Genome-wide profiling of nuclease protected domains reveals physical properties of chromatin

George Spracklin, Pierre-Olivier Estève, and Sriharsa Pradhan
New England Biolabs, Inc. Ipswich, Ma 01938 USA

Abstract

In metazoan cell nuclei, chromatin is functionally divided into transcriptionally active (euchromatin) or inactive (heterochromatin) regions. These heterochromatin regions constitute large chromatin domains that are in close contact with the nuclear lamina. Such lamina-associated domains (LADs) are thought to organize chromosomes inside the nucleus and are enriched for repressive histone modifications. Genome-wide profiling of heterochromatin, especially LADs, is often challenging and warrants a simpler and direct method. Here we developed a new method, Protect-seq, aimed at identifying regions of heterochromatin via resistance to nuclease degradation followed by next-generation sequencing. We performed Protect-seq on the human colon cancer cell line HCT-116 and observed overlap with previously curated LADs. We provide evidence that these protected regions are enriched for the repressive histone modification H3K9me3 and to a lesser extent H3K9me2 and H3K27me3. Moreover, the loss of H3K9me3 in human cells leads to an increase in chromatin accessibility. In sum, we demonstrate a novel technique to identify nuclease inaccessible regions of the genome and our data is consistent with the model that repressive chromatin domains are compacted and targeted to the nuclear lamina, likely via HP1 proteins, which act as scaffolds to maintain chromatin architecture.

