

scOrange: Single-cell Data Mining for Everyone

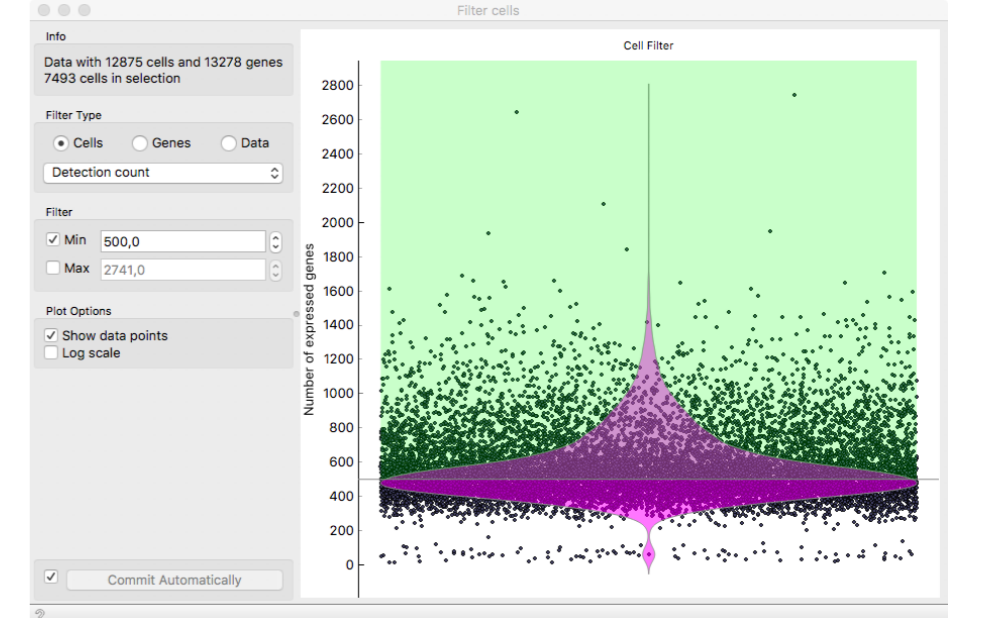
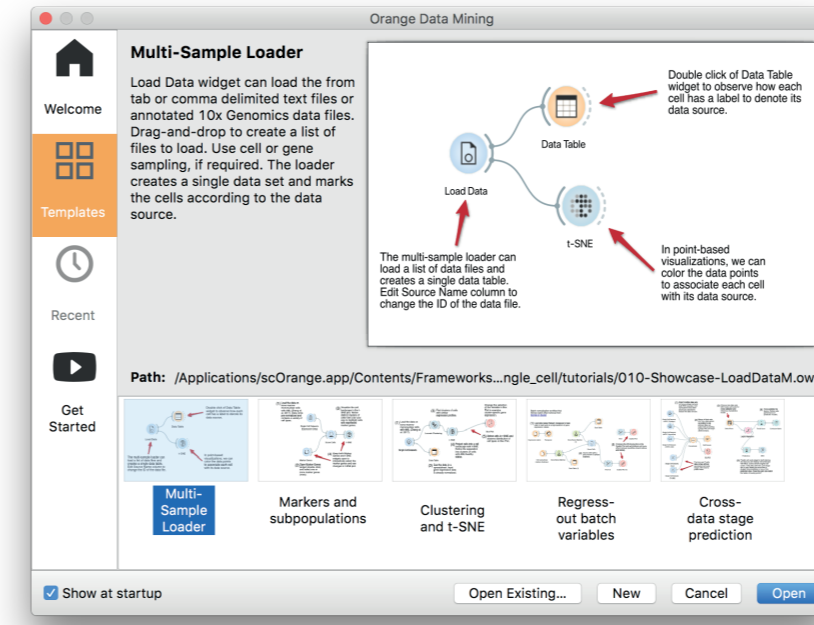
Martin Stražar,¹ Lan Žagar,¹ Jaka Kokošar,¹ Vesna Tanko,¹ Pavlin Poličar,¹ Aleš Erjavec,¹ Anže Starič,¹ Vilas Menon,² Rui Chen,³ Gad Shaulsky,³ Andrew Lemire,² Anup Parikh,⁴ Blaž Zupan^{1,3}

