Learning relationships between histone modifications in single cells

Jake Yeung[1,2]*, Maria Florescu[1]*, Peter Zeller[1]*,
Buys Anton de Barbanson[1], Max D Wellenstein[1], Alexander van Oudenaarden[1]

[1] Hubrecht Institute, Oncode Institute, University Medical Center Utrecht, Netherlands
[2] Institute of Science and Technology Austria (ISTA), Klosterneuburg Austria

Abstract

Regulation of chromatin states involves the dynamic interplay between different histone modifications to control gene expression. Recent advances have enabled mapping of histone marks in single cells, but most methods are constrained to profile only one histone mark per cell. Here we present an integrated experimental and computational framework, scChIX (single-cell chromatin immunocleavage and unmixing), to map multiple histone marks in single cells. scChIX multiplexes two histone marks together in single cells, then computationally deconvolves the signal using training data from respective histone mark profiles. This framework learns the cell type-specific correlation structure between histone marks, and therefore does not require \textit{a priori} assumptions of their genomic distributions. Using scChIX, we demonstrate multimodal analysis of histone marks in single cells across a range of mark combinations. Modeling dynamics of in vitro macrophage differentiation enables integrated analysis of chromatin velocity. Overall, scChIX unlocks systematic interrogation of the interplay between histone modifications in single cells.

Overview

Mapping histone modifications in single cells is still in its infancy, and has the potential to uncover how the chromatin is read to govern gene expression. Currently, most experimental techniques (e.g. scChIC-seq, CUT&RUN , and CUT&TAG) that map single-cell histone modifications are limited to only one histone modification per single cell.

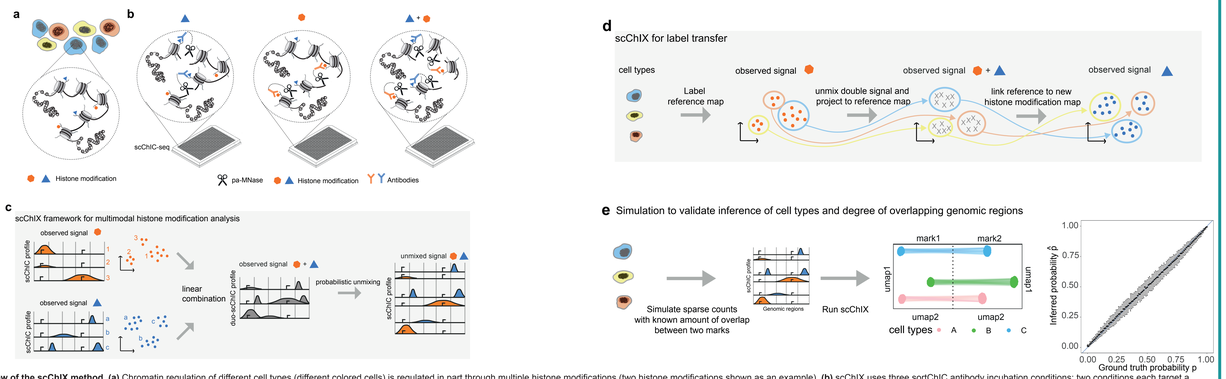
We present an integrated experimental and computational framework for multiplexing histone modifications in single cells. To profile two histone modifications in single cells, we first generate three genome-wide sortChIC datasets: two datasets by incubating cells with one of the two histone modification antibodies separately (single-incubated), and the third by incubating cells with both histone modification antibodies together (double-incubated). We then use our two single-incubated datasets as training data to generate the possible pairs of genome-wide histone modification profiles that, when added together, fit to a single-cell profile from the double-incubated dataset. For each double-incubated cell, we then deconvolve the multiplexed data by probabilistically assigning each fragment back to their respective histone modification.

Methods

Check out the preprint for details:

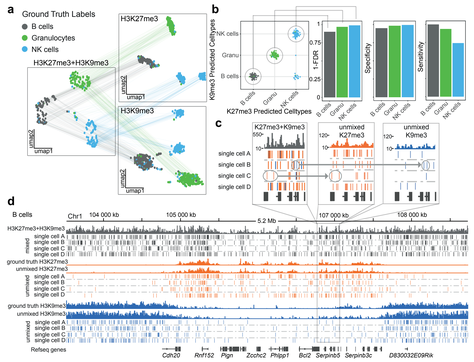
https://www.biorxiv.org/content/10.1101/2021.04.26.440629v1

scChIX combines experimental design with inference to map multiple histone modifications in single cells



