Transcriptomic Evidence for Alterations in Astrocytes and Parvalbumin Interneurons in Bipolar Disorder and Schizophrenia Subjects

Supplemental Information

Contents

Supplementary methods

Datasets

Table S1 contains full information on the datasets analysed. Data from Stanley Medical Research Institute (SMRI) were downloaded from https://www.stanleygenomics.org/stanley. Data for datasets GSE17806, GSE53987, GSE35978, GSE13564, GSE80655 and GSE25219 were downloaded from the Gene Expression Omnibus (GEO) https://www.ncbi.nlm.nih.gov/geo. The McLean dataset was obtained from the Harvard Brain Tissue Resource Center (http://national_databank.mclean.harvard.edu/, accessed in 2004). Information regarding the brain pH and the exact age of the samples in McLean dataset was obtained from an earlier study based on the same cohort of subjects (1). The samples were matched using the common metadata fields (disease information, sex, post-mortem interval, age, 3'/5' GAPDH ratio, 3'/5' ACTB ratio). For two samples, the metadata available did not match the metadata in Veldic et al. (1) so they were excluded from further analysis.

Raw data pre-processing

Microarray data used in the study were pre-processed and normalized with the "rma" function of the "oligo" (Affymetrix gene arrays) or "affy" (Affymetrix 3'IVT arrays) R packages (2). Probeset to gene annotations were obtained from Gemma (3) (http://gemma.msl.ubc.ca/). The data was batch corrected using the "ComBat" function from the "sva" R package (4). The scan date embedded in the CEL files was used to assign batch information to each sample. We used a simple set of heuristics to group samples into batches, which was manually confirmed that the batch assignment was appropriate. We did not perform ComBat if batch is confounded with the Group factor (namely Cont/BP/SCZ). As a result, we did not perform ComBat in dataset Study3Bahn, since one of the batches only included subjects with SCZ. We note that in this particular dataset the CEL files were scanned over a period of a year, and even after binning the scan dates the number of batches remained high (11 batches for 99 samples).

RNAseq data pre-processing (GSE80655): Raw data was downloaded from GEO. Reads were mapped using the STAR Aligner (version 2.4.0h) and then quantified using the RSEM transcript quantifier (version 1.2.31). The transcriptome was prepared using the 'rsem-prepare-reference' script provided by the RSEM project.

We used the expression level of sex-specific genes (*XIST*, *KDM5D*, *RPS4Y1*) to identify misannotated samples as previously described (5). Supplementary Table S1 contains the number of misannotated samples in each of the datasets analysed. A background expression threshold was defined for each dataset as the 95th quantile of signals from probeset annotated to the male-specific genes (*KDM5D, RPS4Y1*) in samples assigned as "female" and probesets annotated to the female-specific gene (*XIST*) in samples assigned as "male". We removed all probesets expressed below this threshold in more than 95% of the samples in the dataset. Next, if multiple remaining probesets were annotated to the same gene, the most variable probeset was selected for the analysis.

Estimation of marker gene profiles

The general approach for estimation of marker gene profiles (MGPs) was previously described (6). Briefly, we first obtained the marker gene-sets for all (mouse) cell types relevant to the brain region of the bulk tissue dataset analysed from NeuroExpresso via neuroexpresso.org. These gene-sets were selected separately for each brain region, based on the specificity of their expression relative to other cells in that region. Supplementary Table S2 provides information regarding all cell types analysed. Human orthologues of the mouse genes in NeuroExpresso were defined using HomoloGene (ftp://ftp.ncbi.nih.gov/pub/HomoloGene/build68/homologene.data) (7). In cases where a mouse gene was mapped to more than one human orthologue, all orthologues were used. MGPs for the individual cell types were calculated based on the corresponding marker gene sets, using Principal component analysis (PCA), as previously described (6).

To reduce the impact of outlier samples, PCA was repeated 100 times on subsampled data, containing equal number of subjects per group (controls, bipolar, schizophrenia). Namely, N samples were randomly selected without replacement from each group, where N represents 90% of subjects in the smallest group, and the mean score for each sample was used for downstream analyses. MGPs were not estimated in cases where less than three cell type marker-genes met the dataset-specific expression level thresholds.

Marker gene-sets for cell types present in multiple brain regions (such as glial cells) were overlapping but not necessarily identical in different brain regions. To exclude the possibility that the region specificity of changes in astrocyte MGPs in psychiatric patients resulted from differences in marker gene-sets used for their calculation, in each subcortical region astrocyte MGPs were also recalculated using the cortical astrocyte marker gene-set.

To ensure that the results of our analysis is not an artifact of the estimation method, we reanalyzed the data using CellCODE (8), demonstrating high correlation between the outcomes of the two methods (Supplementary Fig. S18-20).

QC of MGP analysis

Since MGPs are not actual cell counts, a QC assessment of the outcome was performed to ensure the validity of MGPs as correlates of cell type-specific alterations: For MGPs to be reliable surrogates of cell type-specific changes, they should capture a large proportion of variance in the expression level of the relevant marker genes (namely, the variance explained by the first principal component based on the marker genes). This would indicate that the genes are changing concordantly (in other words, coexpressed), as expected in case when alterations in cellular abundance are taking place. When the variance explained by MGPs is low (in our experience, below 35%), it means that the marker genes are poorly coexpressed indicating that MGPs are not good surrogates of relative cell proportions (Mancarci et al. in preparation).

Statistical analysis of MGPs

For each dataset, a two sided Wilcoxon rank-sum test ("wilcox.test" function from the "stats" package in R, default options) was used to compare MGPs between diagnostic groups. We also used mixed models ("lmer" function from "lme4" R package, using REML = FALSE option) to pool MGPs across the datasets (fixed effects: sex, age, PMI and pH, random effects: study, subject). For each cell type, the significance of each of the covariates (sex, age, PMI and pH) was first tested individually against a null model containing only intersect and random effects, using the "anova" function from "stats" R package and only the significant effects were included in the full model. Confidence intervals were calculated using "confint.merMod" function from "lme4" R package, using method = "boot" option. For visualization purposes, scores were normalized to the range 0-1.

An increase in the abundance of a single cell type is likely to affect the MGP estimates for other cell types in the sample. This is because given a constant amount of RNA per tissue sample, an increase in the proportion of RNA coming from one cell type must be accompanied by a decrease in RNA from other cell types. We therefore also analyzed models in which we included astrocyte

or fsPV MGPs as covariates in the linear mixed models and recalculated the estimated group effects for each of the other MGPs.

Gene-MGP correlation

Spearman's correlation between genes and MGPs of each of the cortical cell types was calculated separately for each of the datasets analysed using "cor.test" function from the "stats" R package. List of probesets with altered expression in the cortex of SCZ subjects was obtained from Supplementary Table 2 in Mistry et al. 2013 (9).

Developmentally regulated genes in fsPV interneurons

Data from a time course of mouse developing fsPV cells (10) was downloaded from GEO (GSE17806). The effect of age on each of the genes on the microarray platform was evaluated using linear models ("lm" function from "stats" R package), treating the age of the animals as an ordered factor. The false discovery rate (FDR) was controlled using "p.adjust" function from the "stats" R package, using method = "BH". Developmentally regulated genes were defined as genes selected while controlling the FDR at 0.01. This identified 725 genes downregulated and 1029 genes upregulated during fsPV maturation, of which 659 and 1029 genes (respectively) had human orthologues and were included in the analysis.

Gene expression in human brain cell types

The expression level of top genes correlated with fsPV and astrocyte MGPs in human cell types was evaluated based on single-cell expression data from Darmanis et al. (GSE67835) (6, 11). In order to match the single cells to our bulk tissue data we only used cells purified from adult subjects, classified by Darmanis et al. into six major cell types: astrocytes, endothelial cells, microglia, oligodendrocytes, oligodendrocyte precursor cells (OPCs) and neurons.

Differential expression analysis

Differential expression based on all 11 datasets (all but the CommonMind data) was tested using mixed models ("lmer" function from "lme4" R package, using REML = FALSE option) by testing the full model (fixed effects: group, sex, age PMI and pH, random effects: study, subject) against the null (all covariates not including the group) using the "anova" function from "stats" R package. The significance of genes was tested separately for BP and SCZ subjects. The false discovery rate

(FDR) was controlled using "p.adjust" function from the "stats" R package, using method = "BH". To test the effect of adjusting for MGPs, differential expression analysis was repeated with MGPs as covariates in the model.

Functional enrichment of cortical cell type transcriptomes

Mouse cell type-specific expression profiles were downloaded from the NeuroExpresso database (6), accessible from http://www.chibi.ubc.ca/supplement-to-mancarci-et-al-neuroexpresso/. Functional enrichment analysis was performed separately for cell types represented by microarray and RNAseq data. For each gene, the signals from individual samples were z-transformed and summarised at the cell type level by taking the average of all samples representing the same cell type (normalized scores), yielding a ranking of genes from high to low expression for each cell type. The functional role of the overexpressed genes was assessed based on GO annotations (Sep 15th 2017 version) and the enrichment of specific terms was evaluated using the precision-recall method implemented in ErmineJ (12, 13) based on the normalized gene scores for each cell type.

Functional enrichment analysis of underexpressed genes

Functional enrichment was evaluated using the precision-recall method implemented in ErmineJ (12, 13) based on one-sided test of the differential expression results, using only "Biological Function" category terms and limiting the gene set size to 20-200. Mitochondria-related terms were identified by filtering GO terms with names matching the regular expression "mitochond|respiratory ele|ATP|oxidative phosph|nucleoside triphosphate", yielding 53/2535 terms. Significant terms were defined as GO terms meeting an enrichment FDR of 0.1.

Variance in gene expression explained by each of the covariates of the mixed model

The variance in gene expression was evaluated using the "fitExtractVarPartModel" function from the "variancePartition" R package. The function extracts the additional variance explained by each covariate based on ANOVA, comparing the full model to model without the covariate.

QC and analysis of the CommonMind and the PsychEncode data

Preprocessed data was downloaded from Synapse (www.synapse.org).

CommonMind data: We did not analyze bipolar disorder samples since their age range was much lower than that of the control and schizophrenia groups; this confound would make it impossible

to attribute effects to diagnosis in the bipolar subset. In addition, we excluded samples from subjects older than 90y, samples that were outliers based on rRNA contamination (DLPFC_RNA_Sequencing_rRNA_Rate), and samples expressing both male and female sex marker genes (three of these samples were identified as cases of Klinefelter syndrome, and the remaining were considered as having quality problems). After filtering we retained data for 164 subjects with schizophrenia and 168 controls.

PsychEncode data: The data included samples from five cohorts of subjects, of these, only SMRI New and SMRI Extra cohorts were included in the analysis. These cohorts represent a total of 26 controls, 27 subjects with BP and 45 subjects with SCZ. The BSHRI cohort was excluded from the analysis since it only included control subjects, and their age range was much higher than of the other cohorts. Two additional cohorts (SMRI Array and SMRI Consortium) were already analyzed as part of the original 13 datasets and thus were excluded from this analysis. The dataset was subjected to QC evaluation similar to the one described for the CommonMind data, with no metadata outliers detected. A single SCZ sample was identified as outlier based on the expression data and removed as part of our general data analysis pipeline.

Supplementary results

This section describes our analysis of the CommonMind and the PsychEncode data along with discussion of the findings.

Analysis of the PsychEncode data

MGP analysis was performed as described in the methods section. Statistical significance of the results was evaluated using linear models, including Age, pH, PMI and RIN as covariates. Sex was not included as a covariate since only one control sample was a female subject. The analysis indicated an increase in the astrocyte MGP and a decrease in fsPV MGPs in subjects with BP and SCZ (Supplementary Fig. S15)

Analysis of the CommonMind data

RNA quality was reported to be a confound in the CommonMind dataset (14). We thus were not surprised to observe poor initial quality control in our MGP analysis. Specifically, the astrocyte MGP captured only 29% of the variance in astrocyte marker genes, much lower than in the other

11 cortical datasets (40-58%, Supplementary Table S7). In addition, we observed high correlation between RIN and fsPV MGP ($|r| = 0.75$, Supplementary Fig. S17, upper plane), suggesting that the sample-sample differences in fsMGP might be driven by differences in RNA quality. Since the reason for poor QC in the MGP analysis appeared to be related to RNA quality of the samples, we first repeated the analysis using subsets of samples exhibiting high RIN values, using different RIN thresholds. However, these subsets still exhibited poor QC in MGP analysis. Thus, since we could not directly identify good quality samples, we repeated the MGP analysis on 10,000 random subsamples of the data, by randomly selecting (without replacement) 30 samples from each diagnostic group. This is similar to the resampling procedure used for the other data sets, with the difference being the number of trials, and the proportion of the dataset sampled. The high number of trials was necessary to obtain multiple trials for which the variance explained by astrocyte MGP was in the range of the other datasets (as described below). For each subsample, the variance captured by the MGP (PC1 based on the cell type marker genes) and the sample metadata were captured. For majority of the subsamples (70%) the astrocyte MGP captured more variance than when analyzing whole dataset (Supplementary Fig. S16, upper panel). In 63/10,000 (0.63%) of the subsamples more than 40% of the variance was captured by the astrocyte MGP (the lower boundary of the other 11 datasets). Overall, the results based on the subsampled data corroborated the outcome of our main analysis, namely, an increase in astrocyte and a decrease in fsPV MGPs in subjects with schizophrenia (Supplementary Fig. S16). For the astrocyte MGP, the effect size was positively correlated with the variance captured by MGPs, with the largest effect observed in all subsamples for which more than 40% of the variance was captured. For the fsPV neuron MGPs, in all subsamples the variance captured by MGP was in the range of other datasets, and in all of the subsamples MGPs were decreased in the schizophrenia group, in agreement with our main analysis (Supplementary Fig. S16, bottom).

Of notice, even in the subsampled data, RIN remained highly correlated with fsPV MGP ($|r|$: 0.45-0.89, median = 0.75), indicating that RNA quality remained a confound even in the subsampled dataset. Importantly, this high correlation was not observed in datasets GSE53987 and PsychEncode, the two other data sets for which RIN values were available (GSE53987 RIN_fsPV $|r|$: 0.09-0.42, median = 0.24; PsychEncode RIN_fsPV $|r|$: 0.01-0.27, median = 0.15, Supplementary Fig. S17).

To summarize, the results of our MGP analysis of the CommonMind and PsychEncode datasets are consistent with and corroborate our findings based on the other 11 data sets. However, the results for the CommonMind data were obtained in the face of substantial data quality concerns, especially a confound with RNA quality, which was not apparent in the other data sets analyzed.

Supplementary figures and tables

Table S1 – Description of the psychiatry expression data analysed in the current study.

