

Luca Pinello, PhD



The focus of **the Pinello laboratory** is to use innovative computational approaches and cutting-edge experimental assays, such as genome editing and single cell sequencing, to systematically analyze sources of genetic and epigenetic variation and gene expression variability that underlie human traits and diseases. The lab uses machine learning, data mining and high performance computing technologies, for instance parallel computing and cloud-oriented architectures, to solve computationally challenging and Big Data problems associated with next generation sequencing data analysis. Our mission is to use computational strategies to further our understanding of disease etiology and to provide a foundation for the development of new drugs and novel targeted treatments.

Pinello Laboratory

Tommaso Andreani*

Huidong Chen, PhD

Kendell Clement, PhD

Jonathan Hsu
shared with Keith Joung lab

Luca Pinello, PhD

Qin Qian, PhD
shared with Dave Langenau lab

Micheal Vinyard
shared with Gad Getz lab

Qiuming Yao, PhD
shared with Daniel Bauer lab

* Visiting PhD student

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.

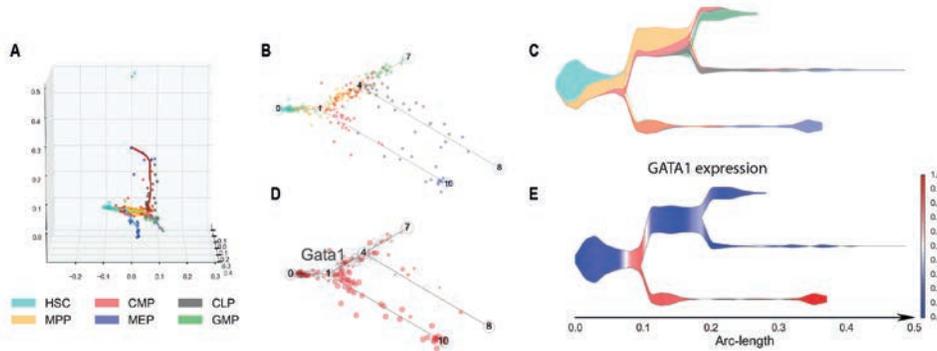
Computational methods for genome editing

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



STREAM on transcriptomic data from the mouse hematopoietic system. A) Dimensionality reduction, reconstructed hierarchical structure composed of curves approximating the inferred trajectories. Single cells are represented as circles and colored according to the FACS sorting labels. B) Flat tree representation at single cell resolution; branches are represented as straight lines, (cells are represented as in A). The length of the branches and the distances between cells and assigned branches are proportional to the original representation in the 3D space. C) Rainbow plot: intuitive visualization to show cell type distribution and density along different branches. D) Single cell resolution expression pattern of GATA1, each circle is red filled proportionally to the relative expression of GATA1 in the whole population. E) Relative expression of GATA1 in each branch using the representation in C.

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.

Selected Publications:

Chen H, Albergante L, Hsu JY, Lareau CA, Lo Bosco G, Guan J, Zhou S, Gorban AN, Bauer DE, Aryee MJ, Langenau DM, Zinovyev A, Buenrostro JD, Yuan GC[†], **Pinello L.** Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. *Nat Commun.* 2019 Apr 23;10(1):1903.

Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, Cole MA, Liu DR, Joung JK, Bauer DE[†], **Pinello L.** CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol.* 2019 Feb 26.

Hsu JY, Fulco CP, Cole MA, Canver MC, Pellin D, Sher F, Farouni R, Clement K, Guo JA, Biasco L, Orkin SH, Engreitz JM, Lander ES, Joung JK, Bauer DE, **Pinello L.** CRISPR-SURF: discovering regulatory elements by deconvolution of CRISPR tiling screen data. *Nat Methods.* 2018 Dec;15(12):992-993.

Canver MC*, Haeussler M*, Bauer DE, Orkin SH, Sanjana NE, Shalem O, Yuan GC, Zhang F, Concordet JP, **Pinello L.** Integrated design, execution, and analysis of arrayed and pooled CRISPR genome-editing experiments. *Nat Protocols.* 2018 May;13(5):946-986.

Pinello L*, Farouni R*, Yuan GC[†]. Haystack: systematic analysis of the variation of epigenetic states and cell-type specific regulatory elements. *Bioinformatics.* 2018 Jan 17.

Canver MC*, Smith EC*, Sher F*, **Pinello L***, Sanjana NE*, Shalem O, Chen DD, Schupp PG, Vinjamur DS, Garcia SP, Luc S, Kurita R, Nakamura Y, Fujiwara Y, Maeda T, Yuan G-C, Zhang F, Orkin SH & Bauer DE. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature.* 2015 Nov 12;527(7577):192-72015 Sep.

*Co-first authors

[†]Co-corresponding authors