OXFORD UNIVERSITY PRESS | Schizophrenia Bulletin

Draft Manuscript for Review. Please review online at http://mc.manuscriptcentral.com/oup/szbltn

Comprehensive gene expression analysis detects global reduction of proteasome subunits in schizophrenia

Journal:	*Schizophrenia Bulletin*
Manuscript ID	SZBLTN-ART-20-0339.R1
Manuscript Type:	Regular Article
Date Submitted by the Author:	14-Sep-2020
Complete List of Authors:	Hertzberg, Libi; Tel Aviv University, Sackler School of Medicine Maggio, Nicola Muler, Inna Yitzhaky, Assif Majer, Michael Haroutaunian, Vahram Zuk, Or; Hebrew University of Jerusalem, Statistics Katsel, Pavel; Icahn School of Medicine Department of Psychiatry, ; JJ Peters VAMC, Domany, Eytan Weiser, Mark; Sheba Medical Center, ;
Keywords:	Gene expression, Integrated analysis, Post mortem brain samples, Ubiquitin proteasome system

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Comprehensive gene expression analysis detects global reduction of proteasome subunits in schizophrenia

Proteasome subunits down-regulation in schizophrenia

Libi Hertzberg[a,b], M.D., Ph.D., Nicola Maggio[c,d], M.D., Ph.D., Inna Muler[e,f], BA, Assif Yitzhaky[a], B.Sc., Michael Majer[a], Vahram Haroutunian[h,i], Ph.D., Or Zuk[j], Ph.D., Pavel Katsel[h], Ph.D., Eytan Domany[a], Ph.D., Mark Weiser[g],M.D.

[a] Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot, Israel
[b] Shalvata Mental Health Center, Affiliated to the Sackler School of Medicine, Tel Aviv University, Israel,
[c] Department of Neurology, The Chaim Sheba Medical Center, Tel Hashomer, Ramat Gan, Israel.
[d] Department of Neurology and Neurosurgery, Sackler Faculty of Medicine and Sagol School of Neuroscience, Tel Aviv University, Israel
[e] Childhood Leukemia Research Institute and the Department of Pediatric Hemato-Oncology, Sheba Medical Center, Tel Hashomer, Ramat Gan, Israel,
[f] Human Molecular Genetics and Biochemistry, Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
[g] Department of Psychiatry, Chaim Sheba Medical Center, Ramat-Gan and the Sackler School of Medicine, Tel-Aviv University, Israel
[h] Departments of Psychiatry and Neuroscience, The Mount Sinai School of Medicine, New York, NY, USA
[i] Department of Psychiatry, James J Peters VA Medical Center, Bronx, NY, USA
[j] Department of Statistics, The Hebrew University of Jerusalem, Jerusalem, Israel

Correspondence: Dr. Libi Hertzberg, Shalvata Mental Health Center, Affiliated to the Sackler School of Medicine, Tel Aviv University, Israel. 13 Aliat Hanoar St. Hod Hasharon 45100, Israel. Email: libi.hertzberg@gmail.com; Tel: +972-9-7478554; Fax: +972-9-7478923

Previous presentation: the paper has been uploaded to a preprint server, BioRxiv, doi: https://doi.org/10.1101/853226

Abstract word count: 244

Text body word count: 4,000

Number of Tables and Figures: 5

ABSTRACT

BACKGROUND: A main challenge in the study of sschizophrenia is its high heterogeneity. While it is generally accepted that there exist several biological mechanisms that may define distinct schizophrenia subtypes, they haven't been identified yet. We performed comprehensive gene expression analysis to search for molecular signals that differentiate schizophrenia patients from healthy controls, and examined whether an identified signal was concentrated in a subgroup of the patients. **METHODS:** Transcriptome sequencing of 14 superior temporal gyrus (STG) samples of subjects with schizophrenia and 15 matched controls from the Stanley Medical Research Institute (SMRI) was performed. Differential expression and pathway enrichment analyses results were compared to an independent cohort. Replicability was tested on six additional independent datasets. **RESULTS:** The two STG cohorts showed high replicability. Pathway enrichment analysis of the down-regulated genes pointed to proteasome-related pathways. Meta-analysis of differential expression identified down-regulation of 12 of 39 proteasome subunit genes in schizophrenia. The signal of proteasome subunits down-regulation was replicated in six additional datasets (overall 8 cohorts with 267 schizophrenia and 266 control samples, from 5 brain regions). The signal was concentrated in a subgroup of the patients with schizophrenia. **CONCLUSIONS:** We detected global down-regulation of proteasome subunits in a subgroup of the patients with schizophrenia. We hypothesize that the down-regulation of proteasome subunits leads to proteasome dysfunction that causes accumulation of ubiquitinated proteins, which has been recently detected in a subgroup of schizophrenia patients. Thus, down-regulation of proteasome subunits might define a biological subtype of schizophrenia.

Key words: gene expression, post mortem brain samples, integrated analysis, ubiquitin proteasome system

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

INTRUDOCTION

Schizophrenia affects 1% of the population and has a complex pathophysiology that is far from being fully understood. A main challenge is its high genetic and clinical heterogeneity (1). While for years several subtypes definitions were in scientific and clinical use, the DSM-5 has omitted them after concluding that they do not predict the course of illness (2). However, it is generally accepted that there exist several mechanisms that may define distinct schizophrenia subtypes, which haven't been identified yet.

Recently, the ubiquitin proteasome system (UPS), governing protein degradation, has been associated with schizophrenia at both transcript (3–6) and protein levels (7,8), with tendency for down-regulation in schizophrenia brain samples. On the genomic level, UPS pathways were enriched with schizophrenia associated copy number variants (9), and the proteasome pathway was enriched in schizophrenia susceptibility genes (10).

Recent findings suggest a more pronounced role of the UPS in schizophrenia. Accumulation of ubiquitinated proteins has been identified in brain samples of a subgroup of schizophrenia patients in the STG, frontal cortex and prefrontal cortex samples (11). Another study detected elevated ubiquitinated proteins levels in the orbitofrontal cortex of schizophrenia patients (12). While ubiquitin binds to proteins (which become "ubiquitinated"), targeting them for proteasome degradation, proteasome dysfunction can cause accumulation of ubiquitinated proteins (13), as has been detected in schizophrenia. Recent studies of proteasome activity in schizophrenia have, however, yielded inconsistent results (12,14). Thus, while elevation of ubiquitinated protein levels seems to play a role in schizophrenia, it is not clear whether this is caused by dysfunction of the proteasome.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Two studies (7,14) have examined protein levels of proteasome subunits in schizophrenia, with three regulatory subunits found to be decreased in both (see Table 2). Several studies reported down-regulation of proteasome subunits genes (4,6,15,16), but only two subunits were found to be down-regulated in more than a single study (see Table 2). Thus, while there is evidence for down-regulation of both transcript and protein levels of proteasome subunits in schizophrenia, the results are currently sporadic.

A basic limitation of gene expression studies of schizophrenia is the fact that brain samples are usually composed of a mixture of cell types, which might dilute authentic changes. In addition, schizophrenia is highly heterogeneous (1) and typical changes in gene expression are modest (fold change range 1.03 – 1.33 (17)), which are thus difficult to detect. A relatively simple way to deal with these limitations is to perform a systematic comparison between independent datasets. Here we performed RNA-sequencing of STG samples from 14 schizophrenia and 15 control subjects from the SMRI. We applied pathway enrichment analysis to the list of genes detected as differentially expressed. We then used an independent cohort from the Mount Sinai School of Medicine (MSSM) to test the replicability of our results. A systematic meta-analysis of the SMRI and MSSM was applied to a subgroup of 39 inter-connected genes, which showed a tendency for down-regulation in schizophrenia. Six additional cohorts of different brain regions were used to further examine the robustness of our results. One of the six datasets was from the same patients as the SMRI data described

1
2
3 above, from a different brain region. Finally, we checked whether the signal
4
5
6 characterizes a subgroup of the patients.
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

METHODS

SMRI subjects

STG postmortem tissues from 15 subjects with schizophrenia and 15 healthy controls were obtained from the SMRI using approved protocols for tissue collection and informed consent (18). Samples were examined by a neuropathologist to exclude cerebral pathologies (19). Diagnoses were performed independently by two psychiatrists according to DSM-IV, and matched by age, gender, post-mortem interval (PMI) and pH (Table 1). RNA-sequencing was applied to 29 out of the 30 STG samples (one sample did not pass quality control – see below).

MSSM subjects

STG samples of 19 schizophrenia and 14 healthy controls were obtained from the Brain Bank of the Department of Psychiatry of the MSSM (Table 1). All cortical dissections and sample preparation were described previously (20–22); see also the Supplementary information. Gene expression was measured using Affymetrix HG-U133A microarrays.

RNA-sequencing

Brain regions were dissected at SMRI and delivered to Israel, where total RNA was isolated using the Trizol method. The concentration of total RNA and RNA Integrity Number (RIN) were measured. Samples with concentration $\geqslant$ 10 ng/μl and RIN $\geqslant$5 were selected for sequencing (one schizophrenia sample was excluded). The mean RIN was 6.3 (± 0.5), and the mean ratio of 260/280 was 1.6 (± 0.14). The mean total RNA yield was 15.4 μg (± 9.7). See Supplementary methods for a description of the libraries preparation protocol. For raw RNA-sequencing data description see Table 1S.

Mapping, quantification of gene expression levels and pre-processing

We used standard software tools for mapping fragments to the genome and for quantification of gene expression levels. See supplementary methods for full description. Pre-processing: Lowess correction was calculated (23). Then expression threshold was set to 6 (log scale) to reduce noise. Filtering: Genes with expression values below 6 in at least 80% of the samples were excluded from the analysis, leaving 16,482 genes after filtering (out of 23,715). We compared this method to filtering by the coefficient of variation (CV). CV was calculated for each of the 23,715 genes. A cutoff of CV=0.73 passed 16,482 genes. The two lists of 16,482 genes had 16,265 in common. Therefore the specific filtering method used did not affect our conclusions.

MSSM microarray pre-processing

MAS-5 algorithm was used for normalization. Lowess correction was then applied, expression levels below 20 were set to 20 and log2-transformation was applied. Probe-sets without assigned gene symbols were removed. 12,033 probe-sets were left for the rest of the analysis after filtering (out of 22,283), representing 8,542 gene symbols. Probe sets of the same gene were combined. For full details see supplementary methods.

Differential gene expression analysis

A linear model was fitted to each gene by a stepwise procedure (24), using the MATLAB function stepwiselm with default parameters. As pH did not differ significantly between schizophrenia and controls (Table 1), at first age, gender and PMI were included as covariates. Later we added pH as well (see Discussion). The model was then refitted using only the selected variables, including diagnosis. Finally, for each

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

gene, the diagnosis coefficient was statistically tested for being nonzero, implying an effect for schizophrenia, beyond any other effect of the covariates. This produced a t-statistic and a corresponding P-value. which were adjusted for multiple hypothesis testing using the false discovery rate (FDR) procedure (25). As the differentially expressed genes are subjected to further pathway enrichment analysis, a non-stringent FDR threshold of 15% was used. A standard 2-sample t-test was also performed; the results were very similar (Figure 1S).

Pathway enrichment analysis using GeneAnalytics

GeneAnalytics tool (26) was used for pathway enrichment analysis. GeneAnalytics leverages PathCards (http://pathcards.genecards.org/), which clusters thousands of pathways from multiple sources into Superpathways, in order to improve inferences and reduce redundancy. Superpathways are scored by log2-transformation of the binomial p-value, which is equivalent to a corrected p-value with significance level <0.05.

Differential expression STRING database network view

Network creation: Given a list of genes, a network is built. A network consists of genes (nodes) and genes' co-expression relations (edges). The co-expression relations data was downloaded from the STRING database, version 10.5 (27). Each such connection has a score between 0 and 1 that indicates the estimated likelihood that a given interaction is "biologically meaningful, specific and reproducible" (27). Only edges with STRING score greater than 0.1 are included in our network.

Differential expression network view: Given a network and gene expression data, of both patients and controls, the following steps are taken, for each gene:

1) The mean expression and standard deviation values, Mc and Sc, are calculated using the control samples only.

2) The mean expression, Mp, is calculated using the patients' samples.

3) Mp-Mc is calculated, the difference in the expression means between the two groups of samples.

4) The deviation from the control group is calculated, by: (Mp-Mc)/Sc

Finally, the network is displayed as an undirected graph, with each node colored according to the deviation described above, (Mp-Mc)/Sc. The edges represent co-expression relations. Only genes that have co-expression relations with other genes in the network are displayed.

RESULTS

UPS related pathways are enriched in the group of genes which are down-regulated in SMRI STG samples of individuals with schizophrenia

Differential expression analysis was performed, yielding 881 up-regulated and 986 down-regulated genes. In order to examine possible connection to antipsychotic medications, alcohol or substance use, we performed correlation analyses between the expression pattern of the differentially expressed genes and Fluphenazine equivalent dosage, substance use and alcohol use measures. Correlation analyses for Fluphenazine equivalent dosage and alcohol use did not reveal any significant association with differential expression. Correlation analysis for substance use detected two down-regulated genes (out of 986) with statistically significant correlated expression (supplementary methods and Figures 2S-4S).

Pathway enrichment analysis was applied separately to the up and down-regulated genes. Results are presented in Tables 3S-4S for the up-regulated and down-regulated genes, respectively. Out of 49 pathways enriched in the down-regulated genes, five are directly UPS related (marked in Table 4S). While several pathways have higher enrichment scores, we focus on the UPS and proteasome-related pathways, since five such pathways were enriched, and eleven closely related additional pathways were also enriched (Table 4S). One of these pathways is Class I MHC Mediated Antigen Processing and Presentation, where proteins degraded by the proteasome are a major source of peptides presented by MHC class I molecules (28). ~~and several closely related additional pathways were also enriched (see Table 4S).~~

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

The UPS signal is highly replicated in the MSSM STG samples

Our findings are replicated in the STG of the independent MSSM cohort of elderly subjects. We first examined whether the two datasets are comparable. Though microarrays differ from RNA-seq in their captured features, there was a significant positive correlation of the t-statistics (schizophrenia vs. controls) between SMRI and MSSM across 7,498 genes common to both platforms (Figure 1A).

We next repeated the differential expression and pathway enrichment analyses in the MSSM cohort. 919 genes and 794 genes were found to be up-regulated and down-regulated in schizophrenia, respectively. MSSM and SMRI differentially expressed genes significantly overlap (hypergeometric P-values: $9.8*10^{-7}$, $1.1*10^{-19}$ for the up-regulated and down-regulated genes, respectively; see Figure 1B).

Pathway enrichment analysis yielded 27 and 48 enriched pathways in up and down-regulated genes, respectively (results are listed in Tables 5S and 6S). Intersecting SMRI 49 enriched pathways with 48 from MSSM in the down-regulated genes yields 30 shared pathways; see Table 4S (hypergeometric p-value: $2.5*10^{-36}$). Four out of the five SMRI enriched UPS pathways were enriched also in the MSSM. A similar analysis of the up-regulated genes yields a hypergeometric p-value of $1.03*10^{-6}$. Interestingly, one of the pathways that were enriched in the MSSM up-regulated genes is Metabolism of Proteins, which contains UPS-related genes. Thus, while a subgroup of the pathway genes is up-regulated, another is down-regulated. For further details see Supplementary Information.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

A network view of the UPS identifies down-regulation of a tightly connected cluster of proteasome subunits

To further explore the UPS differential expression, we used differential expression network view for SMRI (see Methods). It was applied to the Ubiquitin-Proteasome Dependent Proteolysis GeneAnalytics "superpathway" (26), which is representative of the UPS and was significantly enriched in both SMRI and MSSM (see Table 4S). The network view includes all 69 pathway genes for which network data was available from STRING (29), and not only those 27 genes that were found to be down-regulated. As can be seen in Figure 1C, there is a cluster of tightly inter-connected genes which are mostly down-regulated in schizophrenia (bluish colours of the nodes). Interestingly, this cluster is composed of proteasome subunits, as shown in Figure 1D. The same analysis of the MSSM yields a similar view (Figure 5S).

Meta-analysis of SMRI and MSSM datasets identifies down-regulation of multiple proteasome subunits in STG samples of subjects with schizophrenia

We performed a meta-analysis of the expression of each of the 39 proteasome subunit genes, whose expression has been measured by both SMRI and MSSM (see Supplementary Methods). The list of proteasome subunits genes, meta-analysis results and comparisons to previous gene expression and protein-level studies are summarized in Table 2. Overall 12 out of 39 subunit genes were found to be down-regulated.

Down-regulation signal of proteasome subunits in schizophrenia is replicated in 6 independent datasets of 5 different brain regions

To examine whether down-regulation of proteasome subunits is specific to the STG we repeated the differential expression network analysis of the 39 proteasome subunit genes using 6 additional datasets (fully described in the supplementary information):

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

dorso-lateral prefrontal cortex (DLFPC) samples from Arion 2015 (6) and from Ramaker 2017 (30), STG samples from Barnes 2011 (31), Cerebellum samples from Chen 2018 (32), Brodmann area 23 (BA23) samples from the SMRI cohort, and Brodmann area 10 (BA10) samples from Mycox 2009 (33). The results are presented in Figure 2. The DLPFC samples of Arion 2015 (Figure 2A) exhibit pronounced down-regulation, while in the DLPFC samples of Ramaker 2017 (Figure 2B) the signal is weaker, though present in most of the genes; the binomial p-value for the number of genes with (even slightly) reduced expression versus the control group is p= 6.4 10^{-6}). Interestingly, while Ramaker 2017 (30) used brain samples composed of mixture of cells, Arion 2015 (6) used laser microdissection to capture pyramidal neurons. Thus, the difference in down-regulation might be due to dilution of the signal, caused by the mixture of cell types used in Ramaker 2017. The Cerebellum samples from Chen 2018 (32) (Figure 2D), BA10 samples from Mycox 2009 (Figure 2E) and BA23 SMRI (Figure 2F) show clear tendency for down-regulation (binomial p-values 6.9 10^{-7}, 1.2 10^{-5} and 0.04, respectively), with modest magnitude (mostly less than 1 standard deviation). STG samples of Barnes 2011 (31) (Figure 2C) show a similar pattern. Down-regulation might be specific to neurons or subtypes of neurons; as the brain samples in these datasets are of mixture of cells, the signal might be diluted. Overall, this analysis replicates the signal of down-regulation of multiple proteasome subunits, in both the STG and additional 4 brain regions.

