Science Advances
AAAS

Supplementary Materials for

Brain cell-type shifts in Alzheimer's disease, autism, and schizophrenia interrogated using methylomics and genetics

Chloe X. Yap *et al.*

Corresponding author: Chloe X. Yap, chloe.yap@uq.net.au; Michael J. Gandal, michael.gandal@pennmedicine.upenn.edu

Sci. Adv. **10**, eadn7655 (2024)
DOI: 10.1126/sciadv.adn7655

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

Supplementary Note 1: Choice of deconvolution approach

In the absence of brain benchmarking datasets with both directly measured and deconvolution-estimated cell type proportions (CTPs) from matched brain samples, we extensively compared deconvolution pipelines to identify the optimal deconvolution pipeline. This Supplementary Note describes this process.

An overview of the deconvolution pipeline is provided in **Figure 1a** and **Figure 2**. We compared alternative methods at each of the key reference-based deconvolution steps of:

1. Reference data selection
2. Cell-type marker identification
3. Deconvolution algorithm selection
4. Input bulk data QC

An overview of the benchmarking and validation analyses are listed here and discussed in greater detail below:

1. Number of cell-types in the reference – a higher number was prioritised
2. Comparison to different reference cell-type profiles (including reference dataset and marker selection approaches) including comparison of Pearson's correlations (**fig. S5-S6**). For reference panels with fewer cell-types than our primary deconvolution pipeline, we summed cell-types: excitatory and inhibitory neurons to neuronal/NeuN+ CTPs; astrocytes, endothelial cells, microglia, oligodendrocytes, OPCs to glial/NeuN- CTPs.
3. Benchmarking pipelines by comparing to deconvolution of "pure" cell-type populations from FACS (*63*) (**fig. S8**).
4. Comparison to reference-free methods: cell-type composition tends to capture the majority of variance in bulk omics datasets. This is the premise on which reference-free methods have been implemented. We compared smartSVA (*13*) and MethylNet (*30*) (**fig. S9-10**).
5. Overall quality of the deconvolution (Houseman only): we calculated the CETGYO (*85*) score (representing root mean squared error; RMSE) and used a threshold of 0.1 as an indicator of good deconvolution quality (**fig. S11**).
6. Comparison to representations of CTPs from different omics modalities for overlapping samples: for UCLA_ASD (n=49) and LIBD (n=398), we compared the bulk methylation deconvolved CTPs to WGCNA eigengenes from modules representing equivalent cell-types (*69*). For ROSMAP, we used immunohistochemistry (image-based) CTP deconvolution for n=49 matching samples (*22*). We accounted for dissection differences by removing the oligodendrocyte proportion and rescaling the neuronal cell populations to sum to 100% (**fig. S12-S13**).
7. Prior expectations of CTP estimates. For example, we would expect relatively large proportions of neurons and oligodendrocytes, and relatively small proportions of endothelial cells, microglia and OPCs (**Figure 3a**, **fig. S14**).

Reference-based pipelines alternatives

Reference data and alternatives

We compared five reference dataset sources:

1. Single-cell methylation sequencing reference from Luo et al. 2023, involving 15,030 cells from 2 male donors in their 20s. This QC-ed reference dataset includes pseudobulk cell-type methylation profiles for 7 cell-types: excitatory neurons, inhibitory neurons, astrocytes, endothelial cells, microglia, oligodendrocytes, and oligodendrocyte precursor cells (see **Methods** for further details).
2. Whole genome bisulfite sequencing (WGBS) of sorted cell populations from a recently published methylation atlas (used in the "WGBS/FACS" pipeline), with cell-type markers already identified by Loyfer et al. (*27*). This provides 5 cell-types (Neuron, Oligo, Endo, Micro and Other). Note, however, that the endothelial cell reference was not derived from brain samples.
3. Methylation reference based on RNA marker genes from single-cell RNAseq data (used in the "EpiSCORE RNA-based pipeline"), with cell-type markers already identified by Zhu et al. (*28*). This provides 6 cell-types (Neuron, Astro, Endo, Micro, Oligo, OPC).
4. Aggregated array-based methylation profiling of sorted cell populations collated from multiple datasets. This reference had methylation profiles for 7 cell-types: glutamatergic neurons (*20*) (n=5, GSE50853), GABAergic neurons (*20*) (n=5, GSE50853), astrocytes (*81*) (n=2, GSE40699 and GSE92462), endothelial cells (*82*) (n=3, GSE137830), oligodendrocytes (n=45 (*63*)), oligodendrocyte precursor cells (*81*) (n=1, GSE92462). In the absence of publicly available microglia datasets, we used monocytes from the FlowSorted.Blood.450K data available on Bioconductor (*84*) as a proxy, as they share a developmental lineage.
5. Array-based methylation profiling of NeuN+/- cells (*80*). This dataset is the most frequently used in brain cell-type deconvolution, but is limited by its low cell-type resolution. Therefore, we used this NeuN+/- reference dataset for benchmarking. For this, we compared the sums of the higher-resolution deconvolutions to the corresponding NeuN+/- proportion (e.g., excitatory neurons and inhibitory neurons should sum up to the NeuN+ proportion).

Of the two reference datasets with 7 cell-types, we prioritised the single-cell methylation sequencing reference over the aggregated array-based reference for the following reasons:

- The sequencing-based dataset comes from a single dataset, which mitigates batch effects. In contrast, the array-based dataset was aggregated from multiple datasets, introducing possible batch effects between studies.
- Single-cell sequencing with subsequent clustering across millions of methylation sites (including non-CpG sites which have previously been shown to be high-specificity markers of neuronal populations (*104*)), provides greater coverage of the epigenome and allows for unsupervised identification of cellular identity.

One possible limitation of using sequencing-based references to deconvolve bulk array data is that sequencing and array data have different distributions of methylation beta values (*17*). This motivated the introduction of the methylCC (*17*) algorithm (accounting for this cross-platform effects using random effects), which we also tested in our exploratory analyses (see "Alternative reference-based deconvolution algorithms" below).

Marker selection approaches

For the ***Luo et al. 2023 single-cell methylation sequencing reference dataset*** *(24)*, we experimented with two different marker probe selection approaches.

1. "Extremes" selection: We identified cell-type markers based on how extreme their methylation profiles were relative to the other cell-types. For this, we converted the filtered methylation sequencing counts into beta-values (methylated read counts/total read counts). Marker sites were identified where one cell-type had beta≤0.4 whereas all other cell-types had beta≥0.6; or vice versa. We selected this 0.4/0.6 split as this provided at least ~100 marker sites for each of the cell-types. We were left with n=983 excitatory neurons markers, n=99 inhibitory neuron markers, n=499 astrocyte markers, n=682 endothelial cell markers, n=763 microglia markers, n=423 oligodendrocyte markers, and n=838 OPC markers. The majority of marker sites were down-methylated for all cell-types except for inhibitory neurons (majority were up-methylated) which is consistent with previous findings (*20*). The inhibitory neuron reference had markedly fewer marker sites than other cell-type references, but we found that relaxing criteria to increase the number of marker sites specifically for the inhibitory neuron population destabilised the deconvolution. We explored other cut-offs, but found this either resulted in too few markers or poor distinction between cell-type profiles.
2. Chi-squared selection: For this approach, we first identified sites where there were uneven proportions of methylated reads across cell-types. This was achieved using a chi-squared test on a contingency table with 7 columns (one per cell-type) and 2 rows (one each for methylated and unmethylated reads), selecting sites with chi-squared p<0.001 after Bonferroni correction. Then, each marker's cell-type specificity was assigned where one cell-type explained >50% of that marker's chi-squared statistic. Finally, for each cell-type, the top 200 markers were taken, based on the greatest percentage contributions to the chi-squared statistic. For input into methylCC, we generated a matrix from these marker probes (one column per cell-type, and with one row per selected probe), where cells were populated with a 0 for probes that were significantly down-methylated in that cell-type, and populated with a 1 for probes that were significantly up-methylated in that cell-type.

For both of the ***two array-based references***, we performed marker selection using the methylation beta matrix in two ways:

1. Empirical Bayes t-test using the TOAST package (*16*), selecting 100 marker probes per cell-type
2. Row t-test using the minfi package (*79*), selecting a total of 100 marker probes per cell-type (50 up-methylated and 50 down-methylated marker probes)

The overlap between cell-type marker sets is provided in the below table:

Marker selection	**Extremes**	**Chi-squared**	**Empirical Bayes**	**Row t-test**
Extremes	4287			
Chi-squared	709	1400		
Empirical Bayes	3	4	700	
Row t-test	7	4	90	700

For the WGBS/FACS and EpiSCORE RNA-based reference datasets, the cell-type markers had already been selected in the corresponding publications.

Validation of marker selection

We ensured that the "extremes" markers from the sequencing reference dataset had strong effects within the sequencing dataset (**fig. S1**), that they demonstrated concordant patterns in the external array reference dataset (**fig. S2**), and that these markers were not susceptible to strong batch effects within the bulk dataset (**fig. S3**).

Reference-based deconvolution algorithm and alternatives

We applied and compared three reference-based deconvolution algorithms to the Luo et al. 2023 single-cell methylation sequencing reference dataset *(24)*.

1. Houseman (*15*): one of the first and most commonly-used cell-type deconvolution algorithms, which was originally designed for array-based reference and bulk data
2. methylCC (*17*): similar to the Houseman algorithm, except that it uses random variables to account for instances where reference and bulk data are derived from different platforms (e.g., array versus sequencing). For the methylCC algorithm, it was important to deconvolve by batch (with subsequent batch correction as a covariate in statistical modelling), as deconvolving all samples across batches together gave dissimilar estimates (**fig. S4**). This is likely because the bulk methylation datasets used in this study often did have noticeable batch effects, and the methylCC random variables may have also been modelling these rather than platform-specific differences.
3. CelFiE (*78*): originally designed for sequencing-based reference and bulk data, and for more heterogeneous mixtures of cell-types.

The WGBS/FACS dataset used a NNLS-based algorithm similar to the Houseman algorithm *(27)*.
The EpiSCORE RNA-based reference is accompanied by its own deconvolution algorithm *(28)*.

Input bulk methylation data QC: normalisation and batch correction

We applied the deconvolution algorithms to bulk methylation data after functional normalisation using meffil (*74*), and after excluding MASK probes (see **Materials and Methods**).

We considered using ComBat (*76*) to perform batch correction on the bulk methylation data, ultimately taking different approaches to batch correction between reference-based and reference-free approaches. For reference-based deconvolution, ComBat made little overall impact on deconvolution (**fig. S4b**), while also inducing negative methylation beta values that are not handled by reference-based deconvolution algorithms (which expect values between 0 and 1). One option could be to set negative values to zero, though this is not ideal as it changes the statistical distribution. Hence, we dealt with this in the following ways:

- For reference-based deconvolution, we corrected for batch statistically in linear modelling.
- For reference-free deconvolution, we used the ComBat-normalised beta matrix, after setting negative beta values to zero.

Specifically for the methylCC reference-based deconvolution only, we also compared deconvolving all bulk samples together, versus deconvolving per-batch with subsequent aggregation of CTP estimates. These two different approaches gave different estimates, likely because the random effect term captures batch effects rather than only platform-specific effects. This is problematic as the mean and variance is different for each batch (e.g., across 5 batches for the LIBD dataset), whereas the methylCC random effect term specifies a single mean and variance. Hence, we chose to apply methylCC deconvolution by batch, with post-hoc aggregation, and statistical correction.

Evaluation of comparisons

Overall, we compared the following iterations of reference-based deconvolution pipelines:

Reference	**Cell-types**	**Marker selection**	**Deconvolution algorithm**	**Abbreviation in Supplementary Figures**
Single-cell methylation sequencing Luo 2023 *(24)*	Excitatory neuron Inhibitory neuron Astrocytes Endothelial cells Microglia Oligodendrocytes OPCs	Extremes	Houseman	"hseq_*"
		Extremes	methylCC	"mcc_*"
		Chi-squared	methylCC	
		Extremes	CelFiE	"celfie_*"
WGBS of FACS cell populations (*27)*	Neuron Astrocyte Endothelial (not brain-specific) Monocytes + macrophages (proxy for microglia) Oligodendrocytes OPCs	Loyfer et al. 2023 (*27*)	NNLS algorithm	
RNA-based reference *(28)*	Excitatory neuron Inhibitory neuron Astrocytes Endothelial cells Microglia Oligodendrocytes OPCs	Zhu et al. 2022 (*28*)	EpiSCORE	
Array (aggregated) *(20, 63, 81, 82, 84)*	Glutamatergic GABA-ergic Astrocytes Endothelial cells Monocytes + macrophages (proxy for microglia) Oligodendrocytes OPCs	eBayes	Houseman	"h_*"
		Row t-test	Houseman	
Array (NeuN+/-) Guintivano 2013 *(80)*	NeuN+ NeuN-	eBayes	Houseman	

We triangulated the optimal deconvolution pipeline based on the following criteria. Overall, on the basis of the criteria mentioned above, the optimal deconvolution pipeline involved the combination of the Luo et al. 2023 *(24)* sequencing reference, selection of marker probes based on the "extremes" approach, and applying the Houseman (*15*) deconvolution algorithm.

Number of cell-types in the reference panel

The Luo et al. 2023 *(24)*, RNA-based reference from the EpiSCORE paper (*28)* and aggregated array references *(20, 63, 81, 82, 84)* each had 7 cell-types. The WGBS/FACS reference had 6 cell-types *(27)*. The NeuN+/- reference had 2 cell-types *(80)*.

Note that for the WGBS/FACS array reference *(27)*, we used non-brain endothelial cells profiles as a proxy for brain endothelium, and monocytes/macrophage profiles as a proxy for microglia. For the aggregated array reference, we also used monocytes/macrophage profiles as a proxy for microglia. This may limit the efficacy of deconvolution.

Comparison of alternative reference-based deconvolutions

The NeuN+/- Houseman deconvolution was considered the historical gold-standard, having been widely and robustly used over many years. We compared sums of all glial cells (astrocytes, endothelial cells, microglia, oligodendrocytes, OPCs) correlating to the sum of NeuN- cell-type proportions deconvolved using the Houseman algorithm (*15*), and the sums of neurons (excitatory and inhibitory) correlating to the sum of NeuN+ cell-type proportions (**fig. S5**).

We next compared more granular cell-types across deconvolution pipelines (**fig. S5-S6**). We were reassured that the cell-type signal was robust across multiple different reference datasets, marker selection processes and deconvolution algorithms.

Performance in correctly deconvolving sorted (FACS) brain cell populations

fig. S8 shows the results of applying our deconvolution pipeline (Yap et al.), as well as WGBS/FACS and RNA-based deconvolution pipelines to DNA methylation array profiles of FACS sorted cell populations *(63)*. We found excellent concordance between our deconvolution results and the sorted cell-type identities, particularly for "NeuN positive" cell populations (median neuron CTP = 79.9%, SD = 3.1%) and "Sox10 positive" populations (median oligodendrocyte/OPC CTP = 76.8%, SD = 5.9%), whereas the deconvolution of the "Double negative" population was more mixed. However, it is also plausible that the Double negative population captured immature neurons, or even true neurons, given that some populations of neurons are weakly NeuN positive or even NeuN negative (*105*). In contrast, the RNA-based deconvolution pipeline (*28*) performed poorly in accurately identifying NeuN positive and Sox10 positive CTPs (**fig. S8**). These findings lend empirical support to the idea that RNA-based markers produce biased deconvolutions that reflect cellular activity rather than proportions, and also support the notion that our proposed methylation-based CTP deconvolution pipeline is superior to EpiSCORE. The WGBS/FACS deconvolution pipeline had comparable or slightly improved performance to our pipeline; however, this reference dataset as noted above is limited to a smaller number of brain cell-types.

Reference-free deconvolution comparison

While reference-based deconvolution is required to identify labelled CTPs, we compared to reference-free estimates to improve confidence in our results. We experimented with two reference-free algorithms:

1. smartSVA (*13*): this method implements surrogate variable analysis, which identifies drivers of variance in the dataset that are not related to the variable of interest (in this case, diagnosis). We applied smartSVA to ComBat-normalised methylation beta matrices as an unsupervised method could inadvertently capture batch effects over cell-types (**fig. S9**).

2. MethylNet (*30*): this is a machine-learning method, which uses a variational autoencoder (VAE) to "encode" data into a small number of embeddings, and optimises this process by reconstituting the dataset from the embeddings and comparing to the original input dataset. We ran the VAE across the methylation datasets aggregated across the three studies, filtered for variable probes with sd≥0.02. We extracted the VAE embeddings (which are analogous to principal components in a non-linear space), and looked for relationships with the reference-based deconvolutions. Unlike PCA and SVA, the VAE embeddings are not ordered by relative importance, but visualisation indicated strong correlations with deconvolved cell-types from the sequencing reference + "extremes" markers + Houseman algorithm: between embedding 3 and glial/neuronal proportions, and embedding 9 and oligodendrocytes (**fig. S10**).

We found that one of the first couple of sSVs from smartSVA (*13*) invariably had strong correlation with oligodendrocyte proportion, which likely reflects grey matter dissection artefact. There was also strong correlation between cell-types and embeddings from the VAE method.

<u>CETGYO score (Houseman deconvolution algorithms only)</u>
For deconvolutions using the Houseman algorithms: CETGYO score (*85*) (a RMSE measure) generally ≤ 0.1 indicates good quality deconvolution. This was generally the case (**fig. S11**).

Out of the array reference marker selection methods, we favoured the empirical Bayes t-test method implemented in TOAST (*16*) as its deconvolution had a lower CETGYO score (*85*) (**fig. S7a-b**).

<u>Comparison of deconvolved CTPs to other omics modalities</u>
For UCLA_ASD and LIBD, we compared bulk methylation deconvolved CTP estimates to bulk RNA-seq WGCNA modules (*69*) as well as conventional reference-based deconvolution (*7*) of equivalent cell-types. For UCLA_ASD, we also compared to cell counts from single-cell data (*32*). For ROSMAP, we compared deconvolution to both IHC data (*22*) and cell counts from single-cell data (*33*).

It is important to note the limitations of comparing CTP estimates from orthogonal omics technologies. The first issue is one of sampling. Recent work by Toker et al. 2023 (*23*) demonstrated large sampling effects, such that CTP deconvolutions from neighbouring tissue samples may have Pearson correlation as low as r=0.30. The statistical solution to distinguish biological signal from sampling noise is to increase sample size/power (which our study design allows). The second issue is the impact of omics method on CTP estimates. Toker et al. found substantial differences in CTP estimates derived from different omics technologies yet generated from the exact same tissue sample. This was attributed to whether the omics method captures neuropil in addition to nuclear material (*23*). For example, CTP estimates derived from nucleus-focused omics datasets such as DNA methylation and histone acetylation correlated well together, whereas RNA-based estimates (which capture RNA within neuropil) were often negatively correlated. CTP estimates based upon technologies that specifically capture nuclear material are arguably more specific, as neuropil may be composed of neuronal processes originating far away from the brain region that was sampled.

fig. S12-S13 show correlations between bulk methylation CTPs and CTPs from orthogonal omics technologies. Reassuringly, we found reasonable concordance including associations with:

- Bulk RNA-sequencing WGCNA eigengenes corresponding to cell-types (*21*): UCLA_ASD (n=49) for excitatory neurons (Pearson's r=0.31, p=3.0e-2), endothelial cells (r=0.31, p=3.0e-2), microglia (r=0.53, p=1.0e-4); LIBD (n=394) for excitatory neurons (r=0.11, p=2.4e-2), inhibitory

neurons (r=0.17, p=5.0e-4), microglia (r=0.11, p=3.0e-2), and oligodendrocytes (r=0.15, p=2.5e-3);

- Bulk RNA-sequencing reference-based deconvolution (*7*): PsychENCODE with combined UCLA_ASD and LIBD (n=473): neurons (r=0.12, p=8.3e-3), microglia (r=0.37, p=8.2e-17), oligodendrocytes (r=0.40, p=1.8e-19), and OPCs (r=-0.14, p=1.9e-3);
- Cell counts from single-cell RNA-sequencing: UCLA_ASD (n=60) (*34*) microglia (r=0.61, p=9.8e-3); ROSMAP (n=48) (*33*) inhibitory neurons (r=-0.42, p=9.7e-3)); and
- Immunohistochemistry CTPs: ROSMAP (n=49) (*22*) microglia (r=0.38, p=6.7e-3)

CTPs being within expected limits

We favoured deconvolution pipelines whereby the CTP ranges were within expected limits: higher proportions of neurons and oligodendrocytes, and narrow CTP distributions with the exception of oligodendrocytes which we reasoned may be more variable due to grey matter dissection artefacts.

This criterion was useful when comparing "extremes" versus chi-squared test marker selection (**fig. S7c**). The chi-squared test selection overall gave more variable CTP estimates, whereas the "extremes" selection gave a narrower distribution of CTP estimates for microglia and OPCs that were more consistent with expectation. This criterion also favoured the Houseman deconvolution algorithm compared to methylCC (**fig. S14**).

Other comments

The methylCC (*17*) CTP deconvolution was not optimal, even when deconvolving by-batch (versus all batches simultaneously), or using different marker selection algorithms ("extremes" versus chi-squared test). In particular, the "extremes" marker selection with methylCC deconvolution gave reasonable CTP estimates except inhibitory neurons, for which there was large variance between samples (**fig. S11**). Our explanation for this is that the current inhibitory neuron reference markers are relatively poorly defined compared to the other cell-types: there were fewer markers that satisfied the "extremes" criteria (with a 0.4/0.6 split), and the beta values of these markers were less "extreme" than the other cell-types (**fig. S1**). Hence, the less well-defined inhibitory neuron profile violates the assumption of methylCC: that the true methylation value at the selected markers is 0 or 1. We experimented with including a "noise" term in the reference panel (taking markers that were consistently high or low across cell-types, and taking the opposite value as a "noise" reference marker). This showed that less well-defined cell-types with fewer markers had greatest variance with the methylCC deconvolution. We attempted to relax the "extremes" marker selection criteria for inhibitory neurons (e.g., 0.45/0.55 split) to increase the number of markers, but this did not improve the deconvolution. There are a couple of possible explanations for the inhibitory neuron profile being less well-defined. Firstly, there may be substantial inhibitory neuron cell-type heterogeneity, meaning that the distinct profiles are obscured when aggregating them into one population. We attempted to generate more granular inhibitory neuron profiles (caudal vs medial ganglionic eminence); however, there were not enough reference markers for a stable deconvolution. Alternatively, this may be an intrinsic characteristic of inhibitory neurons. It is interesting to note that the majority of inhibitory neuron markers were up-methylated, whereas for all other cell-types, the majority of markers were down-methylated (**fig. S2**), which has been seen in other studies (*20*).

Although the "extremes" marker approach yielded the largest number of marker sites, the CTP estimates were comparable to the eBayes marker approach. This suggests that including more reference marker sites does not necessarily improve deconvolution. Nevertheless, we retained all probes fulfilling the "extremes" criteria for simplicity.

Supplementary Note 2: Compositional data analysis

Motivation

Here, we provide additional information on compositional data analysis, as – outside the field of microbiome research – it is an under-recognised issue in omics analysis. The reference-based deconvolutions used in this analysis provide cell-type proportions which sum up to ~100%. This is a compositional data type, which presents analytical and interpretational challenges of data curvature (*106*). Each composition is correlated to the others, potentially introducing false positives (e.g., if there are two samples with equal absolute values for one cell-type but where the total sums of all cell-types within a sample are different, the relative values can also be markedly different), and making interpretation difficult (e.g., an identified association may simply represent subtle changes in other cell-types). We point the reader to other resources that explain this topic in greater depth (*107*, *108*).

Approaches taken

We took two approaches to analysing the compositional data here, both using functions implemented in the compositions R package (*108*, *109*).

1. Centred-log-ratio (clr) transformation: for each sample, each composition (in this case, cell-type proportion) is divided by the geometric mean of all compositions for that sample, and then the log is taken of this quotient: $clr(x_1, \ldots, x_k) = (log(\frac{x_1}{g(x)}), \ldots, log(\frac{x_k}{g(x)}))$, where $g(x)$ represents the geometric mean of compositions within a sample. The logarithmic function is important as its nonlinear characteristics can model the curvature of compositional datasets. Then, the transformed value is interpreted relative to the geometric mean. The clr-transformation is one of a family of log-ratio transformations; the advantage of comparing to the geometric mean is that there is no loss of variables (for example, additive-log-ratio (alr) transformation compares each composition to one chosen composition, meaning that that one composition is "lost"; there is also the isometric-log-ratio (ilr) transformation which also experiences difficulties in interpretation). One disadvantage of the method is that compositions of zero are not handled (log(0) is undefined), necessitating an offset value, which can shift distributions and affect analysis results (*110*).
2. Compositional PCA: a modified version of principal component analysis that accounts for compositionality (i.e., curvature and constraint to the sum) (*106*). This approach overcomes the offset issue. However, cell-types are represented as ratios or "balances" relative to each other, based on orthogonal axes calculated by maximising variance explained in the dataset. Thus, one limitation is that each PC needs to be back-interpreted to the original compositional variables using the PCA loadings. This method is implemented in the compositions R package (*108*, *109*) by applying the princomp() function to an acomp() object.

Supplementary Note 3: Investigating other clinical associations of CTPs

We additionally investigated whether the CTP associations could be driven by other clinical variables. As an analysis approach, we first tested for associations between CTPs and these additional clinical variables, before adding these clinical variables as a covariate in linear models regressing CTP against diagnosis.

In the UCLA_ASD dataset, there were no associations between microglia CTP and history of seizures or use of psychiatric medications. There was a nominal association between increased microglia and cardiac arrest as the primary cause of death (p=1.4e-2). However, in a model of microglia CTP ~ ASD diagnosis + cause of death + baseline covariates, the association between increased microglia and ASD diagnosis persisted, suggesting that cardiac arrest did not account for the microglia/ASD diagnosis association.

There were n=7 participants with dup15q CNVs (all within the ASD group), which gives limited power to identify statistical trends. Instead, we directly visualised potential outliers by plotting where these individuals sit within the various CTP distributions. We did not identify clear links between outliers and the dup15q participants.

In the LIBD dataset, available phenotypic data included antipsychotic and antidepressant medication, smoking status, and cause of death. Of these phenotypic variables, only the cause of death had statistically significant associations with brain CTPs, with trend-level associations between death by homicide and endothelial cells (b=0.05, SE=0.02, p=1.4e-2), natural death and microglia (b=-0.07, SE=0.03, p=4.8e-3), and death by suicide and oligodendrocytes (b=-0.11, SE=0.08, p=4.8e-2).

Given that oligodendrocytes were the only brain CTP significantly associated with schizophrenia, we jointly modeled diagnosis and mode of death. In this slightly smaller subset of participants with complete data for mode of death (n=347), there was still a marginal association between schizophrenia diagnosis and reduced oligodendrocytes (b=-0.06, SE=0.03, p=5.2e-2).

Unfortunately, for the ROSMAP dataset, no data were collected on medications, cause or death or other comorbidities so we were unable to perform these analyses. However, for clinical associations, we do refer to our extensive sensitivity analyses of alternative measures of Alzheimer's disease severity (eg. clinical, neuropathological) outlined in the Results section and **Fig. 3**.

Supplementary Note 4: Dissecting variance in CTPs explained by polygenic scores and demographic variables

For these statistically significant PGS-brain CTP associations, we quantified the variance explained by the PGS, in addition to variance explained by the covariates (i.e., a conditional analysis whereby the ANOVA model including covariates in the following order: CTP (clr-transformed) ~ age + age^2 + sex + batch + genoPC1-3 + PGS). This analysis is informative where predictor variables may be correlated. The Alzheimer's disease PGS explained R^2=0.38% of the variance in the clr-transformed endothelial CTP (p=3.82e-2 in the full model ANOVA, n=502) (**fig. S28**). The schizophrenia PGS explained R^2=0.51% (p=5.98e-2) and R^2=0.47% (p=3.11e-2) of the variance in clr-transformed oligodendrocyte and astrocyte CTPs, respectively (**fig. S28**).

We sought to further understand the variance explained in endothelial CTP by the Alzheimer's disease PGS, as we had observed both genetic and phenotypic effects. For the association between endothelial cells and Alzheimer's disease, we performed an analysis including all controls plus the ROSMAP participants with an Alzheimer's disease diagnosis. These individuals on average have both higher Alzheimer's disease PGS (**Figure 4a**) and lower endothelial cells (**Figure 3b**), and, as expected, the inclusion of donors with Alzheimer's disease increased the proportion of variance in endothelial cells explained by the Alzheimer's disease PGS. (R^2=0.90%, p=3.17e-4, n=761) (**fig. S28**). When we restricted the variance analysis to ROSMAP participants (which would reduce variance due to age), the Alzheimer's disease PGS explained R^2=1.76%, p=7.62e-4, n=620). This suggests that genetic risk for Alzheimer's disease is most predictive of endothelial CTP among older people, and that – in this more homogeneous group – Alzheimer's disease PGS better predicts endothelial CTP than sex (R^2=1.62%, p=1.23e-3, n=620) or age (R^2=0.87%, p=1.76e-2, n=620) (**fig. S28**).

Supplementary Note 5: Mediation analysis and Mendelian Randomisation

We performed mediation analysis for AZD PGS → endothelial CTP → AZD with the ROSMAP dataset. This tests the question: *does a decrease in endothelial cells lie on the causal pathway between Alzheimer's disease genetic risk and Alzheimer's disease?*

Using the R mediation package, we identified a statistically significant mediation effect in models both including and excluding covariates. In the model including covariates, the average causal mediation effects (ACME) estimate was 8.9e-3, 95%CI (2.6e-3, 2.0e-2), p~0, which represents the mediating effect of endothelial cells on the Alzheimer's PGS/diagnosis association (i.e., Alzheimer's PGS → endothelial CTP → Alzheimer's diagnosis). The average direct effects (ADE) estimate was 0.15, 95%CI (0.11, 0.19), p~0), representing the direct effect of Alzheimer's PGS → Alzheimer's diagnosis.

Regarding Mendelian Randomization, we also performed MR analyses using GWAS summary statistics, though noting that we expect these analyses to be underpowered as the brain CTP GWAS are small with no genome-wide significant SNPs for endothelial cells, thereby violating a core assumption of MR. We did not identify a statistically significant causal relationship between endothelial CTP (exposure) → AZD (outcome) across multiple MR methods (CAUSE, IVW, MR-Egger, Weighted Median, Weighted Mode, GSMR). A power calculation indicated that this analysis was underpowered with only 7% power to detect a causal relationship with $p<0.05$.

We also performed an MR analysis for the reverse direction: AZD (exposure) → endothelial CTP (outcome). This analysis did not identify any statistically significant causal relationships across the multiple tested MR methods. However – as the AZD GWAS is large with robust instrumental SNPs – we found that this analysis was adequately powered. Thus, from this we can conclude that Alzheimer's disease does not causally reduce endothelial CTP.

Taken together, these analyses suggest that endothelial CTP lies on the causal pathway to Alzheimer's disease, whereas Alzheimer's disease does not causally contribute to endothelial CTP.

Supplementary Note 6: Multi-ancestry analysis

The majority of samples across all three datasets were donated by individuals of European ancestry (EUR, n=885 with matched genotyping data, **Figure 1b**). However, there were an additional n=181 individuals of African ancestry (AFR) within the LIBD dataset with matched genotyping data. In line with contemporary efforts to include greater diversity in genetic cohorts (*111*, *112*), we also performed a meta-analysis including the LIBD African participants within the GWAS meta-analysis. We performed a similar meta-analysis to the EUR meta-analysis described in the main text, first performing a linear model GWAS to the LIBD African data (covariates: age, age^2, sex and batch, diagnosis and 5 population stratification genotyping PCs from that dataset), then adding these results to the meta-analysis performed using METAL. We did not include the African participants from the ROSMAP and UCLA_ASD studies as the sample size was too small (i.e., UCLA_ASD n=8; ROSMAP n=1).