Dataset	Brain Region (Org)	Platform	Noise	Sample (Org)	Batch Excluded	Outlier Excluded	MM Excluded	Sample (Final)
Cortex								
GSE21138	BA46	GPL570	7.1	59	1	9	2	47 (25 Cont, 22 SCZ)
GSE35978	Parietal cortex	GPL6244	5.89	147	1	0	1	145 (50 Cont, 44 BP, 51 SCZ)
GSE53987	BA46	GPL570	5.05	51	0	0	1	50 (18 Cont, 17 BP, 15 SCZ)
GSE80655#	DLPFC	GPL11154	0.95	71	0	2	0	69 (24 Cont, 22 BP, 23 SCZ)
McLean	Cortex	GPL96	6.46	62	0	0	1	61 (25 Cont, 17 BP, 19 SCZ)
Study1AltarA	BA46	GPL570	6.64	98	0	9	1	88 (30 Cont, 26 BP, 32 SCZ)
Study2AltarC	BA46/10	GPL570	5.51	33	1	2	5	25 (10 Cont, 7 BP, 8 SCZ)
Study3Bahn	BA46	GPL570	5.9	99	0	0	2	97 (33 Cont, 32 BP, 32 SCZ)
Study4Chen	BA6	GPL96	7.09	27	0	2	0	25 (12 Cont, 13 BP)
Study5Dobrin	BA46	GPL96	6.87	78	0	8	5	65 (20 Cont, 24 BP, 21 SCZ)
Study7Kato	BA46	GPL570	5.97	102	1	7	0	94 (31 Cont, 28 BP, 35 SCZ)
PsychEncode#	BA46	GPL11154	2.08	99	1	1	0	97 (26 Cont, 26 BP, 45 SCZ)
CommonMind#	BA9	GPL16791	2.433	332*	0	0	0	332 (164 Cont, 168 SCZ)
Total								**1195 (468 Cont, 243 BP, 484 SCZ)**
Hippocampus								
GSE53987	HPC	GPL570	5	51	1	0	4	46 (17 Cont, 16 BP, 13 SCZ)
Study17Laeng	HPC (CA1)	GPL96	5.06	61	0	2	0	59 (19 Cont, 20 BP, 20 SCZ)
Total								**105 (36 Cont, 36 BP, 33 SCZ)**
Other								
GSE53987	STR	GPL570	7.29	53	1	6	1	45 (15 Cont, 16 BP, 14 SCZ)
Study16Kemether	thalamus (MD)	GPL96	5.65	36	2	0	0	34 (12 Cont, 9 BP, 13 SCZ)
GSE35978	CRB	GPL6244	6.47	131	1	0	0	130 (50 Cont, 37 BP, 43 SCZ)

Samples (Org) - initial number of samples in the dataset. Batch excluded – number of samples excluded since their analysis date was over a month period from any other sample in the dataset. Outlier excluded – number of outlier samples. MM excluded – number of misannotated samples. Samples (Final) – number of samples included in the MGP analysis. Noise – Log_2 expression value used to define the background expression threshold. DLPFC – dorsolateral prefrontal cortex; HPC – hippocampus; STR – striatum; CRB – cerebellum.
- RNAseq data
* The number of samples remained after exclusion of samples based on metadata QC, as described in supplementary methods

Table S2 – Description of cell types analyzed

Brain region	Cell Type	Description	Calculated MGP	Linear mixed model analysis
Cortex	Astrocyte	Astrocyte cells	Yes (All)	Yes
	Endothelial	Endothelial cells	Yes (All)	Yes
	GabaPV	Fast spiking basket cells	Yes (All)	Yes
	GabaRelnCalb	Martinotti cells	Yes (4/13)	No
	GabaVIPReln	VIP and Reelin positive cells	Yes (All)	Yes
	Layer 2 3 Pyra	Layer 2+3 pyramidal cells	No (0/11)	No
	Layer 4 Pyra	Layer 4 pyramidal cells	Yes (1/13)	No
	Layer 6a Pyra	Layer 6a pyramidal cells	No (0/13)	No
	Layer 6b Pyra	Layer 6a pyramidal cells	Yes (3/13)	No
	Microglia	Microglia (all microglial markers)	Yes (All)	Yes
	Microglia_activation	Genes upregulated in activated microglia	Yes (All)	Yes
	Microglia_deactivation	Genes downregulated in activated microglia	Yes (All)	Yes
	Oligo	Oligodendrocytes	Yes (All)	Yes
	OligoPrecursors	Oligodendrocyte precursor cells	Yes (All)	Yes
	Pyramidal_Glt_25d2	GLT25D positive pyramidal cells	No (0/13)	No
	Pyramidal_S100a10	S100A10 (P11) positive pyramidal cells	No (0/13)	No
	PyramidalCorticoThalam	Cortico-thalamic pyramidal cells	No (0/13)	No
	PyramidalAll	Global markers pyramidal cells	Yes (All)	Yes
Cerebellum	Basket	Cerebellar basket cells	Yes	No
	Bergmann	Bergman glia	Yes	No
	CerebGranule	Cerebellar granule cells	Yes	No
	Golgi	Cerebellar Golgi cells	Yes	No
	Microglia	Microglia (all microglial markers)	Yes	No
	Microglia_activation	Genes upregulated in activated microglia	Yes	No
	Microglia_deactivation	Genes downregulated in activated microglia	Yes	No
	Oligo	Oligodendrocytes	Yes	No

Brain region	Cell Type	Description	Calculated MGP	Linear mixed model analysis
	Purkinje	Purkinje cells	Yes	No
Thalamus	Astrocyte	Astrocytes	Yes	No
	GabaReln	Reelin positive interneurons	Yes	No
	Hypocretinergic	Hypocretinergic cells	Yes	No
	Microglia	Microglia (all microglial markers)	Yes	No
	Microglia_activation	Genes upregulated in activated microglia	Yes	No
	Microglia_deactivation	Genes downregulated in activated microglia	Yes	No
	Oligo	Oligodendrocytes	Yes	No
	ThalamusCholin	Thalamic cholinergic cells	Yes	No
Striatum	Astrocyte	Astrocytes	Yes	No
	ForebrainCholin	Forebrain cholinergic cells	Yes	No
	Microglia	Microglia (all microglial markers)	Yes	No
	Microglia_activation	Genes upregulated in activated microglia	Yes	No
	Microglia_deactivation	Genes downregulated in activated microglia	Yes	No
	Oligo	Oligodendrocytes	Yes	No
	Spiny	Medium spiny neurons (DRD1 or DRD2 positive)	Yes	No
Hippocampus	Astrocyte	Astrocytes	Yes	No
	DentateGranule	Dentate gyrus granule cells	Yes	No
	GabaSSTReln	Somatostatin and Reelin positive interneurons	Yes	No
	Microglia	Microglia (all microglial markers)	Yes	No
	Microglia_activation	Genes upregulated in activated microglia	Yes	No
	Microglia_deactivation	Genes downregulated in activated microglia	Yes	No
	Oligo	Oligodendrocytes	Yes	No
	Pyramidal Thy1 Hipp	Hippocampal pyramidal cells	Yes	No

Full description of the cell types can be accessed through http://neuroexpresso.org/. MGPs were not calculated for cell types with less than three marker genes detected in bulk tissue data. In parenthesis, the number of relevant datasets for which MGP could be calculated. Linear mixed model was applied to cell types for which MGP was calculated for majority of the studies.

Table S3 – Coefficients and 95%CI for pH, age and sex based on linear mixed model including 11 cortical datasets of psychiatric patients.

	Mixed model coefficient				Literature validation	
CellType	pH (95%CI)	Age (95%CI)	SexM (95%CI)	PMI	Age	Sex
Astrocyte	**-0.3 (-0.4,-0.2)**	**0.002 (6e-04,0.004)**	**-0.3 (-0.09,-0.002)**	**-0.003 (-0.004,-0.001)**	(10–13)	(18)
Endothelial	**-0.2 (-0.3,-0.1)**	**0.003 (0.001,0.004)**	**-0.2 (-0.1,-0.02)**	**-0.002 (-0.003,-2e-04)**	(16)	
Oligo	-0.04 (-0.1,0.04)	0.001 (-3e-04,0.003)	-0.04 (-0.05,0.04)	-0.003 (-0.005,-0.002)	(10, 11, 14–16)	
OPC	**-0.3 (-0.3,-0.2)**	**0.002 (5e-04,0.003)**	-0.3 (-0.07,0.001)	**-0.002 (-0.003,-5e-04)**	(16)	
Microglia	**-0.07 (-0.1, -0.02)**	5e-04 (-9e-04,0.002)	-0.07 (-0.05,0.02)	**-0.002 (-0.003,-6e-04)**		
Microglia activation	**-0.2 (-0.2,-0.09**	**0.001 (1e-04,0.003)**	-0.2 (-0.06,0.01)	-0.001 (-0.002,4e-04)	(22)	
Microglia deactivation	**-0.08 (-0.1,-0.02)**	4e-04 (-9e-04,0.002)	-0.08 (-0.05,0.01)	**-0.002 (-0.003,-8e-04)**		
GabaPV	**0.3 (0.2,0.3)**	**-0.002 (-0.004,-3e-04)**	**0.3 (0.02,0.1)**	0.002 (-7e-07,0.003)	(15, 16, 23)	(23)
GabaVIPReln	**0.2 (0.1,0.2)**	**-0.005 (-0.006,-0.003)**	0.2 (-0.02,0.07)	-3e-04 (-0.002,0.001)	(15, 16, 23, 24)	
Layer6bPyra	-0.04 (-0.1,0.02)	**-0.002 (-0.004,-6e-04)**	-0.04 (-0.04,0.03)	5e-04 (-8e-04,0.002)	(15, 16, 23)	
PyramidalAll	**0.2 (0.2,0.3)**	**-0.003 (-0.005,-0.002)**	**0.2 (0.02,0.1)**	**0.002 (3e-04,0.003)**	(15, 16, 23)	(23)

The coefficients are per one pH unit, one year and being a male. Significant coefficients (95%CI not overlapping 0) are highlighted in bold. Literature validation indicate previous studies based on cell counts or computational methods corroborating our finding. OPC – oligodendrocyte precursor cells

Table S4 – Changes in MGPs in bipolar disorder and schizophrenia in sub-cortical datasets

Study (Region)	CellType	BP shift	BP p-value	SCZ shift	SCZ p-value
GSE35978 (Cerebellum)	Basket	-0.08	3.6e-02*	-0.06	0.059
	Bergmann	0.03	0.43	0.07	0.075
	CerebGranule	-0.11	6.1e-04***	-0.02	0.52
	Golgi	-0.04	0.25	-0.08	4.8e-02*
	Microglia	-0.08	0.075	-0.1	1.8e-02*
	Microglia_activation	-0.01	0.84	-0.04	0.18
	Microglia_deactivation	-0.09	0.077	-0.12	1e-02*
	Oligo	-0.01	0.87	0.03	0.32
	Purkinje	-0.05	0.18	-0.02	0.64
	AstrocyteCortex	0.05	0.23	0.1	2.4e-02*
Study16Kemether (Thalamus)	Astrocyte	0.18	0.069	0.25	1.6e-02*
	GabaReln	-0.4	1.3e-03**	-0.35	4.5e-04***
	Hypocretinergic	-0.33	4.9e-02*	-0.34	5.5e-03**
	Microglia	0.09	0.069	-0.03	0.65
	Microglia_activation	0.31	6.6e-04***	0.27	9.7e-04***
	Microglia_deactivation	-0.15	4.9e-02*	-0.31	7.5e-05***
	Oligo	0.19	0.13	0.02	0.94
	ThalamusCholin	-0.52	2e-04***	-0.43	5.3e-05***
	AstrocyteCortex	0.21	2.8e-02*	0.25	9.6e-03**
GSE35987 (Striatum)	Astrocyte	0	0.95	0.08	0.16
	ForebrainCholin	-0.01	0.86	-0.03	0.71
	Microglia	-0.09	0.18	-0.09	4.6e-02*
	Microglia_activation	-0.05	0.42	0.01	0.75
	Microglia_deactivation	-0.07	0.26	-0.14	0.063
	Oligo	0.01	0.98	-0.13	0.22
	Spiny	-0.01	0.77	-0.14	0.23
	AstrocyteCortex	0.04	0.77	0.09	0.077
GSE35987 (Hippocampus)	Astrocyte	0.05	0.63	0.1	0.36
	DentateGranule	-0.1	0.16	-0.32	6e-04***
	GabaSSTReln	-0.05	0.13	-0.36	1.5e-05***
	Microglia	-0.04	0.68	-0.05	0.39
	Microglia_activation	0.01	0.85	0.07	0.18
	Microglia_deactivation	-0.03	0.61	-0.1	0.25
	Oligo	-0.02	0.74	-0.03	0.59
	Pyramidal Thy1 Hipp	0.04	0.61	-0.18	0.1
	AstrocyteCortex	0.04	0.61	0.07	0.46
	Astrocyte	0.04	0.59	-0.02	0.79

Study (Region)	CellType	BP shift	BP p-value	SCZ shift	SCZ p-value
Study17Laeng (Hippocampus)	DentateGranule	0.01	0.97	0.01	0.86
	GabaSSTReln	-0.17	2.6e-02*	-0.23	1.7e-03**
	Microglia	-0.22	1.1e-02*	-0.12	0.17
	Microglia_activation	-0.08	0.19	-0.07	0.37
	Microglia_deactivation	-0.2	9.9e-03**	-0.14	0.061
	Oligo	-0.03	0.63	-0.07	0.4
	Pyramidal Thy1 Hipp	-0.03	0.67	0.03	0.65
	AstrocyteCortex	0.05	0.44	0	0.88

Shifts in MGPs estimated for each study separately based on Wilcoxon signed ranked test in extra-cortical datasets.

Table S5 – GO enrichment in NeuroExpresso cell types represented by microarray data

Name	ID	Gene number	P-value	Corrected P-value	MF P-value	Corrected MF P-value
Astrocyte						
---	---	---	---	---	---	---
FSBasketG42						
inner mitochondrial membrane protein complex	GO:0098800	92	2.00E-04	0.082	1.00E-12	8.20E-10
respiratory chain	GO:0070469	65	1.00E-12	1.60E-09	1.00E-12	9.90E-10
cellular respiration	GO:0045333	97	1.00E-12	2.50E-09	1.00E-12	1.20E-09
mitochondrial ATP synthesis coupled electron transport	GO:0042775	36	1.00E-04	0.055	1.00E-12	1.60E-09
ATP synthesis coupled electron transport	GO:0042773	37	1.00E-04	0.062	1.00E-12	2.50E-09
mitochondrial respiratory chain	GO:0005746	62	1.00E-12	4.90E-09	1.00E-12	4.90E-09
respiratory chain complex	GO:0098803	60	3.00E-04	0.11	1.00E-04	0.049
MartinottiGIN						
cGMP metabolic process	GO:0046068	24	2.00E-04	0.2	1.00E-12	1.20E-09
cyclic nucleotide metabolic process	GO:0009187	36	3.00E-04	0.21	1.00E-12	1.60E-09
cGMP biosynthetic process	GO:0006182	16	2.00E-04	0.25	1.00E-12	2.50E-09
neuropeptide hormone activity	GO:0005184	19	1.00E-12	2.50E-09	1.00E-12	4.90E-09
Microglia						
neutrophil migration	GO:1990266	50	1.00E-12	9.50E-11	1.00E-12	1.30E-10
positive regulation of tumor necrosis factor superfamily cytokine production	GO:1903557	53	1.00E-04	0.0085	1.00E-12	1.30E-10
regulation of tumor necrosis factor superfamily cytokine production	GO:1903555	87	1.00E-12	9.90E-11	1.00E-12	1.40E-10
regulation of lymphocyte chemotaxis	GO:1901623	12	1.00E-12	1.00E-10	1.00E-12	1.40E-10
granulocyte migration	GO:0097530	56	1.00E-12	1.10E-10	1.00E-12	1.50E-10
myeloid leukocyte migration	GO:0097529	73	1.00E-12	1.10E-10	1.00E-12	1.50E-10