Down-regulation of proteasome subunits in schizophrenia is concentrated in a subgroup of the patients

To explore whether the signal is concentrated in a subgroup of the patients, we applied fold change analysis of the 12 down-regulated proteasome subunits (listed in

Table 2) to each of the SMRI schizophrenia samples. The results are plotted in Figure 3A. Half of the patients (7/14, “Group 2”; marked blue along the x-axis) show down-regulation tendency (bluish colors) of most of the 12 proteasome subunits genes, while the others (“Group 1”; marked green) show fold change values closer to 1, for most genes. The same analysis of the Arion 2015 dataset, of microdissected pyramidal neurons, yields even more pronounced distinction (Figure 3B). A similar picture emerges for the other 6 datasets (Figure 7S). Support for this observation comes from a recent study (34), where transcriptomics analysis of 189 DLPFC samples of schizophrenia patients vs. 206 healthy controls identified two molecular subtypes of schizophrenia. In “Type 1” (about half of the patients) four differentially expressed genes (schizophrenia vs. controls) were detected, and in “Type 2” more than 3000. When examining the list of differentially expressed genes (Supplemental Table 3B), 28 proteasome subunits were differentially expressed, all down-regulated, in "Type 2", while no proteasome subunit genes were differentially expressed in "Type 1".

We then applied a similar analysis as in (34), and compared each of “Group 1” and “Group 2” samples to the controls, separately. Differential expression analysis was applied to the 47 SMRI measured proteasome subunit genes, in each of the two groups. While in “Group 1” no differentially expressed genes were found, in “Group 2” 23 proteasome subunits were found to be differentially expressed (FDR < 15%; Table 8S). We conclude that proteasome subunits down-regulation characterizes about half of the patients with schizophrenia.

DISCUSSION

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

The main finding of our study is a global down-regulation of multiple proteasome subunits in post mortem brain samples of individuals with schizophrenia. Although several scenarios may be possible, a reasonable model (Figure 3C) is that given a predisposition to schizophrenia, certain (unknown) factors lead to (1) down-regulation of multiple proteasome subunits in about half the patients. This in turn leads to (2) proteasome dysfunction which causes (3) accumulation of ubiquitinated proteins. We discuss below the evidence relevant to each of the hypotheses (1)-(3) of our model.

Hypothesis (1) is supported by our main finding, which was replicated in 8 datasets of 5 different brain regions. We observed that the signal characterizes about half the patients. This observation is supported by (34), where two molecular subtypes of schizophrenia were detected, one ("Type 2") with 28 down-regulated proteasome subunits genes, and another ("Type 1"), without dysregulation of these genes. In "Type 2" more than 3000 genes (about 25% of those measured) were dysregulated (up- and down-regulated ratio close to 1:1). Thus, the fact that 28 proteasome subunits genes are dysregulated is somewhat less surprising. However, as all the 28 were down-regulated, it makes the concordance with our results significant.

The fact that several studies identified decreased protein levels of proteasome subunit genes (Table 2) supports hypothesis (2), of proteasome dysfunction in schizophrenia. However, it wasn't established whether the lower protein levels are caused by lower expression of the coding genes. Moreover, previous studies of proteasome activity in schizophrenia yielded inconsistent results. While in (14) intra-cellular compartment-specific dysfunction in STG samples was found, no change has been detected in neither blood or brain in (12). A possible explanation of this inconsistency is that the signal is

specific not only to a subgroup of the patients, but also to neurons or subtypes of neurons, and thus diluted. This is supported by our analysis of Arion 2015 dataset, of laser microdissected neurons (6), where higher magnitude of down-regulation was detected (Figure 2). This could also explain why the signal hasn't been detected by many previous relevant gene expression studies. Actually, if we look at some of the datasets (for example, in Figure 2B-E), each proteasome subunit is not pronouncedly down-regulated. Only the analysis of the proteasome subunits as a group, measured in multiple datasets, enabled the detection of the global down-regulation signal.

Hypothesis (3), of accumulation of ubiquitinated proteins in schizophrenia, comes from two recent studies. In (11), accumulation of ubiquitinated proteins has been identified for about half the patients in the STG, frontal cortex and prefrontal cortex samples. In (12), ubiquitinated protein levels were found to be elevated in the orbitofrontal cortex of schizophrenia patients. While the fact that dysfunction of proteasome can cause accumulation of ubiquitinated proteins (13) suggests a causative connection between hypotheses (2) and (3), this link wasn't examined in schizophrenia.

Interestingly, lower pH was detected in brains with accumulation of ubiquitinated proteins in (11) and was also associated with elevated ubiquitinated protein levels in (12). As pH did not differ significantly between schizophrenia and controls in neither SMRI nor MSSM (Table 1), it was not included as a covariate in the differential expression analysis. However, it may have more delicate associations with proteasome subunit genes' expression, possibly in a subgroup of the patients. To examine this we performed a correlation analysis between pH levels and mean fold change of the 12 down-regulated proteasome subunits (listed in Table 2), in 7 datasets (for which pH information was available) and found a clear tendency for positive correlation (Figure

8S). This is concordant with the association between lower pH and accumulation of ubiquitinated proteins shown in (11,12) and gives indirect support to our hypothesis that this accumulation is caused by down-regulation of the proteasome subunits.

In order to further explore the association between pH and proteasome subunits expression, we repeated the step-wise linear regression for both SMRI and MSSM for the 12 down-regulated proteasome subunits, including pH as a covariate (with age, PMI and gender) for both cohorts and RIN was included for the SMRI (Tables 9S-10S). This linear regression analysis gave moderate results in terms of the magnitude and statistical significance of the genes' down-regulation, when compared to standard two-sided t-test (Figure 9S). However, the clear tendency for down-regulation remained, with statistical significance. We thus conclude that pH, or the the other included potential conounding factors, cannot solely explain our observed decreased expression of the proteasome subunits.

In order to further explore the notion of subtypes of schizophrenia, we compared the mean fold change (FC) of the 12 down-regulated proteasome subunits to the polygenic risk score (PRS) of the STG SMRI samples (see Supplementary Information and Figure 10S). No statistically significant correlation was found (p-value 0.56); but when a single outlier sample (with largest PRS and smallest FC) was omitted, we did get a significant Pearson correlation of 0.67 (p-value 0.03), supporting our hypothesis of association between proteasome FC and clinical characteristics. However, since correlation was calculated from only 11 samples, this needs further investigation.

As described in (11), the accumulated ubiquitinated proteins were enriched with nervous system development related pathways, suggesting its possible relation to disease pathogenesis through disruption of relevant pathways. In addition, clinical symptoms were correlated with two ubiquitin conjugation genes' expression in patients' peripheral blood (35). These findings might suggest that our hypothesized model defines a biological and clinical subtype of schizophrenia. In this context we note that Bortezomib, a proteasome inhibitor used in the treatment of cancer, is not known to cause psychosis when given to glioblastoma patients (where the brain-blood-barrier is disrupted) (36–38). This seemingly suggests that proteasome dysfunction is not the cause of the symptoms seen in schizophrenia. In addition, pathways that apparently are not connected to the proteasome/UPS were found to be dysregulated, both by us and by (34), suggesting there are other mechanisms that underlie the pathogenesis of schizophrenia. However, interestingly, four of six pathways repeatedly found as dysregulated in schizophrenia (reviewed in (39)), involve the UPS: presynaptic function (40), signalling (41), oxidative stress (42) and cellular immune mechanisms (43). In addition, it was recently shown that antipsychotics modulate UPS-related protein levels in oligodendrocytes (44). However, it is still not clear whether the UPS has a causal role in schizophrenia and further study is needed to decipher this connection.

~~However, additional studies are needed in order to further explore this connection.~~

Our study is limited by several features. Every postmortem study represents only a snapshot at the end of life. This is especially relevant in schizophrenia, as its pathogenesis is probably rooted in early development (45). The fact that we compare independent cohorts of both relatively young and elderly subjects strengthens the

validity of the results, but doesn't fully overcome this limitation. There is also the question of pharmacotherapy, as exposure to antipsychotics might affect gene expression. We found no significant correlation between Fluphenazine equivalent dose and expression levels. In addition, the fact that the subjects of the cohorts significantly differ in age suggests that duration of exposure to antipsychotics is unlikely to influence proteasome subunits expression substantively. The replication of the detected signal in 8 cohorts from 5 brain regions significantly increases the validity and generalizability of this signal. As gene expression does not always correlate with the levels of the coded proteins, the fact that we measure gene expression alone is a serious limitation, which causes difficulties in making definitive conclusions regarding the biological consequences of the results. While several studies detected decreased protein levels of proteasome subunits (7,14), the results were not fully consistent and the recent proteasome activity studies in schizophrenia were not consistent either, as described above. Thus, further study is needed in order to decipher the consequences of the global down-regulation of proteasome subunits we detect in schizophrenia, in terms of protein levels and proteasome activity.

Overall, we detect global down-regulation of proteasome subunits in schizophrenia, which characterizes about half of the patients. Based on ours and others' recent findings we present a hypothesized model for a mechanism that defines a biological, and maybe also clinical, subtype of schizophrenia. This has the potential to lead to a better understanding of the biological and clinical subtypes of schizophrenia and to finding novel diagnostic and therapeutic tools.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

DISCLOSURES

All authors declare that they have no conflict of interest

FUNDING

This work was supported by a 2014 Brain and Behavior Research Foundation Young Investigator Grant, the Leir Charitable Foundation, the Veterans Administration MIRECC and the National Institutes of Health Neurobiobank (HHSN271201300031).

ACKNOWLEDGMENTS

We thank Professor Shai Izraeli for his advice and for providing the conditions for RNA isolation and quality control measurements. We thank Professor Harker Rhodes for his advice and insight regarding the interpretation of the results, and for sharing the results of the analysis he performed. We also thank Maree Webster from the SMRI for her kind advice regarding the process of RNA isolation and Tian Ge for his helpful advice regarding the calculation of the PRS.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

1. Tsuang MT, Stone WS, Faraone S V. (2001): Genes, environment and schizophrenia. *Br J Psychiatry* 178: s18–s24.

2. Impact of DSM-5 Changes on the Diagnosis and Acute Treatment of Schizophrenia (n.d.): Retrieved February 21, 2020, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4393695/

3. Middleton FA, Mirnics K, Pierri JN, Lewis DA, Levitt P (2002): Gene expression profiling reveals alterations of specific metabolic pathways in schizophrenia. *J Neurosci* 22: 2718–29.

4. Altar CA, Jurata LW, Charles V, Lemire A, Liu P, Bukhman Y, *et al.* (2005): Deficient Hippocampal Neuron Expression of Proteasome, Ubiquitin, and Mitochondrial Genes in Multiple Schizophrenia Cohorts. *Biol Psychiatry* 58: 85–96.

5. Bousman CA, Chana G, Glatt SJ, Chandler SD, Lucero GR, Tatro E, *et al.* (2010): Preliminary evidence of ubiquitin proteasome system dysregulation in schizophrenia and bipolar disorder: convergent pathway analysis findings from two independent samples. *Am J Med Genet B Neuropsychiatr Genet* 153B: 494–502.

6. Arion D, Corradi JP, Tang S, Datta D, Boothe F, He A, *et al.* (2015): Distinctive transcriptome alterations of prefrontal pyramidal neurons in schizophrenia and schizoaffective disorder. *Mol Psychiatry* 20: 1397–1405.

7. Scott MR, Rubio MD, Haroutunian V, Meador-Woodruff JH (2016): Protein Expression of Proteasome Subunits in Elderly Patients with Schizophrenia. *Neuropsychopharmacology* 41: 896–905.

8. Rubio MD, Wood K, Haroutunian V, Meador-Woodruff JH (2013): Dysfunction of

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

the Ubiquitin Proteasome and Ubiquitin-Like Systems in Schizophrenia. *Neuropsychopharmacology* 38: 1910–1920.

9. Pescosolido MF, Gamsiz ED, Nagpal S, Morrow EM (2013): Distribution of Disease-Associated Copy Number Variants Across Distinct Disorders of Cognitive Development. *J Am Acad Child Adolesc Psychiatry* 52: 414-430.e14.

10. Chang X, Lima L de A, Liu Y, Li J, Li Q, Sleiman PMA, Hakonarson H (2018): Common and Rare Genetic Risk Factors Converge in Protein Interaction Networks Underlying Schizophrenia. *Front Genet* 9. https://doi.org/10.3389/fgene.2018.00434

11. Nucifora LG, MacDonald ML, Lee BJ, Peters ME, Norris AL, Orsburn BC, *et al.* (2019): Increased Protein Insolubility in Brains From a Subset of Patients With Schizophrenia. *Am J Psychiatry* appi.ajp.2019.1.

12. Bousman CA, Luza S, Mancuso SG, Kang D, Opazo CM, Mostaid MS, *et al.* (2019): Elevated ubiquitinated proteins in brain and blood of individuals with schizophrenia. *Sci Rep* 9. https://doi.org/10.1038/s41598-019-38490-1

13. Keller JN, Huang FF, Markesbery WR (2000): Decreased levels of proteasome activity and proteasome expression in aging spinal cord. *Neuroscience* 98: 149–56.

14. Scott MR, Meador-Woodruff JH (2019): Intracellular compartment-specific proteasome dysfunction in postmortem cortex in schizophrenia subjects. *Mol Psychiatry*. https://doi.org/10.1038/s41380-019-0359-7

15. Vawter MP, Barrett T, Cheadle C, Sokolov BP, Wood WH, Donovan DM, *et al.* (2001): Application of cDNA microarrays to examine gene expression differences in schizophrenia. *Brain Res Bull* 55: 641–650.

16. Chu TT, Liu Y, Kemether E (2009): Thalamic transcriptome screening in three psychiatric states. *J Hum Genet* 54: 665–675.

17. Fromer M, Roussos P, Sieberts SK, Johnson JS, Kavanagh DH, Perumal TM, *et al.* (2016): Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci* 19: 1442–1453.

18. Torrey EF, Webster M, Knable M, Johnston N, Yolken RH (2000): The stanley foundation brain collection and neuropathology consortium. *Schizophr Res* 44: 151–155.

19. Torrey EFF, Webster M, Knable M, Johnston N, Yolken RH (2000): The stanley foundation brain collection and neuropathology consortium. *Schizophr Res* 44: 151–155.

20. Hakak Y, Walker JR, Li C, Wong WH, Davis KL, Buxbaum JD, *et al.* (2001): Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia, 2001/04/11. *Proc Natl Acad Sci U S A* 98: 4746–4751.

21. Katsel PL, Davis KL, Haroutunian V (2005): Large-scale microarray studies of gene expression in multiple regions of the brain in schizophrenia and Alzheimer's disease, 2005/03/31. *Int Rev Neurobiol* 63: 41–82.

22. Katsel P, Davis KL, Gorman JM, Haroutunian V (2005): Variations in differential gene expression patterns across multiple brain regions in schizophrenia, 2005/06/01. *Schizophr Res* 77: 241–252.

23. Ballman K V., Grill DE, Oberg AL, Therneau TM (2004): Faster cyclic loess: normalizing RNA arrays via linear models. *Bioinformatics* 20: 2778–2786.

24. Pope PT, Webster JT (1972): The Use of an *F* -Statistic in Stepwise Regression Procedures. *Technometrics* 14: 327–340.

25. Benjamini Y, Hochberg Y (1995): Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc* 57: 289–300.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

26. Ben-Ari Fuchs S, Lieder I, Stelzer G, Mazor Y, Buzhor E, Kaplan S, *et al.* (2016): GeneAnalytics: An Integrative Gene Set Analysis Tool for Next Generation Sequencing, RNAseq and Microarray Data. *Omi A J Integr Biol* 20: 139–151.

27. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, *et al.* (2017): The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 45: D362–D368.

28. Stoltze L, Nussbaum AK, Sijts A, Emmerich NPN, Kloetzel PM, Schild H (2000): The function of the proteasome system in MHC class I antigen processing. *Immunology Today*, vol. 21 21: 317–319.

29. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, *et al.* (2015): STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 43: D447–D452.

30. Ramaker RC, Bowling KM, Lasseigne BN, Hagenauer MH, Hardigan AA, Davis NS, *et al.* (2017): Post-mortem molecular profiling of three psychiatric disorders. *Genome Med* 9: 72.

31. Barnes MR, Huxley-Jones J, Maycox PR, Lennon M, Thornber A, Kelly F, *et al.* (2011): Transcription and pathway analysis of the superior temporal cortex and anterior prefrontal cortex in schizophrenia. *J Neurosci Res* 89: 1218–1227.

