Spatial transcriptomics algorithms and the trend of spatial omics

Yuzhou Chang, Ph. D. candidate
Department of Biomedical Informatics
PIIO’s Immuno-Oncology Informatics Group (IOIG), OSUCCC
The Ohio State University
10/18/2022
Background: what is spatially resolved transcriptomics?

Spatially resolved transcriptomics:
Quantifying transcripts while keeping spatial context of samples within tissue or cell.

Nature Methods:
Method of the Year 2020: spatially resolved transcriptomics

Spatially resolved transcriptomics is our Method of the Year 2020, for its ability to provide valuable insights into the biology of cells and tissues while retaining information about spatial context.

About spatial transcriptomics

- Trend of spatial transcriptomics is increasing.
- The number is still going up.
- Gold era for spatial transcriptomics.
Spatial transcriptomics example: Visium

Sample Prep → Imaging → Barcoding & Library Construction → Sequencing → Data Visualization

A coronal mouse brain section
H&E stained

Hpca expression
In hippocampus

https://www.10xgenomics.com/spatial-transcriptomics/
Spatial transcriptomics technologies (before Jan 2022)

Techniques
- HDST
- Slide-seq/Slide-seqV2
- *Visium
- *DSP
- *ST

Ex-situ technologies

In-situ technologies
- seqFISH+/FISSEQ/ISS

Related questions
- Spatial variable gene
- Tissue architecture
- Cell-cell communication

Tools
- Giotto
- Trendsceek
- BayesSpace
- SPOTLight
- STARCH
- STfusion
- MERINGUE
- SpatialDE
- SpaGCN
- Stereoscope
- Cell2location
- SAPRK-X
- STATEGATE
- Tangram

Total detected cells
- In- & Ex-
- Ex-
- Ex-
- Ex-
- Ex-

In- & Ex-
- Giotto
- Trendsceek

Ex-
- BayesSpace
- SPOTLight
- STARCH

Ex-
- STfusion

Ex-
- STfusion

Approximately 5000 genes
≈ 5000 genes
1 cell
1 cell

In situ technologies
- In situ technologies
- Ex situ technologies

In situ technologies
- In situ technologies
- Ex situ technologies

Copy number inference
Fusion genes

Jie Liao, et al., Trends in Biotechnology (2020)
Rania Bassiouni, et al., Molecular Cell (2021)
Golden age for spatial transcriptomics

Method of the Year
2020: spatially resolved transcriptomics

Assess spatial heterogeneity and tissue architecture

Characterize cell-cell communication events in a specific region

Towards spatial omics technologies

Unique questions to be answered
RESEPT: REconstructing and Segmenting Expression mapped RGB images based on sPatially resolved Transcriptomics

1. Input gene expression and spatial location.
2. 3D embedding.
3. Embeddings convert to RGB image.
4. Pseudo-color image segmentation, using 16 human brain datasets which include 14 healthy and 2 Alzheimer’s disease (AD) datasets.

Yuzhou Chang, Computational and Structural Biotechnology Journal, (2022)
16 Datasets used in RESEPT training and testing

- 16 Visium data were used for model training and testing (from CT1 to 151508).
- The data includes health (14) and Alzheimer’s disease sample (2).
- G1 was used for case study.
- CT2 and 151674 were selected to simulate different read depth for stability test.
RESEPT outperformed other computational tools

Performance on 16 real datasets by a fix cluster number
  • Outperform other tools

Performance on simulation datasets (grid sequencing depth)
  • Stable and high performance
Conclusion: RESEPT could confidently reflect layer-specific, cell-type-specific, and pathological region-specific architecture regarding well-studied marker genes, which indicated significant potentials to localize and present important spatial architecture contributing to AD development.

Wei-Ting Chen, Cell (2020)
Glioblastoma case demonstrates RESEPT can be used on cancer tissue

Conclusion:
1. Identify tumor, non-tumor, and infiltrating tumor region.
2. Validate the three regions by pathological features.
3. Validate the three regions by transcriptional features.

Dr. Jose Otero
Dr. Shaoli Sun
Conclusion:
• RESPT is a deep learning framework for tissue heterogeneity visualization and architecture identification.
• The core concept of converting three-dimensional representations to RGB images and being associated with spatially variable genes will potentially enable explainable AI.
• RGB image can associated with certain spatially variable genes which can support main architecture of each RGB channels.
• It can generalize to other tissues (e.g., cancer)
• We apply on Alzheimer's disease and glioblastoma to visualize and reveal pathological region.

Output:
• Tissue architecture identification.
• Distinct cellular sub-populations (cell uncertainty measurement)
Characterize cell-cell communication (CCC) events in a specific region

### Cell-cell communication categories

<table>
<thead>
<tr>
<th>CCC</th>
<th>Intercellular communication</th>
<th>Intracellular communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categories</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mediator</td>
<td>Soluble factors</td>
<td>Surface protein</td>
</tr>
<tr>
<td>Types</td>
<td>Paracrine (P)</td>
<td>Cell contact (CC)</td>
</tr>
</tbody>
</table>

### Cell-cell communication within or across tissue architectures

<table>
<thead>
<tr>
<th>Tissue architecture &amp; CCC</th>
<th><img src="image1" alt="Diagram" /></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image2" alt="Diagram" /></td>
</tr>
</tbody>
</table>

---

**Intercellular communication**

- **Paracrine (P)**
- **Cell contact (CC)**
- **Autocrine (A)**

**Intracellular communication**

- **Soluble factors**
- **Surface protein**
- **Soluble factors**
Hypothesis and observation

**Hypothesis 1:**
S-CEM, R-CEM, and A-CEM will form two patterns in SRT data: paired-CEM (P-CEM) and mixed-CEM (M-CEM).

**Hypothesis 2:**

**Purpose:**

Identify a group of spots that subject to the following conditions:

1. They are M-CEM and P-CEM
2. These spots are spatially clustered
**Hypothesis and observation**

**Conclusion:**

1. M-CEM (mixed co-expression gene modules) capture CD48-CD2 ligand receptor pairs, associating T cell and B cell activation gene signatures.

