



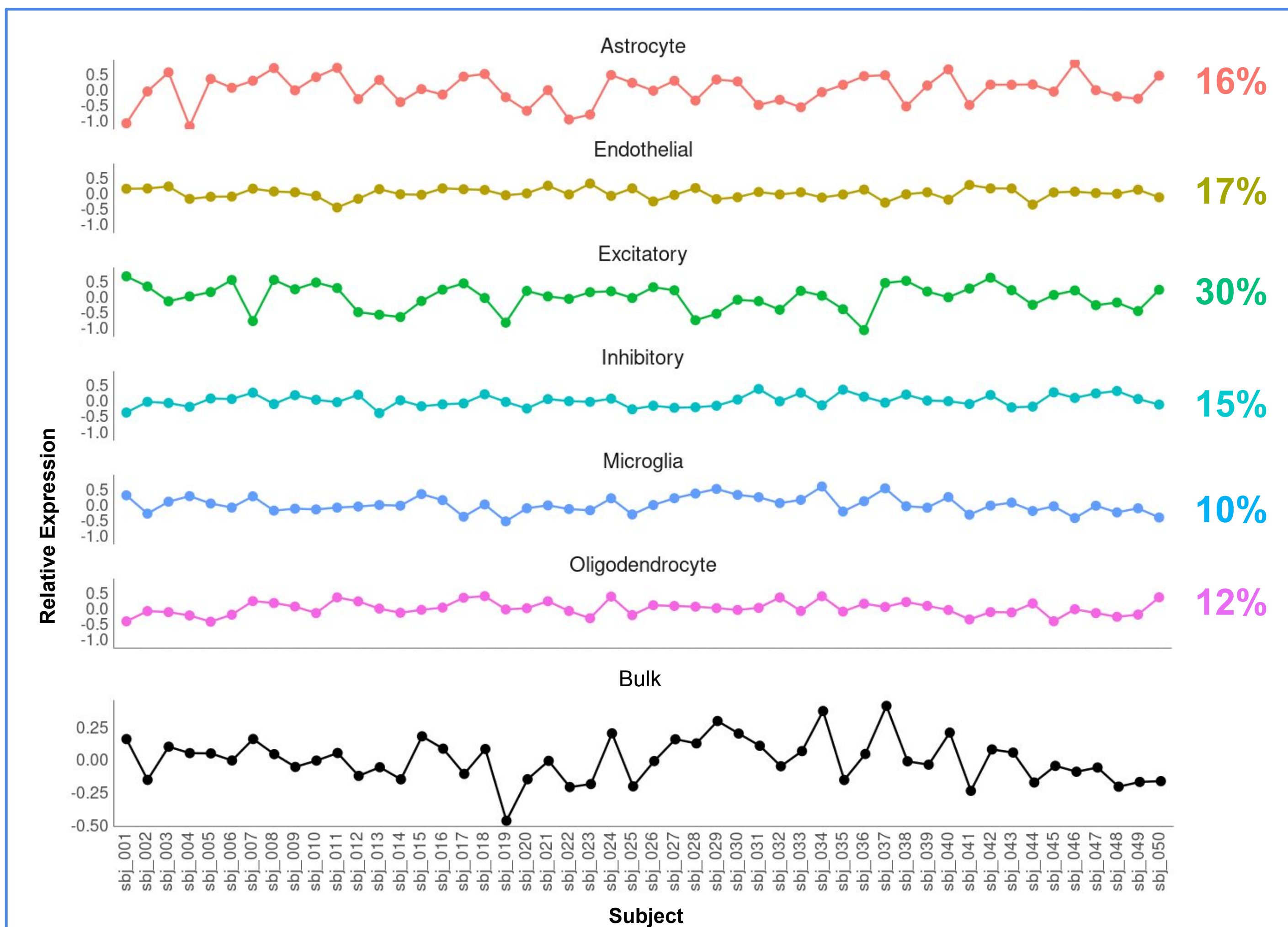
Barriers to detecting cell type specific regulatory signals in gene co-expression

Eric Ching-Pan Chu^{1,2,3}, Alexander Morin^{1,2,3}, Marine Louarn³ Paul Pavlidis^{2,3}
¹UBC Graduate Program in Bioinformatics, ²Michael Smith Laboratories, ³UBC Department of Psychiatry

BACKGROUND

- Bulk tissue co-expression analysis is often carried out for the purpose of inferring intra-cellular gene regulation
- Samples contain multiple different cell types so cell type specific expression patterns would undergo dilution
- We used data simulations as well as analysis of real single cell and bulk tissue data sets to explore conditions required for the propagation of within-cell type co-expression signals to the bulk tissue level

