Our collaborative network includes several groups responding to this RFA. We shared proposals openly on GitHub and discussed their content via Slack. Selected interactions between groups are summarized in the proposal Figure and below. We look forward to these and additional collaborations within this group and the broader HCA consortium.

**Efficient factorization methods.**Several proposals include techniques for efficient factorization methods. For example, **Rob Patro** proposes these techniques to infer relevant features to quantify transcript abundance. We have recently brought together a large group of researchers in this domain to review applications of factorization methods in genomics (in preparation) and propose a JSM session on this topic. We look forward to collaborating with the consortia to develop state of the art factorization methods.

**Modeling transcriptional abundance and variance.** A unique feature of CoGAPS is its ability to infer genomics patterns from both expectation and variance estimates. The algorithm has concurrently found robust patterns associated with dynamics and patterns associated with technical co-variates, such as quantification or batch. Therefore, the algorithm both depends upon and can contribute to benchmarking transcriptional quantification methods such as those developed by **Rob Patro**. We will also consider denoising techniques and filtering, such as those developed by **Smita Krishnaswamy** and **Stephen Piccolo**.

**Sub-sampling methods. Casey Greene’s** deep learning techniques propose training on arbitrary patches are selected and arbitrary rotations of the data. This ensemble approach maps to the parallelization across subsets of samples in our proposal. We plan to work collaborate to develop methods for optimal feature selection across transcripts and samples for efficient and robust pattern inference. The benchmarking in our proposal will provide optimal sample distributions for all such algorithms relying on sub-sampling.

**Benchmark data.** Assessing performance of the proposed algorithm relies on presence of matched, time-course data across bulk and single-cell RNA-sequencing measurement technologies. Benchmark data in the human cell atlas will be critical to optimal algorithm development. We will also benefit from datasets such as the retinal development data in bulk, smart-seq, and 10X platforms provided by **Loyal Goff**. Algorithm development will further benefit from datasets with parallel perturbations in multiple datasets from **Arjun Raj**.

**Model interpretation.** Assessing the performance of unsupervised techniques relies critically on interpretation relative to known covariates, gene function, and independent test datasets. **Loyal Goff** proposes new algorithms for unsupervised model comparison with our shared postdoctoral fellow, Genevieve Stein-O’Brien. These techniques will enable interpreting the function of our inferred transcriptional trajectories relative to the gene signatures, such as those also developed with unsupervised learning in **Casey Greene**’sand **Smita Krisnaswamy**’sproposal. Visualization is also critical to such interpretation. Further collaboration with **Lana Garmire** and **Smita Krisnaswamy** will enhance and enable automated visualization of unsupervised patterns that we learn from transcriptional data with CoGAPS. We look forward to continued interactions with the consortia to optimize unsupervised model interpretation.