Postmortem Human Brain RNA-seq Analyses with Public and Private Data

Peterson, A[1,2*], Collado-Torres L[2], Jaffe AE[2,3,4]
[1]JHSPH, [2]Lieber Institute for Brain Development, [3]Biostatistics and [4]Mental Health JHSPH,
[*]apeter65@jhu.edu

LIEBER INSTITUTE *for* BRAIN DEVELOPMENT
MALTZ RESEARCH LABORATORIES

ABSTRACT

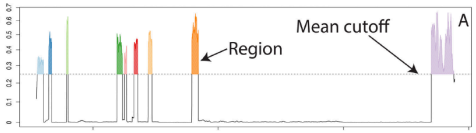
RNA-sequencing (RNA-seq) is a high-throughput method for quantifying gene expression levels that is dependent on high-quality RNA. We used RNA-seq data to explore expression profiles in various brain tissues and to address the effect of confounding caused by RNA quality differences.

First, we assessed RNA expression variability across brain regions through analysis of Genotype-Tissue Expression (GTEx) project data. We computed the mean base-pair level coverage for all brain samples in GTEx and for each of the brain sub-tissues in the dataset using data from the recount2 project (Collado-Torres et al. 2017c, Ellis et al. 2017). We noted differences in the mean expressed region widths in the overall brain compared to sub-tissues at smaller cutoff values. We then compared the width distribution of known exons between the overall brain and the various sub-tissues.

EXPRESSION PROFILES IN GTEx DATA

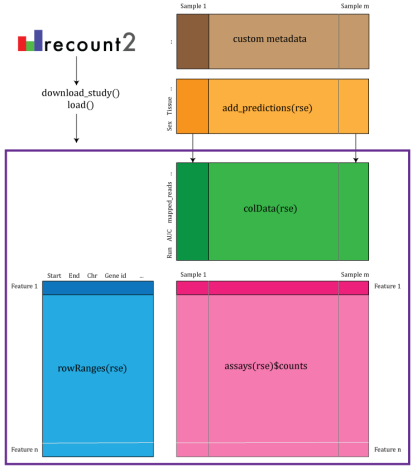
To examine RNA expression variability across brain regions, we first looked at expressed regions defined by a global cutoff.

We computed mean base-pair level coverage for all brain samples and for each of the 13 brain sub-tissues in GTEx using data from recount 2. Reads were scaled by 40 million reads of 100 base-pairs.



(Collado-Torres et al., 2017a)

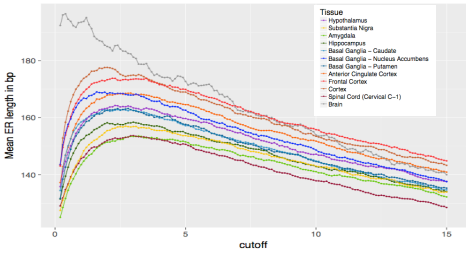
Recount2 contains gene, exon, exon-exon junction, and expressed region data. The data is available as RangedSummarizedExperiment objects that can be accessed by downloading and loading them in an R session.



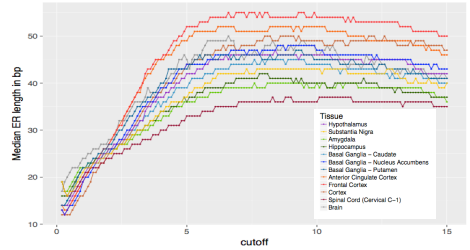
(Collado-Torres et al., 2017b)

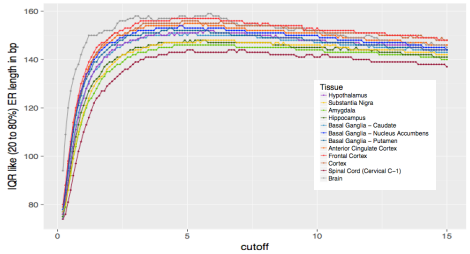
RNA EXPRESSED REGION VARIABILITY

Using GTEx data, we identified mean expressed region width in overall brain samples and samples from 11 sub-tissues.

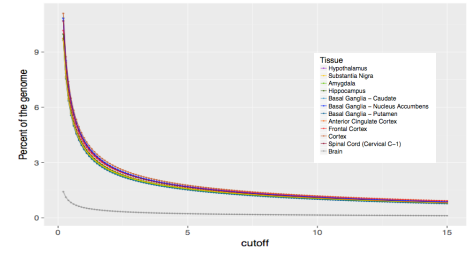


We examined the effect of outliers on the observed differences in mean expressed region width in sub-tissues compared to overall brain samples.



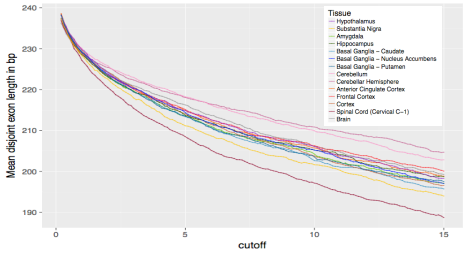


We assessed if the differences in mean expressed regions persisted when comparing the percent of the genome expressed in overall brain and sub-tissues.