Each bulk tissue sample contains a mixture of different cell types



Equation 1:

$$\text{expr}(\text{bulk}) = \sum \alpha(\text{ct}) * \text{expr}(\text{ct})$$

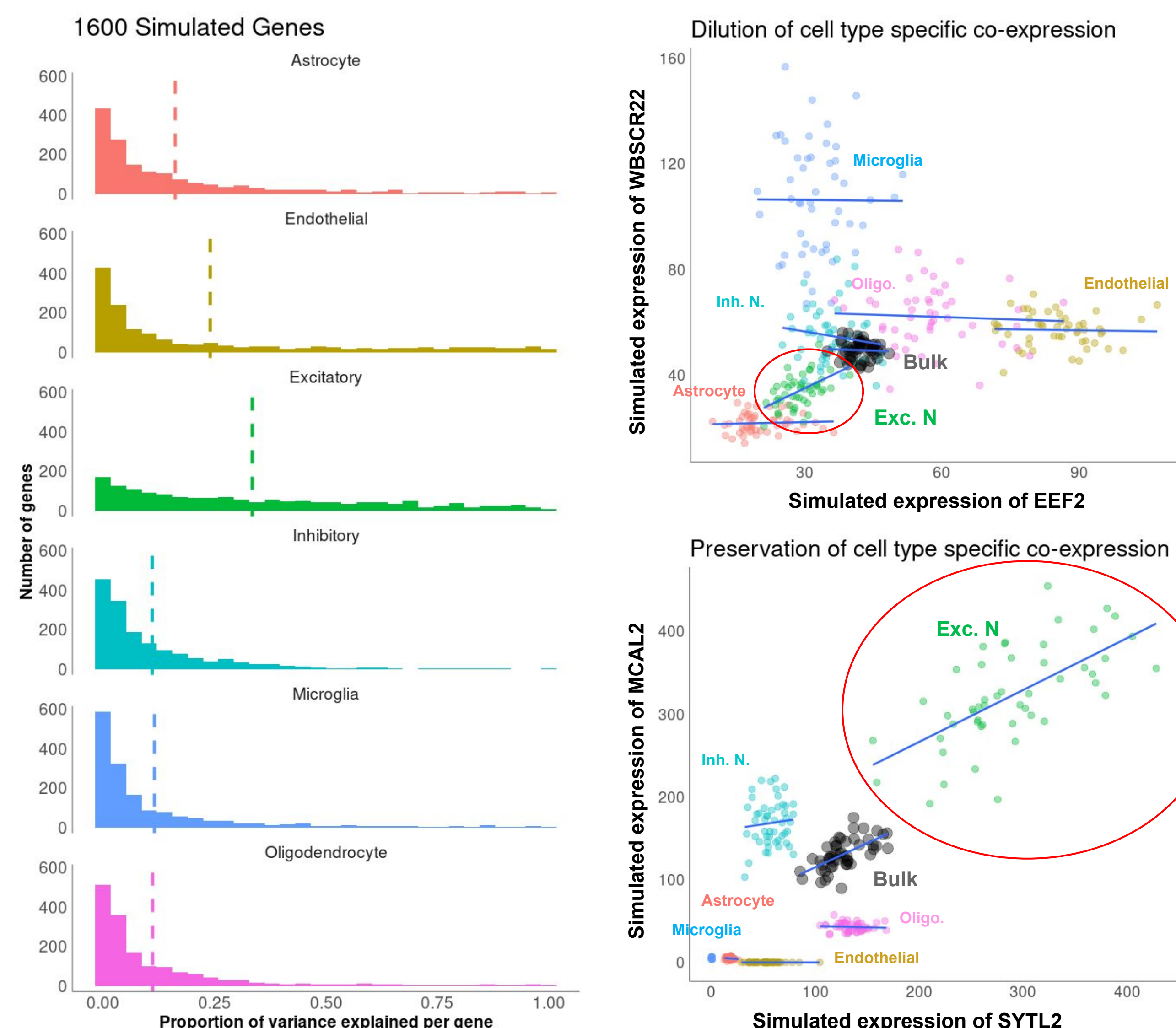
Where

- $\alpha(\text{ct})$ is the proportion of cell type ct
- $\text{expr}(\text{ct})$ is the cell type specific expression vector

Under this formulation, the relative contribution of variability from each cell type is the product of the cell type's proportion - α (ct) and variance of the cell type expression - $\text{expr}(\text{ct})$. *Note that this model ignores the effects of cellular composition variability (CCV). Instead, this model makes the simplifying assumption that bulk tissue level expression patterns could entirely derived from the combination of all cell type specific variation.

