Development of single-cell SPRITE to comprehensively map dynamic organization of DNA in higher-order nuclear structures within single cells

Mary V. Arrastia[1], Joanna Jachowicz[2], Noah Ollikainen[2], Matthew S. Curtis[1], Charlotte A. Lai[2], Sofia Quinodoz[2], David A. Selck[1], Mitchell Guttman[2], Rustem F. Ismagilov[1,2]

1. *Division of Chemistry and Chemical Engineering, California Institute of Technology* 2. *Division of Biology and Biological Engineering, California Institute of Technology*

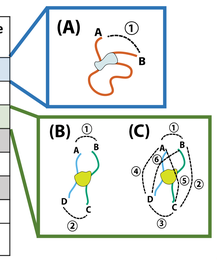
Abstract

An important factor in the control of gene regulation is the 3-dimensional organization of the nucleus, which is dynamically assembled and regulated in different cellular states. Yet, how this nuclear organization is established and how it changes dynamically across single cells is largely unknown. To enable the study of higher-order nuclear organization of thousands of single cells, we have developed a method called scSPRITE (single-cell split-pool recognition of interactions by tag extension) that leverages split-and-pool barcoding of individual cells. Using this method, we have generated deep single-cell maps of approximately 4500 single cells. The scSPRITE heatmap comprising an ensemble of single cells portrays similar features of chromosomal organization when compared against the bulk SPRITE heatmaps (Quinodoz et al. (2018) *Cell*) These same chromosomal features are present when we compare as few as 10 single cells from scSPRITE against original SPRITE. We have also shown that we can obtain high coverage per cell by observing nearly uniform coverage of the genome, demonstrating the robustness of scSPRITE in maintaining intact single cells throughout the procedure. High single cell accuracy was measured from mouse-human cell mixing experiments, with 97% of cells representing a single species. Furthermore, we have initially explored the heterogeneity of single cells by identifying cells in each stage of the cell cycle as previously described (Nagano et al. (2017) *Nature*). This tool will allow us to better understand the heterogeneity of nuclear structure at the single-cell level.

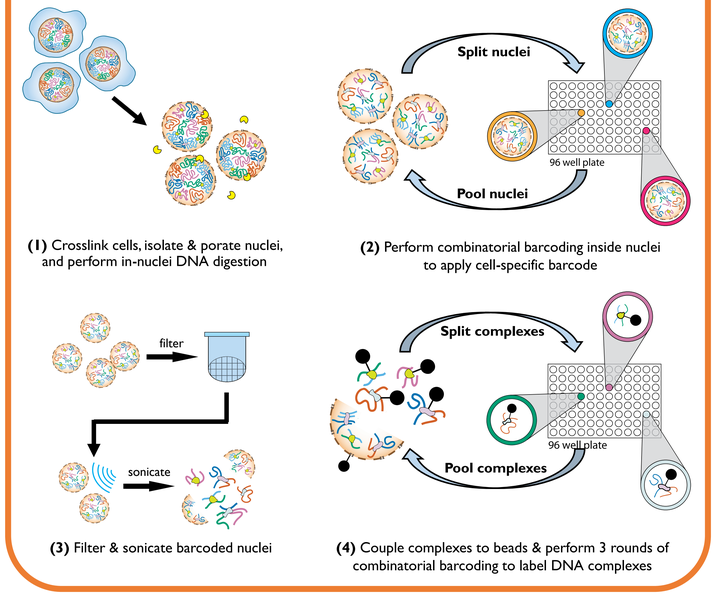
Current single-cell sequencing-based methods miss higher-order nuclear structures

	Single Cell Hi-C	Microscopy	SPRITE
Strengths	•Single-cell resolution •Captures unbiased view of chromosome structure	•Single-cell resolution •Captures higher-order structures	•Captures higher-order structures •Captures unbiased view of chromosome structure
Limitations	•Limited to proximity ligation to view pairwise interactions •Low resolution	•Low throughput •Limited in number of loci to image	•Not at single-cell resolution

Molecules in DNA complex	Maximum # of Pairs by Proximity Ligation	Maximum # of Pairwise Combinations
2	1 **(A)**	1
3	1	3
4	2 **(B)**	6 **(C)**
...	...	...
10	5	45
...	...	...
100	50	4950
n	$\lfloor n/2 \rfloor$	$n(n-1)/2$



scSPRITE methodology



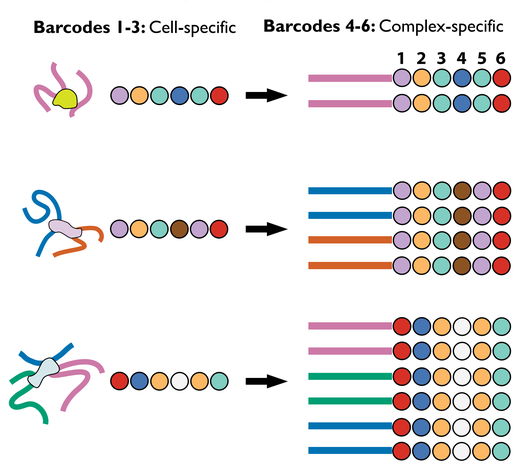
(1) Crosslink cells, isolate & porate nuclei, and perform in-nuclei DNA digestion