We found that the signal for genome-wide significant (GWS) SNPs (i.e., the rs6011327/inhibitory neuron and the rs17025223/astrocyte SNP-trait associations) was attenuated in the LIBD AFR dataset. Hypothesising that demographic differences in the LIBD AFR dataset may be contributing noise to the GWAS meta-analysis, we explored the effect of adjusting for each of the covariates on the GWS SNP effect estimate. For this, we generated sets of forest plots per study (**fig. S33-34**) and per GWS SNP-trait pair, including SNP effect statistics from the following model:

$$CTP \sim SNP + covariate$$

For example, to test the effect of correcting for age on the SNP effect: we plotted the SNP effect and standard error from the model CTP ~ SNP + age.

We found that the LIBD AFR dataset introduced between-dataset heterogeneity. This was highlighted by the impact of age and batch covariates with the SNP effects for rs6011327 (inhibitory neuron) (**fig. S33**).

Overall, we decided against including the LIBD AFR sample in our GWAS meta-analysis, as there appeared to be confounding between the cohort age and brain CTPs, and also because this was a relatively small sample. Alternatively, it is also possible that the top SNP is not the causal variant but instead tags the causal variant, explaining the discrepancies in effect direction between the European and African groups.

Supplementary Figures

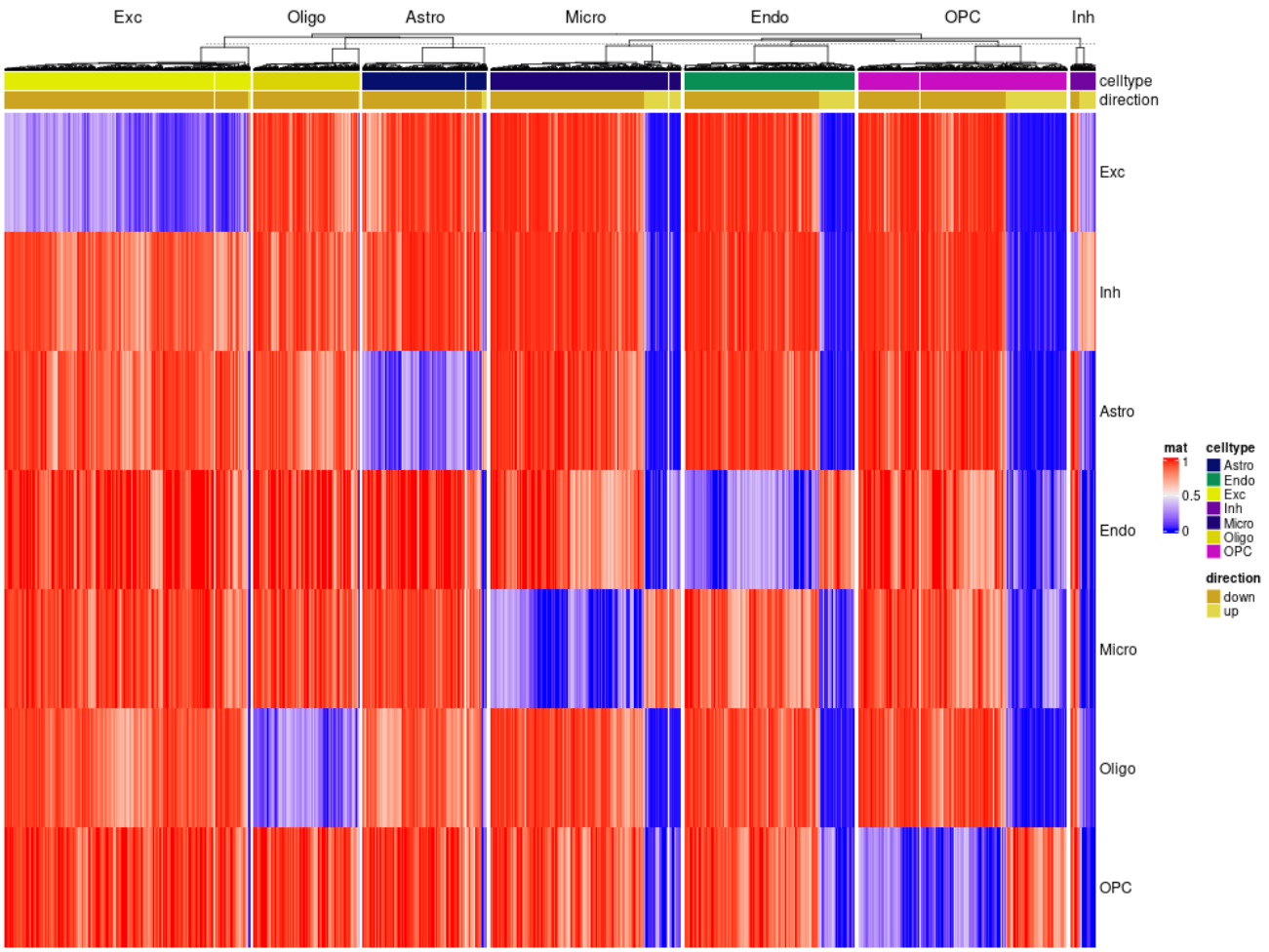


Figure S1: Hierarchically clustered (columns) heatmap of beta values for the sequencing marker probes selected using the "extremes" algorithm (columns), across each of the seven cell-types within the Luo et al. 2023 *(24)* sequencing reference dataset (rows). That is, this heatmap is showing beta values for marker probes selected from within the same dataset.

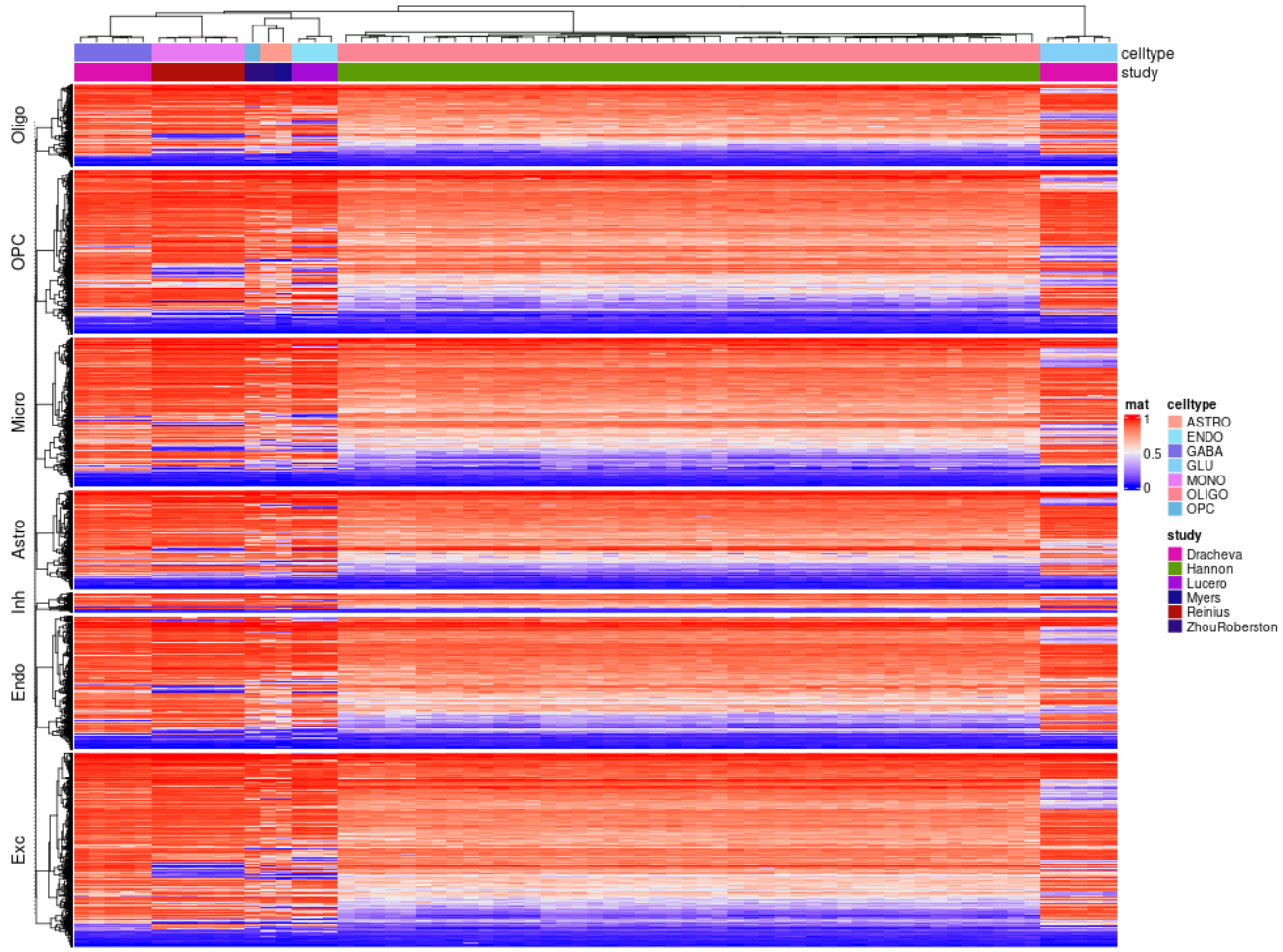


Figure S2: Hierarchically clustered (by rows and columns) heatmap of beta values for the Luo et al. 2023 *(24)* sequencing marker probes selected using the "extremes" algorithm (rows), across each of the seven cell-types within the aggregated array reference dataset (columns). Bars of colours at the top of the heatmap denote the array reference cell-type, and the study from which the data came from. This plot demonstrates consistency in marker probe beta values across sorted cell populations.

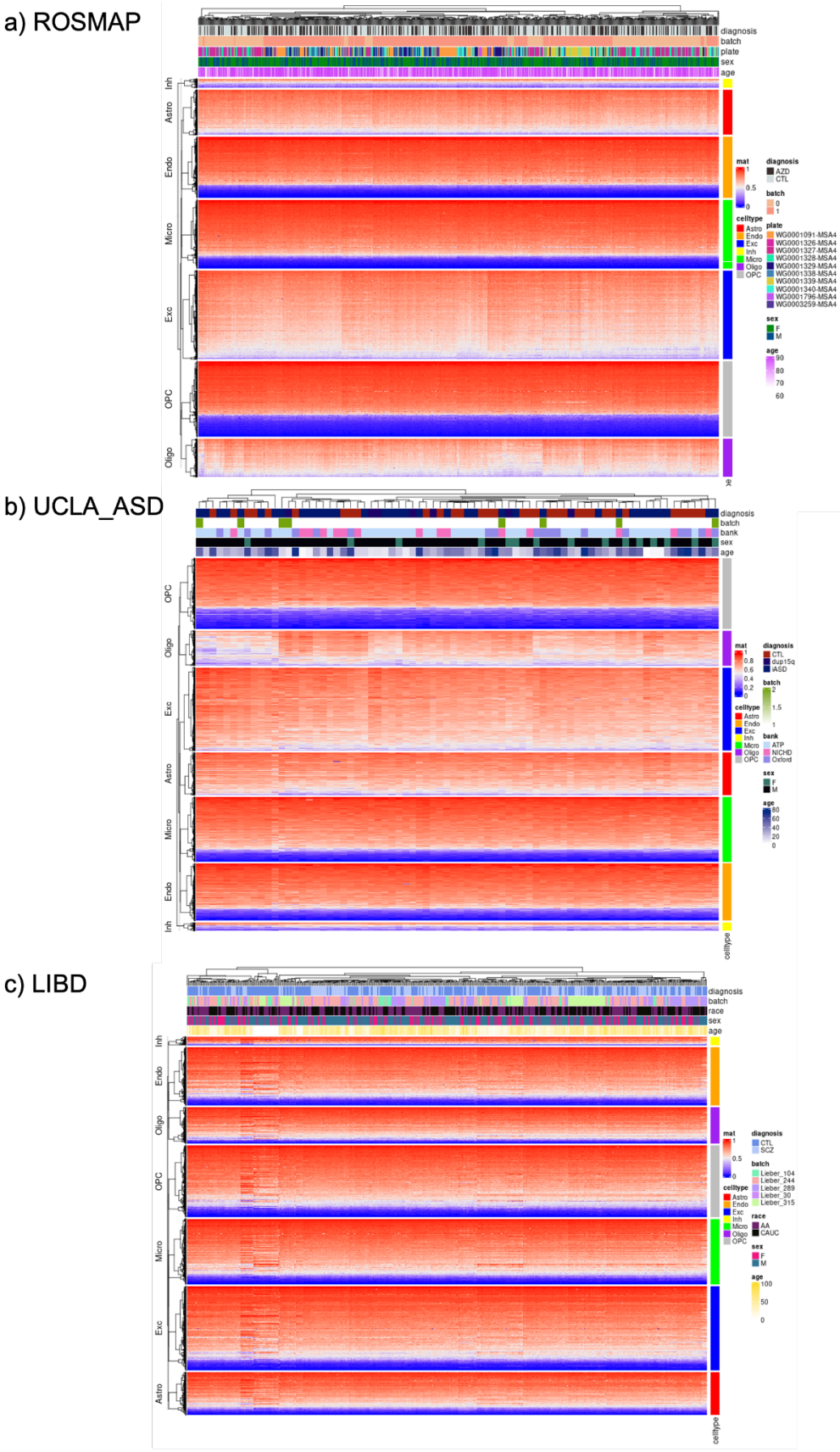
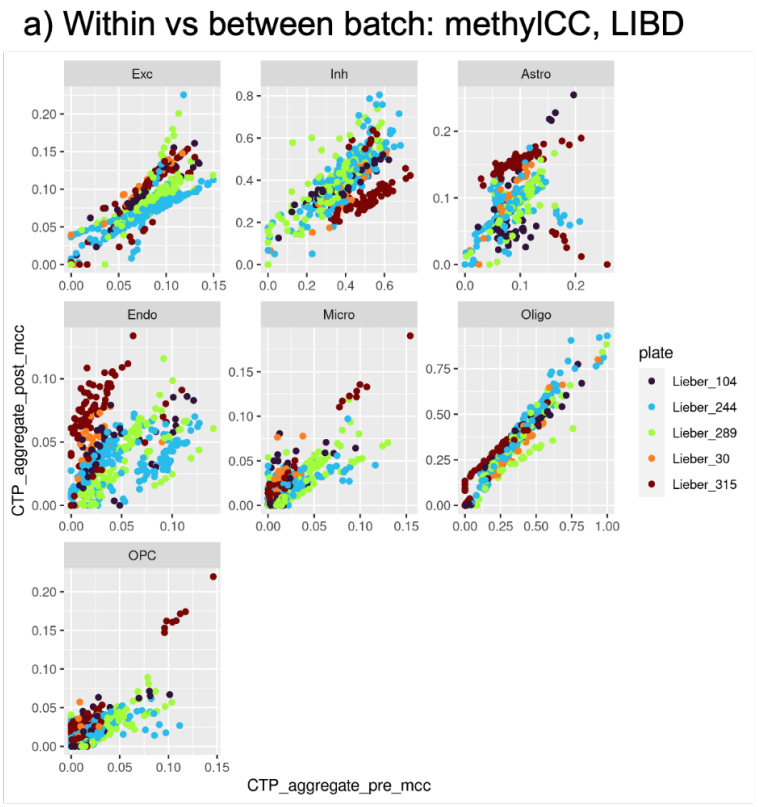


Figure S3: Hierarchically clustered (rows) heatmap of beta-values for the sequencing marker probes selected using the "extremes" algorithm (rows), per sample in a) ROSMAP, b) UCLA_ASD and c) LIBD studies. Colours at the top show per-sample characteristics. This plot shows that these beta-values are consistent across samples, and are unlikely to be driven by batch effects. The greatest variability is seen in oligodendrocyte markers, which likely reflects white matter contamination in dissection.



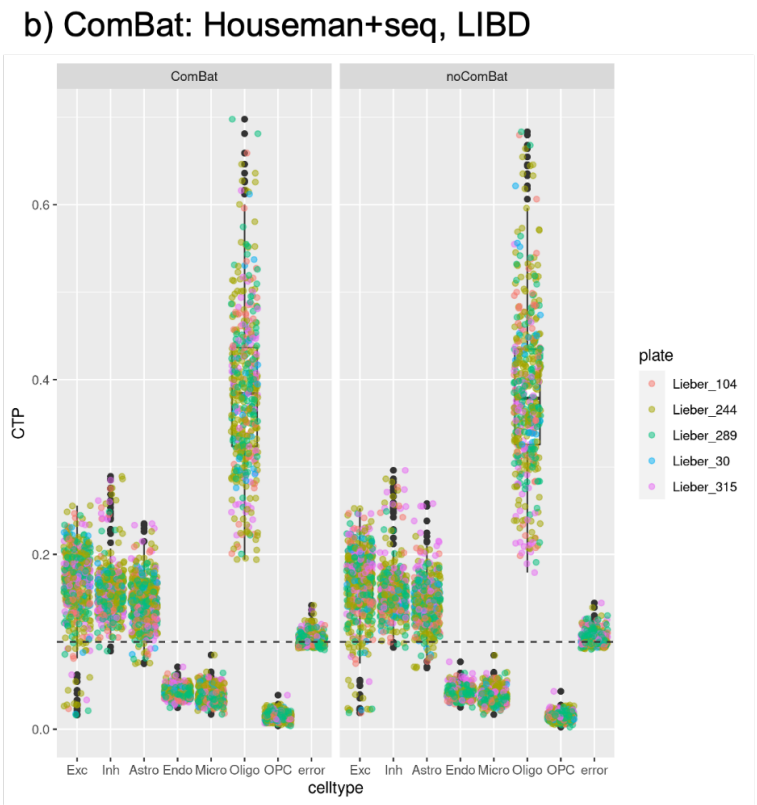


Figure S4: Exploration of methylCC deconvolution using the LIBD dataset. a) Comparison of deconvolution by-batch before aggregation (y-axis), versus deconvolution of all batches at once (x-axis). This shows that batch effects can interfere with the deconvolution. b) Comparison of ComBat batch normalisation (left applies ComBat; right does not apply ComBat), for the deconvolution approach of Luo et al. 2023 sequencing reference + "extremes" marker probe selection + Houseman algorithm. This shows that ComBat batch normalisation has negligible effect on reference-based deconvolution.

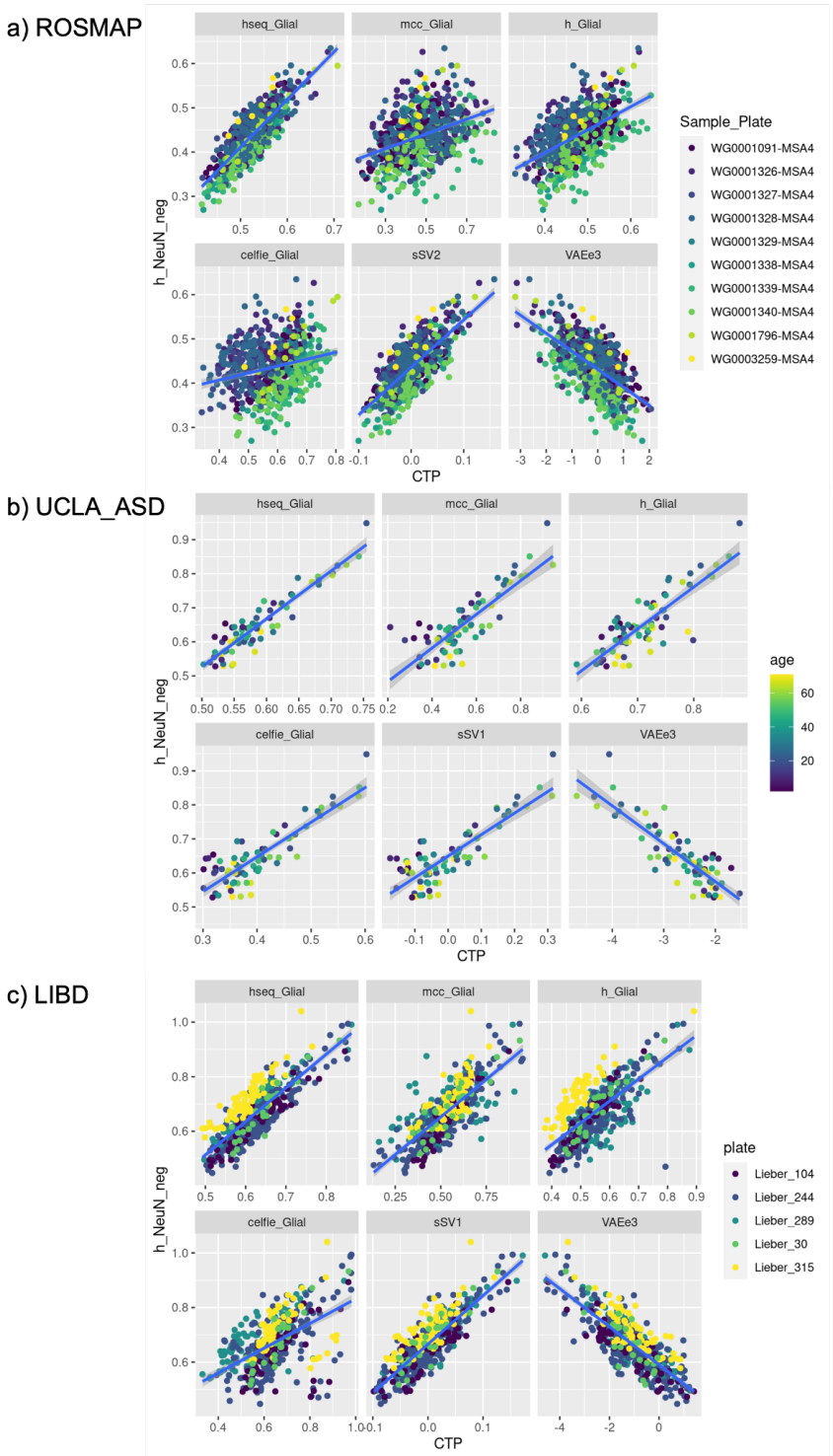


Figure S5: Comparison of the NeuN- deconvolution from Guintivano et al. 2013 *(80)* (considered the historical "gold-standard" methylation array reference; this deconvolution involved marker selection using eBayes t-test and Houseman algorithm) to the sum of non-neuronal cells (astrocytes, endothelial cells, microglia, oligodendrocytes, OPCs) from various deconvolution approaches, for a) ROSMAP, b) UCLA_ASD, and c) LIBD studies. The reference-based approaches include Luo et al. 2023 sequencing reference *(24)* + "extremes" marker selection + Houseman algorithm ("hseq_Glial"), Luo et al. 2023 sequencing reference + "extremes" marker selection + methylCC algorithm ("mcc_Glial"), aggregated array reference + eBayes t-test marker selection + Houseman algorithm ("h_Glial"), Luo et al. 2023 sequencing reference + "extremes" marker selection + CelFiE algorithm ("celfie_Glial"). The reference-free approaches include smartSVA ("sSV") and embeddings from MethylNet, a variational autoencoder method ("VAEe").

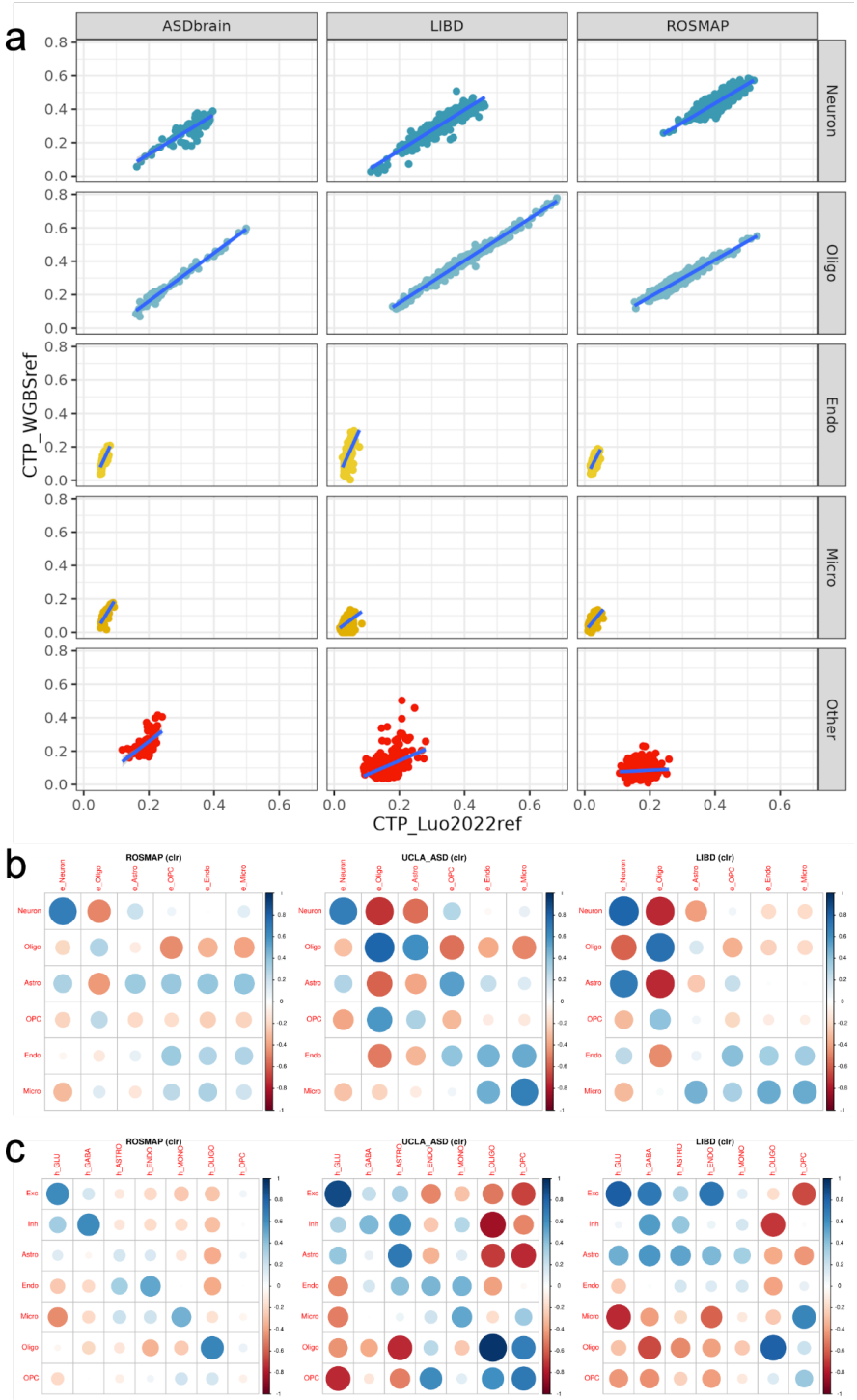


Figure S6: Comparison of alternative reference panels. a) Correlation between CTPs from the Yap et al. deconvolution pipeline (based on Luo et al. 2023 *(24)*) (x-axis) versus CTPs from an independent reference panel derived from WGBS sequencing of fluorescence-activated cell sorting (FACS) from Loyfer et al. 2023 *(27)* (y-axis). From the WGBS/FACS-based deconvolution, we directly compared CTPs for neurons and oligodendrocyte CTPs; as "proxy" cell-types, we also compared vascular endothelial cells (against brain endothelial cells) and macrophages/monocytes (against brain microglia). We also aggregated all other remaining cell-types from the WGBS/FACS-based deconvolution as an "Other" CTP and compared this against the sum of astrocytes and OPCs. b) Heatmap of correlations between CTPs from the Yap et al. deconvolution pipeline (rows) versus CTPs from the EpiSCORE RNA-based pipeline *(28)* (columns). c) Heatmap of correlations between CTPs from the Yap et al. deconvolution pipeline (rows) versus CTPs from the aggregated array-based reference *(20, 63, 81, 82, 84)* pipeline.

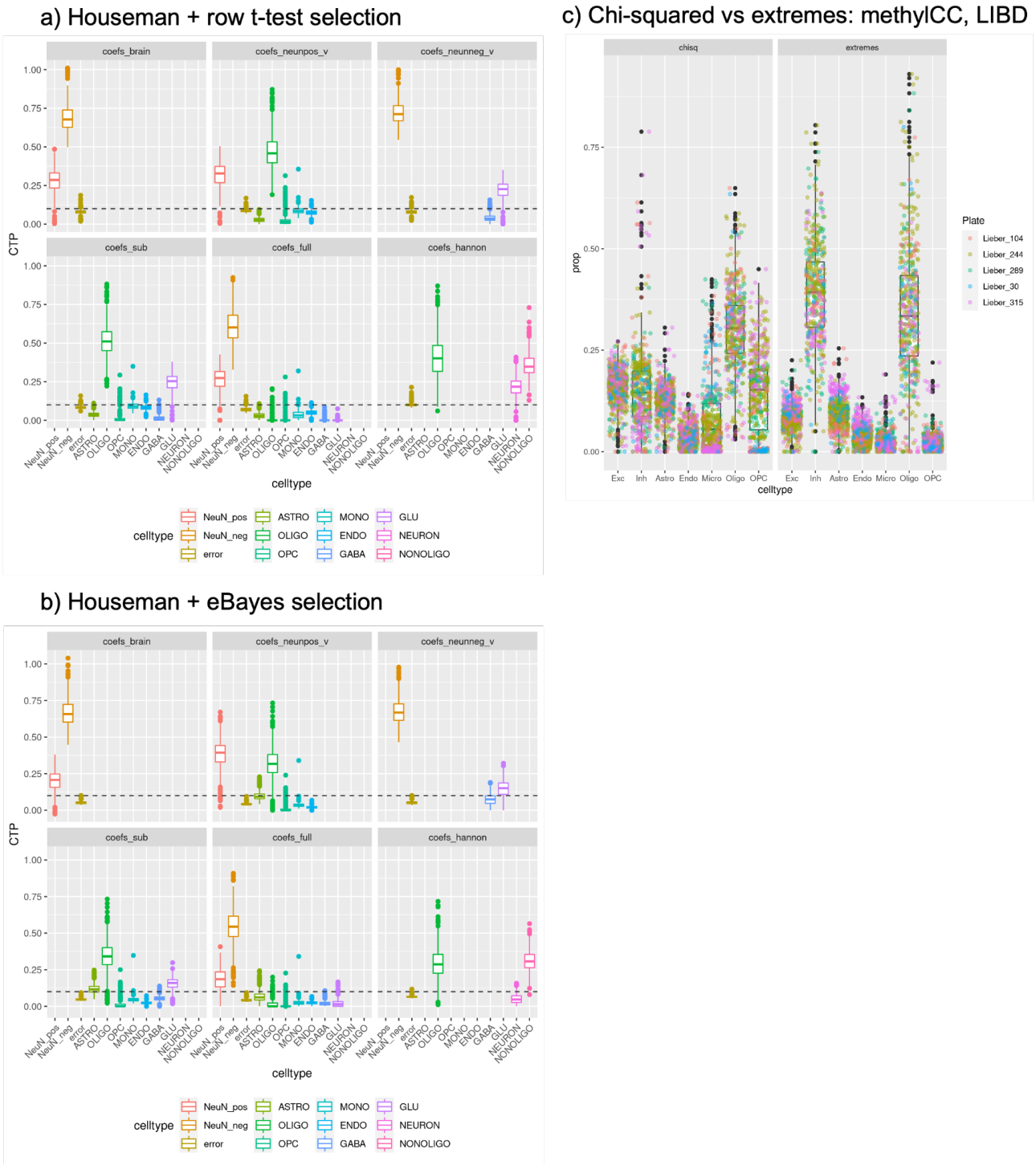


Figure S7: Comparisons of marker selection, using the LIBD dataset for benchmarking. a) Aggregated array reference marker selection, using the combination of a row t-test + Houseman algorithm for deconvolution. The aim of this analysis was to demonstrate the stability of deconvolution when using different combinations of reference cell-types: “coefs_brain” panel: reference includes NeuN+ and NeuN-; “coefs_neumpos_v” panel: reference includes NeuN+, astrocytes, oligodendrocytes, OPCs, microglia using monocytes as a proxy, endothelial cells; “coefs_neunneg_v” panel: reference includes NeuN- reference, excitatory neurons and inhibitory neurons; “coefs_sub” panel: reference includes all seven granular cell-types; “coefs_full” panel: reference includes all of NeuN+, NeuN-, excitatory neurons, inhibitory neurons, astrocytes, oligodendrocytes, OPCs, microglia using monocytes as a proxy, endothelial cells; “coefs_hannon” panel: entirely using the array reference from Hannon et al. 2024 which includes neurons, oligodendrocytes, and other non-neuronal cells. b) Aggregated array reference marker selection, using the combination of eBayes t-test selection + Houseman algorithm. Panels are as described in a). c) Comparison of chi-squared test versus “extremes” marker selection, for the methylCC deconvolution.