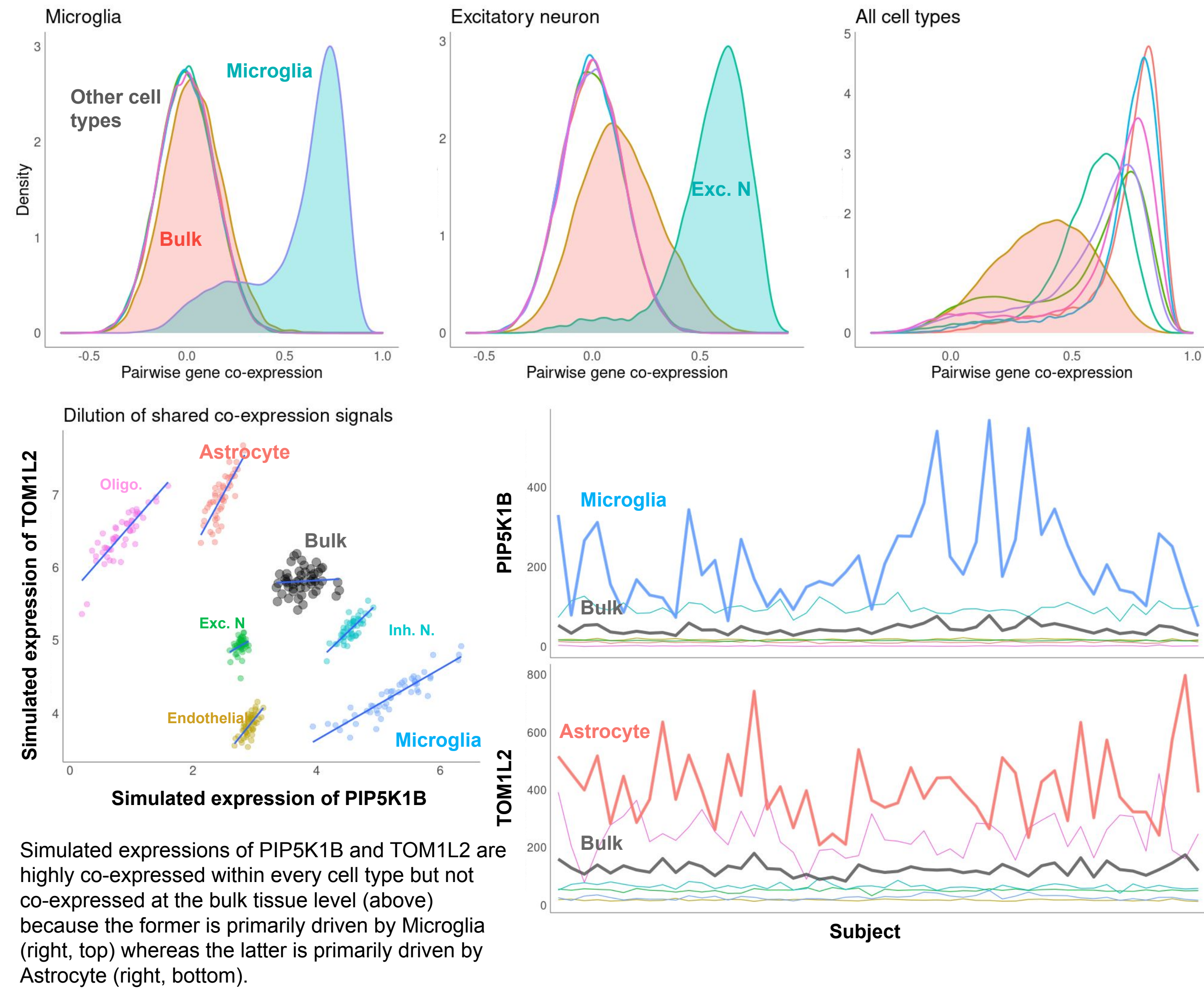
Simulation: Most cell type specific expression patterns are diluted out

We developed a simulation framework to generate single cell expression data. The simulated expressions mimics real data characteristics, and allows artificial imposition of gene co-expression relationships. Bulk level expression profiles are then produced by aggregating single cell simulations using the model depicted above (Equation 1). We simulated 1600 genes. Below: The left panel shows the distribution of proportion of variance explained (PVE) per gene by each cell type. Excitatory neurons, present at 30% per sample, is slightly more resistant to the effect of dilution than other cell types overall. Cell type specific variability is another factor that determines dilution impact, as shown on the right figures.



