences was also looked at.

Following the quality control step, Bowtie2 (version 2.2.5) is used to align the reads to a reference genome (Langmead & Salzberg, 2012). Bowtie2 aligns reads in a multi-step process which allows for accurate and efficient alignment. First, the reference genome is indexed to allow for rapid searching. Once the index is constructed, each read is individually mapped to the genome in a two-part alignment. For the first step in the alignment, the reads are mapped in an ungapped manner. This initial alignment is used as a seed for the final alignment. With this seed, each read is then subjected to a secondary search for gaps. These two steps allow the genome to be rapidly mapped with the memory benefits of an indexed system while managing the

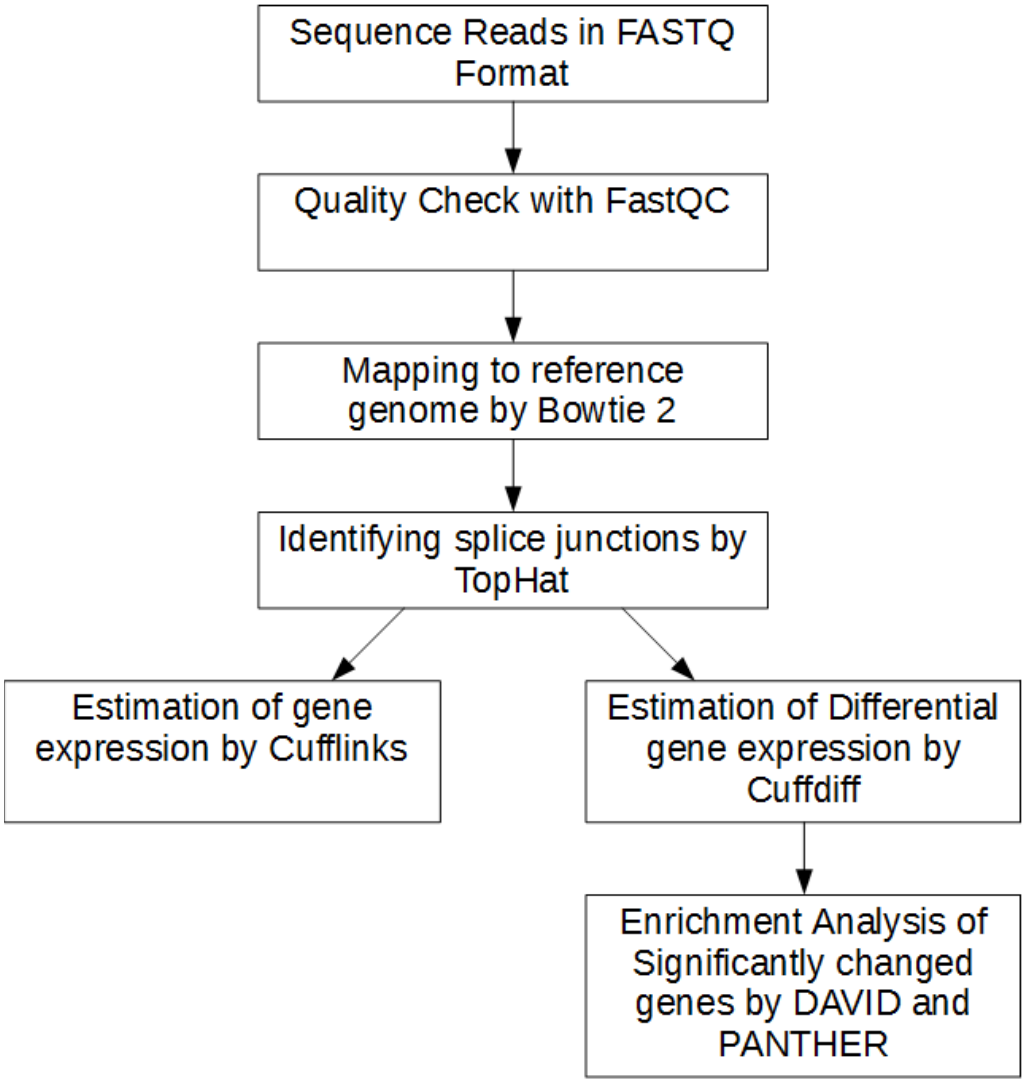


Figure 1: Analysis Pipeline for mRNA data

possibility of gaps.

Once Bowtie2 produces the alignment of reads, TopHat (version 2.1.0) is run (Trapnell, Pachter, & Salzberg, 2009). TopHat is a program that identifies splice-site junctions. It relies upon Bowtie2's alignment. It first uses the alignment to identify the exons based on coverage from the aligned reads that do not have gaps. When the exons are identified, Tophat searches their sequences for splice donor and acceptor sites. Based on these predicted sites, Tophat uses

the gapped reads to identify which sites are actually covered by the reads. This coverage can be used to determine the novel and known splice sites.

With the reads aligned and the splice-junctions identified, the expression levels of genes and isoforms can be measured. Cufflinks (version 2.2.0) is used to estimate the expression levels in each individual sample (Trapnell et al., 2010). To measure isoform expression levels, Cufflinks takes a set of overlapping fragments and determines fragments which are mutually incompatible, meaning they come from different mRNA isoforms. These fragments are used to determine the distinct isoforms. With the isoforms determined, Cufflinks estimates their expression levels using a linear model to determine the probability that a fragment came from a specific isoform. For total gene expression, the relative levels of the isoforms are collapsed together and one level of expression per gene is output.

The alignments are also used in a companion program to Cufflinks called Cuffdiff (version 2.2.0) (Trapnell et al., 2013). Cuffdiff determines the genes and isoforms that are differentially expressed between multiple sample states. In our samples, we had two groups, the wild-type mouse mRNA samples and the mutant mRNA. Rather than simply comparing the total number of reads that map to a given gene, Cuffdiff works by modeling the over dispersion of the alignment of fragments. A beta negative binomial model determines the levels of gene expression across the replicates in each state. With this measure for each state, the two states are statistically tested and corrected for differential gene expression. Cuffdiff reports relative expression level for every gene in each state, corrected p-values, and the fold-change of each gene.

Genes with a corrected p-value less than 0.05 were determined to be significantly differentially expressed. These significantly differentially expressed genes were analyzed for

functional enrichment using two different programs. For both programs, the genes were analyzed in three groups, all significant genes, all significantly upregulated genes, and all significantly down regulated genes. The first program used was DAVID (Database for Annotation, Visualization, and Integrated Discovery)'s gene annotation tool (Huang da, Sherman, & Lempicki, 2009). This tool uses a modified Fisher's exact-test to determine if the set of genes is significantly enriched for a given term compared to the background population. DAVID draws its functional annotations from a number of databases including KEGG's pathways (Kanehisa, Sato, Kawashima, Furumichi, & Tanabe, 2016).

The second functional analysis tool used is the PANTHER (protein annotation through evolutionary relationship) enrichment test (Mi, Muruganujan, Casagrande, & Thomas, 2013). Similar to DAVID, it also compares the given input list to the population to search for significantly enriched terms. Instead of using a modified Fisher's T-test, it relies upon a Mann-Whitney test. The majority of annotations are from the Gene Ontology (GOTerms) database. PANTHER protein class annotations are also used. For both of DAVID and PANTHER outputs, a p-value or FDR threshold of 0.05 was used to determine significance.

miRNA

A modified version of the Cap-miRSeq pipeline was used to analyze the miRNA data (Sun et al., 2014). This adapted analysis pipeline is outlined in Figure 2.

