

Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence *in situ* hybridization data

Qian Zhu¹, Sheel Shah^{2,3}, Ruben Dries¹, Long Cai² & Guo-Cheng Yuan¹

Faculty: Mengjie Chen & Student: Yanyu Liang

GGSB Journal Club

5/2/19

Overview

Background - On the other side of NGS

Main - Paper discussion

Spatial genomics - what else?

FISH and single molecule FISH

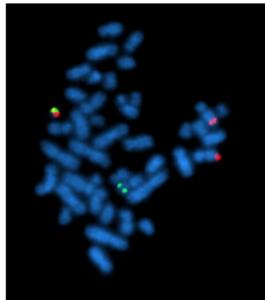


Figure: FISH

[Wikipedia, 2019]

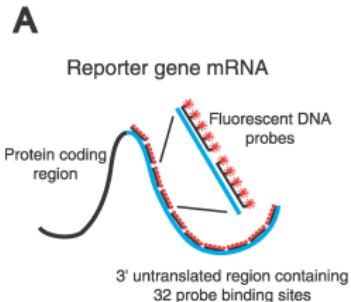


Figure: smFISH

[Raj et al., 2006]

- ▶ FISH: fluorescence in situ hybridization
- ▶ Single molecule FISH: pioneered by [Femino et al., 1998], enhanced version by [Raj et al., 2006]
 - ▶ Multiple short probe/fluorophore → Brighter → Higher sensitivity and specificity

Super-resolution microscopy

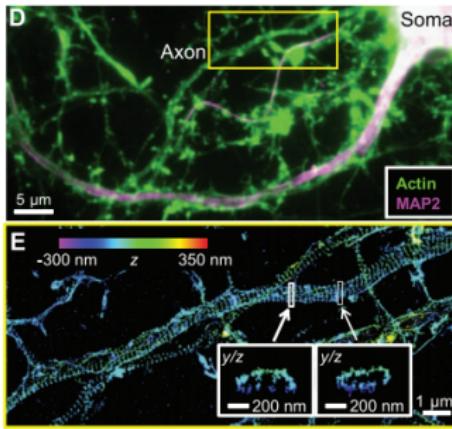


Figure: STORM image of neuron

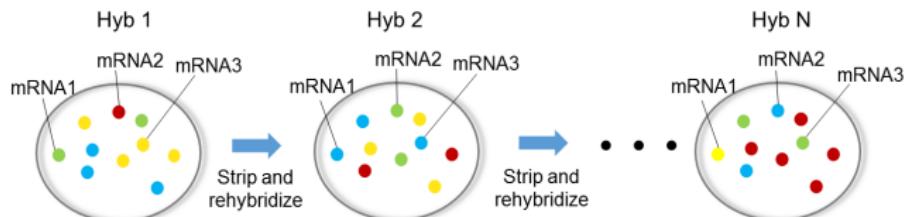
[Xu et al., 2013]

- ▶ Resolution 10-20 nm on x, y and < 50 nm on z direction.
- ▶ It helps to reveal finer structure (under diffraction barrier, ~ 250nm)

Sequencing by hybridization

- ▶ Encode sequence by combination of color [Lubeck and Cai, 2012]
- ▶ Encode sequence by order of color [Lubeck et al., 2014] (seqFISH)
- ▶ Two rounds three channels:
 - ▶ red- cyan → Gene 1
 - ▶ green- cyan → Gene 2
 - ▶ cyan- red → Gene 3
- ▶ The complexity of the target library: F^N where N is number of hybridization and F is number of colors

Sequential FISH: an example



Barcode # scales as F^N

$$5^3=125$$

$$5^6=15,625$$

Barcode Key: (for $N=6$)

mRNA1: ●●●●●●

mRNA2: ●●●●●●

mRNA3: ●●●●●●

Figure: seqFISH barcoding: four colors 6 rounds

[seqFISH, 2019]

What is good about seqFISH?

