In Situ Technologies Enable a Pan-Omic Human Cell Atlas

Summary: Fluorescent *In Situ* Sequencing (FISSeq), Oligopaints, and Expansion Microscopy (ExM) will create a rich human cell atlas including RNA, DNA, and protein, with sub-cellular resolution, at a cost-per-cell at or below conventional single-cell next-generation sequencing (NGS).

NGS is the state-of-the-art, commercially mature method of measuring the whole-genome expression profile of single cells by dissociating cells from their constituent tissue, then isolating and detecting biomolecules in vitro [1-4]. To create an *atlas* of human cells, however, cells and their constituent RNA, DNA and protein molecules must be spatially localized within tissues and organs. Here, efforts have included *in situ* technologies, such as fluorescent *in situ* hybridization (FISH; e.g., Oligopaints, MERFISH, seqFISH) and fluorescent *in situ* sequencing (FISSeq) [5-11], which measure gene expression directly inside intact cells, tissues, and organs. These strategies present an ideal platform to generate the spatial data required for molecular localization. Importantly, spatial molecular data generated by *in situ* technologies can encompass the genome, epigenome, proteome, and transcriptome ("pan-omics"), cell and tissue morphology, and sub-cellular architecture. Thus, the information provided by *in situ* technologies can be far richer than NGS data and requires no additional spatial mapping.

NGS and FISSeq technologies share a common foundation of biochemical reactions, optical fluorescence detection, and fluidics and microscopy automation strategies [12]. With NGS, however, cellular and molecular isolation, such as within a microfluidic chamber or emulsion droplet [13,14], increases the marginal cost per measurement and decreases scalability relative to FISSeq, which can measure all cells and molecules simultaneously in place (Figure 1). Thus, despite its nascent state of commercialization, FISSeq already matches NGS per-cell and per-molecule data acquisition cost [15,16].

Previously published *in situ* RNA multiplexing methods enable detection of hundreds of RNA's per cell with single-nucleotide sensitivity [6], or thousands of RNA's per cell with gene-level sensitivity [7]. Thus fast-paced innovation in FISSeq has nearly achieved parity with NGS in sensitivity. Moreover, in-situ technologies can utilize detection modalities fundamentally more efficient than NGS, such as PCR-free single-molecule detection [17,18]. Furthermore, all these in-situ technologies can utilize Expansion Microscopy (ExM), which physically, isotropically expands a preserved tissue via hydrogel swelling and was demonstrated first with protein identification and localization with improved resolution [19]. The increased intracellular volume will also allow more targets, such as RNA molecules, to be detected within each cell, while still using conventional ultra-fast fluorescent microscopes [20], enabling improvements due to increased effective spatial resolution, as well as increased biomolecule accessibility and physical room for signal amplification [20] and sequencing [21]. In conjunction with Oligopaints, which can generate genome- and transcriptiome-wide renewable FISH and FISSeq probes, these technologies are also being extended to query the genome. These methods will facilitate simultaneous pan-omic measurements of RNA, DNA, and proteins within the same sample, allowing comprehensive capture of cell shapes, spatial configurations, connectome interactions [22], and signaling.

An emerging platform based on ExM, FISSeq, and Oligopaints therefore presents a tractable, cost-effective path towards organ-scale and organism-scale *in situ* pan-omics with sub-cellular resolution.