¹ University of Ljubljana, Ljubljana, Slovenia, ² Howard Hughes Medical Institute, Ashburn, U.S.A., ³ Baylor College of Medicine, Houston, U.S.A., ⁴ Naringi, San Francisco, U.S.A.



singlecell.biolab.si

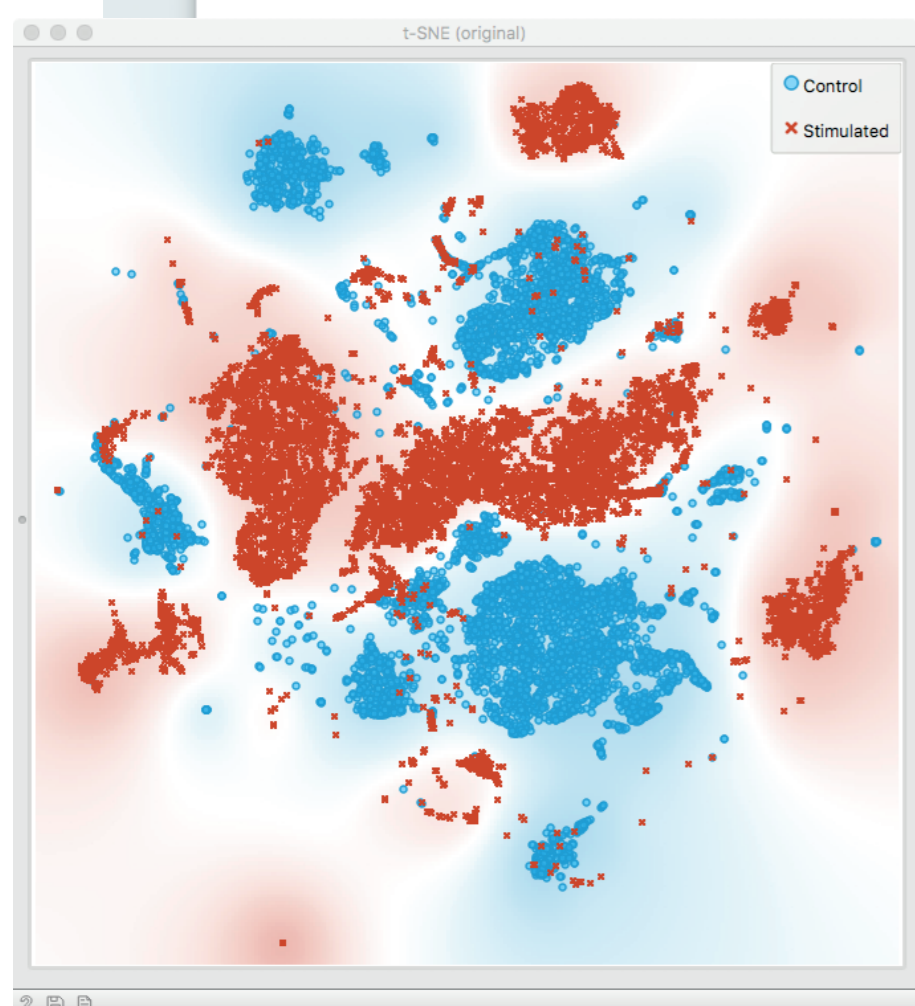
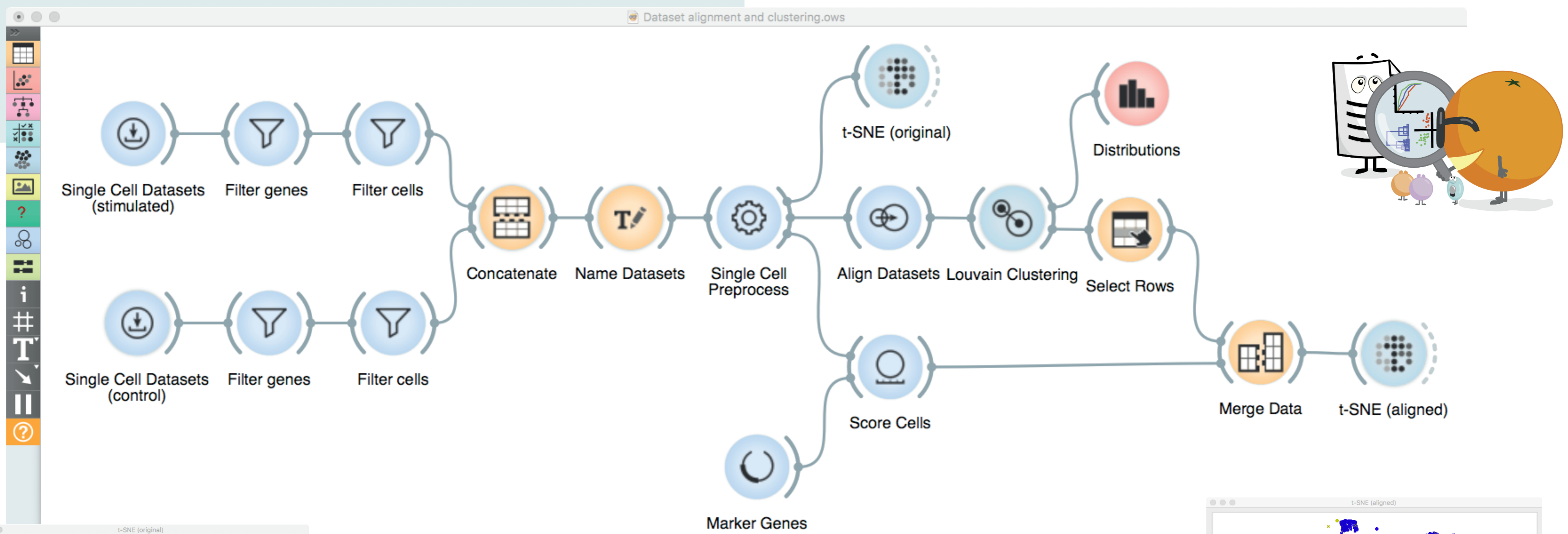
Single-cell RNA sequencing (scRNA-seq) assays pose exciting challenges for computational analysis, visualization and modeling^[1]. We present **scOrange**, enabling data analysis to a broader range of biomedical researchers. It uses visual programming to construct intuitive workflows that deliver fairness, interoperability, accessibility and resuability principles to the single-cell community.



