Caleb Lareau

BST 290 Advanced Computational Biology
Prepared 24 October 2016

Secondary Summary–Roadmap

To compliment the Roadmap study, I presented a Cell paper by John Stam and associates that introduced the hematopoietic differentiation scheme. While the goal of Roadmap and other large consortia (e.g. ENCODE) is to collect a variety of data across several different cell lines, this study focused on collecting 49 results from DNA hypersensitivity assays, which access the regions of the genome associated with open chromatin. A couple of useful bits of biology include: CD34 positive T cells are associated with non-committed cells and most of the cells used were primary from donors. The study compares the open chromatin states of these various cells at different stages of hematopoiesis or that have committed differentially. From Figure 1, we see that regions of open chromatin allow for a very nice clustering of these samples. Notably, transcriptomic analyses provide decent clusters but often aren't as clean as those that are derived from chromatin states.

One point I made concerned Figure 1D where the "hourglass" model associated with differentiation is supported by their DHS data. This hypothesis suggests that the progenitor stages of development (in other words, not pluripotent or committed" have more evolutionary conservation, as suggested by embryos looking similar across organisms. In these data, Stam and colleagues show that the DHS regions accessible in progenitor cells contain sequences that are more conserved evolutionarily. Figure 2 uses external molecular data to support that the DHS regions are identifying enhancer regions.

The heart of where the paper gets interesting is Figure 3/4, that suggest that DHS sites are lost and support this idea of closing chromatin reduces pluripotency. Additionally, as the cell commits, new DHS regions open up, which help the cell 'lock in' the committed phenotype. Figure 5 begins to functionalize what pattern these lost DHS regions have as many auto regulatory factors (i.e. factors that contain their own motif in their promoter). These auto regulatory circuits can occur primarily or with 2 or 3 associated co-factors in the network chain. In class, we discussed how potentially these DHSs that are lost may not contain many loops (evidenced by the capture Hi-C paper losing few loops going form CD34 positive to GM12878) whereas the DHSs gained in commitment may contain many more loops (again evidenced by the capture Hi-C paper). Moreover, as we discussed in the Roadmap paper, there may be room too to examine the intersection of DHS, methylation, and loops and create an epigenetic paradigm that helps explain cellular differentiation.

Finally, the paper wraps up with an analysis of DHS from cancer cells. When looking in a reduced dimensional space, these cancer cells localize to their pluripotent cells rather than the primary cell types that they are associated with. Part of this suggests that the open chromatin landscape in cancers may resemble the stem cells, which may explain some of the regained proliferative properties of the cancers. This potential epigenetic regulation is a tantalizing mechanism that may help provide insights into the cancer and stem cell biology. However, as these experiments were performed in bulk cell populations, inferences are necessarily limited. Advances in scATAC data will allow for a finer-scale examination of the variable chromatin

landscape, which may include accessing tumor purities for variation in the chromatin landscape. The hematopoietic system will provide a useful model system to examine epigenetic variation in determining cellular phenotypes, including cancer. Further, experimental designs that access variable open chromatin may help elucidate the phenotype-specific characteristics associated with these cell states.