- ▶ Low drop out rate ($\sim 94\%$ per round) [Shah et al., 2016]
- ▶ So that it is also easy to correct for this drop out error by one extra round
- ▶ As an *in situ* technology, it preserves spatial information which is erased in scRNA-seq

Bring spatial information into scRNAseq

- ▶ Sources of variation in gene expression:
 - ▶ Cell type?
 - ▶ Cell state?
 - ▶ Cell cycle?
 - ▶ Spatial localization?
- ▶ Benefit: account for spatial variation in differential analysis

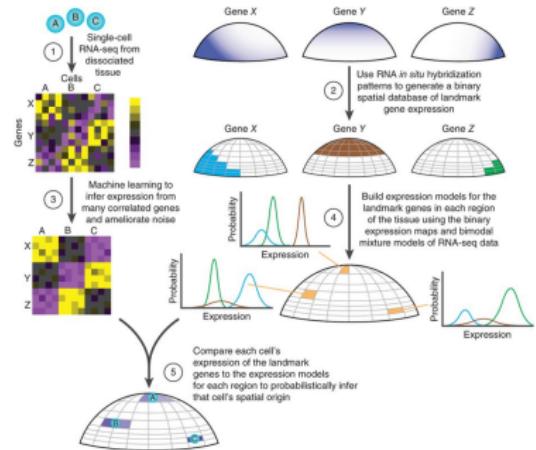
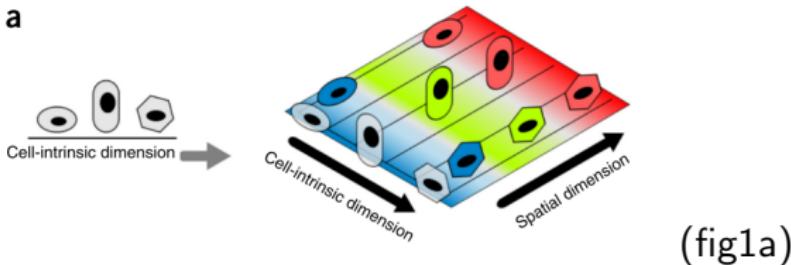


Figure: An example: using spatial marker genes in embryo

[Satija et al., 2015]

Combine seqFISH and scRNAseq



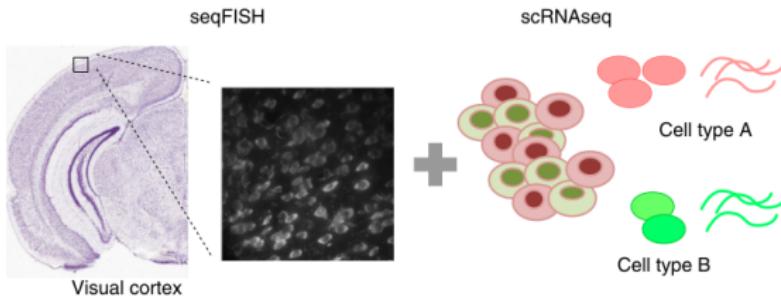
Questions:

- ▶ Is there any spatial pattern/variation in gene expression profile?
- ▶ How to quantify/account for spatial variation if it exists?
- ▶ What do we learn about spatial variation?

Solutions:

- ▶ Detect spatial pattern using HMRF
- ▶ Construct a set of spatial markers
- ▶ Within cell type, how spatial difference contributes to heterogeneity

Collect seqFISH and scRNAseq data



(fig1b)

seqFISH:

- ▶ 1-mm × 1-mm contiguous area of the mouse visual cortex
- ▶ 5 colors, 4 rounds
- ▶ 125 genes were profiled in 1597 cells

scRNAseq:

- ▶ Published scRNAseq data set on the mouse visual cortex region
- ▶ 8 major cell types identified: GABAergic, glutamatergic, astrocytes, three oligodendrocyte groups, microglia and endothelial cells

Identify spatial pattern

- ▶ Spatial pattern of interest:
 - ▶ Sub-region with region-specific expression profile
 - ▶ Nearby cells are likely to be in the same sub-region
- ▶ Data with such spatial pattern can be easily described by hidden Markov random field (HMRF)
 - ▶ C_i region assignment of cell i
 - ▶ E_i expression profile of cell i
 - ▶ Markov random field

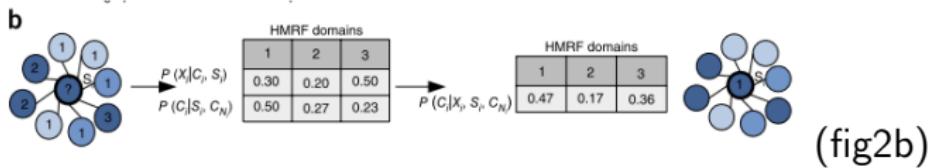
$$\Pr(C_i | C_k : k \text{ neighbours } i)$$

