Identifying Cellular Heterogeneity from Optical Reconstruction of Chromatin Architecture

Noah G. Burget[1,3,4], Sora Yoon[2,3,4], Yeqiao Zhou[1,3,4], Atishay Jay[2,3,4], Yiting Chen[1,3,4], Yimin Sheng[1,3,4], Golnaz Vahedi[2,3,4], and R. Babak Faryabi[1,3,4]

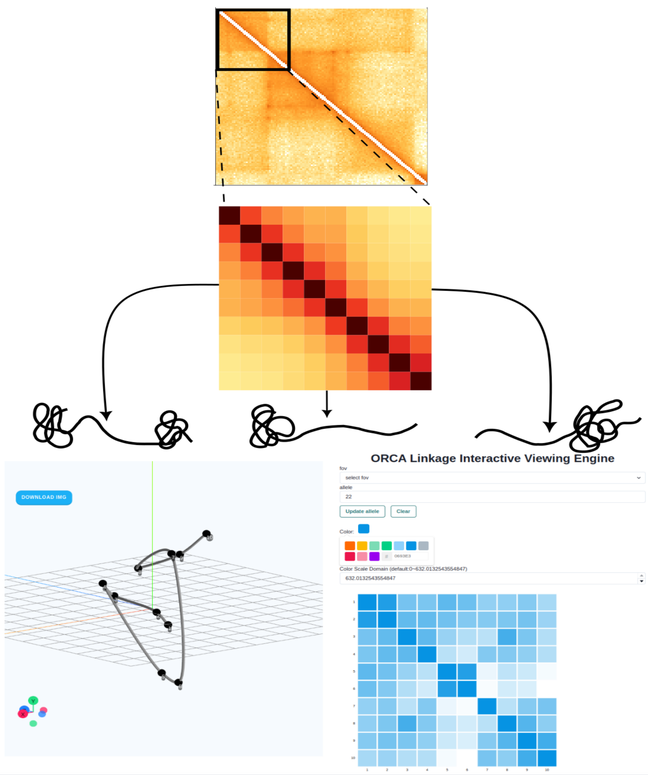
1. Department of Pathology and Laboratory Medicine, 2. Department of Genetics, 3. Institute for Immunology and Immune Health, 4. Epigenetics Institute, University of Pennsylvania, Philadelphia, PA, USA.

FARYABI LAB

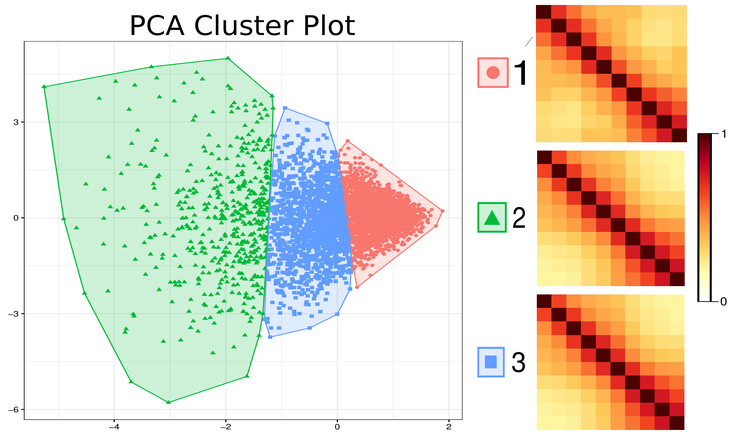
Summary

While sequencing-based chromatin conformation capture (3C) assays measure genome folding at population-level, Optical Reconstruction of Chromatin Architecture (ORCA) allows for examination of folding at individual alleles. ORCA data can be used to study distances between loci on a chromatin fiber as well as interaction frequency at single alleles. We aim to use this method to elucidate heterogeneity of B cell lymphoma cells. To address this question we compared the performance of several unsupervised machine learning algorithms in identifying subpopulations of B cell lymphoma cells with distinct MYC locus folding. Our studies showed that non-negative matrix factorization (NMF) followed by K-Mean clustering outperforms our approaches in detecting cellular heterogeneity from ORCA data. This analysis shows the existence of three distinct populations of B cell lymphomas with different MYC loci folding. Close examination of the largest subpopulation showed that its average MYC locus folding resembles that of 3C experiments. Together, these data suggest that population-based assays fail to capture the heterogeneity of genome folding and NMF can be used to extract this information from ORCA studies.

