Received: 12 May 2020 | Revised: 18 May 2020 | Accepted: 19 May 2020

DOI: 10.1002/dvg.23383

genesis WILEY

REVIEW

A cell type annotation Jamboree—Revival of a communal science forum

Leonid Peshkin | Marc W. Kirschner

Systems Biology, Harvard Medical School, Boston, Massachusetts, USA

Correspondence
Leonid Peshkin, Systems Biology, Harvard Medical School, Boston 02115, MA.
Email: peshkin@gmail.com

Funding information
National Institute of Health (NIH), Grant/Award Number: 5R01HD073104; Howard Hughes Medical Institute

Summary

Cell Atlases are currently being constructed for human tissues as well as several model organisms. New technologies make creation of vast datasets in many species possible, but the value of such data crucially depends on the quality of annotation. The tools of annotating single cell data and creating knowledge representations comparable across organisms have been lagging. We argue that successfully creating Cell Atlases will require a revival of a boot-camp style forum for communal annotation combined with an intensive learning workshop, dubbed a "Jamboree". We report on our experience of successfully developing a structure and curriculum and running such a Jamboree for *Xenopus* Embryonic Cell Types at the Janelia Farms campus of the Howard Hughes Medical Institute.

KEYWORDS

annotation Jamboree, cell atlas, single cell transcriptomics, Xenopus embryology

One important approach to the study of the function of genes responsible for human diseases is to take advantages of experimental opportunities in model organisms. Human disease genes are being rapidly identified, thanks to a revolution in human personal genomics. But functional studies lag behind these identifications. Such functional studies are best done in those model organisms that allow efficient validation and analysis of underlying developmental, cellular and molecular mechanisms. Such studies can lead to predictive disease models to test therapeutic options. Genes do not function in isolation—they are grouped spatially and temporally at multiple nested levels, the most salient functional unit being the cell. Observing biological systems at the cellular level provides unprecedented opportunities to define functional modularity and combinatorial interactions of genes in various physiological contexts. Cell types are highly conserved and thus many of these contexts are conserved in evolution, thus providing a baseline for normal function; deviations will produce new insights into malformations and disease. The Human Cell Atlas, which is already well underway (Regev et al., 2017), is clearly at the center of this emerging single-cell perspective. Yet the best efforts restricted to humans will be severely limited by ethical, technical and practical considerations, even when it focuses only natural cell states in normal individuals. The study of disease and developmental defects will be out of reach for the vast majority of scenarios in humans and therefore must be done in model organisms where we can take advantage of the special tools developed in these organisms. To do this we first have to establish a baseline—characterizing cell type and dynamics of cell state transitions in normal embryonic development and healthy adult. Indeed, Cell Atlases are currently being constructed in several model organisms, most notably mouse and zebrafish. Another important model is *Xenopus*, which has been of monumental utility in studying early development, developmental signaling, and many important stages of organogenesis an oogenesis. It has historically been the most important organism in connecting developmental processes to cell biology and in investigating developmental lineages, cell cycle biology, cell motility, and morphogenesis. Today, nearly every institute in the NIH funds *Xenopus*-based projects, all of which would benefit from a deeply annotated Cell Atlas. The impact of the Cell Atlas in many ways would be similar to the impact of the complete genome, by organizing information at multiple scales of abstraction. Building such resources in an organism like *Xenopus*, fly or fish requires bringing communities together for a collaborative effort of a very special nature, which we had a chance to run as a pilot effort in 2018.

In February 2018, 26 senior developmental biologists (Figure 1) attended a week-long Jamboree at the Janelia laboratory of the Howard Hughes Medical Institute to examine an unpublished and not completely finished very large data set of single-cell transcriptomics in *Xenopus* embryos. This data set has since been published (Briggs et al., 2018; Weinreb, 2018). The importance of this step in biology was highlighted (along with a companion piece on Zebrafish) as Breakthroughs of the Year in all of science for 2018 by Science Magazine (Harland, 2018; Pennisi, 2018). The Janelia Jamobree included people, representing 26 institutions, eight countries, and most importantly expertise in very different areas of embryology. In many ways the rationale for this Jamboree echoed back to that for the first genome Jamborees. Since no single person could know enough to curate the entire genome of a metazoan organism, cooperation, rather than competition, was the only practical path. Similarly, no single person could be expected to characterize the complex developmental lineages and the high-dimensional gene expression patterns of the early *Xenopus* embryo. In both cases, the data would ultimately be essential to everyone working in the field and in both cases this effort would simply be a draft to be further edited and improved in the years ahead.