2. S(sender)-CEM/R(receiver)-CEM capture motility-related LRP coding genes (CXCL12-CXCR4), which were reported to associate with tumor suppression in T cells.

3. Niches 1 and 2 had a higher proportion of CD8+ T cells and B cells and a lower proportion of cancer cells compared to those of all the spots.

4. Pathway of Niche 1 and 2 were associated with T cell and B cell activation functions.

5. The data were unpublished results generated by in-house IRIS-FGM.
Mathematic formulation for CCC

- The mathematic formulation is to find local-low rank matrix from gene expression matrix (row is gene and column is spot).
- Solution of determining local-low rank matrix is NP-hard.
- Approximate solution is to find a set of heavy subgraphs in a weighted graph $G$,
  - nodes is gene,
  - edge is connecting every pair of genes
  - edge weight is determined by spatial distance and transcripts similarity.

\[ W_{i,j} = e^{-\frac{D_{i,j}}{|R_{i,j}| + \epsilon}} \]

Where $D$ is spot-to-spot spatial distance matrix; $R$ is the spot-to-spot similarity matrix computed by Spearman correlation based on gene expression value, and $\epsilon$ is a pseudo-number to improve computational stability.
SAGE: a spatially-guided pattern recognition algorithm for simultaneous detection of CCC and CCC-associated CEM signatures

Downstream analysis of SAGE algorithm

- Determine the potential regulator (i.e., transcriptional factors) for Pair-CEMs and M-CEM using IRIS3.

- Decipher cell type composition using Cell2location.

- Assess CCC directionality using the linear graph neural network-based causality model.

Anjun Ma, et al., Nucleic Acids Research (2020)
David S. Fischer, et al., bioRxiv (2021)
Vitalii Kleshchevnikov, et al., Nature biotechnology (2022)
Spatial omics is coming!

Spatial profiling of chromatin accessibility in mouse and human tissues

Yanxiang Deng, Marek Bartosovic, Sai Ma, Di Zhang, Petra Kukana, Yano Xiao, Graham Su, Yang Liu, Xiaoyu Gao, Gorazd B. Rosodlija, Andrew J. Dwork, J. John Mann, Nina L. Xu, Stephanie Halene, Joseph E. Craft, Kam W. Leong, Maura Baldini, Gonzalo Castelo-Branco & Rong Fan

Nature 609, 375–383 (2022) | Cite this article

29K Accesses | 248 Altmetric | Metrics

Spatial-CUT&Tag: Spatially resolved chromatin modification profiling at the cellular level

Spatial proteomics (e.g., CODEX)
Spatial ATAC-seq
Spatial CITE-seq
Spatial CUT & Tag

Harold Hodgkinson Professor of Biomedical Engineering
Tissue module identification (Ongoing)

**TM ID interpretation**

1. SVG number.
2. Fourier coefficient as unique identifier.
3. Spatial map to show the TM distribution.
4. SVGs and functional enrichment.
5. Overlapped TMs show the interaction with other TMs.
6. Cell proportion (using cell2location) show the cell type composition in this TM.

---

**Tissue Module 5 ID Card (GC)**

- **Spatial map**
- **SVGs**
  - ACTG1
  - TUBB
  - TYMS
  - ATP5MG
  - HMGA1
  - PCNA
  - TOP2A
  - CDC20
- **SVG functional enrichments**
  - GO Biological Process 2021
    - Mitotic spindle organization
    - DNA metabolic process
    - Micrornlubule cytoskeleton organization involved in mitosis
    - Mitotic nuclear division
    - Mitotic sister chromatid segregation
- **Overlap TM with TM 1**
- **Overlap with TM 7**

---

Yuzhou Chang, et al., Nature biotechnology (under revision)
Main idea: Identify new cell type by considering morphology and transcriptional information.

Main idea: de-noise gene expression using protein or histology information.
**Acknowledgement**

**Department of Neuroscience:**
Hongjun Fu
Shuo Chen

**Department of Pathology:**
Dr. Jose Otero
Dr. Shaoli Sun

**Other organizations:**
Dr. Dong Xu
Dr. Fei He
Dr. Juexin Wang
Dr. Bingqiang Liu
Jixin Liu

---

**Construction of cell-specific gene co-regulation signatures based on single-cell transcriptomics analysis (R01-GM131399)**

**Thrombocytes in Cancer Immunity (R01-CA188419)**

**Statistical Power Calculation Framework for Spatially Resolved Transcriptomics Experiments (R21-HG012482-01, Dr. Dongjun Chung, and Dr. Qin Ma)**
THANK YOU