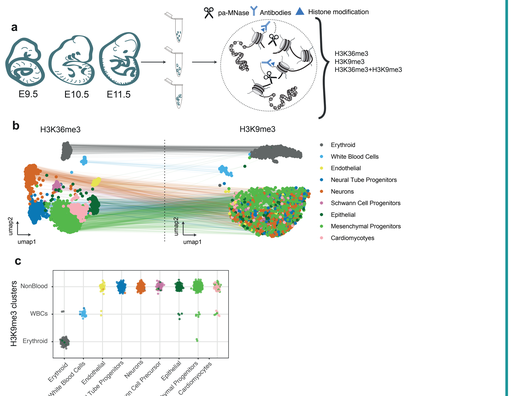
Overview of the scChIX method. (a) Chromatin regulation of different cell types (different colored cells) is regulated in part through multiple histone modifications (two histone modifications shown as an example). **(b)** scChIX uses three sortChIC antibody incubation conditions: two conditions each target a single histone modification (single-incubated) only and the third condition targeting both histone modifications simultaneously (double-incubated). **(c)** Schematic of scChIX for deconvolving multiplexed histone modifications. The two single-incubated sortChIC datasets (one targeting an orange histone modification, the other a blue modification, each modification reveals three clusters) are training data to define the possible pairs of histone modification distributions that can be combined to generate a hypothetical double-incubated cell. For each observed double-incubated cell, we then assign the cell to the most likely pair of cell states, one from each histone modification. We then probabilistically assign each pA-MNase cut into their respective histone modification. **(d)** Label transfer allows joint analysis of two single-incubated sortChIC datasets targeting functionally distinct histone modifications. Information derived from one histone modification, such as cell types, histone mark levels, and pseudotime, can be transferred to another histone modification, using the double-incubated cells as a link. **(e)** Simulation study shows that scChIX can unbiasedly assign reads to each mark regardless of the amount of overlap there is between the two marks across the genome. *p* is the expected fraction of double-incubated reads in a genomic locus that belongs to mark 1. *p-hat* is the estimate of the probability. n=101 simulation datapoints spread evenly between 0 and 1 inclusive. Error bars are 95% confidence intervals, centers are the mean.

scChIX accurately maps multiple histone modifications in single cells



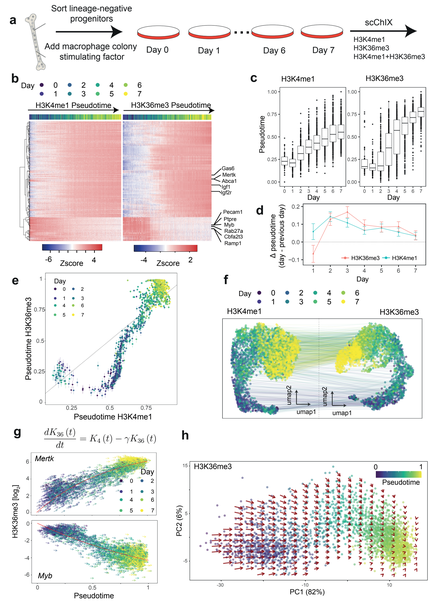
scChIX accurately deconvolves multiplexed histone modifications in single cells. (a) UMAP representation of the H3K27me3 (n=367) and H3K9me3 (n=376) histone modification space derived from the two single-incubated datasets (right two panels), and the H3K27me3+H3K9me3 space (left panel, n=290) derived from the double-incubated data. Cells are colored by their ground truth cell type labels. The cells in the H3K27me3- and H3K9me3-only space have unmixed double-incubated cells whose deconvolved signal has been projected onto their respective UMAPs. Lines connecting across datasets connect where each double-incubated cell is located in each of the three histone modification space. **(b)** Matrix summarizing the cluster pair that scChIX selected for each double-incubated cell. Cells along the diagonal are predicted to be B cells, Granulocytes, and NK cells, respectively. Cells in the off-diagonal are false negatives. Barplots summarizing false discovery rate (FDR), sensitivity, and specificity of assigning each cell type (right). **(c)** Zoom-in coverage plot and single-cell cut fragments in B cells of mixed (H3K27me3+H3K9me3, grey bars), unmixed (H3K27me3 and H3K9me3, orange and blue bars). Positions of cut fragments are shown for four single cells (single cells A, B, C, and D) for H3K27me3+H3K9me3 signal (grey ticks) as well as their unmixed outputs (orange and blue ticks). Circled reads and arrow highlight examples of cut fragments being assigned to either H3\K27me3 (orange) or H3K9me3 (blue). **(d)** Zoom-out of the *Serpinb5* locus. Cut fragments from HK27me3+K9me3 are colored based on whether they have been assigned to H3K27me3 (orange) or H3K9me3 (blue). Ground truth coverage are single-incubated sortChIC data targeting H3K27me3 (orange) and H3K9me3 (blue). **(e)** Heatmap of probabilities *p* of assigning reads to H3K27me3 (p=1, red) or H3K9me3 (p=0, blue) around the *Bcl2* locus. Rows are single cells (ordered by predicted cell type), columns are genomic regions (50 kilobase bins). Transitions between H3K9me3- and H3K27me3-marked chromatin states are independent of cell type. **(f)** Same as (e) but at the *Crim1* locus, where transitions from H3K9me3 to H3K27me3 (blue to red) are cell type-specific.