(2) Perform combinatorial barcoding inside nuclei to apply cell-specific barcode

(3) Filter & sonicate barcoded nuclei

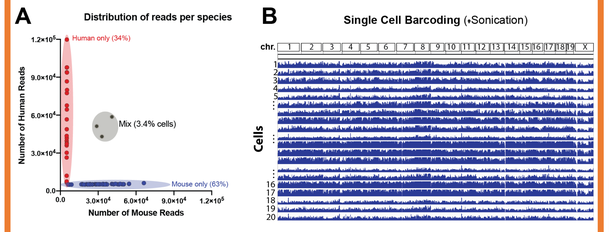
(4) Couple complexes to beads & perform 3 rounds of combinatorial barcoding to label DNA complexes

Six rounds of barcoding preserves cell- and complex-specific information



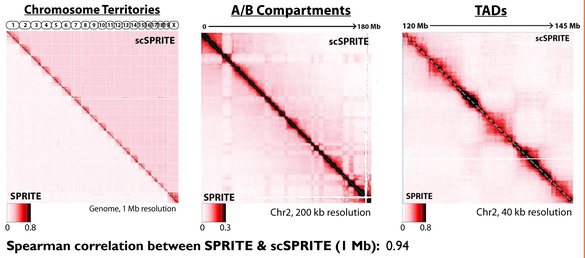
After 6 rounds of combinatorial barcoding, we are able to preserve information concerning cell origin and spatial DNA arrangement in every DNA complex. The first 3 barcodes contain information about which cell the DNA complex originated from. The last 3 barcodes contain the information about which strands of DNA were in close proximity to each other.

High accuracy and coverage of single cells are obtained from scSPRITE



Two of the key features in assessing scSPRITE were high accuracy in identifying single cells and high coverage from each cell. **(A)** To confirm we could identify single cells, we mixed mouse and human cells together and performed scSPRITE as described previously. From this, we can identify ~97% of single cells representing a single species. **(B)** We looked at coverage of our single cell data to determine whether our nuclei remain intact throughout the method. We looked at the genome coverage across 20 single cells. After binning at high (100kb) resolution, we can observe near-uniform coverage across the genome.

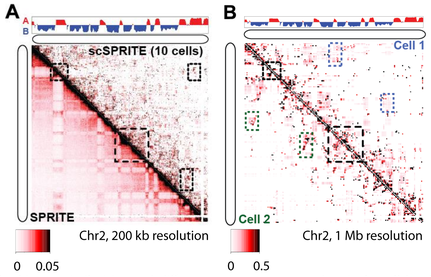
scSPRITE reconstructs known interactions



Spearman correlation between SPRITE & scSPRITE (1 Mb): 0.94

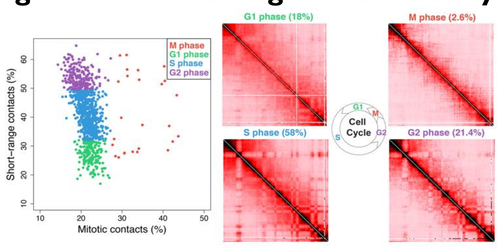
We compared contact maps between our ensemble of cells from scSPRITE (1500 cells) vs bulk SPRITE to assess whether we can observe the similar chromosomal interactions. Visually, we can observe similar chromosome territories, A/B comparments, and TADs in our scSPRITE data when compared to SPRITE. We measured the Spearman correlation between both datasets at 1 Mb resolution, yielding a high coefficient value of 0.94.

Maps containing few cells share similar features



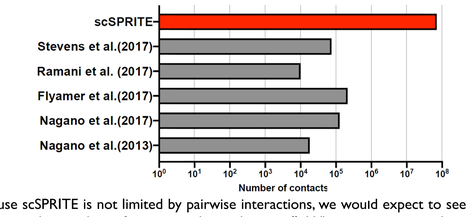
To begin looking at similarities and differences in contacts at the single cell level, we began exploring genome features containing few cells. **(A)** We generated contact maps containing as few as 10 cells pooled together and compared with bulk SPRITE. Even with 10 cells, we can successfully recreate chromosomal contacts. **(B)** When comparing two individual single cells side-by-side, we can observe regions that share similarities in structure (black dashed box) while also observing heterogeneity in contacts (blue and green dashed boxes).

scSPRITE data reveals heterogeneity in single cells across stages of the cell cycle



To explore biological heterogeneity in our single cell data, we analyzed cell cycle progression using the methods previously published. We were able to isolate cells that matched the conditions corresponding to each cell cycle phase. Furthermore, we pooled the cells respective to each stage of the cell cycle to construct heatmaps, showing the condensation and expansion of chromosomes thorughout the cell cycle.

