Combining droplet and full-length sequencing technologies for a complete picture

DATE: March 17, 2020
AUTHOR: Takara Bio Blog Team
CATEGORIES: Research News | NGS | Single-cell

BIO VIEW

The advent of next-generation sequencing techniques in the early twenty-first century revolutionized the way we approach genetic research. As it has become clear that bulk sequencing experiments can mask important biological phenomena of rare cell types, researchers have turned to single-cell sequencing to better understand how individual cells contribute to the complex biology of whole systems.

Droplet-based methods have been favored for single-cell sequencing in part because of the high number of cells they can assay. However, researchers are starting to find that droplet-based methods alone are not enough to fully understand rich single-cell datasets. In a recent study, researchers from the California Institute of Technology and Allen Institute of Brain Science, whose sequencing core director we recently interviewed, revealed that a multipronged approach utilizing SMART-Seq full-length, droplet-based, and spatial scRNA-seq constitutes a complementary, well-rounded approach to single-cell sequencing, especially in the brain (Booeshaghi et al. 2020).

Teaming up

The researchers started by examining preexisting sequencing data from mouse primary motor cortex (MOp) samples generated by the BRAIN Initiative Cell Census Network. These data gave them access to full-length SMART-Seq v4 (SMART-Seq) data for 6,160 cells and 10x Genomics Chromium (10xv3) for 90,031 cells (Figure 1, Panel A). Both technologies were able to identify broad classes of cell types. More careful analysis revealed that while 10xv3 excelled at identifying a few rare cell types due to the greater number of cells assayed, the SMART-Seq method's higher sensitivity and probing of full-length transcripts allowed for isoform quantification as well as refined cell type classification.

Next, the researchers turned their focus to differential analysis of the isoform abundances for glutamatergic and GABAergic neurons determined with SMART-Seq v4. For 260 genes that showed stable expression across the two cell types, 312 isoform markers were observed. The *H3f3b* gene is one such example where one isoform was highly expressed in glutamatergic neurons, while GABAergic neurons showed lower expression of the *H3f3b-204* isoform (Figure 1, Panel B). The authors note, "A gene-level analysis is blind to this isoform shift." SMART-Seq isoform quantifications, cell subclass, and cluster assignments were all validated through comparison with 10xv3 data. Combined, these data allowed the development of an isoform atlas of the MOp (Figure 2).

a

	SMART-Seq	10xv3	MERFISH
reads	15,229,289,828	22,696,236,495	NA
cells	6,295	94,162	243,799
avg # genes detected/cell	10,333	5,891	74
avg # isoforms detected/cell	20,319	NA	NA
clusters	62	147	93