32. Chen C, Meng Q, Xia Y, Ding C, Wang L, Dai R, *et al.* (2018): The transcription factor POU3F2 regulates a gene coexpression network in brain tissue from patients with psychiatric disorders. *Sci Transl Med* 10. https://doi.org/10.1126/scitranslmed.aat8178

33. Maycox PR, Kelly F, Taylor A, Bates S, Reid J, Logendra R, *et al.* (2009): Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function. *Mol Psychiatry* 14: 1083–1094.

34. Bowen EFW, Burgess JL, Granger R, Kleinman JE, Rhodes CH (2019): DLPFC transcriptome defines two molecular subtypes of schizophrenia. *Transl Psychiatry* 9: 1–10.

35. Bousman CA, Chana G, Glatt SJ, Chandler SD, May T, Lohr J, *et al.* (2010): Positive symptoms of psychosis correlate with expression of ubiquitin proteasome genes in peripheral blood. *Am J Med Genet B Neuropsychiatr Genet* 153B: 1336–41.

36. Odia Y, Kreisl TN, Aregawi D, Innis EK, Fine HA (2015): A phase II trial of tamoxifen and bortezomib in patients with recurrent malignant gliomas. *J Neurooncol* 125: 191–195.

37. Raizer JJ, Chandler JP, Ferrarese R, Grimm SA, Levy RM, Muro K, *et al.* (2016): A phase II trial evaluating the effects and intra-tumoral penetration of bortezomib in patients with recurrent malignant gliomas. *J Neurooncol* 129: 139–146.

38. Bota DA, Eroglu Z, Reardon DA, Fu BD, Norfleet J, Desjardins A, *et al.* (2011): Phase II clinical trial of bortezomib and bevacizumab combination in recurrent glioblastoma. *J Clin Oncol* 29: 2056–2056.

39. Kumarasinghe N, Tooney PA, Schall U (2012): Finding the needle in the haystack: a review of microarray gene expression research into schizophrenia. *Aust N Z J Psychiatry* 46: 598–610.

40. Hegde AN (2017): Proteolysis, synaptic plasticity and memory. *Neurobiol Learn Mem* 138: 98–110.

41. Ehlers MD (2003): Activity level controls postsynaptic composition and signaling via the ubiquitin-proteasome system. *Nat Neurosci* 6: 231–242.

42. Shang F, Taylor A (2011, July 1): Ubiquitin-proteasome pathway and cellular responses to oxidative stress. *Free Radical Biology and Medicine*, vol. 51. Elsevier

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Inc., pp 5–16.

43. Kammerl IE, Meiners S (2016): Proteasome function shapes innate and adaptive immune responses. *Am J Physiol Cell Mol Physiol* 311: L328–L336.

44. Seabra G, de Almeida V, Reis-de-Oliveira G, Crunfli F, Antunes ASLM, Martins-de-Souza D (2020): Ubiquitin–proteasome system, lipid metabolism and DNA damage repair are triggered by antipsychotic medication in human oligodendrocytes: implications in schizophrenia. *Sci Rep* 10: 12655.

45. Gogtay N, Vyas NS, Testa R, Wood SJ, Pantelis C (2011): Age of onset of schizophrenia: perspectives from structural neuroimaging studies. *Schizophr Bull* 37: 504–13.

46. Vawter MP, Barrett T, Cheadle C, Sokolov BP, Wood WH, Donovan DM, *et al.* (2001): Application of cDNA microarrays to examine gene expression differences in schizophrenia. *Brain Res Bull* 55: 641–50.

47. Duyckaerts C, Godefroy G (2000): Voronoi tessellation to study the numerical density and the spatial distribution of neurones. *J Chem Neuroanat* 20: 83–92.

48. Tsafrir D, Tsafrir I, Ein-Dor L, Zuk O, Notterman DA, Domany E (2005): Sorting points into neighborhoods (SPIN): data analysis and visualization by ordering distance matrices. *Bioinformatics* 21: 2301–2308.

1
2
3
4
5
6
7
8
9
10

TABLES AND FIGURES

Table 1. Subjects' characteristics. Average values (standard deviation). To compare schizophrenia and controls two-sided t-test p-values were calculated for the continuous variables and Fisher's exact test p-value was calculated for M/F ratio

Characteristics	Schizophrenia	Control	P-value
SMRI subjects			
Number of subjects	14	15	
Gender (M/F)	9/5	9/6	1
Age (years)	43.6 (13)	48.1 (10.6)	0.32
Brain pH	6.2 (0.3)	6.3 (0.2)	0.35
RIN	6.2 (0.5)	6.4 (0.5)	0.17
PMI (minutes)	2052 (900)	1424 (596)	0.03
MSSM subjects			
Number of subjects	19	14	
Gender (M/F)	14/5	5/9	0.04
Age (years)	77.4 (10.9)	82.4 (12.7)	0.23
Brain pH	6.4 (0.2)	6.6 (0.3)	0.08
PMI (minutes)	814 (499)	460 (429)	0.04

11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

Table 2. Proteasome subunits differential gene and protein level expression, in previous studies and in our meta-analysis. Previous gene expression studies' results were listed only for genes which were detected as differentially expressed in more than one study. Down-regulation findings are highlighted in blue. In the meta-analysis, a gene is defined as down-regulated if its summary measure is lower than zero and the confidence interval doesn't cross zero

#	Proteasome subunit genes	Previous gene expression studies	Previous protein level studies (7,14)	Our meta-analysis (SMRI+MSSM)	SMRI+ MSSM meta-analysis summary measure [confidence interval]
	Structural subunits				
	20S core α subunits				
1	PSMA1 (also named 20S α1)	Down-regulated in 2 studies	Not measured	unchanged	-0.37 [-0.97, 0.21]

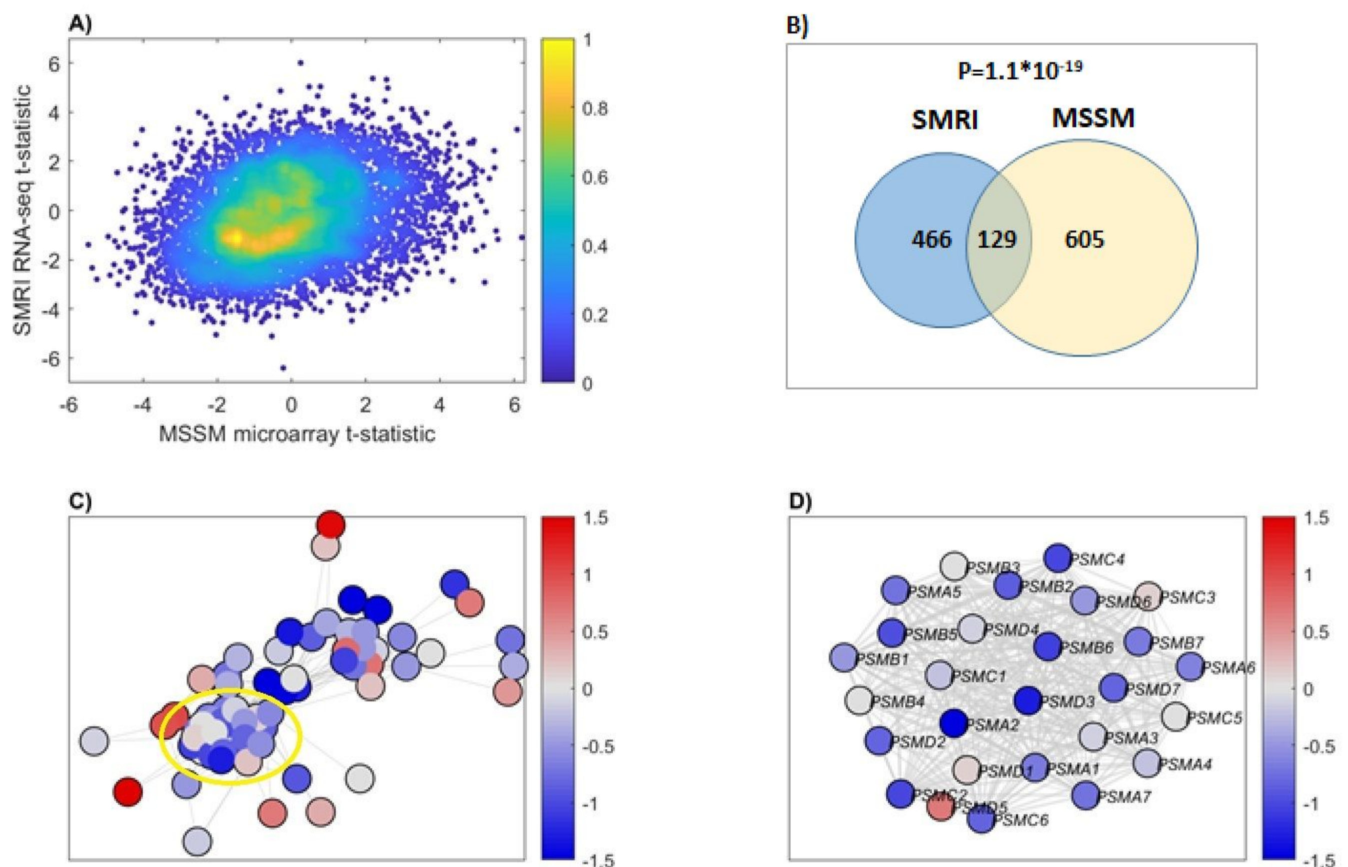
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

		(4,46)			
2	PSMA2 (20S α2)		Not measured	Down-regulation	-1.13 [-1.68, -0.59]
3	PSMA3 (20S α3)		Not measured	unchanged	-0.63 [-1.67, 0.4]
4	PSMA4 (20S α4)		Not measured	unchanged	-0.43 [-0.9, 0.07]
5	PSMA5 (20S α5)		Not measured	Down-regulation	-0.61 [-1.13, -0.09]
6	PSMA6 (20S α6)		unchanged in (14)	Down-regulation	-0.63 [-1.15, -0.12]
7	PSMA7 (20S α7)		Not measured	Down-regulation	-0.79 [-1.32, -0.27]
	Catalytic subunits				
	20S core β subunits				
8	PSMB1 (20S β1)		Not measured	unchanged	-0.17 [-0.73, 0.37]
9	PSMB2 (20S β2)		Down-regulation trend (7) (p=0.08); unchanged in (14)	Down-regulation	-0.62 [-1.13, -0.11]
10	PSMB3 (20S β3)		Not measured	unchanged	-0.25 [-0.75, 0.25]
11	PSMB4 (20S β4)		Not measured	unchanged	-0.03 [-0.53, 0.46]
12	PSMB5 (20S β5)		unchanged (7,14)	Down-regulation	-0.73 [-1.28, -0.18]
13	PSMB6 (20S β6)		Not measured	unchanged	-0.13 [-1.25, 0.97]
14	PSMB7 (20S β7)		Not measured	unchanged	-0.37 [-0.87, 0.13]
	Immunoproteasome β subunit genes				
	PSMB8 (20S β5i)		unchanged (7,14)	unchanged in SMRI; absent in MSSM	
15	PSMB9 (20S β1i)		unchanged (7)	unchanged	-0.04 [-0.45, 0.54]
1	PSMB10 (20S β2i)		unchanged (7,14)	unchanged	0.16 [-0..38,

6					0.71]
	Regulatory subunits				
	19S AAA-ATPase subunits (Rpt)				
17	PSMC1 (19S Rpt2)		unchanged in (7); Down-regulated in (14)	unchanged	0.02 [-0.48, 0.52]
18	PSMC2 (19S Rpt1)		Down-regulated in two studies (7,14)	Down-regulation	-0.93 [-1.46, -0.4]
19	PSMC3 (19S Rpt5)		Unchanged in (7); Down-regulated in (14)	unchanged	-0.11 [-0.62, 0.38]
20	PSMC4 (19S Rpt3)		Down-regulated in two studies (7,14)	Down-regulation	-0.67 [-1.19, -0.15]
21	PSMC5 (19S Rpt6)		Down-regulated in two studies (7,14)	unchanged	0.02 [-0.48, 0.52]
22	PSMC6 (19S Rpt4)	Down-regulated in two studies (4,6)	unchanged (7); Down-regulated in (14)	Down-regulation	-0.83 [-1.36, -0.3]
	19S non-ATPase subunits (Rpn)				
23	PSMD1 (19S Rpn2)		Not measured	unchanged	0.07 [-0.42, 0.58]
24	PSMD2 (19S Rpn1)		Not measured	unchanged	-0.03 [-1.49, 1.41]
25	PSMD3 (19S Rpn3)		Not measured	unchanged	-0.34 [-0.87, 0.18]
26	PSMD4 (19S Rpn10)		unchanged (7)	unchanged	0.09 [-0.41, 0.59]
27	PSMD5		Not measured	unchanged	0.17 [-0.69, 1.05]
28	PSMD6 (19S Rpn7)		Not measured	Down-regulation	-0.62 [-1.13, -0.1]
29	PSMD7 (19S Rpn8)		Not measured	unchanged	-0.07 [-1.02, 0.86]
30	PSMD8 (19S Rpn12)		Not measured	unchanged	-0.39 [-0.9, 0.11]

31	PSMD9 (19S Rpn4)		Not measured	unchanged	-0.39 [-1.42, 0.63]
32	PSMD10		Not measured	unchanged	0.13 [-0.36, 0.63]
33	PSMD11 (19S Rpn6)		unchanged (7)	Down-regulation	-0.73 [-1.25, -0.21]
34	PSMD12 (19S Rpn5)		Not measured	unchanged	-0.23 [-0.74, 0.26]
35	PSMD13 (19S Rpn9)		Not measured	unchanged	0.05 [-0.58, 0.69]
36	PSMD14 (19S Rpn11)		unchanged (7)	Down-regulation	-0.89 [-1.42, -0.37]
	11S subunits				
37	PSME1 (11S α)		Down-regulated in (7); unchanged in (14)	unchanged	-0.33 [-0.84, -0.16]
38	PSME2 (11S β)		unchanged (7,14)	unchanged	0.29 [-0.65, 1.24]
39	PSME3 (11S gamma)		unchanged (7)	unchanged	-0.07 [-0.85, 0.7]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31



32 **Figure 1. A) Binned density scatter plot comparing the t-statistics for case versus control**
33 **differential expression between the independent MSSM replication cohort assayed on**
34 **microarrays and the SMRI RNA-seq data;** correlation between the statistics is 0.28 (P =
35 4.7*10−133). The colorbar represents the density in each cell, calculated by voronoi procedure
36 (47) and normalized to values between 0 (minimal density) and 1 (maximal density). **B)**
37 **Hypergeometric p-value calculation for the intersection between SMRI and MSSM down-**
38 **regulated genes.** The 986 SMRI and 794 MSSM down-regulated genes were intersected with
39 the 7,498 genes that are present in both cohorts, yielding 595 SMRI and 734 MSSM down-
40 regulated genes, with 129 shared genes. **C) SMRI Differential expression network view for**
41 **Ubiquitin-Proteasome Dependent Proteolysis superPathway.** A node's color corresponds to
42 the deviation of expression from the control samples group, in terms of standard deviation units.
43 The edges represent STRING database co-expression relations. Only genes that have co-
44 expression relations with other genes in the network are displayed. A subgroup of highly-
45 interconnected genes, coding for proteasome subunits, is circled **D) Zoom in on proteasome**
46 **subunits.** The same plot as in C), for a subgroup of highly-interconnected genes coding for
47 proteasome subunits (circled in C))
48
49
50
51
52
53
54
55
56
57
58
59
60

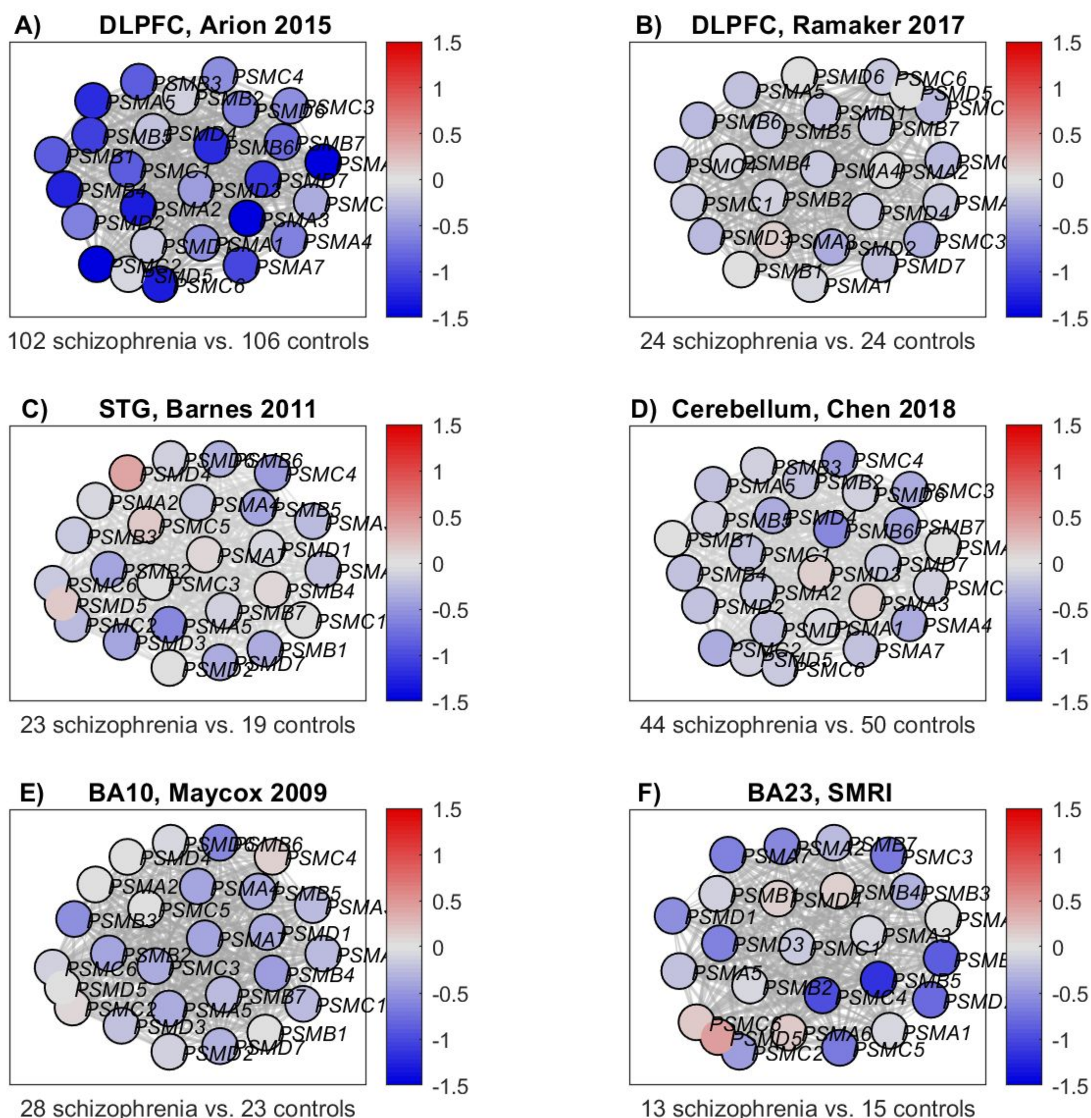


Figure 2. Proteasome subunits differential expression network view: The nodes' colors correspond to the deviation from the group of the control samples, in terms of standard deviation units. The edges represent STRING database co-expression relations. Only genes that have co-expression relations with other genes in the network are displayed. **A) DLPFC, Arion 2015 dataset** (6)**. B) DLPFC, Ramaker 2017 dataset** (30)**. C) STG, Barnes 2011 dataset** (31). **D) Cerebellum, Chen 2018** (32)**. E) BA10, Mycox 2009 dataset** (33) **F) BA23, SMRI dataset**