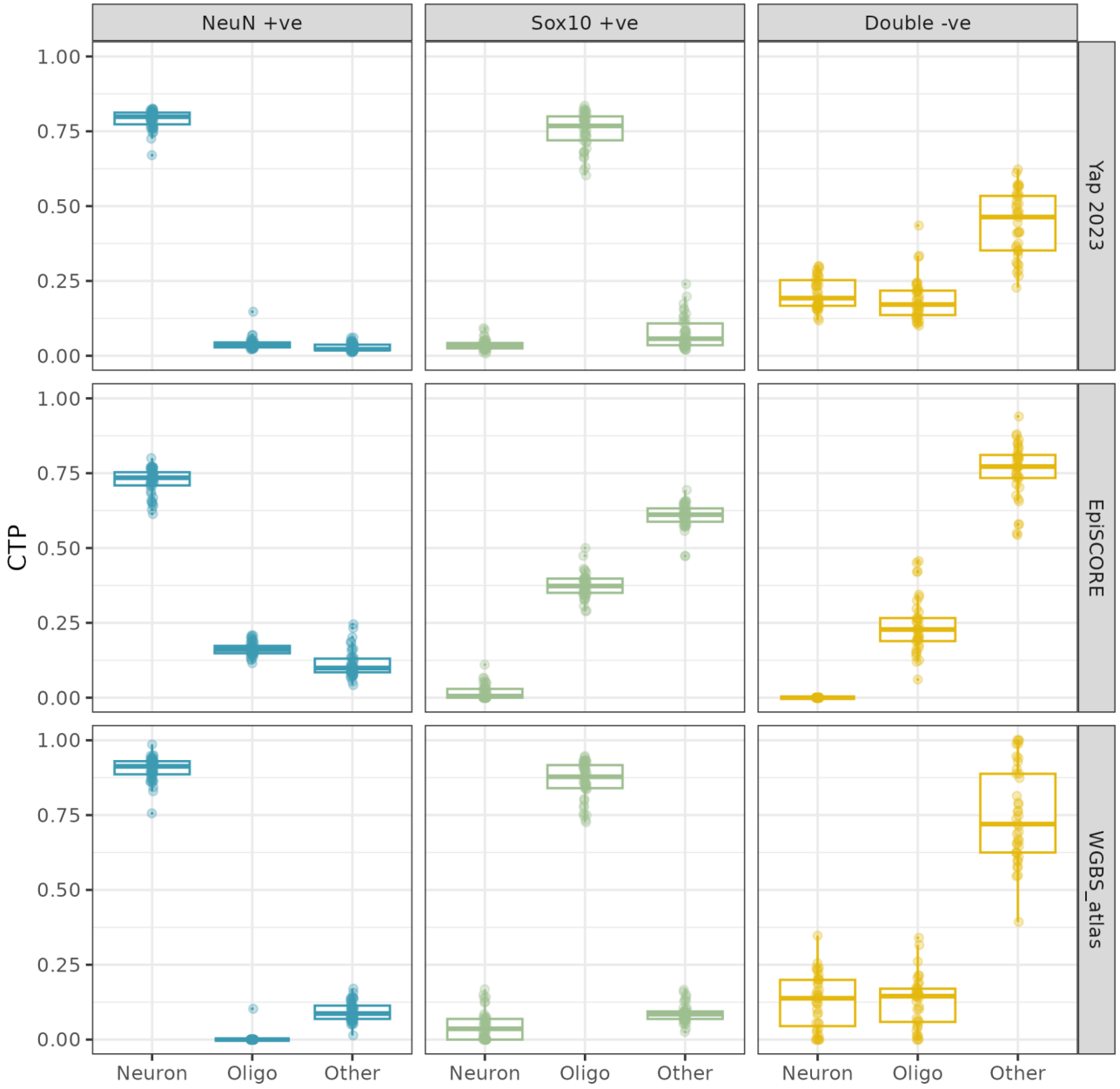


Figure S8: Benchmarking of Yap et al., WGBS/FACS *(27)*, and RNA-based EpiSCORE *(28)* deconvolution algorithms using sorted (FACS) brain cell populations from Hannon et al. 2024 *(63)*, which provide a "ground truth". In an optimal deconvolution pipeline, the "NeuN+" population should be entirely deconvolved as neurons ("Neuron"), the Sox10+ population should be entirely deconvolved as oligodendrocytes (Oligo), and the "Double -ve" population should be entirely deconvolved as "Other" cells.

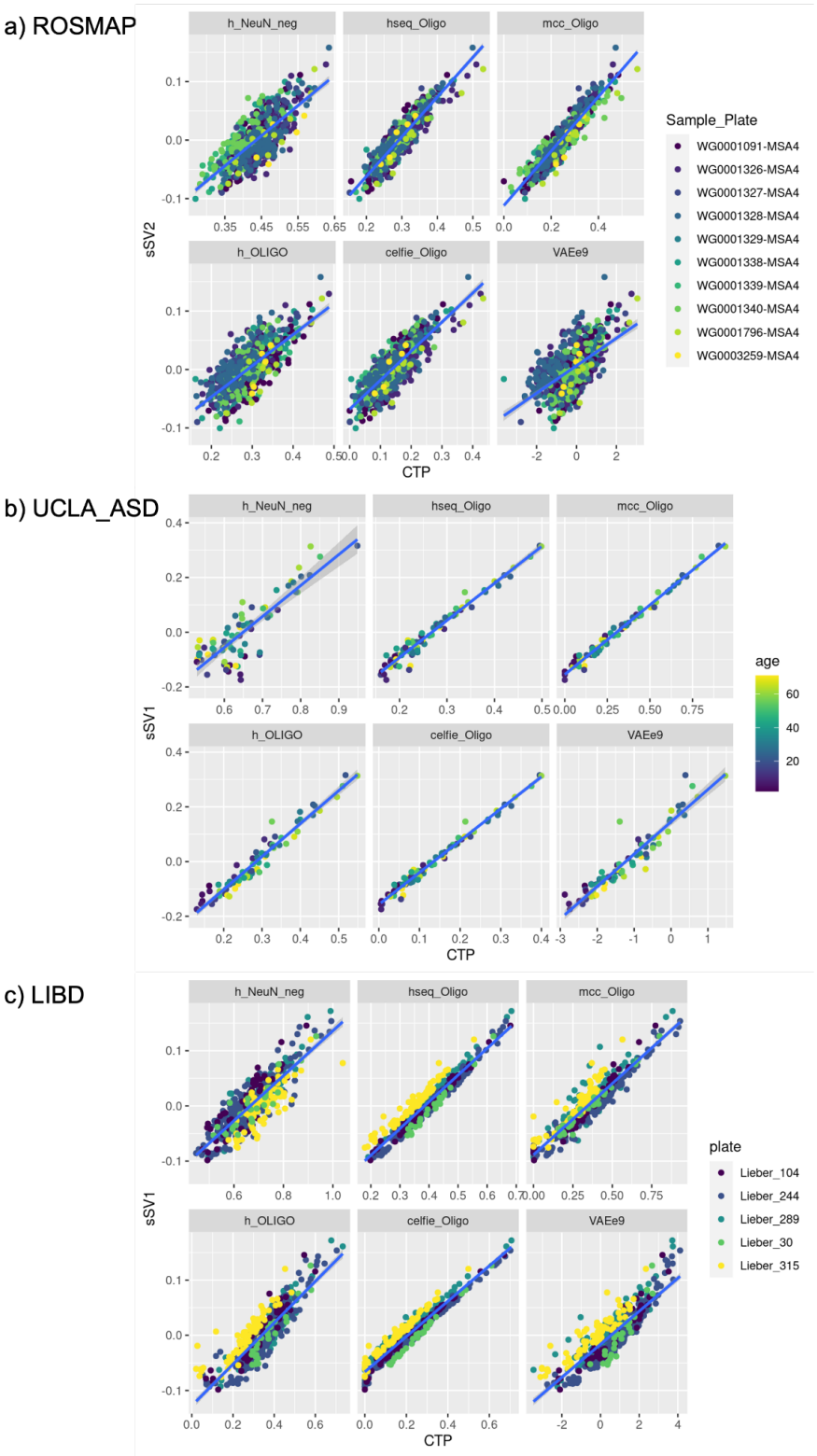


Figure S9: Comparison of surrogate variables from smartSVA (a linear reference-free method) to the oligodendrocyte proportion deconvolved using various algorithms, for a) ROSMAP, b) UCLA_ASD and c) LIBD studies. The reference-based approaches include Luo et al. 2023 sequencing reference *(24)* + "extremes" marker selection + Houseman algorithm ("hseq_Glial"), Luo et al. 2023 sequencing reference + "extremes" marker selection + methylCC algorithm ("mcc_Glial"), aggregated array reference + eBayes t-test marker selection + Houseman algorithm ("h_Glial"), Luo et al. 2023 sequencing reference + "extremes" marker selection + CelFiE algorithm ("celfie_Glial"). The reference-free approaches include smartSVA ("sSV") and embeddings from a variational autoencoder method ("VAEe").

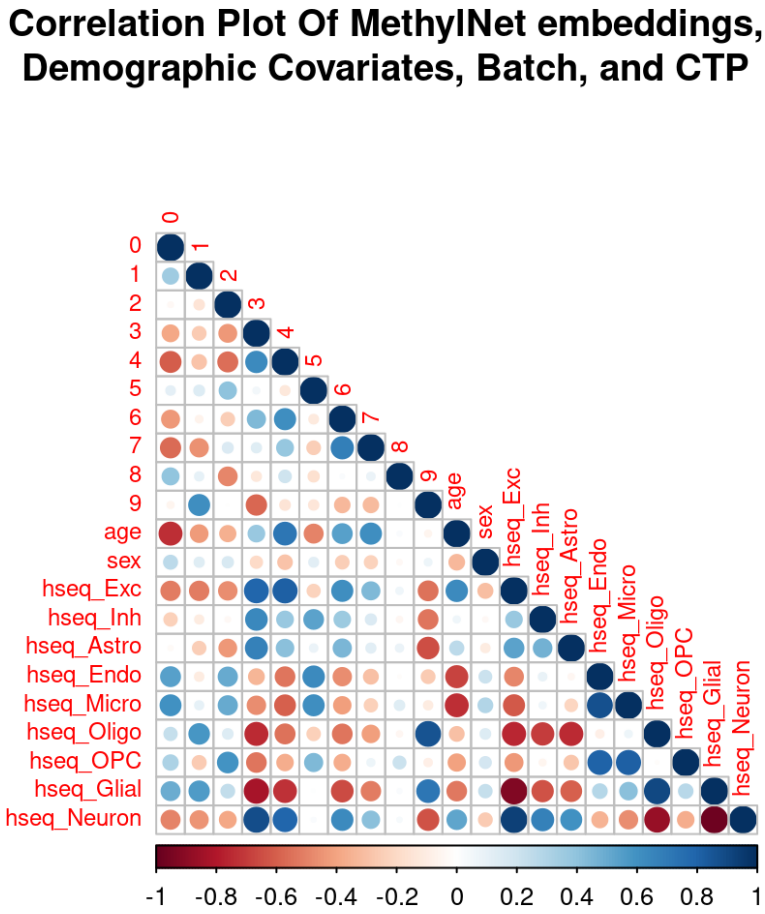


Figure S10: Correlation plot of embeddings (0 to 9) from a non-linear deconvolution method, MethylNet versus brain CTPs deconvolved using the Luo et al. 2023 sequencing reference *(24)* + "extremes" marker probe selection + Houseman algorithm (prefix: "hseq"). MethylNet embeddings were generated for all individuals across all datasets.

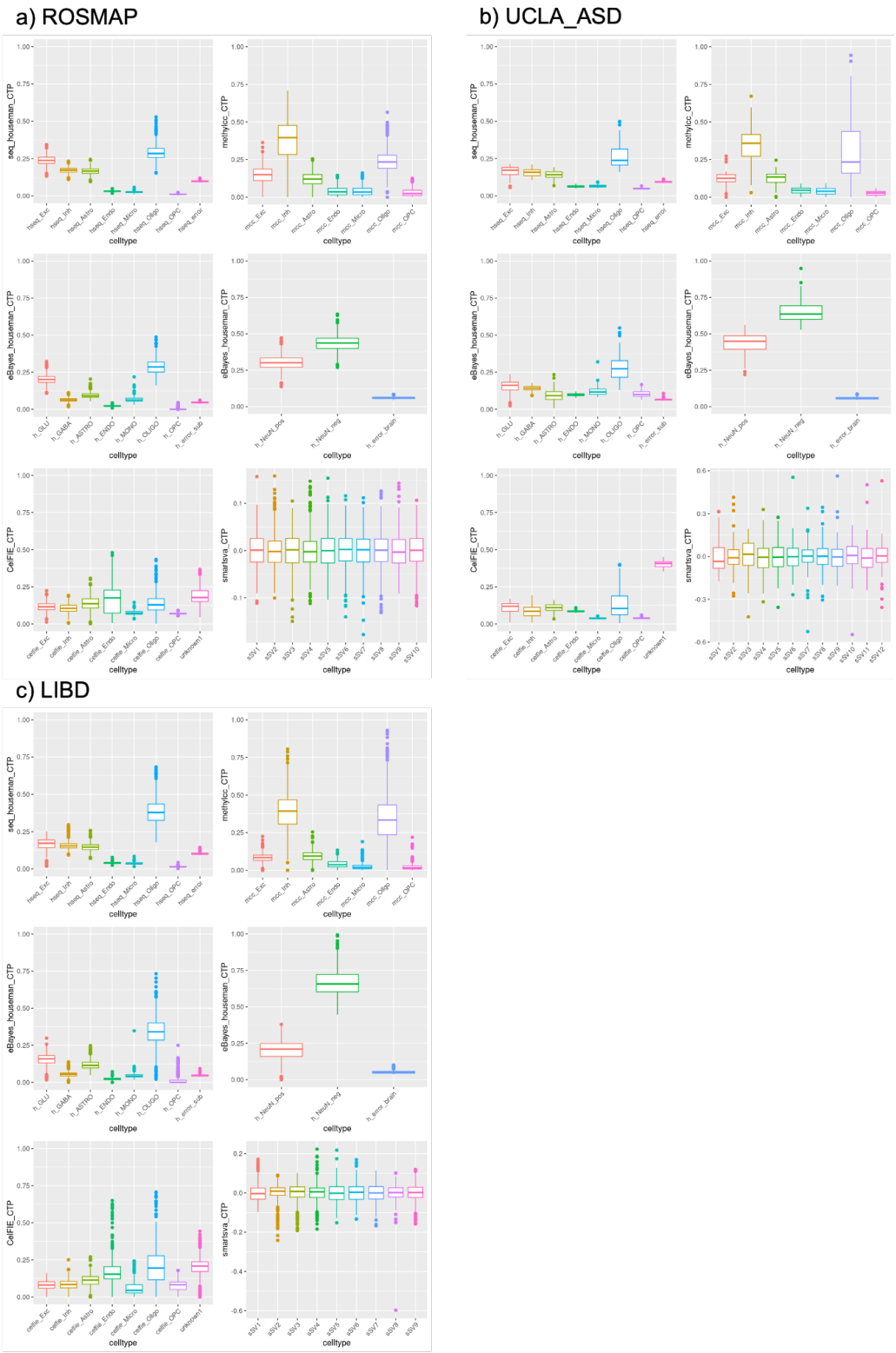


Figure S11: Comparison of brain CTP distributions from different deconvolution methods for a) ROSMAP, b) UCLA_ASD and c) LIBD studies. “seq_houseman_CTP”: Luo et al. 2023 *(24)* sequencing reference + “extremes” marker selection + Houseman algorithm; “methylcc_CTP”: Luo et al. 2023 sequencing reference + “extremes” marker selection + methylCC algorithm; “eBayes_houseman_CTP”: aggregated array reference + eBayes t-test marker selection + Houseman algorithm; “CelFiE_CTP”: Luo et al. 2023 sequencing reference + “extremes” marker selection + CelFiE algorithm; “smartSVA_CTP”: from the reference-free smartSVA method where sSV2 is plotted for ROSMAP and sSV1 is plotted for UCLA_ASD and LIBD studies. For the deconvolutions including the Houseman algorithm, the error term is the CETGYO score. The CelFIE deconvolutions here generated an “unknown” CTP that is not attributable to the specified cell-type reference marker profiles.

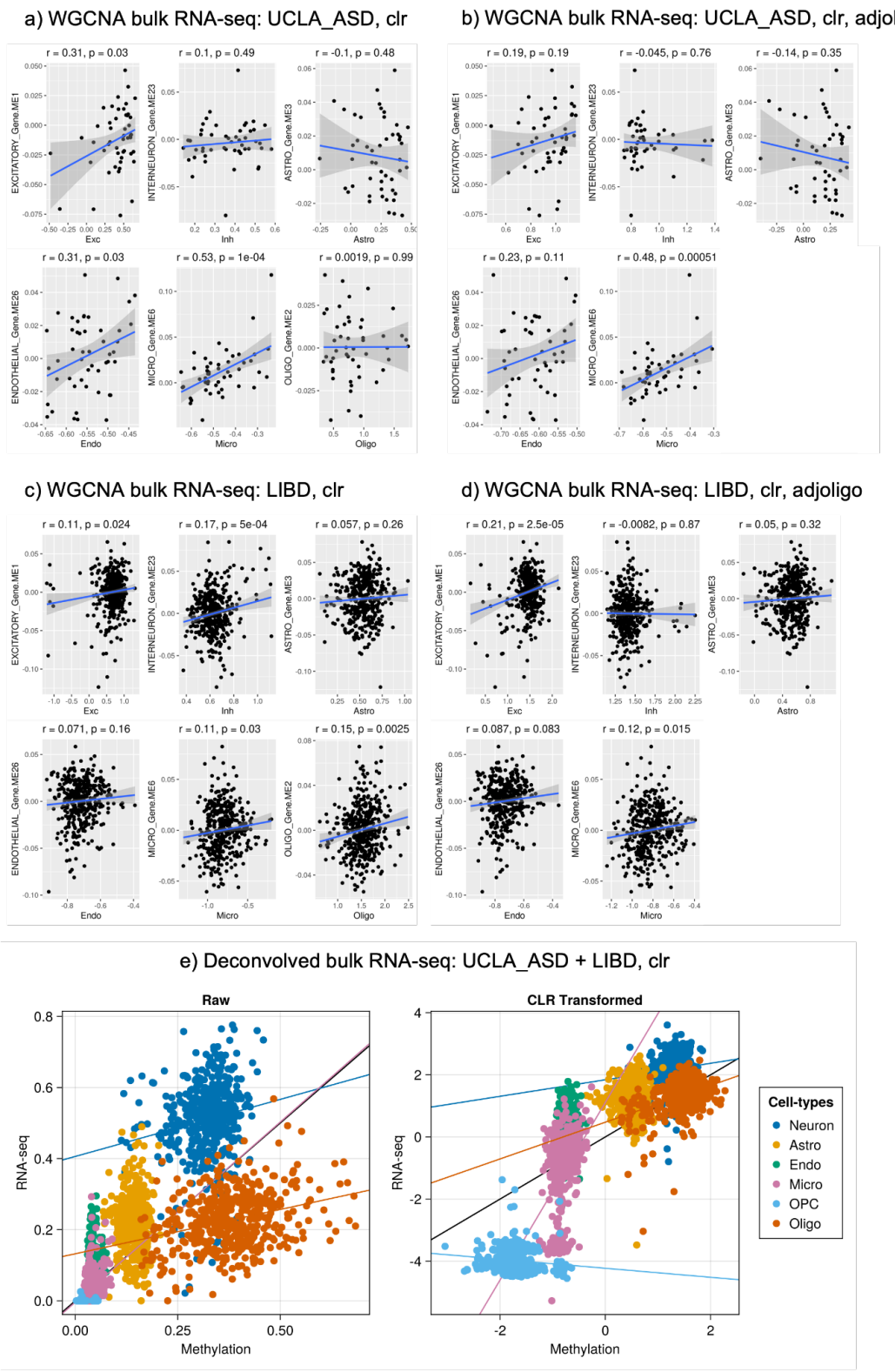


Figure S12: Scatterplots comparing brain CTPs (from the Luo et al. 2023 *(24)* sequencing + "extremes" marker selection + Houseman algorithm) from LIBD and UCLA_ASD to representations of brain CTPs from bulk RNA-seq datasets: a) UCLA_ASD bulk RNA-seq WGCNA modules from Gandal et al. 2018 *(69)*, clr-transformed; b) UCLA_ASD bulk RNA-seq WGCNA modules, clr-transformed and adjusting for oligodendrocyte proportion; c) LIBD bulk RNA-seq WGCNA modules from Gandal et al. 2018, clr-transformed; d) LIBD bulk RNA-seq WGCNA modules, clr-transformed and adjusting for oligodendrocyte proportion; e) Combined UCLA_ASD and LIBD bulk RNA-seq reference based deconvolution from the Wang et al. 2018 combined PsychENCODE dataset with clr-transformed CTPs.

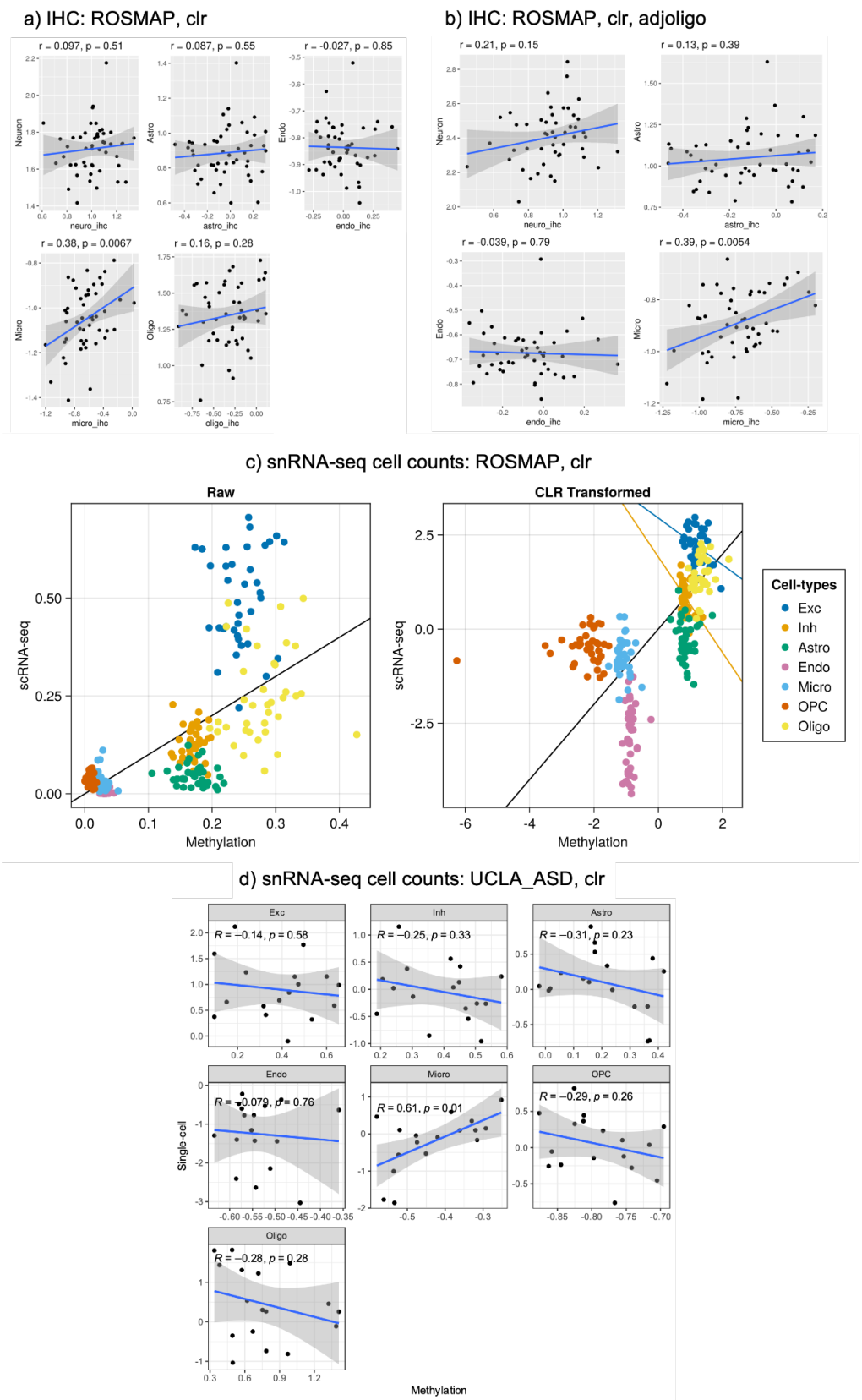


Figure S13: Scatterplots comparing brain CTPs (from the Luo et al. 2023 sequencing *(24)* + "extremes" marker selection + Houseman algorithm) from ROSMAP and UCLA_ASD to representations of brain CTPs from single-cell and immunohistochemistry datasets: a) ROSMAP immunohistochemistry data from Patrick et al. 2020 *(22)*, clr-transformed; b) ROSMAP immunohistochemistry data, clr-transformed and adjusting for oligodendrocyte proportion; c) ROSMAP single-cell RNA-seq cell counts from Mathys et al. 2019 *(33)*, clr-transformed; d) UCLA_ASD single-cell RNA-seq cell counts from Wamsley et al. 2023 *(34)*, clr-transformed.

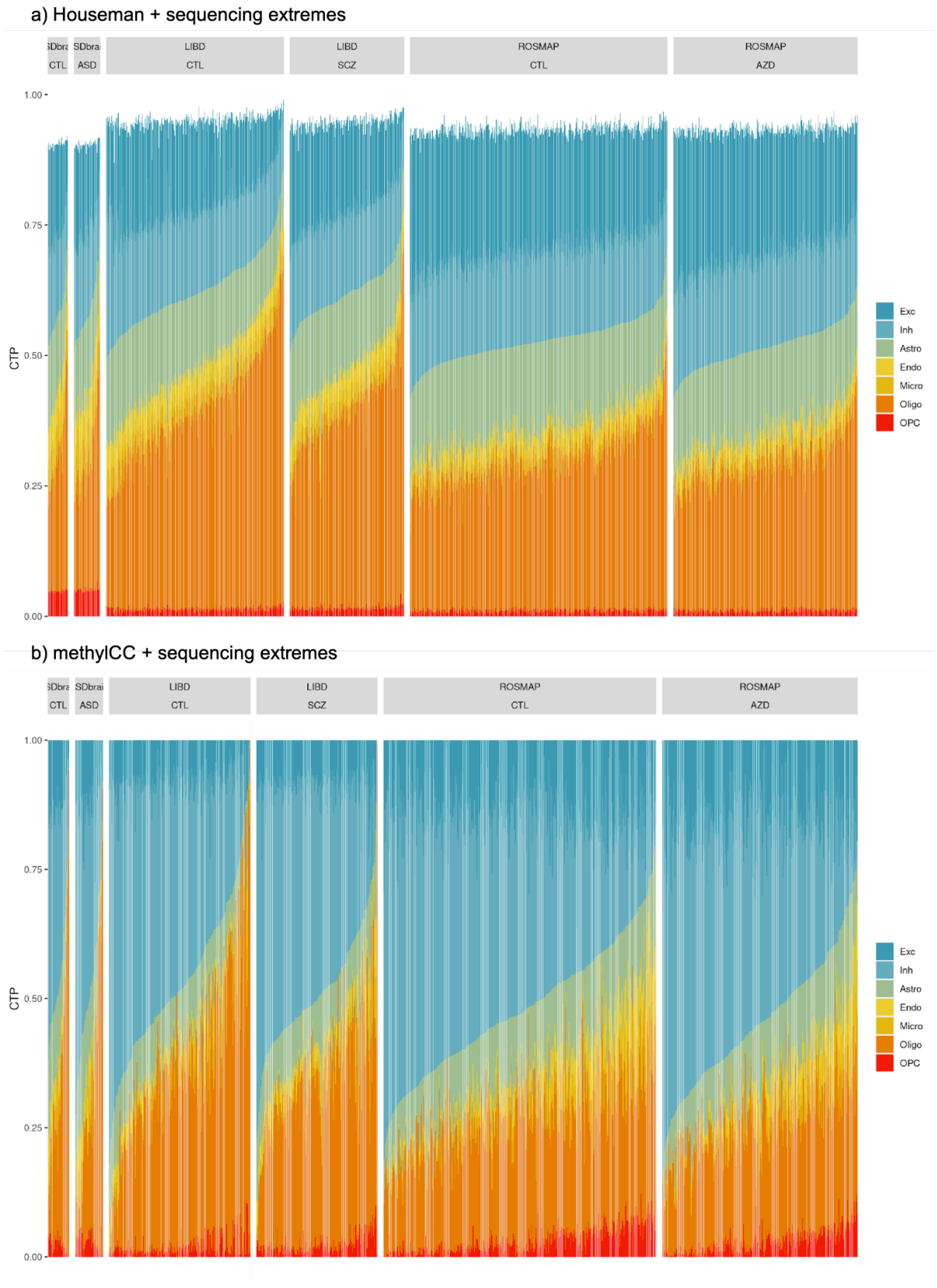


Figure S14: Per-individual (x-axis) deconvolved brain CTP profiles from a) the sequencing reference + "extremes" marker selection + Houseman algorithm approach and b) the Luo et al. 2023 sequencing reference + "extremes" marker selection + methylCC algorithm approach. Note that the Houseman algorithm provides a narrower distribution of CTP estimates than the methylCC algorithm.

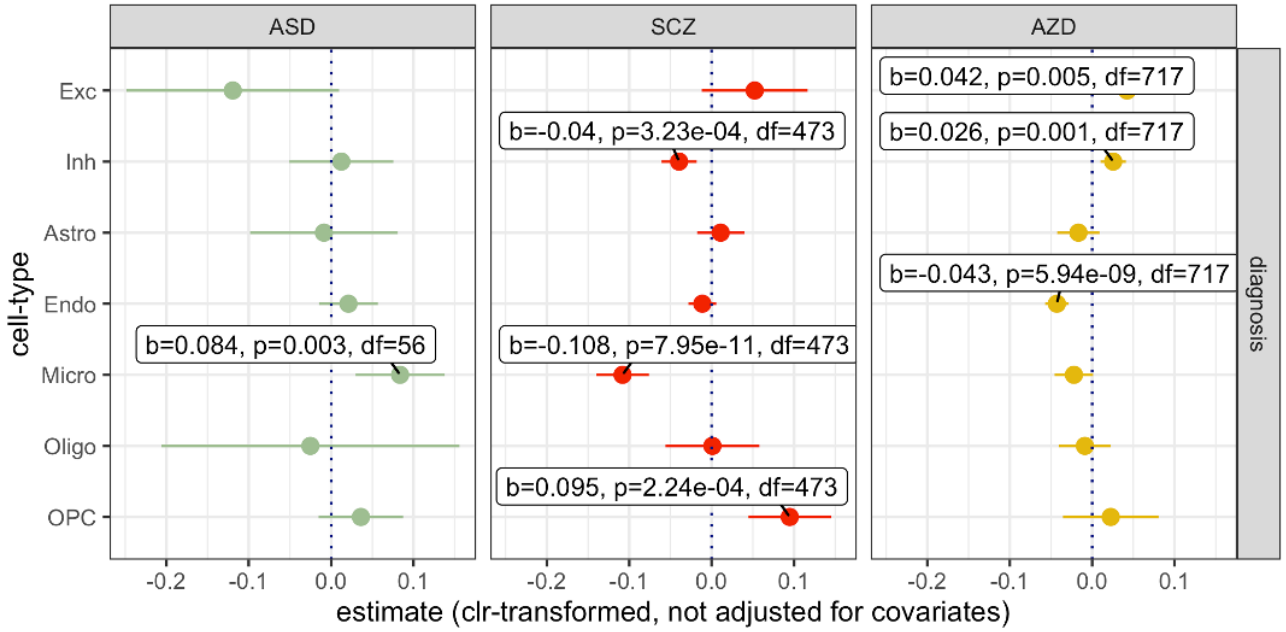


Figure S15: Coefficients +/- 95% CI of a linear model for brain CTP (clr-transformed) ~ diagnosis without covariates. Labelled points show associations with p<0.05.

