bioRxiv preprint doi: https://doi.org/10.1101/116699; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Title page

Title: DLPFC Transcriptome Defines Two Molecular Subtypes of Schizophrenia

Authors: C. Harker Rhodes, MD, PhD
Elijah F. W. Bowen, MS
Caitlyn I. Lee
Jack L. Burgess
Richard Granger, PhD

Contact Information:
C. Harker Rhodes, MD, PhD
144 Sunset Rock Rd
Lebanon, NH 03766

Tel: 603-443-3360
FAX:
E-mail: CHarkerRhodes@gmail.com

bioRxiv preprint doi: https://doi.org/10.1101/116699; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Abstract

The Clinical Brain Disorders Branch of the Intramural Research Program at the National Institutes of Health assembled a large collection of frozen post-mortem human brains from individuals diagnosed with schizophrenia, or other psychiatric diagnoses as well as matched control individuals, collected Illumina HumanHT-12 v4 expression array data from dorsolateral prefrontal cortex (DLPFC) of those brains, and deposited the data at dbGaP (Study ID: phs000979). We report an analysis of the data from the 188 adult schizophrenics and 206 adult controls in that cohort.

Transcripts from 635 genes are differentially expressed in the schizophrenics as compared to the controls at a level of statistical significance which survived Bonferroni correction. Seventeen of those genes are differentially expressed at a level of statistical significance less than 10^{-8} after Bonferroni correction.

WGCNA analysis using gene module eigengenes divided the schizophrenics into two groups, "type 1" and "type 2". There are 2,253 genes differentially expressed in the DLPFC of type 2 schizophrenics at a level of statistical significance which survives Bonferroni correction. Of them 160 have P-values less than 10^{-15} after Bonferroni correction. In the type 1 schizophrenics however, only 13 genes are differentially expressed in the DLPFC and of those 13, only one is also differentially expressed in the type 2 schizophrenics. This striking difference in their DLPFC transcriptomes emphasizes the fundamental biologic difference between these two groups of patients.

The fact that only a few genes are differentially expressed in the DLPFC of type 1 schizophrenics while over two thousand are differentially expressed in the type 2 schizophrenics suggests a testable hypothesis. Although the DLPFC is clearly abnormal in the type 2 schizophrenics, for the type 1 schizophrenics the physiologically significant pathology may be elsewhere, perhaps in the superior temporal or cingulate gyri.

bioRxiv preprint doi: https://doi.org/10.1101/116699; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Introduction

The Clinical Brain Disorders Branch of the Intramural Research Program at the National Institutes of Health assembled a large collection of frozen post-mortem human brains from individuals diagnosed with schizophrenia, or other psychiatric diagnoses as well as matched control individuals, collected Illumina HumanHT-12 v4 expression array data from dorsolateral prefrontal cortex (DLPFC) of those brains, and deposited the data at dbGaP (Study ID: phs000979). We report an analysis of the data from the 188 adult schizophrenics and 206 adult controls in that cohort.

Robust linear mixed effect regression including as a random effect "batch" and as fixed effects Rin, age, sex, age at death, and race identified 635 genes which are differentially expressed in the DLPFC of schizophrenics as compared to the controls at a level of statistical significance which survived Bonferroni correction. As expected, many of those genes have been previously identified in other studies as being differentially expressed in the brains of schizophrenics. Some of them suggest novel druggable targets for the treatment of schizophrenia.

Recognizing the clinical heterogeneity among schizophrenics, a successful attempt was made to divide the patients in this cohort into biologically meaningful schizophrenia subtypes based on their DLPFC transcriptomes. Two subtypes were identified. Unfortunately the limited clinical information available in this Medical Examiner based sample of convenience does not allow a correlation of the subtypes with clinical phenotype, but they have strikingly different sets of differentially expressed genes supporting the conclusion that there are fundamental biologic differences between these two groups of patients.

Methods

Over a period of many years the Clinical Brain Disorders Branch of the NIMH intramural program assembled a large collection of frozen human brains from Medical Examiner patients and conducted detailed post-mortem psychiatric reviews to establish their diagnoses. The human tissue collection, handling, and processing protocols have been previously described [1]. RNA was prepared from dorsolateral prefrontal cortex and hippocampus, Illumina HumanHT-12 v4 expression array data was generated according to the manufacturer's protocols, and that data was made publicly available at dbGaP (Study ID: phs000979).