Name	ID	Gene number	P-value	Corrected P-value	MF P-value	Corrected MF P-value
positive regulation of calcium ion import	GO:0090280	12	1.00E-04	0.0087	1.00E-12	1.50E-10
lymphocyte migration	GO:0072676	34	1.00E-04	0.0088	1.00E-12	1.60E-10
granulocyte chemotaxis	GO:0071621	52	1.00E-12	1.10E-10	1.00E-12	1.60E-10
cellular response to tumor necrosis factor	GO:0071356	76	1.00E-12	1.10E-10	1.00E-12	1.70E-10
Oligodendrocyte						
compact myelin	GO:0043218	16	1.00E-12	8.20E-10	1.00E-12	9.90E-10
myelination	GO:0042552	85	1.00E-12	9.90E-10	1.00E-12	1.20E-09
regulation of gliogenesis	GO:0014013	98	1.00E-12	1.60E-09	1.00E-12	1.60E-09
ensheathment of neurons	GO:0007272	87	1.00E-12	2.50E-09	1.00E-12	2.50E-09
axon ensheathment	GO:0008366	87	1.00E-12	2.50E-09	1.00E-12	2.50E-09
connexin complex	GO:0005922	12	1.00E-12	4.90E-09	1.00E-12	4.90E-09
VIPReln (G30)						
---	---	---	---	---	---	---

Top 10 significant GO terms are presented. Gene number – number of genes annotated with the term and included in the analysis; MF – multifunctionality-adjusted

Table S6 – GO enrichment in NeuroExpresso cell types represented by RNAseq data

Name	ID	Gene number	P-value	Corrected P-value	MF P-value	Corrected MF P-value
Astrocyte						
L-alpha-amino acid transmembrane transport	GO:1902475	30	1.00E-04	0.03	1.00E-12	8.00E-10
L-amino acid import	GO:0043092	10	1.00E-12	5.00E-10	1.00E-12	9.00E-10
amino acid import	GO:0043090	10	1.00E-12	5.00E-10	1.00E-12	1.00E-09
L-glutamate transport	GO:0015813	20	1.00E-12	6.00E-10	1.00E-12	1.00E-09
L-glutamate transmembrane transporter activity	GO:0005313	10	1.00E-12	3.00E-09	1.00E-12	2.00E-09
acidic amino acid transmembrane transporter activity	GO:0015172	10	1.00E-12	3.00E-09	1.00E-12	2.00E-09
L-glutamate transmembrane transport	GO:0089711	10	1.00E-12	3.00E-09	1.00E-12	2.00E-09
dicarboxylic acid transmembrane transporter activity	GO:0005310	30	1.00E-04	0.04	1.00E-12	3.00E-09
neurotransmitter uptake	GO:0001504	10	1.00E-04	0.04	1.00E-12	5.00E-09
acidic amino acid transport	GO:0015800	20	1.00E-12	7.00E-10	1.00E-04	0.05
Endothelial						
negative regulation of viral genome replication	GO:0045071	40	1.00E-04	0.04	1.00E-12	2.00E-09
response to interferon-beta	GO:0035456	20	2.00E-04	0.07	1.00E-12	3.00E-09
response to interferon-alpha	GO:0035455	20	1.00E-12	2.00E-09	1.00E-12	5.00E-09
FSBasketG42						
oxidoreductase complex	GO:1990204	90	1.00E-12	1.00E-10	1.00E-12	1.00E-10
hydrogen ion transmembrane transport	GO:1902600	90	1.00E-12	1.00E-10	1.00E-12	1.00E-10
positive regulation of potassium ion transmembrane transport	GO:1901381	20	1.00E-12	1.00E-10	1.00E-12	1.00E-10
respiratory chain complex	GO:0098803	70	1.00E-12	1.00E-10	1.00E-12	1.00E-10

Name	ID	Gene number	P-value	Corrected P-value	MF P-value	Corrected MF P-value
inner mitochondrial membrane protein complex	GO:0098800	100	1.00E-12	1.00E-10	1.00E-12	1.00E-10
respiratory chain	GO:0070469	80	1.00E-12	1.00E-10	1.00E-12	1.00E-10
cytochrome complex	GO:0070069	30	1.00E-12	1.00E-10	1.00E-12	1.00E-10
neural nucleus development	GO:0048857	60	1.00E-04	0.01	1.00E-12	1.00E-10
proton-transporting ATP synthase complex	GO:0045259	20	1.00E-12	1.00E-10	1.00E-12	2.00E-10
positive regulation of potassium ion transport	GO:0043268	30	1.00E-12	1.00E-10	1.00E-12	2.00E-10
Layer4Pyra						
inactivation of MAPK activity	GO:0000188	20	1.00E-12	5.00E-09	1.00E-12	5.00E-09
Layer6bPyra						
---	---	---	---	---	---	---
Martinotti (GIN)						
---	---	---	---	---	---	---
Microglia						
tertiary granule lumen	GO:1904724	50	1.00E-12	4.00E-10	1.00E-12	5.00E-10
cellular response to thyroid hormone stimulus	GO:0097067	20	1.00E-04	0.03	1.00E-12	6.00E-10
protein targeting to ER	GO:0045047	100	1.00E-12	4.00E-10	1.00E-12	7.00E-10
lysosomal lumen	GO:0043202	90	1.00E-12	4.00E-10	1.00E-12	8.00E-10
endolysosome	GO:0036019	20	1.00E-12	5.00E-10	1.00E-12	9.00E-10
endosome lumen	GO:0031904	20	1.00E-12	5.00E-10	1.00E-12	1.00E-09
response to acidic pH	GO:0010447	20	1.00E-12	8.00E-10	1.00E-12	1.00E-09
SRP-dependent cotranslational protein targeting to membrane	GO:0006614	90	1.00E-12	9.00E-10	1.00E-12	2.00E-09
cotranslational protein targeting to membrane	GO:0006613	90	1.00E-12	1.00E-09	1.00E-12	3.00E-09
toll-like receptor signaling pathway	GO:0002224	90	1.00E-12	5.00E-09	1.00E-12	5.00E-09
Oligodendrocyte						
oligodendrocyte differentiation	GO:0048709	60	1.00E-12	1.00E-09	1.00E-12	1.00E-09
myelination	GO:0042552	80	1.00E-12	2.00E-09	1.00E-12	2.00E-09

Name	ID	Gene number	P-value	Corrected P-value	MF P-value	Corrected MF P-value
glial cell development	GO:0021782	80	1.00E-12	3.00E-09	1.00E-12	3.00E-09
ensheathment of neurons	GO:0007272	80	1.00E-12	5.00E-09	1.00E-12	5.00E-09
axon ensheathment	GO:0008366	80	1.00E-12	5.00E-09	1.00E-12	5.00E-09
Oligodendrocyte Precursor cells						
ceramide catabolic process	GO:0046514	10	1.00E-04	0.05	1.00E-12	9.00E-10
glycosphingolipid catabolic process	GO:0046479	10	1.00E-12	9.00E-10	1.00E-12	1.00E-09
lysosomal lumen	GO:0043202	90	1.00E-12	1.00E-09	1.00E-12	1.00E-09
glycolipid catabolic process	GO:0019377	10	1.00E-12	2.00E-09	1.00E-12	2.00E-09
oligosaccharide catabolic process	GO:0009313	10	1.00E-12	3.00E-09	1.00E-12	3.00E-09
ganglioside metabolic process	GO:0001573	20	1.00E-12	5.00E-09	1.00E-12	5.00E-09
VIPRelnG30						
---	---	---	---	---	---	---

Top 10 significant GO are presented. Gene number – number of genes annotated with the term and included in the analysis; MF – multifunctionality-adjusted

Table S7 – Variance explained by the first three PCs based on fsPV and astrocyte marker genes

	fsPV			Astrocyte		
Dataset	PC1	PC2	PC3	PC1	PC2	PC3
GSE35978	53.80%	17.50%	10.30%	45.70%	10.20%	8.83%
GSE53987	46.20%	21.90%	12.50%	44.00%	8.48%	4.70%
McLeanCortex	52.70%	20.70%	13%	47.10%	13%	5.53%
Study1AltarA	55.30%	24.30%	16%	40.10%	13.80%	7.45%
Study3Bahn	54.90%	18%	13.80%	40.90%	11.70%	5.98%
Study5Dobrin	63.80%	24%	9.20%	42.20%	10.20%	8.10%
Study7Kato	51%	17.50%	12.30%	49.20%	8.39%	4.46%
Study2AltarC	45.60%	25.30%	16.30%	45.50%	14.50%	8.78%
Study4Chen	62.30%	21.10%	11.90%	52.00%	9.17%	6.75%
GSE21138	34.40%	20.70%	15%	57%	7 %	5.26%
GSE80655	36.20%	20%	12.00%	58.50%	10.30%	4.41%
PsychEncode	36.16%	20.27%	16.28%	32.06%	10.26%	8.26%
CommonMind	59.21%	16.81%	9.33%	29.13%	16.58%	10.26%

Table S8 – Linear model results for astrocyte and fsPV MGPs in dataset GSE53987 with or without including RIN as a covariate.

Astrocytes (not RIN adjusted)					fsPV (not RIN adjusted)			
	Estimate	Std. Error	t value	Pr(>|t|)	Estimate	Std. Error	t value	Pr(>|t|)
BP	0.17	0.06	2.85	6.62E-03	-0.13	0.08	-1.68	1.01E-01
SCZ	0.20	0.06	3.22	2.41E-03	-0.20	0.08	-2.57	1.36E-02
pH	-0.21	0.09	-2.38	2.17E-02	0.17	0.11	1.48	1.47E-01
Age	0.00	0.00	1.62	1.13E-01	0.00	0.00	-0.62	5.42E-01
SexM	-0.04	0.05	-0.85	4.00E-01	-0.03	0.06	-0.48	6.32E-01
PMI	0.00	0.00	-0.84	4.05E-01	0.00	0.01	0.42	6.80E-01
Astrocytes (RIN adjusted)					fsPV (RIN adjusted)			
	Estimate	Std. Error	t value	Pr(>|t|)	Estimate	Std. Error	t value	Pr(>|t|)
BP	0.16	0.06	2.83	7.05E-03	-0.12	0.08	-1.61	1.15E-01
SCZ	0.18	0.06	2.98	4.79E-03	-0.19	0.08	-2.36	2.32E-02
pH	-0.18	0.09	-2.09	4.26E-02	0.14	0.11	1.25	2.18E-01
Age	0.00	0.00	1.42	1.62E-01	0.00	0.00	-0.46	6.47E-01
SexM	-0.10	0.06	-1.72	9.28E-02	0.01	0.07	0.17	8.68E-01
PMI	0.00	0.00	-1.15	2.57E-01	0.00	0.01	0.60	5.53E-01
RIN	-0.10	0.05	-2.04	4.79E-02	0.08	0.06	1.30	2.00E-01

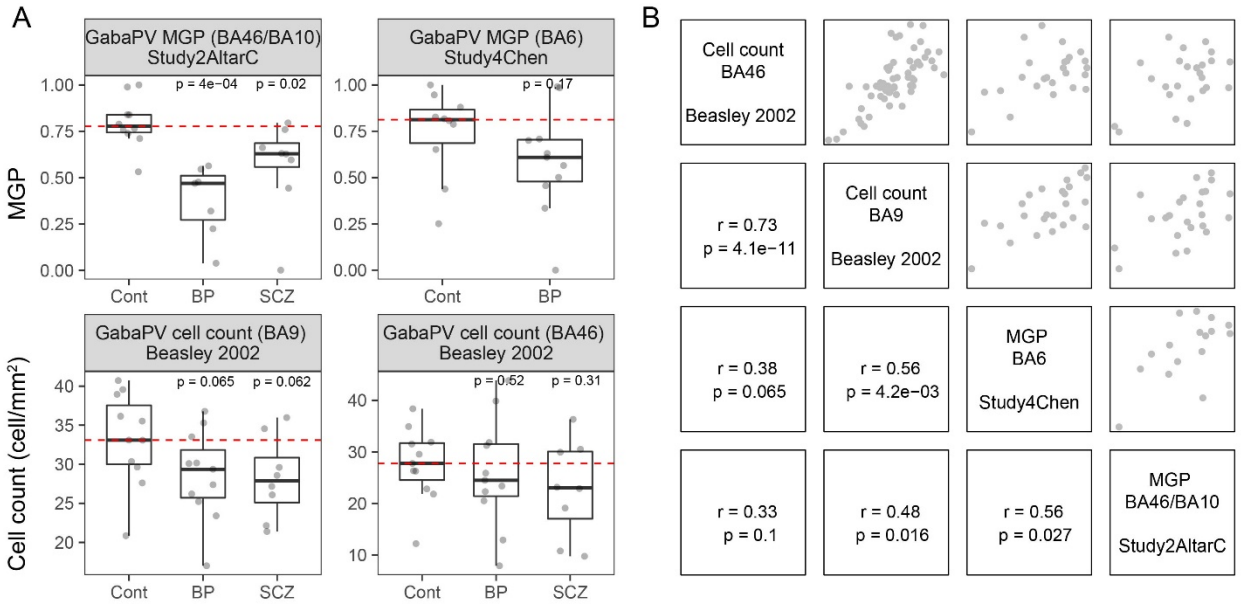

Figure S1. fsPV MGPs are consistent with cell count data from the same cohort of subjects. **A.** Changes in fsPV MGPs from two bulk tissue datasets of psychiatric patients analysed in the current study (upper pannel) and Gaba PV cell counts from Beasley et al. (25) based on the same cohort of subjects. For visualization purposes cell counts were normalized to 0-1 range. Both MGPs and cell counts indicate decrease in fsPV cells in BP and SCZ subjects.
BA – Brodmann area; Cont – controls; BP – bipolar disorder; SCZ - schizophrenia. Red dashed line indicate the median normalized cell count/MGP in the control group. **B.** Correlations of fsPV cell counts and MGPs of the same subjects. Lower triangle shows the Spearman correlation values between any two indicated datasets, upper triangle show the correlation plots. Each dot represents a subject.

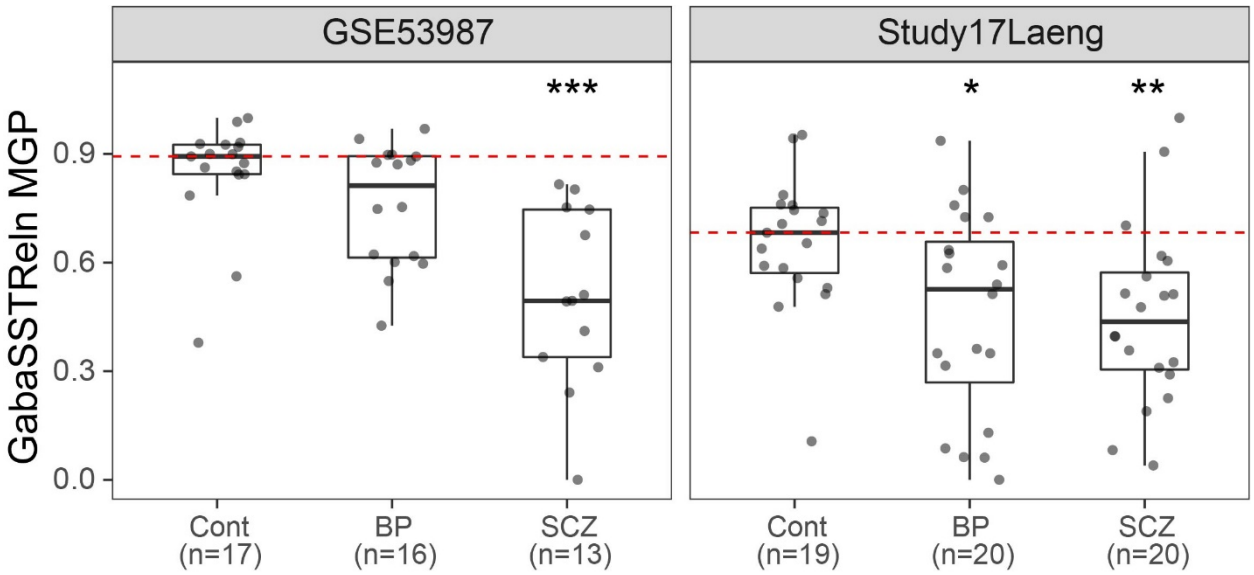

Figure S2. Consistent decrease in GabaSSTReln MGPs in hippocampal samples of subjects with bipolar disorder and schizophrenia.
MGP analysis of two datasets of hippocampal samples indicated consistent decrease in GabaSSTReln MGPs in psychiatric subjects, with stronger effect observed in subjects with schizophrenia. The two datasets contain data from different cohorts of subjects (GSE53987 – ACOME, Study17Laeng – SMRI_A). Each point represents a single sample. Red dashed line indicates the median MGP is the control group. Wilcoxon signed ranked test for comparison to controls: * - $p < 0.05$, ** - $p < 0.01$, *** - $p < 0.001$.

