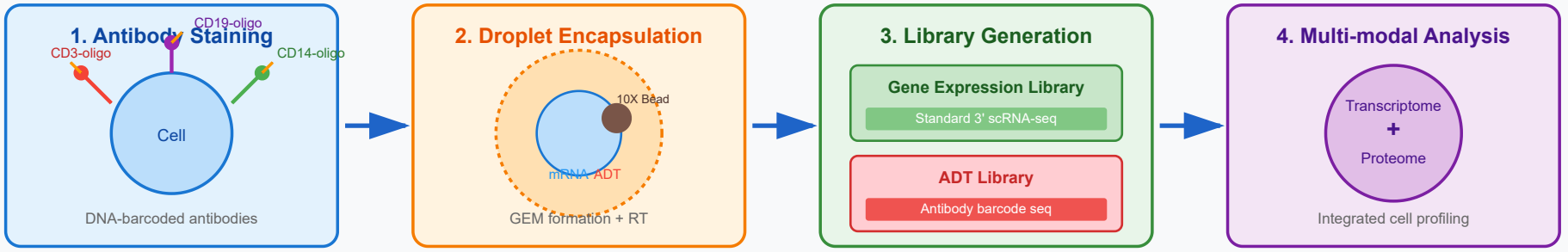
CITE-seq

Cellular Indexing of Transcriptomes and Epitopes by Sequencing

CITE-seq Workflow Overview



Key Benefits of CITE-seq

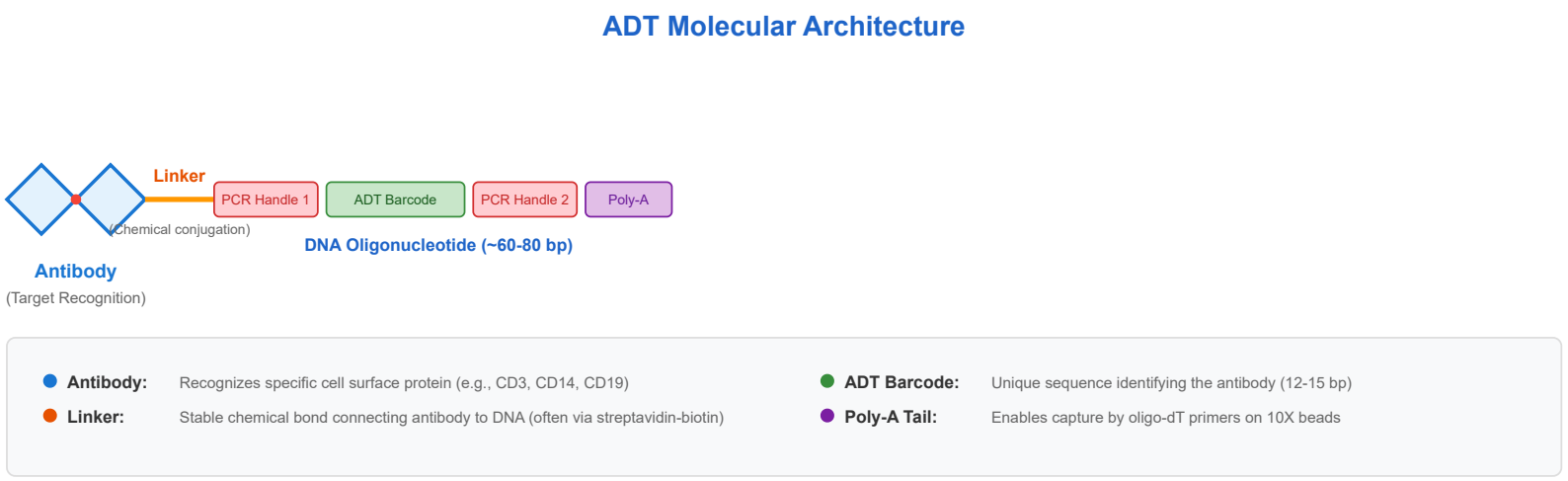
1. **Enhanced Cell Type Resolution**
   Protein markers provide clearer cell identity
2. **Validation of Gene Expression**
   Direct protein measurement validates RNA data
3. **Post-translational Information**
   Captures protein modifications not visible in RNA
4. **Low-abundance Transcripts**
   Proteins detect markers missed by RNA-seq
5. **Platform Compatibility**
   Works with standard 10X Genomics workflow
6. **Same Cell Barcode**
   RNA and protein data share cell identity

> 💡 CITE-seq bridges the gap between transcriptomics and proteomics at single-cell resolution

1. Antibody-Derived Tags (ADT) Technology

CITE-seq's core innovation lies in the use of oligonucleotide-conjugated antibodies, known as Antibody-Derived Tags (ADTs). These specialized reagents enable simultaneous measurement of surface protein expression alongside gene expression in individual cells.