With the appropriate IRB approval and dbGaP authorization that data downloaded to a secure high-performance cluster (two Intel Xeon x5660 2.8GHz CPUs with 6 cores each, with hyperthreading yielding 24 virtual cores, with 72GB of DDR3-1333 RAM). Data analysis done using Rstudio and protocols for reproducible research. HTML-formatted R markdown files which contain the computer code for the entire analysis are included as supplemental material.

Results

Cohort demographics

This cohort is a convenience sample based on Medical Examiner cases for whom the next of kin consented to the use of post-mortem tissue for this purpose. It is, therefore, in many ways not representative of the general population and this study limitation needs to be kept in mind when interpreting these results.

bioRxiv preprint doi: https://doi.org/10.1101/116699; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

As might be expected in a Medical Examiner cohort where the control subjects include accidental death and homicide victims, men are over-represented in the controls. That imbalance is much more prominent in the Caucasian than African American sub-cohorts (table 1). The cohort is, however, reasonably well balanced in terms of both ethnicity (table 2) and age (figure 1).

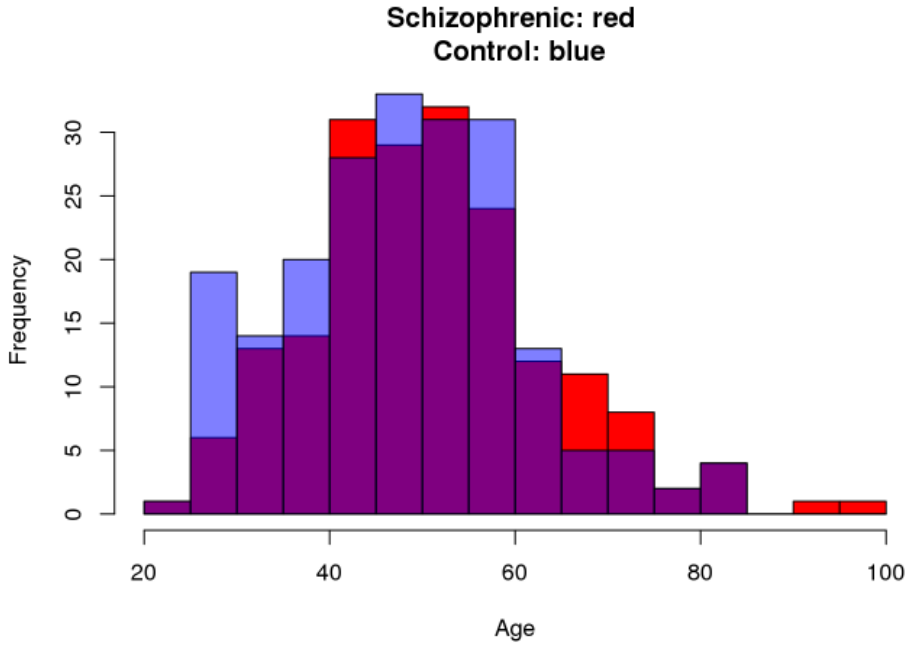
Table 1: Distribution of schizophrenics and controls by gender

	All subjects		Caucasians		African Americans	
	Controls	Sz	Controls	Sz	Controls	Sz
Female	61	72	16	40	43	31
Male	145	117	76	59	58	54

Table 2: Distribution of schizophrenics and controls by ethnicity

	Control	Sz
African American	101	85
Caucasian	92	99
Other	13	5

Figure 1: Distribution of schizophrenics and controls by age.



Differentially expressed genes in schizophrenics as compared to controls

The expression array data included data from 11,883 probes after censoring data from probes which did not detect mRNA in the DLPFC at a statistically significant level or which contained common polymorphisms in the probe sequence.