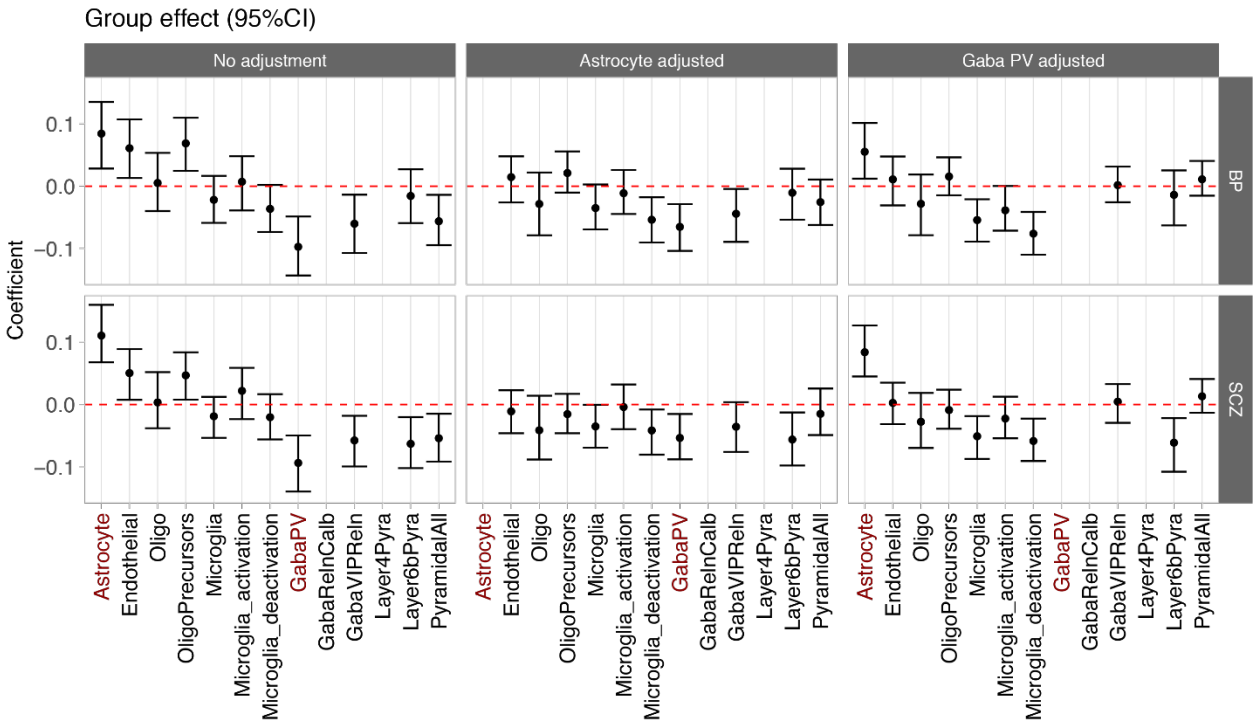

Figure S3. Changes in astrocyte and fsPV MGPs remain significant after adjusting for fsPV and astrocyte MGPs, respectively.
Coefficients and 95%CI for bipolar disorder (upper panel) and schizophrenia (lower panel) based on linear mixed model including 11 cortical datasets of psychiatric patients and controls. Left plots represent coefficients for each cell type MGP (indicated on x axis) unadjusted to astrocyte or fsPV MGP (identical to figure 1B). Middle plots show coefficients based on mixed models adjusted for astrocyte MGP. Right plots show coefficients based on mixed models adjusted for fsPV MGP. Red dashed line indicates zero. Coefficients above zero indicate increased MGP in psychiatric group, coefficients below zero indicate decrease in MGP in psychiatric group.

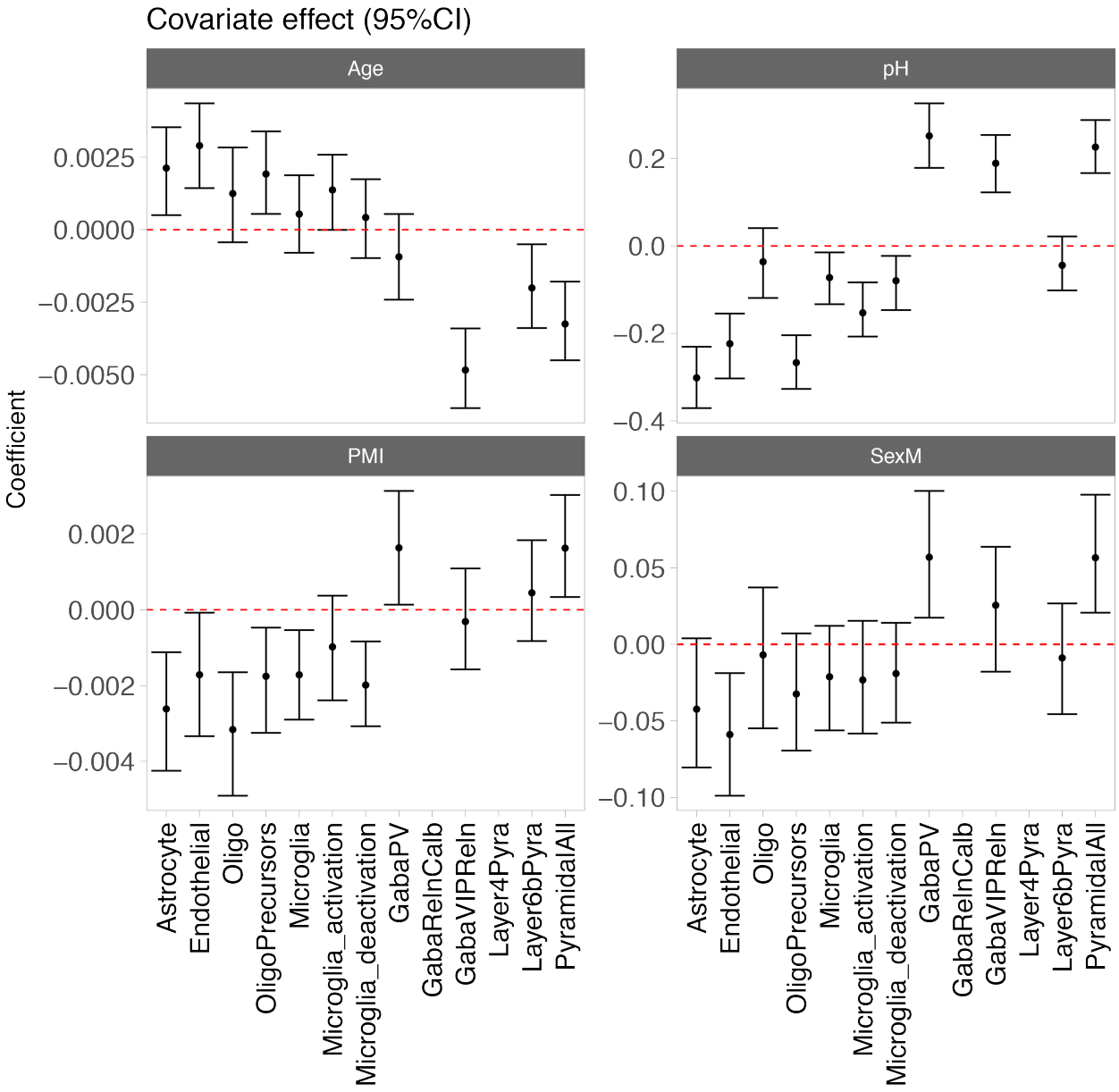

Figure S4. Covariate coefficients

Coefficients and 95%CI for age, pH, PMI and sex, calculated for each cell type based on linear mixed model including 11 cortical datasets of psychiatric patients. The coefficients are per one pH unit, one PMI unit, one year and being a male. GabaRelnCalb and Layer4Pyra cells MGPs were estimated for less than 50% of the datasets and thus were not included in mixed model analysis.

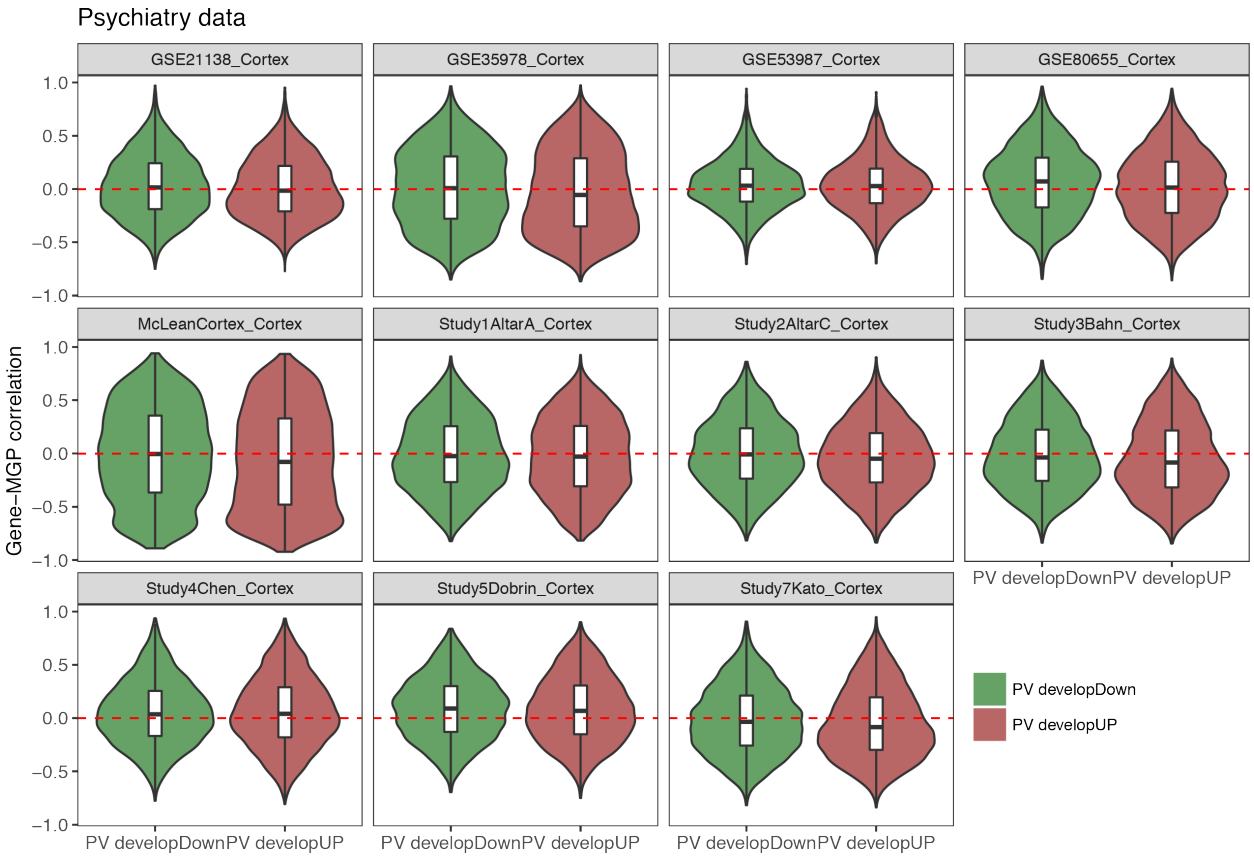

Figure S5. Genes down- and up-regulated during fsPV cell maturation are not correlated with MGPs of these cells in individual psychiatry datasets.
Spearman correlation of genes downregulated (green) and upregulated (red) during maturation of fsPV cells, with fsPV MGPs in psychiatry datasets analysed in this study. In each of the 10 studies analyzed, gene-MGP correlation of both down- and up regulated genes is equally distributed around 0. Red dashed line represents correlation value of 0.

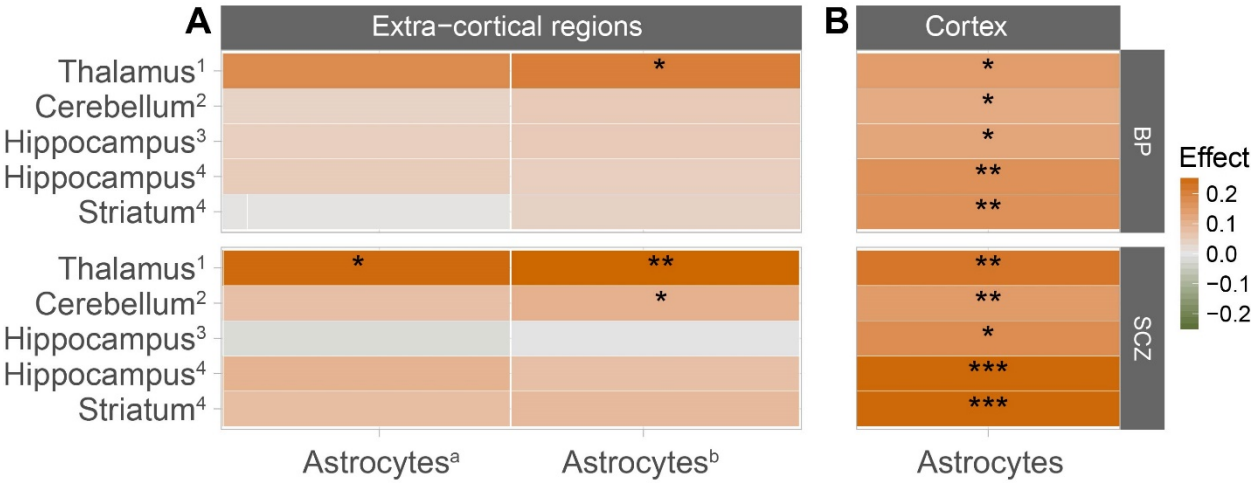

Figure S6. Changes in astrocyte MGPs in bipolar disorder and schizophrenia are region specific.

Shifts in astrocyte MGPs in extra-cortical and cortical samples of subjects with bipolar disorder and schizophrenia. The extra-cortical regions are indicated on y axis. Upper script indicate the study: [1] - Study17Laeng, [2] – GSE35978, [3] - Study16Kemether, [4] – GSE53987.

A. Changes in Astrocyte MGPs in extra-cortical regions from subjects with bipolar disorder (upper panel) and schizophrenia (lower panel). Shifts in MGPs were estimated using Wilcoxon signed ranked test. The right column show the results in each of the extra-cortical datasets when MGPs were calculated bases on cortical astrocyte marker genes. Astrocytes[a] – change in MGPs calculated based on astrocyte (or Bergmann glia, in cerebellum) marker gene list from the region indicated on y axis; Astrocytes[b] – change astrocyte MGPs based on astrocyte marker gene list from cortex.