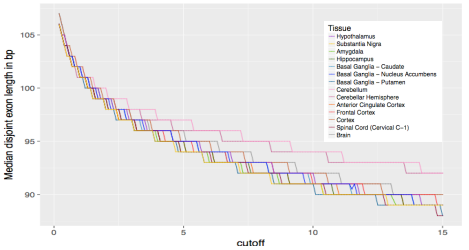


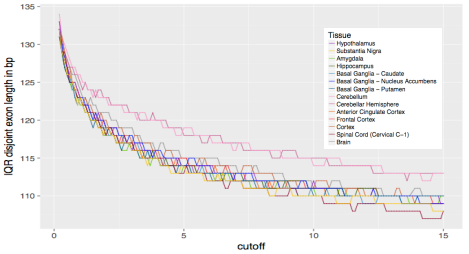
RNA DISJOINT EXON VARIABILITY

We compared the width distribution of annotated exons in recount2 using GTEx data samples for the overall brain and 13 subtissues.

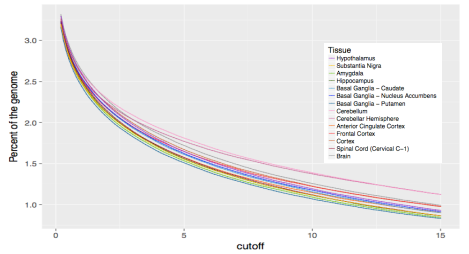


We examined the effect of outliers on annotated exons in overall brain and sub-tissue samples.



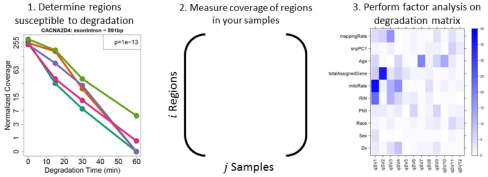


We assessed the percent of the genome that were annotated exons in overall brain samples and the individual sub-tissues.



RNA QUALITY CONFOUNDING

To assess the effects of confounding, we will extend the quality surrogate variable analysis (qSVA) algorithm (Jaffe et al. 2017) and perform a cross-region analysis of RNA-quality tissue in a case-control study, comparing degradation of tissue in patients with schizophrenia to controls using BrainSeq consortium data. Applying this method, we will identify and measure transcript features that are sensitive to tissue degradation in a differential expression degradation dataset, create factors to control for RNA quality confounding in an independent private dataset, BrainSeq, and assess the performance of the modified qSVA algorithm.



(Jaffe et al., 2017)

$$Y = \alpha + \beta\, Dx + \gamma\, region + \delta Dx * region + \varepsilon\, qSVA$$

SUMMARY

We explored RNA expression variability across brain regions through the analysis of Genotype-Tissue Expression (GTEx) project data. Upon noticing pattern differences comparing overall brain samples to sub-tissues, we used annotated exon expression data in recount2 to explore these differences further. Taken together, these results suggest the need to determine an optimal cut off that is specific to each tissue, to ensure minimal inclusion of noise, particularly at lower cut offs.

We are now looking to assess the effect of RNA quality confounding in private data, by performing cross-region brain analyses of RNA-quality tissue, comparing schizophrenic brains to control brains in BrainSeq consortium data. Our initial analysis will include two brain regions that are well-characterized as altered in schizophrenic patients, the dorsolateral prefrontal cortex (DLPFC) and the hippocampus.

REFERENCES

*Collado-Torres, L., Nellore, A., Frazee, A.C., Wilks, C., Love, M.I., Langmead, B., Irizarry, R.A., Leek, J.T., Jaffe, A.E. 2017a. Flexible expressed region analysis for RNA-seq with derfinder. *Nucleic Acids Research, 45(2),* e9. doi: 10.1093/nar/gkw852

*Collado-torres L, Nellore A, and Jaffe, AE. 2017b. recount workflow: Accessing over 70,000 human rNA-seq samples with Bioconductor. *F1000Research*, 6: 1558.

*Collado-Torres, L., Nellore, A., Kammers, K., Ellis, S.E., Taub, M.A., Hansen, K.D., Jaffe, A.E., Langmead, B. and Leek, J.T. 2017c. Reproducible RNA-seq analysis using recount2. *Nature Biotechnology* 35(4), pp. 319–321.

*Ellis, S.E., Collado Torres, L. and Leek, J. 2017. Improving the value of public RNA-seq expression data by phenotype prediction. *BioRxiv*.

*Jaffe, A.E., Tao, R., Norris, A.L., Kealhofer, M., Nellore, A., Shin, J.H., Kim, D., Jia, Y., Hyde, T.M., Kleinman, J.E., Straub, R.E., Leek, J.T. and Weinberger, D.R. 2017. qSVA framework for RNA quality correction in differential expression analysis. *Proceedings of the National Academy of Sciences of the United States of America* 114(27), pp. 7130–7135.

ACKNOWLEDGEMENTS

We would like to acknowledge the members of Andrew Jaffe (Lieber Institute for Brain Development, Johns Hopkins Medical Campus) and Mina Ryten (UCL Neuro) labs for feedback on the explanatory figures.