RNA-Seq reveals differentially expressed genes in post-mortem autism spectrum disorder

by Debra T. Linfield, Raoul R. Wadhwa

Abstract Wright et al. (2017) used gene set enrichment analysis to show significant differential expression of genes in the histamine signaling pathway between matched quadruples of cognitively typical subjects ($n = 39$) and patients diagnosed with autism spectrum disorder ($n = 13$) using postmortem biopsies of the dorsolateral prefrontal cortex in the human brain. Here, we replicate their initial differential gene expression analysis between the two groups, and conduct a biological analysis of the significantly differentially expressed SCARNA genes responsible for development and maintenance of cells in the central nervous system that were not explored by the original paper. To emphasize reproducibility, this report was generated with Rmarkdown that regenerates all figures within code chunks each time the document is knitted, and appendices at the end supply all necessary components for exact recreation of this report from the original paper's data stored in a publicly accessible archive.

Word count: 1958 (excluding appendices, captions)

Introduction

Autism spectrum disorder (ASD) is a set of developmental conditions with a wide variety of symptoms, generally related to social behaviors (Brentani et al., 2013). Studying ASD is complicated by the variety of its clinical manifestations and the existence of many monogenic, Mendelian disorders that cause the symptom of autistic behavior in patients (Ivanov et al., 2015). Specifically, observation of potentially autistic behavior in a patient is insufficient for diagnosis of ASD. This is especially true for pediatric patients, many of whom are not conclusively diagnosed with ASD until they are 5 or 6 years of age (Ellerbeck et al., 2015).

Histamine is a compound released by cells as part of the immune response, with increased levels being observed in cases of neuroinflammation (Jutel et al., 2005). Fernandez et al. (2012) showed that Tourette syndrome (TS), which is the most commonly observed comorbidity in patients diagnosed with ASD, is related to dysregulation of histamine signaling pathways. Moreover, the pathogenetic state of TS has been shown to have significant overlap with ASD (Clarke et al., 2012). Additionally, niaprazine - an antihistaminergic drug - has been used with some success for the treatment of ASD (Rossi et al., 1999). Given the past medical use of an antihistaminergic for ASD treatment and due to the known role of the histamine signaling pathway in the pathobiology of TS, the histamine signaling pathway is a natural target for RNA-Seq studies (Wright et al., 2017).

This report replicates the differential gene expression analysis conducted by Wright et al. (2017) with relation to the histamine signaling pathway. However, rather than replicating the already completed gene set enrichment analysis of this pathway, we explore the set of differentially expressed genes not further analyzed by the original paper.

Materials and Methods

Replication analysis methods

After download of sample data from SRA and subsequent conversion into fastq files using the `fastq-dump` command line tool, skewer (Jiang et al., 2014), HiSat2 (Kim et al., 2015), SAMtools (Li et al., 2009), and featureCounts (Liao et al., 2014) were used for trimming, alignment, formatting, and counting the RNA-Seq output, respectively (see Appendices A and B for details). Due to technical restrictions associated with data processing, all 39 control samples could not be used in this study; with permission of the teaching assistant, a subset of 13 controls was analyzed. The conducted analysis did include all 13 experimental samples from the original study. The R programming language (R Core Team, 2012) was used for further data processing. Specifically, the tidyverse package (Wickham, 2017) was used for data visualization, DESeq2 (Love et al., 2014) was used for differential gene expression analysis, and knitr (Xie, 2014, 2015, 2018) was used for typesetting in conjunction with the xtable (Dahl, 2016), gplots (Warnes et al., 2016), RColorBrewer (Neuwirth, 2014), pheatmap (Kolde, 2015), vsn (Huber et al., 2002), and formatR (Xie, 2017) packages. For the purpose of reproducibility and in support of the Open Science movement, this report was written in Rmarkdown with all plots produced

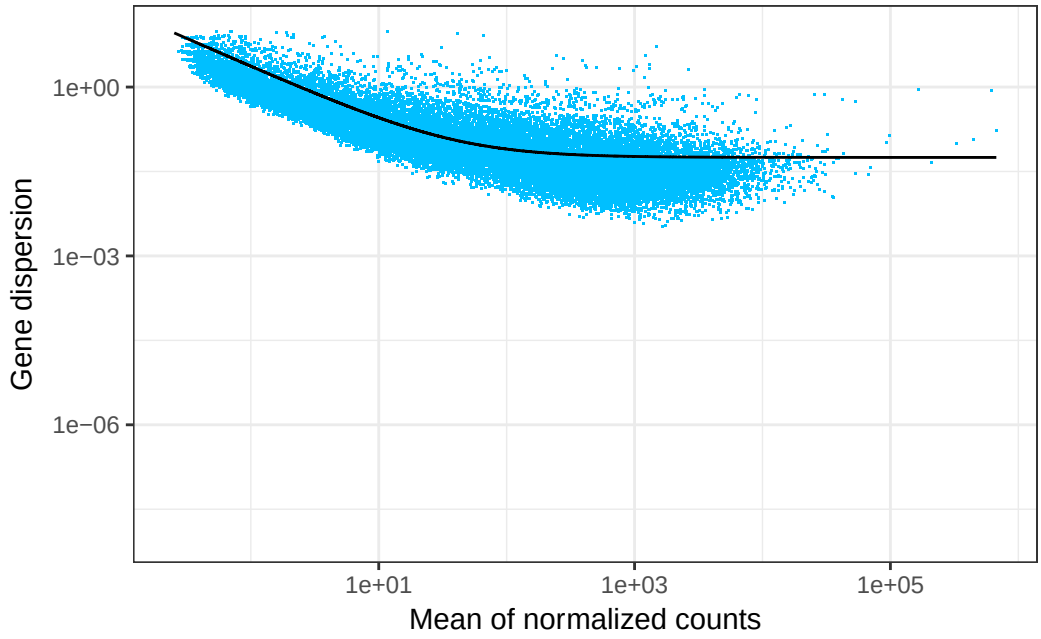

Figure 1: Gene dispersion plot for quality control. Note that raw dispersions are not plotted. Instead, corrected dispersions are graphed as a scatterplot, and dispersion is then fitted as a function of normalized counts by the black line. Log-log axes are used for clarity of visualization.

programmatically within code chunks and reconstructed each time the document was knitted. The Rmarkdown file and all associated dependencies can be found on GitHub (see Appendix A for details).