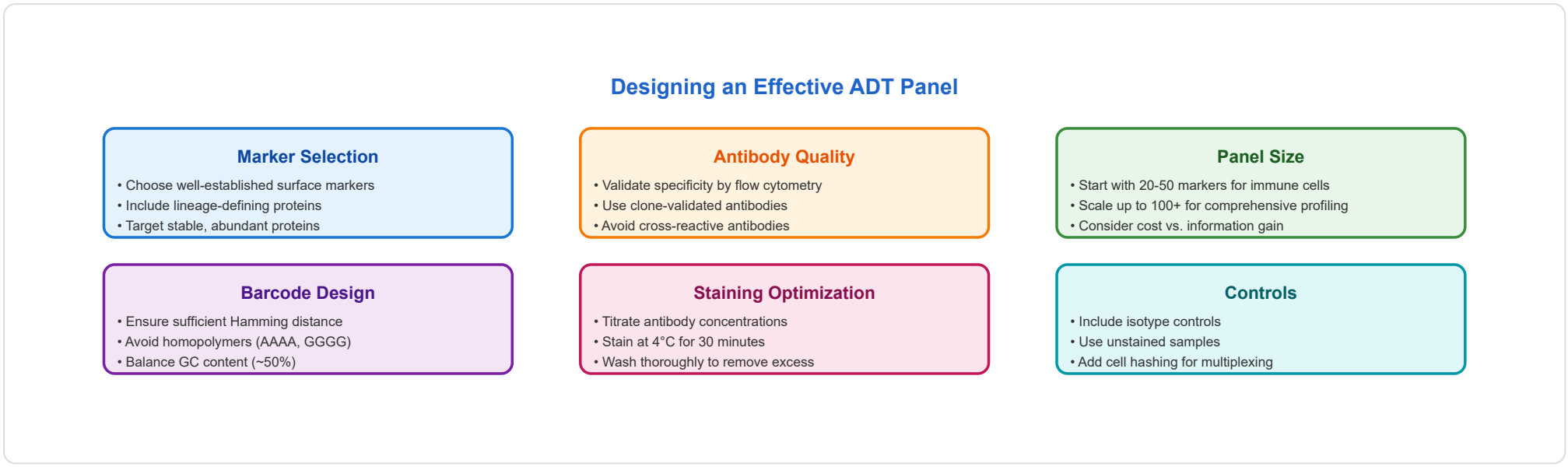
ADT Structure and Design



How ADTs Work

ADTs function by combining the specificity of antibody-antigen recognition with the quantitative power of DNA sequencing. When cells are stained with a cocktail of ADT-conjugated antibodies, each antibody binds to its cognate surface protein. During droplet encapsulation and reverse transcription, the poly-A tail of the ADT oligonucleotide is captured by the same oligo-dT primers that capture mRNA, ensuring that protein and RNA measurements share the same cell barcode.