Protect-seq Methodology

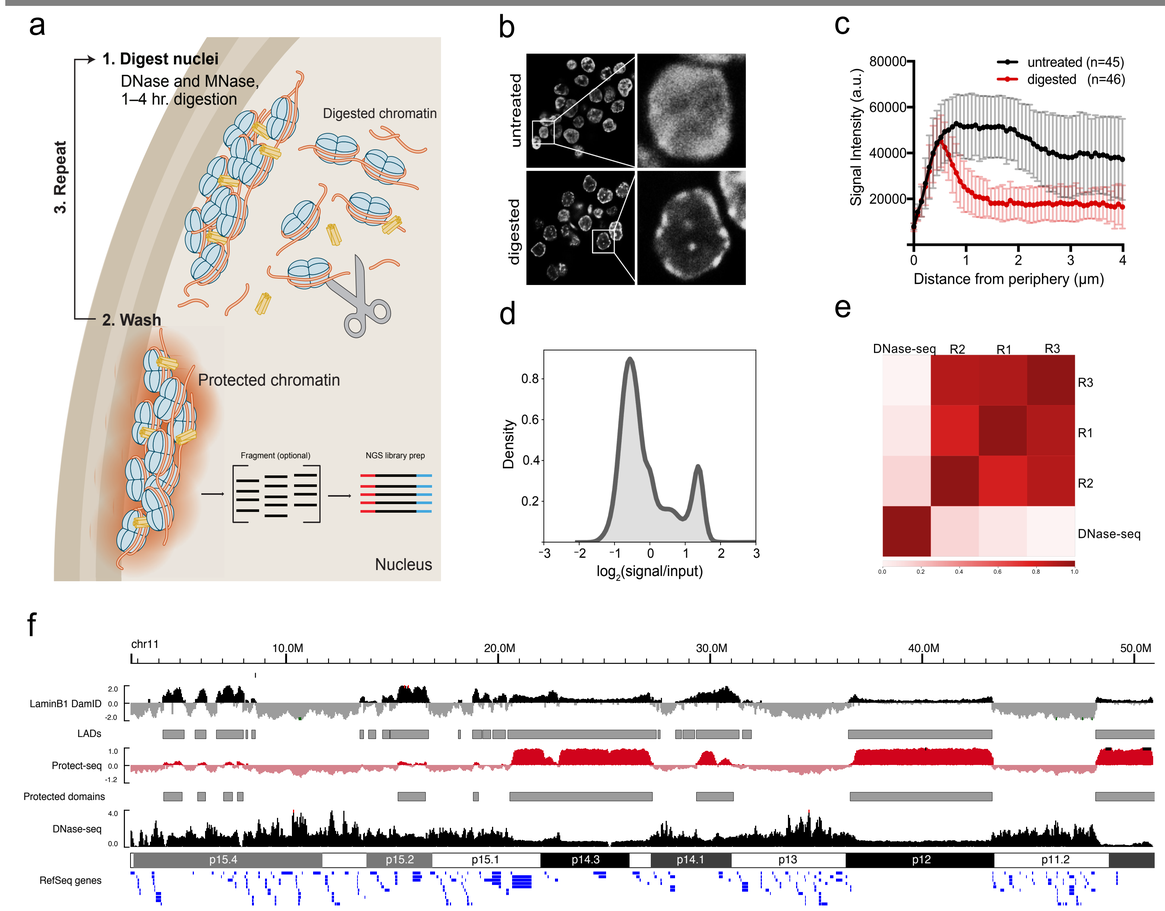

Figure 1 | Protect-seq principle and assay. (a) Schematic of Protect-seq. DNase I and MNase (scissors) degrade accessible chromatin. The remaining 'protected' chromatin is isolated for sequencing. (b) DAPI stain of HCT-116 nuclei with and without nuclease treatment, white boxes represent zoomed image (on right) (c) Spatial quantification of DAPI signal with and without nuclease treatment (d) KDE plot of Protect-seq signal. (e) Pearson correlation of Protect-seq replicates and DNase-seq (ENCSR000ENM)[1] in HCT116. (f) Next-generation sequencing coverage track of DNase-seq (black; accessible chromatin), Protect-seq (red; inaccessible chromatin), and LaminB1 DamID-seq (4DN)[2] (black; LADs). Representative chromosome used (chr11). LaminB1 DamID-seq and Protect-seq data are normalized using Reads Per Genome Coverage (RPGC) and displayed as log2ratio (signal/input).

Domain Calling with Hidden Markov Models

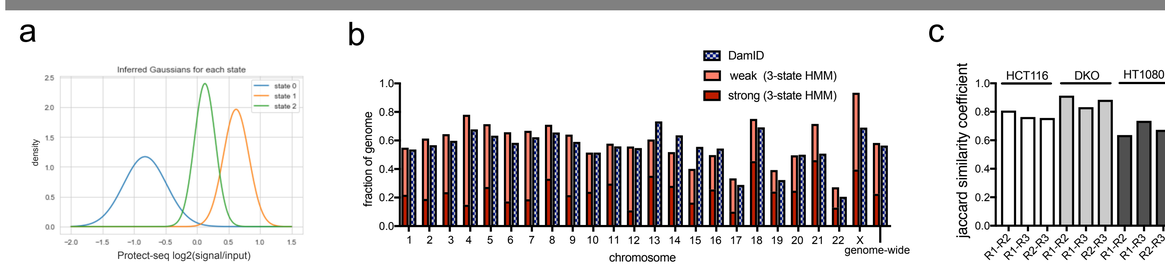

Figure 2 | Using Hidden Markov Models to identify Protect-seq domains. (a) Example of Gaussian distribution of Protect-seq data used in HMM Viterbi calls (3-state HMM depicted). (b) Fraction of each chromosome covered by Protect-seq or LaminB1 DamID-seq domains in HCT116. (c) Jaccard similarity coefficient representing the fraction of domain overlap between replicate pairs.

Comparison with DamID, Repli-seq, and ChIP-seq

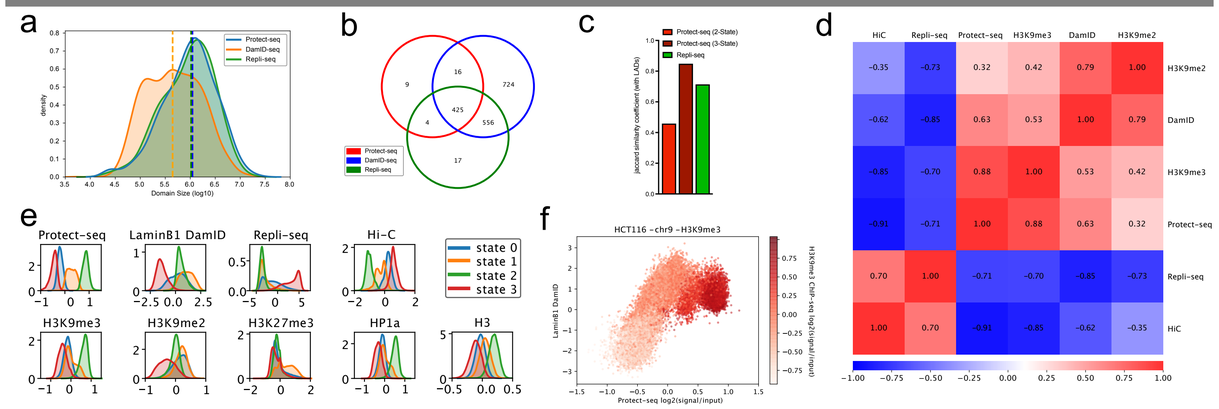

Figure 3 | Protect-seq domains overlap LADs, late-replicating DNA, and repressive histone modifications. (a) KDE plot of Protect-seq, LaminB1 DamID-seq[2], and Repli-seq[2] domain length. Dashed line indicates median domain length. (b) Venn diagram of domain overlap/intersect (c) Jaccard similarity coefficient representing the degree/fraction of domain overlap. (d) Pearson correlation of signal tracks (e) KDE plots of genomic features separated by Protect-seq HMM state. (f) Scatterplot of Protect-seq and LaminB1 DamID-seq (representative chromosomes used). Shading indicates the strength of H3K9me3 signal.