Similar to the mRNA analysis, the quality of the reads is initially checked using FastQC. Unlike the RNA reads, however, the miRNA raw reads include the Illumina adapter tags used in their sequencing. These tags are still included because of the short and variable lengths that occur

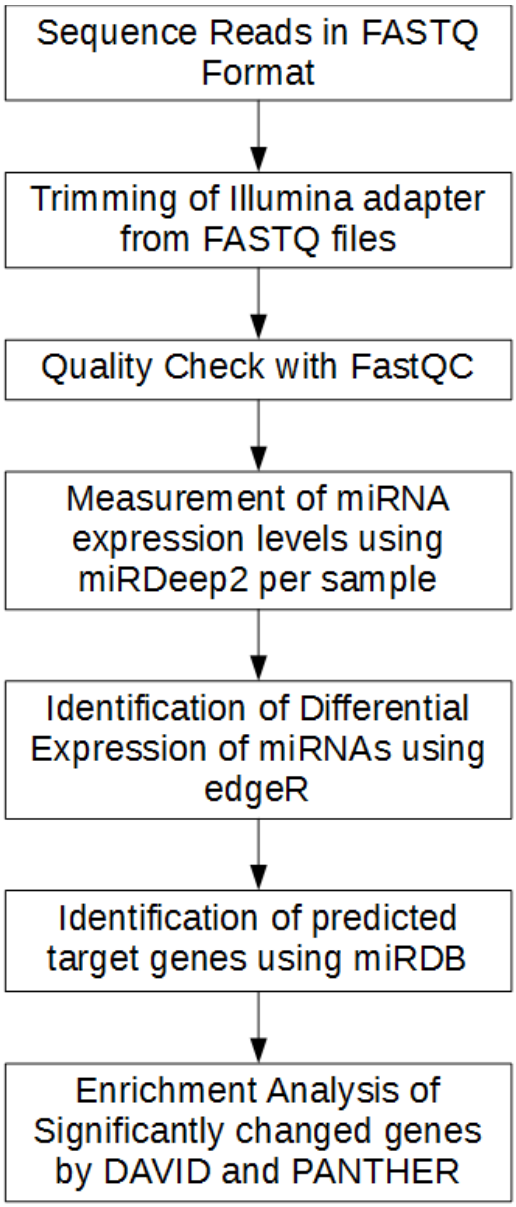


Figure 2: Analysis Pipeline for miRNA data. Adapted from cap-miRSeq pipeline

when sequencing miRNA.

The cutadapt program (version 1.8.3) is used to remove these tags (Martin, 2011). Cutadapt aligns the provided adapter sequence to each read and trims the overlapping region. If

this region is in the middle of the read as opposed to either the 5' or 3' end, cutadapt discards the read.

After the reads are trimmed, FastQC is once again run to confirm that the adapter sequence is no longer present and that the reads have an appropriate distribution around 22 base pairs. 22 base pairs is the expected mean of a miRNA library (Sun et al., 2014).

The reads are then input into miRDeep2 (version 2.0.0.5) (Friedlander, Mackowiak, Li, Chen, & Rajewsky, 2012). miRDeep2 is a program used to identify both new and already known miRNAs. miRDeep2 first uses Bowtie (version 1.1.2) to map the reads to the reference genome (Langmead & Salzberg, 2012; Langmead, Trapnell, Pop, & Salzberg, 2009). Bowtie works in a very similar manner to Bowtie2 but is considered better for the short reads of miRNA data ((Sun et al., 2014). Additionally, Bowtie2's 2-step alignment where an initial seed alignment is used before identifying gaps does not make sense when studying miRNAs and their precursors because they do not undergo splicing.

After aligning the reads to the genome, miRDeep2 quantifies the miRNA expression using the locations of known miRNAs. In particular, miRDeep2 quantifies the levels of both mature miRNAs and precursor miRNA from miRBase. Once the known reads are quantified, miRDeep2 also predicts novel miRNAs by identifying high coverage regions that match the structure of miRNAs.

The known miRNAs are identified using miRBase version 21 (Griffiths-Jones, 2004; Griffiths-Jones, Grocock, van Dongen, Bateman, & Enright, 2006; Griffiths-Jones, Saini, van Dongen, & Enright, 2008; Kozomara & Griffiths-Jones, 2011, 2014). miRBase is a database that provides annotated information on miRNAs from over 80 species. For *Mus musculus*, miRBase

v21 has information on 1915 miRNAs.

After expression levels are quantified for each miRNA, EdgeR is used to test for differential expression between the wild-type and mutant mouse miRNAs (Robinson, McCarthy, & Smyth, 2010). EdgeR is an R package that is part of Bioconductor. EdgeR accounts for the variability between samples by using an over-dispersion Poisson model to determine the distribution of counts across the miRNAs. With this over-dispersion modeled, it uses an empirical Bayes procedure to determine a consensus value for each state. Using these consensus values, a modified Fisher's T-test is then used to determine whether there is a statistically significant difference between the two states. Using an FDR cutoff of 0.05, a list of significant differentially expressed miRNAs is determined.

With these miRNAs identified, predicted targets for each miRNA were identified using miRDB (Wong & Wang, 2015). miRDB is an online database that includes predicted gene targets for a large set of miRNAs from five species including mice and humans. The predictions are generated using the mirTarget algorithm which relies on machine learning techniques in order to identify target sequences in the 3' UTRs of genes. Based on the author's recommendation that genes with scores greater than 80 were highly likely to be actual targets, a cutoff of 80 was used in selecting the genes to consider as targets for each miRNA(X. Wang & El Naqa, 2008).

Once a set of predicted target genes was identified, gene enrichment analysis was performed in a similar manner to that taken for the mRNA. In particular, both DAVID and PANTHER were used to analyze the set of all predicted gene targets of the differentially expressed miRNAs.

Results

mRNA

The FastQC quality control check did not report any problems for the mRNA data. In particular, no contaminating tags were detected, the distribution of sequence lengths was centered on 75, and the majority of reads had phred scores greater than 30. A representative set of these graphs is included for sample A1 in Figure 3. Because of the high similarity between each of the FastQC results, only the results for A1 are shown.