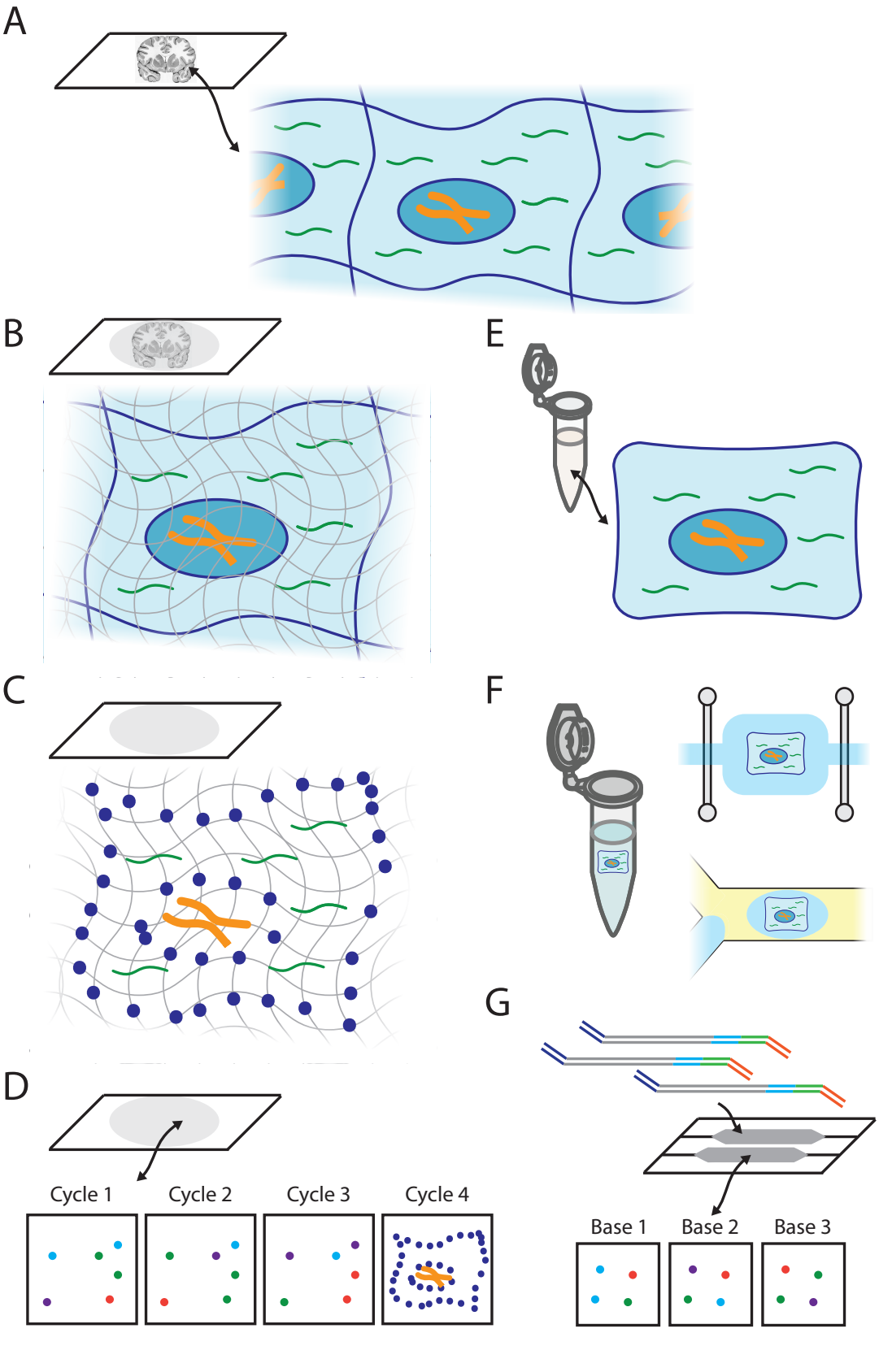
A
B
C
D
Cycle 1
Cycle 2
Cycle 3
Cycle 4
E
F
G
Base 1
Base 2
Base 3

Figure 1: A schematic comparison between *in situ* multiplexing and *in vitro* single-cell sequencing. A) A piece of human tissue, such as a brain section, consists of spatially organized single cells, each comprising a diverse complement of biomolecules including DNA, RNA, and proteins, which constitutes the cell's gene expression signature. B) For *in situ* multiplexing technologies, the sample is fixed and permeabilized, then a hydrogel is polymerized *in situ* [6], capturing and spatially immobilizing biomolecules of interest, such as RNA (green), membrane proteins (blue), and chromatin (orange). The hydrogel may be expanded to provide further resolution [19, 20]. C) The sample may be further permeabilized and clarified, such as by removing lipids, facilitating light microscopy (the process of clearing is a natural byproduct of expansion microscopy, enabling ultrafast imaging [19, 20]). *In situ* biochemical reactions, such as hybridization of oligonucleotide probes to RNA or the genome, or reverse transcription of RNA into cDNA, also facilitate molecular detection. D) The sample is then iteratively probed and imaged, enabling read-out of information rich RNA, cDNA, or genomic sequences, such as by fluorescent sequencing chemistry, molecular beacons and barcodes, and other types of molecular stains. Cycles 1-3 depict three fluorescent DNA sequencing reactions directed to the RNA molecules, with each of the four bases encoded by a distinct fluorescence color. Cycle 4 depicts a protein stain in purple, detecting membrane-bound proteins, and a DNA stain in orange, detecting the chromatin. E) In comparison, for *in vitro* single-cell RNA and DNA sequencing or proteomics, the tissue sample is dissociated into single cells, which may perturb the cell state. F) Single cells are isolated into tubes, microfluidic chambers, or emulsion droplets, and then individually processed into a sequencing library via biochemical reactions such as reverse transcription, adaptor ligation, and PCR. G) The sequencing template molecules, which typically include cell and molecular barcodes, are then pooled onto a solid substrate for fluorescent sequencing. Three bases of the fluorescent sequencing reaction are depicted, as in D with each of the four bases encoded by a distinct fluorescence color.