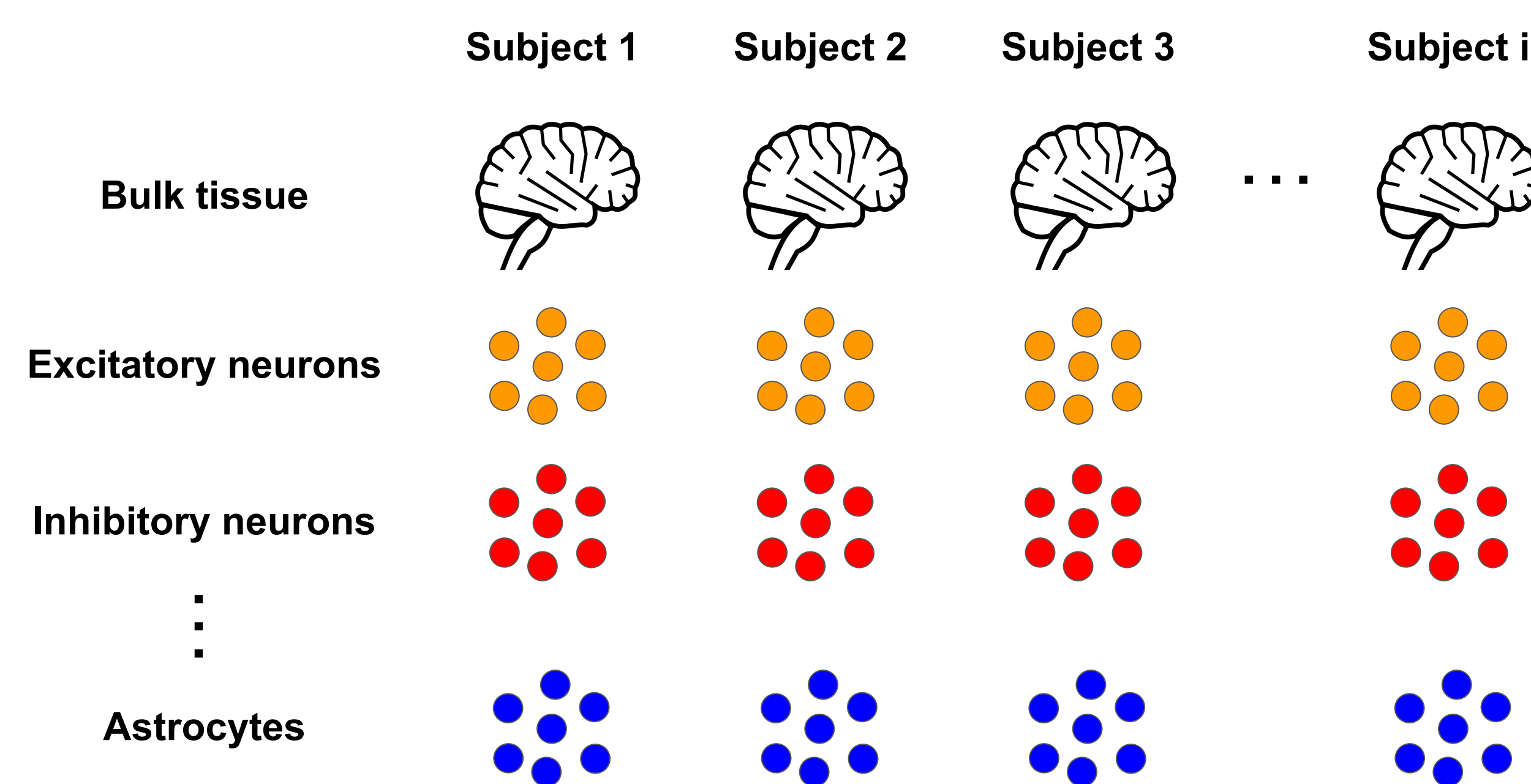
Simulation: Co-expression shared among multiple cell types are not immune to the effects of dilution

Below: Distribution of pairwise co-expressions (Pearson) among genes with strong within cell type co-expression patterns. Left: 200 genes highly co-expressed specifically in microglia. The overwhelming majority of co-expressions are diluted due to the small proportion of microglia (10%) in bulk tissue samples. Centre: 200 genes co-expressed specifically in excitatory neurons. Right: 200 genes highly co-expressed in all cell types. While there is an elevation from the null distribution, it is clear that most signals have undergone some dilution.

