Million Cell Retina Transcriptome Meta-Atlas for All

To create an exhaustively optimized single cell retina single cell transcriptome dataset created from *all* publicly available retina single cell transcriptomes. We use this dataset to provide the basis for a highly responsive reactive web app for querying gene expression across retina cell type, study, species, developmental stage, and other factors. Second, we demonstrate how research groups can project their *private* single cell experiments onto our reference retina single cell atlas with minimal compute resources.

We re-quantified over 1 million single cell transcriptomes across three species, 30 studies, and seven single cell technologies in a unified framework. After a multi layers of quality control we retain over 700,000 high quality transcriptomes. We hand-curate over 300,000 published retina cell type labels from over a dozen published datasets. To optimize the batch correction, we tested 12 independent tools methods and benchmarked them with an un-biased algorithm to identify the optimal method and parameters. After batch correction we create a 2 dimensional UMAP projection, run thousands of differential gene expression tests across cell types, create an xgboost-based machine learning tool to label all cells in our dataset, and calculate developmental trajectories for each major retinal cell type.

After batch correction the 2D UMAP projection places the retina cell types in distinct spaces with the progenitor populations in the center. The photoreceptor progenitors flow from the center into the differentiated rods and cones. Likewise, the amacrine / horizontal precursors lead into the terminally differentiated amacrines and horizontals. The retinal ganglion and muller glia are also in distinct clusters. Importantly these cell type positions are generally consistent across species and studies. Our differential testing confirms that our machine learning cell type tool correctly labels the cells based on the community knowledge.

Despite a high amount of technical variation between published single cell transcriptome atlases, we confirm that we can create a coherent and high-quality meta-atlas. Furthermore, we demonstrate how our meta-atlas is a true community resource by illustrating projection of *independent* and *outside* single cell transcriptome data onto our meta-atlas with minimal compute and disk resources. Our dataset is made available for powerful user-led analysis at plae.nei.nih.gov.