Relevant methods from original study

Post-mortem brain tissue samples from 52 subjects were collected, 39 of which were from non-psychiatric controls subjects, while 13 were from patients diagnosed with ASD. The post-mortem homogenate grey matter samples were collected from Brodmann area 46 and 9. RNA was extracted from the tissue using the RNeasy kit. RNA-Seq was then performed on the extracted RNA using the TruSeq Stranded Total RNA Library Preparation kit in conjunction with Illumina Ribo-Zero Gold ribosomal RNA depletion. All except 5 mismatched control subjects were matched (3:1 control:experimental ratio) to a subject in the experimental group of the same gender, ethnicity, and age (within 6 years) to control for genetic variation as a result of these factors. Differential expression analysis was conducted on log-normalized gene expression values after exclusion of genes with low expression values.

Wright et al. (2017) also performed a replication analysis of the results found by Parikshak et al. (2016); this replication analysis is not repeated within this report due to lack of information necessary for complete reproducibility, specifically, the lack of sample labeling for identification of the overlapping samples between the two studies (see Wright et al. (2017) for details).

Results and Discussion

A gene dispersion plot (Figure 1) is generated for quality control (QC). As expected, higher dispersion is observed for genes with lower normalized counts. The dispersion decreases and quickly flattens for increasing values of normalized counts. This pattern is consistent with expected results as stated by Love et al. (2014). Additionally, the variation plot of genes ranked by mean expression (Figure 2) shows a constant standard deviation across the range of all genes ranked by mean expression, further satisfying the requirements of DESeq2 analysis checked by QC (Love et al., 2014). A clear benefit of Figure 2 over Figure 1 is the use of hexagon binning to convey the density of points in the plot.

An MA plot was generated for QC (Figure 3). Genes that are not differentially expressed (black points) between groups have a fold change close to unity and a log fold change close to zero, whereas differentially expressed genes (red points) are located farther from the horizontal axis, as expected. Thus, the MA plot shows that normalized count data is comparable between the sample groups since normalization was able to correct for artifacts that caused meaningless differences in counts between

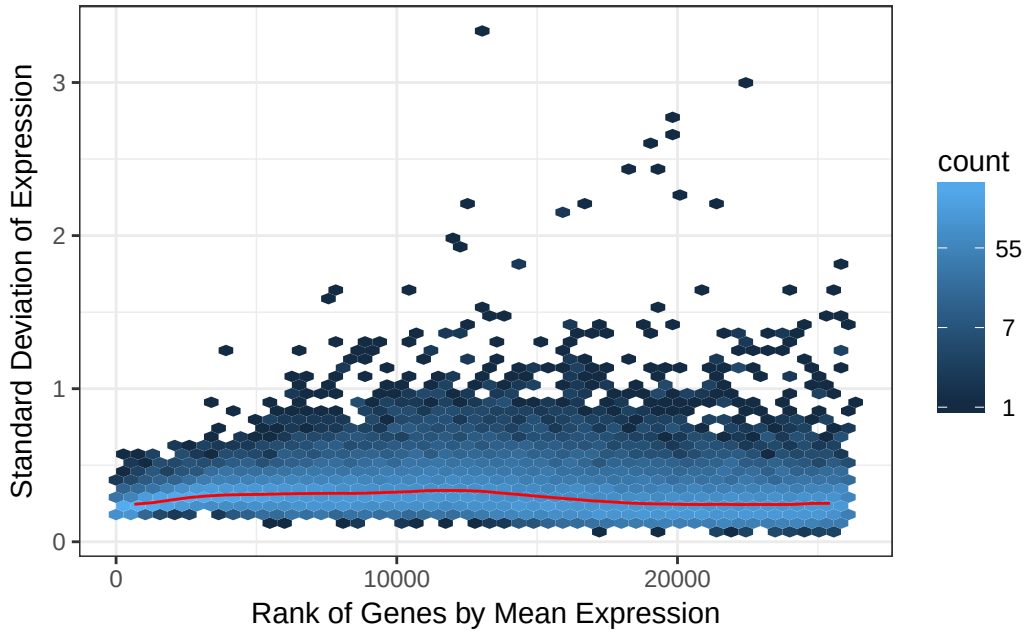

Figure 2: Standard deviation of expression data transformed with the variance stabilizing transformation. The constant variation (measured by standard deviation) across the range of genes ranked by mean expression satisfies the assumptions required of RNA-Seq for analysis by the DESeq2 pipeline.