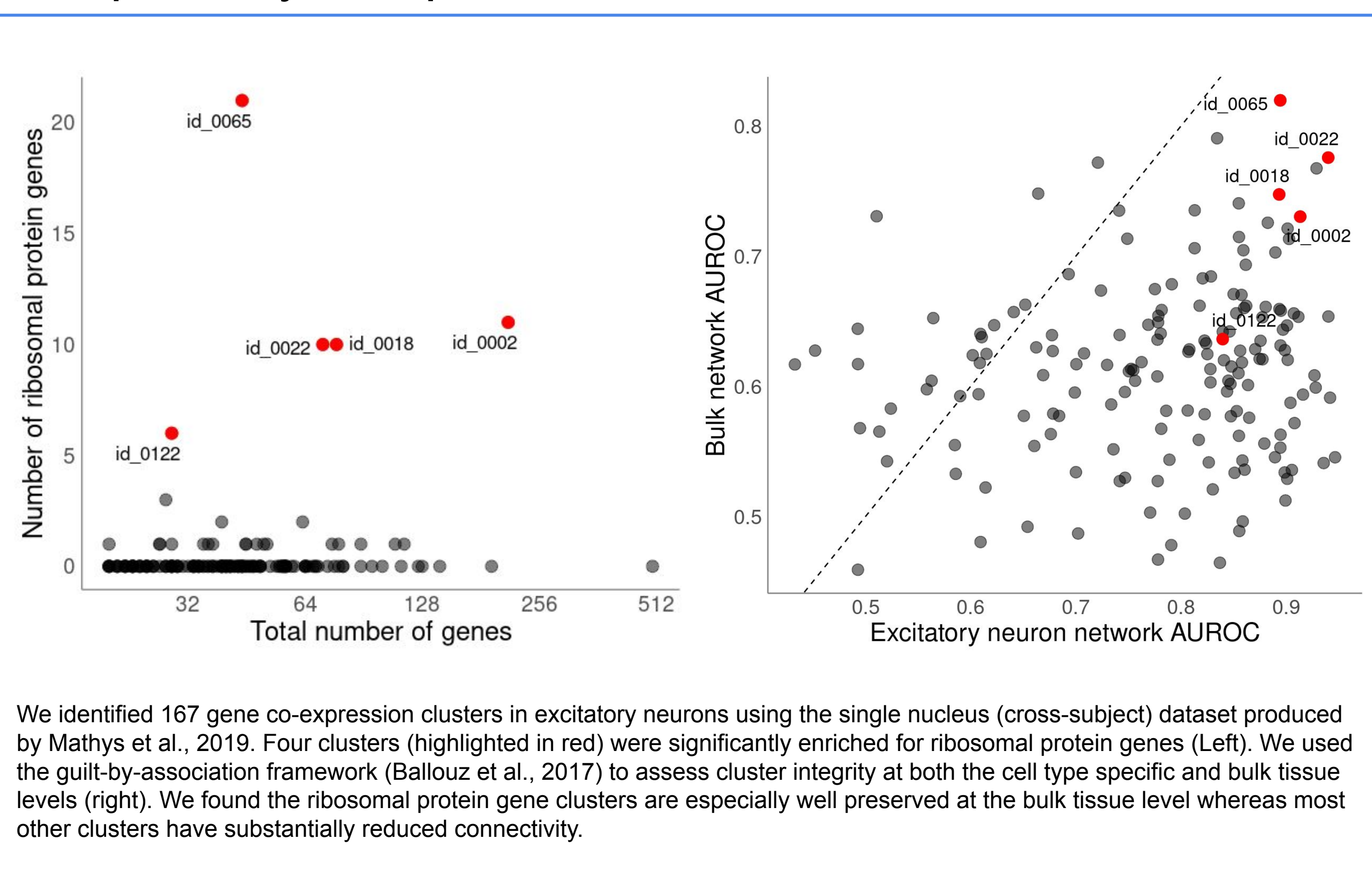


Matching single cell and bulk tissue data enable comparison between cell type specific and bulk level (co-)expression patterns

*Mathys et al., 2019 produced a dataset with 80,660 single-cell transcriptomes across 48 subjects, among which 28 subjects also have bulk tissue data that were produced as part of the ROSMAP cohort (Bennet et al., 2018).

