Analyzing ILD BALF scRNA data

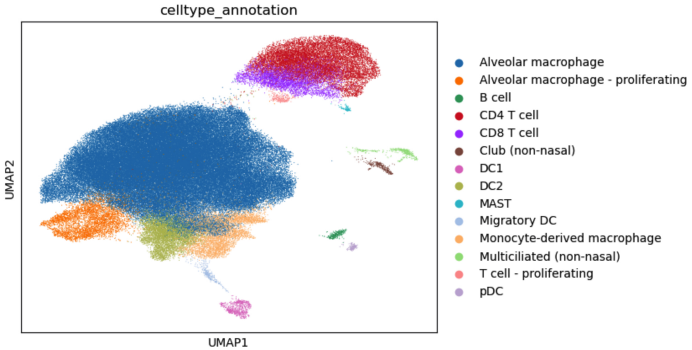
Dr Peter Allen
Research Officer, Translational Genomics Lab | Garvan Institute of Medical Research
Conjoint Lecturer, School of Biomedical Sciences | UNSW Medicine & Health
Research Fellow, Centre of Research Excellence in Pulmonary Fibrosis

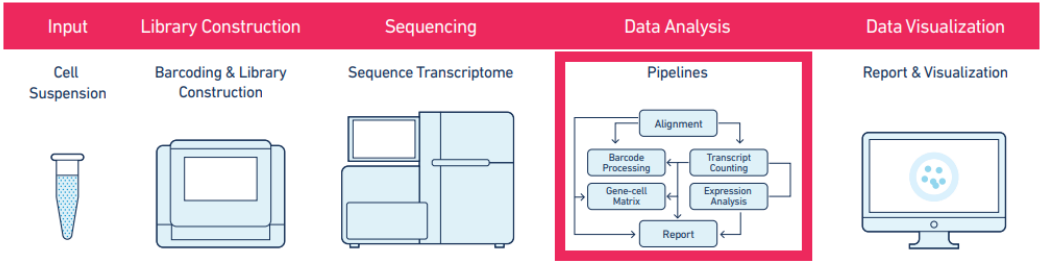
July 25, 2024



scRNA - Ability to Contrast RNA profiles between cell-types

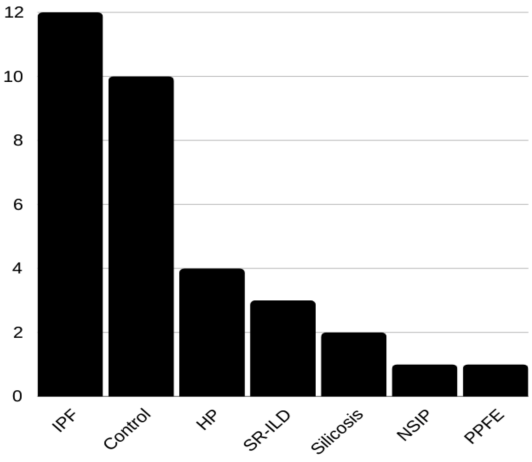
- Captures cellular heterogeneity often lost when performing bulk sequencing
- Additionally gives the power to detect sub-lineages within cell states not previously characterized




Input
Library Construction
Sequencing
Data Analysis
Data Visualization
Cell Suspension
Barcoding & Library Construction
Sequence Transcriptome
Pipelines
Report & Visualization
Alignment
Barcode Processing
Transcript Counting
Gene-cell Matrix
Expression Analysis
Report

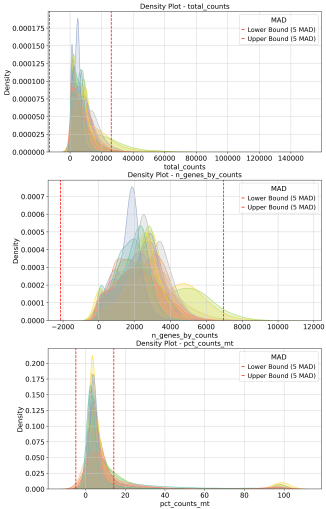
Study-specific Disease Demographics

- ILD scRNA data was generated for 33 BAL samples across 3 phases
- Includes 6 fibrosis-related ILDs



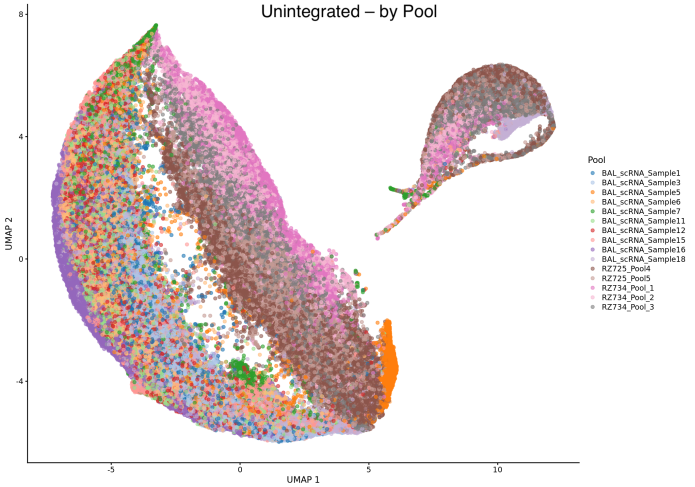
Cleaning and Pre-processing the data through Quality Control

- Assess data quality through metrics like total UMI counts per cell, how many genes within total counts and mitochondrial percentage
- Remove any ambient/background RNA
- Filter out low-quality cells and genes
- Check for batch effects