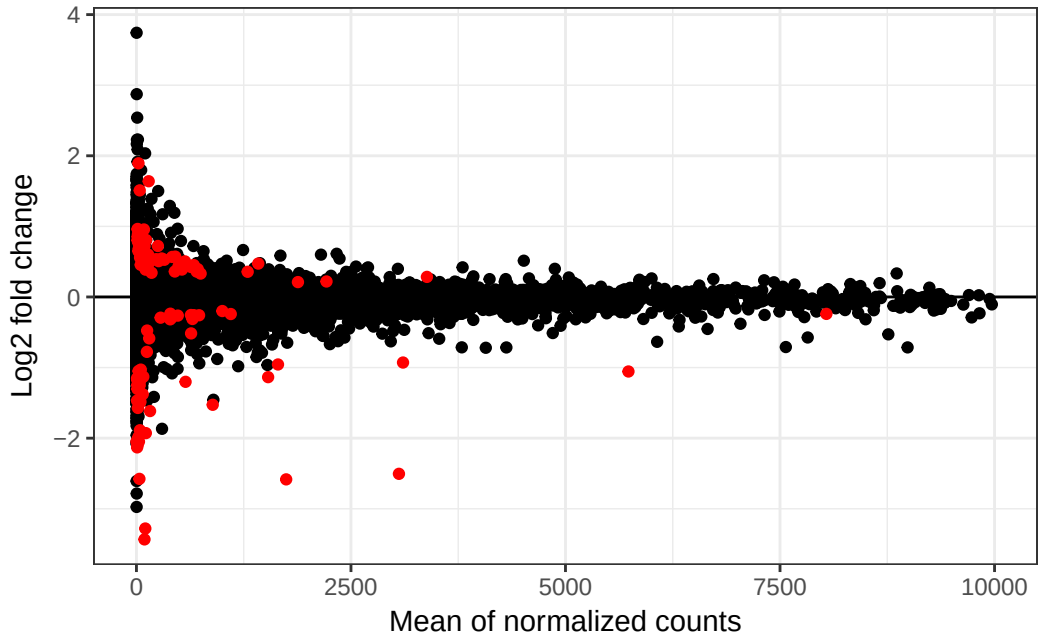

Figure 3: MA plot for QC. Points colored black represent genes with similar expression across groups; points colored red represent differentially expressed genes. As expected, most genes lie near $y = 0$, indicating that only a minority of genes are differentially expressed between the control and experimental groups. The vertical axis is log-transformed for clarity of visualization.

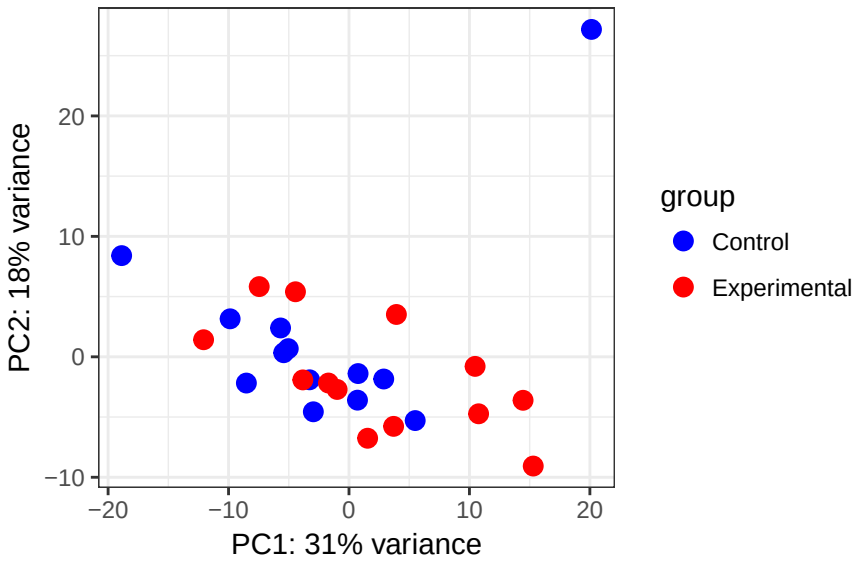

Figure 4: PCA plot of samples graphed by first two PCs. Control ($n = 13$) and experimental ($n = 13$) samples appear interspersed. A single control sample appears to be separated from the rest. See corresponding text for further discussion.

groups. Given that log fold change shrinkage methods were not applied to the MA plot, the visualized data does not raise concerns with regard to QC as stated by Love et al. (2014).

Principal component analysis (PCA) was conducted for QC (Figure 4), and each sample was plotted based on the first two principal components (PCs). It is of note that the control and experimental samples are not trivially clustered based on only the first two PCs. This could be indicative of either non-genetic mechanisms being responsible for ASD diagnoses or of relatively small genetic modifications having large epigenetic effects. In the former case, a lack of genetic changes in the experimental group relative to the control samples would explain interspersion of samples from both groups within the PCA plot. In the latter case, small genetic changes would not create a great deal of variance in the high-dimensional vectors representing the transcriptome. As such, these minor modifications would not be incorporated in the first two PCs, and would thus not prevent interspersion of samples from both groups in the 2-dimensional PCA plot. It is also of note that one sample appears separated from the other twenty-five. It is possible that this separation is an artifact of the study method, with the postmortem biopsy causing alteration of the observed transcriptome. However, Wright et al. (2017) stated that not all 52 samples were able to be matched properly in the matched quadruplet design. Given the robust method of biopsy and RNA extraction used, it is more likely that this sample was simply one of the ones that was not quadruplet matched. It is also possible that since the first two PCs capture less than half (49%) of the variation within the samples' gene expression, this is insufficient to add proper separation between the control and experimental samples, thus causing a slightly higher separation for a single sample to stand out by chance.

A heatmap based on distances between samples was constructed for further QC (Figure 5). Dendrogram clustering of the samples based on Euclidean distance between the vectors representing gene expression of the corresponding samples confirmed the results from the PCA plot (Figure 4): clustering does not split the samples into the control and experimental groups; rather, there is interspersion of samples from both groups in each cluster. The dendrogram and coupled labeling on the heatmap also indicate that the control sample separated from the other samples in the PCA plot (Figure 4) is likely "C21" (SRR5938421).

A volcano plot was generated to determine if any genes were differentially expressed between the experimental and control groups (Figure 6). Points in red are genes that are differentially expressed. While it can be seen that there is differential expression in a small percentage of the genes, the identity of differentially expressed genes cannot be discerned from this graph.