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

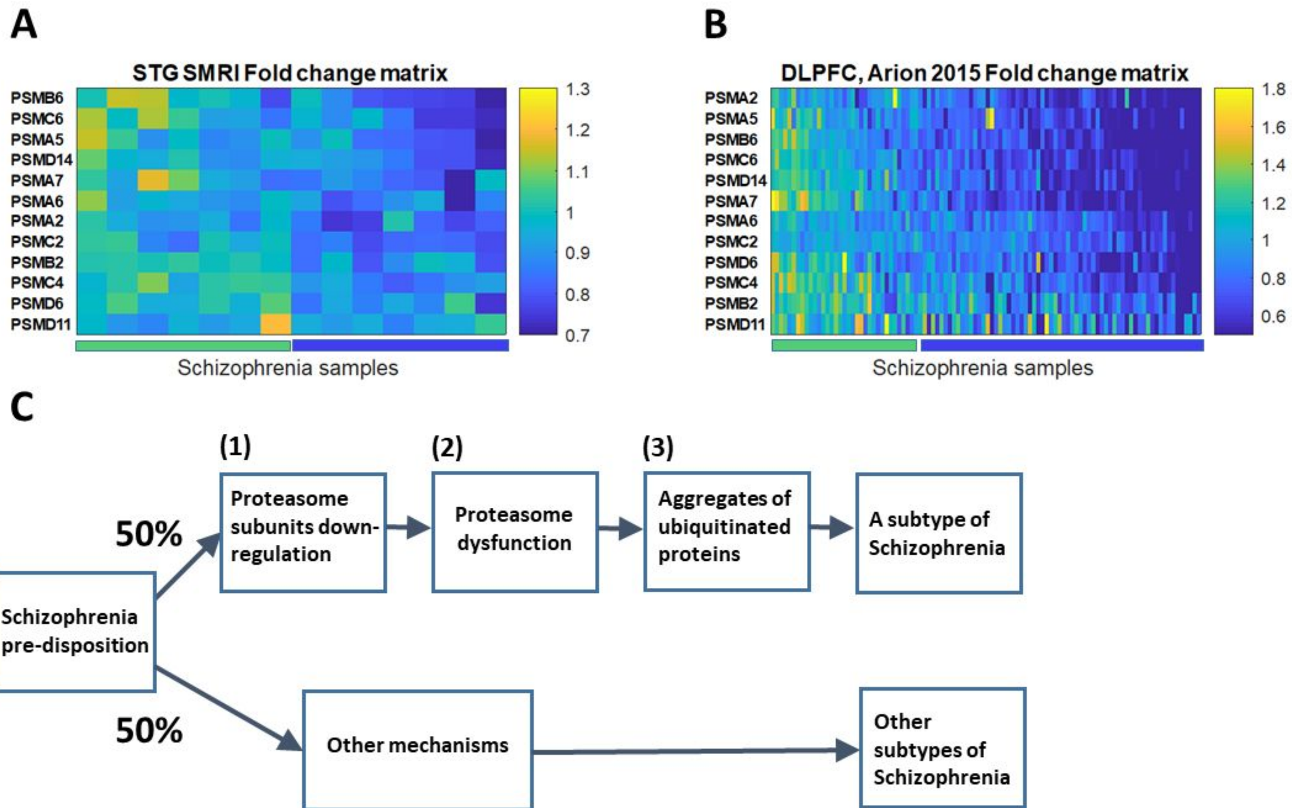


Figure 3. A) SMRI STG schizophrenia samples fold change matrix of proteasome subunits genes. Each row represents one of the 12 proteasome subunits genes that were found to be down-regulated in schizophrenia in the meta-analysis of the SMRI and MSSM datasets. Each column represents one of the SMRI schizophrenia samples. The color code represents the fold change, i.e. the expression value of the proteasome subunit gene in the specific sample, divided by its mean expression in the 15 control samples. Samples and genes locations were sorted by the SPIN tool (48). The left half of the samples, "Group 1", are marked by a green bar along the x-axis and the right half, "Group 2", with blue. **B) DLPFC Arion 2015 schizophrenia samples fold change matrix of proteasome subunits genes.** The same plot as in A) for the DLPFC Arion 2015 dataset. **C) A schematic preliminary model for a biological mechanism based division of schizophrenia patients into subtypes.**

SUPPLEMENTARY INFORMATION

Contents

Mount Sinai School of Medicine (MSSM) subjects...4

MSSM Superior Temporal Gyrus (STG) Gene expression preprocessing...4

MSSM STG Combining probe sets of the same gene...4

Comparison between SMRI STG multiple linear regression and t-test analysis results...5

Correlation analysis between SMRI differential genes' expression and subjects' information regarding medications, substance and alcohol use...5

Gene expression meta-analysis...5

Metabolism of Proteins pathway is enriched in MSSM up-regulated genes...6

Calculation of Polygenic Risk Scores (PRS) for the SMRI data...6

Additional datasets characteristics...7
- SMRI dataset...7
- STG, Barnes 2011 dataset...9
- Cerebellum, Chen 2018 dataset...9
- DLPFC, Arion 2015 dataset...10
- DLPFC, Ramaker 2017 dataset...10
- BA10, Mycox 2009 dataset...11

LIST OF FIGURES

Figure 1S: Comparison between multiple linear regression and t-test analysis resulting SMRI STG up-regulated and down-regulated genes. A) Venn diagram for the intersection between the 881 genes that were found to be up-regulated in schizophrenia STG SMRI samples using multiple linear regression analysis (with FDR Q < 15%) and 855 genes that were found to be up-regulated in schizophrenia STG SMRI samples using t-test analysis (FDR Q < 15%). B) Venn diagram for the intersection between the 986 genes that were found to be down-regulated in schizophrenia STG SMRI samples using multiple linear regression analysis (with FDR Q < 15%) and 944 genes that were found to be down-regulated in schizophrenia STG SMRI samples using t-test analysis (FDR Q < 15%). ...14

Figure 2S: SMRI STG 986 Down-regulated genes: Pearson Correlation Histogram between lifetime quantity of Fluphenazine or equivalent antipsychotic (in mg) and gene expression, along the 14 schizophrenia patients for which this information is available. The X-axis represents the Pearson correlation values, the mean correlation value measured for the 986 down-regulated genes is specified by a black vertical line. ...14

Figure 3S: SMRI STG 986 Down-regulated genes: Pearson Correlation Histogram between Substance use severity (measured 0-5) and gene expression, measured along schizophrenia and control subjects. The X-axis represents the Pearson correlation values, the mean correlation value measured for the 986 down-regulated genes is specified by a black vertical line. ...14

Figure 4S: SMRI STG 986 Down-regulated genes: Pearson Correlation Histogram between Substance use severity (measured 0-5) and gene expression, measured along schizophrenia and control subjects. The X-axis represents the Pearson correlation values, the mean correlation value measured for the 986 down-regulated genes is specified by a black vertical line. ...15

Figure 5S: MSSM STG Differential expression network view: A) Ubiquitin-Proteasome Dependent Proteolysis superPathway. The node's colors correspond to the deviation from the control samples group, in terms of standard deviation units (see Methods). The edges represent STRING database co-expression relations. Only genes that have co-expression relations with other genes in the network are displayed. A subgroup of highly-interconnected genes, coding for proteasome subunits, is circled. B) Zoom in on proteasome subunits. The same plot

as in A), for a subgroup of highly-interconnected genes coding for proteasome subunits (circled in A)). 15

Figure 6S: PSMA5 STG SMRI + MSSM meta-analysis of differential expression. Forest plot was generated using the function "forest" from the "meta" package in R, version 4.9-2 (General Package for Meta-Analysis) (11). The forest plot shows the differences in PSMA5 expression between subjects with schizophrenia and healthy controls, for each of the two studies, SMRI and MSSM. Each square represents the standardized difference (Hedges' g (10)) between schizophrenia and control for that study, with the area of the square reflecting the weight (determined by the sample size) given to that study in the meta-analysis. Each horizontal line represents the 95% confidence interval for the mean difference in that study. The vertical line shows the point of 0 difference. The standardized difference is positive (negative) if the expression is higher (lower) in schizophrenia vs. the control group. The center of the diamond represents the overall difference across both studies and its width represents 95% confidence interval. 16

Figure 7S: Fold change matrix of proteasome subunits genes. A) Cerebellum, Chen 2018 data (16). Each row represents one of the 12 proteasome subunits genes that were found to be down-regulated in schizophrenia in the meta-analysis of the SMRI and MSSM datasets. Each column represents one of the Chen 2018 44 schizophrenia samples. The color code represents the fold change, i.e. the expression value of the proteasome subunit gene in the specific sample, divided by its mean expression in the 50 control samples. B) DLPFC, Ramaker 2017 dataset (19). 24 samples of schizophrenia patients vs. 24 controls. C) DLPFC, Arion 2015 dataset (17). 102 samples of schizophrenia patients vs. 106 controls. D) BA10, Mycox 2009 dataset (20). 28 samples of schizophrenia patients vs. 23 controls. E) STG SMRI dataset. 14 samples of schizophrenia patients vs. 15 controls. F) BA23 SMRI dataset. 13 samples of schizophrenia patients vs. 15 controls. 17

Figure 8S. Scatter plots and Pearson correlation values between pH and mean fold change values of the 12 proteasome subunits down-regulated genes. A) Cerebellum, Chen 2018 data (16). Each blue dot represents a schizophrenia sample (overall 44). The y-axis represents the pH levels and the x-axis represents the mean fold change of the 12 proteasome subunits genes that were found to be down-regulated in schizophrenia in the meta-analysis of the SMRI and MSSM datasets. The Pearson correlation coefficient and p-value are written in the title. B) DLPFC, Ramaker 2017 dataset (19). 24 samples of schizophrenia. C) BA10, Mycox 2009 dataset (20). 28 samples of schizophrenia. D) STG, Barnes 2011 dataset (15). 23 samples of schizophrenia. E) STG MSSM. 19 samples of schizophrenia. F) STG SMRI. 14 samples of 18

Figure 9S. Scatter plot and linear regression between the t-statistics of two differential expression analyses of the 12 proteasome subunits genes that were found to be down-regulated in the STG SMRI and MSSM meta-analysis; step-wise linear regression and two-sided t-test analyses. Left plot: STG SMRI data. Each blue dot represents one of the 12 down-regulated proteasome subunits genes. The x-axis represents two-sided t-test t-statistic of gene expression comparison between schizophrenia and control samples. The y-axis represents the t-statistic of step-wise linear regression analysis, where age, PMI, gender, pH and RIN where included as covariates. Then diagnosis coefficient was then statistically tested for being nonzero, implying an effect for schizophrenia on the expression, beyond any other effect of the covariates. This produced a t-statistic. The dashed red line represents the linear regression line between the t-statistics of the two analyses. The dashed black line represents y = x. Pearson correlation = 0.94; p-value = 4×10^{-6} . Right plot: the same for the STG MSSM data. age, PMI, gender and pH where included as covariates . Pearson correlation = 0.95; p-value = 2.1×10^{-6}........19

Figure 10S. Scatter plot of the mean fold change (FC) of the 12 proteasome subunits genes that were found to be down-regulated in the STG SMRI and MSSM meta-analysis and PRS for the STG SMRI data. Each blue dot represents one of 11 STG SMRI patients with schizophrenia, for which genotyping data was available. The y-axis represents its PRS and the x-axis represents the mean fold change of the 12 down-regulated proteasome subunits genes. Pearson correlation = 0.12, p-value = 0.72 19

Figure 11S. Proteasome and Ribosome subunits differential expression network view: A) SMRI STG differential expression network view. The node's colors correspond to the deviation from the control samples group, in terms of standard deviation units (see Methods). The edges represent STRING database co-expression relations. Proteasome and Ribosome subunits genes are presented. Only genes that have co-expression relations with other genes are

displayed. B) MSSM STG differential expression network view. The same as in A) for the MSSM STG data. 20

LIST OF TABLES

Table 1S. SMRI STG and BA23 RNA-seq data....7

Table 2S: Additional datasets characteristics; Average values ± standard deviation....9

Table 3S: Pathway enrichment analysis of SMRI STG up-regulated genes with corrected p-value < 0.05. GeneAnalytics tool superpathways that were found to be enriched in the list of up-regulated genes are ordered by descending order of their enrichment score. The number of up-regulated genes that belong to each superpathway is given, with the number of genes of each superpathway in parentheses. 17

Table 4S. Pathway enrichment analysis of SMRI STG down-regulated genes with corrected p-value < 0.05. GeneAnalytics tool superpathways that were found to be enriched in the list of down-regulated genes are ordered by descending order of their enrichment score. The enrichment scores are in the second column and the superpathways' names are listed in the third column. The fourth column presents the number of down-regulated genes that belong to each superpathway , with the total number of genes of the superpathway in parentheses. MSSM enrichment score is given in the 5th column, where (-) sign means that the superpathway wasn't enriched in the list of MSSM down-regulated genes. For superpathways that are known to involve the UPS, a reference indicating the UPS involvement is given in the 6th column. Ubiquitin-proteasome directly related pathways are in bold....18

Table 5S: Pathway enrichment analysis of MSSM STG up-regulated genes with corrected p-value < 0.05. GeneAnalytics tool superpathways that were found to be enriched in the list of up-regulated genes are ordered by descending order of their enrichment score. The number of up-regulated genes that belong to each superpathway is given, with the number of genes of each superpathway in parentheses. 21

Table 6S: Pathway enrichment analysis of MSSM STG down-regulated genes with corrected p-value < 0.05. GeneAnalytics tool superpathways that were found to be enriched in the list of down-regulated genes are ordered by descending order of their enrichment score. The number of down-regulated genes that belong to each superpathway is given, with the number of genes of each superpathway in parentheses. UPS directly related pathways are in bold 22

Table 7S: STG down-regulated genes enriched UPS-related pathways; numbers of SMRI and MSSM hits. For each of the 5 UPS-related pathways that were enriched in the SMRI STG down-regulated genes, the number of 'hits' (down-regulated genes that belong to the pathway) shared with MSSM STG down-regulated genes is given in the second column. The numbers of hits specific to SMRI and MSSM down-regulated genes are given in columns 3 and 4, respectively.23

Table 8S: SMRI STG Differential expression analysis of proteasome subunits genes. 7 schizophrenia samples ("Group 2") were compared to the 15 control samples. T-sample t-test was applied; p-values were corrected by the Benjamini-Hochberg procedure (6)....23

Table 9S: SMRI STG comparison between two differential expression analyses of the 12 proteasome subunits genes that were found to be down-regulated in the meta-analysis. Step-wise linear regression analyses was applied, where age, PMI, gender, pH and RIN where included as covariates. Then diagnosis coefficient was then statistically tested for being nonzero, implying an effect for schizophrenia on the expression, beyond any other effect of the covariates. This produced t-statistic and corresponding p-values are listed in columns 3 and 4. The t-statistic and corresponding p-values of standard two-sided t-test analysis are listed in columns 5 and 6. P-values lower than 0.05 are marked with bold....25

Table 10S. MSSM STG comparison between two differential expression analyses of the 12 proteasome subunits genes that were found to be down-regulated in the meta-analysis. Step-wise linear regression analyses was applied, where age, PMI, gender and pH where included as covariates. Then diagnosis coefficient was then statistically tested for being nonzero, implying an effect for schizophrenia on the expression, beyond any other effect of the covariates. This produced t-statistic and corresponding p-values are listed in columns 3 and 4. The t-statistic and corresponding p-values of standard two-sided t-test analysis are listed in columns 5 and 6. P-values lower than 0.05 are marked with bold....25

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Mount Sinai School of Medicine (MSSM) subjects

Human brain samples of 19 schizophrenia and 14 healthy controls of the superior temporal gyrus (STG) were obtained from the Brain Bank of the Department of Psychiatry of the MSSM (Table 1). All cortical dissections and sample preparation were described previously (1,3,5). Brain banking activities were approved by the MSSM Institutional Review Board and written consent for brain donation was obtained from the next-of-kin of all subjects. Cases diagnosed as schizophrenia met the DSM-III/IV criteria, as determined by clinical investigators. None of the samples, of neither subjects with schizophrenia nor controls, showed evidence of any significant neuropathology (6). Whole-genome gene expression was measured using Affymetrix HG-U133A microarrays.