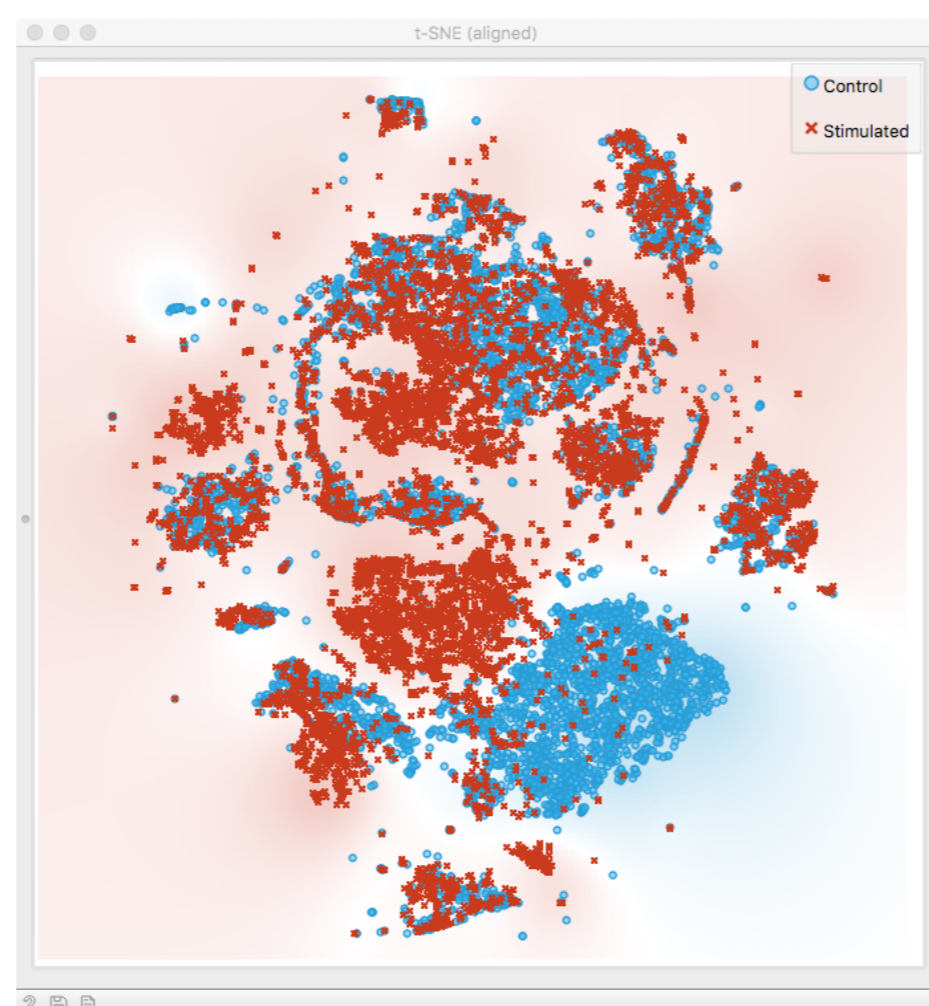
Built-in templates. scOrange comes with example datasets and predefined workflows, based on literature-supported methods and parameters.