Table 1 validates our replication of the study by Wright et al. (2017), showing that our calculated adjusted p-values are similar to those of the original study. These values indicate that without gene set enrichment analysis to highlight the histamine signaling pathway, no single gene in the pathway is significantly differentially expressed. Thus, analysis by Wright et al. (2017) demonstrating the significant change in expression of the histamine signaling pathway in patients diagnosed with ASD indicates that altering expression of a single member of the histamine signaling pathway is likely

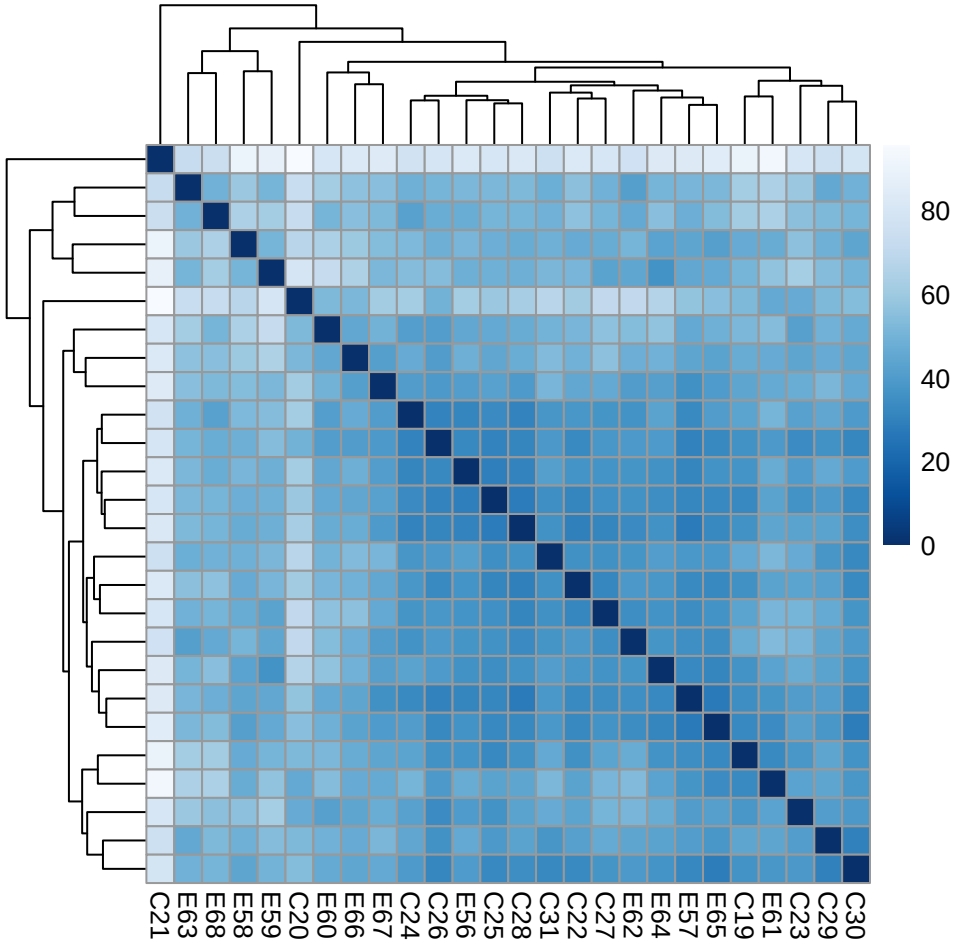

Figure 5: Heatmap of sample distance matrix. Each square is colored based on the distance between the samples labeling the row and column that intersect to form the square; a darker color indicates less distance (higher similarity) between samples. Rows are not explicitly labeled by sample, but labeling can be inferred using column labels and the main diagonal filled with squares of distance zero. Clustering visualized by dendrograms on the heatmap confirm the interspersion of samples between groups illustrated in the PCA plot (Figure 4).

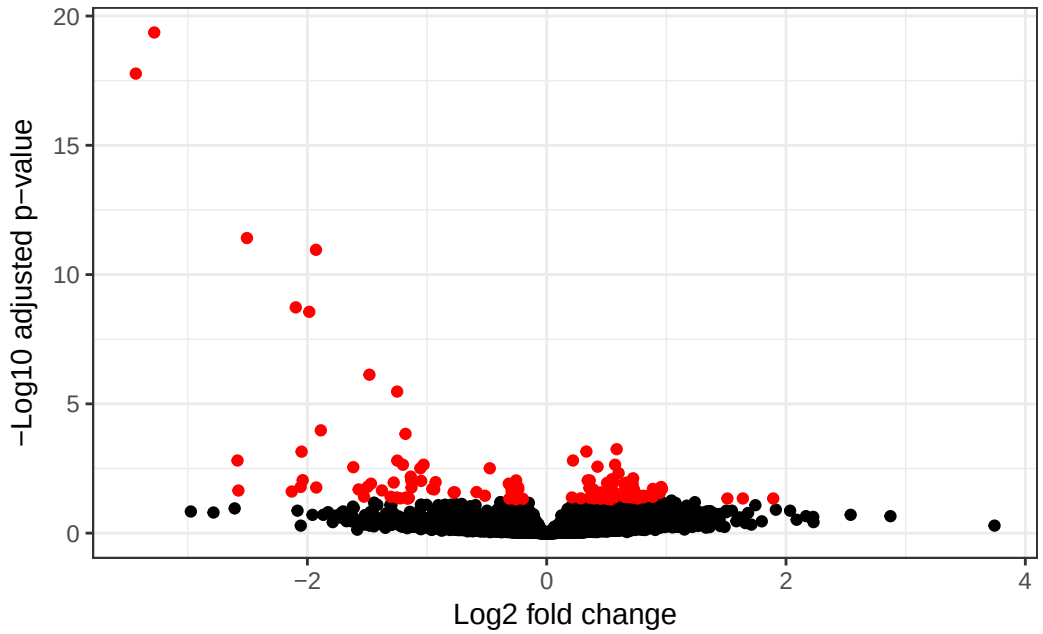

Figure 6: Volcano plot to characterize the number of genes with significantly altered gene expression. Points in black represent genes without significantly altered expression level between groups, and points in red represent genes expressed differentially in patients diagnosed with ASD ($\alpha = 0.01$).