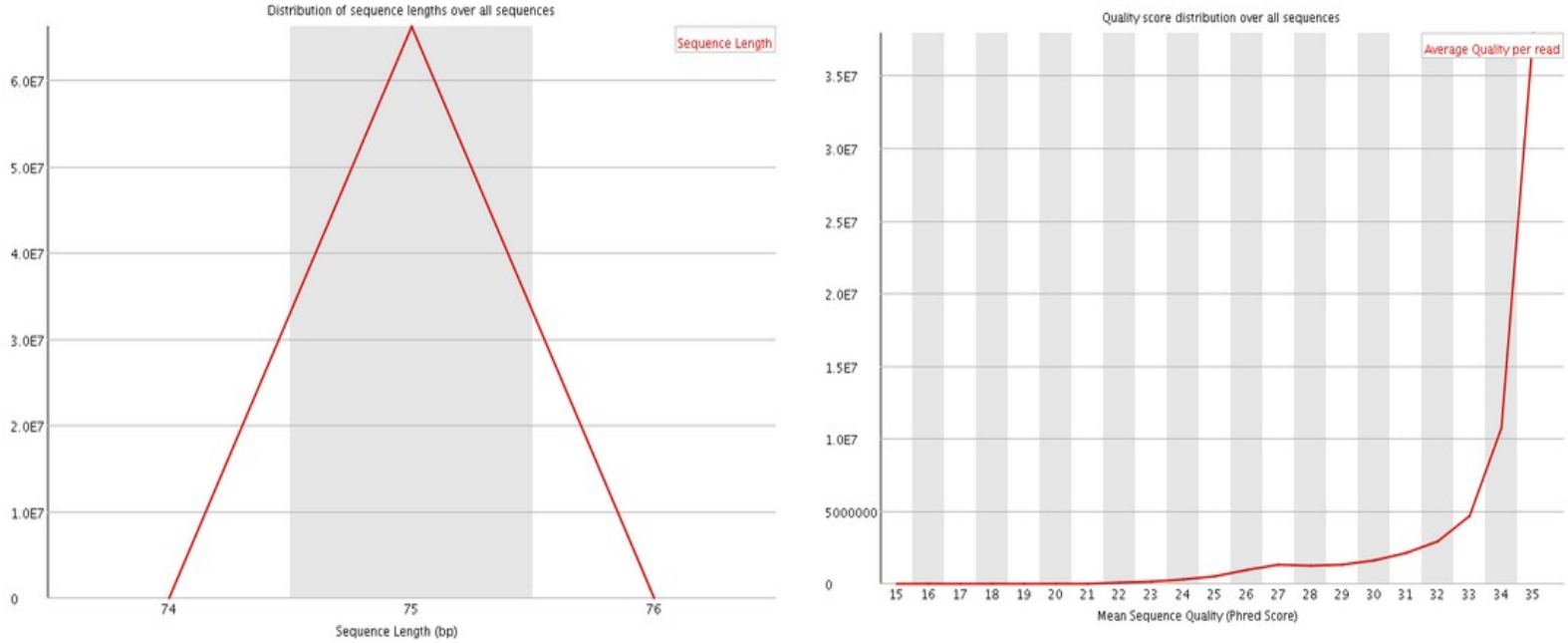


Figure 3: Representative graphs of the results from FastQC for the mRNA fastq file taken from sample A1. Note the high concentration of read lengths centered on 75 bp and the high proportion of reads with Phred scores greater than 30.

The majority of reads mapped to one locus in the genome with a small percentage aligned to multiple loci. The results of the alignment are summarized in Table 1.

Table 1: Summary of mRNA Alignment

Sample	Total Reads	Reads Aligned	Reads with Multiple Alignments	Percentage aligned
A1	66184522	59147564	5994721	0.8936766817
A2	64076289	57390011	6213762	0.8956512915
B1	72003499	64129517	6053677	0.8906444533
B2	68597238	61613192	4934969	0.8981876501
C1	64465291	58057448	5200777	0.9006001074
C2	65050767	58839833	6212492	0.9045217407
D1	65868484	60361604	4725871	0.916395829
D2	70372827	64172217	5047231	0.9118891444
E1	60082666	55119257	3622524	0.9173903335
E2	66234939	60933462	4031545	0.9199595096

Using these alignments, Cuffdiff reported 107 genes that were differentially expressed at a significant level. Of these genes, 87 of the genes were downregulated and 20 were upregulated. These results are summarized in Tables 2 and 3. In addition to the p-values, the table also includes the fold-change and average FPKM values for the wild-type and mutant. Included in the downregulated genes was Slc9a6. This result further confirms the mutant mouse model.

Cuffdiff also identified 89 total isoforms of genes that were differentially expressed. Of these isoforms, 72 were down regulated and 17 were upregulated. Many of these results were isoforms of genes that were also reported in the 107 significant genes already discussed. Because of this significant overlap, these isoforms are not shown here or further discussed.

After identifying the genes, DAVID and PANTHER were run on the combined set of 107 genes as well as the set of 87 downregulated genes and 20 upregulated genes. For PANTHER's enrichment test, 14 processes were identified for the combined set, 14 processes for the down-regulated set, and 2 processes for the upregulated set. For the DAVID results, 25 processes were identified for the combined set, 28 processes for the down-regulated set, and 2 processes for the upregulated set. Selected processes and related information are included in Tables 4, 5, and 6.

Table 2: Significantly Down Regulated Genes

Gene Name	q-value(FDR corrected p-value)	Wild-type Expression (FPKM)	Knockout Expression (FPKM)	Fold Change
Defb9	0.00873111	1.84416	0	0
Olfr570	0.00873111	0.571564	0	0
Pon1	0.00873111	0.492252	0	0
Ttr	0.00873111	2453.17	26.6093	0.0108469042
Prr32	0.00873111	7.10328	0.121435	0.0170956234
Slc4a5	0.00873111	1.76552	0.0380048	0.0215261226
Kcne2	0.00873111	5.12479	0.131308	0.0256221231
Aqp1	0.00873111	5.14732	0.207559	0.0403237024
Tmem72	0.00873111	1.53978	0.0662104	0.0429999091
Clic6	0.00873111	6.07037	0.277462	0.0457075928
Mfrp	0.00873111	2.43242	0.159272	0.0654788236
Wfdc2	0.00873111	3.5252	0.253005	0.071770396
Cdh3	0.00873111	0.28471	0.0215898	0.0758308454
Col8a1	0.00873111	2.47165	0.226053	0.0914583375
Slc9a6	0.00873111	24.3532	2.4047	0.0987426704
Cldn2	0.00873111	2.87168	0.287447	0.1000971557
F5	0.00873111	1.24305	0.125615	0.1010538595
Folr1	0.00873111	11.3632	1.1607	0.1021455224
1500015O10Rik	0.00873111	59.147	6.04914	0.1022729809
Otx2	0.00873111	1.99245	0.218956	0.1098928455
Steap1	0.00873111	3.35355	0.590934	0.1762114774
Abca4	0.00873111	0.625264	0.11339	0.1813473989
Wdr86	0.00873111	1.6629	0.310004	0.1864237176
Tc2n	0.00873111	0.213029	0.0486782	0.2285050392
Trpv4	0.00873111	1.4422	0.341759	0.2369706005
Rbm47	0.00873111	0.589584	0.147038	0.2493927922
Pla2g5	0.00873111	1.53002	0.392534	0.2565548163
Krt18	0.00873111	2.15579	0.561016	0.2602368505
Wfikkn2	0.00873111	1.68395	0.475023	0.2820885418
Cox8b	0.00873111	11.3391	3.22794	0.2846733868
Prlr	0.00873111	1.76093	0.528575	0.300168093
Col4a3	0.00873111	0.273585	0.0913488	0.3338954987
Krt8	0.00873111	1.25459	0.424009	0.3379661882
Kl	0.00873111	7.14049	2.49345	0.3491987245
Enpp2	0.00873111	261.426	91.6439	0.3505538852
Crb3	0.00873111	1.09097	0.391731	0.3590667021
Sulf1	0.00873111	4.28951	1.54786	0.3608477425
Scara5	0.00873111	0.598568	0.221106	0.3693916147
Fap	0.00873111	0.59792	0.227249	0.3800658951
S100a9	0.00873111	7.84036	3.14622	0.4012851451
Trpm3	0.00873111	12.7078	5.45565	0.4293150663
Sostdc1	0.00873111	6.3359	2.73133	0.4310879275
Col8a2	0.00873111	3.18895	1.39114	0.4362376331
Igfbp2	0.00873111	70.6672	30.9927	0.4385726334
Slc16a8	0.00873111	1.31587	0.590409	0.4486833806
Oca2	0.00873111	0.980529	0.440996	0.4497531435
St6galnac2	0.00873111	1.70791	0.790046	0.4625805809