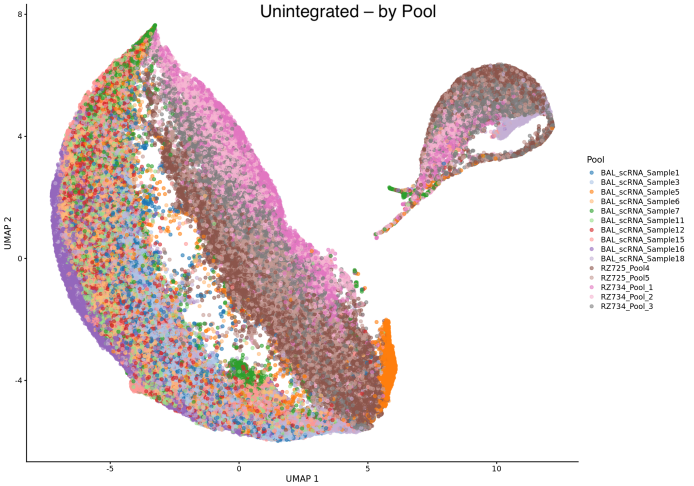
Cleaning and Pre-processing the data through Quality Control

- Assess data quality through metrics like total UMI counts per cell, how many genes within total counts and mitochondrial percentage
- Remove any ambient/background RNA
- Filter out low-quality cells and genes
- Check for batch effects



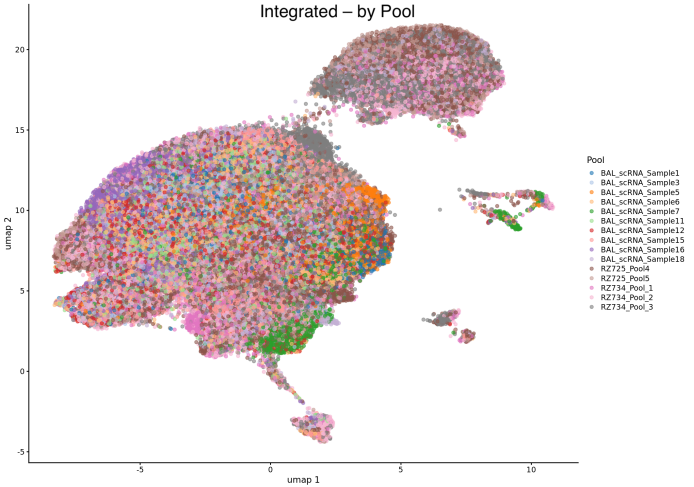
Integrating the data removes batch effects

- Integrating data from multiple samples allows for analyzing whether features are due to biology rather than an artefact of batch
- Ensures cells that are truly similar will be clustered and classified accurately
- Algorithms for integration:
  - scVI ()
  - Harmony or Seurat ()



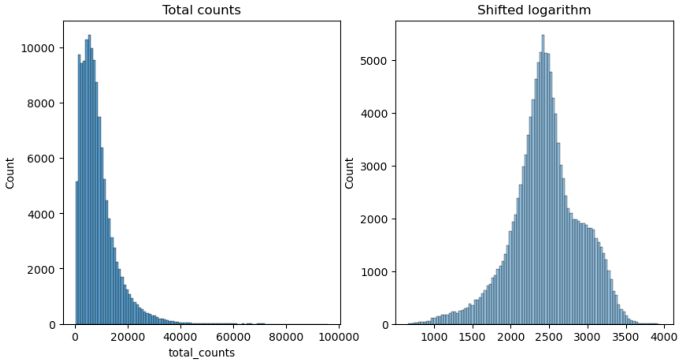
Integrating the data removes batch effects

- Integrating data from multiple samples allows for analyzing whether features are due to biology rather than an artefact of batch
- Ensures cells that are truly similar will be clustered and classified accurately
- Algorithms for integration:
  - scVI ()
  - Harmony or Seurat ()



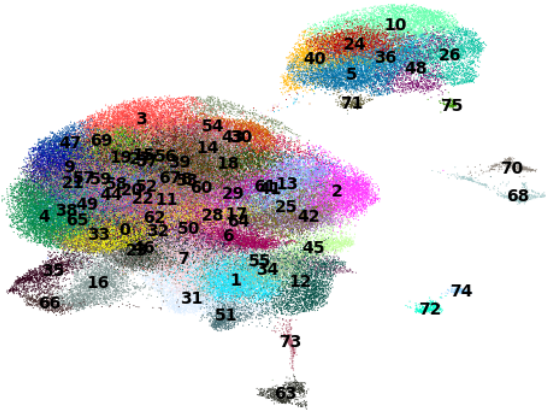
Normalizing Gene Expression

- Smooths differences due to technical variation
- Adjusts for cell-specific differences in RNA content for more robust gene expression comparison across cells



Clustering

- Group similar cells based on their gene expression profile
- Identify individual cell populations
- Clustering Methods:
  - Louvain
  - Leiden ⋆
  - K-means



Annotating the clusters

- Annotation of clusters can be performed in two ways:
  - Unsupervised Machine Learning using reference data
    - Azimuth (R) / Celltypist (Python)
    - Scimilarity (Python)
  - Manual Annotation via known cell markers



Annotating the clusters

- Annotation of clusters can be performed in two ways:
  - Unsupervised Machine Learning using reference data
    - Azimuth (R) / Celltypist (Python)
    - Scimilarity (Python)
  - Manual Annotation via known cell markers