Robust linear mixed effects regression including RIN, gender, ethnicity, and age as fixed effects and "batch" as a random effect identified 696 array probes which detected transcripts from 696 genes

bioRxiv preprint doi: https://doi.org/10.1101/116699; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

which were differentially expressed in the DLPFC of the schizophrenics at a level of statistical significance which survived Bonferroni-correction. For two of those genes, SYNDIG1 (aka TMEM90B, a gene involved in the maturation of excitatory synapses) and PSMB6 (a proteasomal subunit gene), the Bonferroni-corrected P-value was less than 10^{-15}. The complete list of the 696 differentially expressed genes is included in the supplemental material

WGCNA clustering of schizophrenics

Three gene modules were identified when weighted gene co-expression network analysis (WGCNA) was used to cluster the differentially expressed transcripts based on their co-expression in the DLPFC of the schizophrenics. WGCNA was then used a second time, this time to cluster the schizophrenic patients based on the eigengenes for those gene modules. This analysis divides the schizophrenics into two groups, "type 1" and "type 2", each of which is further subdivided into two subgroups. (figure 2).

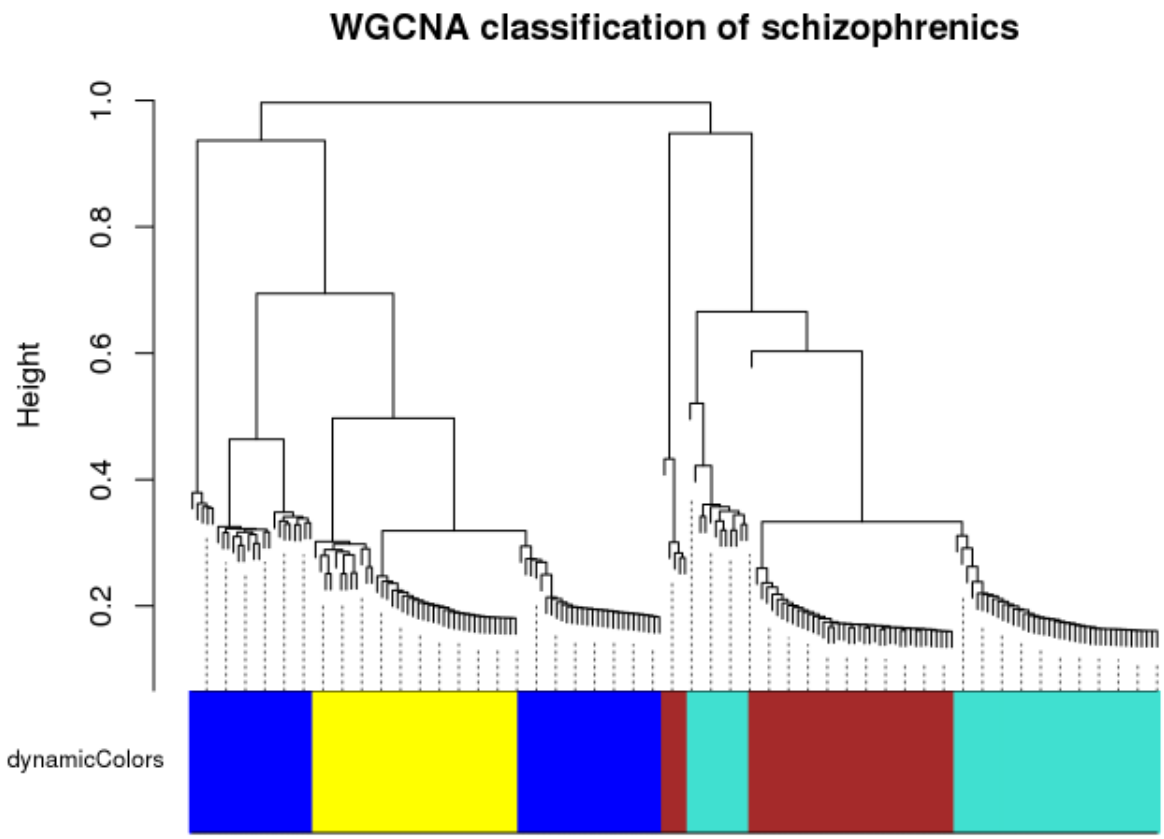


Figure 2: WGCNA clustering of the schizophrenics based on the eigengenes from the three gene modules present in the set of differentially expressed genes. The turquoise and brown classes of patients are the type 1 schizophrenics, while the blue and yellow classes are type 2.