Table 2: Continued

Gene Name	q-value(FDR corrected p-value)	Wild-type Expression (FPKM)	Knockout Expression (FPKM)	Fold Change
Ppp1r3b	0.00873111	0.572651	0.26505	0.4628473538
Serpinb1b	0.00873111	1.15336	0.537629	0.4661415343
S100a8	0.00873111	8.41305	3.9435	0.4687360707
Sema3b	0.00873111	1.5135	0.717894	0.4743270565
Tuba1c	0.00873111	1.27791	0.638096	0.4993278087
Mdfic	0.00873111	1.47163	0.747738	0.5081019006
Hprt	0.00873111	127.137	66.7769	0.5252357693
Ace	0.00873111	5.20419	2.74333	0.5271387094
Col4a4	0.00873111	0.308909	0.163361	0.5288321156
Epn3	0.00873111	0.994995	0.526672	0.5293212529
Frem1	0.00873111	0.820474	0.435238	0.5304714104
Rdh5	0.00873111	3.82925	2.03212	0.5306835542
Lbp	0.00873111	6.27045	3.36398	0.5364814328
Slc13a4	0.00873111	3.65807	1.99318	0.5448720227
Cd59a	0.00873111	4.7992	2.69777	0.5621291049
Cgnl1	0.00873111	2.06613	1.1774	0.5698576566
Lepr	0.00873111	1.21812	0.696717	0.5719608906
Cldn1	0.00873111	2.63739	1.52488	0.5781776681
Slc16a4	0.00873111	2.47973	1.4609	0.5891367205
Atp11c	0.00873111	2.92949	1.74061	0.5941682682
Pcolce2	0.00873111	2.97307	1.77096	0.595667105
Pcolce	0.00873111	5.44896	3.25164	0.5967450669
Mia	0.00873111	21.7837	13.0622	0.5996318348
Igf2	0.00873111	37.4565	22.4894	0.6004138134
Msx1os	0.00873111	4.40595	2.65501	0.6025964888
Car12	0.00873111	11.5183	6.95366	0.6037054079
Cab39l	0.00873111	12.1899	8.11763	0.6659308116
Erdr1	0.00873111	124.784	90.1315	0.7223001346
Mir124a-1hg	0.00873111	18.0786	13.2699	0.7340114832
Cyp2a5	0.0162021	0.509148	0.11575	0.2273405768
Sfrp5	0.0162021	0.564017	0.162421	0.2879718165
Eps8l2	0.0162021	0.80574	0.346005	0.4294251247
Pon3	0.0162021	2.07807	1.20268	0.5787485503
Drc7	0.0162021	2.00461	1.23986	0.6185043475
Vat1l	0.0162021	13.8293	10.4569	0.7561409471
Slc31a1	0.0162021	12.4639	9.46211	0.7591612577
Slco1a5	0.0377788	0.283657	0.0987089	0.3479868292
Itpripl1	0.0440636	0.906185	0.549616	0.6065163294
Slc37a2	0.0440636	1.02852	0.665212	0.6467662272
Spint2	0.0440636	11.8404	8.49163	0.7171742509

Table 3: Significantly Upregulated Genes

Gene Name	q-value(FDR corrected p-value)	Wild-type Expression (FPKM)	Knockout Expression (FPKM)	Fold Change
Oasl2	0.00873111	1.3194	2.57167	1.9491268705
Isg15	0.00873111	3.01847	5.72669	1.8972134483
Ifit3	0.00873111	3.02379	5.44665	1.8012694392
Usp18	0.00873111	1.21586	2.17731	1.7907575195
Cplx3	0.00873111	3.04955	5.25358	1.722737632
Ifit1	0.00873111	1.54852	2.52123	1.6281583288
Nxph3	0.00873111	16.9302	24.6002	1.4530395783
Tmem125	0.00873111	11.9661	16.7294	1.3980655453
Ctgf	0.00873111	6.07622	8.38505	1.3799777443
Gsn	0.00873111	112.269	149.799	1.334288833
Mog	0.00873111	120.491	159.292	1.3220291612
Pdlim2	0.00873111	41.4536	54.7064	1.3197036525
Fa2h	0.00873111	32.5488	42.9253	1.3187983618
Gjc2	0.00873111	18.3144	24.1174	1.3168527199
Adssl1	0.023574	9.36838	12.8614	1.3728499974
Gjb1	0.023574	8.6858	11.7043	1.3475212276
Sox10	0.023574	34.7319	45.0903	1.2982398985
Tmem52	0.0308157	1.82198	3.22739	1.7713655913
Hba-a1	0.0308157	63.4219	84.7613	1.3364677272
Mal	0.0377788	196.956	261.404	1.327216751

Table 4: Selected Pathway Analysis Results for Downregulated mRNAs

DAVID Results

Process	FDR
Ion Transport	0.000933
Antimicrobial	0.0365

PANTHER Results

Process	FDR adjusted p-value
regulation of peptidase activity	0.00992
transmembrane transport	0.0368
regulation of proteolysis	0.0374

Table 5: Selected Pathway Analysis Results for Upregulated mRNAs

DAVID Results

Process	FDR
Interferon-induced 56k Protein	0.0278

PANTHER Results

Process	FDR adjusted p-value
Defense Response to virus	0.0211

Table 6: Selected Pathway Analysis Results for All Differentially Expressed mRNAs

DAVID Results

Process	FDR
ion transport	0.0064
Cell attachment site	0.0137

PANTHER Results

Process	FDR adjusted p-value
epithelial cell apoptotic process	0.0358
regulation of peptidase activity	0.00122
regulation of cellular protein metabolic process	0.0456

miRNA

In the intial run of FastQC, the Illumina Universal adapter tag was detected as expected. After running cutadapt, FastQC no longer detected the adapter and the reads passed the quality

check with majority of reads having phred scores greater than 30. The number of trimmed reads per sample are reported in Table 7. Additionally, Figure 4, includes the graphs from FastQC for sample A1. This figure includes the graphs from before and after the trimming. In particular, it demonstrates the loss of the adapter sequence and the shift in the length of reads to be center on the expected 22 base pairs. Despite these changes, the high quality of the reads is maintained. Due to the similarity in results, only those for A1 are shown.

Having trimmed the reads, miRDeep2 was run. Within miRDeep2, the majority of reads successfully aligned. For each sample, a large number of miRNAs were detected with greater than 5x coverage. The precise statistics for each sample are provided in Table 7.

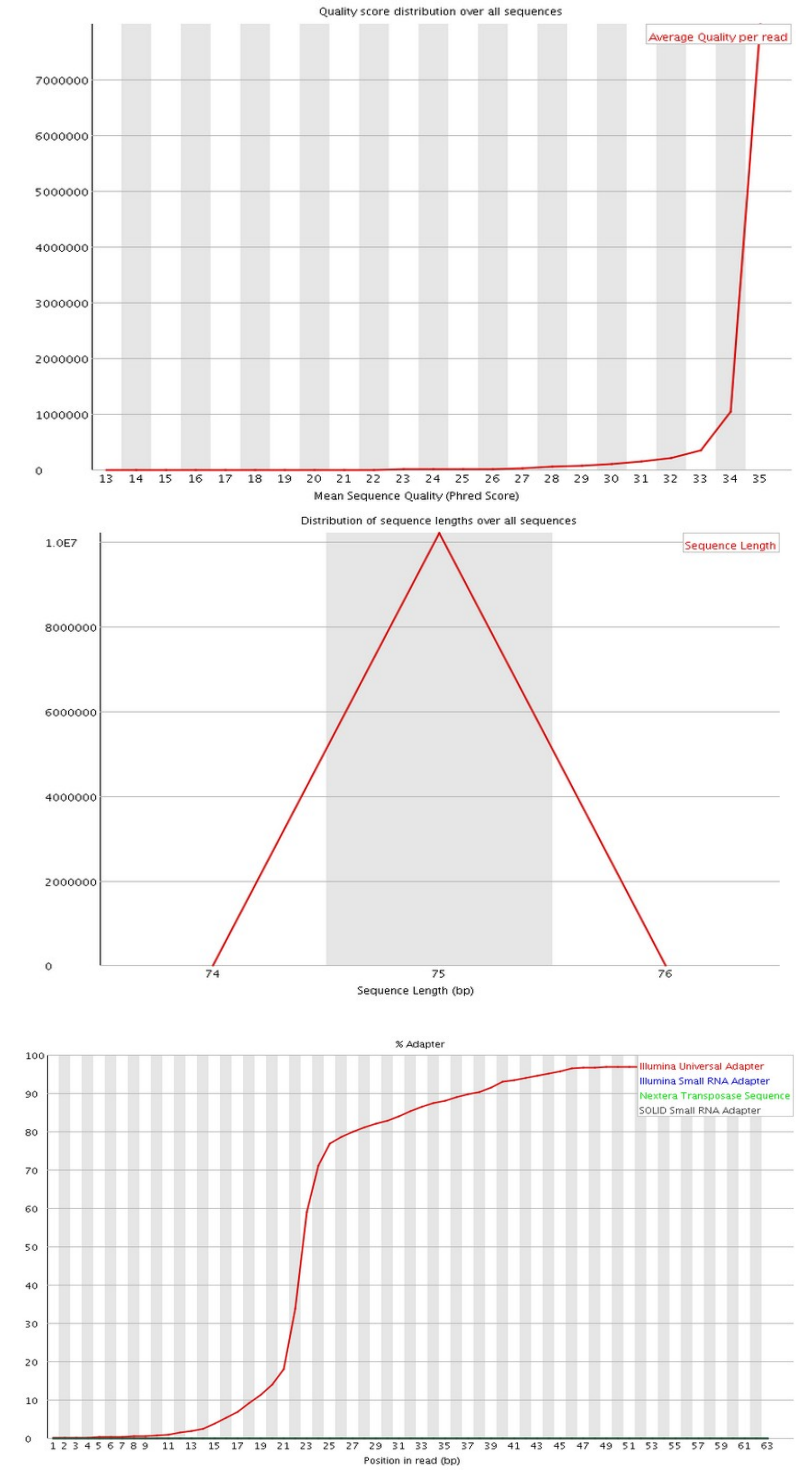
Table 7: Summary of Sample Trimming and Alignment