- ▶ Expression profile signature for each sub-region

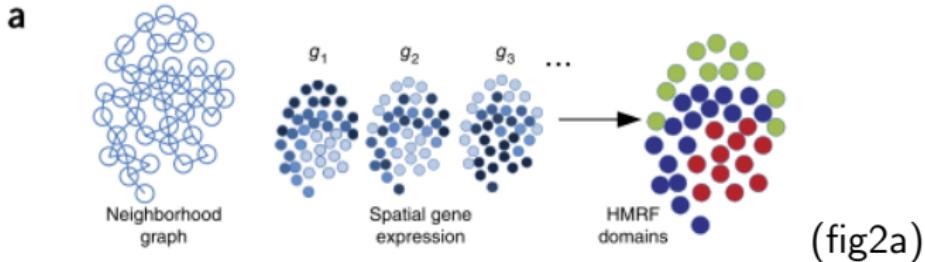
$$\Pr(E_i | C_i)$$

Hidden Markov random field

$$\Pr(C_i|C_k : k \text{ neighbours } i)$$



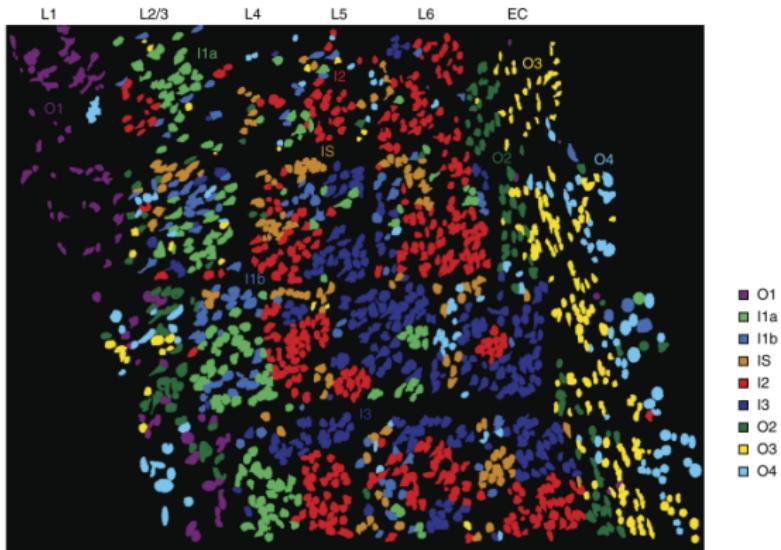
$$\Pr(E_i|C_i)$$



Apply to seqFISH data

- ▶ Pre-select a set of genes with spatial coherence (similar if spatially close), appendix 4
- ▶ Determine the neighbours of a cell by spatial information captured by seqFISH
- ▶ Fit HMRF by EM algorithm (determine the number of sub-regions by Kmeans, appendix 4)

Spatial dissection of seqFISH data



(fig2c)

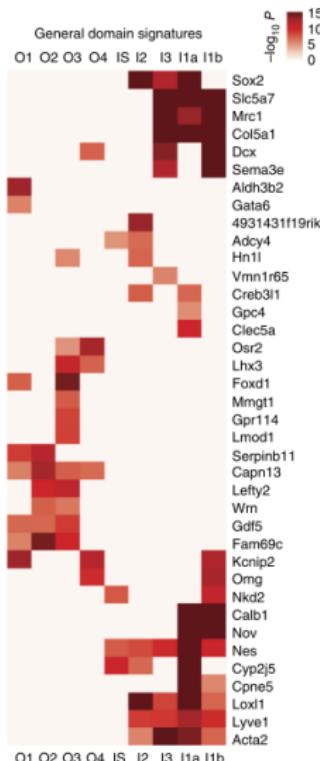
- ▶ 69 out of 125 genes were used
- ▶ 9 sub-regions identified
- ▶ Resembled anatomical structure

Identify region-specific genes

For each sub-region, if a gene is

- ▶ Significantly up-regulated as compared to 7 out of 8 other sub-region
- ▶ Significantly up-regulated as compared to all the rest (p-value shown on the right)