B. For direct comparison with the results in cortical samples, we calculated the shifts in astrocyte MGPs in cortical samples from the same subjects as included in the analyses of each of the extra-cortical tissues. Datasets GSE53987 and GSE35978 included both cortical and extra-cortical samples. Datasets Study16Kemether and Study17Laeng included only thalamic and hippocampal samples (respectively). Since the same subjects were analysed in dataset GSE35978, for each of the subjects included in datasets Study16Kemether and Study17Laeng we extracted the corresponding values of the cortical astrocyte MGPs from dataset GSE35978.

* - $p < 0.05$, ** - $p < 0.01$, *** - $p < 0.001$. Orange – increased MGP; green – decreased MGP. BP – bipolar disorder, SCZ – schizophrenia.

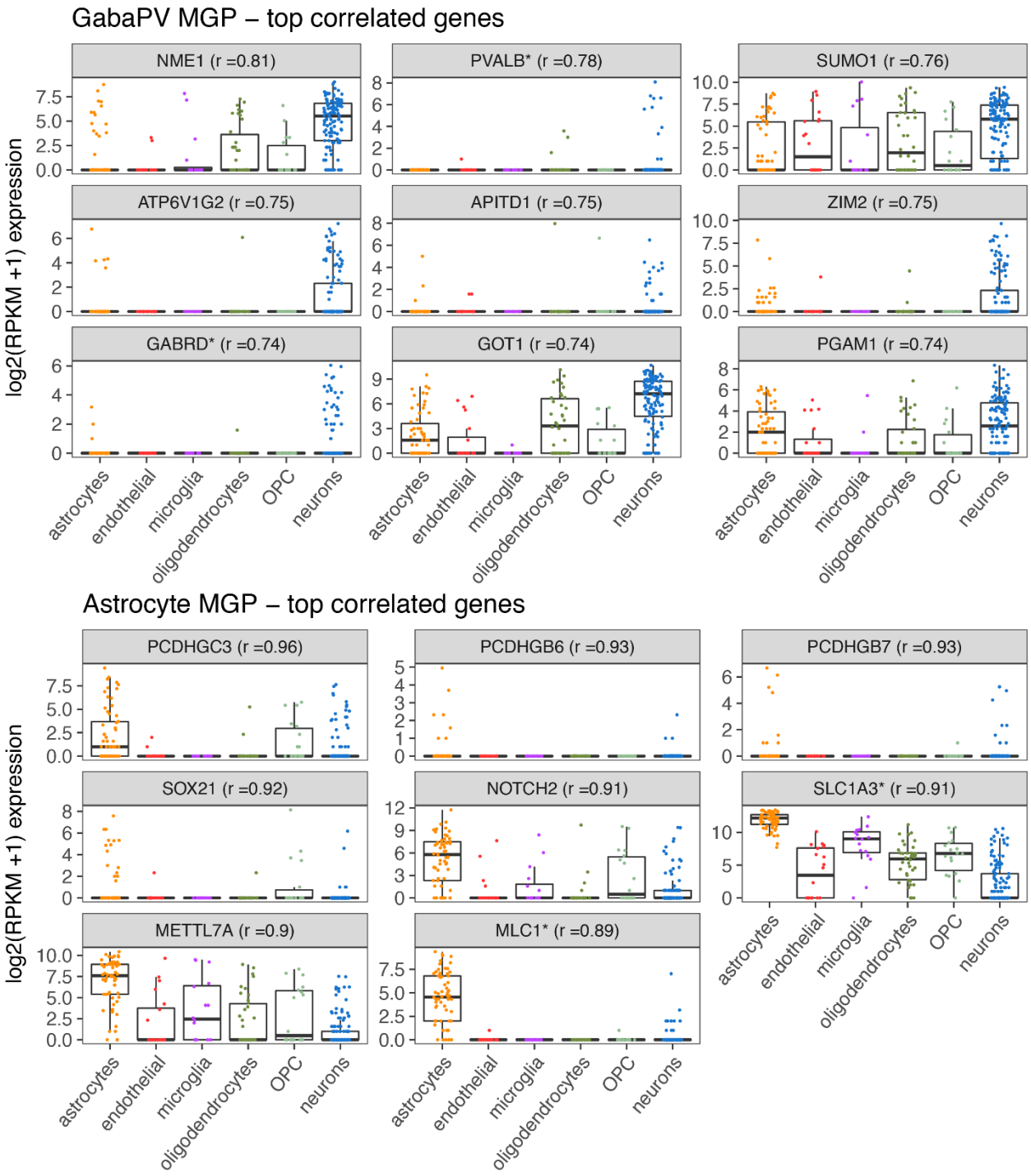

Figure S7. Top genes correlated with astrocyte and fsPV MGPs are enriched in human astrocytes and neurons, respectively.

Expression level of the top 10 genes correlated with astrocyte MGPs (A) and fsPV MGPs (B) in human single cells. Each point represents a single cell from Darmanis et al. The mean Spearman correlation (across 10 cortical psychiatry datasets) is shown in parentheses. Asterisks indicate genes considered as marker genes for the specific cell type in our analysis (and thus were included in MGP calculation).

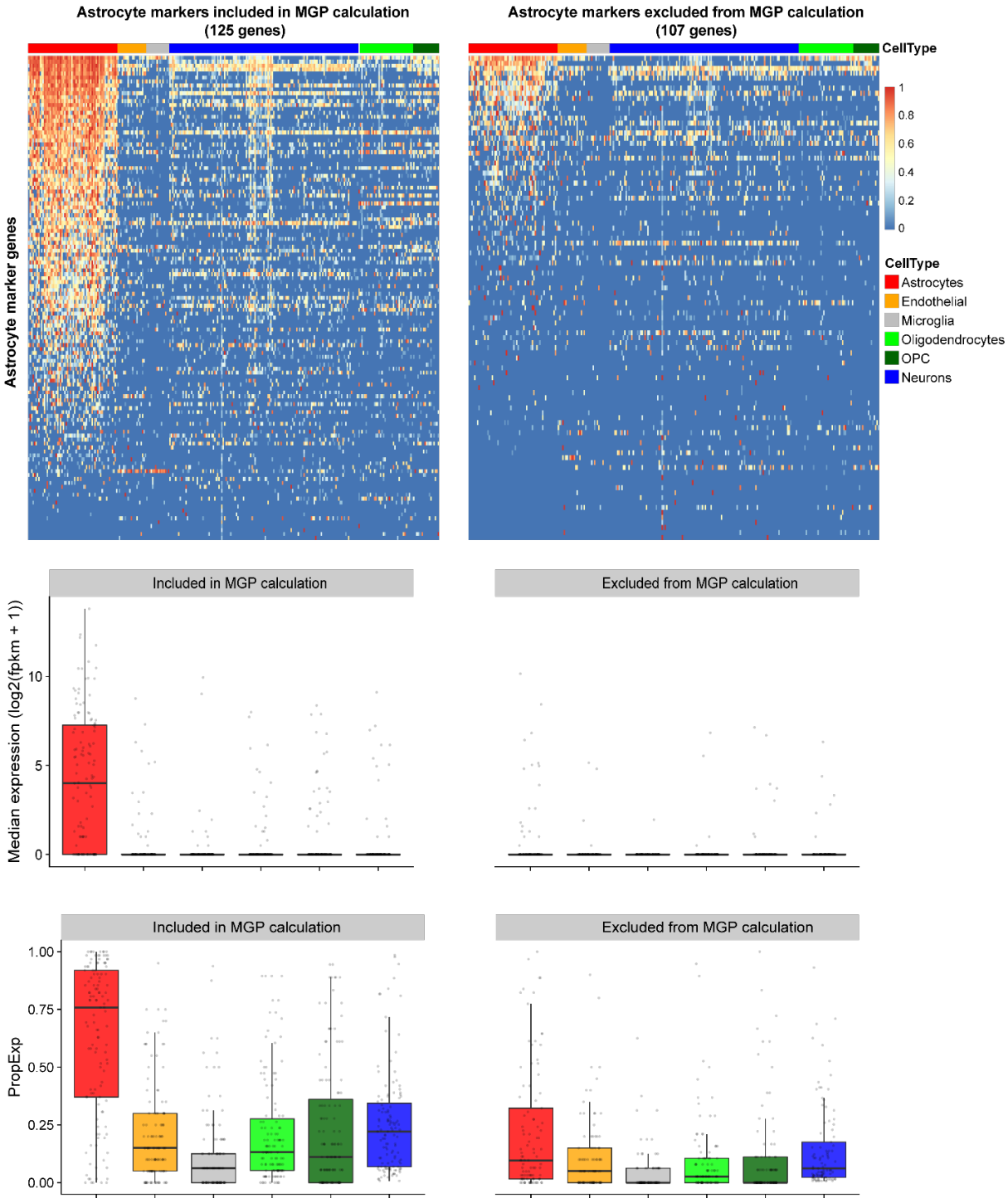

Figure S8. Expression of NeuroExpresso astrocyte markers in human single cells.
Single cell RNAseq data is from Darmanis et al.. Upper plane – expression of marker genes in individual cells, classified to cell types as per Darmanis et al.. On the left, genes that retained for MGP calculation after the exclusion steps (as described in the methods section). On the right – genes that were excluded during these steps. Expression was normalized to (0-1) range for visualization purposes. Middle plane – Median expression of the genes in all cells classified as the corresponding cell type. Lower plane – proportion of cells in each cell types expressing the gene.

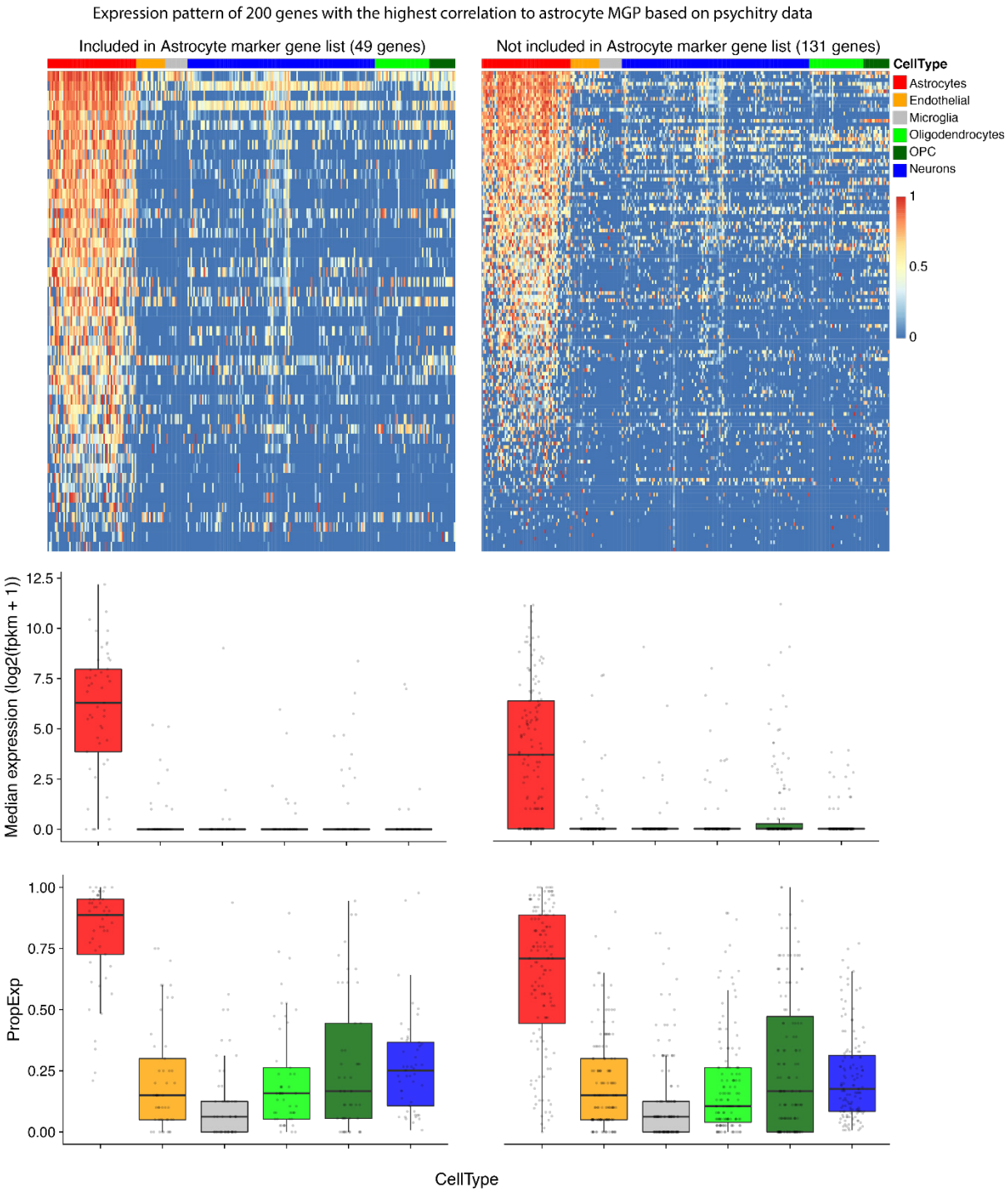

Figure S9. Genes highly correlated with astrocyte MGPs in human bulk tissue data are overexpressed in human astrocytes

Upper plane – expression of marker genes in individual cells, classified to cell types in Darmanis et al. study. On the left, genes that were considered as astrocyte markers in our analysis. On the right – genes not considered as markers. Middle and lower planes show the normalized expression level and proportion of cells expressing the genes.