Gene Name	Log2 Fold Change	Adjusted p-value
HDC	-0.39	0.82
HNMT	0.10	0.82
HRH1	0.15	0.73
HRH2	-0.30	0.66
HRH3	0.03	0.98
HRH4	-0.08	0.97

Table 1: Differential expression of genes in the histamine signaling pathway. Replication of the study by Wright et al. (2017) confirms that individual genes in the histamine signaling pathway are not significantly differentially expressed in patients diagnosed with ASD.

insufficient to explain the pathogenetic mechanism underlying the symptoms of ASD. This is expected for a disorder as complex as ASD.

Given that the individual genes in the histamine signaling pathway do not appear to have altered expression levels between groups, but there is good scientific justification to suspect the expression levels should be altered, we plot the normalized counts of the two groups, faceted for each of the six genes (Figure 7). The bar plot reveals no biological difference in expression levels between groups for genes that were part of the histamine signaling pathway. This validates the idea that gene set enrichment of the histamine signaling pathway is required to observe differential expression of genes in the pathway (Wright et al., 2017).

Table 2 lists the genes with the most significantly altered expression between groups (see Appendix C for complete list), and Figure 8 uses a heatmap to visualize normalized gene expression for each sample. Figure 8 reveals reliably differential expression of SCARNA10 and SCARNA23 between the control and experimental groups, with low variation within groups and significant variation between groups. The SCARNA genes are cajal body-specific molecules found primarily in metabolically active cells, particularly neurons (Gall et al., 1999). Given their importance in the development and maintenance of the central nervous system, differential expression of SCARNA genes could plausibly have a relationship to the onset of ASD. Due to the relatively low variation of SCARNA genes within the experimental group, altered expression of these genes could be a common transcriptomic feature observed across a variety of clinical manifestations of ASD, and could elucidate important pathways involved in the pathogenesis of this complex disorder.

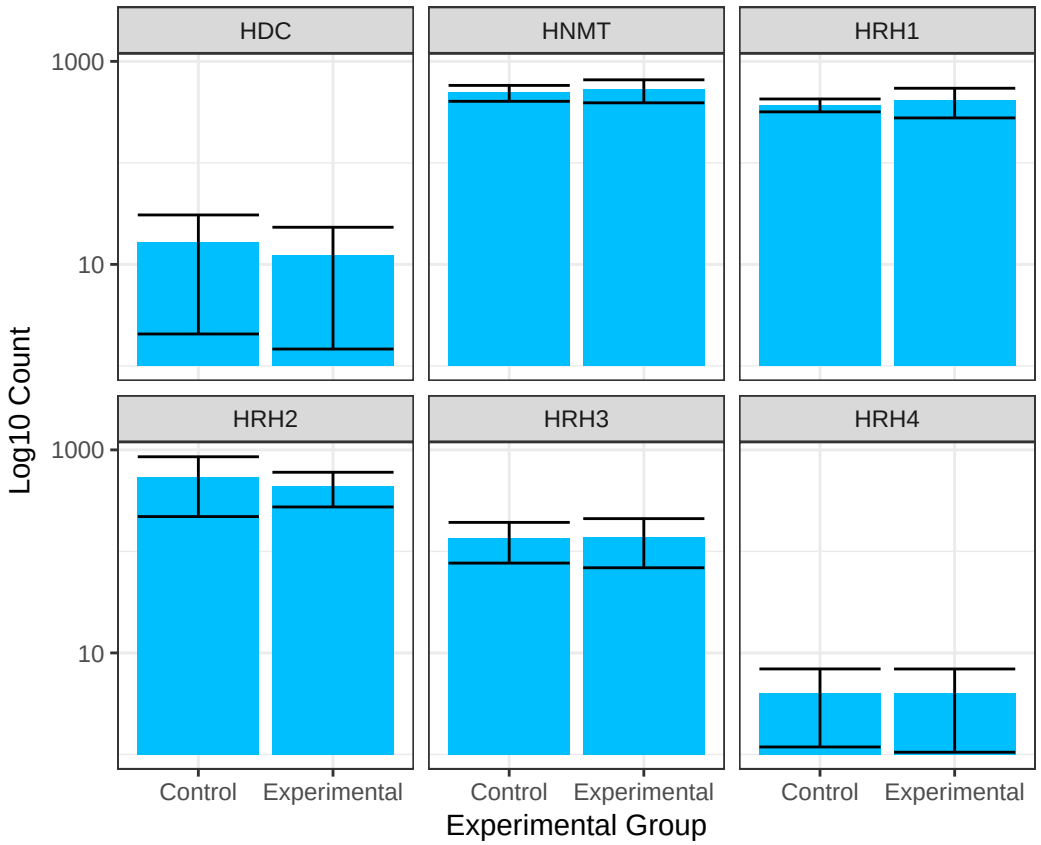

Figure 7: Normalized counts of genes in histamine signaling pathway compared by group and faceted by gene identity. Examination of the normalized counts between conditions confirms the lack of differential gene expression between groups. This likely indicates that lack of significance reflects a lack of biological significance, rather than a lack of statistical significance due to parameters such as sample size. The vertical axes is log-scaled for clarity of visualization. Error bars represent standard deviation around mean.