Over 100 widgets. Together with Orange Data Mining^[3], scOrange widgets implement data manipulation, machine learning and visualization tasks.

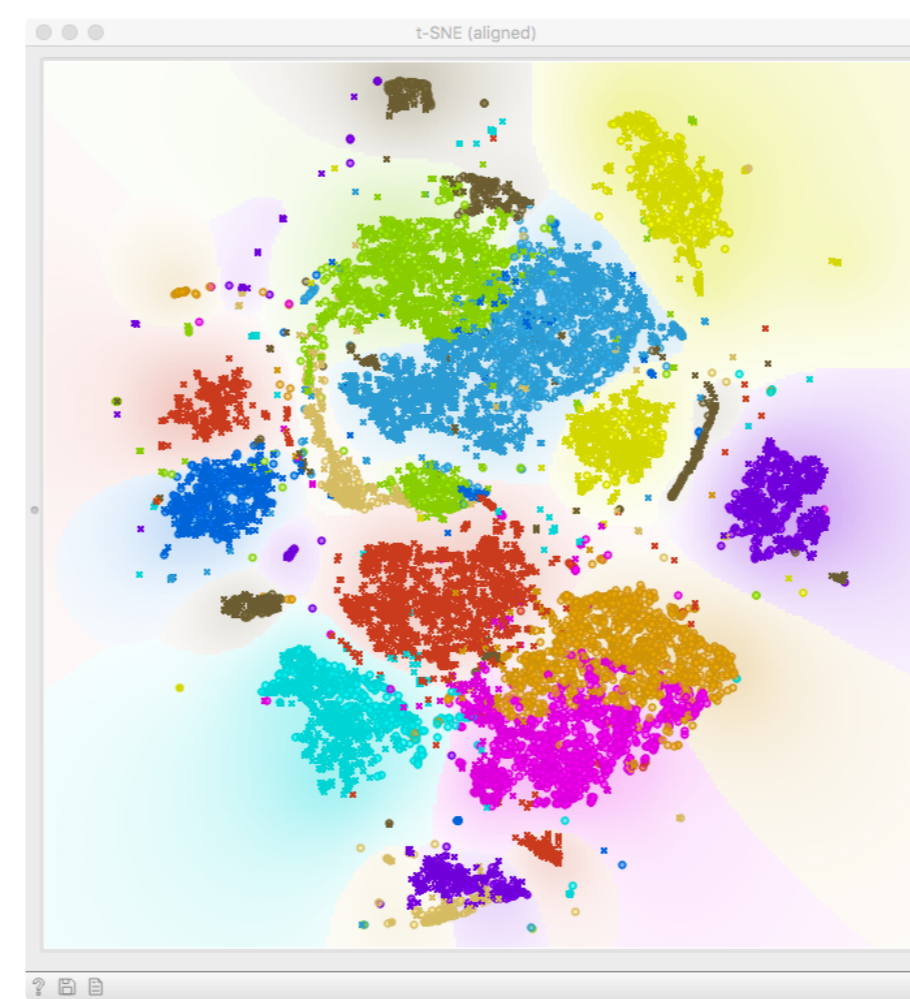
Full interactivity. Users can visually connect hundreds of interoperable **widgets** to construct reusable workflows, designed with emphasis on reproducibility. scOrange assumes no programming knowledge and is accessible to a broader audience than programming language-specific packages^[2].