Genome sequencing was one of the early applications in molecular biology that relied on computer-based visualization tools. But while the genome sequence is linear, the tens of thousands of cells in the embryo form very high-dimensional data sets that cannot be simply interrogated in tabular form. Every gene has a potential relationship with every other gene and this relationship varies with cell type in the embryo and stage of development. Furthermore, developmental biologists as a community have been relatively isolated from a need to use high-dimensional data sets; that is until the advent of a very powerful computer-based analytical tool, called SPRING (Weinreb, Wolock, & Klein, 2018), developed for interactive visualization (Figure 2) of gene expression across stages and across cell types. Using this tool, collections of cells could be singled out, fractionated into groups for further interrogation, and these data could be compared with in situ information and the previous expression information. It is hard to overemphasize a big gap between the computational infrastructure

FIGURE 1 Participants and organizers of the Janelia Jamboree, left to right: Virginia Savova (Instructor, Sanofi); Colin J.H. Brenan (Industry Demo, OneCell, United States); Caleb Weinreb (Instructor, HMS, United States); Alex Lukyanov (Instructor, HMS, United States); Sergei Sokol (MSSM, United States); Roberto Vignali (U of Pisa, Italy); Frank Conlon (UNC, United States); Jose G Abreu (Federal Univ of Rio de Janeiro, Brazil); Leonid Peshkin (Organizer, HMS, United States); Xi He (BWH, United States); Soeren Lienkamp (DFG, Germany); Chris Kintner (Salk, United States); Simon van Heeringen (Radboud University, Nitherlands); Aaron Zorn (CCHMC, United States); Eva Hörmanseder (Helmholtz Institute, Germany); Richard M. Harland (Berkeley, United States); Andre Brandli (Univ of Muenchen, Germany); Enrique Amaya (Univ of Manchester, UK); Ken Cho (UC Irvine, United States); Anna Philpott (Cambridge, UK); Daniel L Weeks (Univ Iowa, United States); Jacques Robert (Univ of Rochester, United States); Nicolas Pollet (CNRS, France); Douglas W Houston (Univ Iowa, United States); Sally Moody (GWU, United States); Gerhard Schlosser (National University of Ireland); Anne-Helene Monsoro-Burq (Inst Curie, France); Mike Kopczynski (Industry Demo, OneCell); Chenbei Chang (Univ of Alabama, United States); Peter Vize (Univ Calgary, Canada); Aldo Ciau-Uitz (Oxford, United Kingdom); Martin Blum (Univ Hohenheim, Germany); Marc Kirschner (Organizer, HMS, United States)

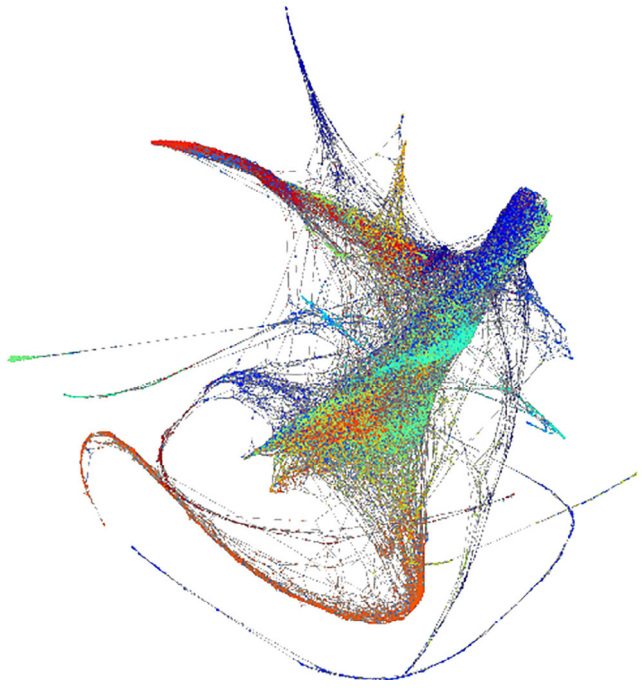

FIGURE 2 A sample representation of *Xenopus* embryonic cells embedded in gene expression space via SPRING tool