Acknowledgements (in alphabetical order): Shahar Alon[1], Edward S. Boyden[1†], Fei Chen[1,2], George M. Church[3,5†], Evan R. Daugharthy[3,4†], Samuel Inverso[4], Adam Marblestone[1], Sam Rodriques[1], Chao-Ting Wu[5†]

1 MIT Media Lab 2 Broad Institute 3 Wyss Institute 4 ReadCoor Inc 5 Harvard Medical School Department of Genetics

† Correspondence:
Edward S. Boyden (esb@media.mit.edu), George M. Church (gmc@harvard.edu), Evan R. Daugharthy (erd@readcoor.com), Chao-Ting Wu (twu@genetics.med.harvard.edu)

References:

[1] D. Ramsköld, et al. "Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells." *Nature biotechnology* 30.8 (2012): 777-782;
[2] AK Shalek, et al. "Single cell RNA Seq reveals dynamic paracrine control of cellular variation." *Nature* 510.7505 (2014): 363;
[3] T. Hashimshony, et al. "CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq." *Genome biology* 17.1 (2016): 1;
[4] S. Picelli, et al. "Smart-seq2 for sensitive full-length transcriptome profiling in single cells." *Nature methods* 10.11 (2013): 1096-1098.
[5] R. Ke, et al. "In situ sequencing for RNA analysis in preserved tissue and cells." *Nature methods* 10.9 (2013): 857-860.
[6] JH Lee, et al. "Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues." *Nature protocols* 10.3 (2015): 442-458.
[7] KH Chen, et al. "Spatially resolved, highly multiplexed RNA profiling in single cells." *Science* 348.6233 (2015): aaa6090.
[8] JR Moffitt, et al. "High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization." *Proceedings of the National Academy of Sciences* 113.39 (2016): 11046-11051.

[9] E. Lubeck, et al. "Single-cell in situ RNA profiling by sequential hybridization." *Nature methods* 11.4 (2014): 360-361.
[10] S. Shah, et al. "In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus." *Neuron* 92.2 (2016): 342-357.
[11] BJ Beliveau, et al. "Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes." *Proceedings of the National Academy of Sciences* 109.52 (2012): 21301-21306.
[12] J. Shendure, et al. "Accurate multiplex polony sequencing of an evolved bacterial genome." *Science* 309.5741 (2005): 1728-1732.
[13] JS Marcus, et al. "Microfluidic single-cell mRNA isolation and analysis." *Analytical chemistry*78.9 (2006): 3084-3089.
[14] EZ Macosko, et al. "Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets." *Cell* 161.5 (2015): 1202-1214.
[15] NGS cost calculation assumes the price of genomic sequencing on Illumina platforms is approximately $0.10/megabase (data collected from public sources, e.g., https://www.genome.gov/sequencingcostsdata/), which is similar to the costs we have obtained from core facility price sheets (personal correspondence). This gives a price of roughly $10^{-5} for a single 100 bp read. Thus an 100 bp sequencing experiment yielding 10^{4} unique RNA molecules (UMIs) per cell, with a conservative raw read-to-UMI ratio of 2, would cost approximately $0.20 per cell.
[16] Quote for FISSEQ services (per cell and per read) provided by ReadCoor, Inc., Cambridge MA 02138.
[17] P. Brennecke, et al. "Accounting for technical noise in single-cell RNA-seq experiments." *Nature methods* 10.11 (2013): 1093-1095.
[18] A. Raj, et al. "Imaging individual mRNA molecules using multiple singly labeled probes." *Nature methods* 5.10 (2008): 877.
[19] F. Chen, et al. "Expansion microscopy." *Science* 347.6221 (2015): 543-548.
[20] F. Chen, et al. "Nanoscale imaging of RNA with expansion microscopy." *Nature Methods* 13.8 (2016): 679-684.
[21] S. Alon, et al. “Expansion Sequencing (ExSEQ): comprehensive *In situ* transcriptome characterization throughout intact brain circuits.” Abstract. *Society for Neuroscience* (2015): 734.13/DD29.
[22] AH Marblestone, et al. "Rosetta Brains: A Strategy for Molecularly-Annotated Connectomics." *arXiv preprint arXiv:1404.5103* (2014).