Epigenetic variability in cellular identity and gene regulation

We are studying the relationship between epigenetic regulators, chromatin structure and DNA sequence and how these factors influence gene expression patterns. We recently proposed an integrative computational pipeline called HAYSTACK (https://github.com/lucapinello/Haystack). HAYSTACK is a software tool to study epigenetic variability, cross-cell-type plasticity of chromatin states and transcription factor motifs and provides mechanistic insights into chromatin structure, cellular identity and gene regulation. By integrating sequence information, histone modification and gene expression data measured across multiple cell-lines, it is possible to identify the most epigenetically variable regions of the genome, to find cell-type specific regulators, and to predict cell-type specific chromatin patterns that are important in normal development and differentiation or potentially involved in diseases such as cancer.

Computational methods for genome editing

Recent genome editing technologies such as CRISPR/Cas9 are revolutionizing functional genomics. However computational methods to analyze and extract biological insights from data generated with these powerful assays are still in an early stage and without standards. We have embraced this revolution by developing cutting-edge computational tools to quantify and visualize the outcome of CRISPR/Cas9 experiments. We created a novel computational tool called CRISPResso2 (http://github.com/pinellolab/CRISPResso2), an integrated software pipeline for the analysis and visualization of CRISPR-Cas9 and base editor outcomes from deep sequencing experiments, as well as a user-friendly web application that can be used by non-bioinformaticians (http://crispresso.rocks). In collaboration with Daniel Bauer’s and Stuart Orkin’s groups, we recently applied CRISPResso and other computational strategies to aid the development of an in situ saturation mutagenesis approach for dissecting enhancer functionality in the blood system with the aim of developing potential therapeutic genome editing applications for hemoglobin disorders.

Exploring single cell gene expression variation in development and cancer

Cancer often starts from mutations occurring in a single cell that results in a
heterogeneous cell population. Although traditional gene expression assays have provided important insights into the transcriptional programs of cancer cells, they often measure a combined signal from a mixed population of cells and hence do not provide adequate information regarding subpopulations of malignant cells. Emerging single cell assays now offer exciting opportunities to isolate and study individual cells in heterogeneous cancer tissues, allowing us to investigate how genes transform one subpopulation into another. Characterizing stochastic variation at the single cell level is crucial to understanding how healthy cells use variation to modulate their gene expression programs, and how these patterns of variation are disrupted in cancer cells. We have developed a method called STREAM to model the variability of gene expression at single cell resolution, and to reconstruct developmental trajectories [see illustrative image] using data from single cell assays such as scRNA-seq, multiplexed qPCR or sc-ATAC-seq. This method can be used for disentangling complex cellular types and states in development, cancer, differentiation or in perturbation studies.