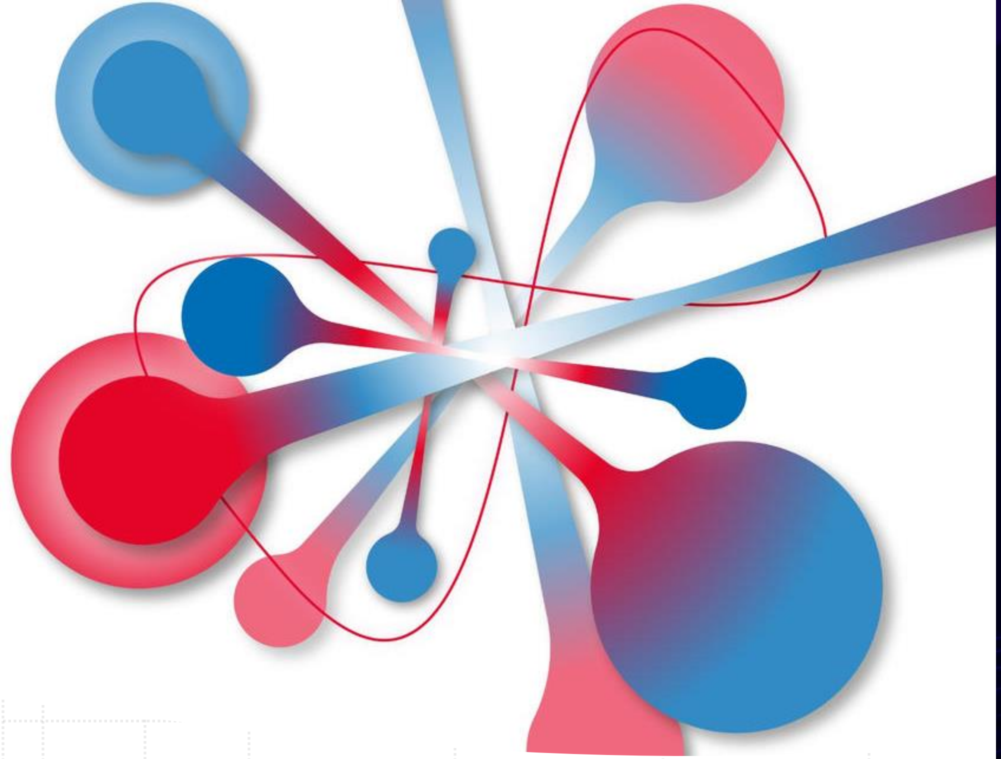

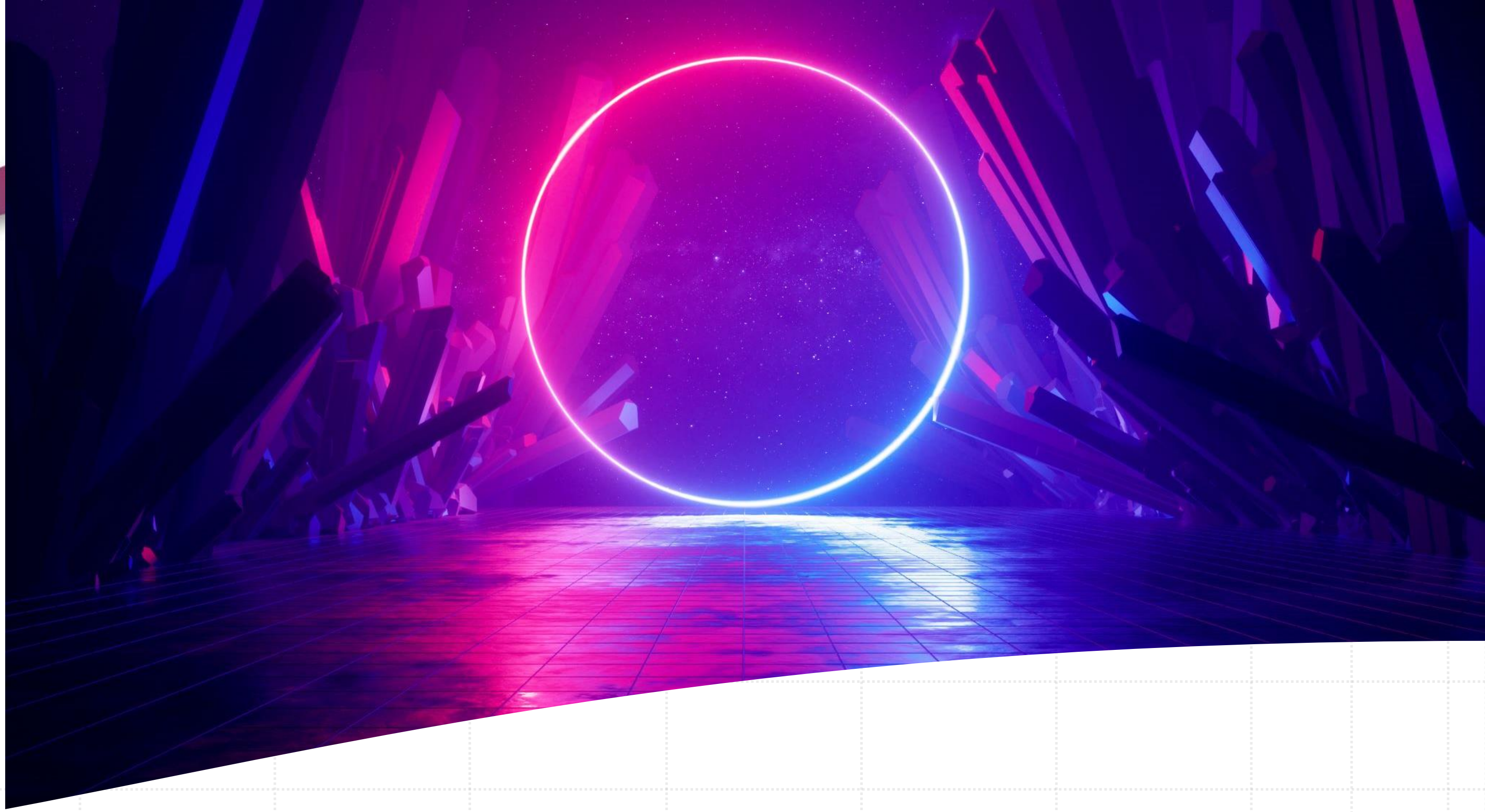

singleCellRNAseq

Juan Jovel

Bioinformatics Unit

Faculty Veterinary Medicine

University of Calgary

Introduction to Single-Cell RNA-seq

Overview of scRNA-seq: What it is, how it differs from bulk RNA-seq, key advantages.

Applications: Understanding cellular heterogeneity, cell atlas projects, disease studies, developmental biology.

Foundations of scRNA-seq

Key Considerations: Number of cells, sequencing depth, tissue preparation.

Challenges: Low RNA content per cell, technical noise, dropout events.

Experimental Design

General Workflow: Cell isolation, library preparation, sequencing, data analysis.

Cell Capture Methods: Microfluidics (e.g., 10X Genomics), FACS, Drop-seq.

Preprocessing of scRNA-seq Data

Initial Steps: FASTQ generation, quality control (QC), and (seudo-) alignment.

Tools: Cell Ranger (10X Genomics), Kallisto/BUStools, Salmon/Alevin, etc.

Quantification of scRNA-seq Libraries

UMI (Unique Molecular Identifiers): Role in counting transcripts.

Gene-Cell Matrices: Creation of the gene-cell count matrix.

Dealing with Dropouts: Technical artifacts and handling sparsity.

Introduction to Seurat and Scanpy

Seurat: Overview of the R-based toolkit.

Scanpy: Python-based counterpart for scRNA-seq.

Comparison: Strengths of each tool.

Quality Control and Filtering

QC Metrics: Mitochondrial content, number of detected genes per cell, total RNA per cell.

Filtering Low-Quality Cells: Threshold-based filtering in Seurat and Scanpy.

Normalization and Scaling

Log-Normalization: Adjusting for differences in sequencing depth.

Scaling Data: Z-score normalization.

Normalization with Seurat and with Scanpy

