Fig. 1

D_{ata} = $A_{mplitudes}$ x $P_{atterns}$

(gene weights for each biological process) (biological process in each sample)

The algorithm looks for patterns that define the biological activity in a set of samples (e.g., a collection of tumors) and the amplitude at which each gene contributes to that biological process.

Fig. 3

HCA Training Data (Goff / Raj)

Feature quantification (Patro)

Image features (Raj)

Reproducible Workflows (Garmire / Greene / Patro / Piccolo)

Efficient factorization

Sensitivity to input features and variance

Denoising / Filtering (Krisnaswamy / Piccolo)

Deep learning

Sub/re-sampling evaluation

Trajectory / temporal patterns (Fertig)

Gene signatures and gene regulation (Greene / Krisnaswamy)

Model interpretation (Goff)

Visualization (Garmire / Krishnaswamy)
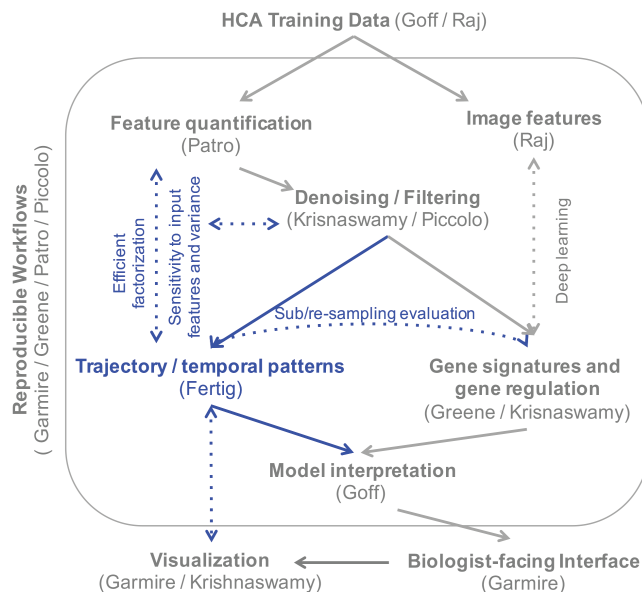
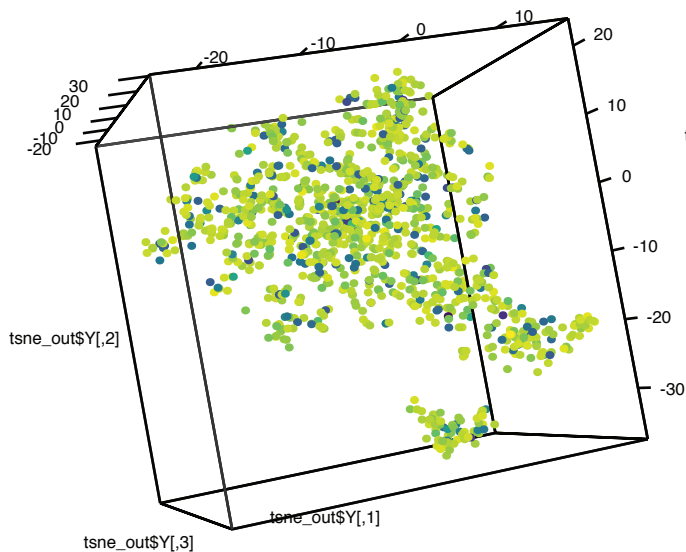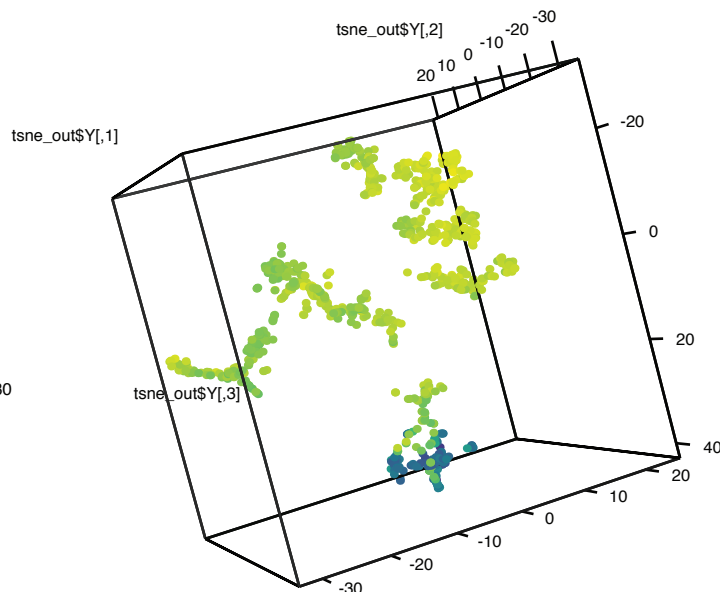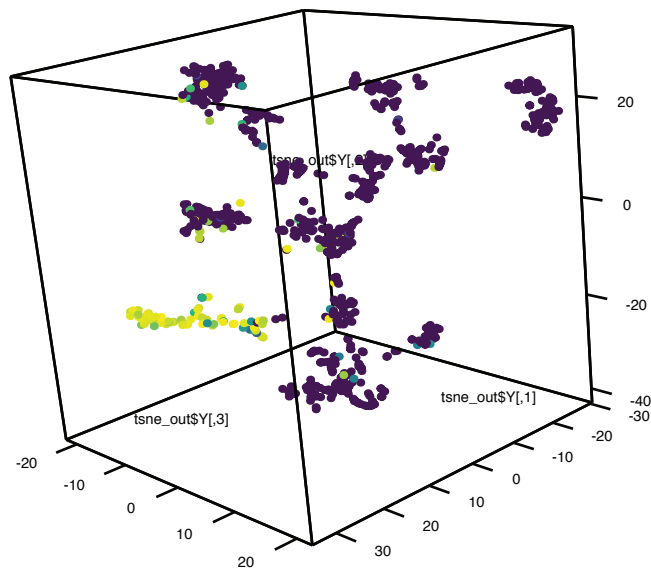Biologist-facing Interface (Garmire)

Fig. 2

A. T-SNE of raw scRNAseq data colored by # genes expressed

B. T-SNE of CoGAPS patterns for scRNAseq colored by # genes expressed

C. T-SNE of CoGAPS patterns colored by Crx expression
(Crx is a marker of photoreceptors)

D. T-SNE of CoGAPS patterns colored by Ccnd1 expression
(Ccnd1 is a marker of cell cycle)