MSSM Superior Temporal Gyrus (STG) Gene expression preprocessing

The HG-U133A Affymetrix chips were pre-processed with the commonly used Affymetrix MicroArray Suite v. 5.0 (MAS-5) algorithm. MAS-5 was criticized for its high false positive rate claimed to stem from making use of mismatches, as opposed to robust multi-array average. However, it was shown that combined with detection calls, MAS-5 is both selective and sensitive (7). Lowess correction was then calculated. As we observed a random-like distribution for probe-sets with low expression levels, we set all the expression levels below 20 to be 20. Data were subjected to log2-transformation. Filtering: Probe-sets that are present in at least 40% of the samples of at least one of the *17* regions of a certain disease type (schizophrenia or control) are kept for the rest of the analysis. Probe-sets without assigned Affymetrix gene symbols annotation were removed. 12,033 probe-sets were left for the rest of the analysis after filtering (out of 22,283), representing 8,542 gene symbols.

MSSM STG Combining probe sets of the same gene

Genes represented by more than one probe-set with the same Affymetrix assigned gene symbol were considered to represent the same gene and the expression was determined as follows: The Pearson correlation coefficient was calculated for each pair in such a group of probe-sets; and the largest subgroup in which each pair of probe-sets had a correlation coefficient higher than 0.5 was found by simple scanning. If the size of the chosen subgroup was larger than 2, the probe-set with the maximal average correlation values (in respect to the rest of the probe-sets in the subgroup) was chosen to represent the gene. Otherwise, in case the size of the chosen subgroup equals 2, one of them is chosen by random.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Comparison between SMRI STG multiple linear regression and t-test analysis results

In addition to multiple linear regression analysis, differentially expressed genes were identified by applying 2-sided t-test for each gene, comparing its expression between the schizophrenia samples and the controls. P-values were then adjusted for multiple hypothesis testing using false discovery rate (FDR) estimation (8), and the differentially expressed genes were determined as those with an estimated FDR ≤ 15%.

The comparison between the resulting lists of up-regulated and down-regulated genes, using multiple linear regression and t-test, is plotted in Figure 1S. It can be seen that the intersection between the lists is very large (calculated hyper-geometric p-value < $1*10^{-50}$, for both up-regulated and down-regulated genes). As a result, the results of pathway enrichment analyses, using the lists obtained from the t-test analysis, were very similar to those using the lists obtained from the multiple linear regression analysis.

Correlation analysis between SMRI differential genes' expression and subjects' information regarding medications, substance and alcohol use

Correlation analyses between the expression of the SMRI 881 up-regulated genes and 986 down-regulated genes and Fluphenazine equivalent dosage, severity of substance use and severity of alcohol use was performed. The Pearson correlation histogram for the SMRI 986 down-regulated genes is plotted in Figures 2S-4S for these parameters. In addition, p-values for each correlation value was calculated. For both the down-regulated genes and up-regulated genes, and for each of the 3 parameters (Fluphenazine equivalent, substance use and alcohol use), FDR(8) was applied and corrected p-values were calculated. While for the 881 up-regulated genes no gene passed FDR of 5% in each of the 3 parameters, for the 986 down-regulated genes 2 genes passed FDR of 5% for substance use, ADSL and C9orf85.

Gene expression meta-analysis

For a given gene, a meta-analysis that integrates its expression in both SMRI and MSSM was applied. To address the differences in study design and platform usage, we applied the Effect size (ES), the standardized difference between the expression in the disease vs. control samples, combined with Random Effect Modeling, which takes both the direction and magnitude of gene expression changes into consideration to generate more biologically consistent results. As was demonstrated in (9), it is superior to other meta-analytic methods in that it has the ability to handle the variability between studies, and highly applicable for gene expression data. ES (Hedges' g (10)) was calculated separately for the SMRI and MSSM datasets. The direction of the effect size was positive if the expression in the disease group was

higher than in the control group. Hedges'g and confidence interval values were calculated for each of the SMRI and MSSM datasets, using the function "metacont" from the "meta" package in R, a general package for meta-analysis, version 4.9-2 (11). The summary measure of the two datasets with its confidence interval was calculated using the same function, using the random effects model (12).

As an example, results of the meta-analysis of proteasome subunit α5, PSMA5, is plotted in Figure 6S. It can be seen that while in each of the SMRI and MSSM separately we can observe only a trend towards down-regulation (95% confidence interval horizontal lines cross the zero), statistical significance was achieved only when the two datasets were integrated.

Metabolism of Proteins pathway is enriched in MSSM up-regulated genes

Interestingly, one of the pathways that were found to be enriched in the MSSM up-regulated genes is Metabolism of Proteins (Table 5S), composed of 1,628 genes, of which 77 belong also to the Ubiquitin-Proteasome Dependent Proteolysis pathway, that was found to be enriched in the down-regulated genes. Thus, a subgroup of genes from this pathway is down-regulated while another is up-regulated. Out of the Metabolism of Proteins pathway we identified a list of tightly interconnected genes that encode for ribosomal subunits, which are mostly up-regulated in schizophrenia (see Figure S11). Metabolism of Proteins pathway was not found to be enriched in the SMRI up-regulated genes (Table 3S). However, the subgroup of ribosomal subunits is contained in another pathway that was found to be enriched in both SMRI and MSSM up-regulated genes, Influenza Viral RNA Transcription and Replication. It can be seen (Figure S11) that ribosomal subunits are mostly up-regulated in schizophrenia in both SMRI and MSSM. In addition, it can be seen that while both the ribosomal and proteasomal subunits are tightly interconnected, the two groups of genes are also highly connected.

Calculation of Polygenic Risk Scores (PRS) for the SMRI data

Genotyping data measured on Affymetrix SNP 5.0 microarray, that contains 492,828 Single Nucleotide Polymorphism (SNP) probes, was available for 11 out of 14 SMRI patients, from the SMRI website (https://www.stanleygenomics.org/). The following Filtering was applied to the genotyping data:

1. Remove SNPs with > 5% missing calls. After this step, 433,647 SNP probes were left.
2. Remove SNPs with Minor Allele Frequency (MAF) < 5%. After this step, 331,978 SNP probes were left.
3. For SNPs with multiple Affymetrix probes, probes with the least number of missing calls were chosen. After this step, 329,773 SNP probes were left.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Posterior SNP effect size (ES) values were inferred for all autosomal SNPs using PRS-CS (2), a Python based command line tool that acts under continuous shrinkage (CS) priors, using GWAS summary statistics (4) and an external Linkage Disequilibrium (LD) reference panel.

Data used as input to PRS-CS:

1. The validation set contains the aforementioned 329,773 SNP probes.
2. The EUR LD reference computed using the 1,000 Genomes samples was downloaded from https://github.com/getian107/PRScs (see "Getting Started" section).
3. Schizophrenia GWAS summary statistics (4) were downloaded from the Psychiatric Genomic Consortium website (https://www.med.unc.edu/pgc/download-results/).

The output of this PRS-CS analysis includes the posterior effect size estimates for 243,046 SNP probes out of the validation set, which are those that exist in both the GWAS summary statistics (4) and the LD reference panel.

Then for each patient j = 1,…, 11, PRSj was calculated as follows, and as described in https://choishingwan.github.io/PRS-Tutorial/plink/:

$$PRS_j = \frac{\sum_{i=1}^{N} ES_i \times G_{ij}}{P \times M_j}$$

Where ES_i is the posterior effect size of SNP i; the number of effect alleles observed in sample j is G_{ij}; the ploidy of the sample is P (is generally 2 for humans); the total number of SNPs included in the PRS is N (here N = 243,046); and the number of non-missing SNPs observed in sample j is M_j. If the sample has a missing genotype for SNP i, then the population minor allele frequency multiplied by the ploidy (MAF_i x P) is used instead of G_{ij}.

Additional datasets characteristics

SMRI dataset

Brodmann Area 23 (BA23) SMRI samples: BA23 postmortem tissues from 13 subjects with schizophrenia and 15 healthy controls were obtained from the SMRI using approved protocols for tissue collection and informed consent (13). All samples were examined by a certified neuropathologist to exclude Alzheimer's disease and other cerebral pathology (13). Diagnoses were performed independently by two psychiatrists according to DSM-IV criteria. See samples' characteristics in Table 2S.

BA23 SMRI RNA extraction and quality control: The brain regions were dissected and total RNA was isolated using the Trizol method by the staff at SMRI. The

concentration of total RNA and RNA Integrity Number value (RIN) were measured. Total RNA samples were delivered on dry ice to The Nancy & Stephen Grand Israel National Center for Personalized Medicine (G-INCPM) for whole transcriptome sequencing. Samples with RIN ⩾5 were selected for sequencing (all 28 samples). Among these samples, the mean RIN was 8.2 (± 0.5).

STG and BA23 SMRI RNA sequencing libraries preparation: Libraries preparation were done using the INCPM-RNA-seq protocol. Briefly, polyA fraction (mRNA) was purified from 500ng of total RNA by oligo(dT) beads following by fragmentation and generation of double stranded cDNA using random hexamers. Then, end repair, A base addition, adapter ligation and PCR amplification steps were performed. Libraries were evaluated by Qubit and TapeStation and pooled in an equimolar ratio. Sequencing libraries were constructed with barcodes to allow multiplexing of samples in a lane. For raw RNA sequencing data description see Table 1S.

STG and BA23 SMRI mapping and quantification of gene expression: Fragment mapped to the genome (hg19) using TopHat version V2.0.5 and Bowtie version 2.2.0. Only fragments with good quality reads (mean Qphred per read>20; corresponding to above 99% probability of a correctly identified base), with both ends uniquely mapped to the genome, were considered (~70% of all fragments). The signals from the 6 lanes were summed. Known Ensembl gene levels were quantified by HTSeq version 0.6.0 in intersection-strict mode. This provides an integral count of reads for each gene in each sample (a sample-by-gene 'read count matrix'). Gene models were downloaded from the UCSC Genome Browser (http://genome.ucsc.edu/; assembly hg19).

BA23 SMRI preprocessing: Lowess correction was calculated (14). Then expression threshold was set to 6 (log scale) to reduce noise. Filtering: Genes with expression values below 6 in at least 80% of the samples (considering both STG and BA23 SMRI samples) were filtered out of the analysis. 16,482 genes were left for the rest of the analysis after filtering (out of 23,715).

Table 1S. SMRI STG and BA23 RNA-seq data

Group	N	Mean total reads/subject	Mapped reads (%)	No. of genes sequenced
STG, SMRI dataset				
Schizophrenia	14	23,559,907	91.2	15,554
Control	15	22,976,936	92.1	15,635
BA23, SMRI dataset				
Schizophrenia	13	29,004,007	92.3	15,818

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Control	15	41,158,436	92.4	15,804

STG, Barnes 2011 dataset

The dataset GSE21935 (15) was downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21935). The dataset consists of 42 superior temporal cortex samples from subjects with schizophrenia (n=23) and healthy controls (n=19). Samples were run on Affymetrix Human Genome U133 Plus 2.0 Arrays. See Table 2S for samples' characteristics. **Normalization method:** Arrays were scanned on a GeneChip Scanner 3000, and fluorescence intensity was obtained by using GeneChip Operating Software (15). As described in GSE21935_series_matrix.txt (available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21935), the data were analyzed with Microarray Suite version 5.0 (MAS 5.0) using Affymetrix default analysis settings and global scaling as normalization method. The trimmed mean target intensity of each array was arbitrarily set to 150. We then applied threshold and logarithm base 2 (log2). The threshold value was determined using scatter plots of healthy control samples, in order to estimate the noise level (the threshold after log2 that was used is 4). **Filtering:** 1) Initial number of probe-sets was 54,675 (45,772 with assigned gene symbols). 2) In case of gene symbol with multiple probe-sets, the probe-set with the highest mean expression over the samples was taken into account and the other probe-sets were discarded. Number of genes after this step: 22,880. 3) Genes that are absent (values equal or lower than the threshold) in more than 70% of both the schizophrenia and the control samples, were filtered out. Number of genes after filtering: 17,464.

Cerebellum, Chen 2018 dataset

The dataset GSE35978 (16) was downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE35978). The initial dataset consisted of 312 brain samples. We used only the cerebellum samples of subjects with schizophrenia (n=44) and healthy controls (n=50). See Table 2S for samples' characteristics. Samples were run on Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]. **Normalization method:** As described in GSE35978_series_matrix.txt (available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE35978), the data were analyzed by Robust Multi-array Average (RMA) (17) using Affymetrix Expression

Console with default analysis settings. **Filtering:** 1) Initial number of probe-sets was 33,297 (25,293 with assigned gene symbols). 2) In case of gene symbol with multiple probe-sets, the probe-set with the highest mean expression over the samples was taken into account and the other probe-sets were discarded. Number of genes after this step: 23,307. No threshold was applied, as the RMA algorithm in includes background correction and quantile normalization (18). Number of genes after filtering: 23,307 (no probes were removed).

DLPFC, Arion 2015 dataset

The dataset GSE93987 (17) was downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE93987). The dataset consists of 208 dorsolateral prefrontal cortex (DLPFC) brain samples from subjects with schizophrenia (n=102) and healthy controls (n=106). See Table 2S for samples' characteristics. Samples were run on Affymetrix HT HG-U133 Arrays + PM Array Plate. **Normalization method:** Affymetrix CEL files were normalized and log2 transformed using RMA (17). **Filtering:** 1) Initial number of probe-sets: 54,613 (44,228 with assigned gene symbols). 2) In case of a gene symbol with multiple probe-sets, the probe-set with the highest mean expression over the samples was taken into account and the other probe-sets were discarded. Number of genes after this step: 21,597. No threshold was applied, as the RMA algorithm in includes background correction and quantile normalization (18). Number of genes after filtering: 21, 597 (no probes were removed).

DLPFC, Ramaker 2017 dataset

The dataset GSE80655 (19) was downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE80655). The initial dataset consisted of 281 brain samples. We used only the DLPFC of subjects with schizophrenia (n=24) and healthy controls (n=24). See Table 2S for samples' characteristics. Samples were run on Illumina HiSeq 2000 Arrays. **Normalization method:** To quantify the expression of each gene, RNA-seq reads were processed with aRNApipe v1.1 using default settings (19). We then applied threshold and log2. The threshold value was determined using scatter plots of control samples in order to estimate the noise level (the threshold after log2 that was used is 3). **Filtering:** 1) Initial number of probe-sets was 57,905 (21,287 with assigned gene symbols). 2) In case of a gene symbol with multiple probe-sets, the probe with the highest mean expression over the samples was taken into account and the other probe-sets were discarded. Number of genes after this step: 20,881. 3) Genes that are absent (values equal or lower than the threshold) in more than 70% of both the schizophrenia and the control samples, were filtered out. Number of genes after filtering: 17,037.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

BA10, Mycox 2009 dataset

The dataset GDS4523 (20) was downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS4523). The dataset consists of 51 BA10 brain samples from subjects with schizophrenia (n=28) and healthy controls (n=23). See Table 2S for samples' characteristics. Samples were run on Affymetrix Human Genome U133 Plus 2.0 Arrays. **Normalization method:** Arrays were scanned on a GeneChip Scanner 3000 and fluorescence intensity for each feature of the array was obtained by using GeneChip Operating Software (Affymetrix) (20). As described in GSE17612_series_matrix.txt (available at https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS4523), the data were analyzed with Microarray Suite version 5.0 (MAS 5.0) using Affymetrix default analysis settings and global scaling as normalization method. The trimmed mean target intensity of each array was arbitrarily set to 150. We then applied threshold and log2. The threshold value was determined using scatter plots of healthy control samples in order to estimate the noise level (the threshold after log2 that was used is 4). **Filtering:** 1) The initial number of probe-sets was 54,613. 2) In case of a gene symbol with multiple probe-sets, the probe-set with the highest mean expression over the samples was taken into account and the other probe-sets were discarded. Number of genes after this step: 30,805. 3) Genes that are absent (values equal or lower than the threshold) in more than 70% of both the schizophrenia and the control samples, were filtered out. Number of genes after filtering: 27,295.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 2S: Additional datasets characteristics; Average values ± standard deviation

Characteristics	Control	Schizophrenia	P-value
BA23, SMRI dataset			
Number of subjects	15	13	
Age (years)	48.06 ±10.6	43.2 ±13.5	0.29
Gender	9M : 6F	9M : 4F	0.71
Brain pH	6.26 ±0.24	6.17 ± 0.26	0.36
PMI	23.7 ±9.9	33.9 ± 15.6	0.05
STG, Barnes 2011 dataset (15), **GSE21935**			
Number of subjects	19	23	
Age (years)	67.68 ± 22	72.17 ± 17	0.46
Gender	10M : 9F	13M : 10F	1
Brain pH	6.489 ± 0.32	6.161 ± 0.17	0.00013
PMI	9.105 ± 4.3	7.13 ± 5.7	0.22
Cerebellum, Chen 2018 dataset (16), **GSE35978**			
Number of subjects	50	44	
Age (years)	45.8 ± 9.3	43.18 ± 9.5	0.18
Gender	31M : 19F	32M : 12F	0.28
Brain pH	6.474 ± 0.32	6.428 ± 0.25	0.44
PMI	27.58 ± 11	33.27 ± 15	0.042
DLPFC, Arion 2015 dataset (17), **GSE93987**			
Number of subjects	106	102	
Age (years)	Not provided	Not provided	
Gender	Not provided	Not provided	
Brain pH	Not provided	Not provided	
PMI	Not provided	Not provided	
DLPFC, Ramaker 2017 dataset (19), **GSE80655**			
Number of subjects	24	24	
Age (years)	50.25 ± 13	42.67 ± 10	0.025
Gender	21M : 3F	21M : 3F	1