The gene is identified as domain signature



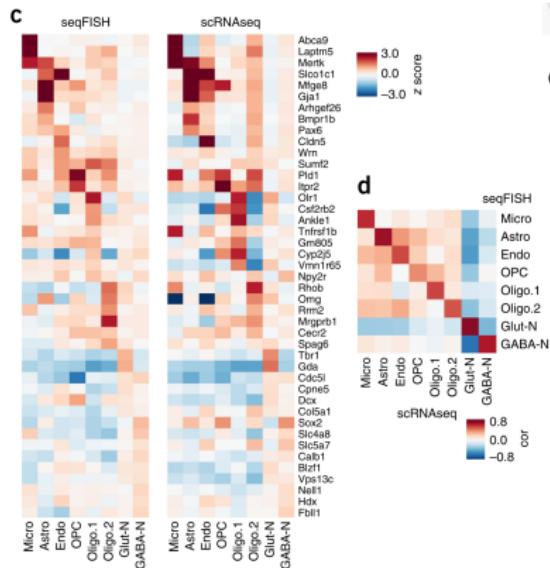
(fig2d)

Impute cell type?

- ▶ Use scRNAseq data to learn the characteristics of each cell type
- ▶ Since genes are highly correlated, using ~ 40 genes have already give 89% accuracy
- ▶ Match the scale of scRNAseq and seqFISH by quantile normalization

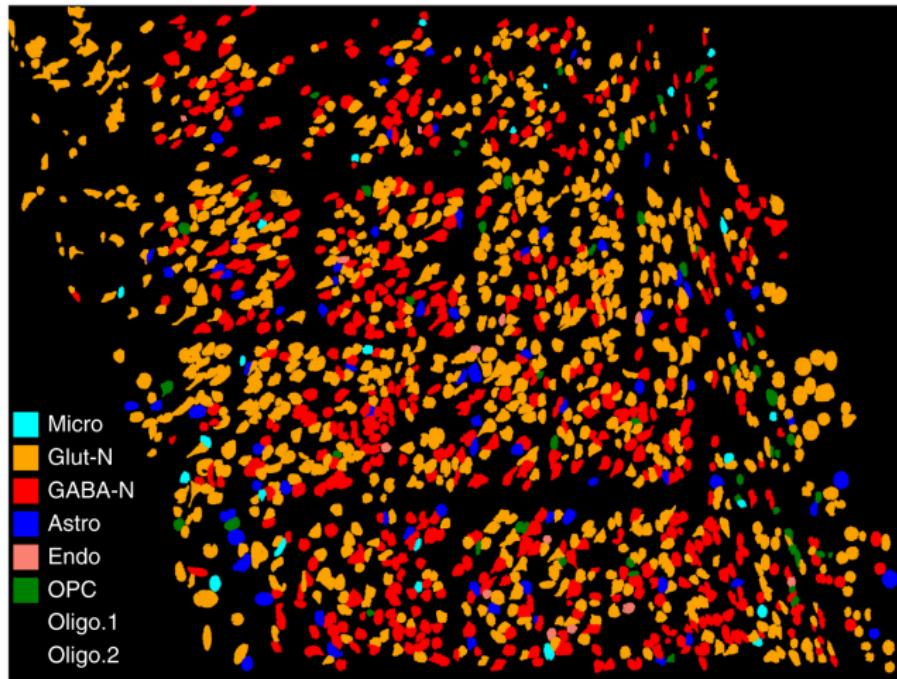
Impute cell type in seqFISH using scRNAseq

- ▶ 43 out of 125 differentially expressed seqFISH genes were used achieving 90% accuracy on scRNAseq
- ▶ 1597 seqFISH cells were classified into: glutamatergic neurons (54%), GABAergic neurons (37%), astrocytes (4.8%), and other glial cell types and endothelial cells (4.2%)



