The transcriptional identity of individual cells is derived from the additive effects of discrete basis vectors that describe dependent and independent contributions to the overall transcriptional state. Many methods exist to deconvolve expression matrices into their constitutive patterns. Most do not scale well to large datasets with complex sources of variation, a defining characteristic of single cell analyses. This limitation necessitates the ability to rapidly explore basis vectors, learned on smaller datasets, across larger datasets, and requires the development of statistical and visualization frameworks upon which to evaluate and compare models derived from different computational approaches.

To address this, we have developed ProjectoR, an R package that uses the relationships defined within a high dimensional data set to rapidly interrogate related phenomena in a new data set using Transfer Learning Methods (TLM). TLMs relax many of the constraints of other methods by using feature mappings that connect the samples and relationships, respectively.

We have established a collaborative network of researchers that aim to resolve and characterize discrete and continuous populations of cells through deconstruction and identification of basis vectors using a variety of approaches. As part of this network, we will establish a curated catalog of benchmark training and test data from the developing mouse and human retina, and evaluate and extend ProjectoR as a model visualization and comparison tool to enable rapid transfer learning of basis vectors across datasets at the scale of the HCA.