Brain pH	6.921 ± 0.11	6.83 ± 0.18	0.043
PMI	21.89 ± 6.6	21.29 ± 9.2	0.8
BA10, Mycox 2009 dataset (20), **GDS4523**			
Number of subjects	23	28	
Age (years)	69.04 ±22	73.32 ±15	0.41
Gender	12M : 11F	19M : 9F	0.39
Brain pH	6.21 ±0.24	6.23 ±0.25	0.82
PMI	9.902 ± 4.4	8.714 ± 7	0.48
Overall number of subjects	237	234	

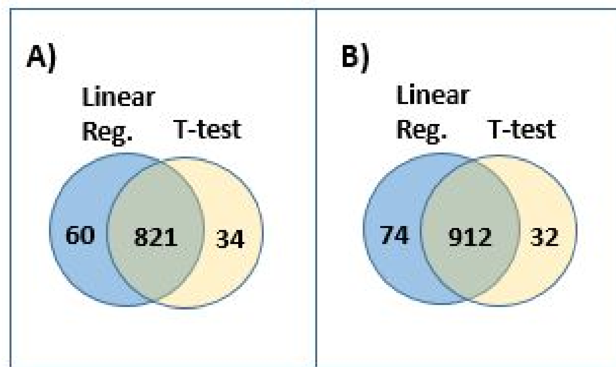


Figure 1S: Comparison between multiple linear regression and t-test analysis resulting SMRI STG up- regulated and down-regulated genes. A) Venn diagram for the intersection between the 881 genes that were found to be up-regulated in schizophrenia STG SMRI samples using multiple linear regression analysis (with FDR Q < 15%) and 855 genes that were found to be up-regulated in schizophrenia STG SMRI samples using t-test analysis (FDR Q < 15%). **B)** Venn diagram for the intersection between the 986 genes that were found to be down-regulated in schizophrenia STG SMRI samples using multiple linear regression analysis (with FDR Q < 15%) and 944 genes that were found to be down-regulated in schizophrenia STG SMRI samples using t-test analysis (FDR Q < 15%).

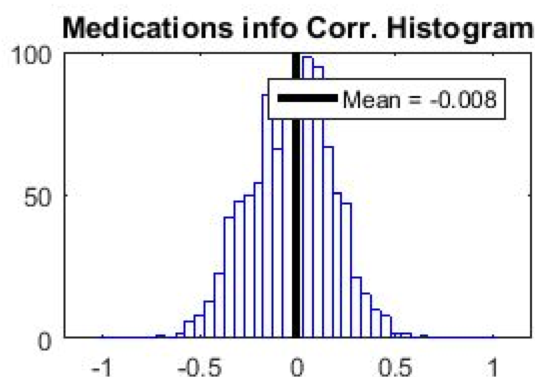


Figure 2S: SMRI STG 986 Down-regulated genes: Pearson Correlation Histogram between lifetime quantity of Fluphenazine or equivalent antipsychotic (in mg) and gene expression, along the 14 schizophrenia patients for which this information is available. The X-axis represents the Pearson correlation values, the mean correlation value measured for the 986 down-regulated genes is specified by a black vertical line.

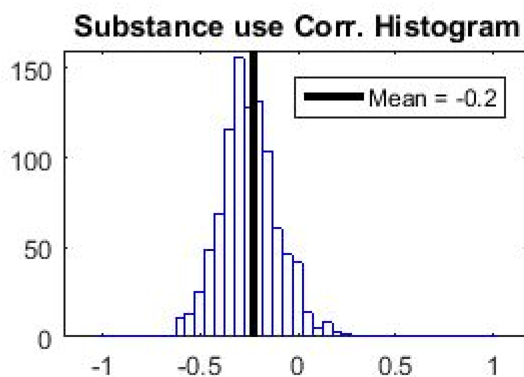
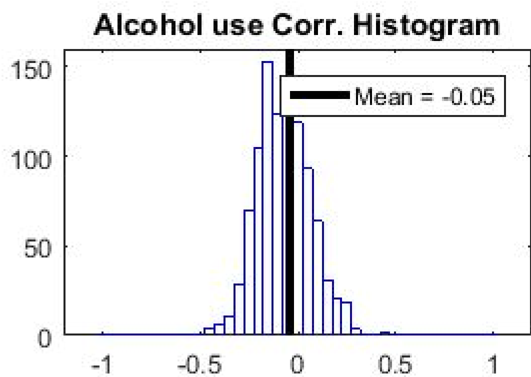


Figure 3S: SMRI STG 986 Down-regulated genes: Pearson Correlation Histogram between Substance use severity (measured 0-5) and gene expression, measured along schizophrenia and control subjects. The X-axis represents the Pearson correlation values, the mean correlation value measured for the 986 down-regulated genes is specified by a black vertical line.

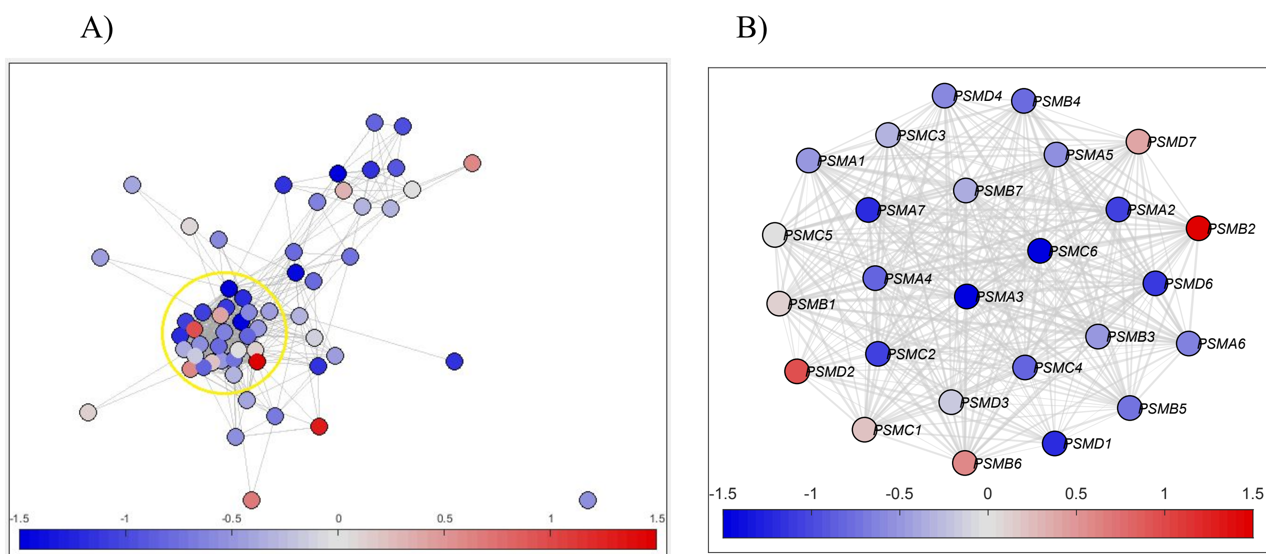
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16



17 18 19 20

Figure 4S: SMRI STG 986 Down-regulated genes: Pearson Correlation Histogram between Substance use severity (measured 0-5) and gene expression, measured along schizophrenia and control subjects. The X-axis represents the Pearson correlation values, the mean correlation value measured for the 986 down-regulated genes is specified by a black vertical line.

21
22
23
24
25
26



27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44

Figure 5S: MSSM STG Differential expression network view: A) Ubiquitin-Proteasome Dependent Proteolysis superPathway. The node's colors correspond to the deviation from the control samples group, in terms of standard deviation units (see Methods). The edges represent STRING database co-expression relations. Only genes that have co-expression relations with other genes in the network are displayed. A subgroup of highly-interconnected genes, coding for proteasome subunits, is circled. **B) Zoom in on proteasome subunits.** The same plot as in A), for a subgroup of highly-interconnected genes coding for proteasome subunits (circled in A)).

45 46 47 48 49 50 51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

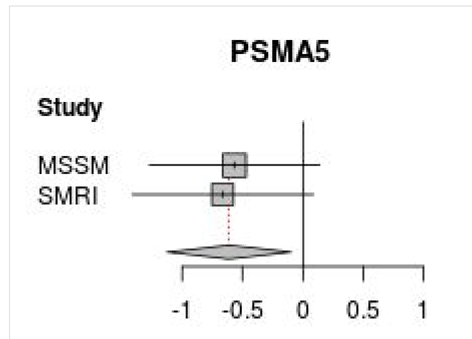
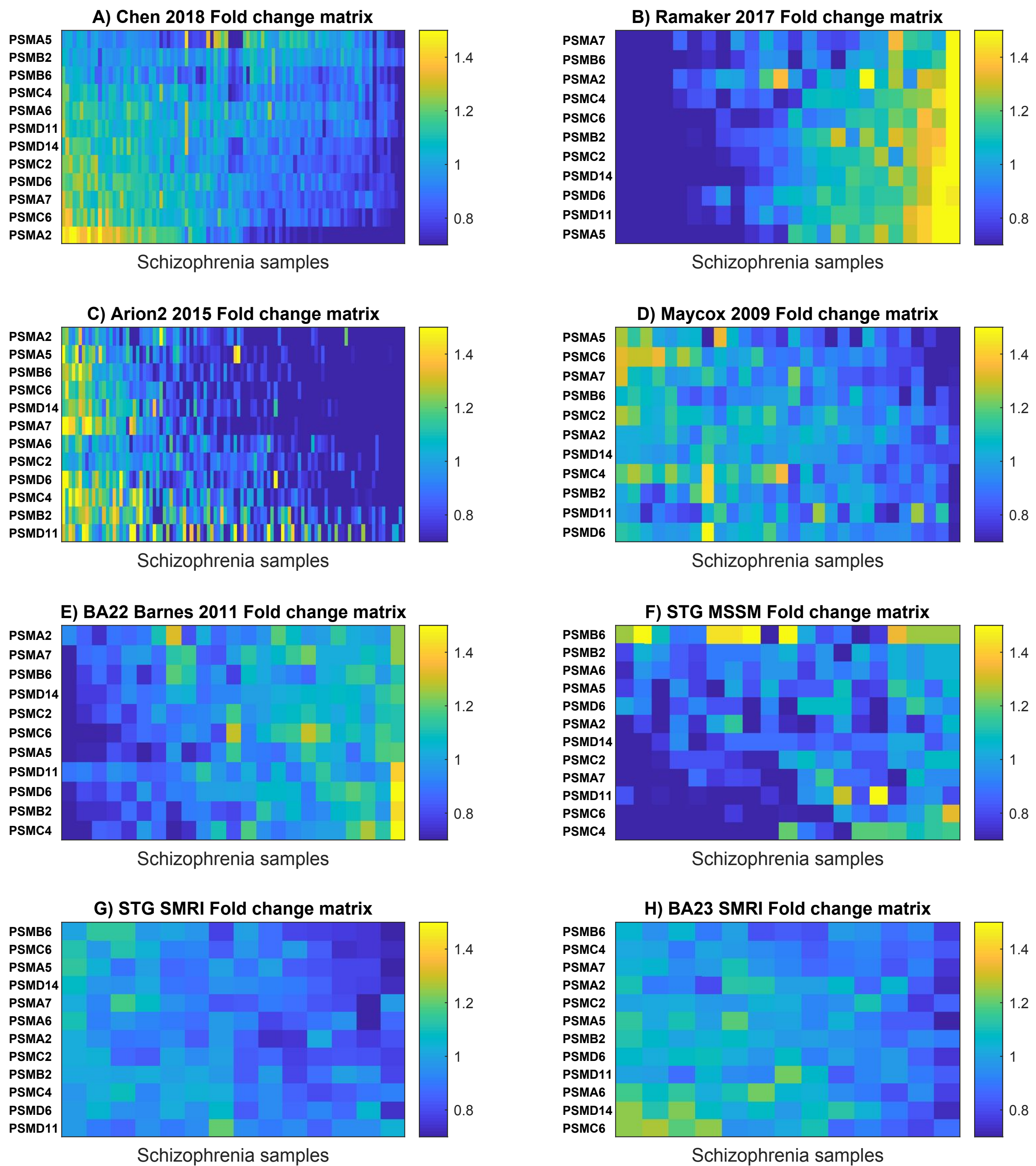


Figure 6S: PSMA5 STG SMRI + MSSM meta-analysis of differential expression. Forest plot was generated using the function “forest” from the “meta” package in R, version 4.9-2 (General Package for Meta-Analysis) (11). The forest plot shows the differences in PSMA5 expression between subjects with schizophrenia and healthy controls, for each of the two studies, SMRI and MSSM. Each square represents the standardized difference (Hedges’ g (10)) between schizophrenia and control for that study, with the area of the square reflecting the weight (determined by the sample size) given to that study in the meta-analysis. Each horizontal line represents the 95% confidence interval for the mean difference in that study. The vertical line shows the point of 0 difference. The standardized difference is positive (negative) if the expression is higher (lower) in schizophrenia vs. the control group. The center of the diamond represents the overall difference across both studies and its width represents 95% confidence interval.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49



50
51
52
53
54
55
56
57
58
59
60

Figure 7S: Fold change matrix of proteasome subunits genes. A) Cerebellum, Chen 2018 data (16). Each row represents one of the 12 proteasome subunits genes that were found to be down-regulated in schizophrenia in the meta-analysis of the SMRI and MSSM datasets. Each column represents one of the Chen 2018 44 schizophrenia samples. The color code represents the fold change, i.e. the expression value of the proteasome subunit gene in the specific sample, divided by its mean expression in the 50 control samples. B) DLPFC, Ramaker 2017 dataset (19). 24 samples of schizophrenia patients vs. 24 controls. C) DLPFC, Arion 2015 dataset (17). 102 samples of schizophrenia patients vs. 106 controls. D) BA10, Mycox 2009 dataset (20). 28 samples of schizophrenia patients vs. 23 controls. E) STG SMRI dataset. 14 samples of schizophrenia patients vs. 15 controls. F) BA23 SMRI dataset. 13 samples of schizophrenia patients vs. 15 controls.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

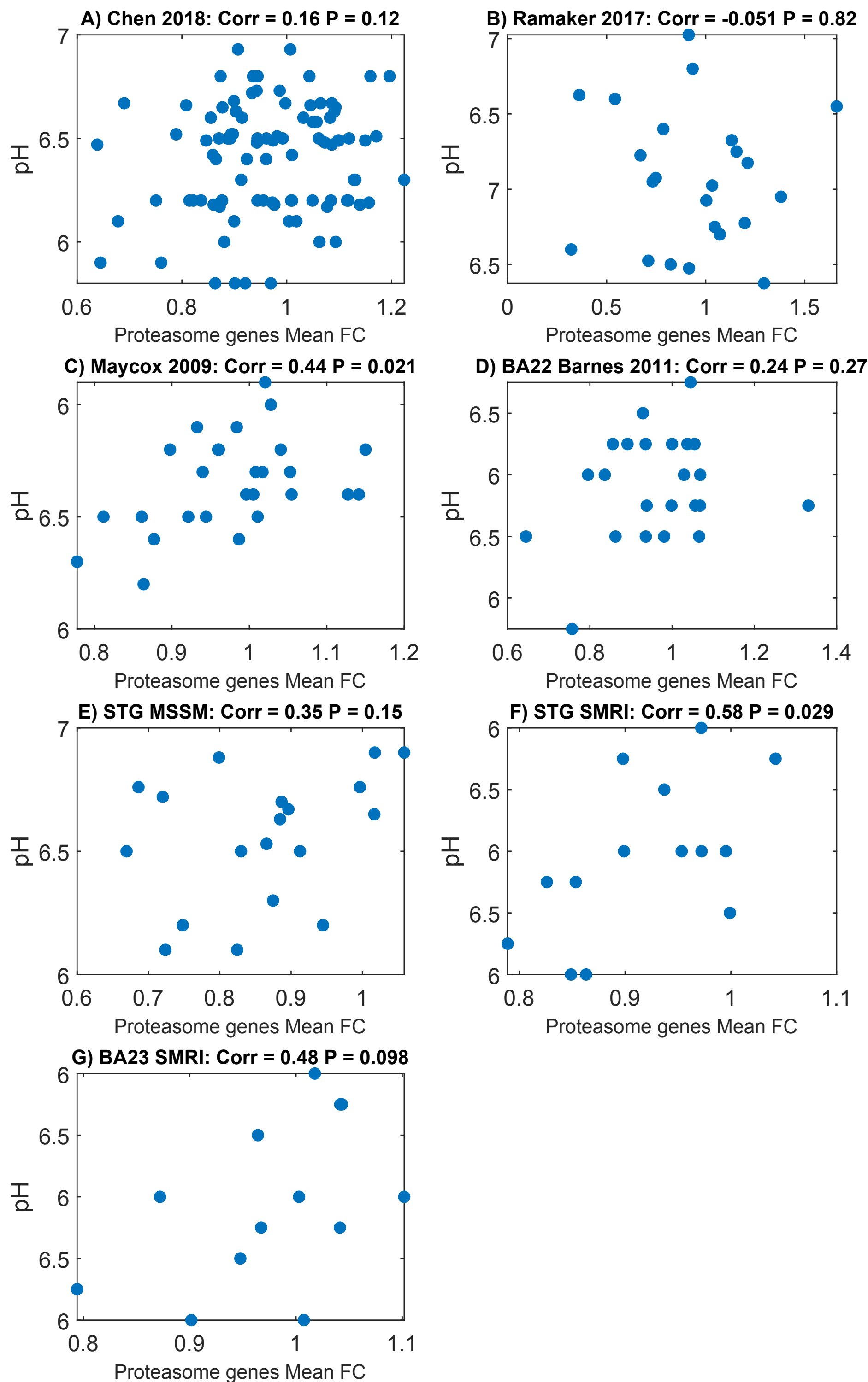


Figure 8S. **Scatter plots and Pearson correlation values between pH and mean fold change values of the 12 proteasome subunits down-regulated genes. A)** Cerebellum, Chen 2018 data (16). Each blue dot represents a schizophrenia sample (overall 44). The y-axis represents the pH levels and the x-axis represents the mean fold change of the 12 proteasome subunits genes that were found to be down-regulated in schizophrenia in the meta-analysis of the SMRI and MSSM datasets. The Pearson correlation coefficient and p-value are written in the title. **B)** DLPFC, Ramaker 2017 dataset (19). 24 samples of schizophrenia. **C)** BA10, Mycox 2009 dataset (20). 28 samples of schizophrenia. **D)** STG, Barnes 2011 dataset (15). 23 samples of schizophrenia. **E)** STG MSSM. 19 samples of schizophrenia. **F)** STG SMRI. 14 samples of

