**Project Summary**

We will develop new bioinformatics pattern detection algorithms to infer developmental trajectories and cell fate decisions from both bulk and single-cell RNA-sequencing data. Gene expression data from healthy human tissues have variation arising from cell fate trajectories, cell cycle, and cell-type specific variation. Single cell RNA-sequencing is a promising data modality to measured these distinct processes in each cell. Computational techniques are essential to infer the transcriptional patterns that represent the underlying biological dynamics from these data. We have developed a sparse, Bayesian non-negative matrix factorization (NMF), CoGAPS. Previous work has demonstrated that CoGAPS can simultaneously distinguish robust, dynamic trajectories and individual variation in bulk gene expression data. Embedding data-specific error models has also enabled CoGAPS to perform multi-modal data integration. In this proposal, we will modify CoGAPS to input the underlying estimates of transcript abundance, transcriptional variance, and sparsity characteristic of single-cell RNA-sequencing data. Preliminary application of CoGAPS to this data has found spiraling patterns associated with rapid oscillations in the cell cycle. Although CoGAPS is robust, the large size of single cell datasets proposed in the HCA is prohibitive or algorithm convergence. Therefore, we propose parallelization of CoGAPS to converge robustly for genome-wide data of the scope of the human cell atlas without requiring data compaction. Additional modeling of sparsity in distinct single cell and bulk sequencing technologies will also enable unprecedented, multimodal data integration. Active participating with the collaborative network and consortium will yield biological and technical advances to further improve algorithm development.