Best Practices for Single-Cell Research: a Guide to Orthogonal Validation Methods

Single-cell research has rapidly emerged as a revolutionary technique to reveal insights into the cellular heterogeneity governing many biological systems. The advent of single-cell technologies such as the C1™ system that simplify and streamline the workflow has led to an explosion of papers in a variety of fields. In particular, single-cell RNA sequencing is now commonly used to profile and classify the cell types present in a tissue, discover new gene signatures for a cell type or pathway, elucidate the developmental lineage of a cell type, and compare different biological states or conditions.

However, the ability to validate the findings from a broad, hypothesis-free approach such as RNA seq is critical to drawing accurate conclusions. Because single-cell RNA seq is a rapidly developing field, consensus on the best validation methods is still evolving. But with adoption of single-cell RNA seq becoming widespread, the need for standards for validation grows.

Single-cell RNA seq has provided a new method to cast a broad net to study a large number of genes from a large number of cells simultaneously. Many targeted, lower-throughput methods can offer excellent confirmation of RNA seq results. These methods would often be too laborious to carry out on the entire transcriptome, but RNA seq results can be used to guide more targeted follow-up efforts. In this note, we summarize commonly used orthogonal validation techniques to verify single-cell discoveries.

Methods for orthogonal validation

Single-cell questions such as the identification of new cell types are frequently answered by the discovery of new marker genes or the identification of multiple genes co-expressed in the same cell. It is therefore critical that any validation method demonstrates the presence of the gene or combination of genes within the cellular context. Table 1 summarizes the most common methods, several of which are discussed here.

Imaging techniques offer an excellent approach to validate localization or co-expression and can provide additional spatial information. Spatial information, such as the brain layer or tissue type in which a gene functions, is often critical to interpreting results. Imaging methods can use either genomic targets, such as in situ hybridization (ISH) or fluorescent in situ hybridization (FISH), or use protein targets, such as immunohistochemistry (IHC), or immunofluorescent imaging. These methods all require the generation of suitable probes for each target, which can be challenging for antibody-based methods. Additionally, all microscopy-based methods are limited by the number of colors or markers that can be detected simultaneously. Imaging is frequently used to effectively validate single-cell findings (2, 4, 5). For example, Zeisel et al. (5) identified new types of cortical neurons by single-cell RNA seq, then confirmed the co-expression of key markers by FISH.

Flow cytometry is another orthogonal technique that can accurately confirm the co-expression of newly discovered targets. Flow cytometry can utilize either RNA or antibody-based targets and provides quantitative single-cell information on the expression of multiple genes or proteins. Flow cytometry has been used by many important papers as a validation method (1, 3). Mahata et al. identified a new surface marker by single-cell RNA seq, then used fluorescence-activated cell sorting to both validate the co-expression and purify the subpopulation.

However, like imaging, flow cytometry is limited in the number and the availability of probes that can be used simultaneously—frequently only 4 to 12 targets. The Fluidigm Helios™, a CyTOF® system, employs mass cytometry rather than fluorescence, which allows for the detection of dozens of proteins simultaneously. RNA seq results typically produce a large number of hits, so the use of such a high-throughput, protein-based, and single-cell system offers an excellent orthogonal method to validate entire pathways or panels.

While traditional protein-based assays are not sensitive to the single-cell level, they can also be useful when confirming gene expression results as long as the caveats of bulk analysis are included in the interpretation. Western blot can demonstrate the co-expression of targets within a sample, though it is also limited by antibody reagents and is fairly low-throughput. Mass spectrometry can provide another broad, unbiased method for orthogonal validation and can provide another large dataset to cross-reference with RNA seq results. Mass spectrometry was used to cross-validate single-cell RNA seq results by Romanov et al. (4), who were able to define a new subtype of neurons through combining multiple approaches.

Functional assays can unequivocally confirm the role of a set of genes in a biological system. Functional assays often require significant development but provide the strongest validation. Genetic manipulations such as siRNA knockdown, overexpression, or CRISPR assays can be very effective in providing functional information of newly discovered targets. Mahata et al. (3) identified a previously unknown pathway involved in steroid biosynthesis in Th2 cells through standard

(bulk) RNA seq, then identified key markers of a Th2 subpopulation through single-cell RNA seq. They performed several functional assays on the newly discovered subpopulation to directly validate their findings. Similarly, Behrens et al. (1) developed an inducible overexpression system to unequivocally demonstrate the relationship between two genes discovered by single-cell qPCR.

