February 21, 2023

To Laura Zahn, Cell Genomics:

We are pleased to submit the enclosed manuscript, “Combinatorial expression motifs in signaling pathways” for your consideration as a Research Article in Cell Genomics, with co-submission to Cell Systems. I previously discussed this work with you during your recent visit to Caltech.

In multicellular organisms, a handful of core pathways operate across an astonishing variety of contexts, tissues, and organs. These pathways include signaling systems such as TGF-β, Notch, and Wnt, sets of splicing regulators such as SRSF, and others. In each case, members of a family of protein variants are expressed in different combinations in different cell states. Currently, biologists tend to focus on specific components expressed in processes of interest or in whole-transcriptome analyses that can obscure the contribution of smaller sets of genes. However, the organizational principles that govern which pathway components are expressed by which cell types have remained unclear. To address this issue, we took advantage of public scRNA-seq atlas datasets to systematically analyze expression profiles for pathways across all cell types.

Remarkably, we find that most cell types utilize a limited set of pathway expression profiles, dramatically smaller than the theoretically possible number of component combinations. Many of these profiles recur in distantly related cell types. Conversely, similar cell types often express distinct profiles for the same pathway. We term these profiles *pathway expression motifs.* We identify the strongest motifs across four key pathways and show how they vary across cell types and during developmental trajectories. We also identify a smaller number of non-motif “private” profiles that are limited to sets of closely related cell types. Together, these results support a mosaic view of genome expression, in which each cell type effectively selects a different motif from a common “menu” for each of its major pathways.

More specifically, we show:

* **Pathway expression motifs can be identified in scRNAseq cell atlas data.** We integrate multiple single-cell mouse atlas data sets and introduce an algorithm that identifies pathway expression motifs. We show that motifs are prevalent and highly combinatorial, with seemingly equivalent components (e.g. receptor variants) exhibiting overlapping but distinct profiles of expression across cell types. Critically, the motif structure reduces the potential space of signaling pathway component combinations, while allowing a single pathway to operate in multiple functionally distinct “modes.”
* **TGF-β, Notch, SRSF, and Wnt exhibit strong motif signatures.** We show how each of these four central pathways exhibits qualitatively similar features in their motif organization. We also identify motifs in several other signaling and non-signaling pathways, suggesting they are a general (although not universal) feature of mammalian pathway architecture.
* **Pathway profiles appear and change independently and dynamically.** The expression profiles of different pathways are largely uncorrelated across cell types. Developmental scRNAseq datasets further indicate that pathways switch motifs in an independent and uncorrelated manner through changes in multiple pathway components. These results provide insight into the ways in which cells dynamically regulate pathway motif expression during development.

In addition to identifying a general feature of cellular pathway expression, these results provide a timely demonstration of how scRNAseq cell atlas data sets can be used to identify unexpected features of cellular regulation. They should therefore be of interest to researchers in systems biology, genomics, single cell biology, and quantitative biology.

As reviewers, we would suggest:

* Arjun Raj (University of Pennsylvania) - quantitative, single cell biology.
* Olga Troyanskaya (Princeton University and Flatiron Institute) - genomics, single-cell biology, and computational methods.
* Uri Alon (Weizmann Institute) - systems biology and the identification of motifs in biological systems.
* Naama Barkai (Weizmann Institute) - systems biology, including protein level motifs
* Cole Trapnell (University of Washington) - single cell biology, reconstruction of developmental pseudo-temporal trajectories.
* Jay Shendure (University of Washington) - single cell biology and genomics.
* Alex Schier (Biozentrum) - developmental biology, including core signaling pathways.

Thank you for your consideration.

Sincerely,



Michael Elowitz