Loss of heterochromatin

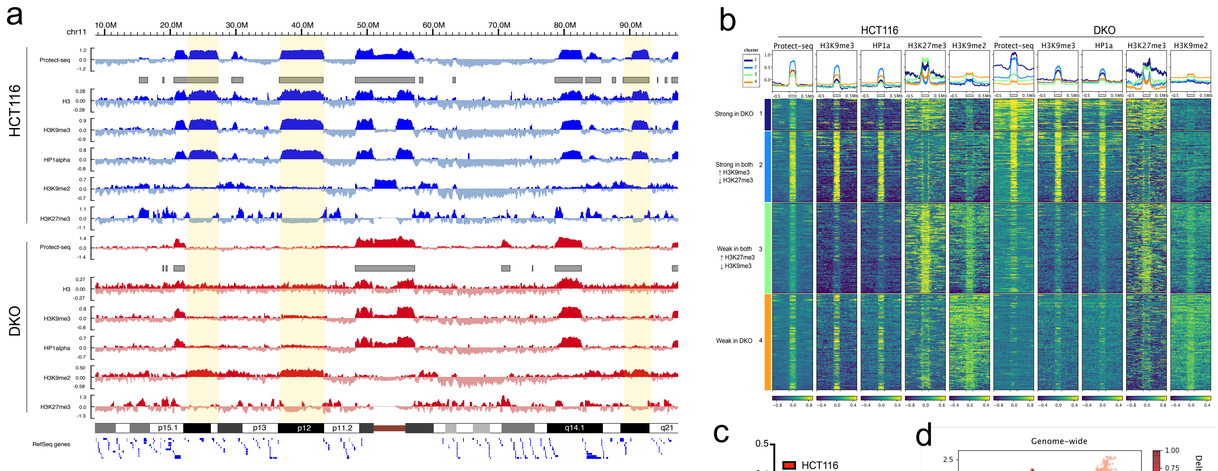

Figure 4 | Protect-seq domains are dependent on H3K9me3 and HP1 heterochromatin domains. (a) Genome browser view of various signal tracks. Grey boxes indicate domain calls using HMMs. (b) Heatmap of k-means clustered Protect-seq domains. Top, meta-profile plots for each cluster. (c) Fraction of genome covered by Protect-seq domains in HCT116 and DKO with light shading representing the fraction lost or gained in DKO cells. (d) Genome-wide scatterplot of Protect-seq signal in HCT116 and DKO. Shading indicates the strength of H3K9me3 signal in DKO cells.

Protect-seq in HT1080

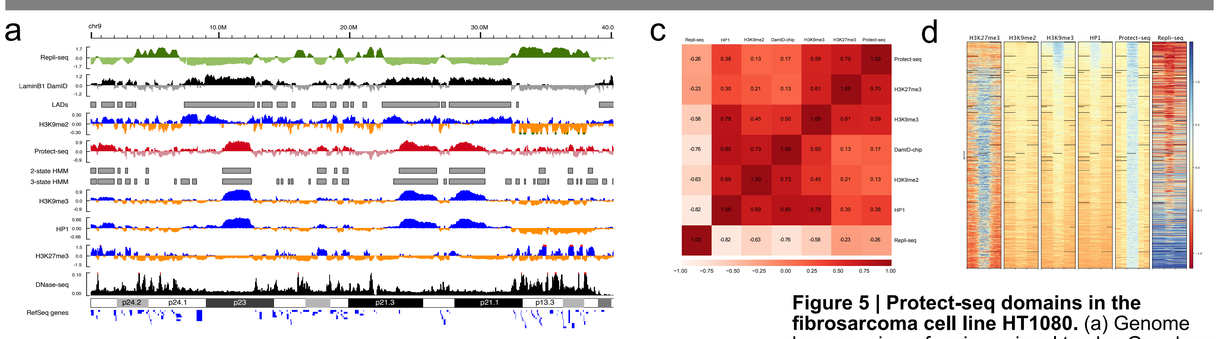

Figure 5 | Protect-seq domains in the fibrosarcoma cell line HT1080. (a) Genome browser view of various signal tracks. Grey boxes indicate domain calls using HMMs. (b) Fraction of chromosome covered by Protect-seq or LaminB1 DamID-seq[3] domain. (c) Pearson Correlation of signal track data (d) Heatmap displaying enriched genomic features within and outside Protected domains. (e) KDE plots of genomic features separated by Protect-seq HMM state.

Conclusions, References, and Acknowledgements

- **Protect-seq is a novel technique to identify nuclease resistant chromatin domains**
- **Protect-seq is a simple, easy-to-use, cost-and-time effective method which does not require actively dividing cells**
- **Protected domains correlate with LADs, late replicating domains, and repressive histone marks**

We thank members of the Mirny lab for helpful discussions. We thank Tilde Carlow, Tom Evans, Rich Roberts, Jim Ellard and Don Comb at NEB for supporting basic research.

References

1. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature, 489(7414), 57–74. doi:10.1038/nature11247
2. Dekker, J., Belmont, A. S., Guttman, M., Leshyk, V. O., Lis, J. T., Lomvardas, S., … 4D Nucleome Network (2017). The 4D nucleome project. Nature, 549(7671), 219–226. doi:10.1038/nature23884
3. Meuleman, W., Peric-Hupkes, D., Kind, J., Beaudry, J. B., Pagie, L., Kellis, M., … van Steensel, B. (2013). Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. Genome research, 23(2), 270–280. doi:10.1101/gr.141028.112