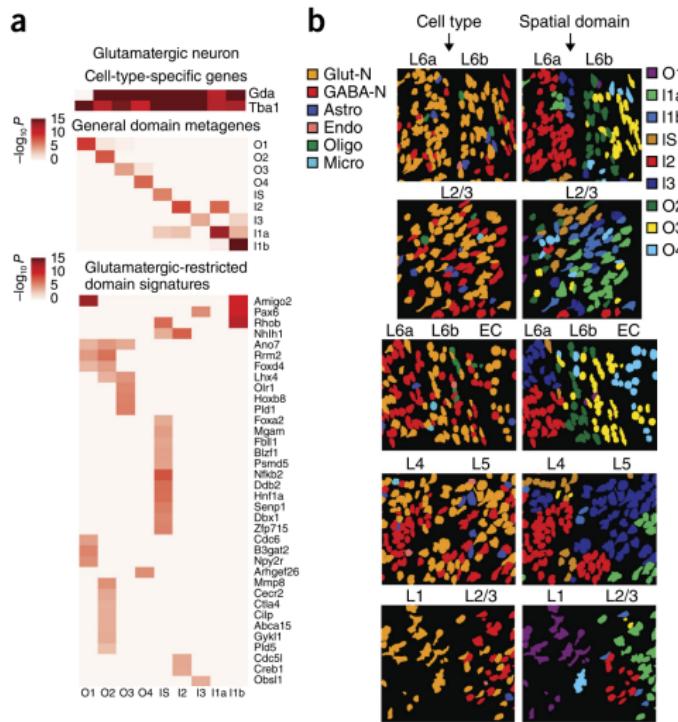
(fig1c,d)

Spatial distribution of cell type



(fig1e)

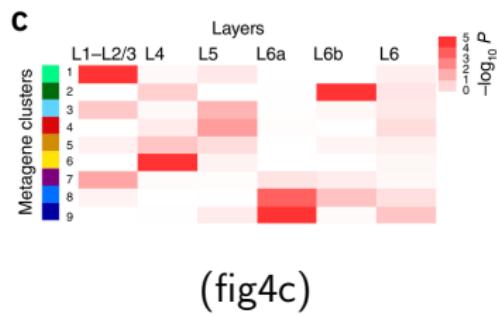
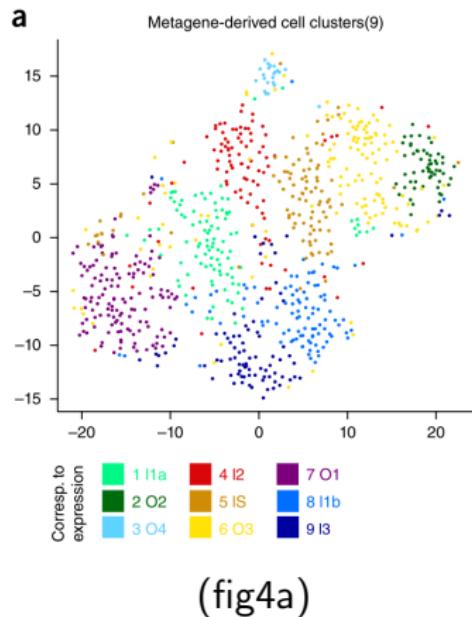
How does spatial variation contribute to cellular heterogeneity?



(fig3)

- ▶ Within cell type, identify more domain markers
- ▶ How spatial domain is associated with morphology

Re-analyze scRNAseq with imputed spatial assignment: spatial pattern



Spatial pattern also appears in scRNAseq data. Marker gene match the layer structure.

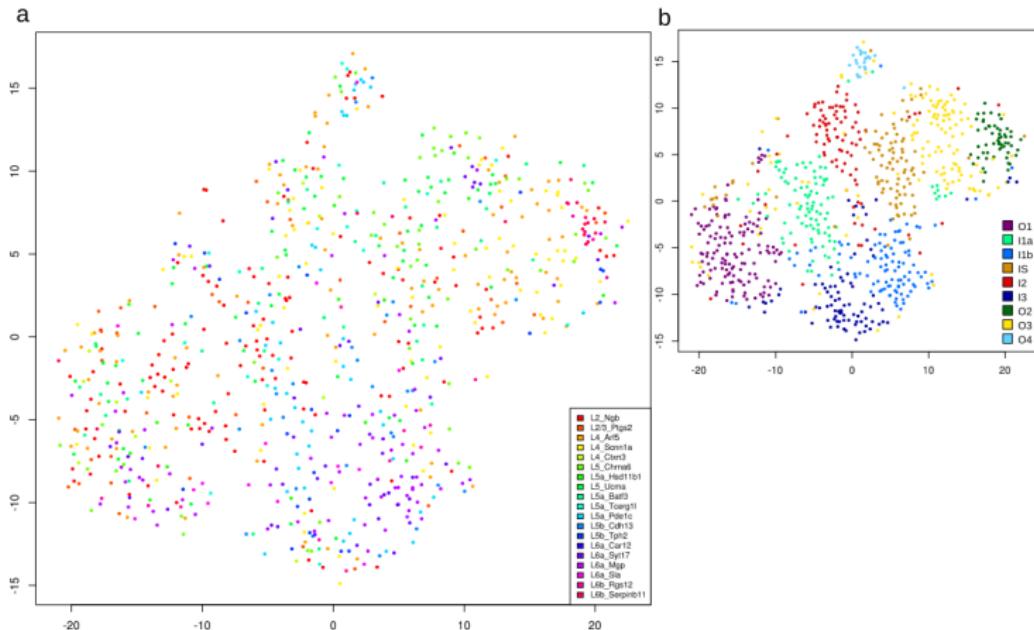
Re-analyze scRNAseq with imputed spatial assignment: spatial markers

