Overview of Single Cell Omics

CSE 590C Winter 2021
Jan 11, 2020

Ayse B. Dincer and Walter L. Ruzzo
Overview of Single Cell Omics

Single-Cell Genomics

Carmela Paolillo,¹ Eric Londin,² and Paolo Fortina³,4✳
1. Introduction: Next-generation sequencing

Next-generation sequencing (NGS) to study genomes and transcriptomes

Image from www.ebi.ac.uk, Functional genomics I Course
1. Introduction: Limitations of bulk approaches

• "Bulk" approaches
  • Combining thousands of cells from a single tissue
  • Treating them as a single unit
  • Ignoring the role of each individual cell

• We now appreciate the heterogeneous nature of various tissues
  • different cell populations may have vastly different transcriptomes and different contributions to cellular processes
1. Introduction: Single-Cell Omics

Development of advanced methods to capture single cells and amplify genome-wide DNA and RNA -> Single-Cell Genomics (SCG)

SCG aims to provide a new understanding of genetics and transcriptomics at the single-cell level

Fig. 1. Applications of SCG in cancer. Cells, CTCs, primary tumor cells, and cell of origin all can be subjected to a variety of analyses involving SCG.
1. Introduction: History of single-cell genomics
2.1 SCG Technology: Isolation of Single Cells

1\textsuperscript{st} step = isolation of an intact single cell

**Micromanipulation**
- Microscope
- Capillary pipette

**Flow cytometry**
- FACS
- Laser
- Multispectral detector + Electronics

**Microdissection**
- LCM
- Cell

2.1 SCG Technology: Isolation of Single Cells

**Microfluidics**

Microparticle and lysis buffer → Oil

Cells from suspension

Droplet with single cell → Oil

---

**Chromium Single Cell 3’ Solution (10x Genomics)**

Barcoded gel beads → Cells → Oil → Single-cell GEMs → Collection → Retro-transcription → Break emulsion → Library generation → Sequencing → Data analysis and report visualization

---

2.1 SCG Technology: Isolation of Single Cells

Blood collection

Anti-EpCAM antibody with magnetic particle

CTC enrichment

Clinical applications of SCG involves the isolation of single cells from body fluids

CellSearch semiautomated cell-isolation system
Enables analysis of circulating tumor cells (CTCs)

2.2 SCG Technology: Amplification

1st step = isolation of an intact single cell

For sequencing we need enough genetic material to be analyzed

So, amplification is a crucial step

2nd step = amplification (WTA or WGA)

Amplification remains a major challenge for SCG

Goal = minimize artifacts without loss in sensitivity and specificity
2.2 SCG Technology: Whole-Transcriptome Amplification (WTA)

1. Full-length methods
   - Full representation of poly A-containing transcripts
   - Examples:
     - SMART-seqQuartz-Seq

2. Tag-based methods
   - Enable sequencing of either the 5’ or 3’ end
   - Allow strand specificity
   - CEL-seq is one of the most common

More recent protocols:
Label cells by using molecular barcodes

Examples:
- CytoSeq (Illumina) (oligonucleotide beads)
- inDrop and Drop-seq (Illumina) (droplets using a microfluidic device)
- SPLiT-seq (does not require encapsulating a single cell in droplets or microwells)
2.2 SCG Technology: Whole-Transcriptome Amplification (WTA)

Drop-Seq pipeline
2.2 SCG Technology: Whole-Transcriptome Amplification (WTA)

**SPLiT-seq pipeline**

1. **Split** (Reverse transcription)
   - Cells from sample
   - Cells from each well are pooled back together.

2. **Pool**

3. **Split** (Ligation)
   - Cells are then distributed into 96 wells, and an in-cell ligation reaction appends a second well-specific barcode to the cDNA.

4. **Sequence**
   - After sequencing, each single cell transcriptome is assembled by combining reads containing the same four-barcode combination.
   - Barcodes: Genes 1 2 3 4
   - Cell 1
   - Cell 2
   - Cell 3

Split Biosciences
### 2.2 SCG Technology: Whole-Genome Amplification (WGA)

<table>
<thead>
<tr>
<th>1. PCR amplification</th>
<th>2. Isothermal Amplification (MDA)</th>
<th>3. Hybrid Methods</th>
</tr>
</thead>
</table>
| • First WGA method developed for SC  
  • Degenerate oligonucleotide primers (DOP-PCR) are used | • Most widely used protocol for WGA  
  • Has lower error rate  
  • Better genome coverage and reduced false-positive rates | • Combines advantages of the first 2 and reduces associated biases  
Examples:  
• PicoPLEX  
• MALBAC |