The differential gene expression in the DLPFC is strikingly different in these two groups of patients. There are 2,253 genes differentially expressed in the DLPFC of type 2 schizophrenics at a level of statistical significance which survives Bonferroni correction. Of them 160 have P-values less than 10^{-15} after Bonferroni correction. In the type 1 schizophrenics however, only 13 genes are differentially expressed in the DLPFC and of those 13, only one is also differentially expressed in the type 2

bioRxiv preprint doi: https://doi.org/10.1101/116699; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

schizophrenics. This striking difference in their DLPFC transcriptomes emphasizes the fundamental biologic difference between these two groups of patients.

The difference between the type 1 and type 2 schizophrenics is also evident in their sex ratios. In this Medical Examiner cohort, men are over-represented in the controls presumably because accidental death and homicide victims are more likely to be men than women. However, the study cohort also includes bipolar patients and they have a sex ratio of close to 1:1. Interestingly, the type 1 schizophrenics have a sex ratio similar to that of the controls while the type 2 schizophrenics have a sex ratio similar to that of the bipolar patients (see table 3).

Table 3: Sex ratio in the controls, bipolar patients, and type 1 and type 2 schizophrenics in this study cohort.

	Controls	Bipolar patients	Type 1 Sz	Type 2 Sz
F	61	41	27	45
M	145	47	70	47

Based on a chi-squared test, the difference in the sex ratios of the type 1 and type 2 schizophrenics is significant at a level $P < 0.005$.

The complete list of genes differentially expressed in type 1 and type 2 schizophrenics is included in the supplemental materials, but the differences between these groups of patients is best illustrated by specific examples of the differentially expressed genes. For example, the nuclear-encoded mitochondrial gene NDUFB2 (NADH:ubiquinone oxidoreductase subunit B2) is down regulated in the DLPFC of type 2 schizophrenics but not type 1 schizophrenics and the effect size is so large that there is little overlap between the two groups of patients (figure 3).

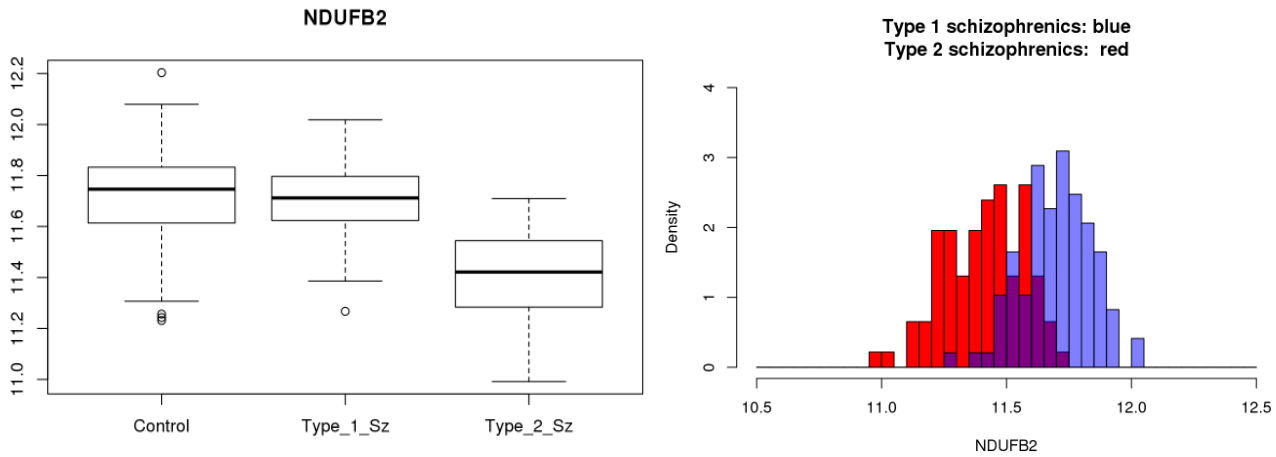


Figure 3: Expression of C (NADH:ubiquinone oxidoreductase subunit B2) in DLPFC.
Left: Boxplot of NDUFB2 in the DLPFC of controls, type 1 schizophrenics and type 2 schizophrenics
Right: Histogram of NDUFB2 in the DLPFC of controls, type 1 schizophrenics and type 2 schizophrenics
Based on a two-tailed t-test, the NDUFB2 expression in the type 1 and type 2 schizophrenics is different at a statistical significance level of $P < 2.2e-16$ (the smallest floating point number that can be represented by the software we are using).