Sample	Total Reads	Reads After Trim	Reads Aligned	Percentage aligned	miRNAs detected
A1	10207523	9478408	7296168	0.7697672436	765
A2	7926152	7669221	5967508	0.7781113623	709
B1	10541152	9030058	6508754	0.7207876184	744
B2	10144530	9835439	7538065	0.7664187638	742
C1	10533373	9932475	7292271	0.7341846821	771
C2	10063024	9477082	5541726	0.5847502427	646
D1	10969697	10559078	7985380	0.7562573172	785
D2	10471115	9941941	8295560	0.8344004455	785
E1	9486295	9007172	6711919	0.7451749561	741
E2	10205785	9684134	6781778	0.7002978274	751

Using miRDeep2's data on expression levels, edgeR detected 9 differentially expressed miRNAs with an FDR of 0.05 or less. Of these miRNA, 8 were expressed at lower levels in the mutant and 1 was expressed at higher levels. Using miRDB, predicted targets for the 9 miRNAs were identified. Between all of the miRNAs, a total of 940 genes were identified. These miRNAs and their fold-changes are shown in Table 8. A complete list of the predicted target genes for each miRNA is included in the excel spreadsheet supplementary Table S1.

Before Trimming

After Trimming



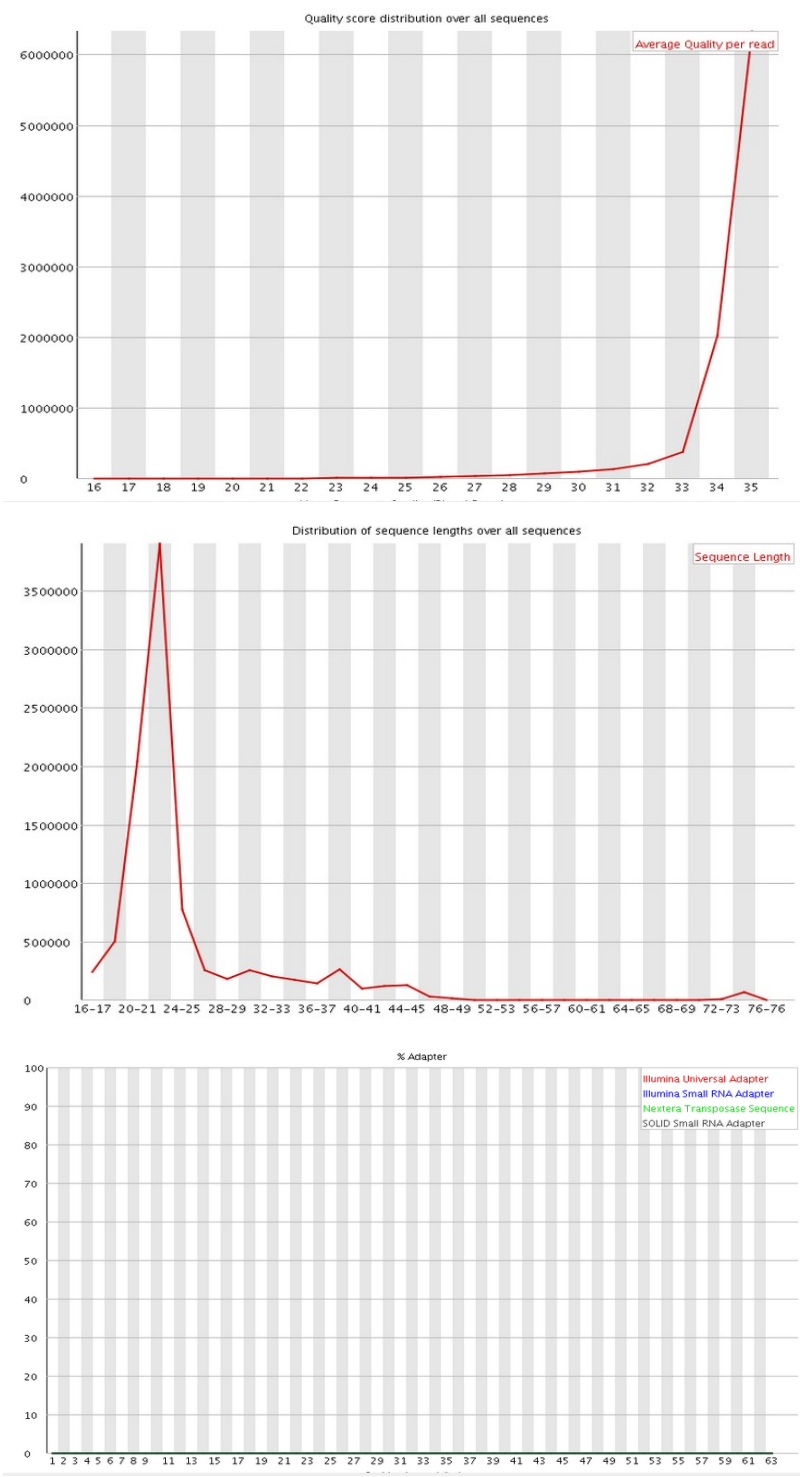


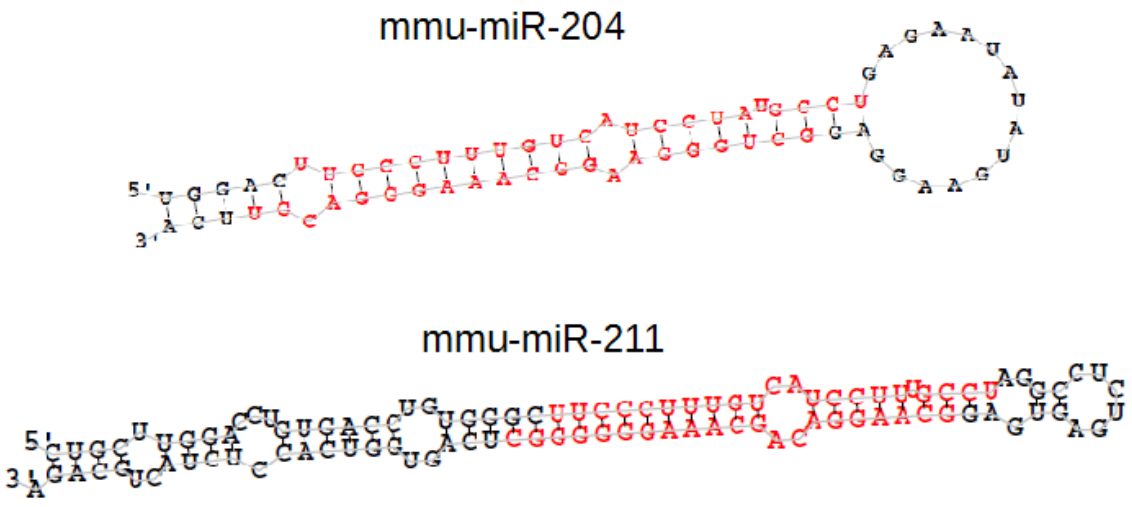
Figure 4: A representative comparison of the FastQC results for the miRNA data before and after the trimming of the adapter sequence for sample A1. The top graph shows the Phred Quality Scores which does not significantly change as expected. The middle graphs show the sequence length distribution which shifts from almost exclusively 75 to centered on 22, the expected mean of a miRNA library. The bottom graphs show the percentage of adapter sequence which in the before is highly contaminated with the Illumina Adapter sequence but shifts to no presence after trimming.