ADT Panel Design Considerations



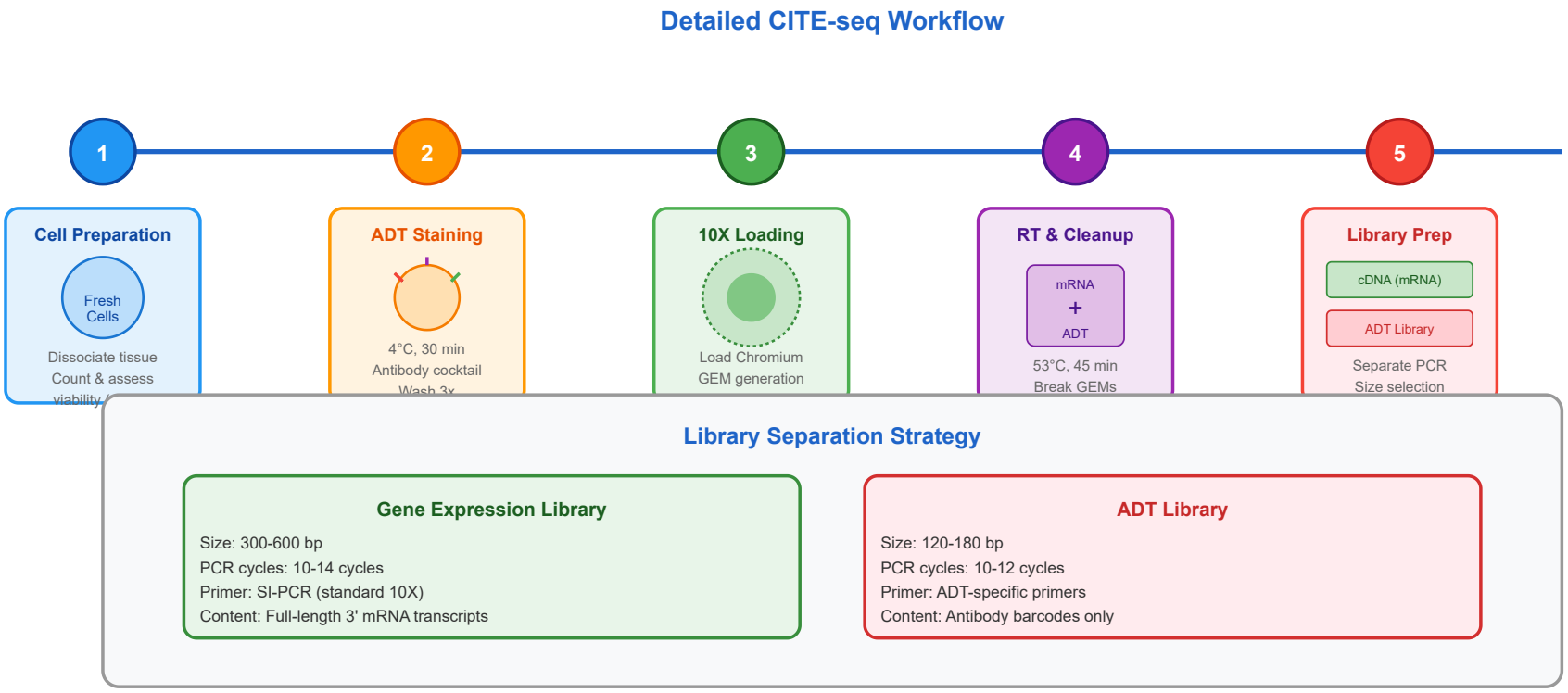
Key Advantages of ADT Technology

- **Direct Protein Measurement:** Captures actual protein expression, not just mRNA abundance
- **Higher Sensitivity:** Antibodies can detect low-abundance proteins that may not be captured in RNA data
- **Same Cell Resolution:** Both RNA and protein data are linked by the same cell barcode
- **Reduced Dropout:** Protein measurements are less affected by technical dropout compared to sparse scRNA-seq data
- **Functional States:** Captures activation markers and post-translational modifications

2. Technical Workflow and Implementation

CITE-seq integrates seamlessly with standard droplet-based single-cell RNA-seq platforms, particularly the 10X Genomics Chromium system. The workflow requires minimal modifications to existing protocols while adding a powerful protein measurement dimension.