c

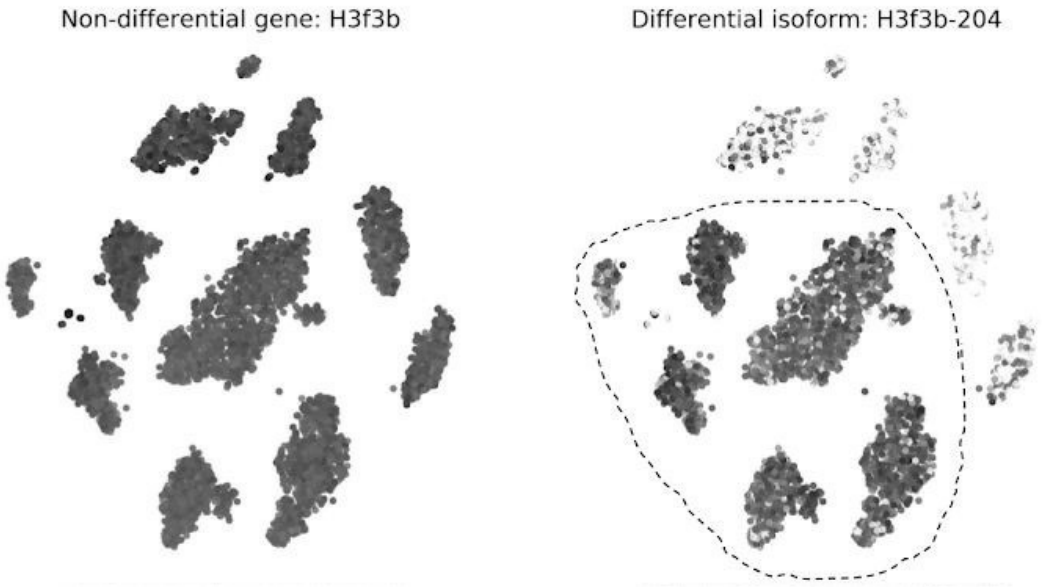

Figure 1. Isoform specificity in the absence of gene specificity. Panel A. Overview of data analyzed. **Panel C.** The *H3f3b* gene abundance distribution across cells (left) and *H3f3b-204* isoform distribution across cells (right). Image and caption adapted from Figure 1 of Booeshaghi et al. 2020 (some panels not shown) and used under a Creative Commons Attribution 4.0 International License.

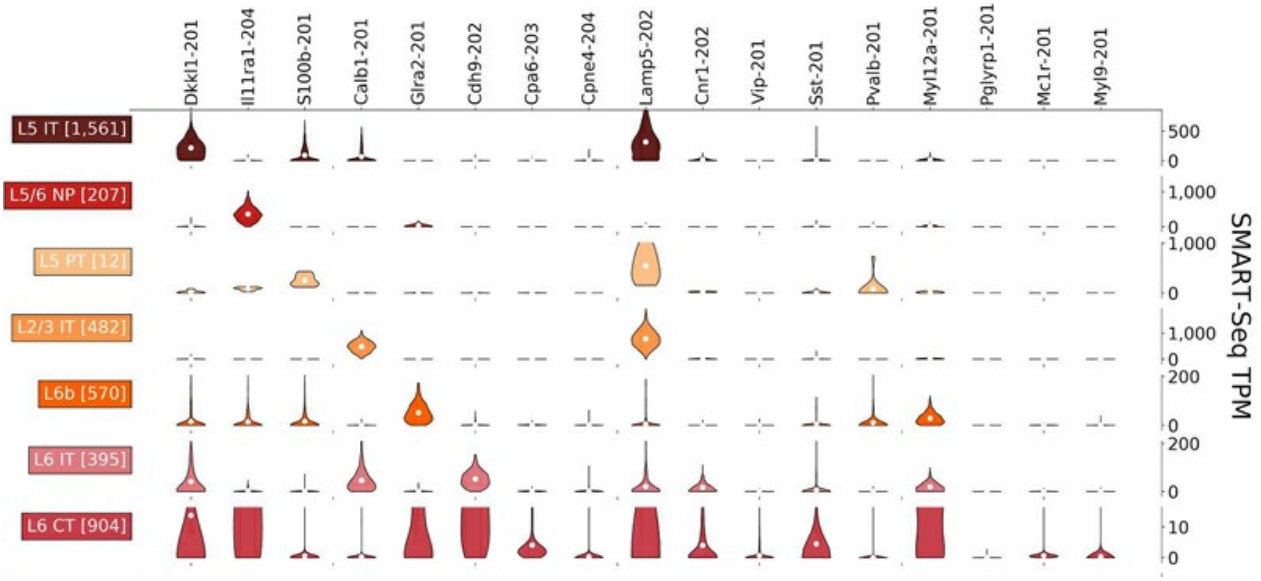

Figure 2. Isoform atlas. A sample from an isoform atlas displaying isoform markers differential with respect to subclasses. Each row corresponds to one subclass, and each column corresponds to one isoform. SMART-Seq isoform abundance estimates are in TPM units, and each column is scaled so that the maximum TPM is four times the mean of the isoform specific for that row's cluster. Image and caption adapted from Figure 2 of Booeshaghi et al. 2020 (subset of figure not shown) and used under a Creative Commons Attribution 4.0 International License.

To further their understanding of isoform expression, they next sought to use the SMART-Seq isoform quantifications to refine data derived from spatial MERFISH data. The authors took a close look at the *Pvalb* gene and found that MERFISH results show it to span all layers of motor cortex slices analyzed. Quantification of the SMART-Seq data, however, indicates that only one of the two isoforms for the gene is expressed. This spatial expression pattern of the isoforms would have been missed if examination had been performed only with the gene-level analysis provided by MERFISH.