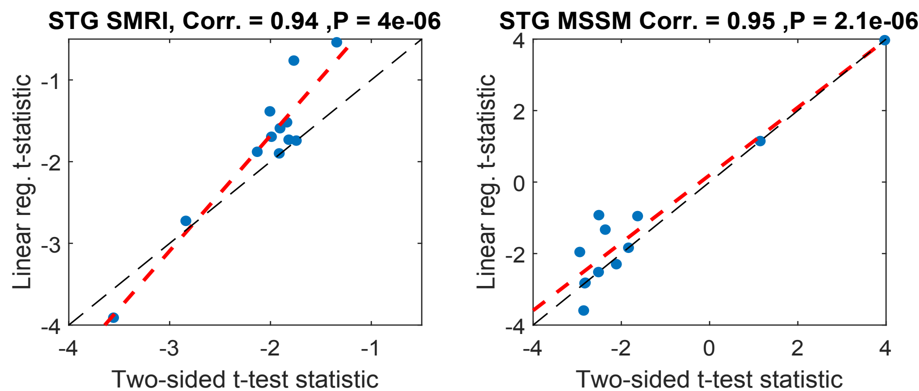


Figure 9S. Scatter plot and linear regression between the t-statistics of two differential expression analyses of the 12 proteasome subunits genes that were found to be down-regulated in the STG SMRI and MSSM meta-analysis; step-wise linear regression and two-sided t-test analyses. Left plot: STG SMRI data. Each blue dot represents one of the 12 down-regulated proteasome subunits genes. The x-axis represents two-sided t-test t-statistic of gene expression comparison between schizophrenia and control samples. The y-axis represents the t-statistic of step-wise linear regression analysis, where age, PMI, gender, pH and RIN where included as covariates. Then diagnosis coefficient was then statistically tested for being nonzero, implying an effect for schizophrenia on the expression, beyond any other effect of the covariates. This produced a t-statistic. The dashed red line represents the linear regression line between the t-statistics of the two analyses. The dashed black line represents y = x. Pearson correlation = 0.94; p-value = 4×10^{-6} **. Right plot: the same for the STG MSSM data.** age, PMI, gender and pH where included as covariates . Pearson correlation = 0.95; p-value = 2.1×10^{-6}

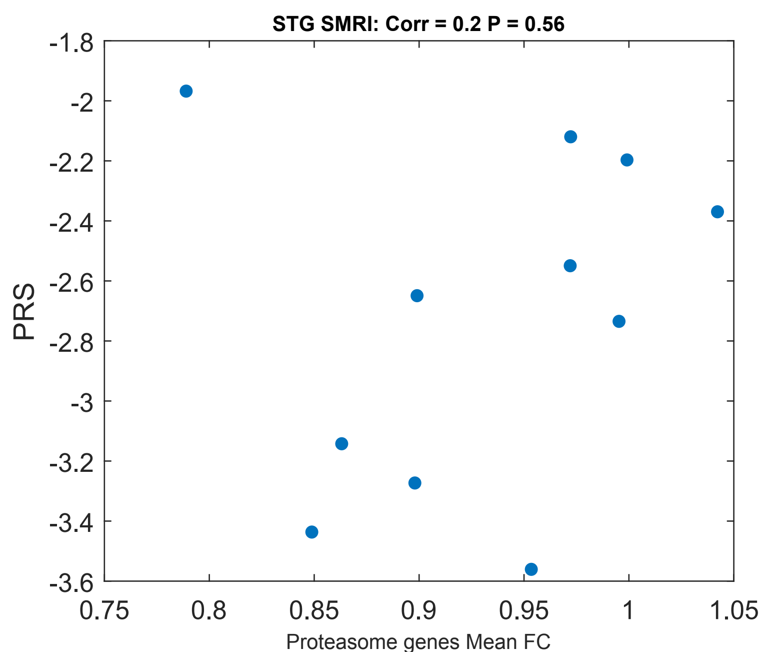


Figure 10S. Scatter plot of the mean fold change (FC) of the 12 proteasome subunits genes that were found to be down-regulated in the STG SMRI and MSSM meta-analysis and PRS for the STG SMRI data. Each blue dot represents one of 11 STG SMRI patients with schizophrenia, for which genotyping data was available. The y-axis represents its PRS and the x-axis represents the mean fold change of the 12 down-regulated proteasome subunits genes. Pearson correlation = 0.2, p-value = 0.56

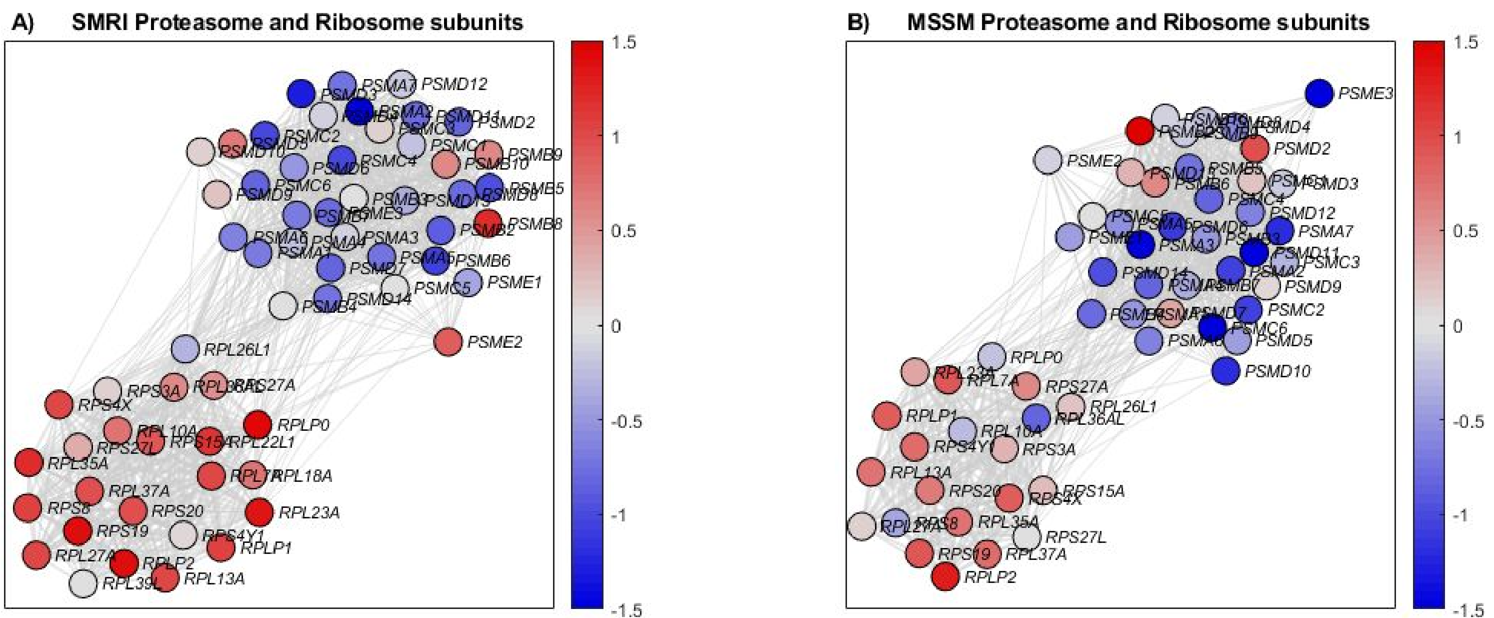


Figure 11S. Proteasome and Ribosome subunits differential expression network view: A) SMRI STG differential expression network view. The node's colors correspond to the deviation from the control samples group, in terms of standard deviation units (see Methods). The edges represent STRING database co-expression relations. Proteasome and Ribosome subunits genes are presented. Only genes that have co-expression relations with other genes are displayed. **B) MSSM STG differential expression network view**. The same as in A) for the MSSM STG data.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 3S: Pathway enrichment analysis of SMRI STG up-regulated genes with corrected p-value < 0.05. GeneAnalytics tool superpathways that were found to be enriched in the list of up-regulated genes are ordered by descending order of their enrichment score. The number of up-regulated genes that belong to each superpathway is given, with the number of genes of each superpathway in parentheses.

#	Score	SuperPath Name	Num Matched (SuperPath) genes	MSSM Enrichment Score
1	13.43	Metallothioneins Bind Metals	5 (11)	-
2	13.2	Axon Guidance	19 (175)	-
3	12.4	Protein-protein Interactions at Synapses	11 (72)	-
4	11.42	ERK Signaling	71 (1177)	18.14
5	10.38	Influenza Viral RNA Transcription and Replication	16 (158)	18.85
6	10.29	NFAT and Cardiac Hypertrophy	26 (326)	-
7	10.25	Hedgehog Pathway	10 (73)	-
8	10.2	4-hydroxytamoxifen, Dexamethasone, and Retinoic Acids Regulation of P27 Expression	5 (18)	-
9	10.14	P38 MAPK Signaling Pathway (sino)	9 (61)	-
10	9.92	RET Signaling	59 (974)	-
11	9.4	CREB Pathway	36 (528)	13.49
12	9.25	LKB1 Signaling Events	7 (42)	-
13	9.01	Phospholipase-C Pathway	34 (498)	10.15
14	8.8	Signaling By NOTCH1	12 (113)	-
15	8.76	Circadian Entrainment	21 (262)	-
16	8.74	MAPK Signaling Pathway	24 (316)	-
17	8.74	HIV Life Cycle	52 (865)	12.88
18	8.7	Regulation of Lipid Metabolism Insulin Signaling-generic Cascades	21 (263)	-
19	8.57	P38 MAPK Signaling Pathway (WikiPathways)	6 (34)	-
20	8.56	Focal Adhesion	22 (283)	11.03

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 4S. Pathway enrichment analysis of SMRI STG down-regulated genes with corrected p-value < 0.05. GeneAnalytics tool superpathways that were found to be enriched in the list of down-regulated genes are ordered by descending order of their enrichment score. The enrichment scores are in the second column and the superpathways' names are listed in the third column. The fourth column presents the number of down-regulated genes that belong to each superpathway , with the total number of genes of the superpathway in parentheses. MSSM enrichment score is given in the 5th column, where (-) sign means that the superpathway wasn't enriched in the list of MSSM down-regulated genes. For superpathways that are known to involve the UPS, a reference indicating the UPS involvement is given in the 6th column. Ubiquitin-proteasome directly related pathways are in bold

#	Score	SuperPath Name	Num Matched (SuperPath) genes	MSSM Enrichment Score	Evidence for UPS involvement
1	41.07	MRNA Splicing - Major Pathway	65 (307)	33.59	
2	28.4	Chks in Checkpoint Regulation	46 (224)	18.43	
3	27.72	Translational Control	41 (189)	24.61	
4	26.05	Vesicle-mediated Transport	93 (660)	18.05	
5	25.02	CDK-mediated Phosphorylation and Removal of Cdc6`	114 (880)	18.12	The UPS plays a central role (21)
6	24.58	Gene Expression	203 (1841)	29.46	
7	21.52	Protein Processing in Endoplasmic Reticulum	34 (166)	20.01	Integrally involved in the UPS (22)
8	21.33	DNA Damage	49 (292)	13.79	Closely involve the UPS (23)
9	21.24	Cell Cycle, Mitotic	84 (622)	13.65	Tightly regulated by the UPS (24)
10	**19.58**	**Ubiquitin-Proteasome Dependent Proteolysis**	**27 (122)**	**18.4**	
11	**19.06**	**Metabolism of Proteins**	**175 (1628)**	**25.5**	
12	18.59	Regulation of Degradation of DeltaF508 CFTR in CF	18 (63)	10.93	Dominated by the UPS (25)
13	16.6	Cell Cycle	28 (145)	-	
14	**16.59**	**Ubiquitin Mediated Proteolysis**	**27 (137)**	**10.94**	
15	16.23	Signaling By Hedgehog	27 (139)	12.4	
16	**15.57**	**Proteolysis_Putative Ubiquitin Pathway**	**12 (35)**	**-**	
17	15.39	Cellular Response to Heat Stress	20 (89)	13.16	Heat shock proteins recognize misfolded proteins and incorporate the UPS (26)
18	15.35	Transcription-Coupled Nucleotide Excision Repair (TC-NER)	23 (112)	-	
19	15.29	Nucleotide Excision Repair	16 (61)	-	
20	14.6	Class I MHC Mediated Antigen Processing and Presentation	95 (823)	15.89	Proteins degraded by proteasomes are a major source of peptides presented by MHC class I

					molecules (27)
21	14.52	Innate Immune System	210 (2132)	22.32	
22	14.01	HIV Life Cycle	98 (865)	11.41	
23	13.78	Clathrin-mediated Endocytosis	25 (137)	-	A key process that transports a wide range of molecules from the cell surface to the interior and is closely regulated by the UPS (28)
24	13.68	Signaling By NOTCH1	22 (113)	-	
25	13.47	Copper Homeostasis	14 (54)	-	
26	13.46	Mitotic G1-G1/S Phases	27 (156)	-	
27	13.29	Transport to The Golgi and Subsequent Modification	41 (285)	17.23	
28	13.2	Regulation of Cholesterol Biosynthesis By SREBP (SREBF)	14 (55)	-	
29	13.13	Telomere C-strand (Lagging Strand) Synthesis	20 (100)	-	
30	13	Terpenoid Backbone Biosynthesis	11 (36)	-	
31	12.61	Processing of Capped Intronless Pre-mRNA	10 (31)	-	
32	12.56	Mitotic Metaphase and Anaphase	29 (180)	11.56	Tightly regulated by the UPS (29)
33	12.45	Transport of The SLBP Independent Mature MRNA	31 (199)	12.88	
34	12.31	Remodeling of Adherens Junctions	22 (121)	12.71	Cadherin, the main adhesion molecule in adherens junctions, is tightly regulated by the UPS (30)
35	12.08	Cell Cycle Checkpoints	31 (202)	-	
36	11.92	Presenilin Action in Notch and Wnt Signaling	12 (46)	-	
37	11.5	Circadian Rythm Related Genes	31 (207)	13.68	
38	11.4	RNA Transport	27 (171)	14.03	
39	11.3	CLEC7A (Dectin-1) Signaling	24 (145)	11.77	
40	11.17	Formation of HIV Elongation Complex in The Absence of HIV Tat	28 (182)	-	
41	10.96	Cellular Senescence	55 (452)	14.27	
42	10.95	RNA Polymerase II Transcription Termination	15 (72)	-	
43	10.94	Calnexin/calreticulin Cycle	10 (36)	-	
44	10.86	Mechanisms of CFTR Activation By S-nitrosoglutathione (normal and CF)	11 (43)	9.96	
45	**10.75**	**Proteolysis Role of Parkin in The Ubiquitin-Proteasomal Pathway**	**15 (73)**	**15.85**	
46	10.75	Sterol Regulatory Element-Binding	15 (73)	-	

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

		Proteins (SREBP) Signalling			
47	10.67	Metabolism	235 (2543)	10.35	
48	10.52	Cytoskeletal Signaling	40 (304)	12.9	
49	10.24	Brain-Derived Neurotrophic Factor (BDNF) Signaling Pathway	23 (144)	11.87	

Table 5S: Pathway enrichment analysis of MSSM STG up-regulated genes with corrected p-value < 0.05. GeneAnalytics tool superpathways that were found to be enriched in the list of up-regulated genes are ordered by descending order of their enrichment score. The number of up-regulated genes that belong to each superpathway is given, with the number of genes of each superpathway in parentheses.