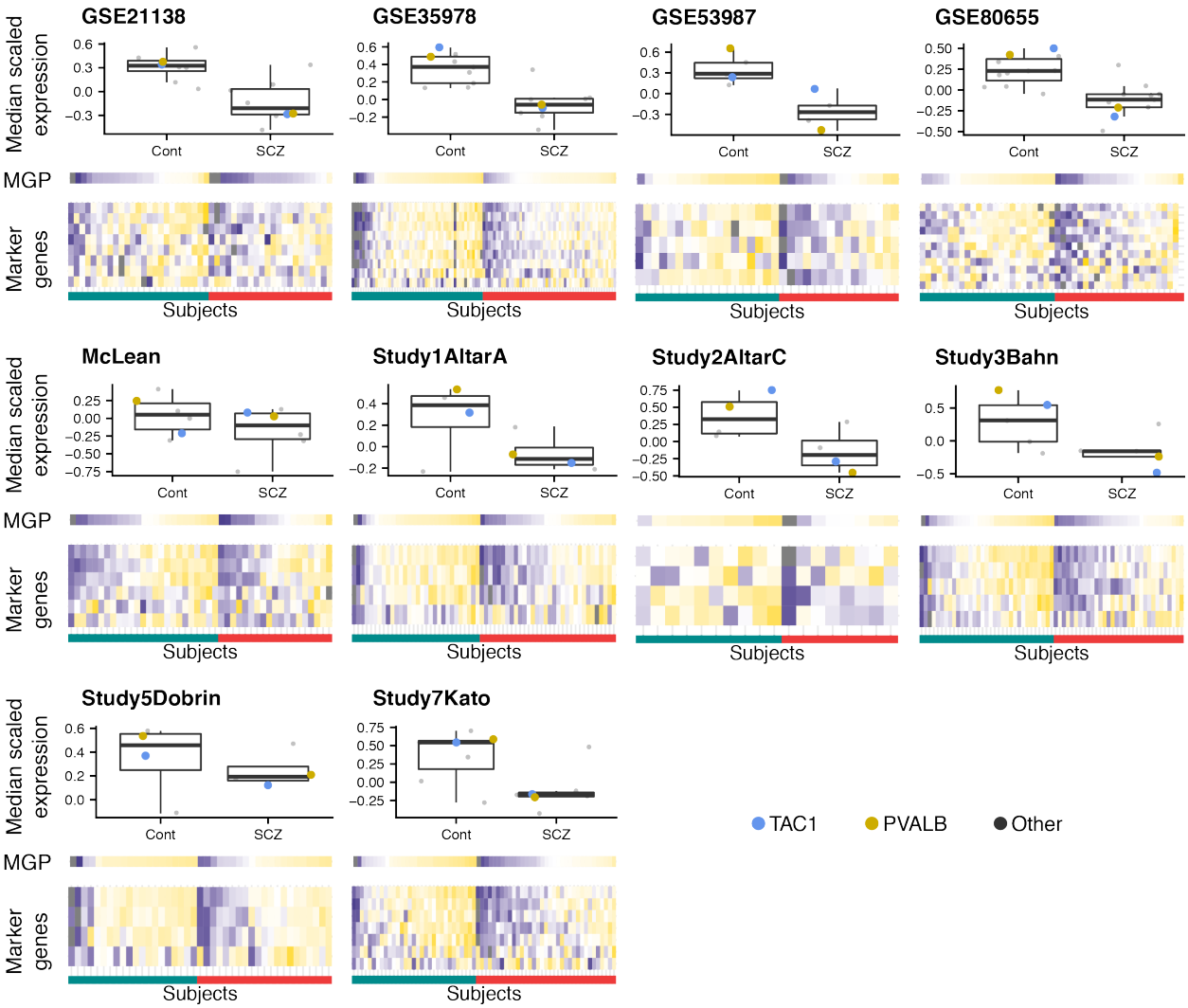

Figure S10. Expression pattern of individual fsPV markers in cortical datasets analyses. For each dataset, the upper plane shows the median scaled expression of each fsPV marker in control or schizophrenia groups. Highlighted are known fsPV markers validated to be expressed in human fsPV cells. The middle plane indicates the calculated MGP in each subject and the lower plane indicates the scaled expression of each marker gene (rows) in each subject (columns).

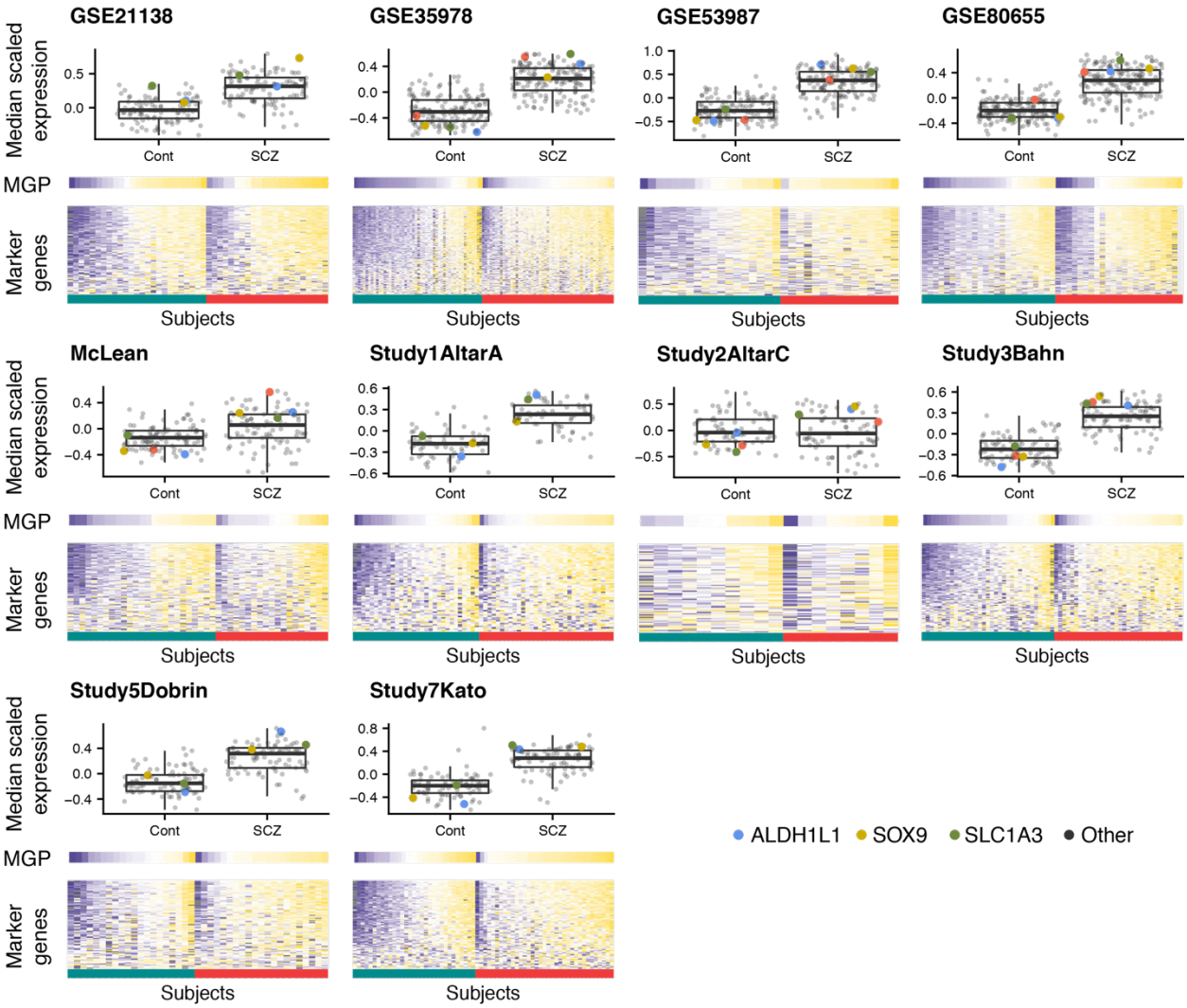

Figure S11. Expression pattern of individual astrocyte markers in cortical datasets analyses. For each dataset, the upper plane shows the median scaled expression of each astrocyte marker in control or schizophrenia groups. Highlighted are known astrocyte markers validated to be expressed in human astrocyte cells. The middle plane indicates the calculated MGP in each subject and the lower plane indicates the scaled expression of each marker gene (rows) in each subject (columns).

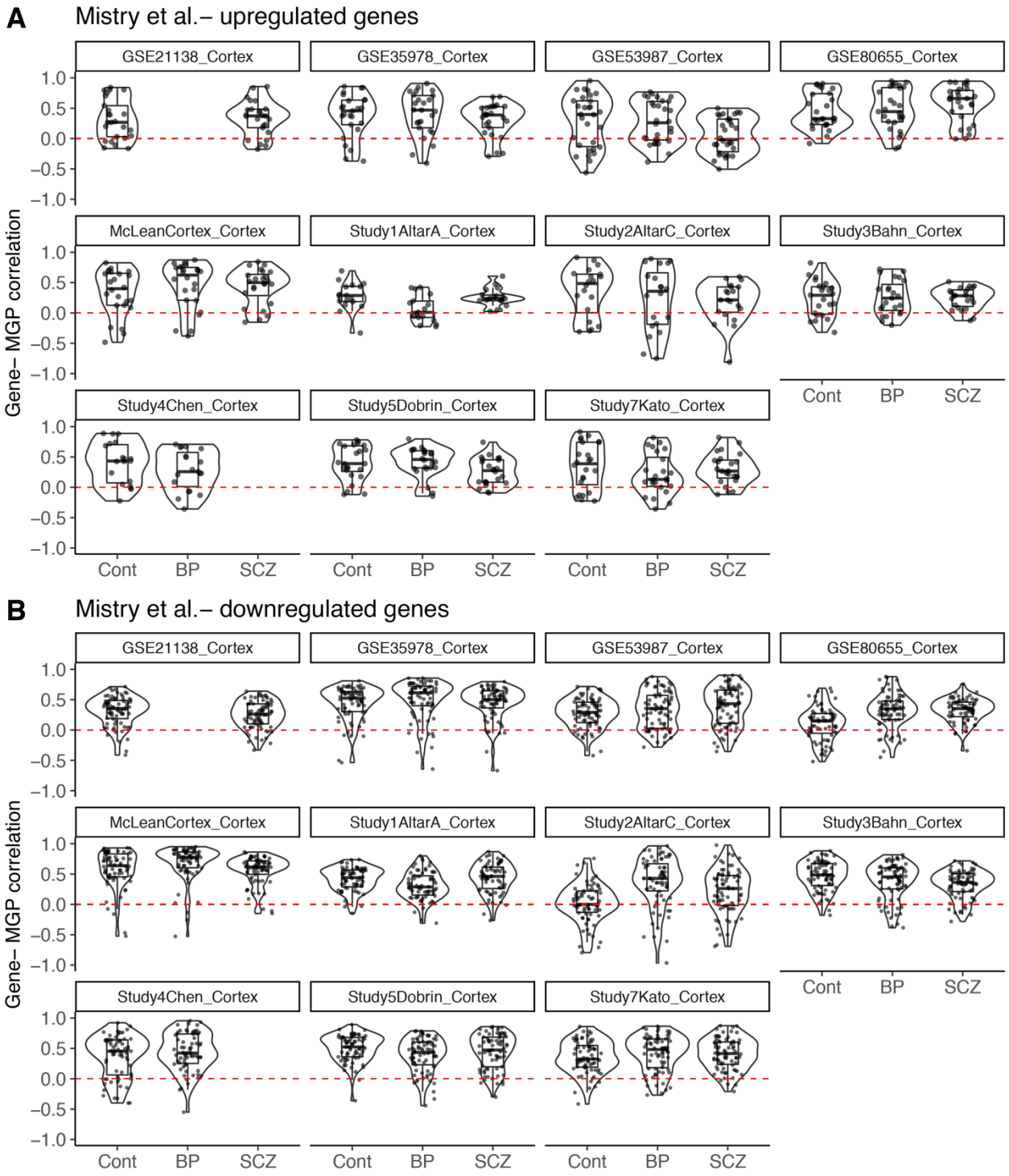

Figure S12. Correlation of genes with altered expression in schizophrenia with astrocyte and fsPV MGPs.

Spearman correlation of genes upregulated in schizophrenia based on Mistry et al. (22) with astrocyte MGPs (A) and genes downregulated in Mistry et al. with fsPV MGPs (B). Each point represents a single gene.

Figure S13. Correlation of astrocyte and fsPV markers with their corresponding MGPs. Spearman correlation of astrocyte marker genes with astrocyte MGPs (A) and fsPV marker genes with fsPV MGPs (B). Each point represents a single gene.

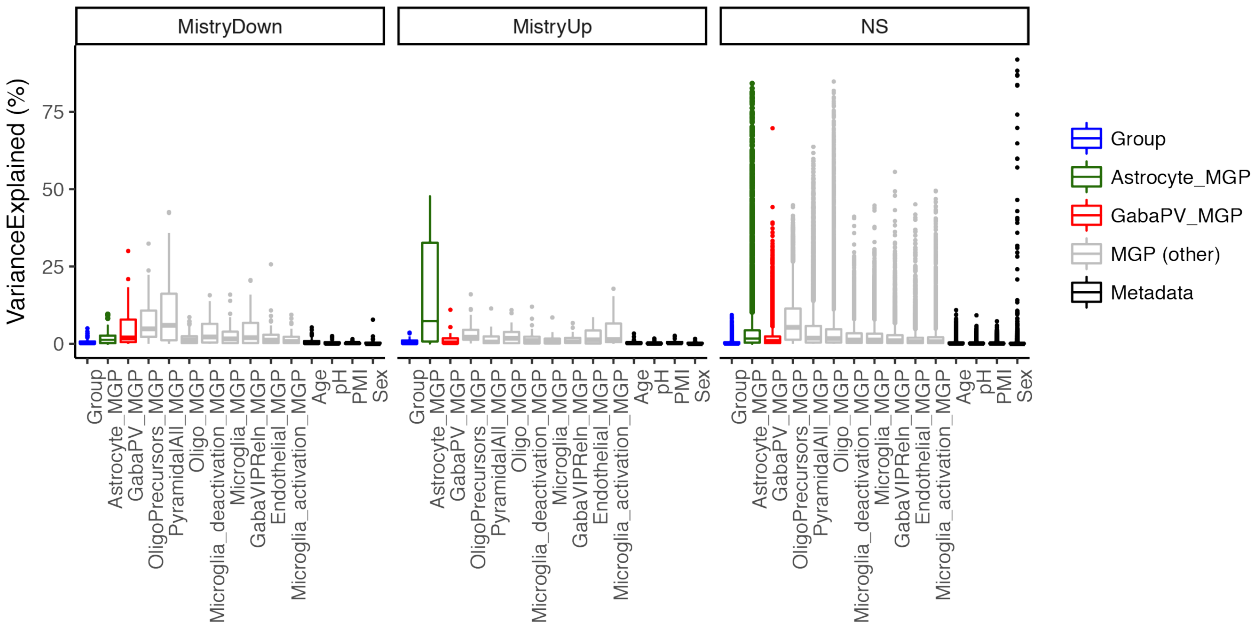

Figure S14. Variance in gene expression explained by covariates of the mixed model. Proportion of variance in the expression level of individual genes explained by different covariates of the mixed model. The proportion shown is after adjusting for the variance explained by the random effects. Each data point represents individual gene. The genes were grouped based on the meta-analysis from Mistry et al. MistryDown – genes underexpressed based on Mistry et al.; MistryUP – genes overexpressed based on Mistry et al.; NS – not significant genes based on Mistry et al.

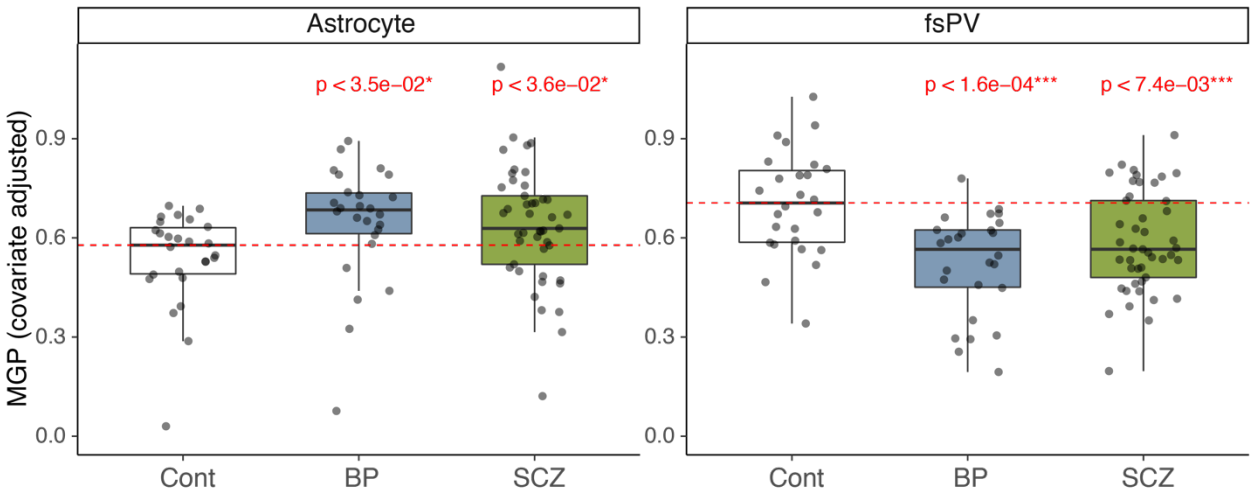

Figure S15. MGP analysis of PsychEncode data replicate the results from the original 11 datasets.