Step-by-Step Protocol

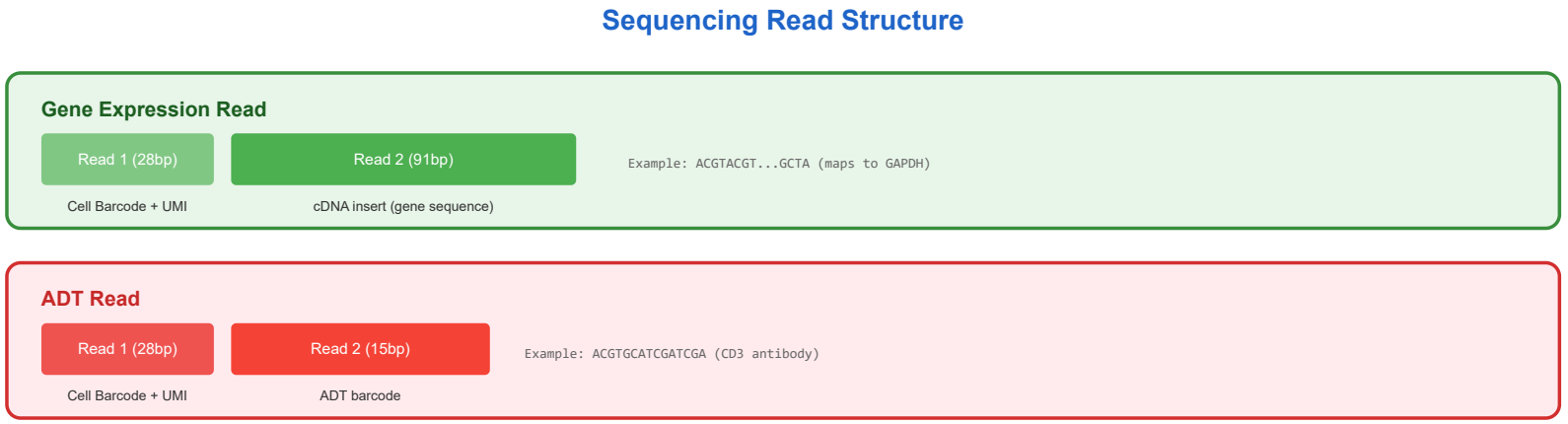


Critical Technical Considerations

- **ADT Library Optimization:** The ADT library is much smaller and simpler than the cDNA library, requiring careful titration to avoid over-amplification. Typically, ADT libraries are sequenced with 5-10% of total sequencing reads.

- **Size Selection:** Proper size selection is crucial to separate ADT libraries (120-180 bp) from cDNA libraries (300-600 bp). SPRI bead cleanup at 0.6x ratio effectively separates these populations.
- **Washing Stringency:** Thorough washing after antibody staining is essential to remove unbound ADTs, which can create high background signal. Three washes with PBS + 1% BSA are recommended.
- **Cell Concentration:** Maintain cell concentration at recommended levels (700-1,200 cells/µL) to ensure proper encapsulation efficiency and minimize doublets.

Sequencing and Data Output



Sequencing Depth Recommendations

Gene Expression Library

Recommended depth: 20,000-50,000 reads per cell

ADT Library

Recommended depth: 1,000-5,000 reads per cell

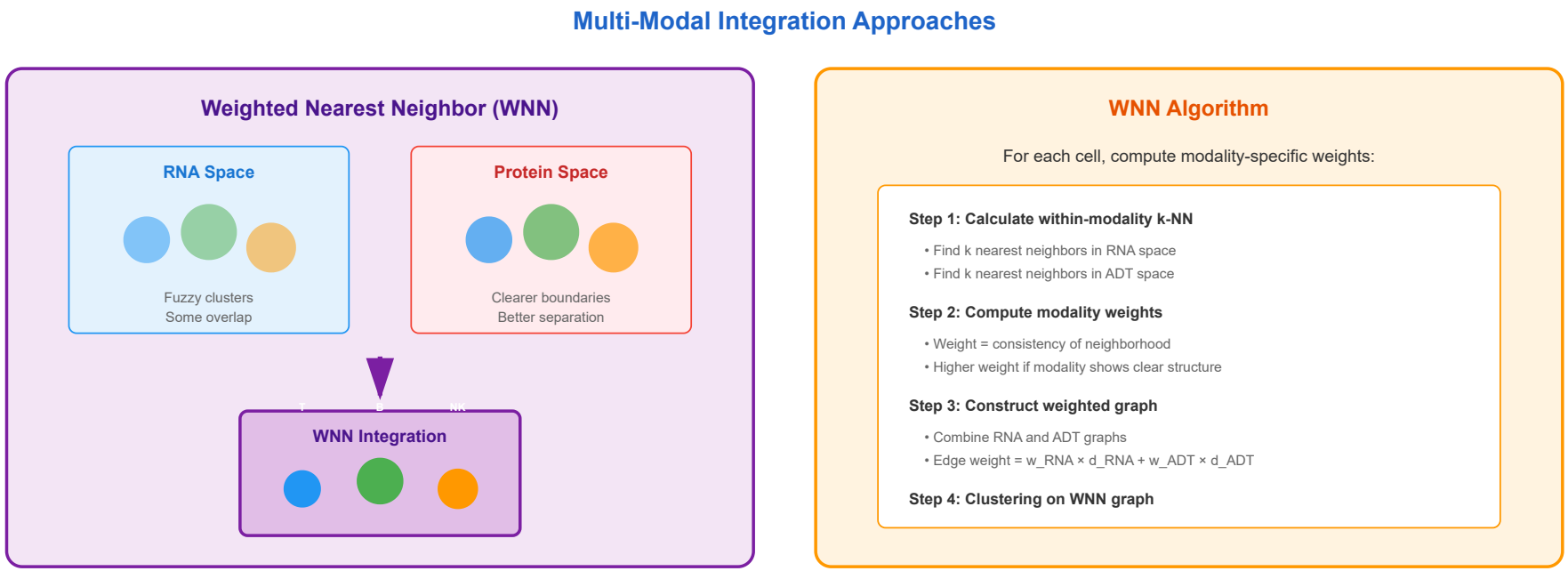
Purpose: Capture transcriptome diversity
Typical output: 2,000-5,000 genes per cell

Purpose: Quantify protein markers
Typical output: 20-200 proteins per cell

3. Applications and Multi-Modal Data Integration

CITE-seq enables sophisticated multi-modal analysis by integrating transcriptomic and proteomic data from the same single cells. This integration reveals biological insights that would be impossible to obtain from either data type alone.