Feature Selection

Identifying Highly Variable Genes:
Why it's important, how it's done.

Feature selection in single-cell RNA-seq is crucial for reducing noise and focusing on the most informative genes, enabling more accurate identification of cell types and states.

Dimensionality Reduction

Principal Component Analysis (PCA): First step for reducing data complexity.

t-SNE and UMAP: Techniques for visualizing high-dimensional data.

Clustering Cells

Clustering Algorithms: Louvain and Leiden methods.

Choosing Resolution: Impact of resolution on cluster granularity.

Differential Expression Analysis

Identifying Marker Genes: Per-cluster analysis.

Comparing DE Methods: Wilcoxon Rank Sum, t-test, etc.

Annotation of Cell Types

Using Reference Datasets: Cell type identification using tools like SingleR, Azimuth.

Manual Annotation: Using marker gene databases and literature.

Integration of Multiple Datasets

Why Integration Matters: Correcting for batch effects, combining multiple datasets.

Seurat's Integration Workflow

Scanpy's Harmony Method:

Trajectory Analysis and Pseudotime

Overview of Trajectory Analysis: Understanding lineage progression and differentiation.

Tools: Monocle3, Slingshot.

Seurat/Scanpy Integration: Compatibility and workflows.

Advanced Topics in scRNA-seq

RNA Velocity: Predicting future states of cells using spliced and unspliced mRNA.

Seurat/Scanpy Integration: scVelo in Scanpy.

Conclusion and Future Directions

Emerging Trends: Multi-omics integration (CITE-seq, scATAC-seq), spatial transcriptomics.

Challenges and Opportunities: Scalability, data interpretation, computational demands.

Join us for this workshop…

- It will be a lot of fun!!!