Distinct cell types can share heterochromatin



Applying scChIX to mouse organogenesis reveals shared heterochromatin landscapes and cell type-specific differences in H3K36me3:H3K9me3 ratios. (a) Schematic of mouse organogenesis to study H3K36me3 and H3K9me3 in single cells. **(b)** Joint UMAP of mouse organogenesis after deconvolution from scChIX (n=2911 H3K36me3 cells, n=2166 H3K9me3 cells). **(c)** Assignment of multiple H3K36me3 cell types to one H3K9me3 cluster. The H3K36me3 (columns) and H3K9me3 (rows) label for each double-incubated (n=1197 cells) is plotted onto a matrix to H3K36me3 cell types to H3K9me3 clusters. Cells are colored by their cell type label as in (b). **(d)** Subclustering of non-blood cells for H3K9me3, represented as a UMAP. Arrow denotes a pseudotime axis. Pseudotime defined as the first principal component of the 2D UMAP. **(e)** Joint UMAP of deconvolved double-incubated cells (n=1197 cells), colored by the log ratio of number of H3K36me3 cuts versus number of H3K9me3 cuts. **(f)** Boxplot of H3K36me3:H3K9me3 ratio across cell types. Number of double-incubated cells for each cell type: n=163 erythroid, n=17 white blood cells, n=24 endothelial, n=136 neural tube progenitors, n=197 neurons, n=46 Schwann cell precursors, n=73 epithelial, n=458 mesenchymal progenitors, n=83 cardiomyocytes. Boxplots show 25th percentile, median and 75th percentile, with the whiskers spanning 97% of the data.

Different histone modifications progress at different rates during differentiation



Applying scChIX to two active marks reveals chromatin velocity during in vitro macrophage differentiation. (a) Schematic of mouse macrophage in vitro differentiation time course experiment to study H3K4me1 and H3K36me3 in single cells. **(b)** Heatmap of histone modification signal on the bodies of dynamic genes over pseudotime. Rows are gene bodies and columns are single-incubated cells ordered along pseudotime. Color labels of columns are days from which the cells were recovered during the time course. **(c)** Boxplots of pseudotime estimates of single-incubated cells along the time course. Number of cells per day for H3K4me1: n=58 day 0, n=148 day 1, n=249 day 2, n=350 day 3, n=369 day 4, n=383 day 5, n=488 day 6, n=519 day 7. For H3K36me3: n=42 day 0, n=125 day 1, n=176 day 2, n=301 day 3, n=384 day 4, n=366 day 5, n=522 day 6, n=567 day 7. Boxplots show 25th percentile, median and 75th percentile, with the whiskers spanning 97\% of the data. **(d)** Estimate of the average difference of pseudotime from one day to the next. Error bars indicate 95\% confidence intervals, calculated by a linear model of the pseudotime differences between days. Statistics derived from number of cells indicated in (c). **(e)** Estimates of two different pseudotimes from a single cell. Error bars are 95\% confidence intervals of the estimates. Each point is a double-incubated cell. **(f)** Joint UMAP of H3K4me1 and H3K36me3 from scChIX, lines connect single cells with multimodal information. **(g)** Chromatin velocity estimates of an up-regulated gene (above) and a downregulated gene (below). Red curve is the exponential relaxation fit according to the first-order differentiation equation. **(h)** High-dimensional chromatin velocities of dynamic genes projected onto the principal components one and two. Vector field estimated by smoothing across nearest neighbors of cells.

Take home messages

scChIX maps multiple histone modifications in single cells

Multimodal analysis of histone modifications unlocks new analyses to discover new regulatory principles

Time-series experiment with scChIX reveals distinct temporal dynamics and chromatin velocity