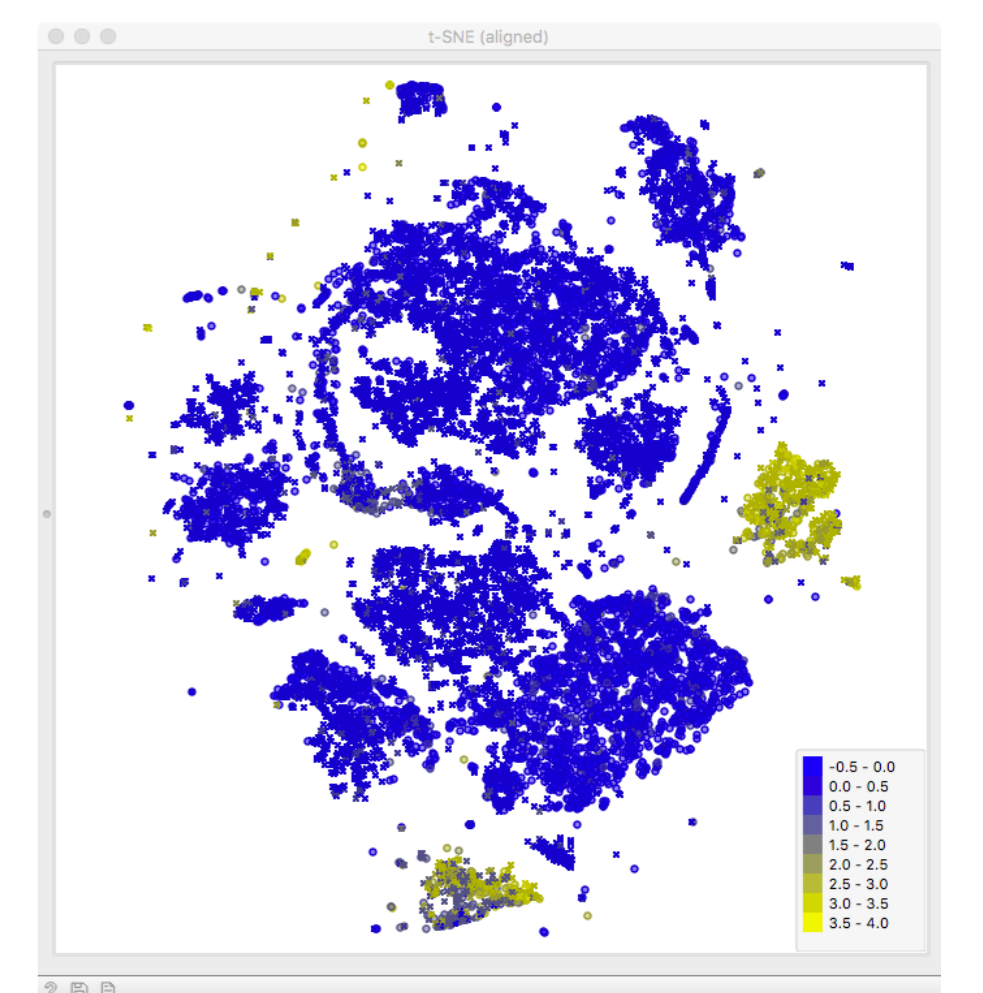
Projections. Run t-SNE, PCA, MDS to visualize cells in a 2D plane.