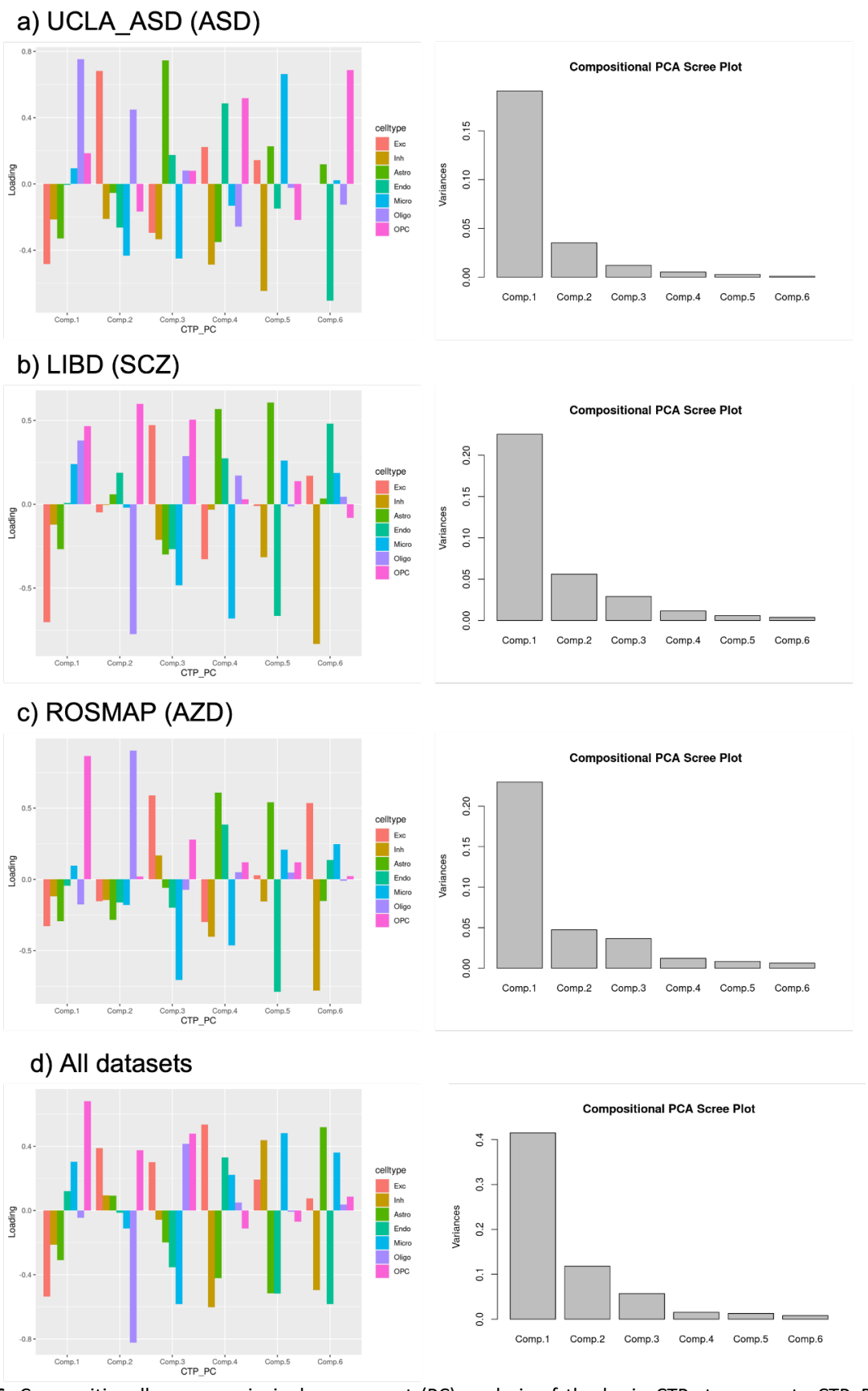


Figure S16: Compositionally aware principal component (PC) analysis of the brain CTPs to generate CTP_PCs, for a) UCLA_ASD, b) LIBD, c) ROSMAP, and d) all studies combined. Depicted are the loadings of each of CTP_PCs (left), and scree plot for variance explained by each of the PCs (right).

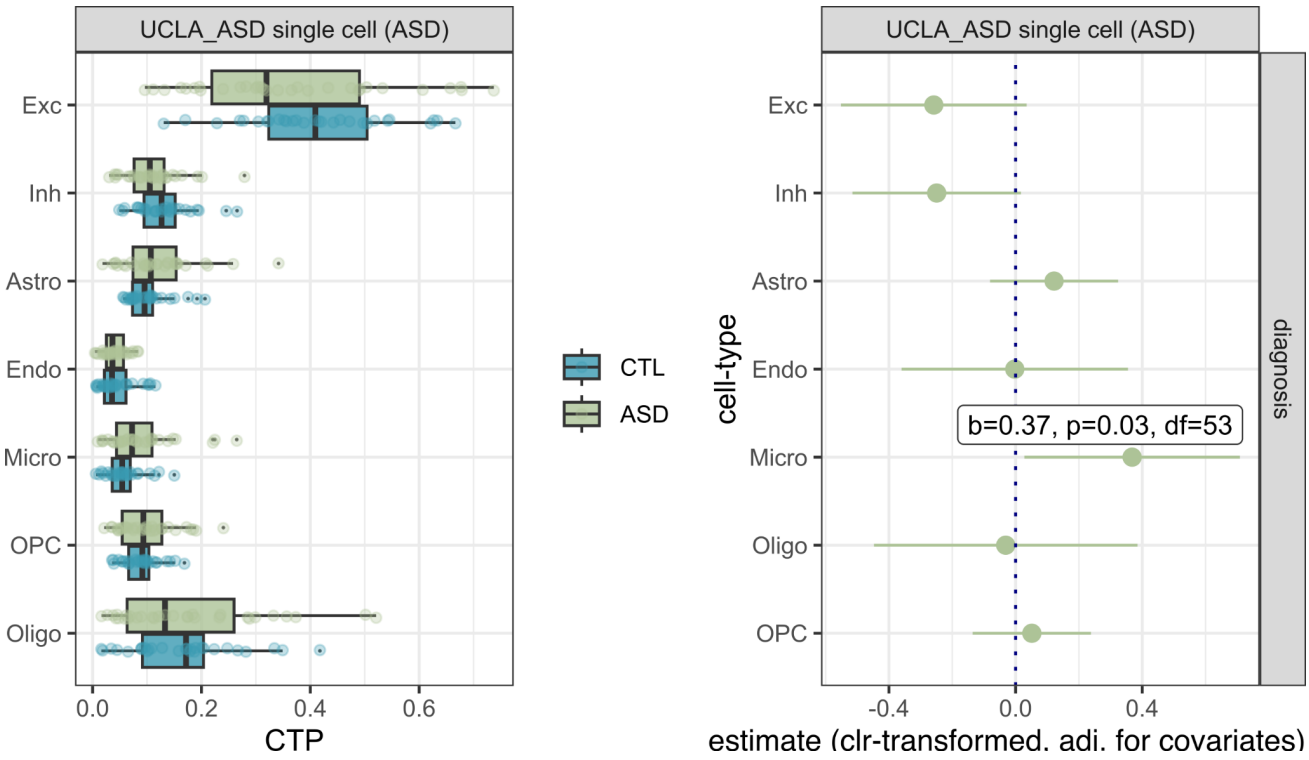


Figure S17: Re-analysis of single cell counts of brain cell-types from n=60 UCLA_ASD participants from Wamsley et al. 2023 *(34)* . Left panel: Boxplot of brain CTPs. Right panels: Diagnosis coefficients for AZD relative to controls (+/- 95% CI) from linear models of brain CTP (clr-transformed) ~ diagnosis + age + age^2 + sex + brain region. Statistically significant results are labelled (p<0.05).

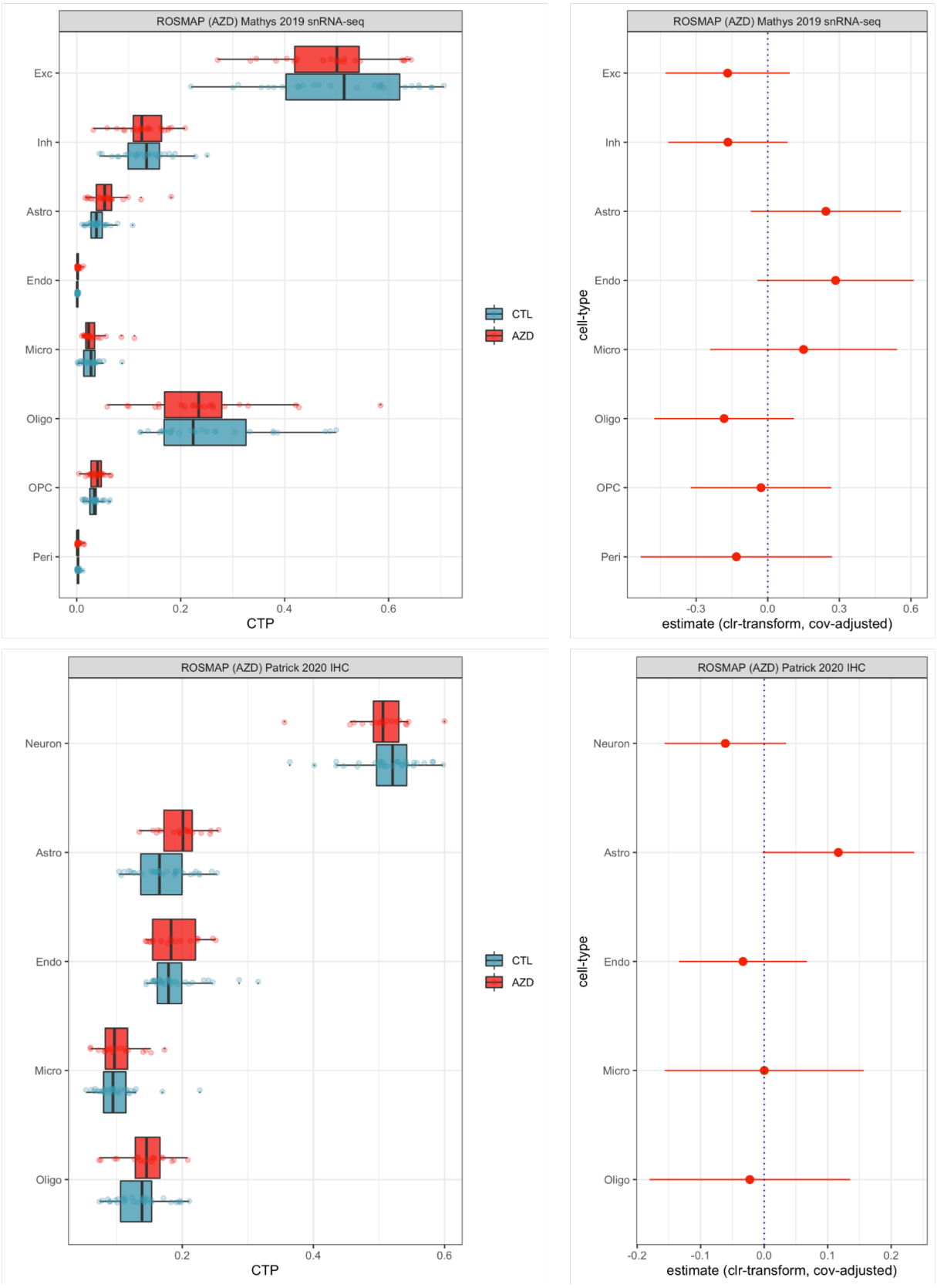


Figure S18: Re-analysis of snRNA-seq-derived brain CTPs from n=48 ROSMAP participants (n=24 AZD, n=24 undiagnosed) from Mathys et al. 2019 *(33)* (upper) and n=49 ROSMAP participants (n=18 AZD, n=31 undiagnosed) from Patrick et al. 2020 *(22)* (lower). Left panels: Boxplot of brain CTPs. Right panels: Diagnosis coefficients for AZD relative to controls (+/- 95% CI) from linear models of brain CTP (clr-transformed) ~ diagnosis + age + age^2 + sex. No results are nominally-significant ($p<0.05$).

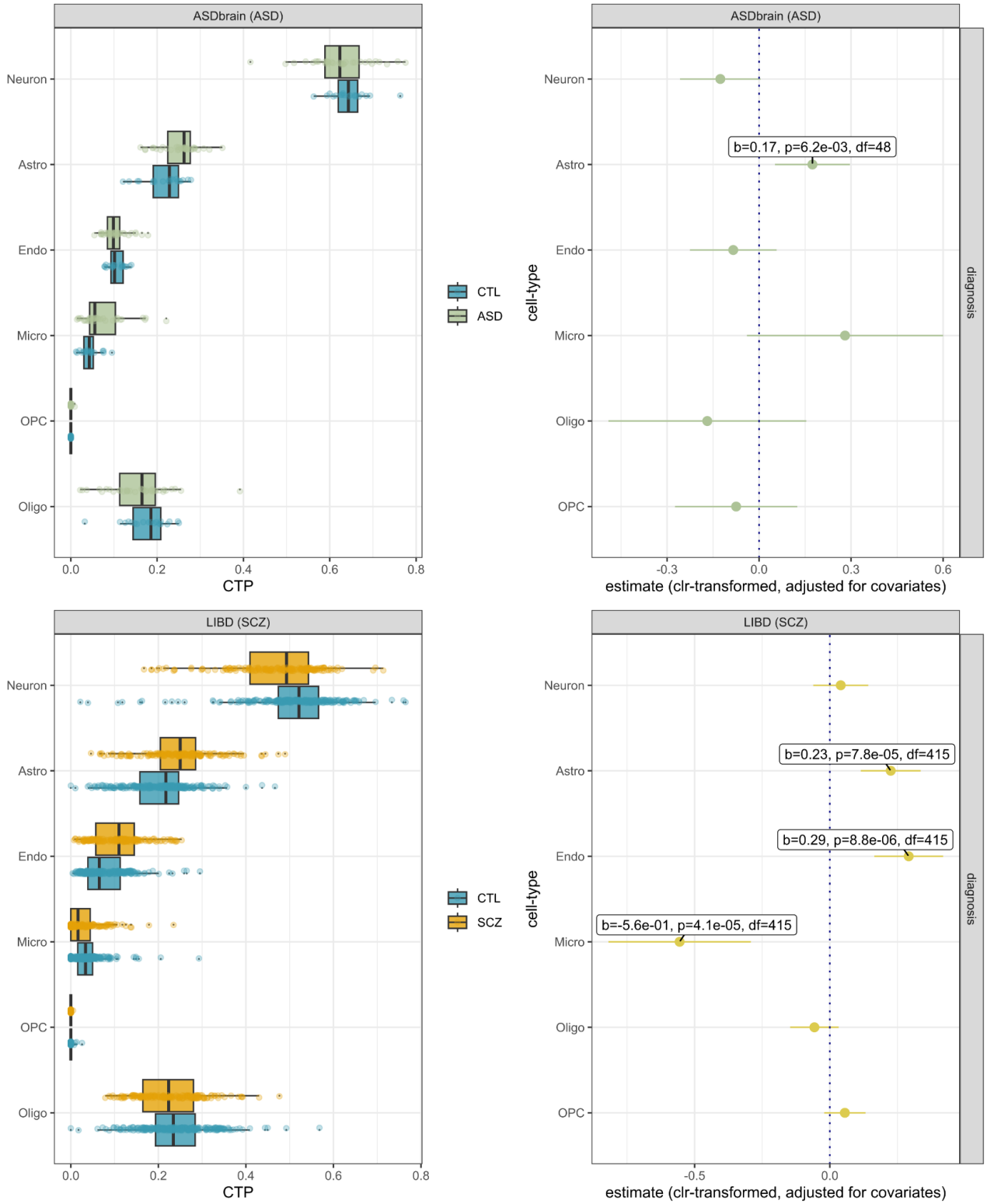


Figure S19: Diagnostic associations using CTPs deconvolved from bulk RNA-seq from Wang et al. 2018 *(7)* including samples from UCLA_ASD (upper) and LIBD (lower). Left panels: Boxplot of brain CTPs. Right panels: Diagnosis coefficients for diagnosis relative to undiagnosed controls (95% CI) from linear models of brain CTP (clr-transformed) ~ diagnosis + age + age^2 + sex. Statistically significant results ($p<0.05$) are labelled.

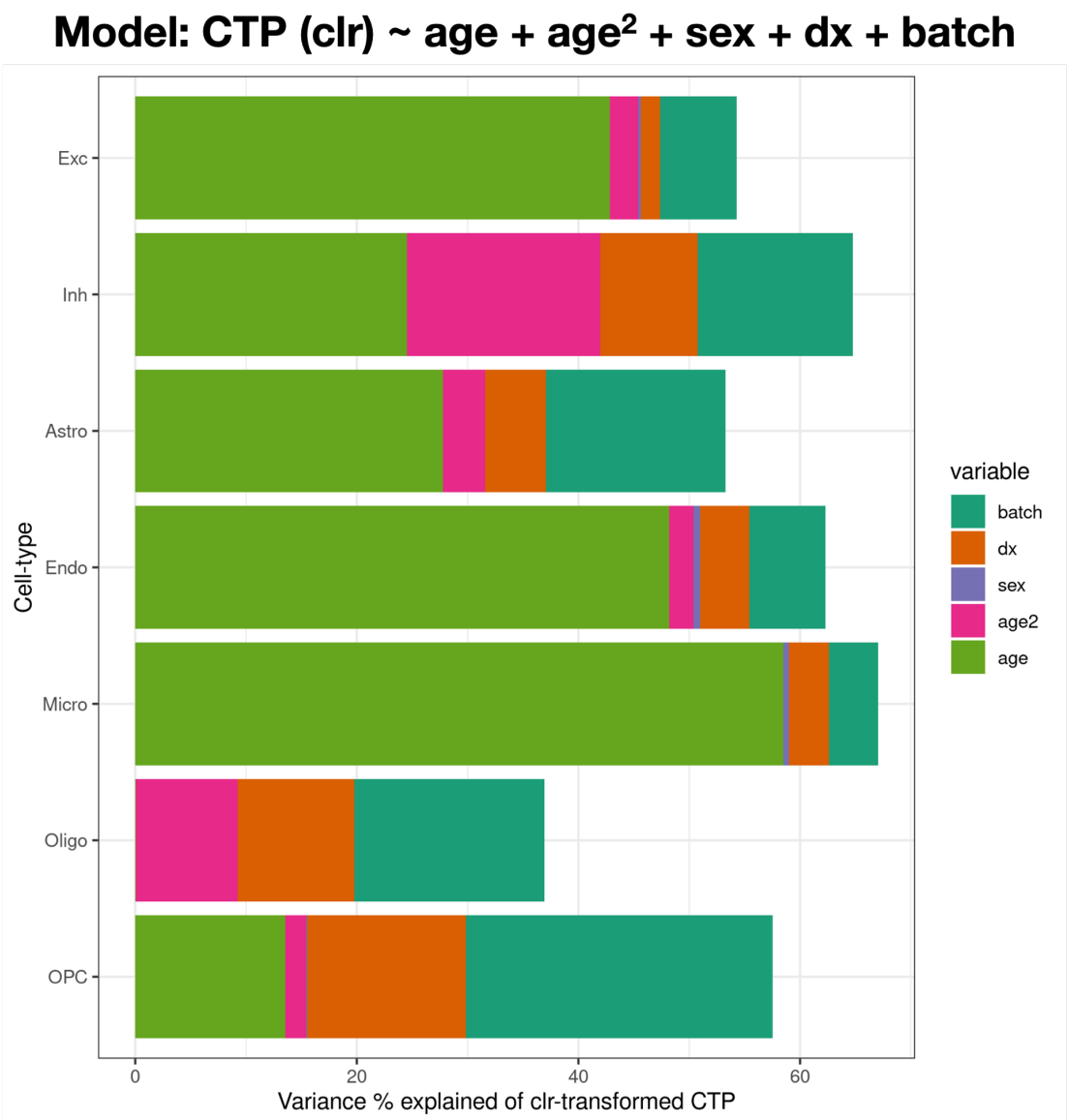


Figure S20: Variance explained of clr-transformed brain CTPs by each variable included in the linear models: in order, age, age^2, sex, diagnosis (Alzheimer's disease, schizophrenia, autism, or control) and batch in the full dataset (ROSMAP + UCLA_ASD + LIBD). Barplots show variance components from an ANOVA model of brain CTP (clr-transformed) ~ age + age^2 + sex + diagnosis + batch. Note that diagnosis is confounded with batch (as each study ascertained for a single diagnosis), and also with age (particularly for ROSMAP which focused on older individuals). Thus, we included batch as the last term here as the order of terms is relevant in an ANOVA model. Note that in other analyses reported in the main text, we jointly modelled batch and diagnosis.

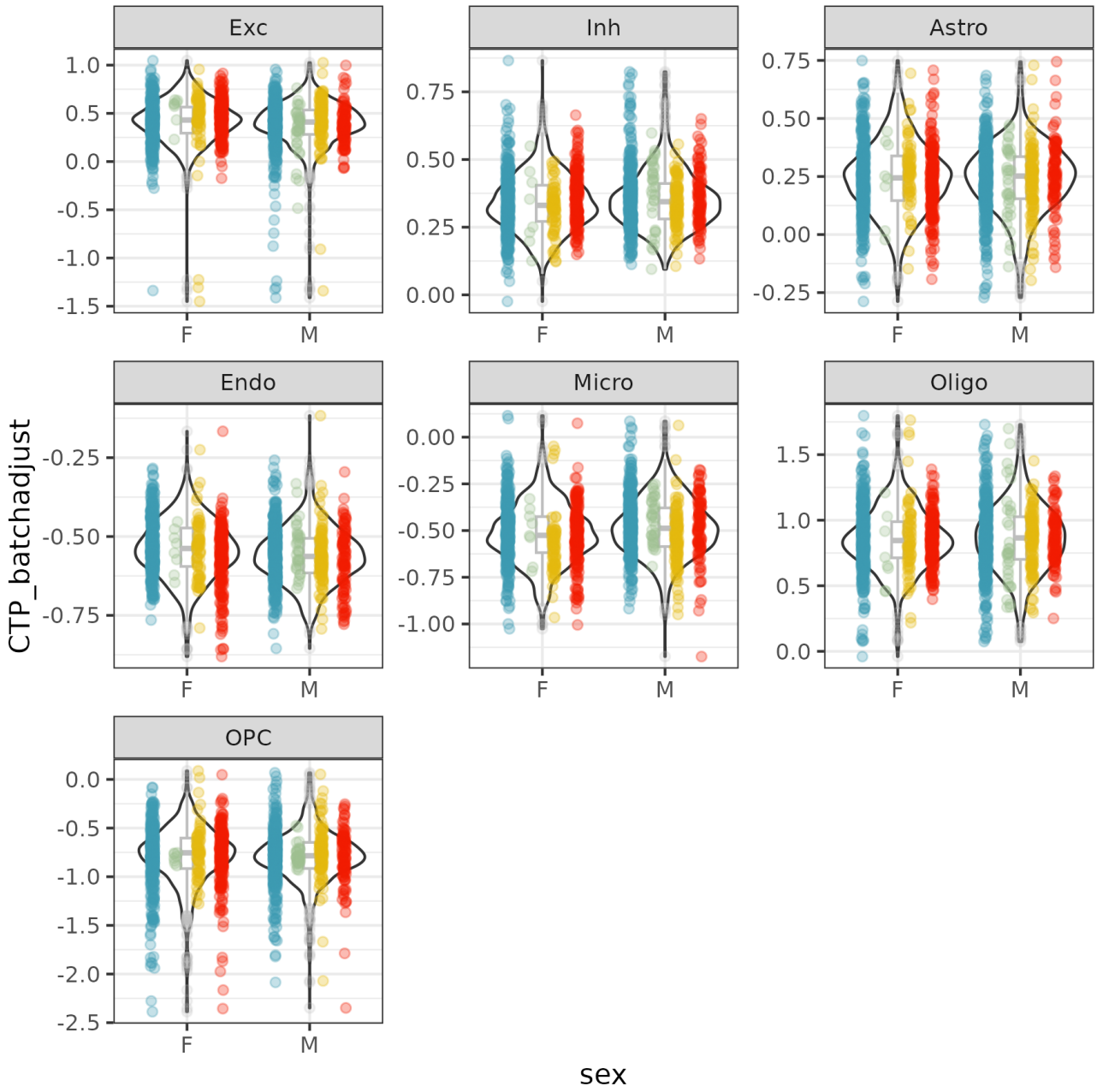


Figure S21: Violin plot showing distribution of brain CTPs between sexes (female versus male). Colours denote diagnosis (blue: undiagnosed participants; green: ASD; yellow: schizophrenia; red: Alzheimer's disease). Brain CTPs are adjusted for batch.

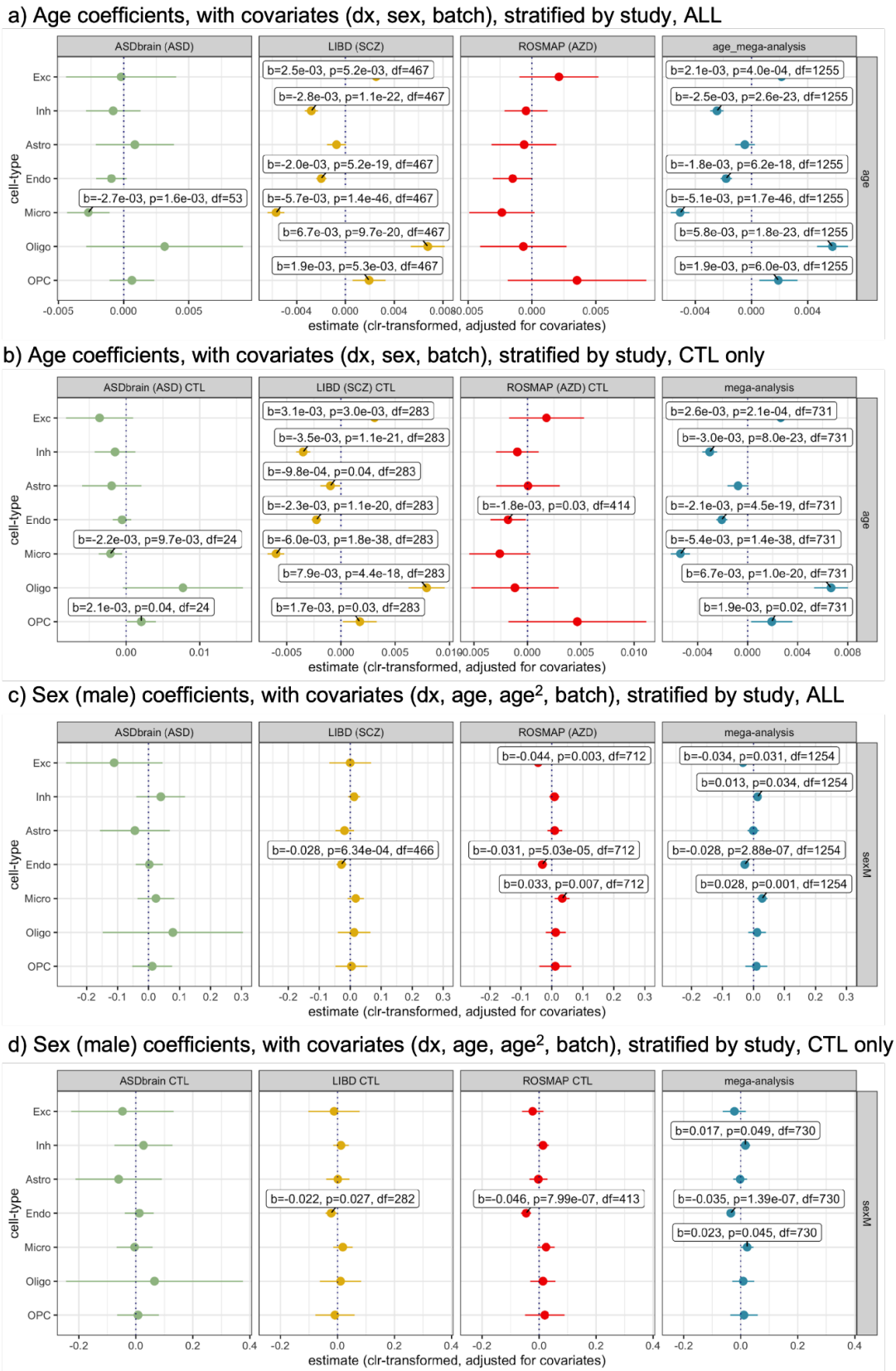


Figure S22: Sensitivity analyses for the relationship between brain CTPs versus age and sex. Age coefficients (+/- 95% CI) for a linear model of brain CTP (clr-transformed) ~ age + sex + batch + diagnosis, including all individuals (a) and only including undiagnosed individuals (b). Sex (effects shown for male sex) coefficients (+/- 95% CI) for a linear model of brain CTP (clr-transformed) ~ sex + age + age^2 + batch + diagnosis, including all individuals (c) and only including undiagnosed individuals (d). Panels show stratification by different studies, with "mega-analysis" referring to the aggregated analysis. These results show that directions of age and sex effects are usually concordant across studies.

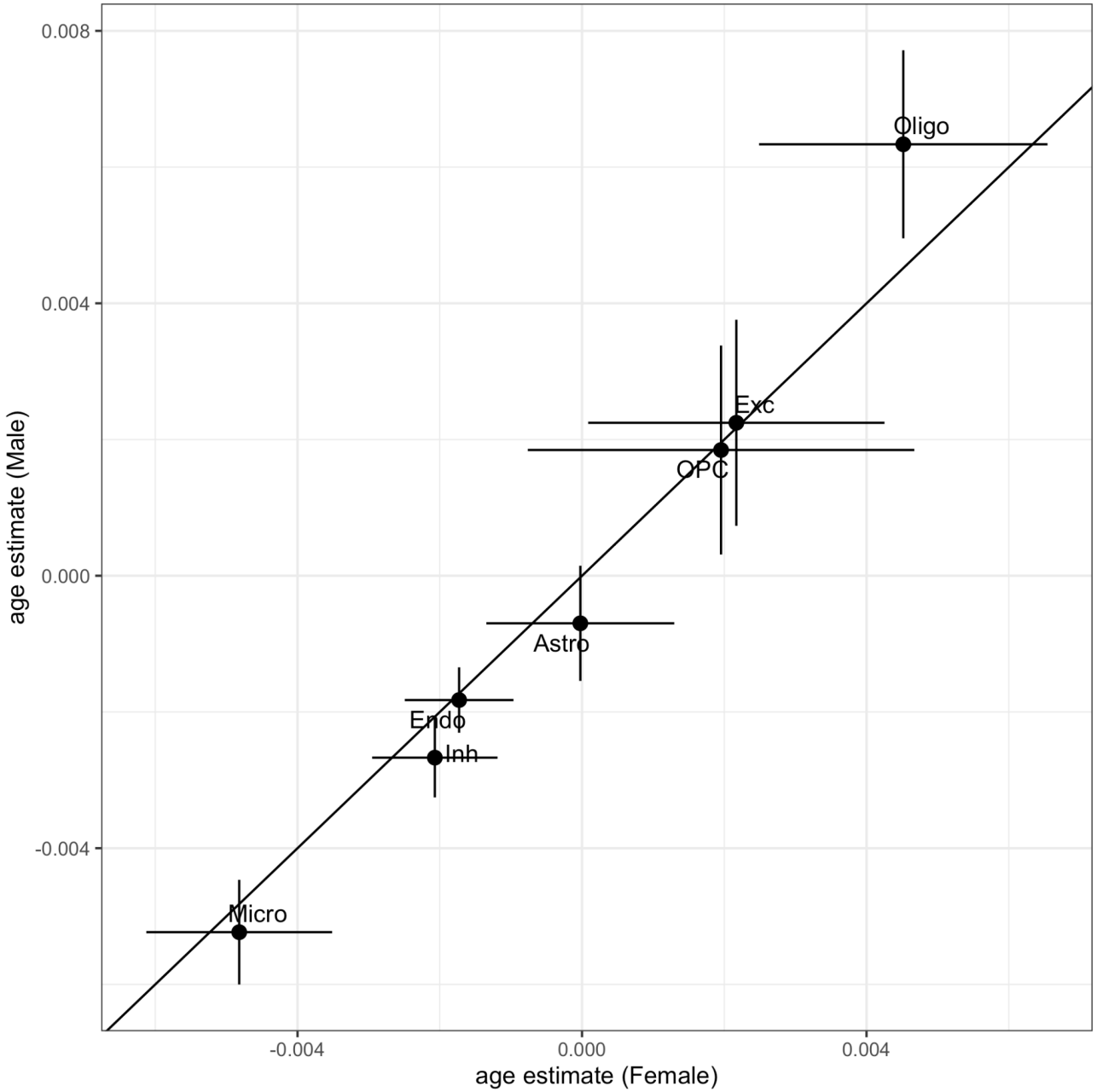


Figure S23: Age-CTP effect sizes from linear models, stratified by sex. Point denotes relationship between male-specific and female-specific estimates. Bars denote 95% confidence interval. Diagonal line represents a 1:1 relationship between the female and male age-CTP estimates.