With enlarged gene sets, more domain-specific genes were identified. And they are enriched in distinct Gene Ontology biological processes



(fig4d)

Does spatial variation explain cell subtypes?

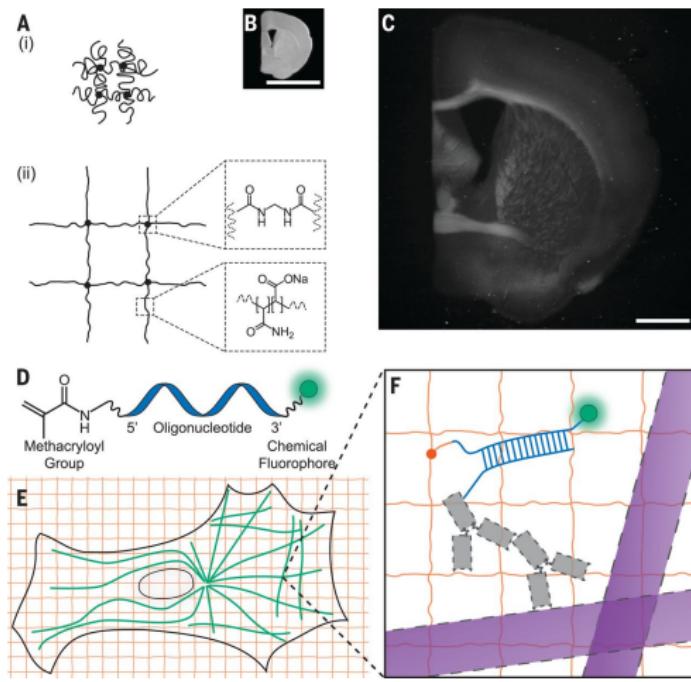


(figS14)

Take-aways

- ▶ There is clear spatial pattern in visual cortex of mouse
- ▶ Such spatial variation contributes to heterogeneity of cells (expression and morphology)
- ▶ And it is distinct source variation as compared to cell subtype (based on expression profile)

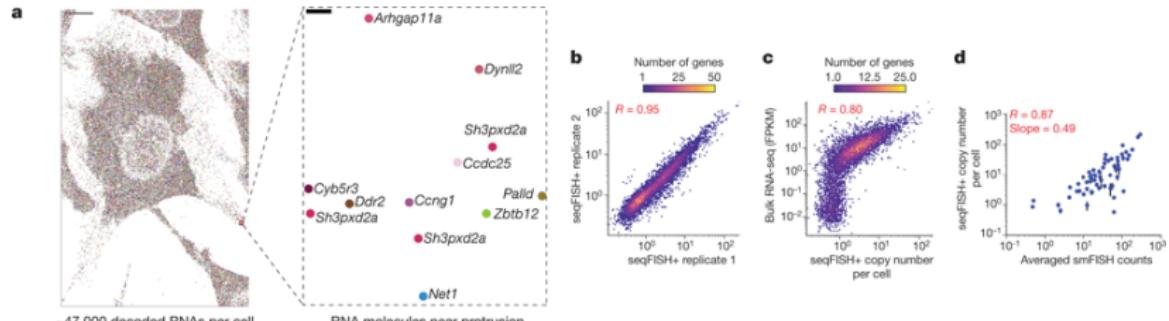
Technology side - expansion microscopy



[Chen et al., 2015]

Fixing molecule on gel and expand gel so that the object of interest is enlarged isotropically.