bioRxiv preprint doi: https://doi.org/10.1101/116699; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Another instructive example is the pair of proteasomal subunit genes PSMB6 and PSMB10. Many studies have implicated ubiquitin proteasome system (UPS) in the pathophysiology of schizophrenia and described the differential expression of UPS genes in the brains of schizophrenics [2-5]. There are, however, inconsistencies in the published results, perhaps in part because none of the previous studies distinguished between type 1 and type 2 schizophrenics. As can be seen in figure 4, there is a small, but statistically significant down regulation of PSMB10, the gene for one of the proteasome beta subunits, in type 1, but not type 2 schizophrenics. On the other hand, PSMB6, a gene for a different beta subunit is dramatically down regulated in the type 2 schizophrenics.

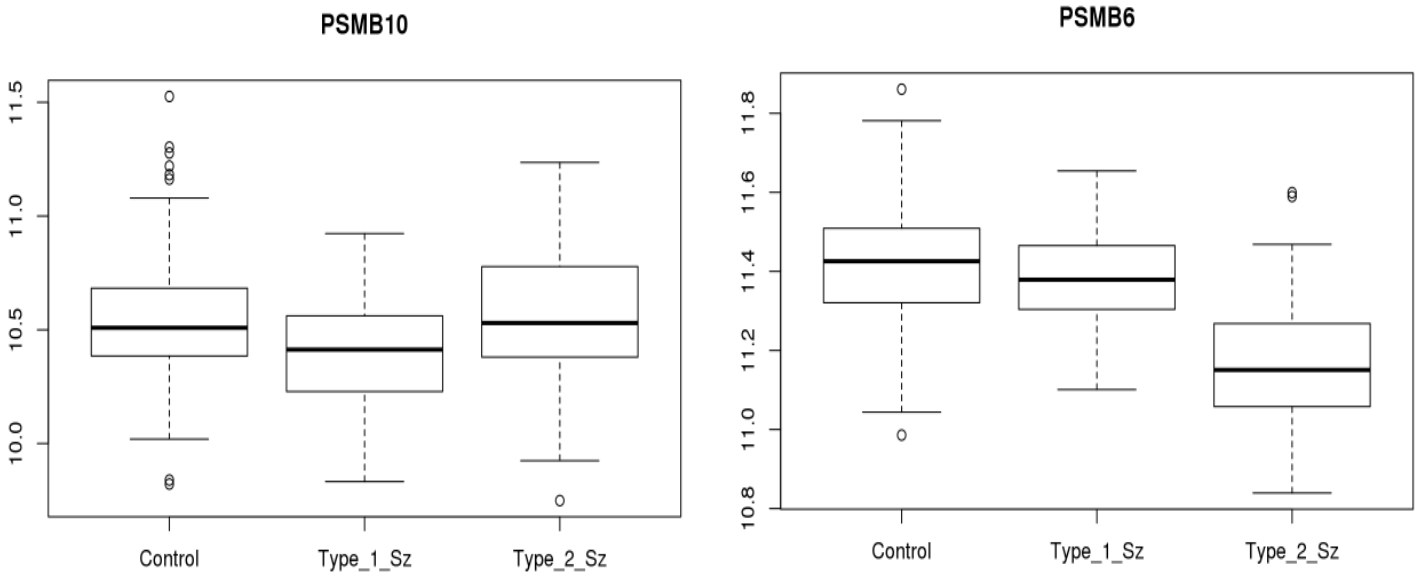


Figure 4: Differential expression of the proteasomal subunit genes PSMB10 and PSMB6 in type 1 and type 2 schizophrenics.

Left: PSMB10 is down regulated in type 1 schizophrenics ($P < 0.01$ after Bonferroni correction) but not in type 2 schizophrenics.

Right: PSMB6 is dramatically down regulated in type 2 schizophrenics ($P < 10^{-15}$). It is also very slightly down regulated in type 1 schizophrenics, but not at a level of statistical significance which survives Bonferroni correction.

Relationship to "low GABA marker" phenotype

About half of all schizophrenics, schizoaffective patients, and bipolar patients have what has been described as a "low GABA marker" molecular phenotype based on the expression of GABA neuron markers [6, 7]. Specifically, this subset of schizophrenic patients have reduced expression of GAD67, parvalbumin, somatostatin, and the transcription factor LHX6 in their DLPFC.