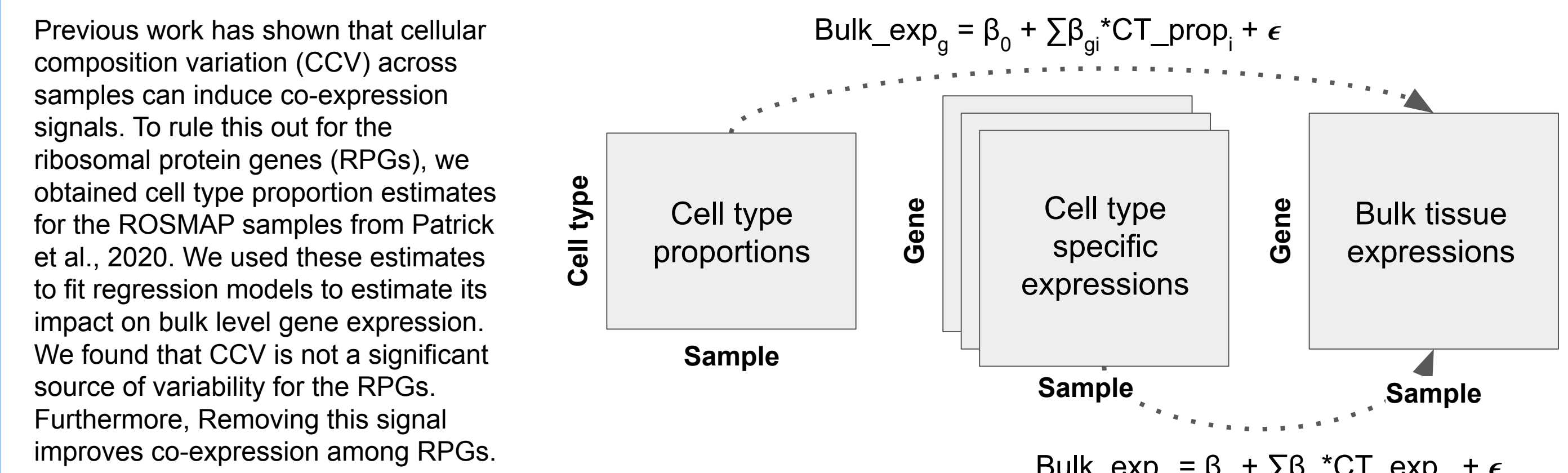


Co-expressions among ribosomal protein genes are exceptionally well preserved

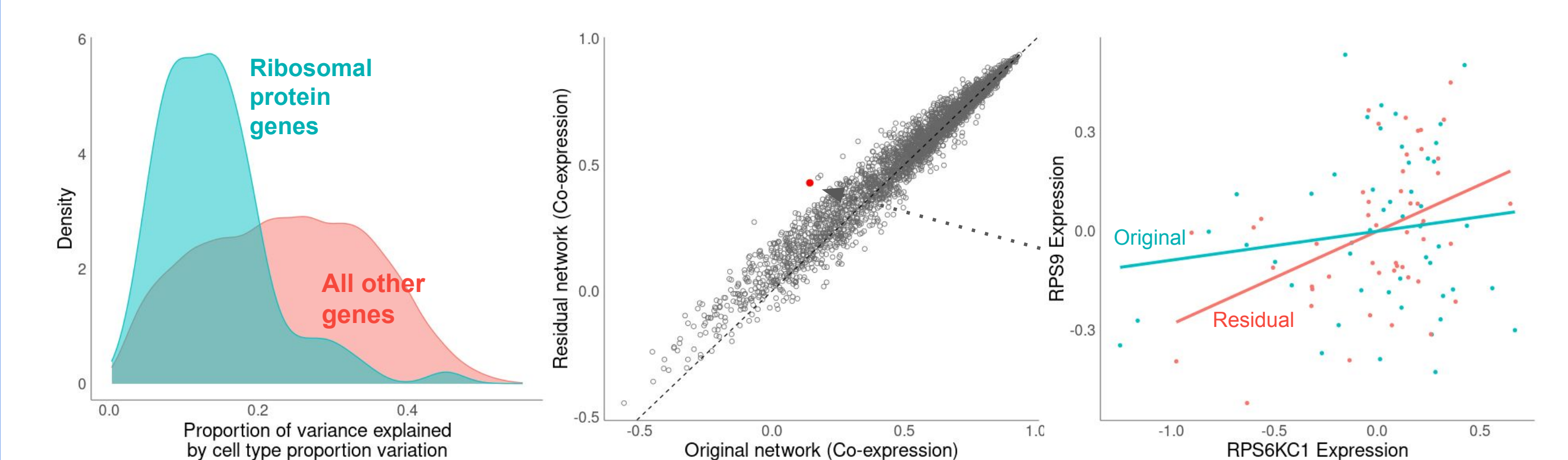


We identified 167 gene co-expression clusters in excitatory neurons using the single nucleus (cross-subject) dataset produced by Mathys et al., 2019. Four clusters (highlighted in red) were significantly enriched for ribosomal protein genes (Left). We used the guilt-by-association