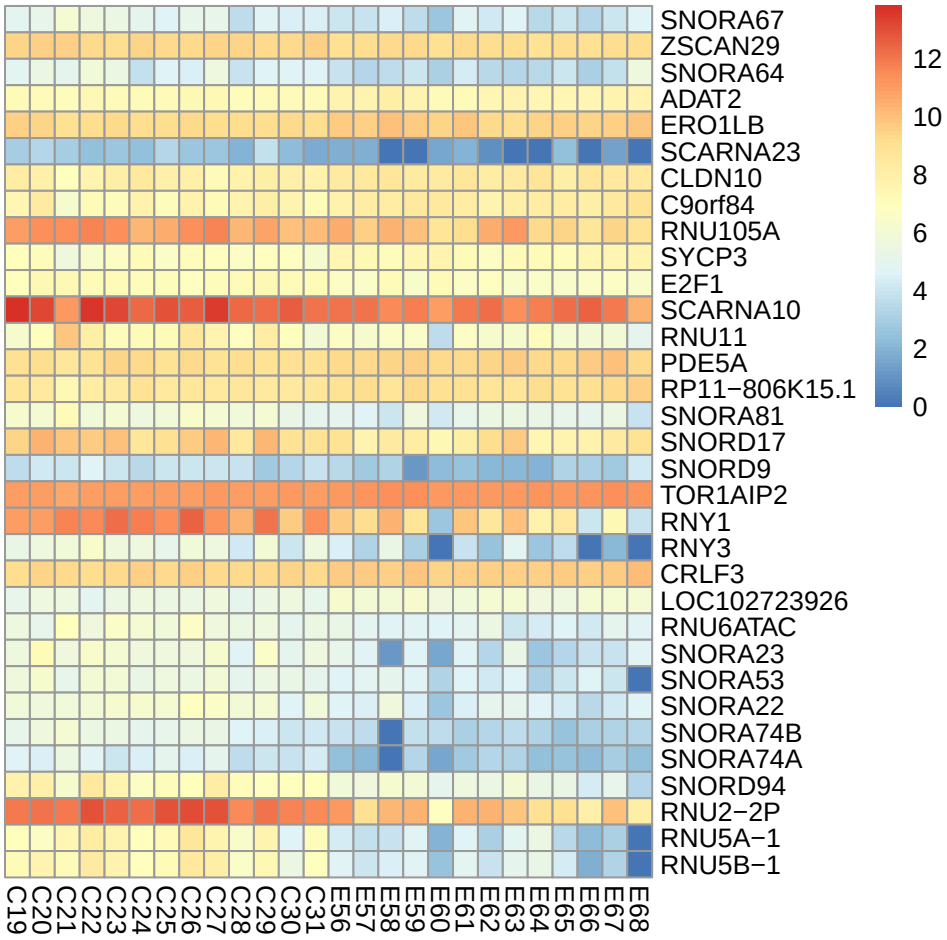

Figure 8: Heatmap of differentially expressed genes between cognitively typical subjects (left half, prefix 'C') and patients diagnosed with ASD (right half, prefix 'E'). Heterogeneous gene expression within groups (e.g. expression of RNY and SNOR genes within experimental group) likely indicates expression dependent on particular clinical manifestations of ASD, reducing the power of this RNA-Seq to detect altered gene expression between groups. However, the SCARNA genes (SCARNA10 and SCARNA23) appear to have relatively homogeneous expression within groups and with significantly reduced expression in the experimental group ($p \ll 0.01$).

Gene Name	Log2 Fold Change	Adjusted p-value
SNORA74B	-1.98	0.00
SNORA74A	-2.10	0.00
SNORD94	-1.93	0.00
RNU2-2P	-2.51	0.00
RNU5A-1	-3.43	0.00
RNU5B-1	-3.28	0.00

Table 2: Most significantly differentially expressed genes between cognitively typical subjects and subjects diagnosed with autism spectrum disorder (ASD). Summary data for the six genes with the most significantly altered expression between groups is shown, with a complete table given in Appendix C. A significance threshold of 0.01 post-correction for false discovery rate was used to identify differentially expressed genes.

Conclusions

We found similar results to Wright et al. (2017). The potential issues in the quality control (QC) plots did not cause any unexpected results. The SCARNA genes were significantly differentially expressed between groups (Figure 8), while there were no significant differences in expression of histamine signaling genes between patients diagnosed with ASD and cognitively typical subjects (Figure 7). Further elucidation of the genetic pathogenesis of ASD would require dedicated wet-lab experiments to isolate causal relationships between reduced expression of the SCARNA10 and SCARNA23 genes and the onset of ASD.

Sample selection was one of the primary strengths of the study conducted by Wright et al. (2017). First, the sample size of 52 is larger than most RNA-Seq studies, and grants more power for identification of differentially expressed genes. This gain in power is particularly important when gene set enrichment analysis is conducted as it involves a more in-depth look at a gene set picked *a priori*. Additionally, in the original study, the samples were cross-matched in a 3:1 control:experimental ratio by age at death, cause of death, and ethnicity. This matched quadruplet design helped control for incidental differential gene expression as a result of the aforementioned factors. It should be noted that although the original datasets were provided through the Sequence Read Archive (SRA), the analysis pipeline was not easily found and the sample labels for cross-matching were not publicly available. Given that only a subset of the control samples were used, the matched quadruplet design could not be replicated. However, the replicated differential expression values match relatively closely to the values reported by Wright et al. (2017).

One limitation of the study by Wright et al. (2017) was the selected operationalization of autism spectrum disorder (ASD). Given that ASD is a class of NDDs, blocking for a distinctive etiology of autism would have garnered more meaningful results. Due to the wide variety of manifestations of ASD, there is a loss of power in differential gene expression analysis if each of the experimental samples differentially expresses a unique gene pathway. This is particularly clear for one matched quadruplet of samples (3 controls, 1 experimental) where the age of death was less than 5 years old. Although some clinical manifestations of ASD can be diagnosed at less than 18 months of age, the absence of a clear test (e.g. blood test) for ASD makes diagnosis tricky for younger subjects. However, it may not be realistic to find 13 post-mortem brain biopsies from subjects exhibiting identical forms of ASD, so this was a justifiable compromise made by Wright et al. (2017).