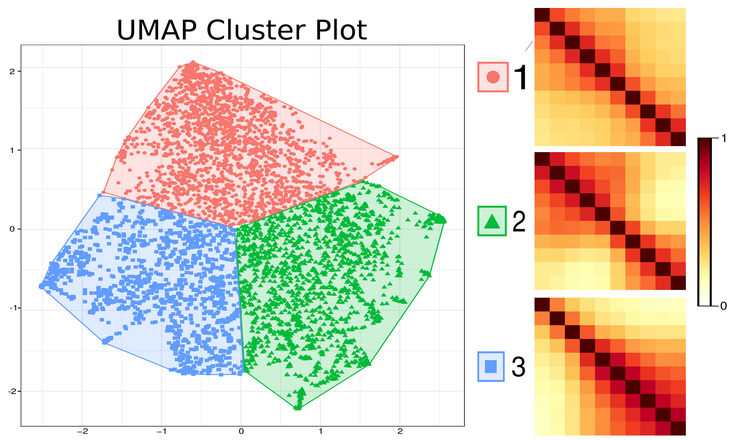
Background



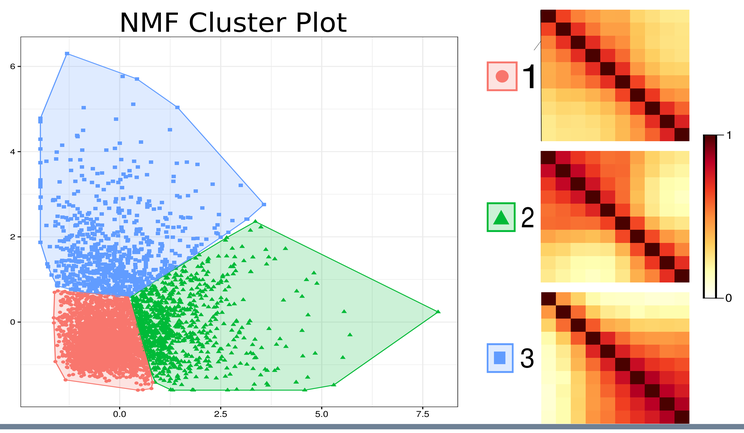
Method #1



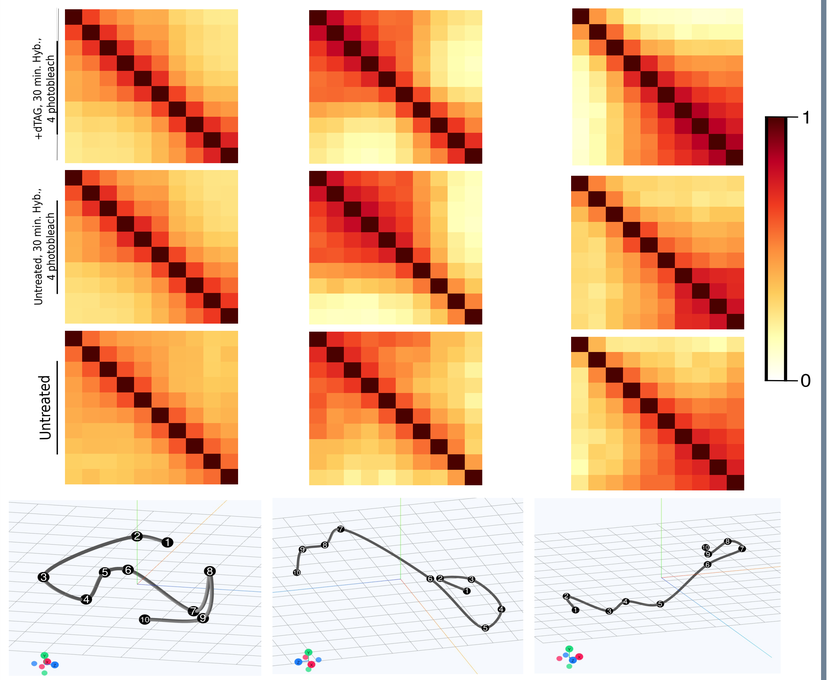
Method #2



Method #3



NMF reveals distinct MYC locus folding



Heterogeneity independent of condition