**Disadvantages:**
- Amplification bias
- Variability in efficiency
- High error rate owing to the thermostable polymerase
- Reduced uniformity and a high rate of allelic dropout
2.2 SCG Technology: WTA and WGA

Particular approach used should be based on

- specific application
- relative scarcity or abundance of the single cell to be isolated
- ease of DNA and/or RNA isolation from the target cell
- required depth of sequencing coverage
3. Applications of SCG: Epigenomics

Epigenomics = genome-wide studies of

- DNA accessibility
- higher-order chromosome organization
- DNA/protein modifications

Epigenetic marks on chromatin across the genome -> changes in gene expression

https://www.whatisepigenetics.com/fundamentals/
How can SCG help?

Information about chromatin modifications and their regulatory effects can be analyzed at single-cell resolution.
3. Applications of SGC: Cancer Genomics

Cancer = high inter-tumor and intratumor variability
The ability to molecularly phenotype every clone or population within a tumor -> key to treatment

How can SCG help?
- Monitoring disease progression and predict treatment resistance
- Isolation and characterization of CTCs
- Tumor heterogeneity, tumor microenvironment, and clonal evolution
- Identification of cancer cells of origin, rare cell subpopulations, tumor–tissue hierarchies
3. Applications of SGC: Cancer Genomics

Drug resistant clone identification

Tumor cell heterogeneity

Treat EGFR-TKI

(a) Minor resistant subpopulation

(b) Drug tolerant states

(c) Microenvironment

Resistant cells

Sensitive cells

Dying cells

Fibroblast

Hwang et al., Single-cell RNA sequencing technologies and bioinformatics pipelines
3. Applications of SGC: Immunology

**Immune system** = high degree of genetic diversity representing multiple different cell types

**How can SCG help?**
- Cellular complexity of the immune system
- Characterize/classify the different cell types
- Immunotherapy
- Neurodegenerative disease
3. Applications of SGC

Microbiology
• Evaluating microbiome biodiversity

Prenatal Screening
• Detecting clinically significant genetic alterations with noninvasive prenatal diagnostic procedures

Cell Lineage and Differential Markers
• Inferring lineage information from the early developmental stage
• Identifying novel differential markers
4. Bioinformatics for SGC

Challenges in organizing, archiving, and mining the data

Sparsity (dropouts) and high rates of technical artifacts (low-quality sequence data or batch effects)

Hwang et al., Single-cell RNA sequencing technologies and bioinformatics pipelines
4. Bioinformatics for SGC

Basic approach for analyzing SCG data:

1. Quality control and mapping of the sequence reads to the genome (as done with bulk sequencing)
2. Counting the number of genes
3. Normalization of the molecular counts
4. Clustering cells by the expressed genes
4. Bioinformatics for SGC: Example from scRNA-seq

4. Bioinformatics for SGC: Example from scRNA-seq

Problem of quantification of expression in scRNA-seq

\[
\text{RPK} = 10^6 \cdot \frac{n_i}{l_i}
\]

\[
\text{RPKM} = 10^6 \cdot \frac{n_i}{l_i \cdot \sum n_i}
\]

\[
\text{TPM} = 10^6 \cdot \frac{\text{RPK}_i}{\sum \text{RPK}_i} = 10^6 \cdot \frac{n_i / l_i}{\sum n_i / l_i}
\]

4. Bioinformatics for SGC: Example from scRNA-seq

REMINDER

4. Bioinformatics for SGC: Example from scRNA-seq

Discussion and Conclusions

• Use of single-cell genomics technology has grown rapidly
  • Major advances in the SCG methodology
  • Applications to diverse fields of biological research

• SCG allows:
  • Analysis of single-cell variants in both normal and disease states
  • Identification of new biomarkers for disease diagnosis, staging of disease progression, and prediction of optimal treatment regimen

• Has potential in the future to generate a much more comprehensive understanding of hereditary and somatic genetics
Discussion and Conclusions

What is next?

• **Standardized protocols** for reproducibility of studies and comparison of data sets generated by different groups

• **Improved sequencing technologies and better amplification methods** that minimize errors

• Development of improved methods **coupling single-cell and bulk NGS data**

• **SCG coupled with other multiomics technologies**
Discussion and Conclusions

Limitations:

• Various artifacts caused by suboptimal sample sizes, nonrepresentative starting samples, and other issues
• Difficult to interpret increasingly larger data sets

Discussion Questions:

What other limitations exist?
What do you think is the next step?
How do we select between bulk vs single-cell?
Can we combine bulk and single-cell?
Can we combine different omics data?