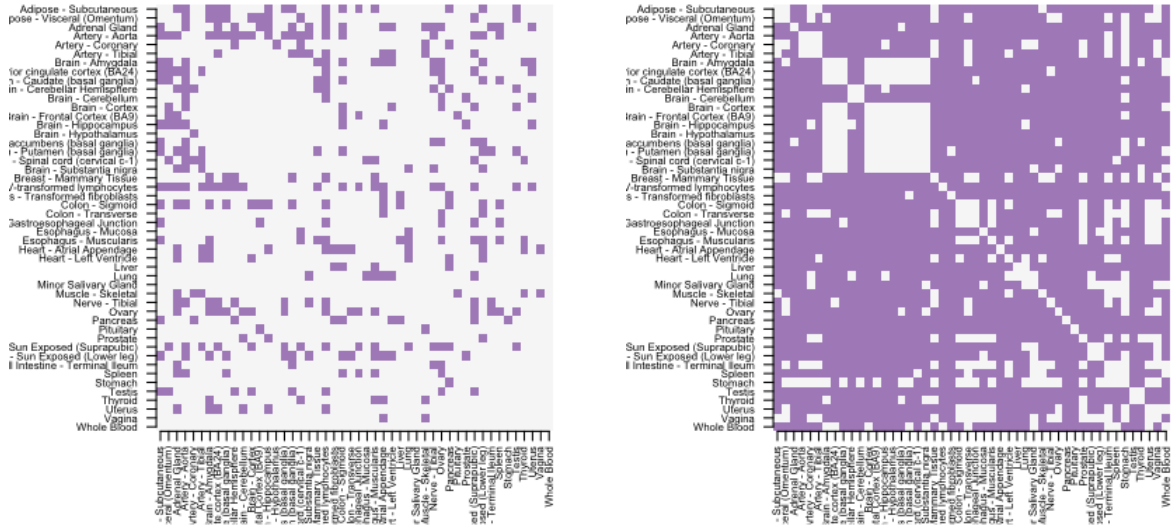


Fig 1. (a): Structure plot of estimated membership proportions for GoM model with $K = 20$ clusters fit to 8555 tissue samples from 53 tissues in GTEx data. Each horizontal bar shows the cluster membership proportions for a single sample, ordered so that samples from the same tissue are adjacent to one another. Within each tissue, the samples are sorted by the proportional representation of the underlying clusters. **(b):** Structure plot of estimated membership proportions for $K = 4$ clusters fit to only the brain tissue samples. This analysis highlights finer-scale structure among the brain samples that is missed by the global analysis in (a).



(a) hierarchy method

(b) GoM method

Fig 2. A comparison of accuracy of GoM model vs hierarchical clustering. For each pair of tissues from the GTEX data we assessed whether or not each method (with $K = 2$ clusters) separated the samples precisely according to their actual tissue of origin, with successful separation indicated by a filled square. Some pairs of tissues (e.g. pairs of brain tissues) are more difficult to distinguish than others. Overall the GoM model is successful in 86% comparisons and the hierarchical clustering in 39% comparisons.

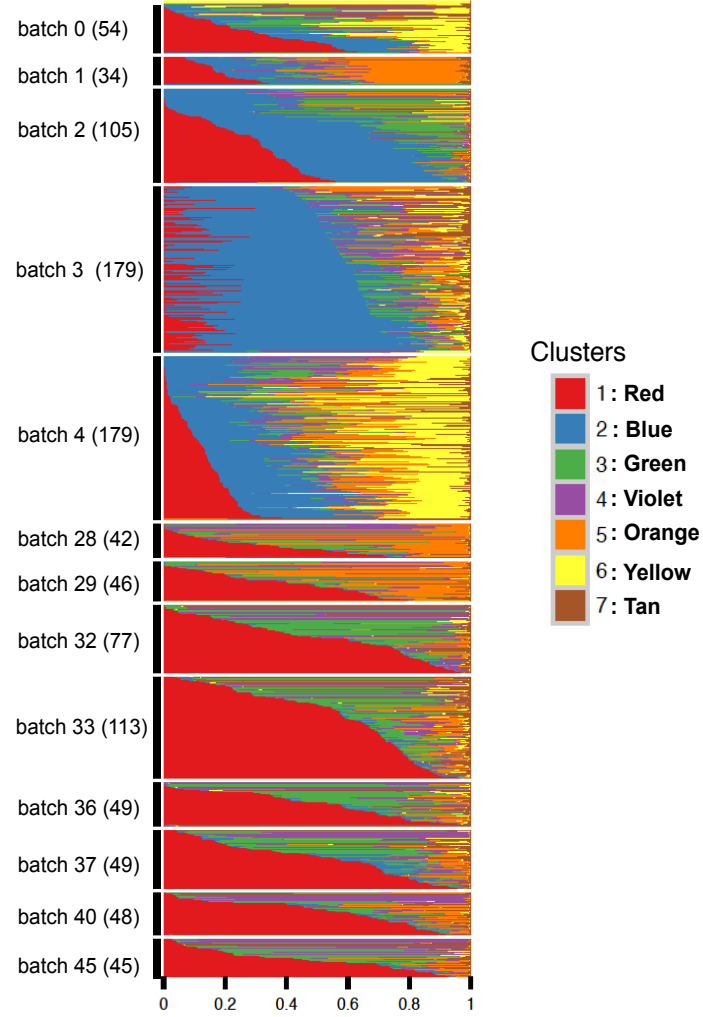


Fig 3. Structure plot of estimated membership proportions for GoM model with $K = 7$ clusters fit to 1,041 single cells from [?]. The samples (cells) are ordered so that samples from the same amplification batch are adjacent and within each batch, the samples are sorted by the proportional representation of the underlying clusters. In this analysis the samples do not appear to form clearly-defined clusters, with each sample being allocated membership in several “clusters”. Membership proportions are correlated with batch, and some groups of batches (e.g. 28-29; 32-45) show similar palettes. These results suggest that batch effects are likely influencing the inferred structure in these data.