Data Integration Methods



Key Applications

CITE-seq Applications Across Biology

1 Immune Cell Profiling

- Resolve fine T cell subsets (CD4+, CD8+, Tregs, exhausted T cells)
- Characterize activation states (CD25, CD69, HLA-DR)
- Identify rare populations (double-negative T cells, γδ T cells)
- Study immune responses in vaccination, infection, and cancer

Example: COVID-19 immune profiling, CAR-T cell monitoring

2 Tumor Microenvironment

- Map tumor-infiltrating lymphocytes (TILs)
- Identify immunosuppressive cells (MDSCs, Tregs, TAMs)
- Measure checkpoint molecules (PD-1, PD-L1, CTLA-4, TIM-3)
- Correlate immune infiltration with tumor genetics

Example: Predicting immunotherapy response, resistance mechanisms

3 Developmental Biology

- Track cell lineage markers during differentiation
- Validate developmental trajectories with surface markers
- Study stem cell heterogeneity (HSCs, MSCs, iPSCs)
- Measure differentiation stage markers (CD34, CD38, CD117)

Example: Hematopoiesis mapping, organoid development

4 Autoimmune & Inflammatory Disease

- Profile immune dysregulation in disease vs. healthy tissue
- Identify pathogenic cell states
- Track inflammatory markers (TNF-α, IL-6, IFN-γ receptors)
- Monitor treatment response over time

Example: Rheumatoid arthritis, inflammatory bowel disease, lupus

5 Neuroscience

- Classify neuronal subtypes using surface receptors
- Study neurotransmitter receptor expression
- Profile microglia activation states
- Investigate neuroinflammation in disease

Example: Alzheimer's disease, neurodegeneration, brain development

6 Sample Multiplexing (Cell Hashing)

- Pool multiple samples in one 10X run
- Use hashtag oligos (HTOs) for sample identification
- Reduce batch effects and costs
- Detect and remove doublets computationally

Example: Patient cohorts, treatment time courses, genetic perturbations

Advantages Over RNA-seq Alone

- **Resolving Ambiguous Clusters:** Protein markers can distinguish cell types that appear similar at the transcriptome level. For example, CD4+ and CD8+ T cells may cluster together based on RNA but are cleanly separated by CD4/CD8 protein expression.

- **Validation of Annotations:** Surface protein measurements provide independent validation of RNA-based cell type assignments, increasing confidence in biological interpretations.
- **Functional Markers:** Many activation markers (PD-1, Ki-67, phospho-proteins) are post-translationally regulated and not reliably detected at the RNA level.
- **Low-Dropout Measurements:** Protein counts are less sparse than RNA counts, providing more robust quantification of key markers that may be subject to dropout in scRNA-seq.

Computational Tools for CITE-seq Analysis

- **Seurat (R):** Integrated workflow with WNN analysis, multi-modal clustering, and visualization. Most widely used for CITE-seq data.
- **MUON (Python):** Multi-modal omics analysis framework with support for CITE-seq integration.
- **CiteFuse (R):** Specialized package for CITE-seq analysis including doublet detection and data normalization.
- **totalVI (Python/scvi-tools):** Probabilistic model for joint analysis of RNA and protein data with batch correction.
- **Azimuth:** Reference-based cell type annotation using CITE-seq reference atlases.

Future Directions and Extensions

CITE-seq technology continues to evolve with new applications and extensions including REAP-seq (RNA expression and protein sequencing), ASAP-seq (ATAC-seq with protein), TEA-seq (Transcriptome, Epitope, and Accessibility sequencing), and DOGMA-seq (DNA, Oligo, Gene expression, Methylation, and Accessibility sequencing). These multi-omic approaches promise even deeper insights into cellular identity and function by layering additional modalities onto single-cell measurements.

🎯 CITE-seq represents a paradigm shift in single-cell biology by enabling simultaneous measurement of transcriptome and proteome, providing unprecedented resolution for cell type identification, functional state characterization, and disease profiling.