Lastly, all findings should be interpreted with the background of the field in mind; literature and historical data can provide much-needed context for interpreting new results. Historical data was used effectively by Zeisel et al. (5), who in addition to using FISH to verify novel targets were also able to correlate their sequencing data for new neuronal markers with the known expression information found in the Allen Brain Atlas.

Future methods: functional genomics to streamline discovery and validation

The need to confirm new findings by orthogonal methods is clear. Ideally, the validation method can also provide additional functional information. However, as discussed above, most techniques are low-throughput in the number of samples or parameters and distinct from hypothesis-generating methods. The Fluidigm Polaris™ system is well-suited to streamline both validation and functional genomics. The Polaris system integrates single-cell selection, dosing, and culturing upstream of RNA seq. Researchers can now quickly and easily validate new markers through the selection feature while also testing and developing new hypotheses through the dosing and culturing aspects in the same experiment. The combination of existing orthogonal techniques such as imaging with new technology such as the Polaris system will enable confident cross-validation of exciting new single-cell discoveries.

Table 1. Summary of validation methods

	Validation method	Pros	Cons	Sample throughput	Number of parameters
GENOMIC	RNA flow	Fast; easily ID co-expression	Limited probes and parameters	High	Low (1–3)
	In situ hybridization (ISH)	ID novel genes in spatial context	1 probe, limited co-expression	Low	Low (1–2)
	Fluorescent in situ hybridization (FISH)	ID co-expression in spatial context	Limited probes and parameters	Low	Low (1–4)
	Target manipulation (CRISPR, siRNA, overexpression)	Causal information, functional information	Off-target effects, probe design, time-consuming	Low	Low to variable
PROTEIN	Immunohistochemistry (IHC)	ID novel target in spatial context	1 probe, limited co-expression	Low	Low (1)
	Immunofluorescence imaging	ID co-expression; can be in spatial context	Antibodies, limited parameters	Low to medium	Low (1–4)
	Flow cytometry	Fast; easily ID co-expression	Antibodies, limited parameters	High	Medium (4–10)
	Helios™	High-parameter; validates dozens of targets at the single-cell level	Antibody availability	High	High (10–50)
	Western blot	ID co-expression	Limited parameters, bulk only	Low	Low (1–3)
	Mass spectrometry	Unbiased validation of dozens of targets	Bulk only	Low	High (thousands)
OTHER	Biological assay	Definitive, functional verification of hypothesis	Difficult, non-standardized; may require development	Variable	Variable

Table 2. Key validation methods used in publications

Validation methods	Reference
Flow cytometry, western blot, overexpression assay, chromatin immune-precipitation	Behrens et al. 2014
Fluorescent imaging	Darmanis et al. 2015
Flow cytometry, functional assay	Mahata et al. 2014
Fluorescent imaging, mass spectrometry, siRNA	Romanov et al. 2015
FISH, IHC, functional assay, comparison to databases	Zeisel et al. 2015

References

1 Behrens, A.N., Zierold, C., Shi, X. et al. “Sox7 is regulated by ETV2 during cardiovascular development.” *Stem Cells and Development* 23(17) 2014: 2,004–2,013.

2 Darmanis, S., Sloan, S.A., Zhang, Y. et al. “A survey of human brain transcriptome diversity at the single cell level.” *Proceedings of the National Academy of Sciences of the United States of America* 112(23) (2015): 7,285–7,290.

3 Mahata, B., Zhang, X., Kolodziejczyk, A.A. et al. “Single-cell RNA sequencing reveals T helper cells synthesizing steroids de novo to contribute to immune homeostasis.” *Cell Reports* 7(4) (2014): 1,130–1,142.

4 Romanov, R.A., Alpár, A., Zhang, M.D. et al. “A secretagogin locus of the mammalian hypothalamus controls stress hormone release.” *The EMBO Journal* 34(1) (2015): 36–54.

5 Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S. et al. “Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq.” *Science* 347(6226) (2015): 1,138–1,142.

CORPORATE HEADQUARTERS
7000 Shoreline Court, Suite 100
South San Francisco, CA 94080 USA
Toll-free: 866 359 4354 in the US
Fax: 650 871 7152
fluidigm.com

SALES
North America | +1 650 266 6170 | info-us@fluidigm.com
Europe/EMEA | +33 1 60 92 42 40 | info-europe@fluidigm.com
China (excluding Hong Kong) | +86 21 3255 8368 | info-china@fluidigm.com
Japan | +81 3 3662 2150 | info-japan@fluidigm.com
All other Asian countries | +1 650 266 6000 | info-asia@fluidigm.com
Latin America | +1 650 266 6000 | info-latinamerica@fluidigm.com