framework (Ballouz et al., 2017) to assess cluster integrity at both the cell type specific and bulk tissue levels (right). We found the ribosomal protein gene clusters are especially well preserved at the bulk tissue level whereas most other clusters have substantially reduced connectivity.

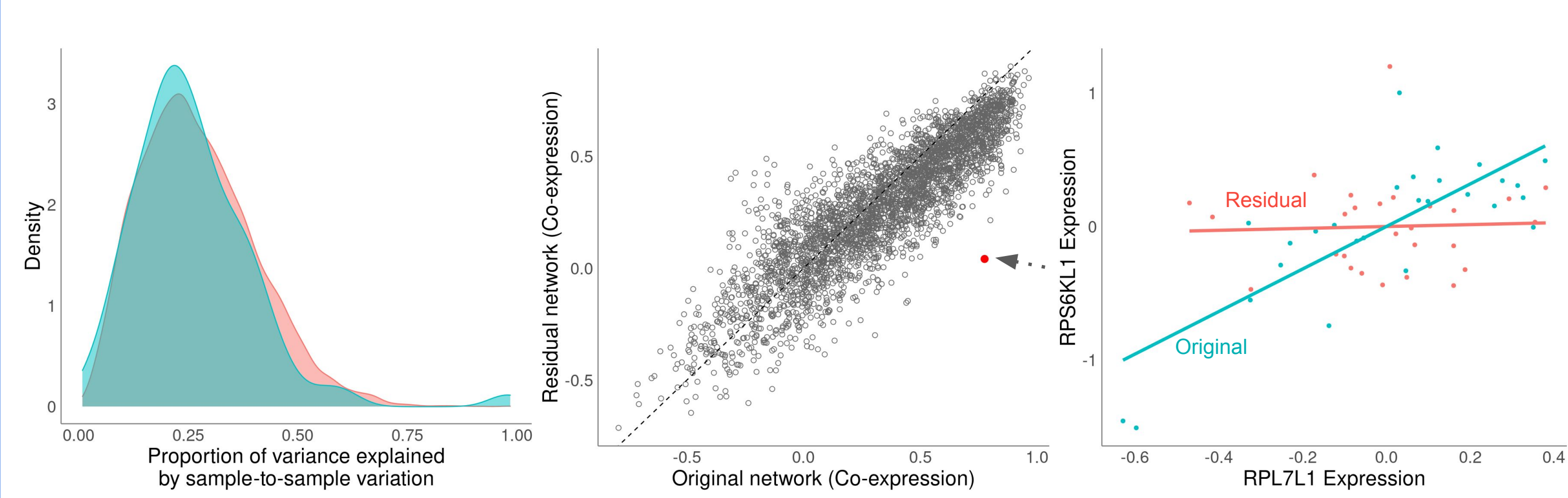
Co-expression among ribosomal protein genes (RPGs) are unlikely driven by cellular composition variation (CCV)