Median # of contacts per cell from scSPRITE exceeds that of scHiC



Because scSPRITE is not limited by pairwise interactions, we would expect to see increased counts in the number of contacts obtained per cell. When we compare the number of contacts per cell in scSPRITE compared with the previous scHiC datasets, we generate at least 100 times more contacts per cell compared with these previous methods. More contacts is useful as it lays the foundation to observe higher-order complexes in single cells.

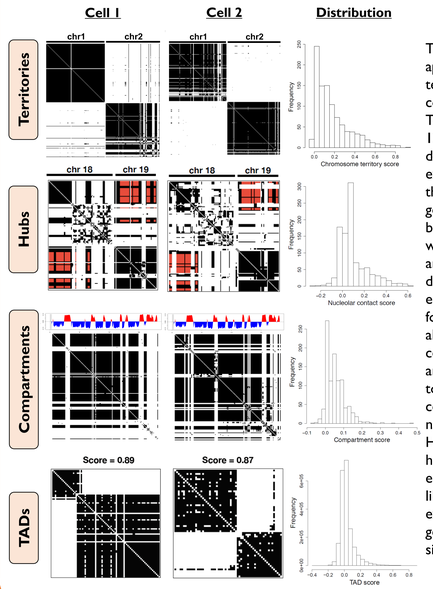
Calculating contact score for single cells

$$Contact\ Score = \left(\frac{\#\ of\ X\ contacts\ observed}{Total\ \#\ X\ contacts} - \frac{\#\ of\ non{-}X\ contacts\ observed}{Total\ \#\ of\ non{-}X\ contacts}\right)$$

X = the specific structure being studied (i.e. chromosome territories, compartments, TADs, etc.)

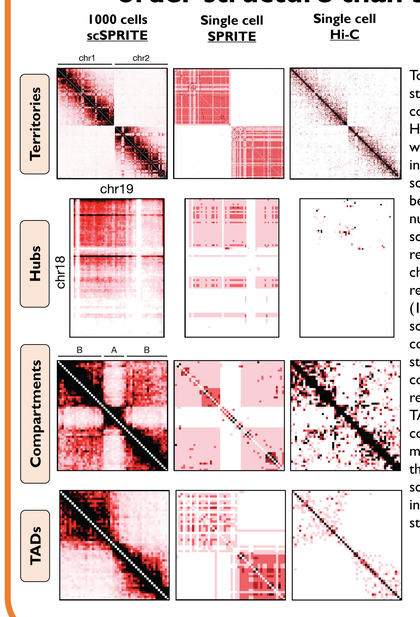
We defined a contact score metric that allows us to identify genomic structures in our single-cell dataset. For each structure, there exists a known number of contacts within a given region at a specific resolution (Total # X contacts). We count the number of contacts in that same region for each single cell (Total # X contacts observed) and take the ratio of those two values to determine the fraction of specific contacts. We perform the same calculation for any non-specific contacts that might exist in that same region for the same cell. These two ratios are then subtracted, giving us a value ranging from -1 to 1.

scSPRITE shows top-down view of genome structure



The contact score was applied to identify territories, hubs, A/B compartments, and TADs within the top 1000 single cells in our dataset. We identified examples of single cells that demonstrate each genomic feature, and binary heatmaps of cells with the highest scores are displayed. A distribution of scores for each genomic structure for the top 1000 cells is also shown. A fraction of cells are centered around zero, either due to low coverage of those cells or a high ratio of non-specific contacts. However, there are many high-scoring cells for each genomic feature, likely indicating the existence of each genomic structure in single cells.

scSPRITE reveals more about higher-order structure than scHiC



To see how our single cell structures from scSPRITE compared against single-cell HiC. For all the comparison, we selected the most informative single cell from scHiC and compared it to the best cell conveying each nuclear structure from scSPRITE. For lower resolution structures such as chromosome territories (1Mb res) and A/B compartments (1Mb res), both scHiC and scSPRITE show similar coverage. Long-range structures like nucleolar hub contacts (1Mb res) and high-resolution structures like TADs (40kb res), scHiC contacts are more sparse, making it difficult to reveal these structures clearly. scSPRITE contains more information about these structures.

Acknowledgements

This work was funded by the National Institutes of Health (NIH) as part of the NIH Common Fund's 4D Nucleome Program (grant number 5U01HL130007-02) and the National Science Foundation Graduate Research Fellowship Program. We would like to additionally thank Elizabeth Detmar, Elizabeth Soehalim, Chris Chen, and Vickie Trinh for their respective contributions to this work. We would also like to thank Igor Antoshechkin for his assistance in performing sequencing at the Millard and Muriel Jacobs Genetics and Genomics Laboratory and Fao Gao with bioinformatics assistance through the Caltech Bioinformatics Resource Center.