#	Score	SuperPath Name	Num Matched (SuperPath) genes
1	23.36	GPCR Pathway	59 (708)
2	20.4	Metabolism	148 (2544)
3	18.85	Influenza Viral RNA Transcription and Replication	21 (158)
4	18.14	ERK Signaling	79 (1177)
5	17.08	Metabolism of Proteins	100 (1628)
6	17.07	PEDF Induced Signaling	54 (721)
7	16.36	Degradation of The Extracellular Matrix	29 (298)
8	15.13	RRNA Processing in The Nucleus and Cytosol	22 (203)
9	14.93	Influenza A	29 (315)
10	14.54	Pathways in Cancer	41 (528)
11	13.8	TGF-Beta Pathway	47 (652)
12	13.49	CREB Pathway	40 (528)
13	13.31	Integrin Pathway	42 (568)
14	12.88	HIV Life Cycle	57 (865)
15	12.2	Regulation of Insulin-like Growth Factor (IGF) Transport and Uptake By Insulin-like Growth Factor Binding Proteins (IGFBPs)	6 (21)
16	11.18	Naphthalene Metabolism	6 (24)
17	11.06	Neuropathic Pain-Signaling in Dorsal Horn Neurons	21 (232)
18	11.03	Focal Adhesion	24 (283)
19	10.66	Actin Nucleation By ARP-WASP Complex	27 (341)
20	10.64	Akt Signaling	45 (681)
21	10.6	PAK Pathway	45 (682)
22	10.46	Apoptotic Pathways in Synovial Fibroblasts	47 (725)
23	10.37	FOXM1 Transcription Factor Network	7 (37)
24	10.29	NRF2 Pathway	15 (145)
25	10.24	Cell Adhesion_ECM Remodeling	9 (61)
26	10.18	G-Beta Gamma Signaling	27 (349)
27	10.15	Phospholipase-C Pathway	35 (498)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 6S: Pathway enrichment analysis of MSSM STG down-regulated genes with corrected p-value < 0.05. GeneAnalytics tool superpathways that were found to be enriched in the list of down-regulated genes are ordered by descending order of their enrichment score. The number of down-regulated genes that belong to each superpathway is given, with the number of genes of each superpathway in parentheses. UPS directly related pathways are in bold

#	Score	SuperPath Name	Num Matched (SuperPath) genes
1	33.59	MRNA Splicing - Major Pathway	38 (307)
2	29.46	Gene Expression	116 (1841)
3	**25.5**	**Metabolism of Proteins**	**102 (1628)**
4	24.61	Translational Control	25 (189)
5	22.32	Innate Immune System	121 (2132)
6	20.01	Protein Processing in Endoplasmic Reticulum	21 (166)
7	18.43	Chks in Checkpoint Regulation	24 (224)
8	**18.4**	**Ubiquitin-Proteasome Dependent Proteolysis**	**17 (122)**
9	18.12	CDK-mediated Phosphorylation and Removal of Cdc6	59 (880)
10	18.05	Vesicle-mediated Transport	48 (660)
11	17.23	Transport to The Golgi and Subsequent Modification	27 (285)
12	15.89	Class I MHC Mediated Antigen Processing and Presentation	54 (823)
13	**15.85**	**Proteolysis Role of Parkin in The Ubiquitin-Proteasomal Pathway**	**12 (73)**
14	14.27	Cellular Senescence	34 (452)
15	14.03	RNA Transport	18 (171)
16	13.93	Beta-Adrenergic Signaling	26 (308)
17	13.79	DNA Damage	25 (292)
18	13.75	Telomere Extension By Telomerase	6 (19)
19	13.68	Circadian Rythm Related Genes	20 (207)
20	13.65	Cell Cycle, Mitotic	42 (622)
21	13.16	Cellular Response to Heat Stress	12 (89)
22	12.9	Cytoskeletal Signaling	25 (304)
23	12.88	Transport of The SLBP Independent Mature MRNA	19 (199)
24	12.71	Remodeling of Adherens Junctions	14 (121)
25	12.49	Integrated Breast Cancer Pathway	16 (154)
26	12.4	Signaling By Hedgehog	15 (139)
27	12.19	Immune Response_IL-6 Signaling Pathway	7 (33)
28	12.1	Ran Pathway	5 (15)
29	11.87	Brain-Derived Neurotrophic Factor (BDNF) Signaling Pathway	15 (144)
30	11.86	Signaling By Wnt	26 (338)
31	11.77	CLEC7A (Dectin-1) Signaling	15 (145)
32	11.67	RNA Polymerase II Transcription Initiation And Promoter Clearance	19 (213)
33	11.56	Mitotic Metaphase and Anaphase	17 (180)
34	11.41	HIV Life Cycle	51 (865)
35	11.39	ErbB1 Downstream Signaling	12 (102)
36	11.3	Viral MRNA Translation	36 (547)
37	11.21	Cytokine Signaling in Immune System	46 (761)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

38	10.95	Antigen Processing-Cross Presentation	13 (121)
39	**10.94**	**Ubiquitin Mediated Proteolysis**	**14 (137)**
40	10.93	Regulation of Degradation of DeltaF508 CFTR in CF	9 (63)
41	**10.59**	**Deubiquitination**	**23 (301)**
42	10.56	TRNA Processing	12 (109)
43	10.35	Metabolism	120 (2543)
44	10.25	Sulfur Amino Acid Metabolism	8 (54)
45	10.25	Translation Factors	8 (54)
46	10.24	WNT Signaling	20 (250)
47	10.07	Apoptotic Pathways in Synovial Fibroblasts	43 (725)
48	9.96	Mechanisms of CFTR Activation By S-nitrosoglutathione (normal and CF)	7 (43)

Table 7S: STG down-regulated genes enriched UPS-related pathways; numbers of SMRI and MSSM hits. For each of the 5 UPS-related pathways that were enriched in the SMRI STG down-regulated genes, the number of 'hits' (down-regulated genes that belong to the pathway) shared with MSSM STG down-regulated genes is given in the second column. The numbers of hits specific to SMRI and MSSM down-regulated genes are given in columns 3 and 4, respectively.

UPS pathway	# SMRI+MSSM hits	# SMRI only hits	# MSSM only hits
Ubiquitin-Proteasome Dependent Proteolysis	4	27	17
Metabolism of Proteins	27	235	120
Ubiquitin Mediated Proteolysis	5	27	14
Proteolysis Role of Parkin in The Ubiquitin-Proteasomal Pathway	2	15	12
Proteolysis_Putative Ubiquitin Pathway	0	12	0 (pathway wasn't enriched in MSSM)

Table 8S: SMRI STG Differential expression analysis of proteasome subunits genes. 7 schizophrenia samples ("Group 2") were compared to the 15 control samples. T-sample t-test was applied; p-values were corrected by the Benjamini-Hochberg procedure (8)

#	Gene Symbol	t-statistic	p-value	Corrected p-value
1	PSMC4	-5.08	5.75E-05	0.00282
2	PSMA2	-4.51	0.000215	0.00527
3	PSMC2	-4.28	0.000366	0.00597
4	PSMB5	-3.77	0.00121	0.0123
5	PSMB6	-3.75	0.00125	0.0123
6	PSMD2	-3.45	0.00253	0.0174
7	PSMB2	-3.42	0.00268	0.0174
8	PSMD3	-3.4	0.00285	0.0174
9	PSMC6	-3.35	0.0032	0.0174
10	PSMD8	-3.27	0.00386	0.0189
11	PSME4	-3.11	0.0055	0.0241
12	PSMA7	-3.08	0.0059	0.0241

13	PSMD14	-2.84	0.0101	0.0381
14	PSMB7	-2.79	0.0112	0.0393
15	PSMA5	-2.74	0.0127	0.0394
16	PSMA1	-2.73	0.0129	0.0394
17	PSMD13	-2.48	0.0221	0.0636
18	PSMA6	-2.39	0.0266	0.0723
19	PSME3	-2.34	0.0298	0.0768
20	PSMD6	-2.21	0.0393	0.0963
21	PSME2	2.17	0.0426	0.0995
22	PSMD11	-2.1	0.0484	0.107
23	PSMB10	1.74	0.0975	0.199
24	PSMG2	1.7	0.104	0.203
25	PSMB9	1.57	0.132	0.25
26	PSMB8	1.51	0.148	0.268
27	PSMC1	-1.48	0.155	0.271
28	PSMG1	-1.39	0.179	0.298
29	PSMB1	-1.38	0.182	0.298
30	PSMA4	-1.33	0.199	0.315
31	PSMD7	-1.23	0.234	0.349
32	PSMC5	-1.22	0.235	0.349
33	PSMD5	1.14	0.27	0.389
34	PSMA3	-1.11	0.279	0.39
35	PSMG3	-0.95	0.353	0.481
36	PSMG4	-0.835	0.413	0.548
37	PSMD9	-0.75	0.462	0.595
38	PSMB4	-0.71	0.486	0.61
39	PSMD1	0.645	0.526	0.645
40	PSMD12	-0.537	0.597	0.701
41	PSMD4	-0.532	0.601	0.701
42	PSMD10	0.486	0.632	0.708
43	PSMB3	-0.481	0.636	0.708
44	PSMF1	-0.458	0.652	0.71
45	PSME1	0.258	0.799	0.852
46	PSMC3	0.157	0.877	0.914
47	PSMC3IP	0.0853	0.933	0.936

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 9S: **SMRI STG comparison between two differential expression analyses of the 12 proteasome subunits genes that were found to be down-regulated in the meta-analysis.** Step-wise linear regression analyses was applied, where age, PMI, gender, pH and RIN where included as covariates. Then diagnosis coefficient was then statistically tested for being nonzero, implying an effect for schizophrenia on the expression, beyond any other effect of the covariates. This produced t-statistic and corresponding p-values are listed in columns 3 and 4. The t-statistic and corresponding p-values of standard two-sided t-test analysis are listed in columns 5 and 6. P-values lower than 0.05 are marked with bold

#	Gene Symbol	SMRI STG Reg. t-stat	P-value	SMRI STG Two-sided t-stat	P-value
1	PSMA2	-3.90989	**0.000591**	-3.55651	**0.001412**
2	PSMA5	-1.73142	0.095702	-1.81818	0.080148
3	PSMA6	-1.8989	0.069177	-1.91308	0.066398
4	PSMA7	-1.51848	0.140961	-1.83716	0.077217
5	PSMB2	-0.7631	0.452841	-1.76824	0.08832
6	PSMB6	-1.69528	0.101968	-1.99045	0.056753
7	PSMC2	-2.72472	**0.011816**	-2.8391	**0.008491**
8	PSMC4	-1.38437	0.178478	-2.00554	0.055022
9	PSMC6	-1.59099	0.123699	-1.90557	0.067406
10	PSMD6	-0.53925	0.594488	-1.34314	0.19041
11	PSMD11	-1.74248	0.092806	-1.74396	0.092542
12	PSMD14	-1.87932	0.072397	-2.13114	**0.042333**

Table 10S. **MSSM STG comparison between two differential expression analyses of the 12 proteasome subunits genes that were found to be down-regulated in the meta-analysis.** Step-wise linear regression analyses was applied, where age, PMI, gender and pH where included as covariates. Then diagnosis coefficient was then statistically tested for being nonzero, implying an effect for schizophrenia on the expression, beyond any other effect of the covariates. This produced t-statistic and corresponding p-values are listed in columns 3 and 4. The t-statistic and corresponding p-values of standard two-sided t-test analysis are listed in columns 5 and 6. P-values lower than 0.05 are marked with bold

#	Gene Symbol	MSSM STG Reg. t-stat	P-value	MSSM STG 2-tstat	P-value
1	PSMA2	-1.95628	0.059799	-2.94196	**0.006123**
2	PSMA5	-0.95177	0.34908	-1.63121	0.112966
3	PSMA6	-2.29707	**0.029025**	-2.11272	**0.042778**
4	PSMA7	-2.81585	**0.008384**	-2.81605	**0.00838**
5	PSMB2	3.969405	**0.000398**	3.969429	**0.000398**
6	PSMB6	1.147369	0.260007	1.147403	0.259994
7	PSMC2	-2.51747	**0.017202**	-2.51781	**0.017188**
8	PSMC4	-1.83756	0.07573	-1.83758	0.075726
9	PSMC6	-2.83044	**0.008087**	-2.83057	**0.008085**
10	PSMD6	-1.32987	0.193583	-2.36405	**0.024523**
11	PSMD11	-0.92428	0.362967	-2.50922	**0.017538**
12	PSMD14	-3.59292	**0.001153**	-2.85374	**0.007634**

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

1. Hakak Y, Walker JR, Li C, Wong WH, Davis KL, Buxbaum JD, *et al.* (2001): Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia, 2001/04/11. *Proc Natl Acad Sci U S A* 98: 4746–4751.
2. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW (2019): Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 10: 1–10.
3. Katsel PL, Davis KL, Haroutunian V (2005): Large-scale microarray studies of gene expression in multiple regions of the brain in schizophrenia and Alzheimer's disease, 2005/03/31. *Int Rev Neurobiol* 63: 41–82.
4. Schizophrenia Working Group of the Psychiatric Genomics C (2014): Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511: 421–427.
5. Katsel P, Davis KL, Gorman JM, Haroutunian V (2005): Variations in differential gene expression patterns across multiple brain regions in schizophrenia, 2005/06/01. *Schizophr Res* 77: 241–252.
6. Purohit DP, Perl DP, Haroutunian V, Powchik P, Davidson M, Davis KL (1998): Alzheimer disease and related neurodegenerative diseases in elderly patients with schizophrenia: a postmortem neuropathologic study of 100 cases, 1998/03/24. *Arch Gen Psychiatry* 55: 205–211.
7. Pepper SD, Saunders EK, Edwards LE, Wilson CL, Miller CJ (2007): The utility of MAS5 expression summary and detection call algorithms, 2007/08/01. *BMC Bioinformatics* 8: 273.
8. Benjamini Y, Hochberg Y (1995): Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc* 57: 289–300.
9. Choi JK, Yu U, Kim S, Yoo OJ (2003): Combining multiple microarray studies and modeling interstudy variation. *Bioinformatics* 19: i84–i90.
10. Hedges L V. (1981): Distribution Theory for Glass's Estimator of Effect Size and Related Estimators. *J Educ Stat* 6: 107.
11. Schwarzer G (2007): Meta: An R package for meta-analysis. *R news* 7: 40–45.
12. Fleiss J (1993): Review papers : The statistical basis of meta-analysis. *Stat Methods Med Res* 2: 121–145.
13. Torrey EF, Webster M, Knable M, Johnston N, Yolken RH (2000): The stanley foundation brain collection and neuropathology consortium. *Schizophr Res* 44: 151–155.
14. Ballman K V., Grill DE, Oberg AL, Therneau TM (2004): Faster cyclic loess: normalizing RNA arrays via linear models. *Bioinformatics* 20: 2778–2786.
15. Barnes MR, Huxley-Jones J, Maycox PR, Lennon M, Thornber A, Kelly F, *et al.* (2011): Transcription and pathway analysis of the superior temporal cortex and anterior prefrontal cortex in schizophrenia. *J Neurosci Res* 89: 1218–1227.
16. Chen C, Meng Q, Xia Y, Ding C, Wang L, Dai R, *et al.* (2018): The transcription factor POU3F2 regulates a gene coexpression network in brain tissue from patients with psychiatric disorders. *Sci Transl Med* 10. https://doi.org/10.1126/scitranslmed.aat8178
17. Arion D, Corradi JP, Tang S, Datta D, Boothe F, He A, *et al.* (2015): Distinctive transcriptome alterations of prefrontal pyramidal neurons in schizophrenia and schizoaffective disorder. *Mol Psychiatry* 20: 1397–1405.
18. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003): Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4: 249–264.
19. Ramaker RC, Bowling KM, Lasseigne BN, Hagenauer MH, Hardigan AA, Davis

NS, *et al.* (2017): Post-mortem molecular profiling of three psychiatric disorders. *Genome Med* 9: 72.
20. Maycox PR, Kelly F, Taylor A, Bates S, Reid J, Logendra R, *et al.* (2009): Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function. *Mol Psychiatry* 14: 1083–1094.
21. Méndez J, Stillman B (2000): Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. *Mol Cell Biol* 20: 8602–12.
22. Kostova Z, Wolf DH (2003): For whom the bell tolls: protein quality control of the endoplasmic reticulum and the ubiquitin-proteasome connection. *EMBO J* 22: 2309–17.
23. Bergink S, Jentsch S (2009): Principles of ubiquitin and SUMO modifications in DNA repair. 458: 461–467.
24. Bassermann F, Eichner R, Pagano M (2014): The ubiquitin proteasome system — Implications for cell cycle control and the targeted treatment of cancer. *Biochim Biophys Acta - Mol Cell Res* 1843: 150–162.
25. Ahner A, Gong X, Frizzell RA (2013): Cystic fibrosis transmembrane conductance regulator degradation: cross-talk between the ubiquitylation and SUMOylation pathways. *FEBS J* 280: 4430–8.
26. Flick K, Kaiser P (2012): Protein degradation and the stress response. *Semin Cell Dev Biol* 23: 515–22.
27. Stoltze L, Nussbaum AK, Sijts A, Emmerich NPN, Kloetzel PM, Schild H (2000): The function of the proteasome system in MHC class I antigen processing. *Immunology Today*, vol. 21 21: 317–319.
28. Piper RC, Dikic I, Lukacs GL (2014): Ubiquitin-dependent sorting in endocytosis. *Cold Spring Harb Perspect Biol* 6. https://doi.org/10.1101/cshperspect.a016808
29. Stegmeier F, Rape M, Draviam VM, Nalepa G, Sowa ME, Ang XL, *et al.* (2007): Anaphase initiation is regulated by antagonistic ubiquitination and deubiquitination activities. *Nature* 446: 876–881.
30. Kowalczyk AP, Nanes BA (2012): Adherens junction turnover: regulating adhesion through cadherin endocytosis, degradation, and recycling. *Subcell Biochem* 60: 197–222.