MGPs of astrocyte (left) and fsPV (right) cells in dorsolateral prefrontal cortex samples from subjects with bipolar disorder (BP) and schizophrenia (SCZ). Each point represents MGP in individual sample, after adjustment for pH, Age, PMI and RIN. The p-values are based on linear models including pH, Age, PMI and RIN as covariates. Sex was not included as a covariate in the model since the control samples included only a single female sample.

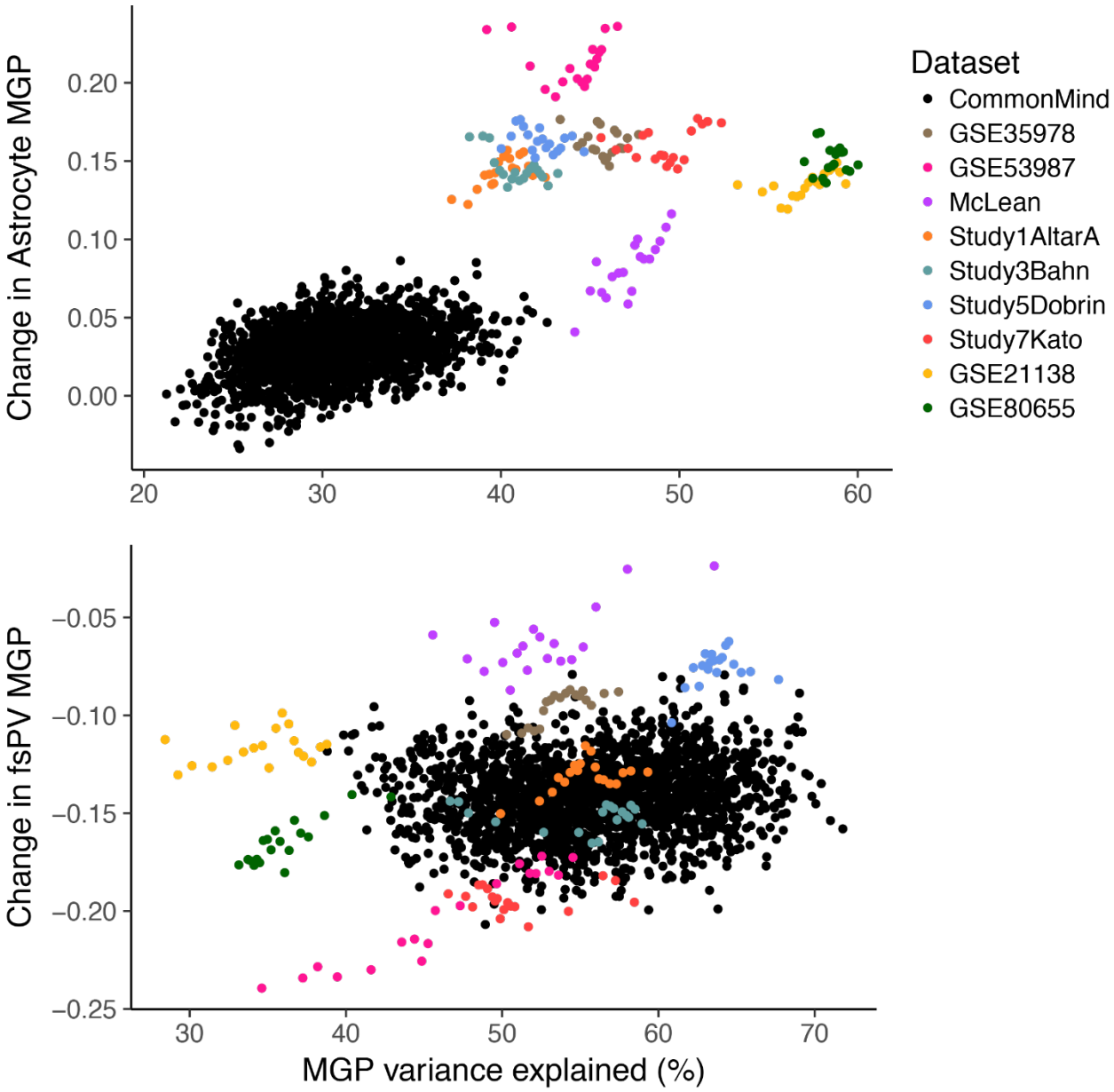

Figure S16. Variation explained by PC1 of astrocyte and fsPV marker genes (namely, MGPs) in subsamples of each dataset analyzed.

MGP analysis was repeated 100 times on subsamples of each dataset (as described in the methods section), with the exception of the CommonMind dataset where due to the large sample size, the subsampling was performed 10,000 times, subsampling 30 subjects/group. Each point represents the value of a sliding window based on 10 subsamples. X axis, indicates the variance explained by the astrocyte (upper plane) and the fsPV (lower plane) MGPs. Y axis indicates the calculated shift in astrocyte (upper plane) and fsPV (lower plane) MGP in the SCZ group (two-sided Wilcoxon's test). In the vast majority of subsamples from the CommonMind data, the variance explained by PC1 based on astrocyte genes is much lower than in subsamples of all the other 11 datasets (upper panel). In all datasets, the percentage of variance explained is positively correlated with the calculated shift in astrocyte MGP. The shifts in astrocyte and fsPV MGPs calculated for each subsampled data point are consistent with the results of the main analysis, that is, an increase in astrocyte MGPs and a decrease in fsPV MGPs in subjects with schizophrenia.

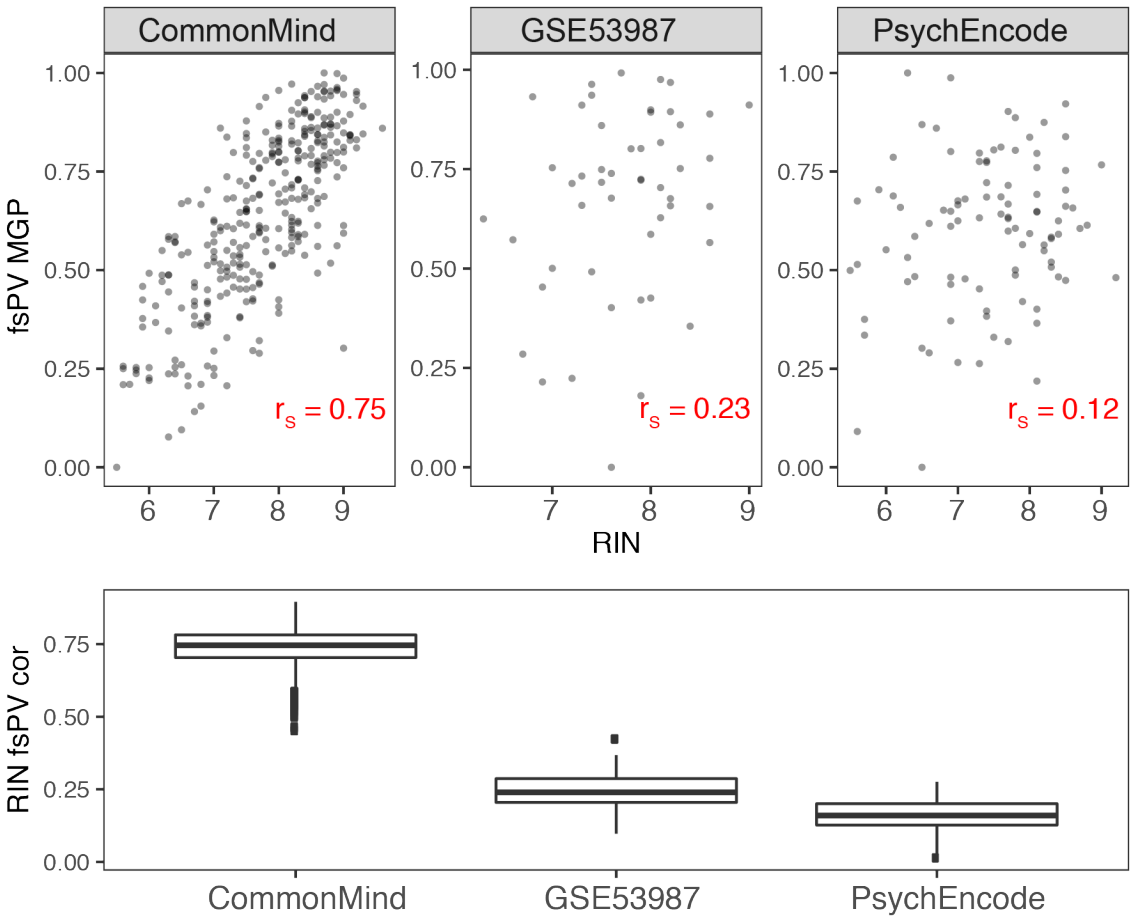

Figure S17. fsMGP are highly correlated with RIN in the CommonMind dataset.
RIN values were available for three of the datasets analysed – CommonMind, GSE53987 and PsychEncode. Upper plane shows correlation between fsMGP and RIN in the full datasets. Each point in the upper plane represents individual sample. Lower plane shows fsMGP and RIN in subsamples of the datasets (as described in methods and supplementary methods), each point represents Spearman's correlation from individual subsample of the data. fsMGP exhibit high correlation with RIN in the CommonMind dataset but not in GSE53987 or PsychEncode, indicating that in the CommonMind dataset changes in MGPs are at least partially induced by differences in RNA quality between the samples.

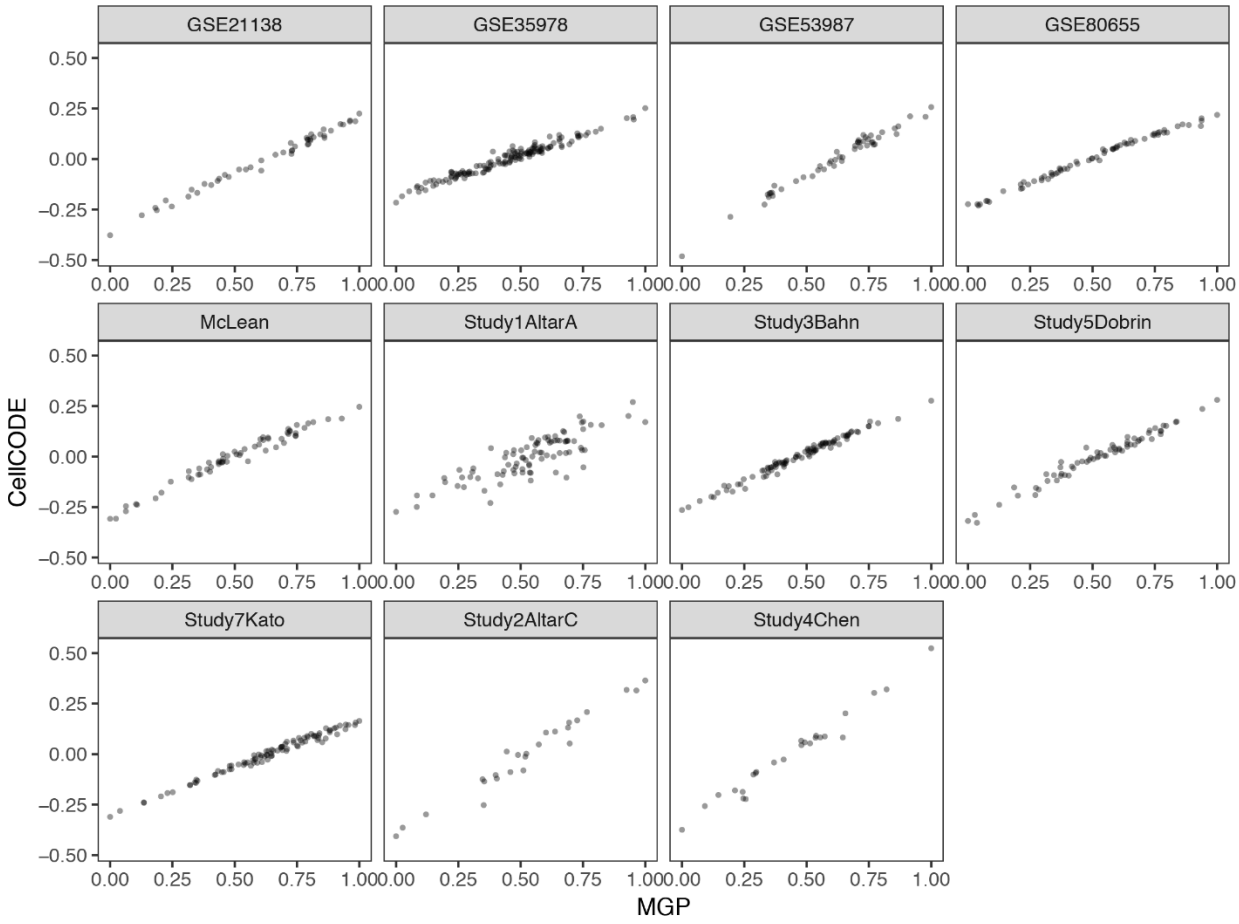

Figure S18. Correlation between estimated relative change in astrocytes using CellCODE or MGP analysis.

In each dataset, we repeated the estimation of astrocyte using CellCODE, and compared the estimates to the calculated astrocyte MGP, using our method. Each point represents individual samples. The correlation between the two methods is extremely high, indicating that the results of our analysis are not an artifact of the estimation method.

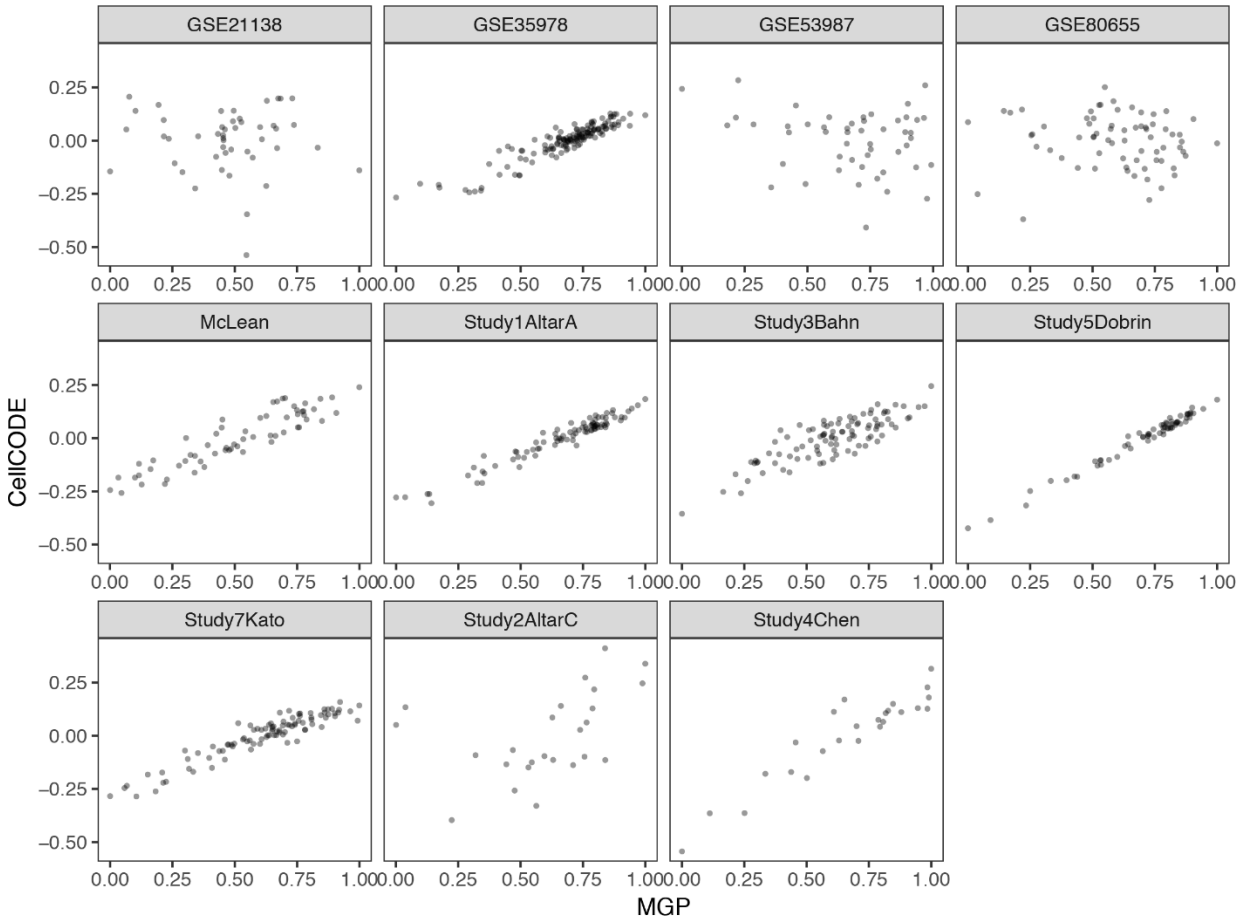

Figure S19. Correlation between estimated relative change in fsPV cells using CellCODE method or MGP analysis.
Similar to Figure S18, but for fsPV cells.