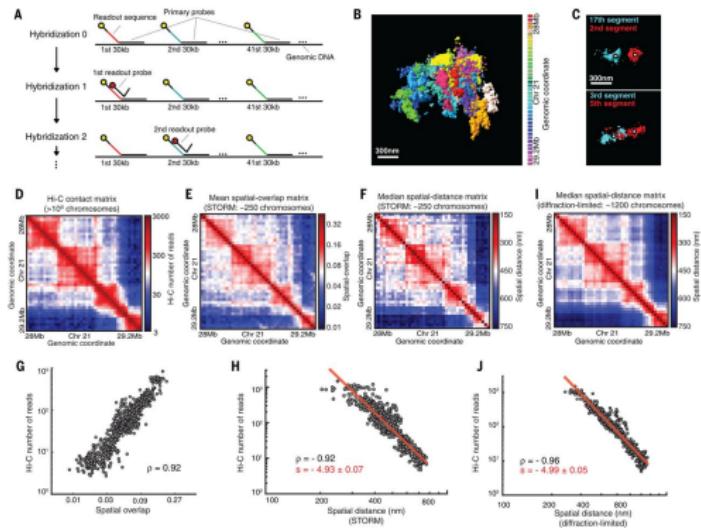
Technology side - SeqFISH+



[Eng et al., 2019]

- ▶ Dilute signal by 60 psuedo-colors (in three channels); 3 rounds plus one for error correction ($\underbrace{20^3}_{\text{each channel}} \times \underbrace{3}_{\text{number of channels}}$)
- ▶ 10000 genes, ~ 3000 cells, efficacy 49% (compared to smFISH)

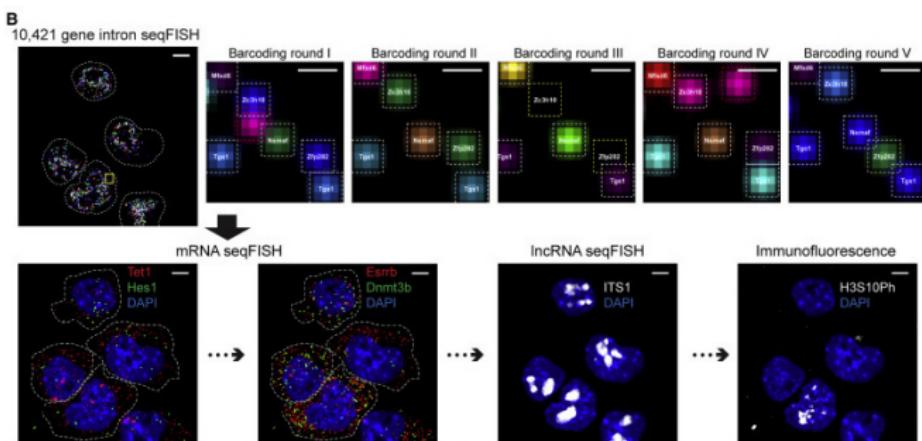
Application side - chromatin interaction



[Bintu et al., 2018]

- ▶ Sequence resolution in kb, spatial resolution in nanometer, profiled thousands of single cells
- ▶ Recover Hi-C result by population averaging
- ▶ BUT with single cell resolution and higher-order interaction