Table 8: Differentially Expressed miRNAs

Gene Name	FDR	Fold Change	Predicted Target Genes
mmu-miR-1298-5p	1.08679E-010	0.059719738	4
mmu-miR-448-3p	1.95083E-008	0.1226571787	159
mmu-miR-204-3p	0.0001663138	0.3144024127	156
mmu-miR-335-3p	0.000915842	2.3046457786	299
mmu-miR-211-5p	0.0078502323	0.4142191885	204
mmu-miR-204-5p	0.0078502323	0.432691189	204
mmu-miR-1264-5p	0.0195547763	0.4657391832	41
mmu-miR-34c-3p	0.0263124376	0.5245821254	41
mmu-miR-1264-3p	0.0430982929	0.4547691266	197

It should be noted that two pairs of the predicted mature miRNAs have the same number of target genes. For mmu-miR-1264-5p and mmu-miR-34c-3p, this match is simply by chance. Their target gene lists are different. For mmu-miR-211-5p and mmu-miR-204-5p, however, the predicted lists are the same. The sequences for the two mature miRNAs only differ by one base. As a result, when predicting miRNAs, the possibility for imperfect binding leads to the same predicted targets (X. Wang & El Naqa, 2008). Despite the similarities in the mature miRNAs, their precursor miRNAs are significantly different. This precursor difference explains the difference in fold change. The reads that align will not always be the same due to reads which overlap the differing precursor sections. These differences are highlighted in Figure 5.

DAVID and PANTHER were run on the set of 940 genes that resulted from combining the lists of predicted genes for each of the differentially expressed miRNAs. DAVID indentified 24 enriched terms with an FDR of 0.05. PANTHER detected 278 enriched terms with an FDR. This large difference can be explained by PANTHER's larger database of terms to search for. Selected enriched pathways are reported in Table 9.



	Mature sequence
mmu-miR-204-5p	uucccuuugucauccu**a**ugccu
mmu-miR-211-5p	uucccuuugucauccu**u**ugccu

Figure 5: A summary of the similarity between the mature sequences but the differences between the precursor sequences. The base that differs in the mature sequences is highlighted in red in the table.

Table 9: Selected Pathway Analysis Results for miRNA Target Genes

DAVID Results

Process	FDR	Genes in Set
transcription regulation	9.03724761691649E-010	133
neurogenesis	0.00002375	23
synapse	0.0039439159	24

PANTHER Results

Process	FDR adjusted p-value	Genes in Set
regulation of primary metabolic process	1.31E-020	401
regulation of gene expression	6.08E-020	269
nervous system development	1.71E-018	169
neuron projection development	7.64E-010	63
neuron differentiation	8.51E-010	85
neuron development	1.57E-009	73
regulation of neurogenesis	0.000048	64
axon development	0.000128	36

Discussion

Our transcriptome based approach to studying CS has resulted in many potentially interesting results that could be used to direct future research into CS. However, understanding the limitations of this kind of study is important when interpreting these results. First, the RNA used for both the mRNA and miRNA were extracted from the hippocampus of the mice. This region is an appropriate choice as NHE6 has been shown to be expressed in this brain region starting on post-natal day 0 (Ouyang et al., 2013). However, using a hippocampal extract for the sequencing means that the RNA comes from a mixture of all the cell types present in this brain region. This mixture includes both neurons and glia. These cell types have both been shown to express NHE6 but given their different functions, mixing their RNAs may weaken signals unique to a single cell type and hide potentially valuable insights (Ouyang et al., 2013). Studies have increasingly suggested a role for glia in many neurodegenerative diseases including ALS and certain tauopathies, so by including this mixture of cell types, understanding the cell types' unique roles is not possible(Re et al., 2014; Yoshiyama et al., 2007). Beyond the brain region and cell types, the experimental design does provide an advantage in that both the mRNA and miRNA data comes from the same samples.

Another potential confound comes from the use of a mouse model. This model system is particularly noteworthy when considering the role of miRNAs. While miRNAs are often conserved between species, the genes they target may show distinct patterns that differ between species although they are usually similar (Roux, Gonzalez-Porta, & Robinson-Rechavi, 2012; X. Wang & El Naqa, 2008). This means that while the miRNA results presented here may provide potentially valuable insight into CS, they should also be further studied in better model systems

such as patient derived induced pluripotent stem cells.

Despite these limitations, our results still provide some potentially valuable insight in CS. Starting with the mRNA results, some of the gene set enrichment results provide interesting pathways to further investigate. In particular, the two pathways, regulation of peptidase activity and regulation of proteolysis, both appear linked to CS because of Slc9a6's role in the endosomal/lysosomal degradative pathway. It has been shown that Slc9a6 leads to the overacidification of the endosomes (Ouyang et al., 2013). This change in pH's effect is not fully understood, but in non-neuronal cells, pH has been shown to be distinct in different parts of the endosomal/lysosomal system (Prasad & Rao, 2015; Yap & Winckler, 2012). Genes that were reported in these pathways include Ctgf, Aqp1, Pcolce2, Gsn, Serpinb1b, Kcne2, Cd59a, and Col4a3. Ctgf has been reported as a down-stream target of mTor which is upregulated in autophagy (Bernard et al., 2014). Aqp1 is a water channel, Serpinb1b is a serine peptidase inhibitor, and Kcne2 is a potassium gated voltage channel (Pruitt et al., 2014).

Beyond the pathway analysis for the mRNA, a few other genes stand out. Both Igf2 and Igfbp2 are down-regulated. These genes are part of the insulin growth factor signaling process (Brouwer-Visser & Huang, 2015). Igf2 is a peptide that is secreted by most tissues and is known to increase cell growth (Livingstone, 2013). Igfbp2 encodes insulin-like growth fact binding partner 2 which when phosphorylated binds Igf2 and increases its translation (Brouwer-Visser & Huang, 2015). This process has been implicated in regulating cell survival, growth, and metabolism. In particular, it has been shown to be upregulated in cancers (Brouwer-Visser & Huang, 2015). This pathway has also been shown to interact with mTOR and is a compenent in the mTOR signaling pathway (Dai et al., 2011). In particular, mTOR phosphorylates Igfbp2

(Brouwer-Visser & Huang, 2015). mTOR has been demonstrated to be a key regulator of autophagy. mTOR inhibition leads to increased autophagy (Rubinsztein, Codogno, & Levine, 2012). As members of this signaling pathway, the misregulation of Igf2 and Igfbp2 could have significant impacts on the development of the cells and could potentially be involved in cell death. Both Igf2 and Igfbp2 are significantly down-regulated in the observed data. This may suggest a potential decrease in phosphorylation of Igfbp2 by mTOR as phosphorylation increases the levels of Igf2. However, this would need to be studied through more direct biological experiments. Understanding how CS leads to changes in their mRNA levels may provide greater insight into the disease.