needed to organize the data and the deep biological understanding necessary to integrate it into a developmental context. The computer scientists and the biologists recognized this and approached things with optimism and humility. Once the biologists learned to "drive" this interrogation tool (please refer to the recorded tutorials [Peshkin & Kirschner, 2018]), the developmental biologists confirmed experimental information, including fate maps, lineages (Figure 3), as well as already identified marker genes. This activity could be interactively performed with different assumptions or with different subsets of genes. This approach confirmed many previously discovered developmental trajectories and importantly discovered many new ones. Many of these findings, which appeared frequently in the various embryonic lineages, were reported as a "show and tell" lecture to the entire group (Figure 1), followed by intensive community discussion of the genes and the lineages.

The genome Jamborees of the 1990s, and the spate of genome sequencing efforts that followed, reflect an abrupt transition in science. Gone were the days that data of a certain sort would be the property of any one laboratory. It was widely understood that science would progress faster if the sequences were released quickly, and the analytical tools disseminated so that anyone could make use of them. In the case of the Janelia Jamboree, massive high-dimensional data

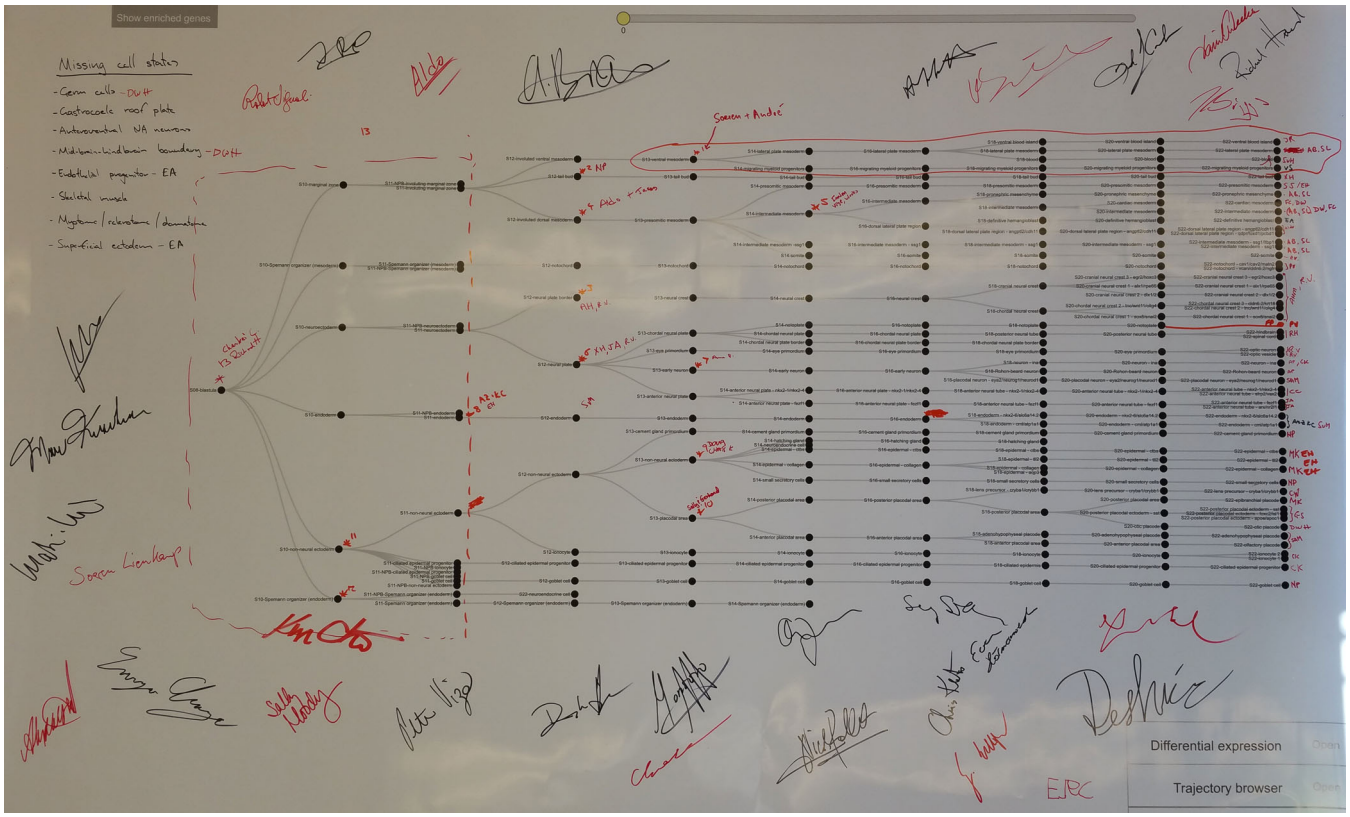

FIGURE 3 All participants autographed a giant poster which served as a dash-board, showing *Xenopus* early embryogenesis as hierarchically linked cell clusters across 10 developmental stages. It was used during Jamboree to distribute tissues to the respective expert sub-groups

sets would be relatively useless without complex computational tools to visualize them, and would still be nearly useless if biologists could not learn how to manipulate the tools on their own. Even with that, science would progress more slowly if there was not ample opportunity for peer discussion and the development of community spirit.