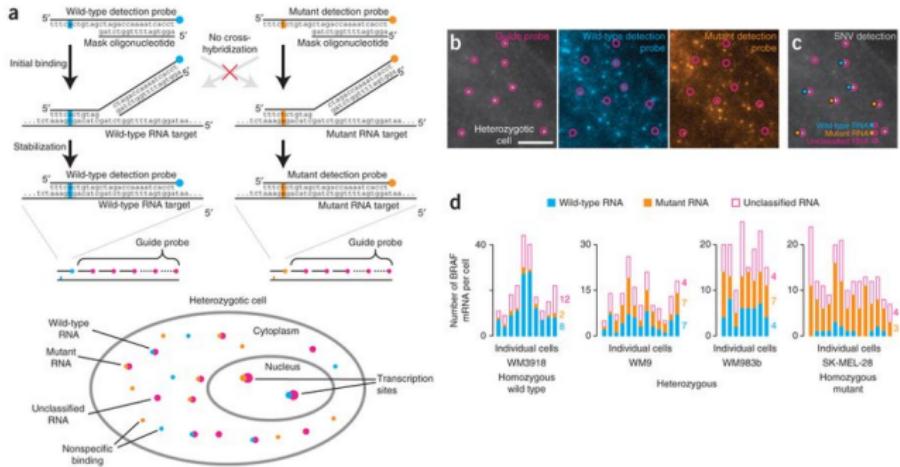
Application side - beyond mRNA



[Shah et al., 2018]

- ▶ Nascent transcriptome with mRNAs and lncRNAs profiled in the same cell

Application side - imaging allele-specific expression



[Levesque et al., 2013]

- ▶ Clever design on probe to enable allele-specific detection

Any downside? Of course

- ▶ Protocol is hard to implement ...
- ▶ To image real-world sample can be very challenging (transparency, thickness)
- ▶ Lack of isoform information (although it is possible)

Thank You

References |

-  Bintu, B., Mateo, L. J., Su, J.-H., Sinnott-Armstrong, N. A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A. N., and Zhuang, X. (2018).
Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells.
Science, 362(6413):eaau1783.
-  Chen, F., Tillberg, P. W., and Boyden, E. S. (2015).
Expansion microscopy.
Science, 347(6221):543–548.
-  Eng, C.-H. L., Lawson, M., Zhu, Q., Dries, R., Koulen, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.-C., et al. (2019).
Transcriptome-scale super-resolved imaging in tissues by rna seqfish+.
Nature, page 1.
-  Femino, A. M., Fay, F. S., Fogarty, K., and Singer, R. H. (1998).
Visualization of single rna transcripts in situ.
Science, 280(5363):585–590.

References II

-  Levesque, M. J., Ginart, P., Wei, Y., and Raj, A. (2013). Visualizing snvs to quantify allele-specific expression in single cells. *Nature methods*, 10(9):865.
-  Lubeck, E. and Cai, L. (2012). Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nature methods*, 9(7):743.
-  Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M., and Cai, L. (2014). Single-cell *in situ* rna profiling by sequential hybridization. *Nature methods*, 11(4):360.
-  Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., and Tyagi, S. (2006). Stochastic mrna synthesis in mammalian cells. *PLoS biology*, 4(10):e309.

References III

-  Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., and Regev, A. (2015).
 Spatial reconstruction of single-cell gene expression data.
Nature Biotechnology, 33:495.
-  seqFISH (2019).
 What is seqfish?
<https://www.seqfish.com/technology>.
-  Shah, S., Lubeck, E., Zhou, W., and Cai, L. (2016).
 In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus.
Neuron, 92(2):342–357.
-  Shah, S., Takei, Y., Zhou, W., Lubeck, E., Yun, J., Eng, C.-H. L., Koulena, N., Cronin, C., Karp, C., Liaw, E. J., et al. (2018).
 Dynamics and spatial genomics of the nascent transcriptome by intron seqfish.
Cell, 174(2):363–376.

References IV



Wikipedia (2019).

Fluorescence in situ hybridization — Wikipedia, the free encyclopedia.

<http://en.wikipedia.org/w/index.php?title=Fluorescence%20in%20situ%20hybridization&oldid=892650336>.

[Online; accessed 24-April-2019].



Xu, K., Zhong, G., and Zhuang, X. (2013).

Actin, spectrin, and associated proteins form a periodic cytoskeletal structure in axons.

Science, 339(6118):452–456.

Coherent score: $\delta_g = \frac{1}{|L_1|} \sum_{s_i \in L_1} \frac{m_i - n_i}{\max(m_i, n_i)}$, L_1 highly expressed cells, m_i average 'distance' to L_0 , n_i average 'distance' to L_1 .
Significance of δ_g is done by permutation.

Select the number of clusters: gap statistic,
 $gap(k) = \mathbb{E}[\log W_k] - \log W_k$, W_k is the sum of average within cluster squared distance. Expectation is taken under the null (no cluster structure; obtained from Bootstrap). Criteria: smallest k such that $gap(k + 1) - sd(k + 1) < gap(k)$ (namely no significant increase). Select k by gap statistic under k-means