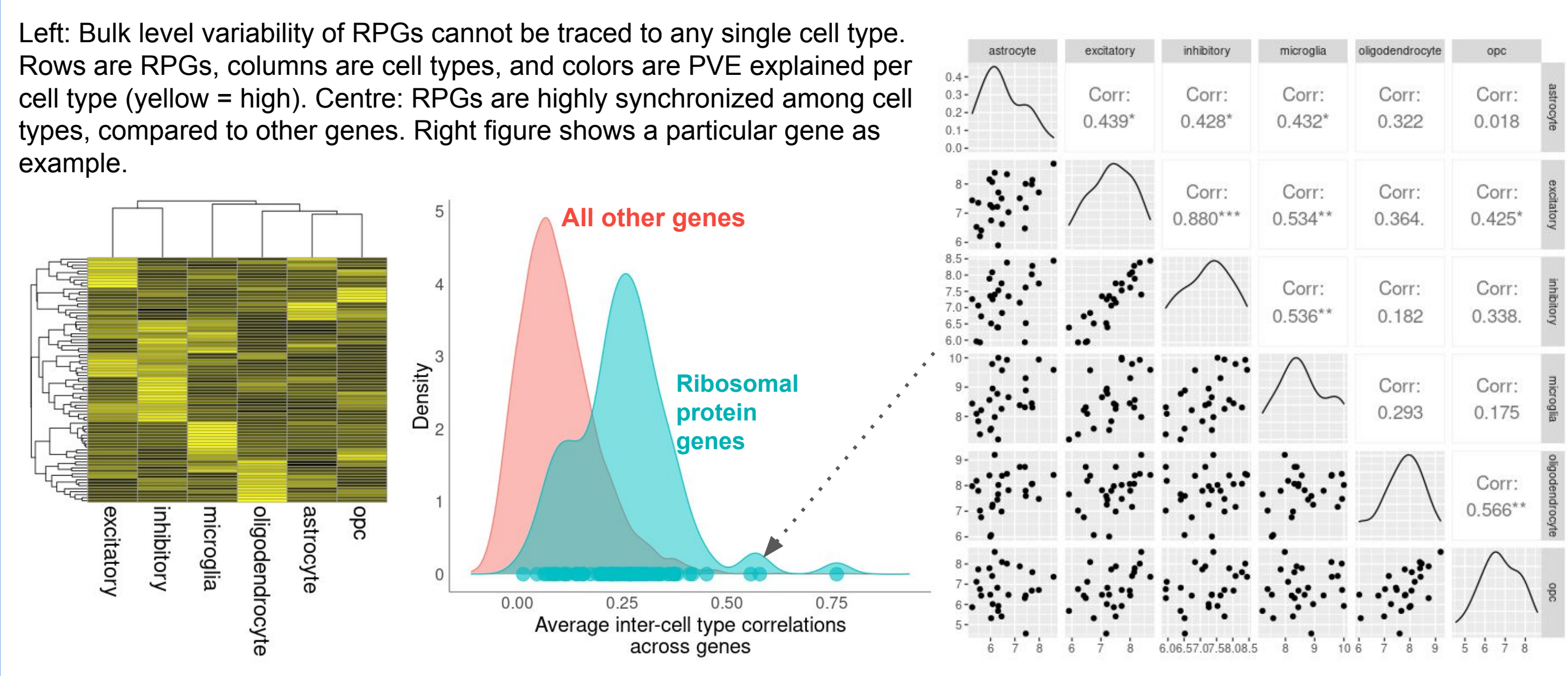
Below: Explaining level expression patterns as a function of CCV across bulk tissue samples. Left: Distribution of proportion of variance explained (PVE) per gene. Centre: The impact on co-expression among RPGs upon removal CCV as a source of variability. Right: An example of a pairwise gene co-expression that improves upon CCV removal.



Below: Explaining bulk level expression patterns as a function of cell type specific expression profile across subjects. Left: Distribution of PVE per gene. Centre: The impact on co-expression among RPGs upon removal the underlying cell type specific expression profiles. Right: An example of a pairwise gene co-expression that decreases upon signal removal.



Inter-cell type synchrony may explain the preservation of co-expression among ribosomal protein genes



SUMMARY

- Dilution among cell types likely removes most of cell type specific expression patterns
- Ribosomal protein genes are a exceptional group of genes whose co-expressions may be preserved via inter-cell type synchrony
- These findings were corroborated by analysis of other data
- The simulation framework will be made available

References

Ballouz S, Weber M, Pavlidis P, Gillis J. EGAD: ultra-fast functional analysis of gene networks. *Bioinformatics*. 2017 Feb 15;33(4):612-4.
Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, Menon M, He L, Abdurrob F, Jiang X, Martorell AJ. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*. 2019 Jun;570(7761):332-7.
Patrick E, Taga M, Ergun A, Ng B, Casazza W, Cimpean M, Yung C, Schneider JA, Bennett DA, Gaiteri C, De Jager PL. Deconvolving the contributions of cell-type heterogeneity on cortical gene expression. *PLOS Computational Biology*. 2020 Aug 17;16(8):e1008120.
Bennett DA, Buchman AS, Boyle PA, Barnes LL, Wilson RS, Schneider JA. Religious orders study and rush memory and aging project. *Journal of Alzheimer's disease*. 2018 Jan 1;64(s1):S161-89.