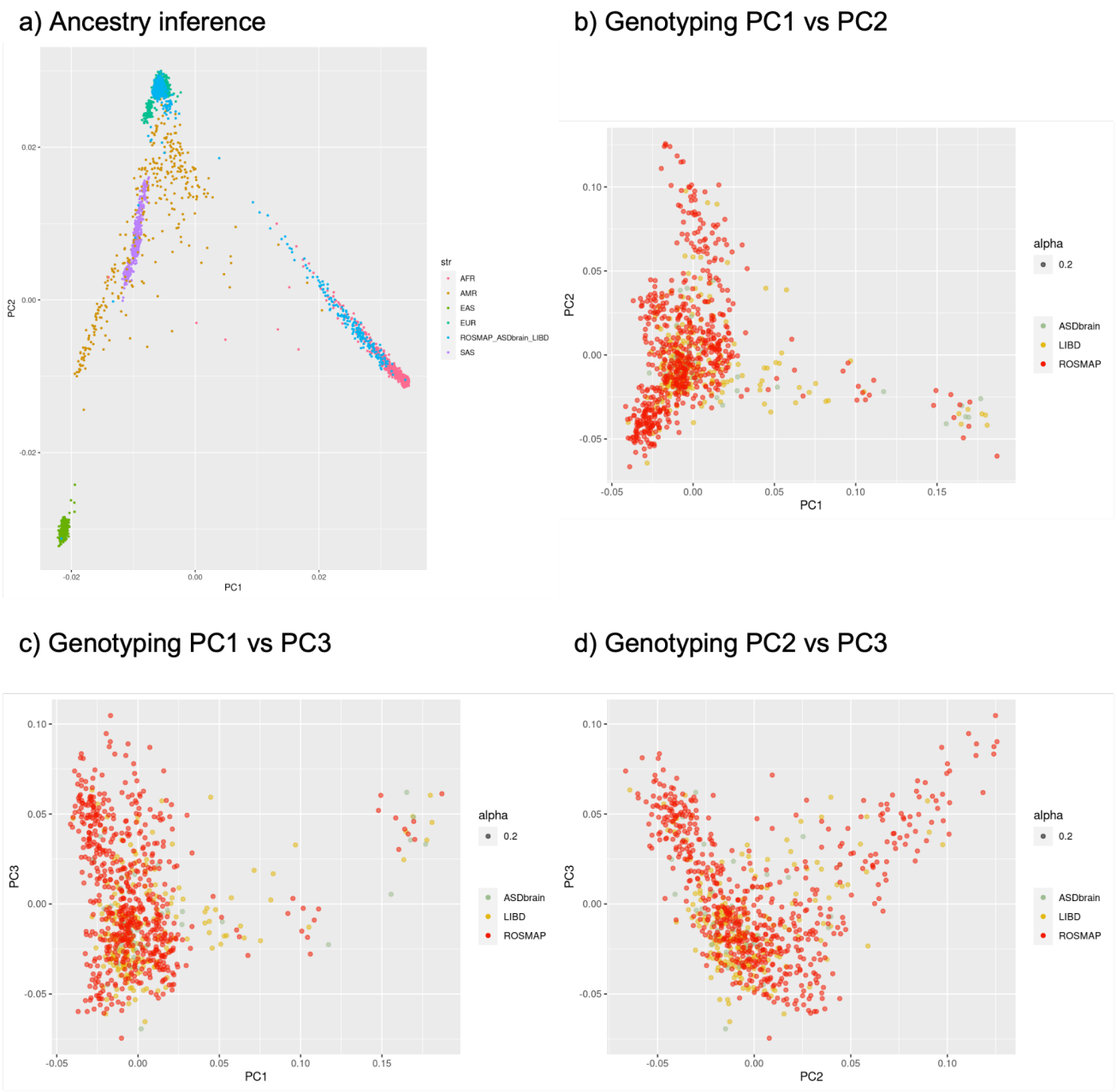


Figure S24: Ancestry plots. a) Genetic relatedness matrix (GRM) of SNP genotypes projected onto the first two principal components of the 1000G reference dataset (HapMap 3 SNPs, minor allele frequency > 0.05). b-d) Plots visualising the first 3 principal components of the European dataset, which were used to represent population stratification in downstream analyses. The GRM used to generate these PCs was filtered for Europeans, LD pruned, and all included individuals had genetic relatedness < 0.05.

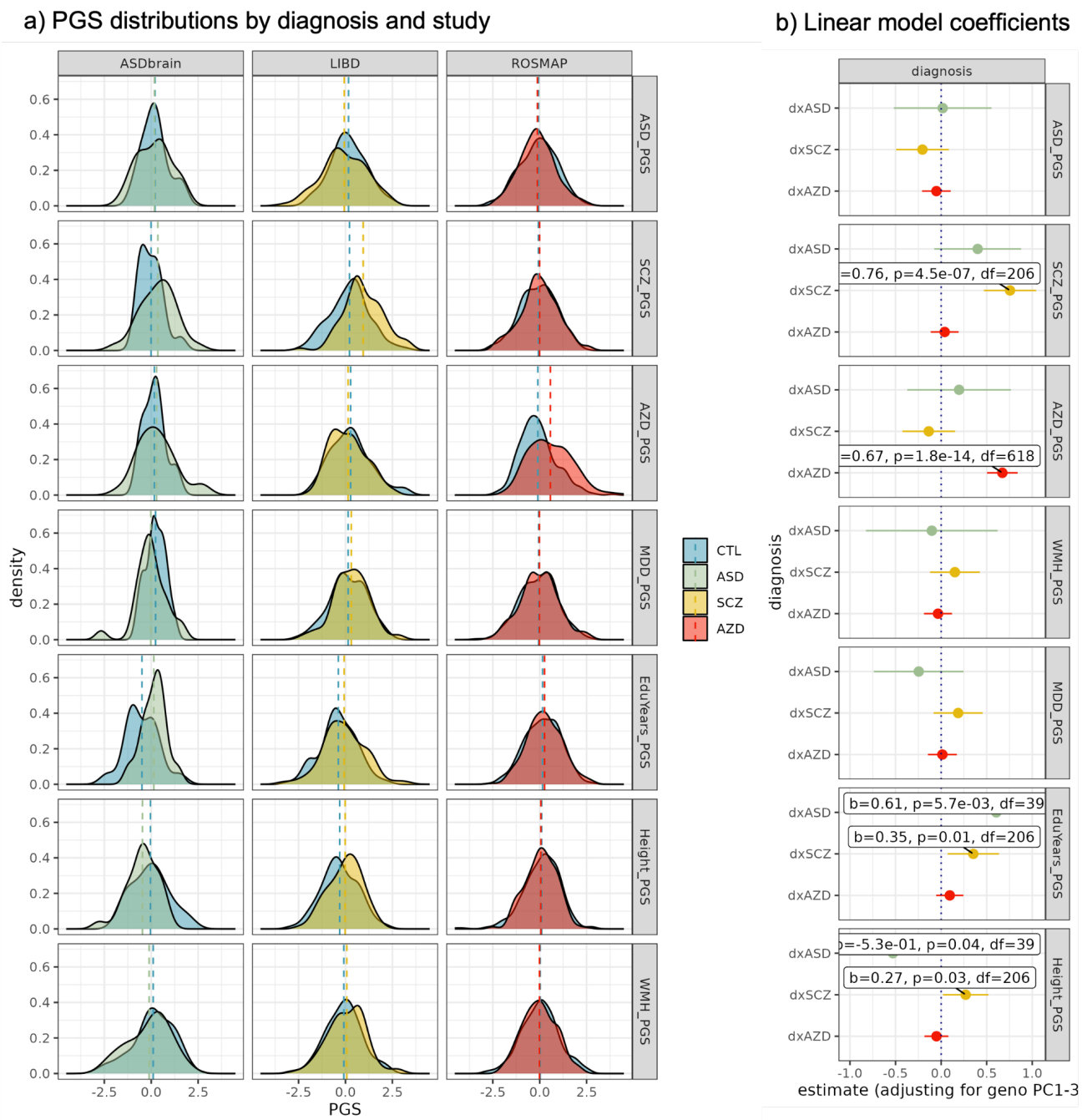


Figure S25: Polygenic score (PGS) prediction of diagnosis. a) PGS distributions for diagnosed versus undiagnosed individuals, stratified by study, and b) Diagnosis coefficients (+/- 95% CI) for a linear model of diagnosis ~ PGS + genotyping PC1-3. Models were run separately for each diagnosis. We found nominal associations between reduced height PGS in the individuals diagnosed with ASD and with schizophrenia, though these associations were weak in comparison to PGS prediction of schizophrenia and Alzheimer's disease. We also found that participants diagnosed with ASD and schizophrenia had significantly higher educational attainment PGS after adjusting for genotyping PCs.

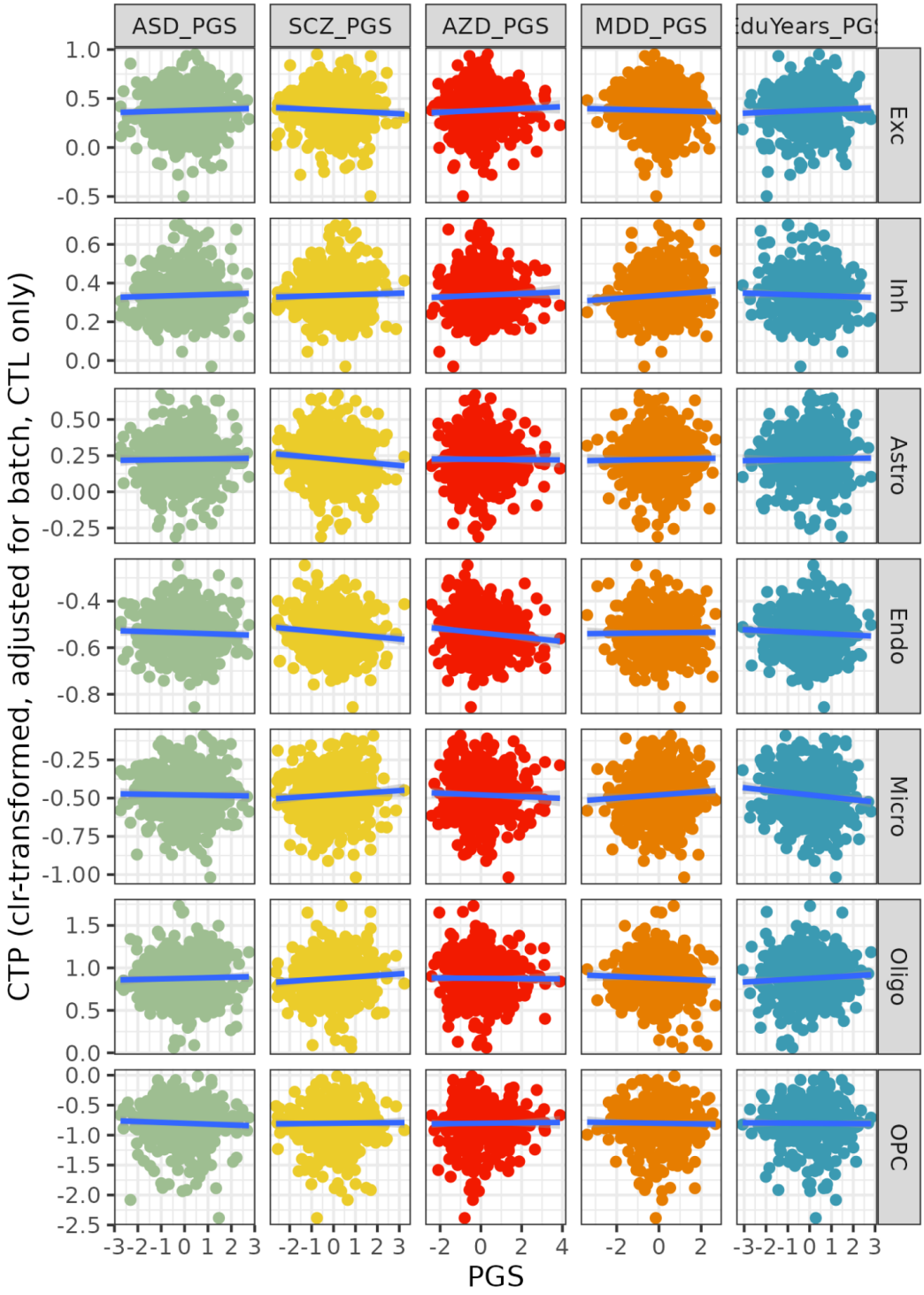


Figure S26: Scatterplot of brain CTP in rows (clr-transformed, adjusted for batch, undiagnosed participants only) versus neuropsychiatric PGS in columns.

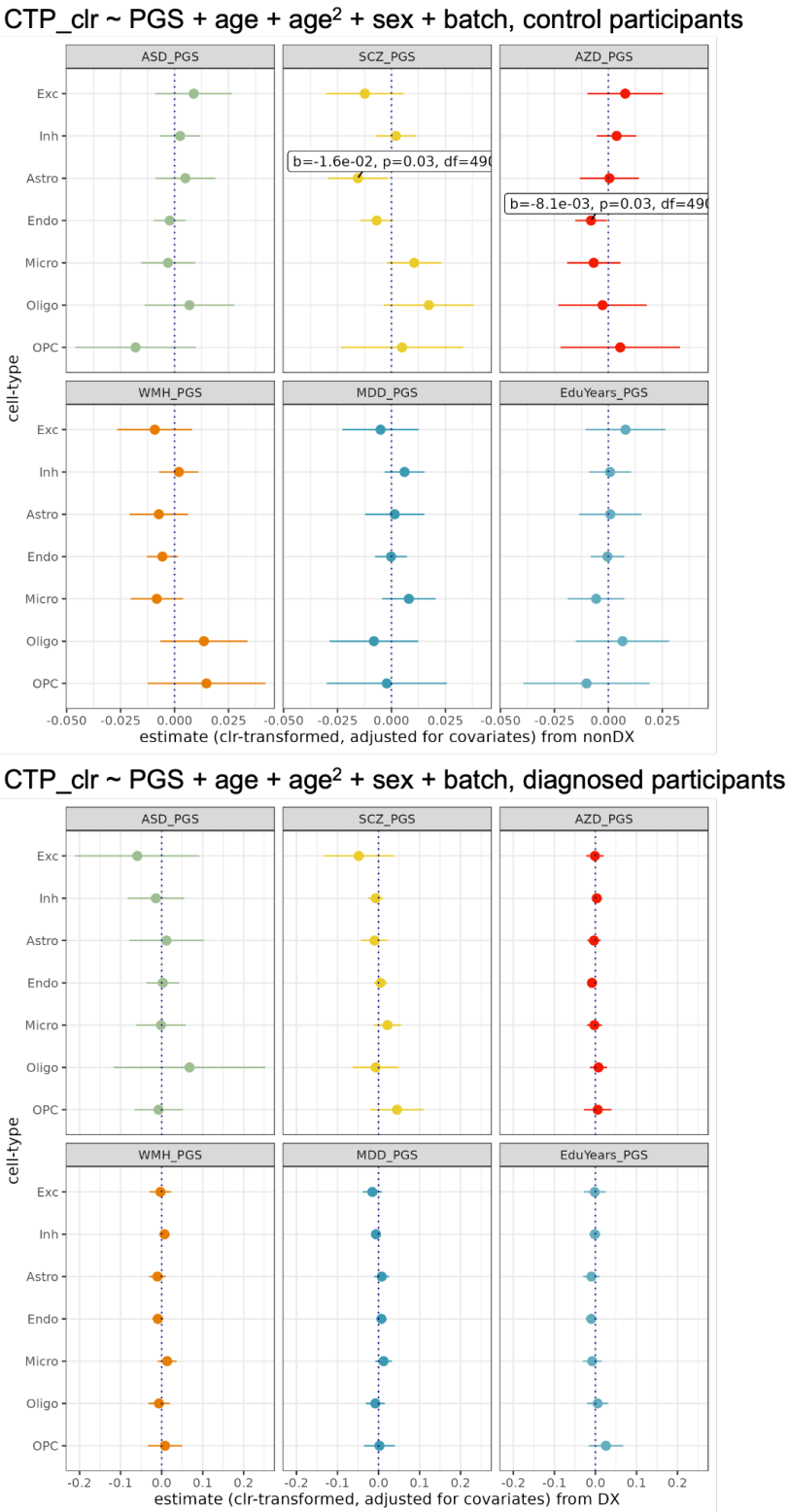


Figure S27: Sensitivity analyses for associations between brain CTPs and neuropsychiatric trait PGS in linear models including a) undiagnosed participants only, and b) diagnosed participants only. PGS coefficients (+/- 95% CI) from a linear model of brain CTPs (clr-transformed) ~ neuropsychiatric PGS + age + age^2 + sex + batch + genotyping PC1-3. Panels represent the different neuropsychiatric PGS being tested. For the analysis including diagnosed participants only: the ASD_PGS, SCZ_PGS and AZD_PGS analysis only included individuals with that respective neuropsychiatric diagnosis. For the WMH_PGS, MDD_PGS and EduYears_PGS analysis, the sensitivity analysis restricted the sample to individuals diagnosed with Alzheimer's disease due to their relevance to this diagnosis.

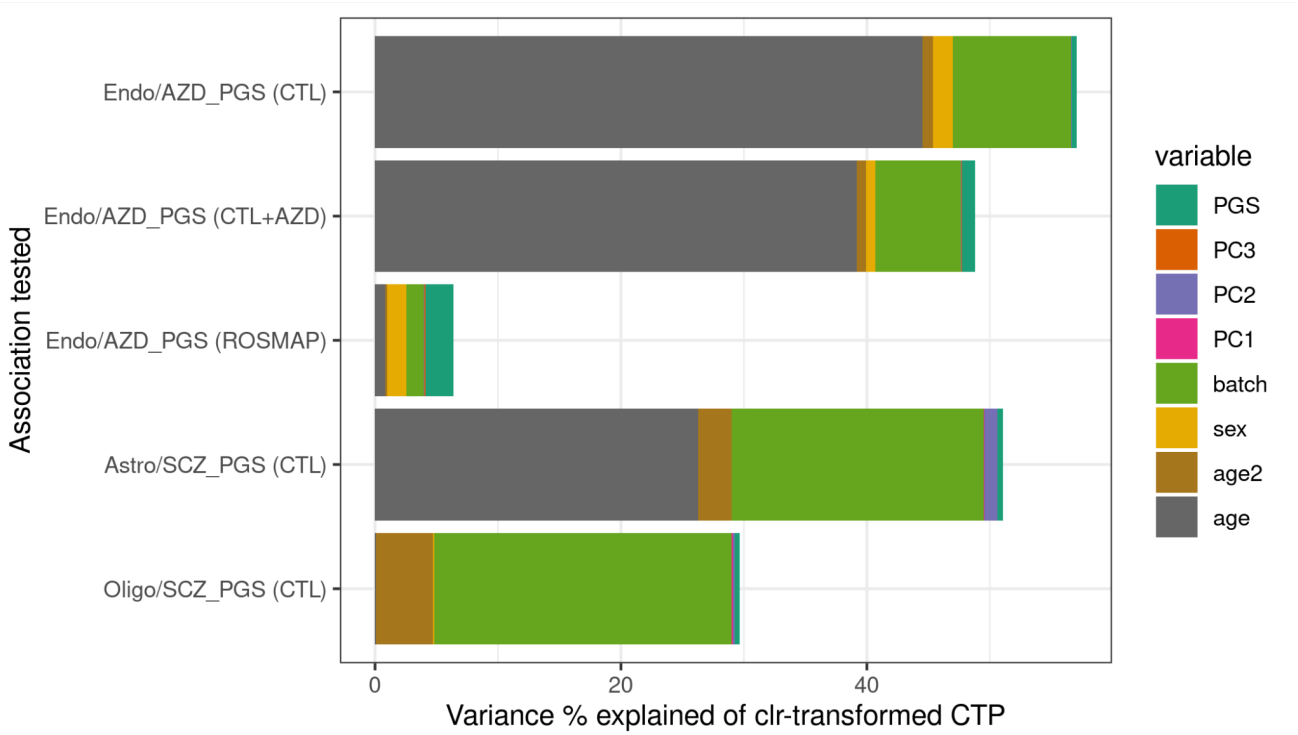


Figure S28: Variance explained of clr-transformed brain CTPs from each of the variables in the model, for significant CTP-PGS associations ($p<0.05$) in Figure 3b. The ANOVA model was constructed as brain CTP (clr-transformed) ~ age + age^2 + sex + batch + genotyping PC1-3 + neuropsychiatric PGS. Text in brackets denotes which individuals were included in that ANOVA model, as the demographic variables of the sample changes the variance explained by each covariate.

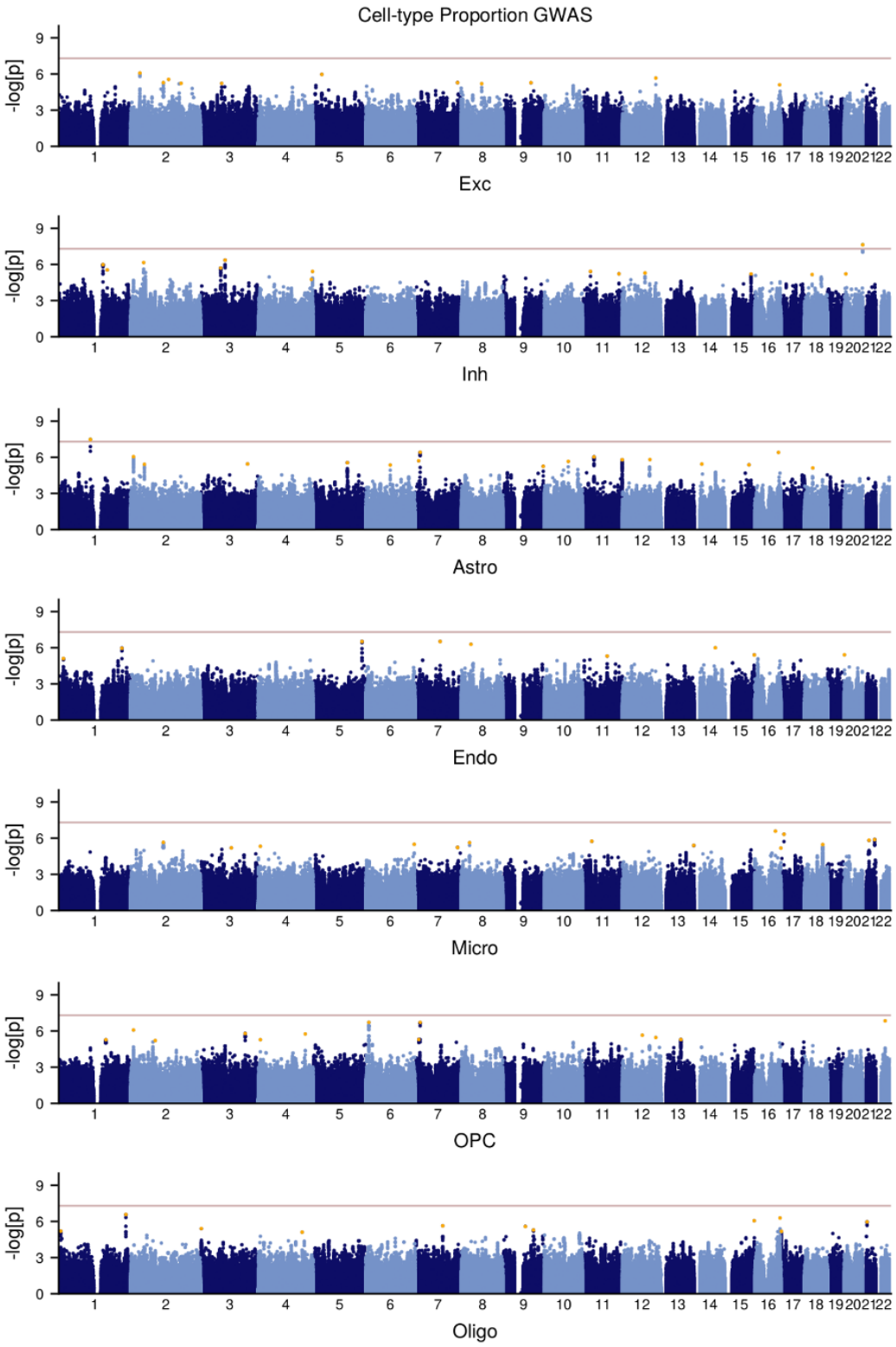


Figure S29: Manhattan plots for linear genome-wide association studies for each brain CTP (inverse normal transformation (INT) after CLR-transformation), taking brain CTPs as the response variable, and including all n=873 European individuals with matched genotyping and DNA methylation data. Fixed effect covariates were age + age^2 + sex + batch + diagnosis. Red line denotes p<5e-8 threshold, orange dots are independent loci (identified using GCTA-COJO) using a less stringent threshold of p<1e-5.

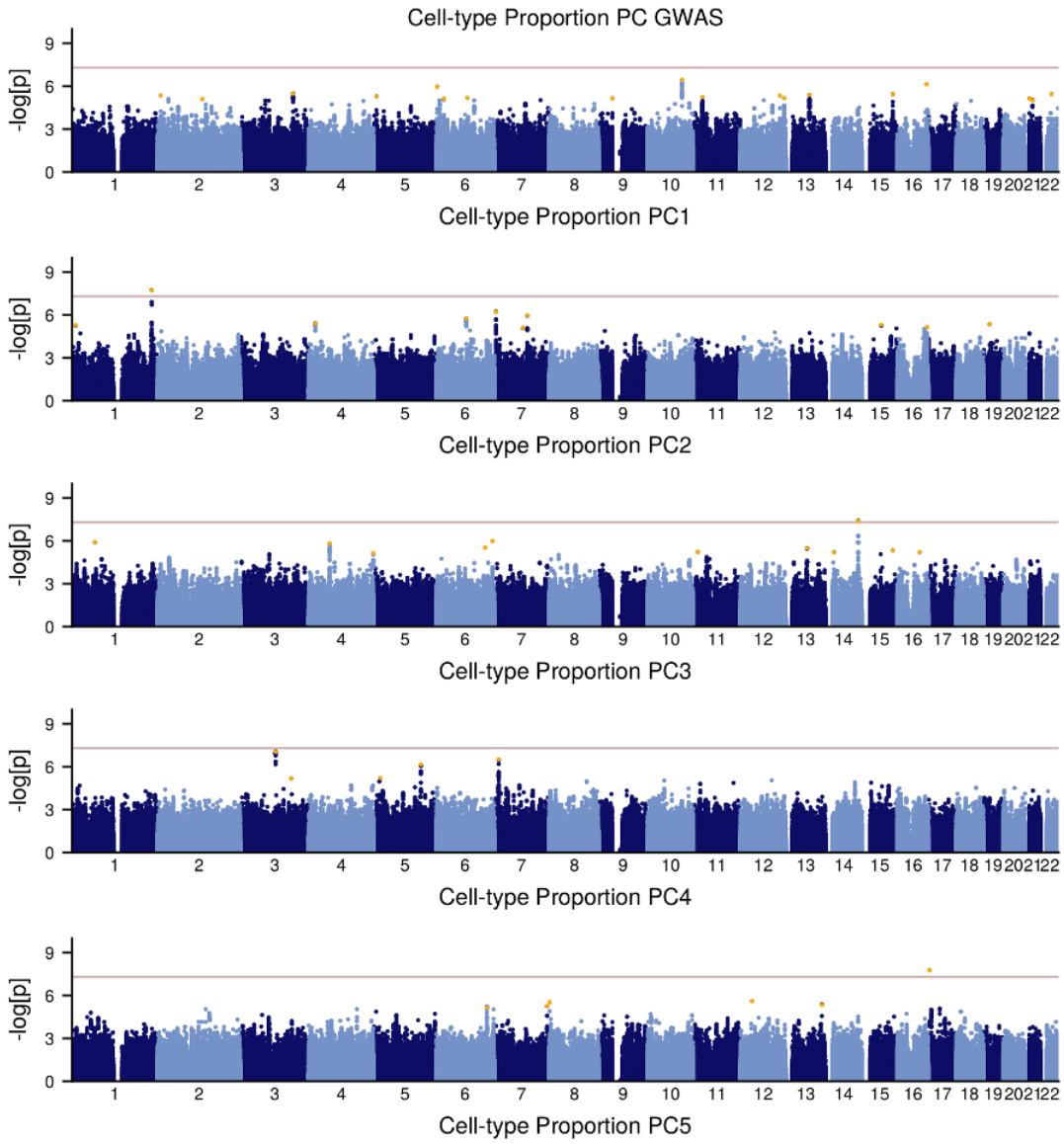


Figure S30: Manhattan plots for linear genome-wide association studies for each brain CTP, taking the inverse normal transformation (INT) of compositionally-aware principal components of the brain CTPs ("CTP_PCs") as the response variable, and including all n=873 European individuals with matched genotyping and DNA methylation data. Fixed effect covariates were age + age^2 + sex + batch + diagnosis. Red line denotes p<5e-8 threshold, orange dots are independent loci (identified using GCTA-COJO) using a less stringent threshold of p<1e-5.