There was palpable elation at a rediscoveries that emerged hourly when a new aspect some well appreciated gem of science, such as the lineage that forms the placodes in the head or the precursors to lineages in the skin was confirmed by such an orthogonal method to fate mapping or transplantation. But even more exciting still was the new putative cell sub-types, marked by genes that the attendees expert in their respective tissues had no knowledge of; or the revelation of some early unappreciated ordering of events of differentiation, such as the ordered appearance of structural components in ciliated cells, well before they were previously detectable by standard methods. Those were good feelings. But perhaps just as special, was the joy at unfettered collaboration, even when it meant spending 18-hr days. The re-forming of groups around questions, the teaching of techniques that were presented earlier but not fully grasped, the larger questions that were raised and perhaps someday pursued, such as ways to try to find the underlying variability of cells within a tissue or the significance of cells embodying more than one lineage—all of this made the grueling effort exhilarating. This common experience inevitably leads to questions of how to maintain the excitement and collegiality over the long term, and how to extend it to more than the 26 people who attended the Jamboree. We think everyone felt as Anna Philpott (Cambridge University, United Kingdom) stated, "I do not remember ever working so hard around the clock for a week, having such an intellectually stimulating time, and such an enjoyable one as well". It was an all-at-once glimpse into a new high-tech future of embryology and at the same time a reversion to an older time when people worked with each other (often for summers at marine labs) with their own hands, and where publication was not an individual life-and-death knightly combat with an inchoate dark force, but merely a way of summarizing for posterity what had already been shared with colleagues. We believe that is what it will take to create a meaningful Cell Atlas in a model organism. It is a task beyond the individual participant but which promises to enrich science for everyone and into the future.