Acknowledgments

We thank Dr. Gürkan Bebek and Peter Wilkinson for assistance with this project. This work made use of the High Performance Computing Resource in the Core Facility for Advanced Research Computing at Case Western Reserve University.

Bibliography

H. Brentani, C. S. Paula, D. Bordini, D. Rolim, F. Sato, J. Portolese, M. C. Pacifico, and J. T. McCracken. Autism spectrum disorders: an overview of diagnosis and treatment. *Revista Brasileira de Psiquiatria*, 35(Suppl 1):S62–S72, 2013. [p1]

R. A. Clarke, S. Lee, and V. Eapen. Pathogenetic model for tourette syndrome delineates overlap with related neurodevelopmental disorders including autism. *Translational Psychiatry*, 2:e158, 2012. [p1]

D. B. Dahl. *xtable: Export Tables to LaTeX or HTML*, 2016. URL https://CRAN.R-project.org/package=xtable. R package version 1.8-2. [p1]

K. Ellerbeck, C. Smith, and A. Courtemanche. Care of children with autism spectrum disorder. *Primary Care*, 42(1):85–98, 2015. [p1]

T. V. Fernandez, S. J. Sanders, I. R. Yurkiewicz, A. G. Ercan-Sencicek, Y. S. Kim, D. O. Fishman, M. J. Raubeson, Y. Song, K. Yasuno, W. S. Ho, K. Bilguvar, J. Glessner, S. H. Chu, J. F. Leckman, R. A. King, D. L. Gilbert, G. A. Heiman, J. A. Tischfield, P. J. Hoekstra, B. Devlin, H. Hakonarson, S. M. Mane, M. Gunel, and M. W. State. Rare copy number variants in tourette syndrome disrupt genes in histaminergic pathways and overlap with autism. *Biological Psychiatry*, 71(5):392–402, 2012. [p1]

J. G. Gall, M. Bellini, Z. Wu, and C. Murphy. Assembly of the nuclear transcription and processing machinery: Cajal bodies and transcriptosomes. *Mol Biol Cell*, 10(12):4385–4402, 1999. [p6]

W. Huber, A. von Heydebrek, H. Sueltmann, A. Poustka, and M. Vingron. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, 18:S96–S104, 2002. [p1]

H. Y. Ivanov, V. K. Stoyanova, N. T. Popov, and T. I. Vachev. Autism spectrum disorder - a complex genetic disorder. *Folia Medica*, 57(1):19–28, 2015. [p1]

H. Jiang, R. Lei, S.-W. Ding, and S. Zhu. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*, 15:182, 2014. [p1]

M. Jutel, K. Blaser, and C. A. Akdis. Histamine in allergic inflammation and immune modulation. *International Archives of Allergy and Immunology*, 137(1):82–92, 2005. [p1]

D. Kim, B. Langmead, and S. L. Salzberg. Hisat: a fast spliced aligner with low memory requirements. *Nature Methods*, 12:357–360, 2015. [p1]

R. Kolde. *pheatmap: Pretty Heatmaps*, 2015. URL https://CRAN.R-project.org/package=pheatmap. R package version 1.0.8. [p1]

H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–2079, 2009. [p1]

Y. Liao, G. K. Smyth, and W. Shi. featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923–930, 2014. [p1]

M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biology*, 15(12):550, 2014. [p1, 2, 4]

E. Neuwirth. *RColorBrewer: ColorBrewer Palettes*, 2014. URL https://CRAN.R-project.org/package=RColorBrewer. R package version 1.1-2. [p1]

N. N. Parikshak, V. Swarup, T. G. Belgard, M. Irimia, G. Ramaswami, M. J. Gandal, C. Hartl, V. Leppa, L. de la Torre Ubieta, J. Huang, J. K. Lowe, B. J. Blencowe, S. Horvath, and D. H. Geschwind. Genome-wide changes in lncrna, splicing, and regional gene expression patterns in autism. *Nature*, 540:423–427, 2016. [p2]

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL http://www.R-project.org/. ISBN 3-900051-07-0. [p1]

P. G. Rossi, A. Posar, A. Parmeggiani, E. Pipitone, and M. D'Agata. Niaprazine in the treatment of autistic disorder. *Journal of Child Neurology*, 14(8):547–550, 1999. [p1]

G. R. Warnes, B. Bolker, L. Bonebakker, R. Gentleman, W. H. A. Liaw, T. Lumley, M. Maechler, A. Magnusson, S. Moeller, M. Schwartz, and B. Venables. *gplots: Various R Programming Tools for Plotting Data*, 2016. URL https://CRAN.R-project.org/package=gplots. R package version 3.0.1. [p1]

H. Wickham. *tidyverse: Easily Install and Load the 'Tidyverse'*, 2017. URL https://CRAN.R-project.org/package=tidyverse. R package version 1.2.1. [p1]

C. Wright, J. H. Shin, A. Rajpurohit, A. Deep-Soboslay, L. Collado-Torres, N. J. Brandon, T. M. Hyde, J. E. Kleinman, A. E. Jaffe, A. J. Cross, and D. R. Weinberger. Altered expression of histamine signaling genes in autism spectrum disorder. *Translational Psychiatry*, 7:e1126, 2017. [p1, 2, 4, 6, 9, 12]

Y. Xie. knitr: A comprehensive tool for reproducible research in R. In V. Stodden, F. Leisch, and R. D. Peng, editors, *Implementing Reproducible Computational Research*. Chapman and Hall/CRC, 2014. [p1]

Y. Xie. *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition, 2015. [p1]

Y. Xie. *formatR: Format R Code Automatically*, 2017. URL https://CRAN.R-project.org/package=formatR. R package version 1.5. [p1]

Y. Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*, 2018. URL https://yihui.name/knitr. R package version 1.20. [p1]

Appendix A: Reproducible Scripts for Replicating Results

The Rmarkdown file knitted to generate this report can be found at https://github.com/rrrlw/RNASeq-ASD. The code chunks in this Rmarkdown file contain all necessary elements to recreate all of the plots; note that the working directory in the GitHub repository also contains the counts file (output from subread's featureCounts) that is read in and processed by the Rmarkdown. However, since R was not used for primary processing of the RNA-Seq data, the next paragraph and associated code chunk include details and code for reproducing that portion of the pipeline.