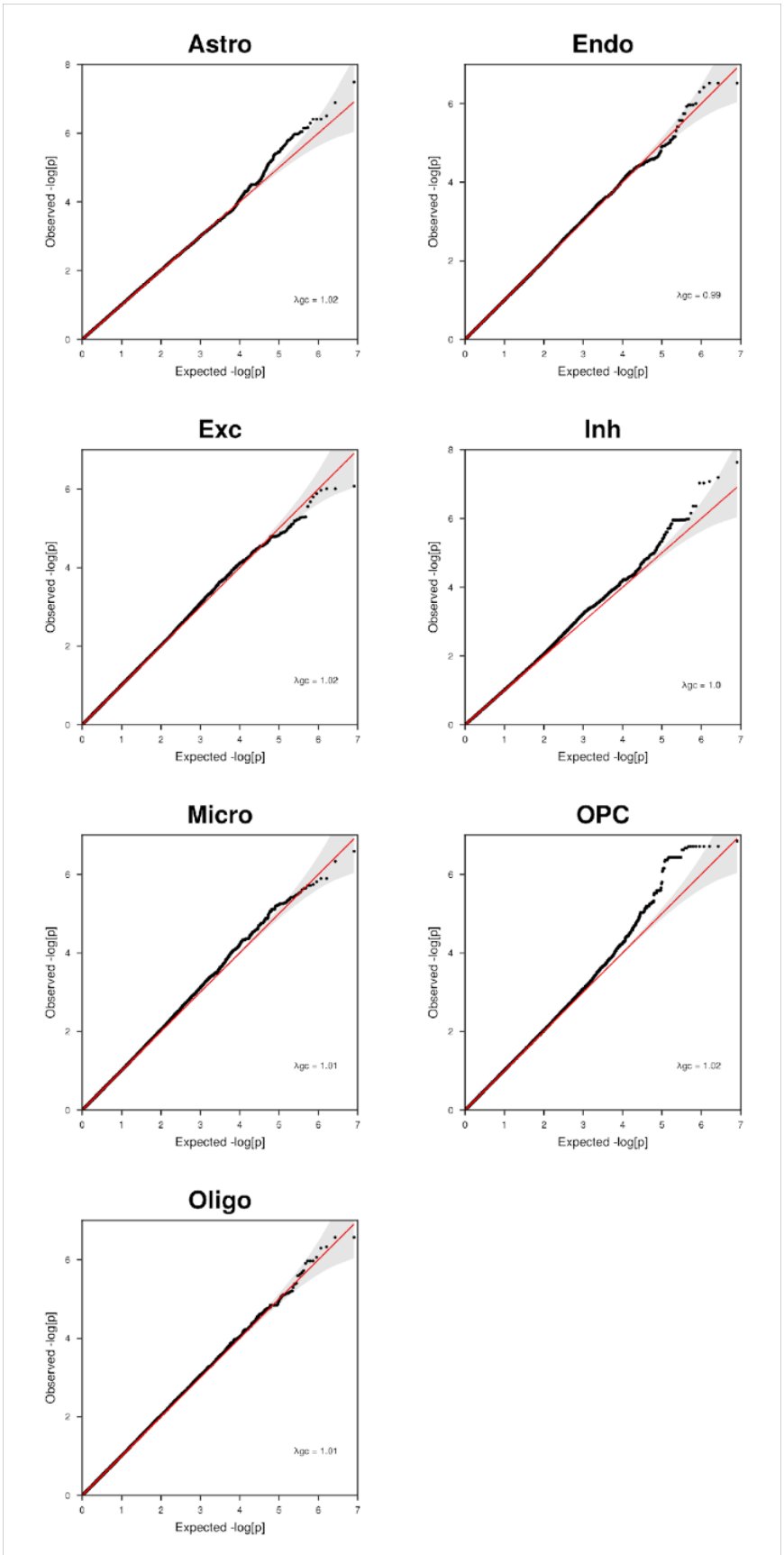


Figure S31: QQ-plots for linear genome-wide association studies, taking clr-transformed brain CTPs (with subsequent inverse normal transformation) as the response variable, and including all individuals. Fixed effect covariates were age + age^2 + sex + batch + diagnosis.

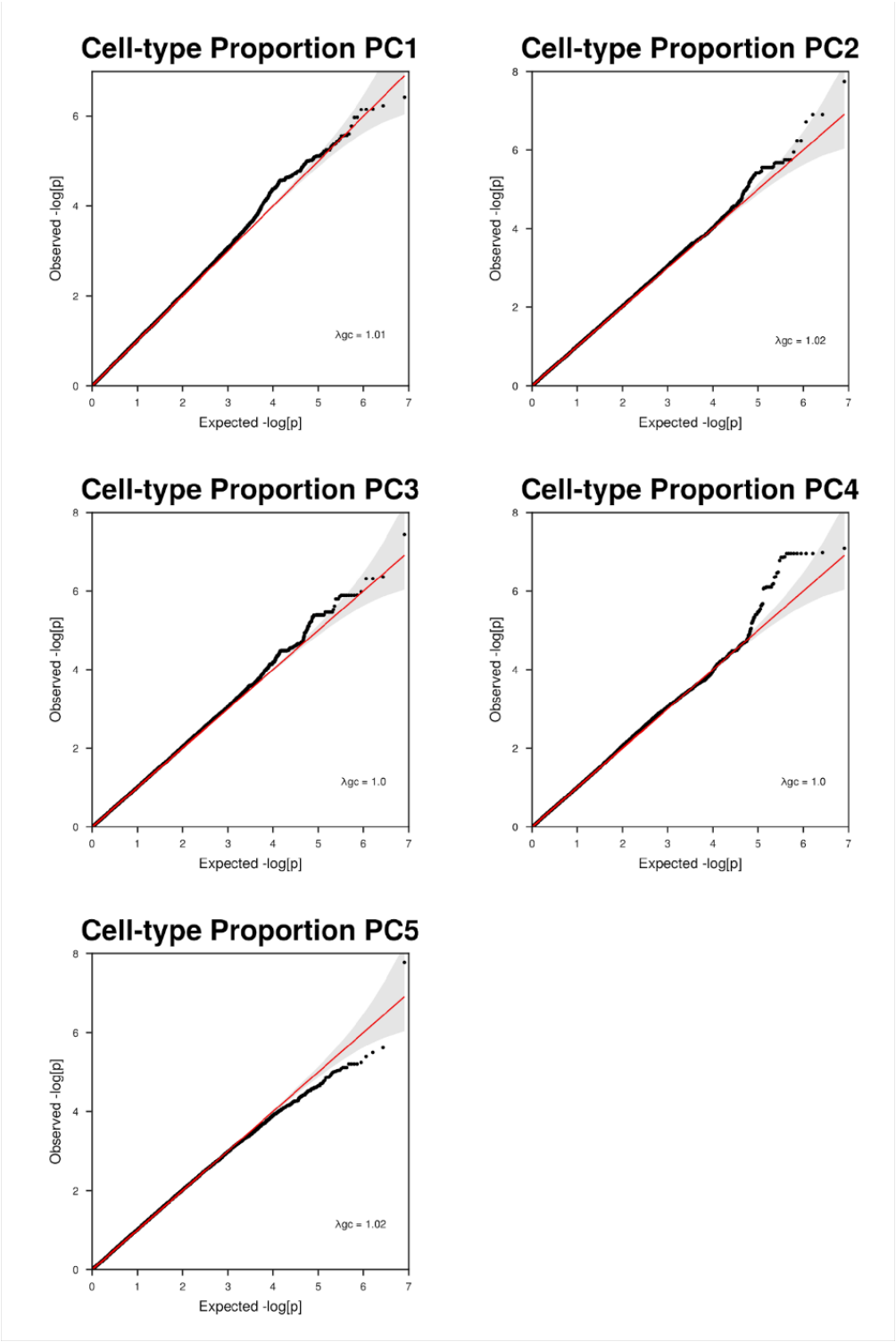


Figure S32: QQ-plots for linear model genome-wide association studies, taking CTP_PCs (with subsequent inverse normal transformation) as the response variable, and including all individuals. Fixed effect covariates were age + age^2 + sex + batch + diagnosis.

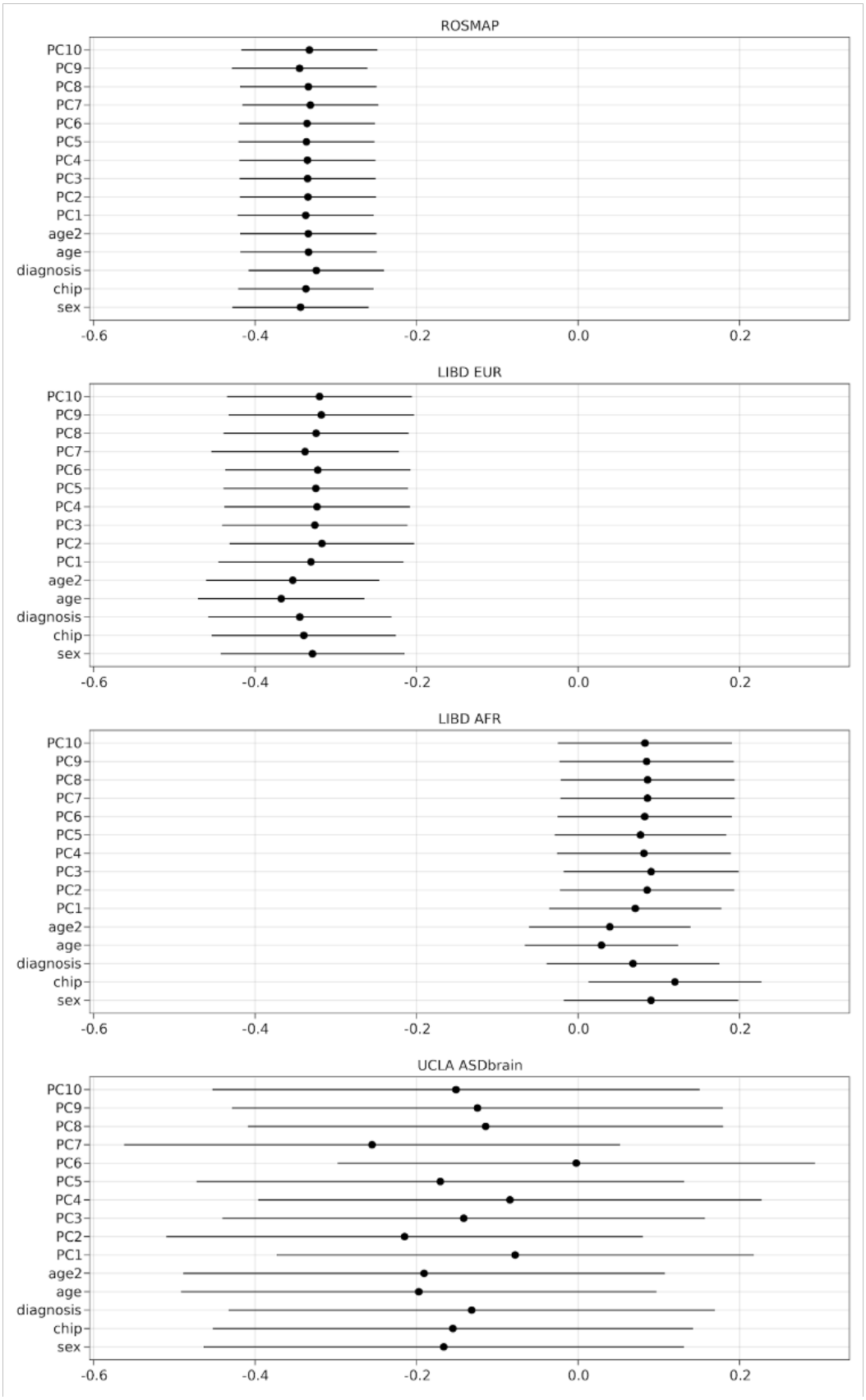


Figure S33: Meta-analysis forest plot for rs6011327 genotype regressed against inhibitory neuron CTP adjusting for each single covariate shown on the left axis, with one panel for each of the ROSMAP (all EUR), LIBD_EUR, LIBD_AFR, UCLA_ASD studies.

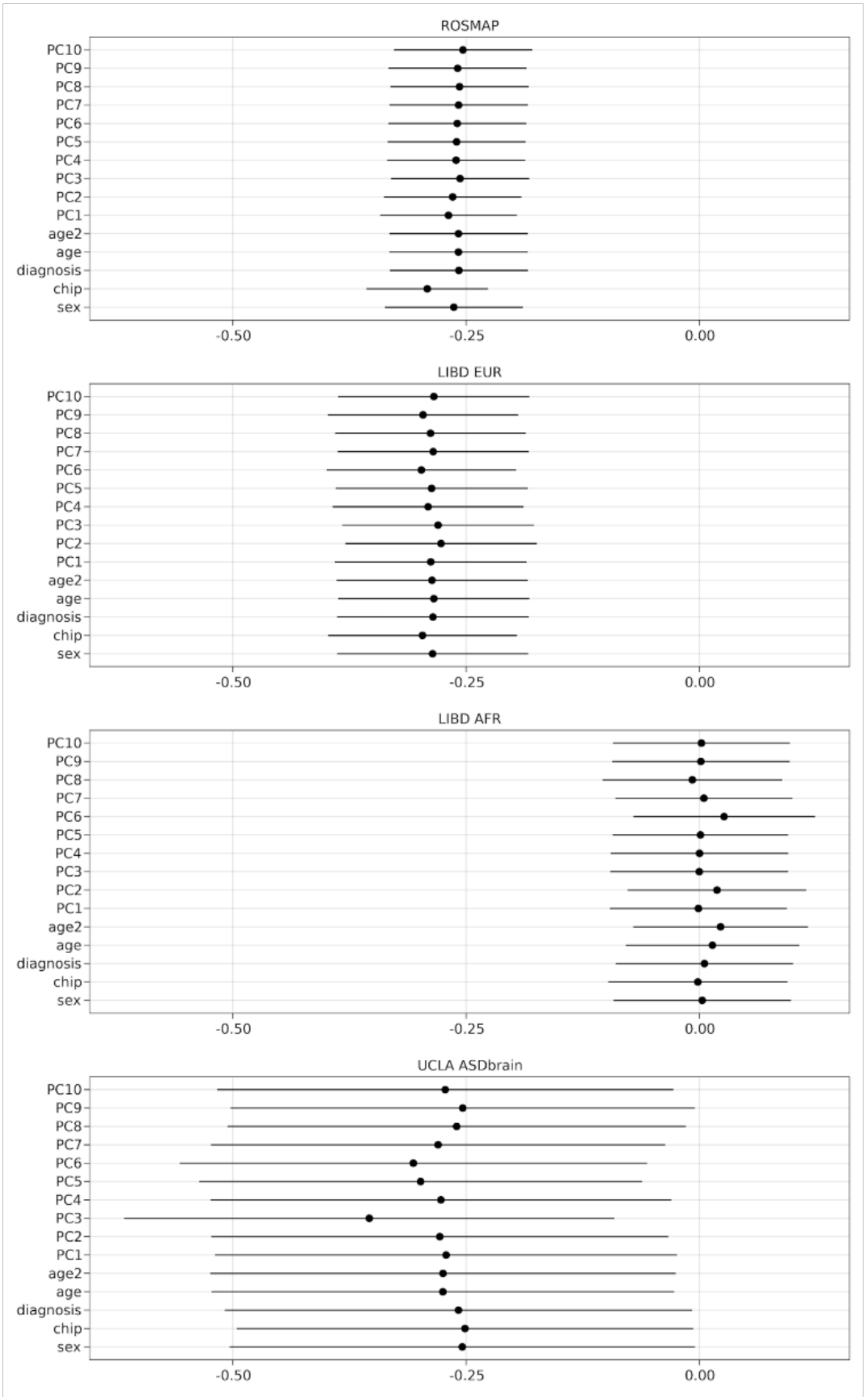


Figure S34: Meta-analysis forest plot for r17025223 genotype regressed against astrocyte CTP adjusting for each single covariate shown on the left axis, with one panel for each of the ROSMAP (all EUR), LIBD_EUR, LIBD_AFR, UCLA_ASD studies.

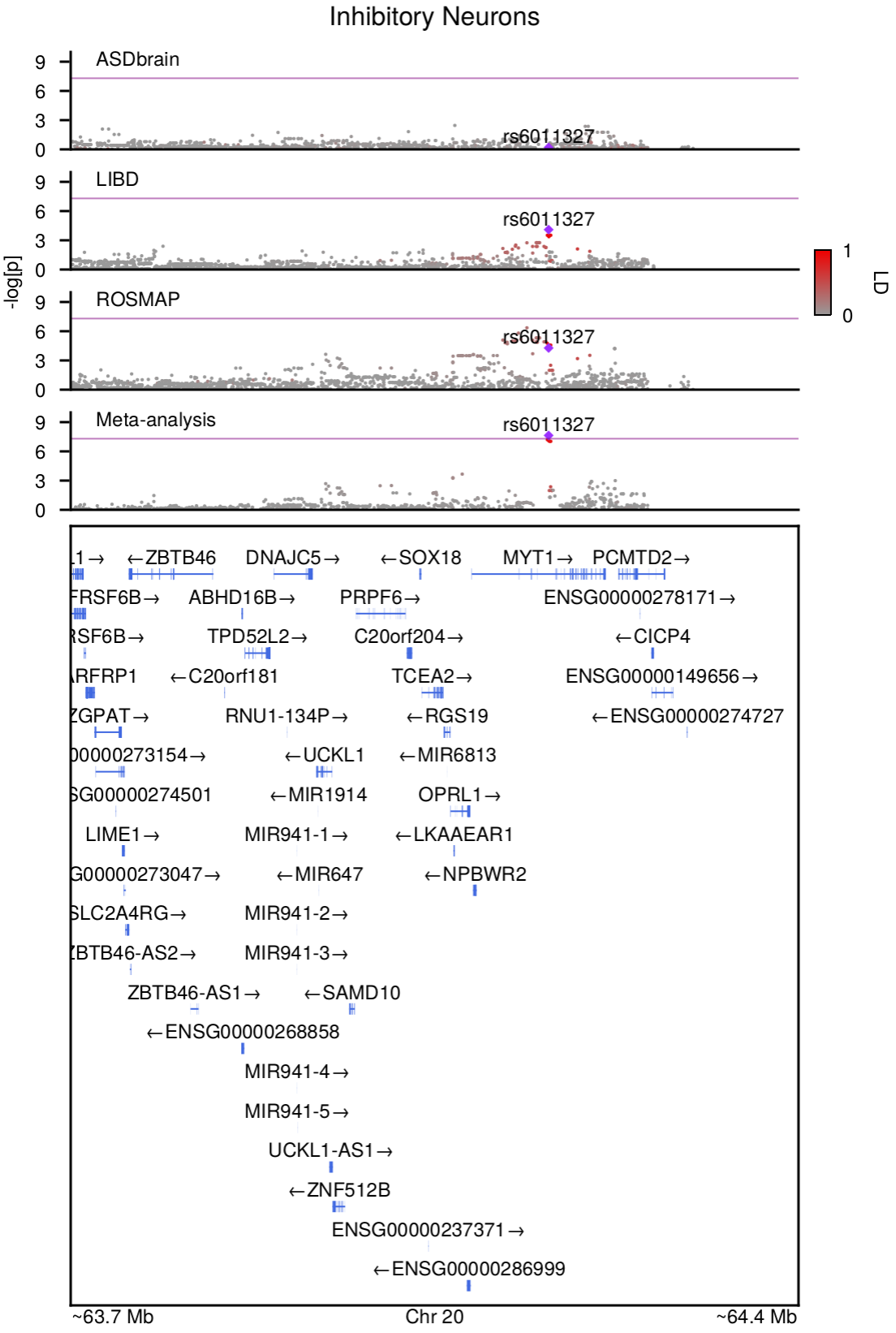


Figure S35: LocusZoom plot for the inhibitory neuron CTP GWS locus, with tracks for each meta-analyzed study and the overall meta-analysis. Purple point indicates the top SNP and red-grey gradient denotes linkage disequilibrium.

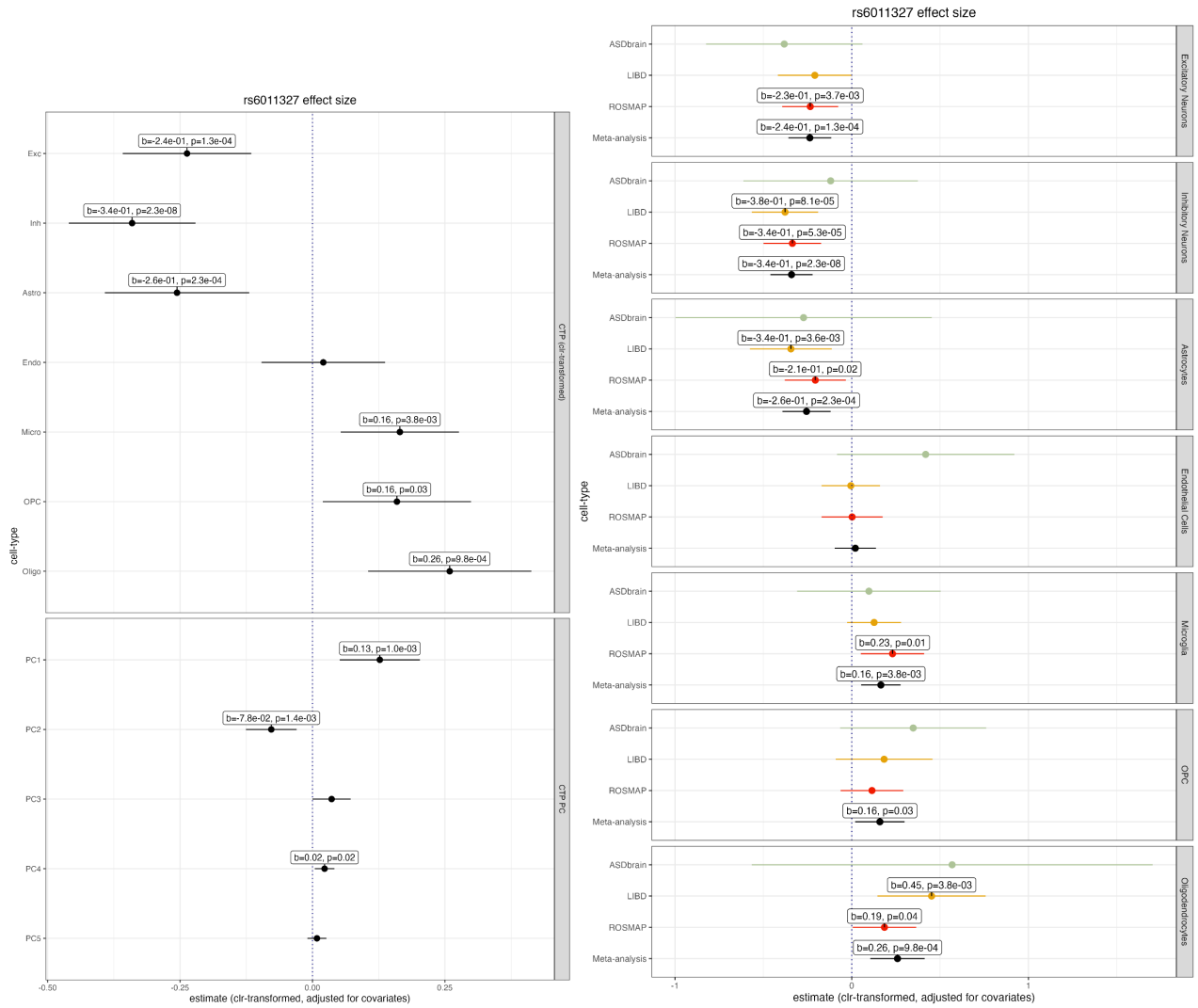


Figure S36: Forest plot for the inhibitory neuron CTP GWS locus. Left panel shows effect sizes from meta-analysis of each CTP and CTP_PC. Right panel shows effect sizes from the individual studies alongside the meta-analysis result for each CTP.

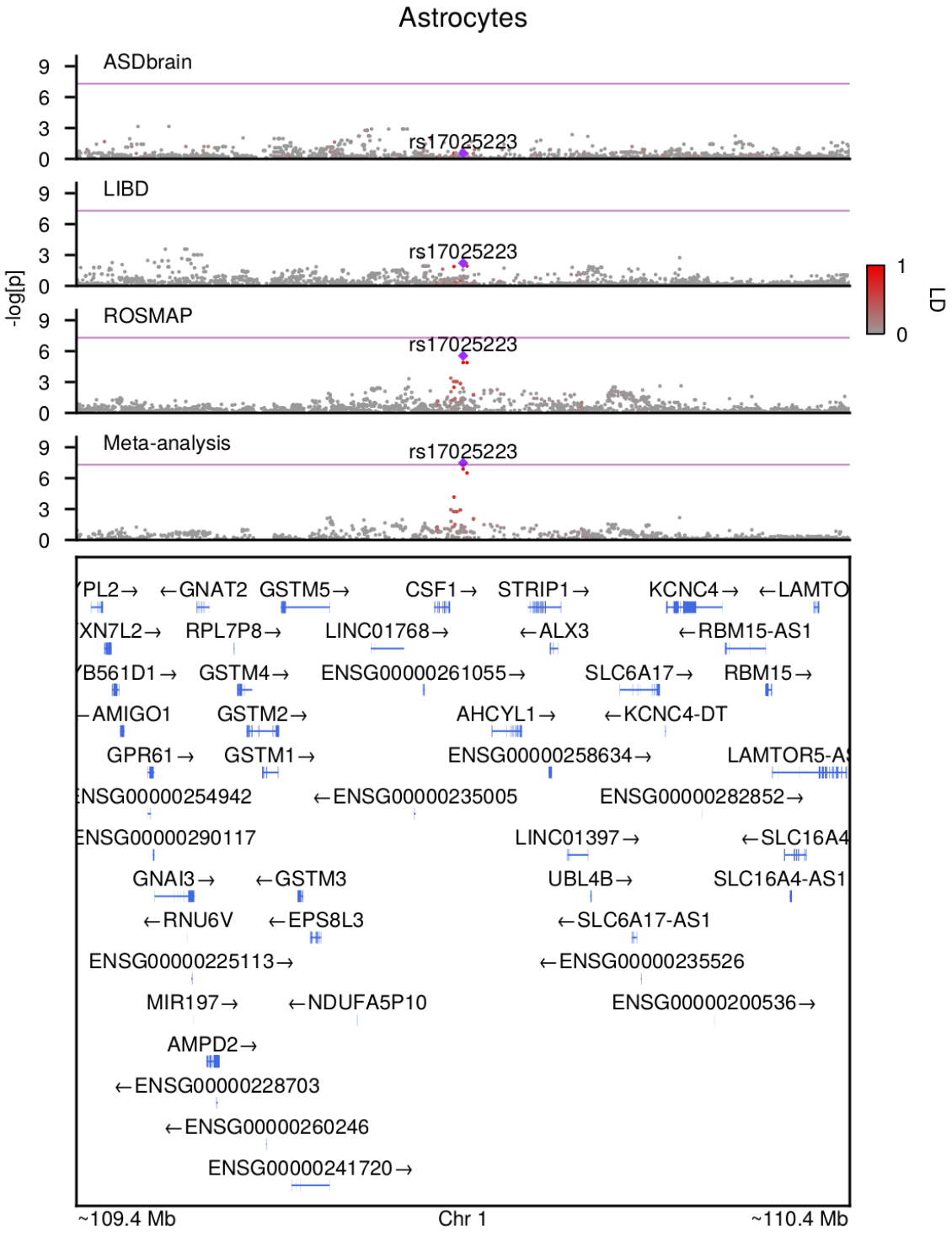


Figure S37: LocusZoom plot for the astrocyte CTP GWS locus, with tracks for each meta-analyzed study and the overall meta-analysis. Purple point indicates the top SNP and red-grey gradient denotes linkage disequilibrium.

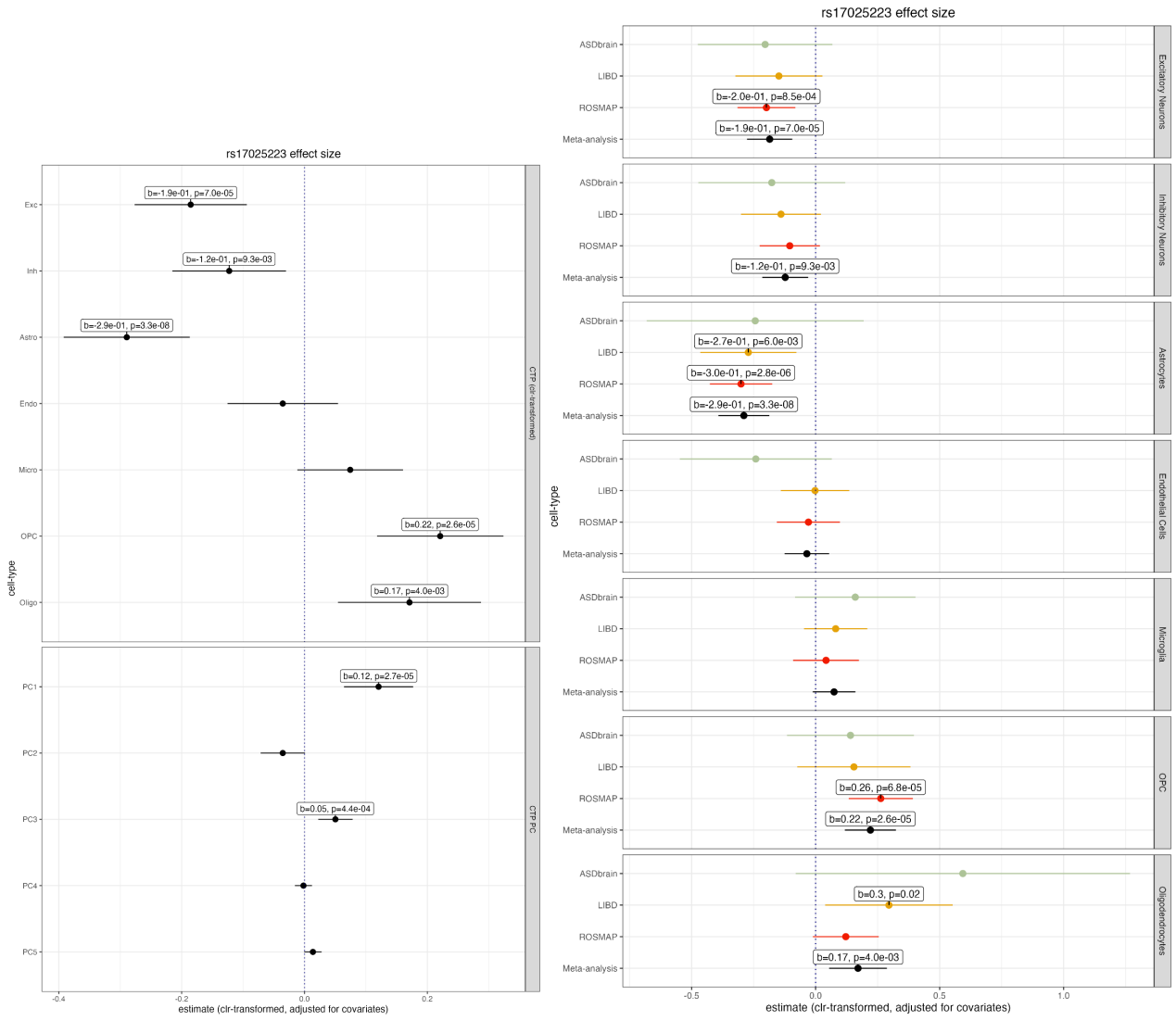


Figure S38: Forest plot for the astrocyte CTP GWS locus. Left panel shows effect sizes from meta-analysis of each CTP and CTP_PC. Right panel shows effect sizes from the individual studies alongside the meta-analysis result for each CTP.