The miRNA results also include some interesting results. In particular, miR-35c-3p has been studied as a tumor suppressor that can induce apoptosis (Lopez & Alvarez-Salas, 2011). miR-204-5p has also been shown to be a tumor suppressor that induces apoptosis in human gastric cancers (Sacconi et al., 2012). It also promoted apoptosis in rat Schwann cells (Gao et al., 2014). In addition to apoptosis, miR-204-5p has also been shown to cause autophagy in human renal cancer cells (Hall et al., 2014). Finally, miR-211-5p has also been implicated in apoptosis (Chitnis et al., 2012). Unlike the other two miRNAs discussed above, miR-211-5p actually inhibits apoptosis.

By studying the results from both the miRNA and mRNA data, the loss of Slc9a6 seems to have a significant effect on cell's degradative pathways and potentially on larger cell death processes. These pathways are particularly implicated by the number of differentially expressed miRNAs that have been demonstrated to affect apoptosis and autophagy. Additionally, the mRNA pathway analysis further supports the link to degradative processes with the enrichment

proteins related to proteolysis and peptidase activity, both necessary processes for protein degradation. While this enrichment may be expected to some degree because Slc9a6 is an endosomal protein, the loss of this specific protein may cause broader changes to the degradative system.

The pathway analysis of the predicted target genes for all of the miRNAs also suggests a potentially interesting role for miRNAs in CS. In particular, there is a significant enrichment for genes related to different aspects of neuronal development including axon development, neuron projection development and neurogenesis. Decreased neuronal arborization has been seen in Slc9a6 null cultured mouse neurons (Ouyang et al., 2013). While this paper suggests a mechanism for the decreased arborization as a result of changes to BDNF/TrkB signalling, miRNAs may also play a role in either a compensatory manner or a pathogenic manner. Understanding what role miRNAs play in the neuronal problems related to CS may help provide insight into the disease's pathogenesis. The large number of predicted target genes may also help provide connections to other related diseases and potential shared mechanisms.

In order to improve these results and remove possible uncertainty, this experiment could potentially be reproduced using RNA from patient derived induced pluripotent stem cells (iPSCs). Even if these results were confirmed in iPSCs, it would still be important to conduct further biological experiments to understand what role these miRNAs are playing. Are they pathogenic or are they the cell attempting to respond to the problems caused by the Slc9a6 mutation?

Another future direction that is currently being pursued is looking for potential overlap between this transcriptome data and proteomics data generated from high-throughput mass-

spectrometry. This work could also be potentially expanded to include metabolomics data as well.

Conclusion

Overall, both the mRNA and miRNA data provide potentially valuable starting points for further research into understanding CS. In particular, they serve to highlight the potential importance of understanding what role miRNA regulation plays in CS and the potential role of cell death in the disease. Both data-sets showed differentially expressed genes or miRNAs that were enriched for degradative processes including proteolysis and primary metabolic process. Additionally, genes were identified from the RNA data that are part of the insulin-like growth factor signaling pathway which is directly tied to mTOR and autophagy. A number of the miRNAs have also been demonstrated to regulate autophagy and apoptosis. These results may help guide further study of CS.

Acknowledgments

I would like to thank my advisor Eric Morrow for all of his help, support, and guidance throughout this project, Ece Gasmiz for working closely with me and providing constant support and guidance, Qing Ouyang for her work and data that were crucial to this project, the rest of the Morrow lab for all the support over the last year, Brown's Center for Computation and Visualization, and my second reader Louis Lapierre.

References

Ambros, V., Bartel, B., Bartel, D. P., Burge, C. B., Carrington, J. C., Chen, X., . . . Tuschl, T. (2003). A uniform system for microRNA annotation. *Rna, 9*(3), 277-279.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc: Babraham's Bioinformatics.

Bernard, M., Dieude, M., Yang, B., Hamelin, K., Underwood, K., & Hebert, M. J. (2014). Autophagy fosters myofibroblast differentiation through MTORC2 activation and downstream upregulation of CTGF. *Autophagy, 10*(12), 2193-2207. doi:10.4161/15548627.2014.981786

Brouwer-Visser, J., & Huang, G. S. (2015). IGF2 signaling and regulation in cancer. *Cytokine Growth Factor Rev, 26*(3), 371-377. doi:10.1016/j.cytogfr.2015.01.002

Chalfie, M., Horvitz, H. R., & Sulston, J. E. (1981). Mutations that lead to reiterations in the cell lineages of C. elegans. *Cell, 24*(1), 59-69.

Chao, M. V., & Lee, F. S. (2004). Neurotrophin survival signaling mechanisms. *J Alzheimers Dis, 6*(6 Suppl), S7-11.

Chen, K., & Rajewsky, N. (2007). The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet, 8*(2), 93-103. doi:10.1038/nrg1990

Chitnis, N. S., Pytel, D., Bobrovnikova-Marjon, E., Pant, D., Zheng, H., Maas, N. L., . . . Diehl, J. A. (2012). miR-211 is a prosurvival microRNA that regulates chop expression in a PERK-dependent manner. *Mol Cell, 48*(3), 353-364. doi:10.1016/j.molcel.2012.08.025

Christianson, A. L., Stevenson, R. E., van der Meyden, C. H., Pelser, J., Theron, F. W., van Rensburg, P. L., . . . Schwartz, C. E. (1999). X linked severe mental retardation, craniofacial dysmorphology, epilepsy, ophthalmoplegia, and cerebellar atrophy in a large South African kindred is localised to Xq24-q27. *J Med Genet, 36*(10), 759-766.

Dai, N., Rapley, J., Angel, M., Yanik, M. F., Blower, M. D., & Avruch, J. (2011). mTOR phosphorylates IMP2 to promote IGF2 mRNA translation by internal ribosomal entry. *Genes Dev, 25*(11), 1159-1172. doi:10.1101/gad.2042311

Danzer, S. C., Crooks, K. R., Lo, D. C., & McNamara, J. O. (2002). Increased expression of brain-derived neurotrophic factor induces formation of basal dendrites and axonal branching in dentate granule cells in hippocampal explant cultures. *J Neurosci, 22*(22), 9754-9763.

Eminaga, S., Christodoulou, D. C., Vigneault, F., Church, G. M., & Seidman, J. G. (2013). Quantification of microRNA expression with next-generation sequencing. *Curr Protoc Mol Biol, Chapter 4*, Unit 4.17. doi:10.1002/0471142727.mb0417s103

Fafournoux, P., Noel, J., & Pouyssegur, J. (1994). Evidence that Na+/H+ exchanger isoforms NHE1 and NHE3 exist as stable dimers in membranes with a high degree of specificity for homodimers. *J Biol Chem, 269*(4), 2589-2596.