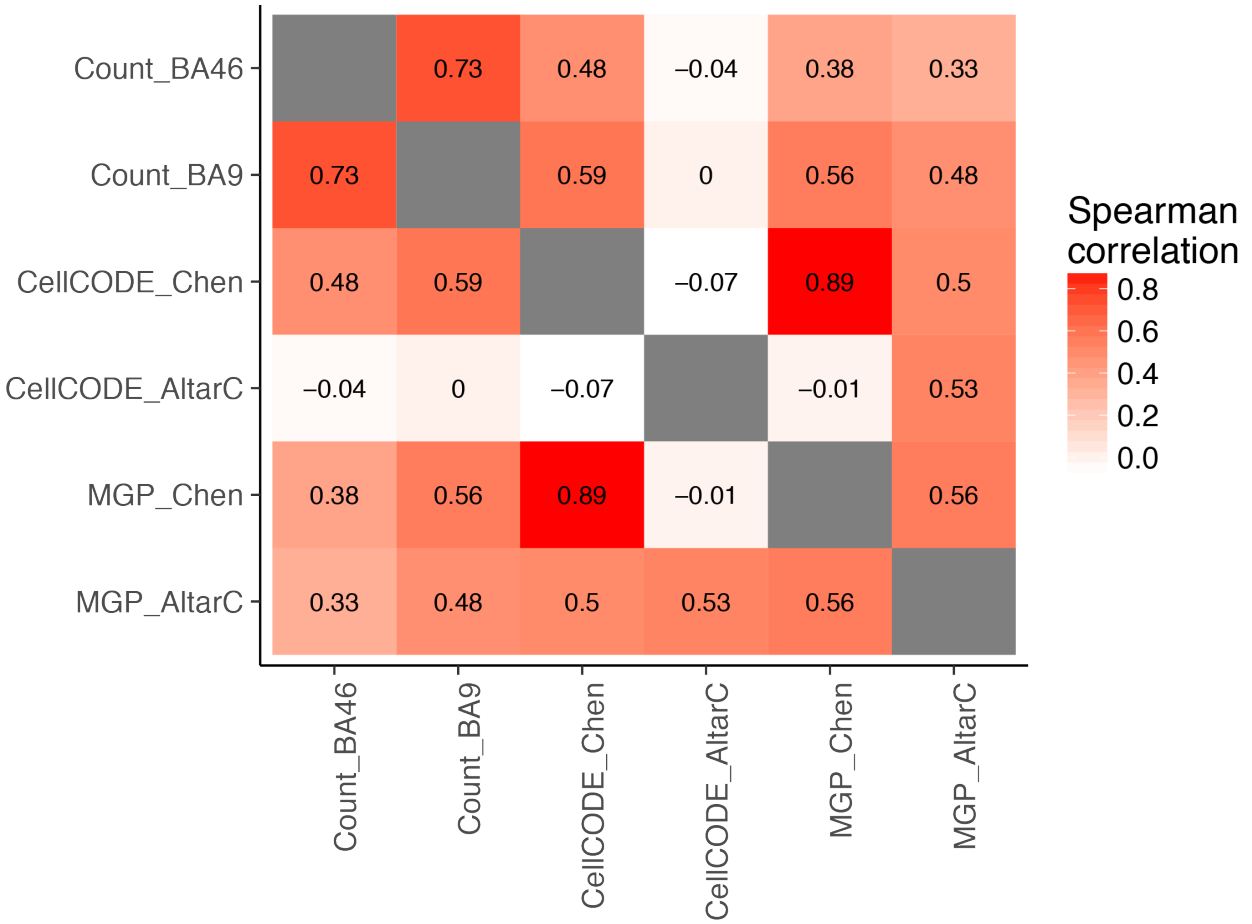

Figure S20. For datasets with low correlation between CellCODE and MGP analysis estimates, MGPs are better correlates of changes in fsPV cell counts.
As indicated in the methods sections, fsPV cell counts from two cortical regions (BA9 and BA46) were available for one cohort of subjects - Stanley Consortium, represented by two expression datasetsChen and AltarC. We thus calculated Spearman correlation between the actual cell counts and the estimates from the two datasets based on CellCODE or MGP analysis methods. Similar correlations between the estimates of the two methods and cell counts are observed for the Chen dataset, in which the correlation between the estimated of the two methods is high. In contrast, in dataset AltarC, where the correlation between the estimates of the two methods is relatively low, only MGP-based estimates are correlated with the cell counts. Moreover, CellCODE estimates are not correlated between AltarC and Chen datasets, despite the fact that they represent the same subjects.

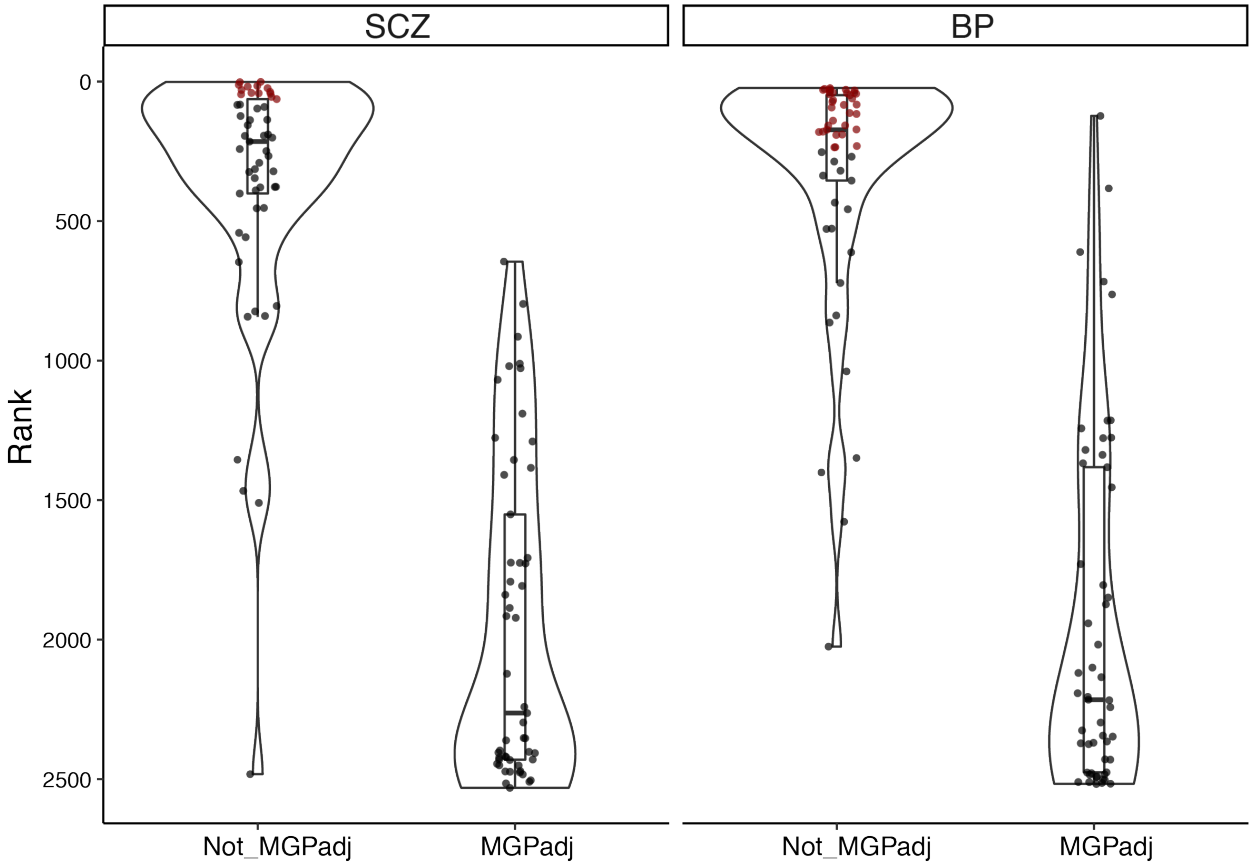

Figure S21. Enrichment of mitochondria-related terms among under-expressed genes is not observed after adjustment for MGPs.
Ranks of 53 mitochondria-related GO terms (out of 2535 terms tested) among genes under-expressed in subjects with schizophrenia (left) and bipolar disorder (right). Each point represents individual mitochondria-related GO term. Significant terms are indicated in red. Not_MGPadj – enrichment of underexpressed genes based on model not adjusted for MGPs; MGPadj - enrichment of underexpressed genes based on model adjusted for MGPs; SCZ – schizophrenia; BP – bipolar disorder.

Supplemental References

1. Veldic M, Guidotti A, Maloku E, Davis JM, Costa E (2005): In psychosis, cortical interneurons overexpress DNA-methyltransferase 1. *Proc Natl Acad Sci U S A*. 102: 2152–2157.

2. Carvalho BS, Irizarry RA (2010): A framework for oligonucleotide microarray preprocessing. *Bioinforma Oxf Engl*. 26: 2363–2367.

3. Zoubarev A, Hamer KM, Keshav KD, McCarthy EL, Santos JRC, Van Rossum T, *et al.* (2012): Gemma: a resource for the reuse, sharing and meta-analysis of expression profiling data. *Bioinforma Oxf Engl*. 28: 2272–2273.

4. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD (2012): The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinforma Oxf Engl*. 28: 882–883.

5. Toker L, Feng M, Pavlidis P (2016): Whose sample is it anyway? Widespread misannotation of samples in transcriptomics studies. *F1000Research*. 5. doi: 10.12688/f1000research.9471.2.

6. Mancarci BO, Toker L, Tripathy SJ, Li B, Rocco B, Sibille E, Pavlidis P (2017): Cross-Laboratory Analysis of Brain Cell Type Transcriptomes with Applications to Interpretation of Bulk Tissue Data. *eNeuro*. ENEURO.0212-17.2017.

7. NCBI Resource Coordinators (2017): Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 45: D12–D17.

8. Chikina M, Zaslavsky E, Sealfon SC (2015): CellCODE: A robust latent variable approach to differential expression analysis for heterogeneous cell populations. *Bioinformatics*. btv015.

9. Mistry M, Gillis J, Pavlidis P (2013): Genome-wide expression profiling of schizophrenia using a large combined cohort. *Mol Psychiatry*. 18: 215–225.

10. Okaty BW, Miller MN, Sugino K, Hempel CM, Nelson SB (2009): Transcriptional and electrophysiological maturation of neocortical fast-spiking GABAergic interneurons. *J Neurosci Off J Soc Neurosci*. 29: 7040–7052.

11. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, *et al.* (2015): A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci*. 201507125.

12. Lee HK, Braynen W, Keshav K, Pavlidis P (2005): ErmineJ: tool for functional analysis of gene expression data sets. *BMC Bioinformatics*. 6: 269.

13. Gillis J, Mistry M, Pavlidis P (2010): Gene function analysis in complex data sets using ErmineJ. *Nat Protoc*. 5: 1148–1159.

14. Jaffe AE, Tao R, Norris AL, Kealhofer M, Nellore A, Shin JH, *et al.* (2017): qSVA framework for RNA quality correction in differential expression analysis. *Proc Natl Acad Sci*. 201617384.

15. Loerch PM, Lu T, Dakin KA, Vann JM, Isaacs A, Geula C, *et al.* (2008): Evolution of the Aging Brain Transcriptome and Synaptic Regulation. *PLOS ONE*. 3: e3329.

16. French L, Ma T, Oh H, Tseng GC, Sibille E (2017): Age-Related Gene Expression in the Frontal Cortex Suggests Synaptic Function Changes in Specific Inhibitory Neuron Subtypes. *Front Aging Neurosci*. 9. doi: 10.3389/fnagi.2017.00162.

17. Damadzic R, Bigelow LB, Krimer LS, Goldenson DA, Saunders RC, Kleinman JE, Herman MM (2001): A quantitative immunohistochemical study of astrocytes in the entorhinal cortex in schizophrenia, bipolar disorder and major depression: absence of significant astrocytosis. *Brain Res Bull*, Neuropathology of Severe Mental Illness: Studies from the Stanley Foundation Neuropathology Consortium. 55: 611–618.

18. Mouton PR, Long JM, Lei D-L, Howard V, Jucker M, Calhoun ME, Ingram DK (2002): Age and gender effects on microglia and astrocyte numbers in brains of mice. *Brain Res*. 956: 30–35.

19. Byne W, Tatusov A, Yiannoulos G, Marcus S (2008): Effects of Mental Illness and Aging in Two Thalamic Nuclei. *Schizophr Res*. 106: 172–181.

20. Vostrikov V, Uranova N (2011): Age-Related Increase in the Number of Oligodendrocytes Is Dysregulated in Schizophrenia and Mood Disorders. *Schizophr Res Treat*. 2011: e174689.

21. Peters A, Sethares C (2004): Oligodendrocytes, their Progenitors and other Neuroglial Cells in the Aging Primate Cerebral Cortex. *Cereb Cortex*. 14: 995–1007.

22. Schuitemaker A, van der Doef TF, Boellaard R, van der Flier WM, Yaqub M, Windhorst AD, *et al.* (2012): Microglial activation in healthy aging. *Neurobiol Aging*. 33: 1067–1072.

23. Pakkenberg B, Gundersen HJ (1997): Neocortical neuron number in humans: effect of sex and age. *J Comp Neurol*. 384: 312–320.

24. Cha CI, Lee YI, Lee EY, Park KH, Baik SH (1997): Age-related changes of VIP, NPY and somatostatin-immunoreactive neurons in the cerebral cortex of aged rats. *Brain Res*. 753: 235–244.

25. Beasley CL, Zhang ZJ, Patten I, Reynolds GP (2002): Selective deficits in prefrontal cortical GABAergic neurons in schizophrenia defined by the presence of calcium-binding proteins. *Biol Psychiatry*. 52: 708–715.