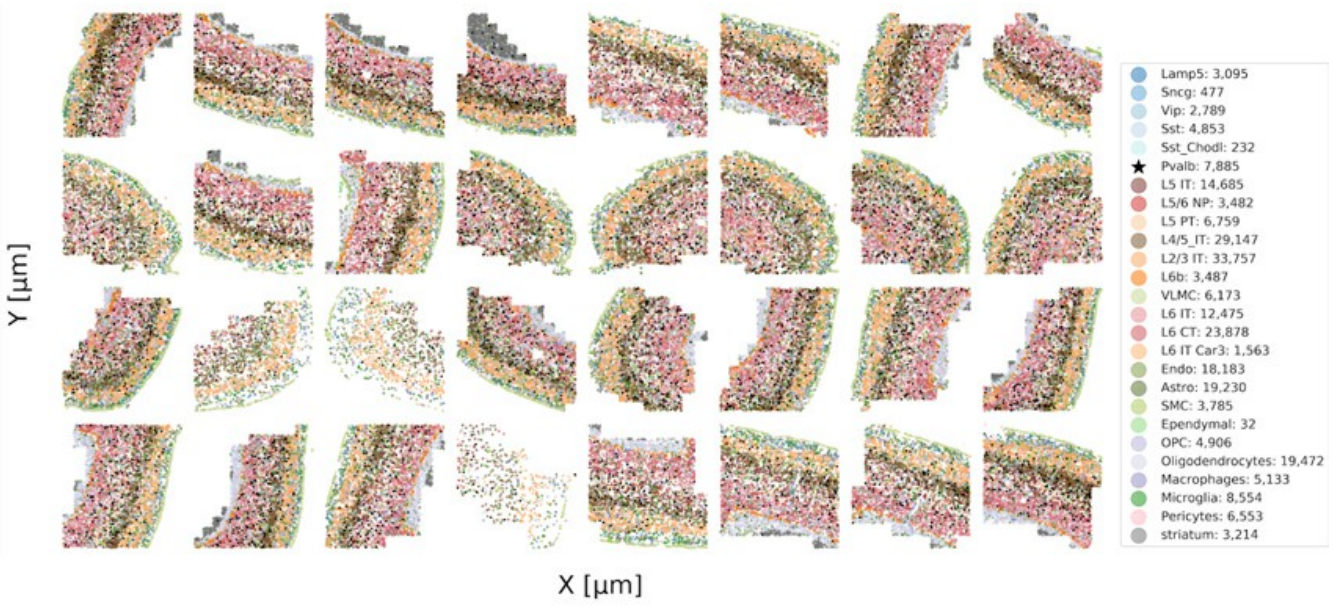

Figure 3. Spatial location of cells in all subclasses across slices from the MOp, the Pvalb cells are starred. Image and caption adapted from Extended Data Figure 7, Panel B of Booeshaghi et al. 2020 (some data not shown) and used under a Creative Commons Attribution 4.0 International License.

A new age for scRNA-seq analysis

Aristotle once said, "The whole is greater than the sum of its parts." This study is an excellent example of how using different technologies to approach a research question can advance our understanding to much greater heights than focusing on just one. For scRNA-seq specifically, the authors provide the following perspective:

> The [SMART-Seq] technology should be viewed as a complement, rather than competitor, to droplet or spatial single-cell RNA-seq. Our analyses suggest that a workflow consisting of droplet-based single-cell RNA-seq to identify cell types, then SMART-Seq [v4 chemistry] for isoform analysis, and finally spatial RNA-seq with a panel based on isoform-specific markers identified by SMART-Seq [v4 chemistry], would effectively leverage different technologies' strengths."

And we think that's good science!

References

Booeshaghi, A. S. *et al.* Isoform cell type specificity in the mouse primary motor cortex. *bioRxiv* 2020.03.05.977991 (2020).

Takara Bio USA, Inc.
United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.565.6999