Once created, a Cell Atlas will enhance the value of unique methods already available in that model organism. Moreover, a Cell Atlas will be critical in complementing other emerging approaches: genome editing and proteomics. In *Xenopus*, for example, another spectacular recent single-cell study uncovered regeneration-organizing cells (Aztekin et al., 2019). Genome-edited *Xenopus* mutants can be characterized in development and in adult function at the single-cell level by contrasting to a baseline atlas both the expression profile of a particular cell type of interest and a change in populational proportions among various cell types, where rare or hard-to-detect cell types could be amplified and characterized. In addition, proteomics methods have been undergoing rapid development, particularly in *Xenopus*. Though lagging far behind single-cell transcriptomics, single-cell proteomics has already been demonstrated (Lombard-Banek, Moody, Manzini, & Nemes, 2019), ahead of other organisms, specifically in Xenopus development during which cells are large and protein levels are tantalizing close to what could be seen by mass spectrometry. The principles of multiplexing and barcoding, instrumentation and signal amplification are very different between proteomics and transcriptomics. Yet, the data resulting from single-cell proteomics will face many of the same challenges of interpretation. Moreover, combining what we learn from single-cell transcriptomics with both the deep bulk-level proteomics available today and single-cell proteomics of tomorrow will offer a whole new promise of a biological insight. The success of the 2018 single-cell transcriptomics Jamboree in creating new resources, training researchers in high-tech tools and building a community of collaborators encourages us to propose the same style approach in creating Cell Atlases for model organisms generally and continuing the tradition in Xenopus, particularly since in October 2019 the Xenopus Resources and Emerging Technologies conference, which brought together United States-based PIs, unanimously identified (Wallingford, Horb, Zorn, Peshkin, & Khokha, 2020) Xenopus Cell Atlas as next "essential resource." We hope in the not too distant future the Xenopus advantages in experimental embryology and its large protein complement, will allow us to go beyond the genome to the proteome and to the discoveries that await us at the level of post-translational modifications. As these happen, we will remember the "Spirit of Janelia" and leverage the energy of a community that can provide new insights for human biology.

ACKNOWLEDGEMENTS

We are grateful to Janelia Farm Research Campus of the Howard Hughes Medical Institute, for funding, hosting and supporting the Jamboree. During the preparation of the workshop LP and MK were funded by 5R01HD073104 award from the NIH.

ORCID

Leonid Peshkin https://orcid.org/0000-0002-6420-848X

REFERENCES

Aztekin, C., Hiscock, T. W., Marioni, J. C., Gurdon, J. B., Simons, B. D., & Jullien, J. (2019). Identification of a regeneration-organizing cell in the Xenopus tail. *Science*, *364*, 653–658. https://doi.org/10.1126/science.aav9996

Briggs, J. A., Weinreb, C., Wagner, D. E., Megason, S., Peshkin, L., Kirschner, M. W., & Klein, A. M. (2018). The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science*, *360*, eaar5780. https://doi.org/10.1126/science.aar5780

Harland, R. M. (2018). A new view of embryo development and regeneration. *Science*, *360*, 967–968. https://doi.org/10.1126/science.aat8413

Lombard-Banek, C., Moody, S. A., Manzini, M. C., & Nemes, P. (2019). Microsampling capillary electrophoresis mass spectrometry enables single-cell proteomics in complex tissues: Developing cell clones in live *Xenopus laevis* and zebrafish embryos. *Analytical Chemistry*, *91*, 4797–4805. https://doi.org/10.1021/acs.analchem.9b00345

Pennisi, E. (2018). Development cell by cell. *Science*, *362*, 1344–1345. https://doi.org/10.1126/science.362.6421.1344

Peshkin, L., Kirschner, M. (2018). The videos of tutorials recorded during Janelia farms Xenopus single cell jamboree: A lecture on single-cell transcriptomics in Xenopus Embryo (J Briggs) https://youtu.be/g1gZG223N0I;

An overview of the single-cell field in light of embryology projects (A Klein) https://youtu.be/J75H5KqGqLY; A tutorial on using SPRING layout tool in the context of Xenopus embryology (Caleb Weinreb) https://youtu.be/S8nf56EvkK8.

Regev, A., Teichmann, S. A., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., … Yosef, N. (2017). Human cell atlas meeting participants. The Human Cell Atlas. *Elife*, *6*. https://doi.org/10.7554/eLife.27041

Wallingford, J., Horb, M., Zorn, A., Peshkin, L., Khokha, M. (2020). *Xenopus community white paper*. Retrieved from https://www.xenbase.org/entry/doNewsRead.do?id=699.

Weinreb, C. (2018). *Xenopus single cell browser*. Retrieved from http://tinyurl.com/scXen2018.

Weinreb, C., Wolock, S., & Klein, A. M. (2018). SPRING: A kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics*, *34*, 1246–1248. https://doi.org/10.1093/bioinformatics/btx792

How to cite this article: Peshkin L, Kirschner MW. A cell type annotation Jamboree—Revival of a communal science forum. *genesis*. 2020;e23383. https://doi.org/10.1002/dvg.23383