In this expression array dataset, the Illumina probes for somatostatin and parvalbumin do not detect transcripts at a level significantly different from zero. Neither GAD67 (GAD1) nor the LHX6 transcripts are differentially expressed at a level of statistical significance which survives Bonferroni correction. However, since we are now testing the specific hypotheses that those transcripts are down

bioRxiv preprint doi: https://doi.org/10.1101/116699; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

regulated in schizophrenics Bonferroni correction is no longer appropriate. Both of those genes are down regulated in the type 2 schizophrenics, but not the type 1 schizophrenics. (The P-value without Bonferroni correction for differential expression of GAD67 in type 2 schizophrenics is 1×10^{-7}; for LHX6 the P-value is 1×10^{-6}.

Discussion

This reanalysis of a publicly available expression array dataset identifies 635 genes which are differentially expressed in the dorsolateral prefrontal cortex of schizophrenics as compared to controls at a level of statistical significance which survives Bonferroni correction. More importantly, it demonstrates that schizophrenics can be divided into two molecularly distinct subgroups based on their DLPFC transcriptomes. Future studies of gene expression in schizophrenia need to take into account this molecularly defined patient heterogeneity.

The fact that only a few genes are differentially expressed in the DLPFC of type 1 schizophrenics while over two thousand are differentially expressed in the type 2 schizophrenics suggests a testable hypothesis. Although the DLPFC is clearly abnormal in the type 2 schizophrenics, for the type 1 schizophrenics the physiologically significant pathology may be elsewhere, perhaps in the superior temporal or cingulate gyri.

These results are consistent with the description by Volk et al. (2016, 2012) of a "low GABA marker" (LGM) molecular phenotype in about half of all schizophrenics [6, 7]. That phenotype was described based on gene expression levels in the DLPFC and in this study of the DLPFC transcriptome the type 2, but not the type 1, schizophrenics have reduced expression of GABA neuron markers. Note, however, that both this study, and the studies of Volk et al. examined only DLPFC. It may be that schizophrenics who do not have the LGM phenotype in DLPFC may express that phenotype in other cortical areas and visa-versa.

The limitations of this study are obvious. The results are based on the study of a single cohort and need to be replicated in other patient/control cohorts. As with any observational study any speculation about causation is only speculation – there is no way to know whether the transcriptome differences described here are related to the pathogenesis of schizophrenia or are the consequence of the presence of the disease and whatever therapy the patients received. And finally, like any expression array study, the results of this study must be regarded as preliminary until they are confirmed by qPCR or some other technique.

Supplemental Information

Online supplemental files 1-9: R markdown files containing the computer code for the analyses described in this manuscript.

Online supplemental file 10: A csv file containing a list of the genes differentially expressed in the DLPFC of schizophrenics (without separating the type 1 and type 2 patients).

bioRxiv preprint doi: https://doi.org/10.1101/116699; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Literature cited

1. Lipska, B.K., et al., *Critical factors in gene expression in postmortem human brain: Focus on studies in schizophrenia.* Biol Psychiatry, 2006. **60**(6): p. 650-8.
2. Andrews, J.L., et al., *Alterations of ubiquitin related proteins in the pathology and development of schizophrenia: Evidence from human and animal studies.* J Psychiatr Res, 2017. **90**: p. 31-39.
3. Altar, C.A., et al., *Deficient hippocampal neuron expression of proteasome, ubiquitin, and mitochondrial genes in multiple schizophrenia cohorts.* Biol Psychiatry, 2005. **58**(2): p. 85-96.
4. Arion, D., et al., *Distinctive transcriptome alterations of prefrontal pyramidal neurons in schizophrenia and schizoaffective disorder.* Mol Psychiatry, 2015. **20**(11): p. 1397-405.
5. Scott, M.R., et al., *Protein Expression of Proteasome Subunits in Elderly Patients with Schizophrenia.* Neuropsychopharmacology, 2016. **41**(3): p. 896-905.
6. Volk, D.W., et al., *Deficits in transcriptional regulators of cortical parvalbumin neurons in schizophrenia.* Am J Psychiatry, 2012. **169**(10): p. 1082-91.
7. Volk, D.W., et al., *Cortical GABA markers identify a molecular subtype of psychotic and bipolar disorders.* Psychol Med, 2016. **46**(12): p. 2501-12.