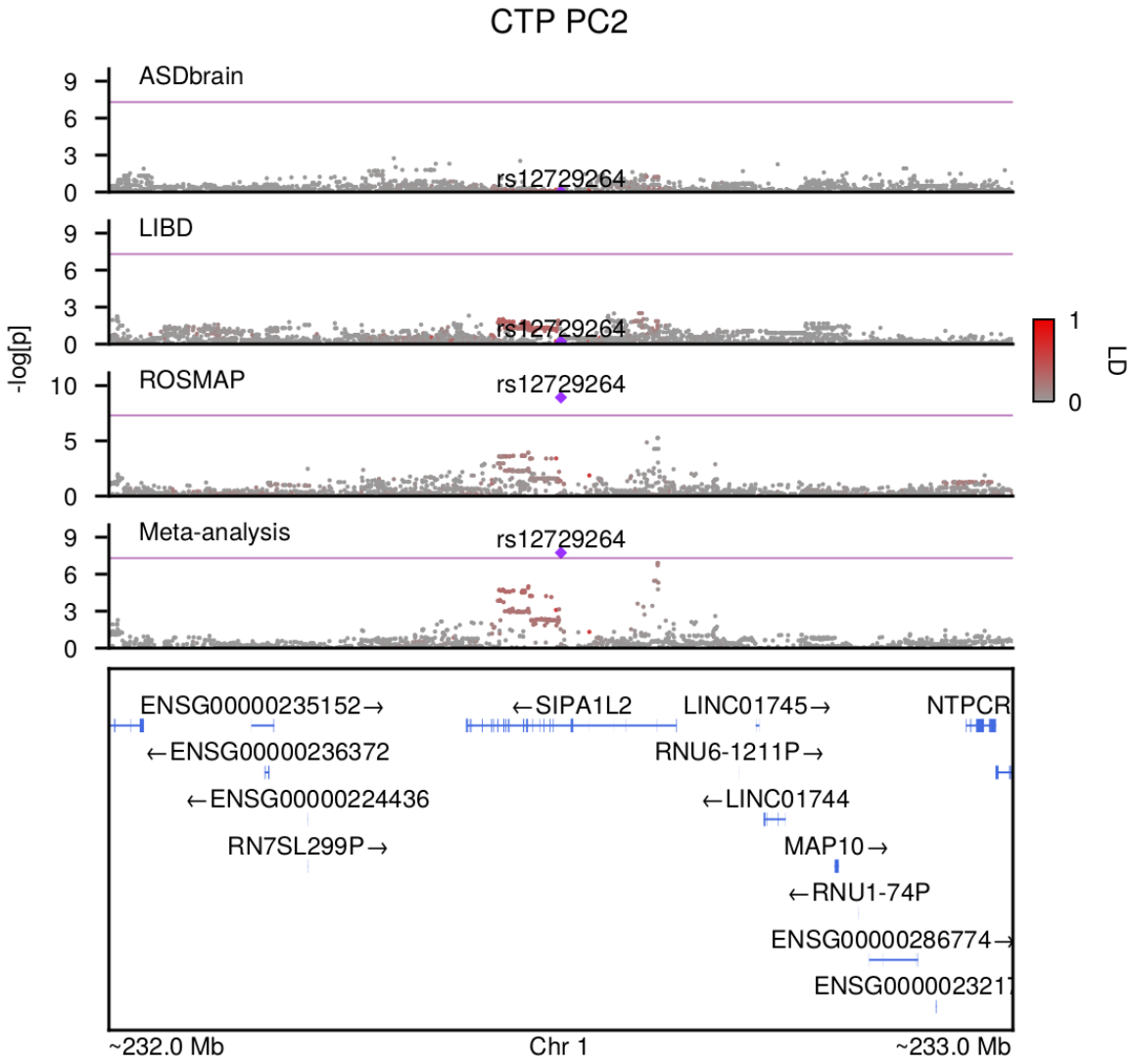


Figure S39: LocusZoom plot for the CTP_PC2 GWS locus, with tracks for each meta-analyzed study and the overall meta-analysis. Purple point indicates the top SNP and red-grey gradient denotes linkage disequilibrium.

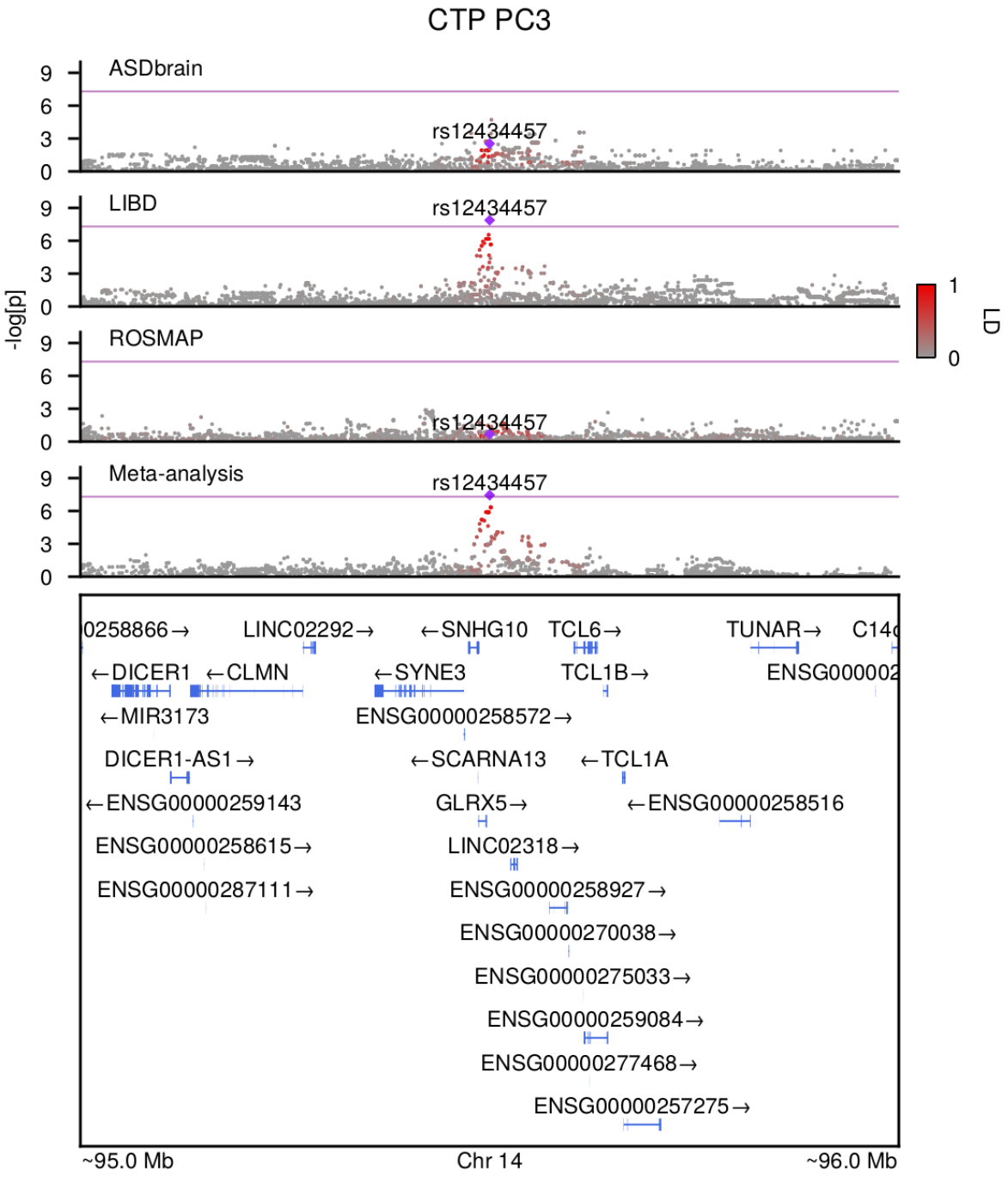


Figure S40: LocusZoom plot for the CTP_PC3 GWS locus, with tracks for each meta-analyzed study and the overall meta-analysis. Purple point indicates the top SNP and red-grey gradient denotes linkage disequilibrium.

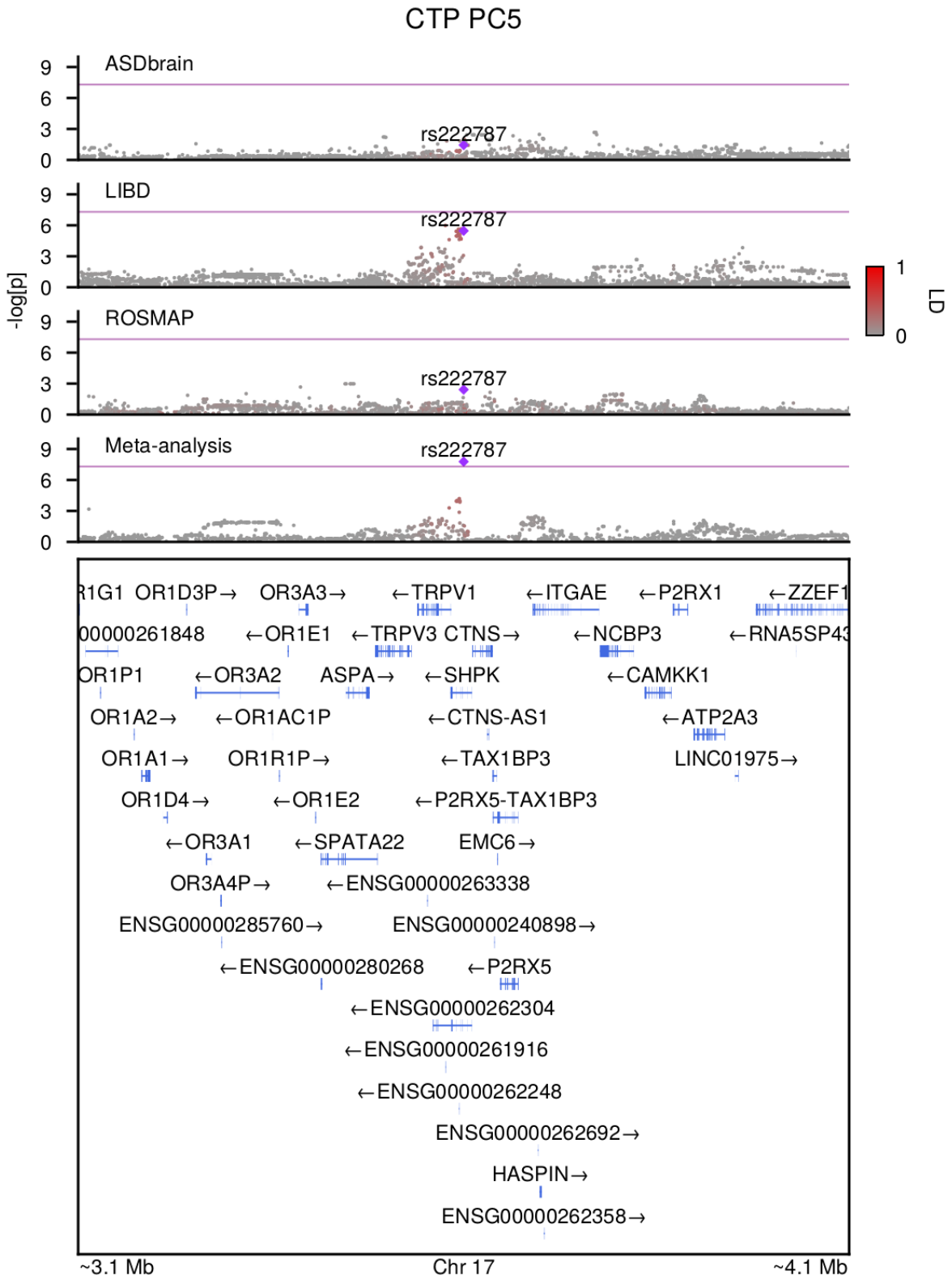


Figure S41: LocusZoom plot for the CTP_PC5 GWS locus, with tracks for each meta-analyzed study and the overall meta-analysis. Purple point indicates the top SNP and red-grey gradient denotes linkage disequilibrium.

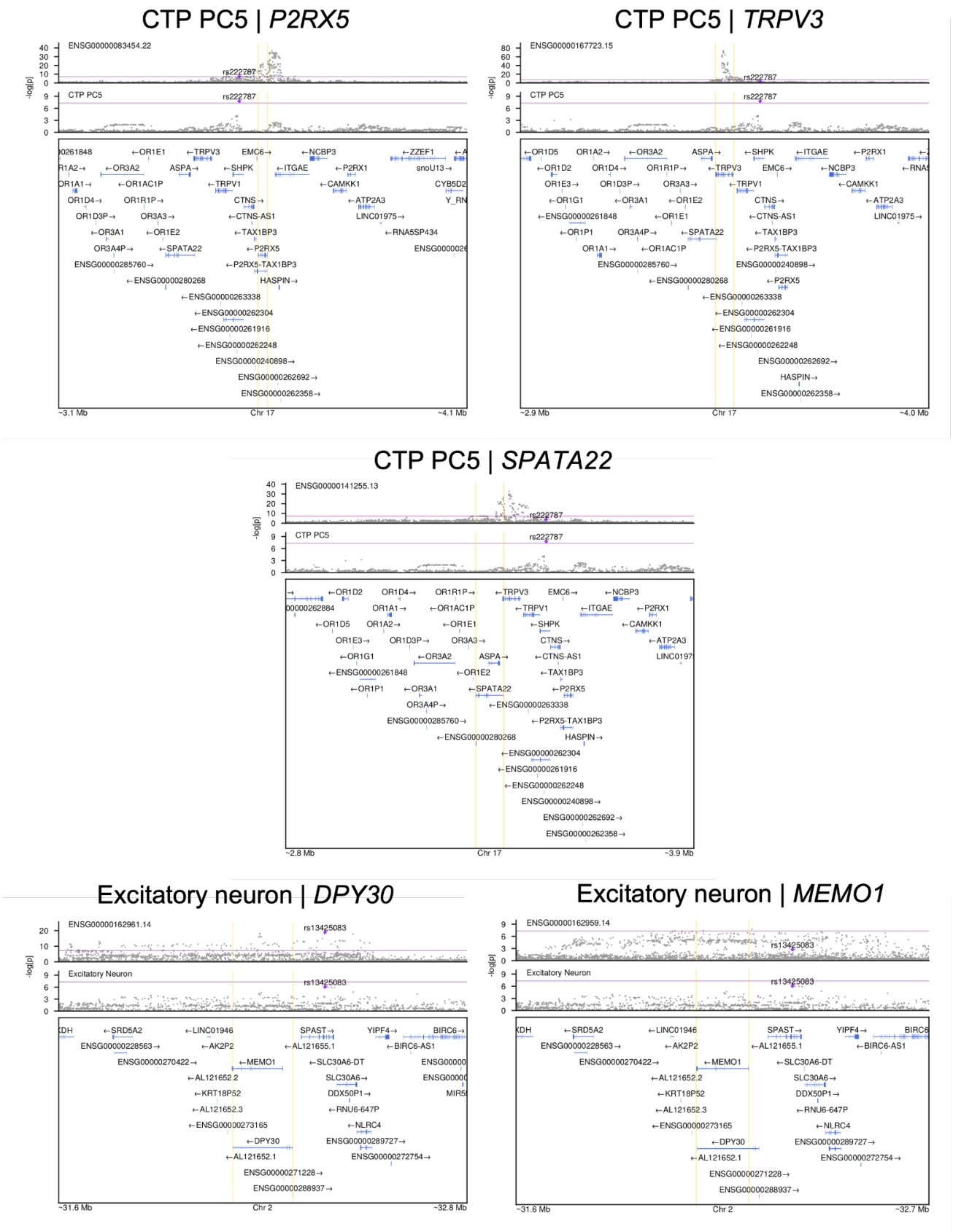


Figure S42: LocusZoom plots for notable genes suggested to have a potential causal role in brain CTP or CTP_PC by Summary Mendelian Randomisation (SMR) analysis. Upper track provides SMR-significant cis-eQTLs for gene expression; middle track provides summary statistics for the relevant CTP; bottom track illustrates gene position. Purple dot indicates the focal SNP used in SMR. Yellow lines indicate the gene start and end coordinates.

Supplementary Tables

Table S1: Summary table of phenotypes, stratified by study and diagnosis. Unless otherwise specified, numbers denote the means of the subgroup, and numbers in brackets denote the corresponding standard deviation.

Table S2: Comparison of CTP deconvolution methods in the ROSMAP dataset. mcc* denotes methylCC deconvolution; sSV* denotes smartSVA; h* indicates Houseman using array reference data; hseq* indicates Houseman with sequencing reference data; celfie* indicates CelFIE (and includes an output titled "unknown1"); VAEe* indicates variational autoencoder embeddings.

Table S3: Comparison of CTP deconvolution methods in the LIBD dataset. mcc* denotes methylCC deconvolution; sSV* denotes smartSVA; h* indicates Houseman using array reference data; hseq* indicates Houseman with sequencing reference data; celfie* indicates CelFIE (and includes an output titled "unknown1"); VAEe* indicates variational autoencoder embeddings.

Table S4: Comparison of CTP deconvolution methods in the UCLA_ASD dataset. mcc* denotes methylCC deconvolution; sSV* denotes smartSVA; h* indicates Houseman using array reference data; hseq* indicates Houseman with sequencing reference data; celfie* indicates CelFIE (and includes an output titled "unknown1"); VAEe* indicates variational autoencoder embeddings.

Table S5: Deconvolved brain CTPs (raw), with covariates.

Table S6: Deconvolved brain CTPs (clr-transform, offset 1e-3), with covariates and raw PGS for all ancestries. CTP_PC* indicates compositionally-aware principal components of the CTP data; *pgs_raw indicates PGS calculated for using genotyping across all ancestries. This table has a total of n=1,098 across all ancestries, including n=885 EUR. After applying a rel<0.05 threshold on the n=885 EUR, there were n=878 EUR which were used in the PGS analysis so that population stratification PCs did not simply capture family structure.

Table S7: Deconvolved brain CTPs (clr-transform, offset 1e-3), with covariates and standardised PGS for n=878 Europeans. *pgs_raw indicates unstandardised PGS, PC* indicates genotyping PCs within the European dataset, *_PGS indicates standardised PGS within the European subset.

Table S8: Summary Mendelian Randomisation results.

Table S9: Deconvolved brain CTPs (clr-transform, offset 1e-3), adjusted for oligodendrocyte proportions, with covariates.

Table S10: Deconvolved brain CTPs (raw), adjusted for oligodendrocyte proportions, with covariates.

Table S11: APOE genotype imputation table (from https://www.snpedia.com/index.php/APOE)

REFERENCES AND NOTES

1. G. Venkatasubramanian, M. S. Keshavan, Biomarkers in psychiatry—A critique. *Ann. Neurosci.* **23**, 3–5 (2016).

2. H. Sarlus, M. T. Heneka, Microglia in Alzheimer's disease. *J. Clin. Invest.* **127**, 3240–3249 (2017).

3. S. Jäkel, E. Agirre, A. Mendanha Falcão, D. van Bruggen, K. W. Lee, I. Knuesel, D. Malhotra, C. ffrench-Constant, A. Williams, G. Castelo-Branco, Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543–547 (2019).

4. J.-B. Pingault, P. F. O'Reilly, T. Schoeler, G. B. Ploubidis, F. Rijsdijk, F. Dudbridge, Using genetic data to strengthen causal inference in observational research. *Nat. Rev. Genet.* **19**, 566–580 (2018).

5. P. L. De Jager, G. Srivastava, K. Lunnon, J. Burgess, L. C. Schalkwyk, L. Yu, M. L. Eaton, B. T. Keenan, J. Ernst, C. McCabe, A. Tang, T. Raj, J. Replogle, W. Brodeur, S. Gabriel, H. S. Chai, C. Younkin, S. G. Younkin, F. Zou, M. Szyf, C. B. Epstein, J. A. Schneider, B. E. Bernstein, A. Meissner, N. Ertekin-Taner, L. B. Chibnik, M. Kellis, J. Mill, D. A. Bennett, Alzheimer's disease: Early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.* **17**, 1156–1163 (2014).

6. A. E. Jaffe, Y. Gao, A. Deep-Soboslay, R. Tao, T. M. Hyde, D. R. Weinberger, J. E. Kleinman, Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat. Neurosci.* **19**, 40–47 (2016).

7. D. Wang, S. Liu, J. Warrell, H. Won, X. Shi, F. C. P. Navarro, D. Clarke, M. Gu, P. Emani, Y. T. Yang, M. Xu, M. J. Gandal, S. Lou, J. Zhang, J. J. Park, C. Yan, S. K. Rhie, K. Manakongtreecheep, H. Zhou, A. Nathan, M. Peters, E. Mattei, D. Fitzgerald, T. Brunetti, J. Moore, Y. Jiang, K. Girdhar, G. E. Hoffman, S. Kalayci, Z. H. Gümüş, G. E. Crawford, P. Roussos, S. Akbarian, A. E. Jaffe, K. P. White, Z. Weng, N. Sestan, D. H. Geschwind, J. A. Knowles, M. B. Gerstein, Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018).

8. G. E. Hoffman, J. Bendl, G. Voloudakis, K. S. Montgomery, L. Sloofman, Y.-C. Wang, H. R. Shah, M. E. Hauberg, J. S. Johnson, K. Girdhar, L. Song, J. F. Fullard, R. Kramer, C.-G. Hahn, R. Gur, S.

Marenco, B. K. Lipska, D. A. Lewis, V. Haroutunian, S. Hemby, P. Sullivan, S. Akbarian, A. Chess, J. D. Buxbaum, G. E. Crawford, E. Domenici, B. Devlin, S. K. Sieberts, M. A. Peters, P. Roussos, CommonMind Consortium provides transcriptomic and epigenomic data for schizophrenia and bipolar disorder. *Sci. Data* **6**, 180 (2019).

9. B. Ng, C. C. White, H.-U. Klein, S. K. Sieberts, C. McCabe, E. Patrick, J. Xu, L. Yu, C. Gaiteri, D. A. Bennett, S. Mostafavi, P. L. De Jager, An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.* **20**, 1418–1426 (2017).

10. A. E. Teschendorff, S. C. Zheng, Cell-type deconvolution in epigenome-wide association studies: A review and recommendations. *Epigenomics* **9**, 757–768 (2017).

11. F. Avila Cobos, J. Alquicira-Hernandez, J. E. Powell, P. Mestdagh, K. De Preter, Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nat. Commun.* **11**, 5650–5650 (2020).

12. E. Rahmani, N. Zaitlen, Y. Baran, C. Eng, D. Hu, J. Galanter, S. Oh, E. G. Burchard, E. Eskin, J. Zou, E. Halperin, Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies. *Nat. Methods* **13**, 443–445 (2016).

13. J. Chen, E. Behnam, J. Huang, M. F. Moffatt, D. J. Schaid, L. Liang, X. Lin, Fast and robust adjustment of cell mixtures in epigenome-wide association studies with SmartSVA. *BMC Genomics* **18**, 413 (2017).

14. K. B. Michels, A. M. Binder, S. Dedeurwaerder, C. B. Epstein, J. M. Greally, I. Gut, E. A. Houseman, B. Izzi, K. T. Kelsey, A. Meissner, A. Milosavljevic, K. D. Siegmund, C. Bock, R. A. Irizarry, Recommendations for the design and analysis of epigenome-wide association studies. *Nat. Methods* **10**, 949–955 (2013).

15. E. A. Houseman, W. P. Accomando, D. C. Koestler, B. C. Christensen, C. J. Marsit, H. H. Nelson, J. K. Wiencke, K. T. Kelsey, DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86 (2012).

16. Z. Li, H. Wu, TOAST: Improving reference-free cell composition estimation by cross-cell type differential analysis. *Genome Biol.* **20**, 190 (2019).

17. S. C. Hicks, R. A. Irizarry, methylCC: Technology-independent estimation of cell type composition using differentially methylated regions. *Genome Biol.* **20**, 261 (2019).

18. O. A. Sosina, M. N. Tran, K. R. Maynard, R. Tao, M. A. Taub, K. Martinowich, S. A. Semick, B. C. Quach, D. R. Weinberger, T. Hyde, D. B. Hancock, J. E. Kleinman, J. T. Leek, A. E. Jaffe, Strategies for cellular deconvolution in human brain RNA sequencing data [version 1; peer review: 1 approved, 1 approved with reservations] (F1000Research, 2021); https://doi.org/10.12688/f1000research.50858.1.

19. J. R. Ecker, D. H. Geschwind, A. R. Kriegstein, J. Ngai, P. Osten, D. Polioudakis, A. Regev, N. Sestan, I. R. Wickersham, H. Zeng, The BRAIN initiative cell census consortium: Lessons learned toward generating a comprehensive brain cell atlas. *Neuron* **96**, 542–557 (2017).

20. A. Kozlenkov, M. Wang, P. Roussos, S. Rudchenko, M. Barbu, M. Bibikova, B. Klotzle, A. J. Dwork, B. Zhang, Y. L. Hurd, E. V. Koonin, M. Wegner, S. Dracheva, Substantial DNA methylation differences between two major neuronal subtypes in human brain. *Nucleic Acids Res.* **44**, 2593–2612 (2016).

21. C. Luo, C. L. Keown, L. Kurihara, J. Zhou, Y. He, J. Li, R. Castanon, J. Lucero, J. R. Nery, J. P. Sandoval, B. Bui, T. J. Sejnowski, T. T. Harkins, E. A. Mukamel, M. M. Behrens, J. R. Ecker, Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600–604 (2017).

22. E. Patrick, M. Taga, A. Ergun, B. Ng, W. Casazza, M. Cimpean, C. Yung, J. A. Schneider, D. A. Bennett, C. Gaiteri, P. L. D. Jager, E. M. Bradshaw, S. Mostafavi, Deconvolving the contributions of cell-type heterogeneity on cortical gene expression. *PLOS Comput. Biol.* **16**, e1008120 (2020).

23. L. Toker, G. S. Nido, C. Tzoulis, Not every estimate counts—Evaluation of cell composition estimation approaches in brain bulk tissue data. *Genome Med.* **15**, 41 (2023).

24. C. Luo, H. Liu, F. Xie, E. J. Armand, K. Siletti, T. E. Bakken, R. Fang, W. I. Doyle, T. Stuart, R. D. Hodge, L. Hu, B.-A. Wang, Z. Zhang, S. Preissl, D.-S. Lee, J. Zhou, S.-Y. Niu, R. Castanon, A. Bartlett, A. Rivkin, X. Wang, J. Lucero, J. R. Nery, D. A. Davis, D. C. Mash, R. Satija, J. R. Dixon, S. Linnarsson, E. Lein, M. M. Behrens, B. Ren, E. A. Mukamel, J. R. Ecker, Single nucleus multi-omics identifies human cortical cell regulatory genome diversity. *Cell Genom* **2**, 100107 (2022).

25. K. McGregor, S. Bernatsky, I. Colmegna, M. Hudson, T. Pastinen, A. Labbe, C. M. T. Greenwood, An evaluation of methods correcting for cell-type heterogeneity in DNA methylation studies. *Genome Biol.* **17**, 84 (2016).

26. C. C. Y. Wong, R. Smith, E. Hannon, G. Ramaswami, N. N. Parikshak, E. Assary, C. Troakes, J. Poschmann, L. C. Schalkwyk, W. Sun, S. Prabhakar, D. H. Geschwind, J. Mill, Genome-wide DNA methylation profiling identifies convergent molecular signatures associated with idiopathic and syndromic autism in post-mortem human brain tissue. *Hum. Mol. Genet.* **28**, 2201–2211 (2019).

27. N. Loyfer, J. Magenheim, A. Peretz, G. Cann, J. Bredno, A. Klochendler, I. Fox-Fisher, S. Shabi-Porat, M. Hecht, T. Pelet, J. Moss, Z. Drawshy, H. Amini, P. Moradi, S. Nagaraju, D. Bauman, D. Shveiky, S. Porat, U. Dior, G. Rivkin, O. Or, N. Hirshoren, E. Carmon, A. Pikarsky, A. Khalaileh, G. Zamir, R. Grinbaum, M. Abu Gazala, I. Mizrahi, N. Shussman, A. Korach, O. Wald, U. Izhar, E. Erez, V. Yutkin, Y. Samet, D. Rotnemer Golinkin, K. L. Spalding, H. Druid, P. Arner, A. M. J. Shapiro, M. Grompe, A. Aravanis, O. Venn, A. Jamshidi, R. Shemer, Y. Dor, B. Glaser, T. Kaplan, A DNA methylation atlas of normal human cell types. *Nature* **613**, 355–364 (2023).

28. T. Zhu, J. Liu, S. Beck, S. Pan, D. Capper, M. Lechner, C. Thirlwell, C. E. Breeze, A. E. Teschendorff, A pan-tissue DNA methylation atlas enables in silico decomposition of human tissue methylomes at cell-type resolution. *Nat. Methods* **19**, 296–306 (2022).

29. G. J. Sutton, D. Poppe, R. K. Simmons, K. Walsh, U. Nawaz, R. Lister, J. A. Gagnon-Bartsch, I. Voineagu, Comprehensive evaluation of deconvolution methods for human brain gene expression. *Nat. Commun.* **13**, 1358 (2022).

30. J. J. Levy, A. J. Titus, C. L. Petersen, Y. Chen, L. A. Salas, B. C. Christensen, MethylNet: An automated and modular deep learning approach for DNA methylation analysis. *BMC Bioinformatics* **21**, 108 (2020).

31. G. Shireby, E. L. Dempster, S. Policicchio, R. G. Smith, E. Pishva, B. Chioza, J. P. Davies, J. Burrage, K. Lunnon, D. Seiler Vellame, S. Love, A. Thomas, K. Brookes, K. Morgan, P. Francis, E. Hannon, J. Mill, DNA methylation signatures of Alzheimer's disease neuropathology in the cortex are primarily driven by variation in non-neuronal cell-types. *Nat. Commun.* **13**, 5620 (2022).

32. M. J. Gandal, J. R. Haney, B. Wamsley, C. X. Yap, S. Parhami, P. S. Emani, N. Chang, G. T. Chen, G. D. Hoftman, D. de Alba, G. Ramaswami, C. L. Hartl, A. Bhattacharya, C. Luo, T. Jin, D. Wang, R. Kawaguchi, D. Quintero, J. Ou, Y. E. Wu, N. N. Parikshak, V. Swarup, T. G. Belgard, M. Gerstein, B. Pasaniuc, D. H. Geschwind, Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. *Nature* **611**, 532–539 (2022).

33. H. Mathys, J. Davila-Velderrain, Z. Peng, F. Gao, S. Mohammadi, J. Z. Young, M. Menon, L. He, F. Abdurrob, X. Jiang, A. J. Martorell, R. M. Ransohoff, B. P. Hafler, D. A. Bennett, M. Kellis, L.-H. Tsai, Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).

34. B. Wamsley, L. Bicks, Y. Cheng, R. Kawaguchi, D. Quintero, J. Grundman, J. Liu, S. Xiao, N. Hawken, M. Margolis, S. Mazariegos, D. H. Geschwind, Molecular cascades and cell-type specific signatures in ASD revealed by single cell genomics. bioRxiv 530869 [**Preprint**]. 10 March 2023. https://doi.org/10.1101/2023.03.10.530869.

35. A. C. Yang, R. T. Vest, F. Kern, D. P. Lee, M. Agam, C. A. Maat, P. M. Losada, M. B. Chen, N. Schaum, N. Khoury, A. Toland, K. Calcuttawala, H. Shin, R. Pálovics, A. Shin, E. Y. Wang, J. Luo, D. Gate, W. J. Schulz-Schaeffer, P. Chu, J. A. Siegenthaler, M. W. McNerney, A. Keller, T. Wyss-Coray, A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. *Nature* **603**, 885–892 (2022).

36. J. Grove, S. Ripke, T. D. Als, M. Mattheisen, R. K. Walters, H. Won, J. Pallesen, E. Agerbo, O. A. Andreassen, R. Anney, S. Awashti, R. Belliveau, F. Bettella, J. D. Buxbaum, J. Bybjerg-Grauholm, M. Bækvad-Hansen, F. Cerrato, K. Chambert, J. H. Christensen, C. Churchhouse, K. Dellenvall, D. Demontis, S. De Rubeis, B. Devlin, S. Djurovic, A. L. Dumont, J. I. Goldstein, C. S. Hansen, M. E. Hauberg, M. V. Hollegaard, S. Hope, D. P. Howrigan, H. Huang, C. M. Hultman, L. Klei, J. Maller, J. Martin, A. R. Martin, J. L. Moran, M. Nyegaard, T. Nærland, D. S. Palmer, A. Palotie, C. B. Pedersen, M. G. Pedersen, T. dPoterba, J. B. Poulsen, B. S. Pourcain, P. Qvist, K. Rehnström, A. Reichenberg, J. Reichert, E. B. Robinson, K. Roeder, P. Roussos, E. Saemundsen, S. Sandin, F. K. Satterstrom, G. D. Smith, H. Stefansson, S. Steinberg, C. R. Stevens, P. F. Sullivan, P. Turley, G. Bragi Walters, X. Xu; Autism Spectrum Disorder Working Group of the Psychiatric Genomics Consortium; BUPGEN; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium; 23andMe Research

Team, K. Stefansson, D. H. Geschwind, M. Nordentoft, D. M. Hougaard, T. Werge, O. Mors, P. B. Mortensen, B. M. Neale, M. J. Daly, A. D. Børglum, Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).