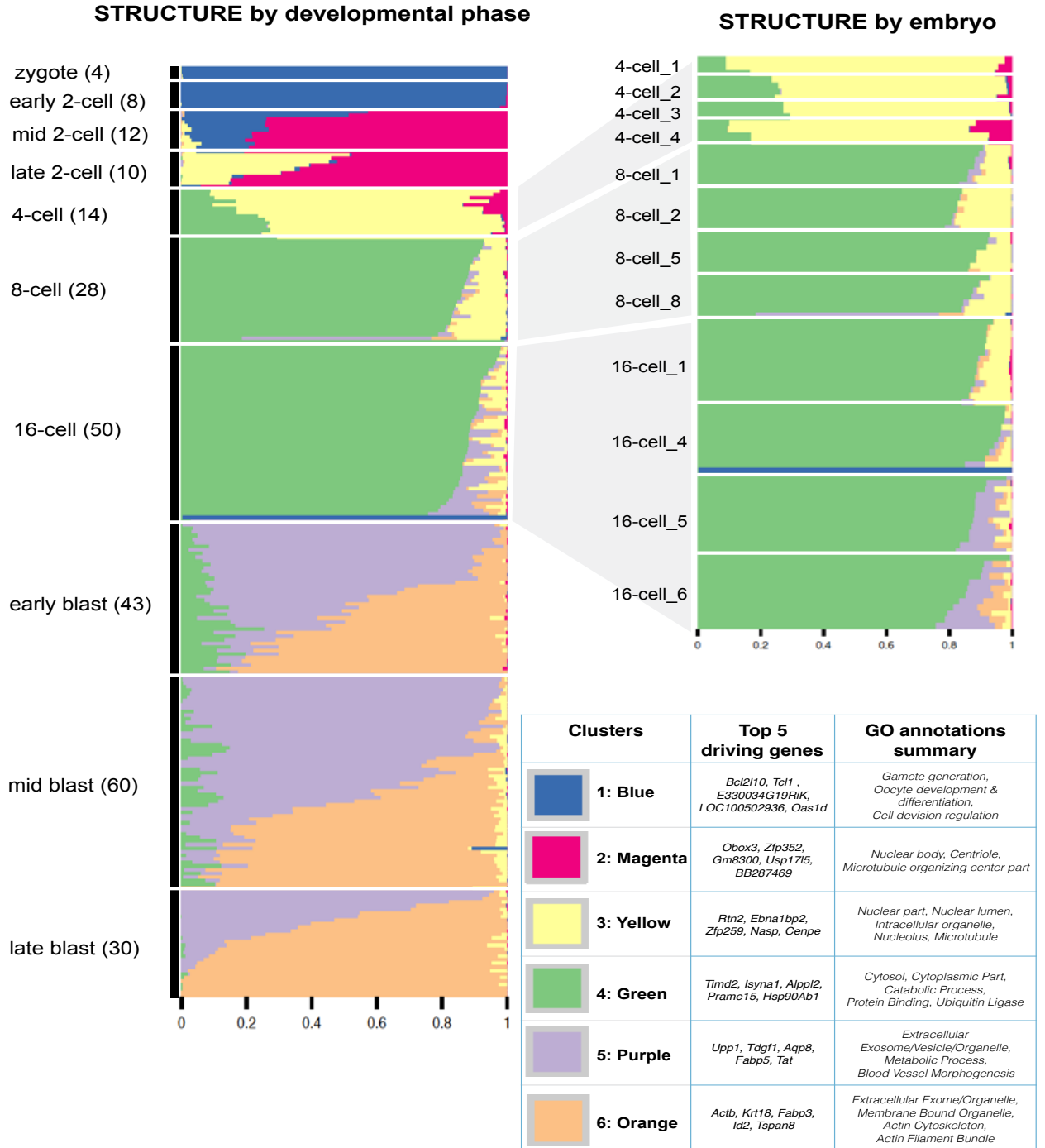


Fig 4. Structure plot of estimated membership proportions for GoM model with $K = 6$ clusters fit to 259 single cells from [?]. The cells are ordered by their preimplantation development phase (and within each phase, sorted by the proportional representation of the clusters). While the very earliest developmental phases (zygote and early 2-cell) are essentially assigned to a single cluster, others have membership in multiple clusters. Each cluster is annotated by the genes that are most distinctively expressed in that cluster, and by the gene ontology categories for which these distinctive genes are most enriched (see Table ?? for more extensive annotation results). See text for discussion of biological processes driving these results.