The bash script below reproduces the processing steps of the RNA-Seq analysis pipeline prior to input of counts into R and analysis by DESeq2. The fastq-dump tool can be used to extract fastq files from the files obtained from the SRA. See GEO page with ID GSE102741 for details on which files on the SRA correspond to the work done by Wright et al. (2017). Note that the bash script completes processing for all 52 samples used in the original study, not just the subset examined in this paper. Also note that the below code assumes that three empty directories named `readsTrimmed`, `alignments`, and `counts` exist in the working directory; additionally, the appropriate HiSat2 indexing file should be located in `HiSatIndex/human` and the genome gtf for the counting step should be a file named `genes.gtf` in the working directory.

```
##### TRIMMING (skewer) #####
module load skewer

# trim all samples
for i in {19..70}
do
    skewer --mode pe --threads 8 --mean-quality 30 --min 36 -q 30 --output
        readsTrimmed/S${i}.fastq --compress -y AGATCGGAAGAGC -x
        AGATCGGAAGAGC SRR59384${i}_1.fastq.gz SRR59384${i}_2.fastq.gz
done

##### ALIGNING (HiSat2) #####
module load hisat2
module add samtools

for i in {19..70}
do
    # SAM file creation
    hisat2 -p 12 -x HiSatIndex/human -1 readsTrimmed/S${i}.fastq-trimmed-pair1.fastq.gz
        -2 readsTrimmed/S${i}.fastq-trimmed-pair2.fastq.gz -S alignments/S${i}.sam

    # conversion from SAM to BAM
    samtools view -bS alignments/S${i}.sam | samtools sort -o alignments/S${i}.bam

    # BAM indexing
    samtools index alignments/S${i}.bam alignments/S${i}.bam.bai

    # flagstat to validate BAM file content
    samtools flagstat alignments/S${i}.bam > alignments/C${i}.flagstat

done

##### COUNTING (featureCounts) #####
module load subread

# count features in all BAM files
featureCounts -T 8 -s 0 -p -a genes.gtf -o counts/gene_id.counts alignments/*.bam
```

Appendix B: Version Numbers of Tools used for Analysis

Tool Name	Version	Reference
Skewer	0.2.2	Jiang et al. (2014)
HiSat2	2.1.0	Kim et al. (2015)
SAMtools	1.8	Li et al. (2009)
featureCounts	1.5.0-p2	Liao et al. (2014)
DESeq2	1.16.1	Love et al. (2014)
tidyverse	1.2.1	Wickham (2017)
knitr	1.20	Xie (2018)
RColorBrewer	1.1-2	Neuwirth (2014)
gplots	3.0.1	Warnes et al. (2016)
xtable	1.8-2	Dahl (2016)
pheatmap	1.0.8	Kolde (2015)
formatR	1.5	Xie (2017)
vsn	3.6	Huber et al. (2002)

Table 3: For reproducibility, this table lists the version numbers of every tool or package used in this report. To find the appropriate software repository/archive for each tool, see the associated reference.

Appendix C: Complete List of Differentially Expressed Genes

Gene Name	Log2 Fold Change	Standard Error (LFC)	Adjusted p-value
RNU5B-1	-3.28	0.32	0.00
RNU5A-1	-3.43	0.35	0.00
RNU2-2P	-2.51	0.31	0.00
SNORD94	-1.93	0.24	0.00
SNORA74A	-2.10	0.29	0.00
SNORA74B	-1.98	0.28	0.00
SNORA22	-1.48	0.23	0.00
SNORA53	-1.25	0.21	0.00
SNORA23	-1.89	0.34	0.00
RNU6ATAC	-1.18	0.22	0.00
LOC102723926	0.58	0.11	0.00
CRLF3	0.33	0.07	0.00
RNY3	-2.05	0.40	0.00
TOR1AIP2	0.22	0.04	0.00
RNY1	-2.58	0.53	0.00
SNORD9	-1.25	0.26	0.00
SNORD17	-1.20	0.25	0.00
SNORA81	-1.03	0.22	0.00
RP11-806K15.1	0.57	0.12	0.00
PDE5A	0.42	0.09	0.00
RNU11	-1.62	0.34	0.00
SCARNA10	-1.06	0.23	0.00
E2F1	-0.48	0.10	0.00
SYCP3	0.60	0.13	0.00
RNU105A	-1.14	0.25	0.01
C9orf84	0.72	0.16	0.01
CLDN10	0.55	0.12	0.01
SCARNA23	-2.04	0.46	0.01
ERO1LB	0.36	0.08	0.01
ZSCAN29	-0.26	0.06	0.01
SNORA64	-1.13	0.26	0.01
ADAT2	0.34	0.08	0.01
SNORA67	-1.05	0.24	0.01

Table 4: Complete list of differentially expressed genes. The complete list of differentially expressed genes for significance level 0.01; these genes exhibited a significantly altered expression level in patients diagnosed with ASD.

Author Contact and Affiliations

Debra T. Linfield
Department of Systems Biology and Bioinformatics, Case Western Reserve University, Cleveland, OH 44106, U.S.A.

debra.linfield@case.edu

Raoul R. Wadhwa
Department of Systems Biology and Bioinformatics, Case Western Reserve University, Cleveland, OH 44016, U.S.A.

raoul.wadhwa@case.edu