37. A. F. Pardiñas, P. Holmans, A. J. Pocklington, V. Escott-Price, S. Ripke, N. Carrera, S. E. Legge, S. Bishop, D. Cameron, M. L. Hamshere, J. Han, L. Hubbard, A. Lynham, K. Mantripragada, E. Rees, J. H. MacCabe, S. A. McCarroll, B. T. Baune, G. Breen, E. M. Byrne, U. Dannlowski, T. C. Eley, C. Hayward, N. G. Martin, A. M. McIntosh, R. Plomin, D. J. Porteous, N. R. Wray, A. Caballero, D. H. Geschwind, L. M. Huckins, D. M. Ruderfer, E. Santiago, P. Sklar, E. A. Stahl, H. Won, E. Agerbo, T. D. Als, O. A. Andreassen, M. Bækvad-Hansen, P. B. Mortensen, C. B. Pedersen, A. D. Børglum, J. Bybjerg-Grauholm, S. Djurovic, N. Durmishi, M. G. Pedersen, V. Golimbet, J. Grove, D. M. Hougaard, M. Mattheisen, E. Molden, O. Mors, M. Nordentoft, M. Pejovic-Milovancevic, E. Sigurdsson, T. Silagadze, C. S. Hansen, K. Stefansson, H. Stefansson, S. Steinberg, S. Tosato, T. Werge, D. A. Collier, D. Rujescu, G. Kirov, M. J. Owen, M. C. O'Donovan, J. T. R. Walters, Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).

38. R. E. Marioni, S. E. Harris, Q. Zhang, A. F. McRae, S. P. Hagenaars, W. D. Hill, G. Davies, C. W. Ritchie, C. R. Gale, J. M. Starr, A. M. Goate, D. J. Porteous, J. Yang, K. L. Evans, I. J. Deary, N. R. Wray, P. M. Visscher, GWAS on family history of Alzheimer's disease. *Transl. Psychiatry* **8**, 99 (2018).

39. D. M. Howard, M. J. Adams, T.-K. Clarke, J. D. Hafferty, J. Gibson, M. Shirali, J. R. I. Coleman, S. P. Hagenaars, J. Ward, E. M. Wigmore, C. Alloza, X. Shen, M. C. Barbu, E. Y. Xu, H. C. Whalley, R. E. Marioni, D. J. Porteous, G. Davies, I. J. Deary, G. Hemani, K. Berger, H. Teismann, R. Rawal, V. Arolt, B. T. Baune, U. Dannlowski, K. Domschke, C. Tian, D. A. Hinds, M. Trzaskowski, E. M. Byrne, S. Ripke, D. J. Smith, P. F. Sullivan, N. R. Wray, G. Breen, C. M. Lewis, A. M. McIntosh, Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).

40. J. J. Lee, R. Wedow, A. Okbay, E. Kong, O. Maghzian, M. Zacher, T. A. Nguyen-Viet, P. Bowers, J. Sidorenko, R. Karlsson Linnér, M. A. Fontana, T. Kundu, C. Lee, H. Li, R. Li, R. Royer, P. N. Timshel, R. K. Walters, E. A. Willoughby, L. Yengo, M. Alver, Y. Bao, D. W. Clark, F. R. Day, N. A. Furlotte,

P. K. Joshi, K. E. Kemper, A. Kleinman, C. Langenberg, R. Mägi, J. W. Trampush, S. S. Verma, Y. Wu, M. Lam, J. H. Zhao, Z. Zheng, J. D. Boardman, H. Campbell, J. Freese, K. M. Harris, C. Hayward, P. Herd, M. Kumari, T. Lencz, J. Luan, A. K. Malhotra, A. Metspalu, L. Milani, K. K. Ong, J. R. B. Perry, D. J. Porteous, M. D. Ritchie, M. C. Smart, B. H. Smith, J. Y. Tung, N. J. Wareham, J. F. Wilson, J. P. Beauchamp, D. C. Conley, T. Esko, S. F. Lehrer, P. K. E. Magnusson, S. Oskarsson, T. H. Pers, M. R. Robinson, K. Thom, C. Watson, C. F. Chabris, M. N. Meyer, D. I. Laibson, J. Yang, M. Johannesson, P. D. Koellinger, P. Turley, P. M. Visscher, D. J. Benjamin, D. Cesarini, Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).

41. L. Yengo, J. Sidorenko, K. E. Kemper, Z. Zheng, A. R. Wood, M. N. Weedon, T. M. Frayling, J. Hirschhorn, J. Yang, P. M. Visscher; GIANT Consortium, Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).

42. E. Persyn, K. B. Hanscombe, J. M. M. Howson, C. M. Lewis, M. Traylor, H. S. Markus, Genome-wide association study of MRI markers of cerebral small vessel disease in 42,310 participants. *Nat. Commun.* **11**, 2175 (2020).

43. N. D. Prins, P. Scheltens, White matter hyperintensities, cognitive impairment and dementia: An update. *Nat. Rev. Neurol.* **11**, 157–165 (2015).

44. S. Debette, H. S. Markus, The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: Systematic review and meta-analysis. *BMJ* **341**, c3666 (2010).

45. J. M. Wardlaw, M. C. Valdés Hernández, S. Muñoz-Maniega, What are white matter hyperintensities made of?. *J. Am. Heart Assoc.* **4**, e001140 (2015).

46. J. Yang, T. Ferreira, A. P. Morris, S. E. Medland; Genetic Investigation of ANthropometric Traits (GIANT) Consortium; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, P. A. F. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. N. Weedon, R. J. Loos, T. M. Frayling, M. I. McCarthy, J. N. Hirschhorn, M. E. Goddard, P. M. Visscher, Conditional and joint

multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).

47. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

48. Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard, N. R. Wray, P. M. Visscher, J. Yang, Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).

49. Y. Zhang, P. Babczyk, A. Pansky, M. U. Kassack, E. Tobiasch, P2 receptors influence hMSCs differentiation towards endothelial cell and smooth muscle cell lineages. *Int. J. Mol. Sci.* **21**, 6210 (2020).

50. L. M. Schwiebert, W. C. Rice, B. A. Kudlow, A. L. Taylor, E. M. Schwiebert, Extracellular ATP signaling and P2X nucleotide receptors in monolayers of primary human vascular endothelial cells. *Am. J. Physiol. Cell Physiol.* **282**, C289–C301 (2002).

51. U. Lalo, Y. Pankratov, S. P. Wichert, M. J. Rossner, R. A. North, F. Kirchhoff, A. Verkhratsky, P2X1 and P2X5 subunits form the functional P2X receptor in mouse cortical astrocytes. *J. Neurosci.* **28**, 5473–5480 (2008).

52. B. F. King, Rehabilitation of the P2X5 receptor: A re-evaluation of structure and function. *Purinergic Signal* **19**, 421–439 (2023).

53. S. Earley, A. L. Gonzales, Z. I. Garcia, A dietary agonist of transient receptor potential cation channel V3 elicits endothelium-dependent vasodilation. *Mol. Pharmacol.* **77**, 612–620 (2010).

54. P. W. Pires, M. N. Sullivan, H. A. T. Pritchard, J. J. Robinson, S. Earley, Unitary TRPV3 channel Ca^{2+} influx events elicit endothelium-dependent dilation of cerebral parenchymal arterioles. *Am. J. Physiol. Heart Circ. Physiol.* **309**, H2031–H2041 (2015).

55. X. Chen, J. Zhang, K. Wang, Inhibition of intracellular proton-sensitive Ca^{2+}-permeable TRPV3 channels protects against ischemic brain injury. *Acta Pharm. Sin. B* **12**, 2330–2347 (2022).

56. K. Shah, G. D. King, H. Jiang, A chromatin modulator sustains self-renewal and enables differentiation of postnatal neural stem and progenitor cells. *J. Mol. Cell Biol.* **12**, 4–16 (2020).

57. N. Nakagawa, C. Plestant, K. Yabuno-Nakagawa, J. Li, J. Lee, C.-W. Huang, A. Lee, O. Krupa, A. Adhikari, S. Thompson, T. Rhynes, V. Arevalo, J. L. Stein, Z. Molnár, A. Badache, E. S. Anton, Memo1-mediated tiling of radial glial cells facilitates cerebral cortical development. *Neuron* **103**, 836–852.e5 (2019).

58. C. Wallace, A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLOS Genet.* **17**, e1009440 (2021).

59. J. Bryois, D. Calini, W. Macnair, L. Foo, E. Urich, W. Ortmann, V. A. Iglesias, S. Selvaraj, E. Nutma, M. Marzin, S. Amor, A. Williams, G. Castelo-Branco, V. Menon, P. De Jager, D. Malhotra, Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nat. Neurosci.* **25**, 1104–1112 (2022).

60. Z. Li, F. H. G. Farias, U. Dube, J. L. Del-Aguila, K. A. Mihindukulasuriya, M. V. Fernandez, L. Ibanez, J. P. Budde, F. Wang, A. M. Lake, Y. Deming, J. Perez, C. Yang, J. A. Bahena, W. Qin, J. L. Bradley, R. Davenport, K. Bergmann, J. C. Morris, R. J. Perrin, B. A. Benitez, J. D. Dougherty, O. Harari, C. Cruchaga, The TMEM106B FTLD-protective variant, rs1990621, is also associated with increased neuronal proportion. *Acta Neuropathol.* **139**, 45–61 (2020).

61. M. Fujita, Z. Gao, L. Zeng, C. McCabe, C. C. White, B. Ng, G. S. Green, O. Rozenblatt-Rosen, D. Phillips, L. Amir-Zilberstein, H. Lee, R. V. Pearse, A. Khan, B. N. Vardarajan, K. Kiryluk, C. J. Ye, H.-U. Klein, G. Wang, A. Regev, N. Habib, J. A. Schneider, Y. Wang, T. Young-Pearse, S. Mostafavi, D. A. Bennett, V. Menon, P. L. De Jager, Cell-subtype specific effects of genetic variation in the aging and Alzheimer cortex. bioRxiv 515446 [**Preprint**]. 08 November 2022.

62. H. Yousef, C. J. Czupalla, D. Lee, M. B. Chen, A. N. Burke, K. A. Zera, J. Zandstra, E. Berber, B. Lehallier, V. Mathur, R. V. Nair, L. N. Bonanno, A. C. Yang, T. Peterson, H. Hadeiba, T. Merkel, J. Körbelin, M. Schwaninger, M. S. Buckwalter, S. R. Quake, E. C. Butcher, T. Wyss-Coray, Aged blood impairs hippocampal neural precursor activity and activates microglia via brain endothelial cell VCAM1. *Nat. Med.* **25**, 988–1000 (2019).

63. E. Hannon, E. L. Dempster, J. P. Davies, B. Chioza, G. E. T. Blake, J. Burrage, S. Policicchio, A. Franklin, E. M. Walker, R. A. Bamford, L. C. Schalkwyk, J. Mill, Quantifying the proportion of different cell types in the human cortex using DNA methylation profiles. *BMC Biol.* **22**, 17 (2024).

64. F. J. Raabe, L. Slapakova, M. J. Rossner, L. Cantuti-Castelvetri, M. Simons, P. G. Falkai, A. Schmitt, Oligodendrocytes as a new therapeutic target in schizophrenia: From histopathological findings to neuron-oligodendrocyte interaction. *Cells* **8**, 1496 (2019).

65. V. M. Vostrikov, N. A. Uranova, D. D. Orlovskaya, Deficit of perineuronal oligodendrocytes in the prefrontal cortex in schizophrenia and mood disorders. *Schizophr. Res.* **94**, 273–280 (2007).

66. P. R. Hof, V. Haroutunian, V. L. Friedrich, W. Byne, C. Buitron, D. P. Perl, K. L. Davis, Loss and altered spatial distribution of oligodendrocytes in the superior frontal gyrus in schizophrenia. *Biol. Psychiatry* **53**, 1075–1085 (2003).

67. S. Kelly, N. Jahanshad, A. Zalesky, P. Kochunov, I. Agartz, C. Alloza, O. A. Andreassen, C. Arango, N. Banaj, S. Bouix, C. A. Bousman, R. M. Brouwer, J. Bruggemann, J. Bustillo, W. Cahn, V. Calhoun, D. Cannon, V. Carr, S. Catts, J. Chen, J.-x. Chen, X. Chen, C. Chiapponi, K. K. Cho, V. Ciullo, A. S. Corvin, B. Crespo-Facorro, V. Cropley, P. De Rossi, C. M. Diaz-Caneja, E. W. Dickie, S. Ehrlich, F.-m. Fan, J. Faskowitz, H. Fatouros-Bergman, L. Flyckt, J. M. Ford, J.-P. Fouche, M. Fukunaga, M. Gill, D. C. Glahn, R. Gollub, E. D. Goudzwaard, H. Guo, R. E. Gur, R. C. Gur, T. P. Gurholt, R. Hashimoto, S. N. Hatton, F. A. Henskens, D. P. Hibar, I. B. Hickie, L. E. Hong, J. Horacek, F. M. Howells, H. E. Hulshoff Pol, C. L. Hyde, D. Isaev, A. Jablensky, P. R. Jansen, J. Janssen, E. G. Jönsson, L. A. Jung, R. S. Kahn, Z. Kikinis, K. Liu, P. Klauser, C. Knöchel, M. Kubicki, J. Lagopoulos, C. Langen, S. Lawrie, R. K. Lenroot, K. O. Lim, C. Lopez-Jaramillo, A. Lyall, V. Magnotta, R. C. W. Mandl, D. H. Mathalon, R. W. McCarley, S. McCarthy-Jones, C. McDonald, S. McEwen, A. McIntosh, T. Melicher, R. I. Mesholam-Gately, P. T. Michie, B. Mowry, B. A. Mueller, D. T. Newell, P. O'Donnell, V. Oertel-Knöchel, L. Oestreich, S. A. Paciga, C. Pantelis, O. Pasternak, G. Pearlson, G. R. Pellicano, A. Pereira, J. Pineda Zapata, F. Piras, S. G. Potkin, A. Preda, P. E. Rasser, D. R. Roalf, R. Roiz, A. Roos, D. Rotenberg, T. D. Satterthwaite, P. Savadjiev, U. Schall, R. J. Scott, M. L. Seal, L. J. Seidman, C. S. Weickert, C. D. Whelan, M. E. Shenton, J. S. Kwon, G. Spalletta, F. Spaniel, E. Sprooten, M. Stäblein, D. J. Stein, S. Sundram, Y. Tan, S. Tan, S. Tang, H. S. Temmingh, L. T. Westlye, S. Tønnesen, D.

Tordesillas-Gutierrez, N. T. Doan, J. Vaidya, N. E. M. van Haren, C. D. Vargas, D. Vecchio, D. Velakoulis, A. Voineskos, J. Q. Voyvodic, Z. Wang, P. Wan, D. Wei, T. W. Weickert, H. Whalley, T. White, T. J. Whitford, J. D. Wojcik, H. Xiang, Z. Xie, H. Yamamori, F. Yang, N. Yao, G. Zhang, J. Zhao, T. G. M. van Erp, J. Turner, P. M. Thompson, G. Donohoe, Widespread white matter microstructural differences in schizophrenia across 4322 individuals: Results from the ENIGMA Schizophrenia DTI Working Group. *Mol. Psychiatry* **23**, 1261–1269 (2018).

68. M. J. Gandal, J. R. Haney, N. N. Parikshak, V. Leppa, G. Ramaswami, C. Hartl, A. J. Schork, V. Appadurai, A. Buil, T. M. Werge, C. Liu, K. P. White; CommonMind Consortium; PsychENCODE Consortium; iPSYCH-BROAD Working Group, S. Horvath, D. H. Geschwind, Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693–697 (2018).

69. M. J. Gandal, P. Zhang, E. Hadjimichael, R. L. Walker, C. Chen, S. Liu, H. Won, H. van Bakel, M. Varghese, Y. Wang, A. W. Shieh, J. Haney, S. Parhami, J. Belmont, M. Kim, P. M. Losada, Z. Khan, J. Mleczko, Y. Xia, R. Dai, D. Wang, Y. T. Yang, M. Xu, K. Fish, P. R. Hof, J. Warrell, D. Fitzgerald, K. White, A. E. Jaffe; PsychENCODE Consortium, M. A. Peters, M. Gerstein, C. Liu, L. M. Iakoucheva, D. Pinto, D. H. Geschwind, Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaa8127 (2018).

70. I. Voineagu, X. Wang, P. Johnston, J. K. Lowe, Y. Tian, S. Horvath, J. Mill, R. M. Cantor, B. J. Blencowe, D. H. Geschwind, Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011).

71. J. M. Williamson, D. A. Lyons, Myelin dynamics throughout life: An ever-changing landscape? *Front. Cell. Neurosci.* **12**, 424 (2018).

72. T. S. Johnson, S. Xiang, T. Dong, Z. Huang, M. Cheng, T. Wang, K. Yang, D. Ni, K. Huang, J. Zhang, Combinatorial analyses reveal cellular composition changes have different impacts on transcriptomic changes of cell type specific genes in Alzheimer's Disease. *Sci. Rep.* **11**, 353 (2021).

73. Z. Li, J. L. Del-Aguila, U. Dube, J. Budde, R. Martinez, K. Black, Q. Xiao, N. J. Cairns, J. D. Dougherty, J.-M. Lee, J. C. Morris, R. J. Bateman, C. M. Karch, C. Cruchaga, O. Harari, Genetic

variants associated with Alzheimer's disease confer different cerebral cortex cell-type population structure. *Genome Med.* **10**, 43 (2018).

74. J. L. Min, G. Hemani, G. Davey Smith, C. Relton, M. Suderman, Meffil: Efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics* **34**, 3983–3989 (2018).

75. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).

76. J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, J. D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).

77. W. Zhou, P. W. Laird, H. Shen, Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res.* **45**, e22 (2017).

78. C. Caggiano, B. Celona, F. Garton, J. Mefford, B. L. Black, R. Henderson, C. Lomen-Hoerth, A. Dahl, N. Zaitlen, Comprehensive cell type decomposition of circulating cell-free DNA with CelFiE. *Nat. Commun.* **12**, 2717 (2021).

79. M. J. Aryee, A. E. Jaffe, H. Corrada-Bravo, C. Ladd-Acosta, A. P. Feinberg, K. D. Hansen, R. A. Irizarry, Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).

80. J. Guintivano, M. J. Aryee, Z. A. Kaminsky, A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. *Epigenetics* **8**, 290–302 (2013).

81. D. Zhou, B. M. Alver, S. Li, R. A. Hlady, J. J. Thompson, M. A. Schroeder, J.-H. Lee, J. Qiu, P. H. Schwartz, J. N. Sarkaria, K. D. Robertson, Distinctive epigenomes characterize glioma stem cells and their response to differentiation cues. *Genome Biol.* **19**, 43 (2018).

82. R. Lucero, V. Zappulli, A. Sammarco, O. D. Murillo, P. S. Cheah, S. Srinivasan, E. Tai, D. T. Ting, Z. Wei, M. E. Roth, L. C. Laurent, A. M. Krichevsky, X. O. Breakefield, A. Milosavljevic, Glioma-derived

miRNA-containing extracellular vesicles induce angiogenesis by reprogramming brain endothelial cells. *Cell Rep.* **30**, 2065–2074.e4 (2020).

83. S. Policicchio, J. P. Davies, B. Chioza, J. Burrage, J. Mill, E. Dempster, Fluorescence-activated nuclei sorting (FANS) on human post-mortem cortex tissue enabling the isolation of distinct neural cell populations for multiple omic profiling V.1 (2020); www.protocols.io/view/fluorescence-activated-nuclei-sorting-fans-on-huma-bmh2k38e.

84. L. E. Reinius, N. Acevedo, M. Joerink, G. Pershagen, S.-E. Dahlén, D. Greco, C. Söderhäll, A. Scheynius, J. Kere, Differential DNA methylation in purified human blood cells: Implications for cell lineage and studies on disease susceptibility. *PLOS ONE* **7**, e41361 (2012).

85. D. S. Vellame, G. Shireby, A. MacCalman, E. L. Dempster, J. Burrage, T. Gorrie-Stone, L. S. Schalkwyk, J. Mill, E. Hannon, Uncertainty quantification of reference-based cellular deconvolution algorithms. *Epigenetics* **18**, 2137659 (2023).

86. C. Fuchsberger, G. R. Abecasis, D. A. Hinds, minimac2: Faster genotype imputation. *Bioinformatics* **31**, 782–784 (2015).

87. B. Howie, C. Fuchsberger, M. Stephens, J. Marchini, G. R. Abecasis, Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).

88. S. McCarthy, S. Das, W. Kretzschmar, O. Delaneau, A. R. Wood, A. Teumer, H. M. Kang, C. Fuchsberger, P. Danecek, K. Sharp, Y. Luo, C. Sidore, A. Kwong, N. Timpson, S. Koskinen, S. Vrieze, L. J. Scott, H. Zhang, A. Mahajan, J. Veldink, U. Peters, C. Pato, C. M. van Duijn, C. E. Gillies, I. Gandin, M. Mezzavilla, A. Gilly, M. Cocca, M. Traglia, A. Angius, J. C. Barrett, D. Boomsma, K. Branham, G. Breen, C. M. Brummett, F. Busonero, H. Campbell, A. Chan, S. Chen, E. Chew, F. S. Collins, L. J. Corbin, G. D. Smith, G. Dedoussis, M. Dorr, A.-E. Farmaki, L. Ferrucci, L. Forer, R. M. Fraser, S. Gabriel, S. Levy, L. Groop, T. Harrison, A. Hattersley, O. L. Holmen, K. Hveem, M. Kretzler, J. C. Lee, M. McGue, T. Meitinger, D. Melzer, J. L. Min, K. L. Mohlke, J. B. Vincent, M. Nauck, D. Nickerson, A. Palotie, M. Pato, N. Pirastu, M. McInnis, J. B. Richards, C. Sala, V. Salomaa, D. Schlessinger, S. Schoenherr, P. E. Slagboom, K. Small, T. Spector, D. Stambolian, M. Tuke, J. Tuomilehto, L. H. Van den Berg, W. Van Rheenen, U. Volker, C. Wijmenga, D. Toniolo, E. Zeggini, P.

Gasparini, M. G. Sampson, J. F. Wilson, T. Frayling, P. I. W. de Bakker, M. A. Swertz, S. McCarroll, C. Kooperberg, A. Dekker, D. Altshuler, C. Willer, W. Iacono, S. Ripatti, N. Soranzo, K. Walter, A. Swaroop, F. Cucca, C. A. Anderson, R. M. Myers, M. Boehnke, M. I. McCarthy, R. Durbin, G. Abecasis, J. Marchini; Haplotype Reference Consortium, A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).

89. D. Taliun, D. N. Harris, M. D. Kessler, J. Carlson, Z. A. Szpiech, R. Torres, S. A. G. Taliun, A. Corvelo, S. M. Gogarten, H. M. Kang, A. N. Pitsillides, J. LeFaive, S. Lee, X. Tian, B. L. Browning, S. Das, A.-K. Emde, W. E. Clarke, D. P. Loesch, A. C. Shetty, T. W. Blackwell, A. V. Smith, Q. Wong, X. Liu, M. P. Conomos, D. M. Bobo, F. Aguet, C. Albert, A. Alonso, K. G. Ardlie, D. E. Arking, S. Aslibekyan, P. L. Auer, J. Barnard, R. G. Barr, L. Barwick, L. C. Becker, R. L. Beer, E. J. Benjamin, L. F. Bielak, J. Blangero, M. Boehnke, D. W. Bowden, J. A. Brody, E. G. Burchard, B. E. Cade, J. F. Casella, B. Chalazan, D. I. Chasman, Y.-D. I. Chen, M. H. Cho, S. H. Choi, M. K. Chung, C. B. Clish, A. Correa, J. E. Curran, B. Custer, D. Darbar, M. Daya, M. de Andrade, D. L. DeMeo, S. K. Dutcher, P. T. Ellinor, L. S. Emery, C. Eng, D. Fatkin, T. Fingerlin, L. Forer, M. Fornage, N. Franceschini, C. Fuchsberger, S. M. Fullerton, S. Germer, M. T. Gladwin, D. J. Gottlieb, X. Guo, M. E. Hall, J. He, N. L. Heard-Costa, S. R. Heckbert, M. R. Irvin, J. M. Johnsen, A. D. Johnson, R. Kaplan, S. L. R. Kardia, T. Kelly, S. Kelly, E. E. Kenny, D. P. Kiel, R. Klemmer, B. A. Konkle, C. Kooperberg, A. Köttgen, L. A. Lange, J. Lasky-Su, D. Levy, X. Lin, K.-H. Lin, C. Liu, R. J. F. Loos, L. Garman, R. Gerszten, S. A. Lubitz, K. L. Lunetta, A. C. Y. Mak, A. Manichaikul, A. K. Manning, R. A. Mathias, D. D. McManus, S. T. McGarvey, J. B. Meigs, D. A. Meyers, J. L. Mikulla, M. A. Minear, B. D. Mitchell, S. Mohanty, M. E. Montasser, C. Montgomery, A. C. Morrison, J. M. Murabito, A. Natale, P. Natarajan, S. C. Nelson, K. E. North, J. R. O'Connell, N. D. Palmer, N. Pankratz, G. M. Peloso, P. A. Peyser, J. Pleiness, W. S. Post, B. M. Psaty, D. C. Rao, S. Redline, A. P. Reiner, D. Roden, J. I. Rotter, I. Ruczinski, C. Sarnowski, S. Schoenherr, D. A. Schwartz, J.-S. Seo, S. Seshadri, V. A. Sheehan, W. H. Sheu, M. B. Shoemaker, N. L. Smith, J. A. Smith, N. Sotoodehnia, A. M. Stilp, W. Tang, K. D. Taylor, M. Telen, T. A. Thornton, R. P. Tracy, D. J. Van Den Berg, R. S. Vasan, K. A. Viaud-Martinez, S. Vrieze, D. E. Weeks, B. S. Weir, S. T. Weiss, L.-C. Weng, C. J. Willer, Y. Zhang, X. Zhao, D. K. Arnett, A. E. Ashley-Koch, K. C. Barnes, E. Boerwinkle, S. Gabriel, R. Gibbs, K. M. Rice, S. S. Rich, E. K. Silverman, P. Qasba, W. Gan; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, G. J. Papanicolaou, D. A. Nickerson, S. R. Browning, M. C. Zody, S. Zöllner, J. G. Wilson, L. A. Cupples,

C. C. Laurie, C. E. Jaquish, R. D. Hernandez, T. D. O'Connor, G. R. Abecasis, Sequencing of 53,831 diverse genomes from the NHLBI TOPMed program. *Nature* **590**, 290–299 (2021).

90. C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

91. Q. Zhang, J. Sidorenko, B. Couvy-Duchesne, R. E. Marioni, M. J. Wright, A. M. Goate, E. Marcora, K. Huang, T. Porter, S. M. Laws, P. S. Sachdev, K. A. Mather, N. J. Armstrong, A. Thalamuthu, H. Brodaty, L. Yengo, J. Yang, N. R. Wray, A. F. McRae, P. M. Visscher, Risk prediction of late-onset Alzheimer's disease implies an oligogenic architecture. *Nat. Commun.* **11**, 4799 (2020).

92. L. R. Lloyd-Jones, J. Zeng, J. Sidorenko, L. Yengo, G. Moser, K. E. Kemper, H. Wang, Z. Zheng, R. Magi, T. Esko, A. Metspalu, N. R. Wray, M. E. Goddard, J. Yang, P. M. Visscher, Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat. Commun.* **10**, 5086 (2019).

93. J. Morrison, N. Knoblauch, J. H. Marcus, M. Stephens, X. He, Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nat. Genet.* **52**, 740–747 (2020).

94. S. Burgess, A. Butterworth, S. G. Thompson, Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).

95. J. Bowden, G. Davey Smith, S. Burgess, Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).

96. J. Bowden, G. Davey Smith, P. C. Haycock, S. Burgess, Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).

97. F. P. Hartwig, G. Davey Smith, J. Bowden, Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998 (2017).

98. Z. Zhu, Z. Zheng, F. Zhang, Y. Wu, M. Trzaskowski, R. Maier, M. R. Robinson, J. J. McGrath, P. M. Visscher, N. R. Wray, J. Yang, Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**, 224 (2018).

99. C. J. Willer, Y. Li, G. R. Abecasis, METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

100. M. Kim, D. D. Vo, M. E. Kumagai, C. T. Jops, M. J. Gandal, GeneticsMakie.jl: A versatile and scalable toolkit for visualizing locus-level genetic and genomic data. *Bioinformatics* **39**, btac786 (2023).

101. T. Qi, Y. Wu, H. Fang, F. Zhang, S. Liu, J. Zeng, J. Yang, Genetic control of RNA splicing and its distinct role in complex trait variation. *Nat. Genet.* **54**, 1355–1363 (2022).

102. C. Giambartolomei, D. Vukcevic, E. E. Schadt, L. Franke, A. D. Hingorani, C. Wallace, V. Plagnol, Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLOS Genet.* **10**, e1004383 (2014).

103. N. de Klein, E. A. Tsai, M. Vochteloo, D. Baird, Y. Huang, C.-Y. Chen, S. van Dam, R. Oelen, P. Deelen, O. B. Bakker, O. El Garwany, Z. Ouyang, E. E. Marshall, M. I. Zavodszky, W. van Rheenen, M. K. Bakker, J. Veldink, T. R. Gaunt, H. Runz, L. Franke, H.-J. Westra, Brain expression quantitative trait locus and network analyses reveal downstream effects and putative drivers for brain-related diseases. *Nat. Genet.* **55**, 377–388 (2023).

104. R. Lister, E. A. Mukamel, J. R. Nery, M. Urich, C. A. Puddifoot, N. D. Johnson, J. Lucero, Y. Huang, A. J. Dwork, M. D. Schultz, M. Yu, J. Tonti-Filippini, H. Heyn, S. Hu, J. C. Wu, A. Rao, M. Esteller, C. He, F. G. Haghighi, T. J. Sejnowski, M. M. Behrens, J. R. Ecker, Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).

105. V. V. Gusel'nikova, D. E. Korzhevskiy, NeuN As a neuronal nuclear antigen and neuron differentiation marker. *Acta Naturae* **7**, 42–47 (2015).

106. J. Aitchison, Principal component analysis of compositional data. *Biometrika* **70**, 57–65 (1983).

107. T. P. Quinn, I. Erb, G. Gloor, C. Notredame, M. F. Richardson, T. M. Crowley, A field guide for the compositional analysis of any-omics data. *GigaScience* **8**, giz107 (2019).

108. K. G. Van den Boogaart, R. Tolosana-Delgado, *Analyzing Compositional Data with R* (Springer, 2013), vol. 122.

109. K. G. van den Boogaart, R. Tolosana-Delgado, "compositions": A unified R package to analyze compositional data. *Comput. Geosci.* **34**, 320–338 (2008).

110. P. I. Costea, G. Zeller, S. Sunagawa, P. Bork, A fair comparison. *Nat. Methods* **11**, 359–359 (2014).

111. R. E. Peterson, K. Kuchenbaecker, R. K. Walters, C.-Y. Chen, A. B. Popejoy, S. Periyasamy, M. Lam, C. Iyegbe, R. J. Strawbridge, L. Brick, C. E. Carey, A. R. Martin, J. L. Meyers, J. Su, J. Chen, A. C. Edwards, A. Kalungi, N. Koen, L. Majara, E. Schwarz, J. W. Smoller, E. A. Stahl, P. F. Sullivan, E. Vassos, B. Mowry, M. L. Prieto, A. Cuellar-Barboza, T. B. Bigdeli, H. J. Edenberg, H. Huang, L. E. Duncan, Genome-wide association studies in ancestrally diverse populations: Opportunities, methods, pitfalls, and recommendations. *Cell* **179**, 589–603 (2019).

112. A. R. Martin, M. Kanai, Y. Kamatani, Y. Okada, B. M. Neale, M. J. Daly, Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).