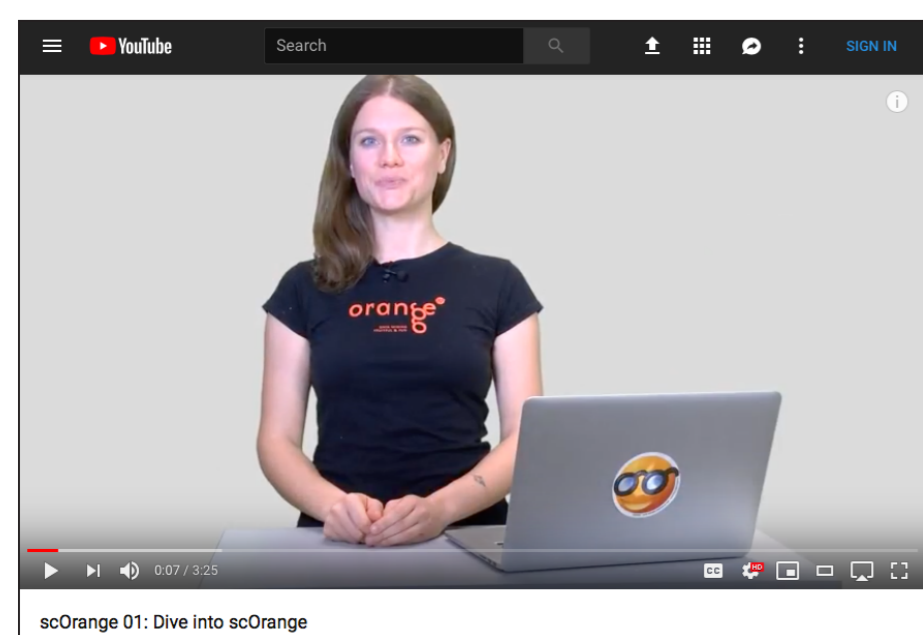
Batch effect removal. Use canonical correlation analysis to remove unwanted, dataset-specific variation^[4].



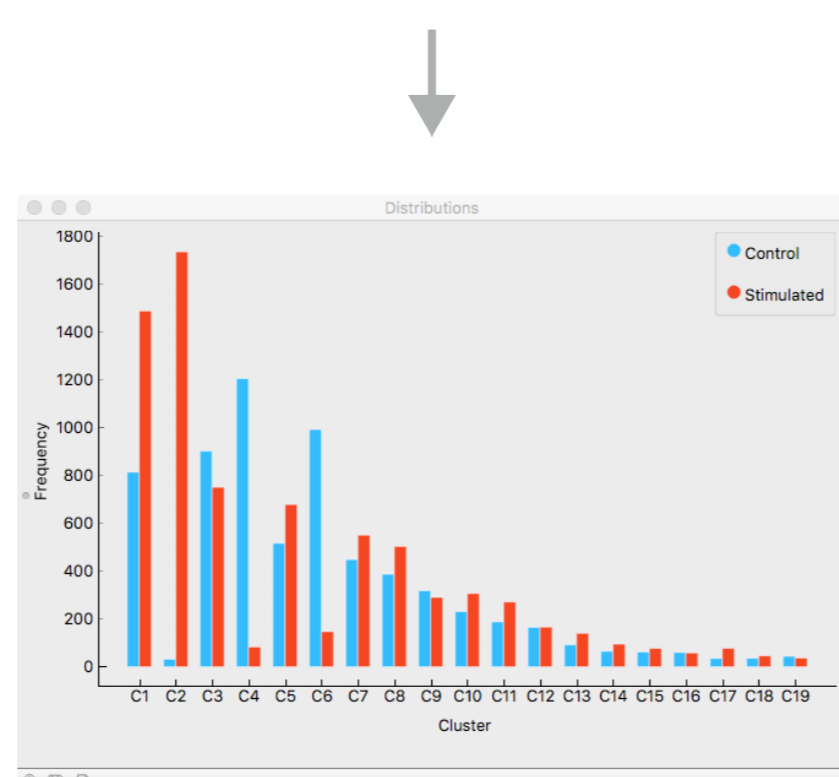
Clustering. Identify known cell populations or discover previously unknown subpopulations.



Marker genes scoring & discovery. Validate the clustering with known markers or use statistical tests to find new ones.



Active community. Video tutorials, blogs and docs enable a quick dive straight into action.



Moreover, characterize cell subpopulations through analysis of differentially expressed genes and gene set enrichment, find candidate marker genes, and predict cell types or cell cycle stage^[5].

Support for common data formats (text, 10X, Loom, Excel, ...). Fully-reproducible, standardized analyses without a single line of code. Custom extensions possible through open source.

Visit us for a live demo!