Friedlander, M. R., Mackowiak, S. D., Li, N., Chen, W., & Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res, 40*(1), 37-52. doi:10.1093/nar/gkr688

Gamsiz, E. D., Ouyang, Q., Schmidt, M., Nagpal, S., & Morrow, E. M. (2012). Genome-wide transcriptome analysis in murine neural retina using high-throughput RNA sequencing. *Genomics, 99*(1), 44-51. doi:10.1016/j.ygeno.2011.09.003

Gao, R., Wang, L., Sun, J., Nie, K., Jian, H., Gao, L., . . . Gan, S. (2014). MiR-204 promotes apoptosis in oxidative stress-induced rat Schwann cells by suppressing neuritin expression. *FEBS Lett, 588*(17), 3225-3232. doi:10.1016/j.febslet.2014.07.004

Gilfillan, G. D., Selmer, K. K., Roxrud, I., Smith, R., Kyllerman, M., Eiklid, K., . . . Stromme, P. (2008). SLC9A6 mutations cause X-linked mental retardation, microcephaly, epilepsy, and ataxia, a phenotype mimicking Angelman syndrome. *Am J Hum Genet, 82*(4), 1003-1010. doi:10.1016/j.ajhg.2008.01.013

Griffiths-Jones, S. (2004). The microRNA Registry. *Nucleic Acids Res, 32*(Database issue), D109-111. doi:10.1093/nar/gkh023

Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., & Enright, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res, 34*(Database issue), D140-144. doi:10.1093/nar/gkj112

Griffiths-Jones, S., Saini, H. K., van Dongen, S., & Enright, A. J. (2008). miRBase: tools for microRNA genomics. *Nucleic Acids Res, 36*(Database issue), D154-158. doi:10.1093/nar/gkm952

Hall, D. P., Cost, N. G., Hegde, S., Kellner, E., Mikhaylova, O., Stratton, Y., . . . Czyzyk-Krzeska, M. F. (2014). TRPM3 and miR-204 establish a regulatory circuit that controls oncogenic autophagy in clear cell renal cell carcinoma. *Cancer Cell, 26*(5), 738-753. doi:10.1016/j.ccell.2014.09.015

He, L., & Hannon, G. J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet, 5*(7), 522-531. doi:10.1038/nrg1379

Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc, 4*(1), 44-57. doi:10.1038/nprot.2008.211

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res, 44*(D1), D457-462. doi:10.1093/nar/gkv1070

Kozomara, A., & Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res, 39*(Database issue), D152-157. doi:10.1093/nar/gkq1027

Kozomara, A., & Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res, 42*(Database issue), D68-73. doi:10.1093/nar/gkt1181

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods, 9*(4), 357-359. doi:10.1038/nmeth.1923

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol, 10*(3), R25. doi:10.1186/gb-2009-10-3-r25

Livingstone, C. (2013). IGF2 and cancer. *Endocr Relat Cancer, 20*(6), R321-339. doi:10.1530/erc-13-0231

Lopez, J. A., & Alvarez-Salas, L. M. (2011). Differential effects of miR-34c-3p and miR-34c-5p on SiHa cells proliferation apoptosis, migration and invasion. *Biochem Biophys Res Commun, 409*(3), 513-519. doi:10.1016/j.bbrc.2011.05.036

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal, 17*(1), 10-12. doi:DOI:10.14806/ej.17.1.200

Mi, H., Muruganujan, A., Casagrande, J. T., & Thomas, P. D. (2013). Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc, 8*(8), 1551-1566. doi:10.1038/nprot.2013.092

Murtazina, R., Booth, B. J., Bullis, B. L., Singh, D. N., & Fliegel, L. (2001). Functional analysis of polar amino-acid residues in membrane associated regions of the NHE1 isoform of the mammalian Na+/H+ exchanger. *Eur J Biochem, 268*(17), 4674-4685.

Nakamura, N., Tanaka, S., Teko, Y., Mitsui, K., & Kanazawa, H. (2005). Four Na+/H+ exchanger isoforms are distributed to Golgi and post-Golgi compartments and are involved in organelle pH regulation. *J Biol Chem, 280*(2), 1561-1572. doi:10.1074/jbc.M410041200

Ouyang, Q., Lizarraga, S. B., Schmidt, M., Yang, U., Gong, J., Ellisor, D., . . . Morrow, E. M. (2013). Christianson syndrome protein NHE6 modulates TrkB endosomal signaling required for neuronal circuit development. *Neuron, 80*(1), 97-112. doi:10.1016/j.neuron.2013.07.043

Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., . . . Ruvkun, G. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature, 408*(6808), 86-89. doi:10.1038/35040556

Pescosolido, M. F., Stein, D. M., Schmidt, M., El Achkar, C. M., Sabbagh, M., Rogg, J. M., . . . Morrow, E. M. (2014). Genetic and phenotypic diversity of NHE6 mutations in Christianson syndrome. *Ann Neurol, 76*(4), 581-593. doi:10.1002/ana.24225

Prasad, H., & Rao, R. (2015). The Na+/H+ exchanger NHE6 modulates endosomal pH to control processing of amyloid precursor protein in a cell culture model of Alzheimer disease. *J Biol Chem, 290*(9), 5311-5327. doi:10.1074/jbc.M114.602219

Pruitt, K. D., Brown, G. R., Hiatt, S. M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., . . . Ostell, J. M. (2014). RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res, 42*(Database issue), D756-763. doi:10.1093/nar/gkt1114

Re, D. B., Le Verche, V., Yu, C., Amoroso, M. W., Politi, K. A., Phani, S., . . . Przedborski, S. (2014). Necroptosis drives motor neuron death in models of both sporadic and familial ALS. *Neuron, 81*(5), 1001-1008. doi:10.1016/j.neuron.2014.01.011

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140. doi:10.1093/bioinformatics/btp616

Roux, J., Gonzalez-Porta, M., & Robinson-Rechavi, M. (2012). Comparative analysis of human and mouse expression data illuminates tissue-specific evolutionary patterns of miRNAs. *Nucleic Acids Res, 40*(13), 5890-5900. doi:10.1093/nar/gks279

Rubinsztein, D. C., Codogno, P., & Levine, B. (2012). Autophagy modulation as a potential therapeutic target for diverse diseases. *Nat Rev Drug Discov, 11*(9), 709-730. doi:10.1038/nrd3802

Sacconi, A., Biagioni, F., Canu, V., Mori, F., Di Benedetto, A., Lorenzon, L., . . . Blandino, G. (2012). miR-204 targets Bcl-2 expression and enhances responsiveness of gastric cancer. *Cell Death Dis, 3*, e423. doi:10.1038/cddis.2012.160

Stromme, P., Dobrenis, K., Sillitoe, R. V., Gulinello, M., Ali, N. F., Davidson, C., . . . Walkley, S. U. (2011). X-linked Angelman-like syndrome caused by Slc9a6 knockout in mice exhibits evidence of endosomal-lysosomal dysfunction. *Brain, 134*(Pt 11), 3369-3383. doi:10.1093/brain/awr250

Sun, Z., Evans, J., Bhagwate, A., Middha, S., Bockol, M., Yan, H., & Kocher, J. P. (2014). CAP-miRSeq: a comprehensive analysis pipeline for microRNA sequencing data. *BMC Genomics, 15*, 423. doi:10.1186/1471-2164-15-423

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol, 31*(1), 46-53. doi:10.1038/nbt.2450

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics, 25*(9), 1105-1111. doi:10.1093/bioinformatics/btp120

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., . . . Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol, 28*(5), 511-515. doi:10.1038/nbt.1621

Walkley, S. U. (2009). Pathogenic cascades in lysosomal disease-Why so complex? *J Inherit Metab Dis, 32*(2), 181-189. doi:10.1007/s10545-008-1040-5

Wang, X., & El Naqa, I. M. (2008). Prediction of both conserved and nonconserved microRNA targets in animals. *Bioinformatics, 24*(3), 325-332. doi:10.1093/bioinformatics/btm595

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet, 10*(1), 57-63. doi:10.1038/nrg2484

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., . . . Lander, E. S. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature, 420*(6915), 520-562. doi:10.1038/nature01262

Wong, N., & Wang, X. (2015). miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res, 43*(Database issue), D146-152. doi:10.1093/nar/gku1104

Yap, C. C., & Winckler, B. (2012). Harnessing the power of the endosome to regulate neural development. *Neuron, 74*(3), 440-451. doi:10.1016/j.neuron.2012.04.015

Yoshiyama, Y., Higuchi, M., Zhang, B., Huang, S. M., Iwata, N., Saido, T. C., . . . Lee, V. M. (2007). Synapse loss and microglial activation precede tangles in a P301S tauopathy mouse model. *Neuron, 53*(3), 337-351. doi:10.1016/j.neuron.2007.01.010