Probing Protein-Directed Chromatin Architecture Using Dovetail™ HiChIP Assays

Technical Note

Product Highlights:

- **Capture ChIP-seq and Hi-C data together in a single library**
- **8 ready-to-use antibodies available with more on the way**
- **Map chromatin interactions at nucleosome level resolution**

Introduction

Three -dimensional (3-D) chromatin conformation, mediated through protein factors, is known to contribute to gene regulation. Chromatin conformation capture (3C) methods are commonly employed to survey higher-order features of genome topology, aiding in the explanation of how distal loci, megabases (Mbp) apart, may be co-regulated. To understand the protein factors contributing to mediating 3-D structure of the genome, ChIP-seq has been an indispensible tool for the genome-wide mapping of DNA-binding factor sites. However, while ChIP-seq enables a whole-genome approach to *de novo* discovery of protein binding sites, researchers remain largely blind to the distal interactions that are facilitated through the protein-protein interaction of two or more DNA-bound factors.

The Dovetail™ HiChIP *MNase* Kit combines the benefits of ChIP-seq with Hi-C, a proximity ligation method that captures long-range interactions using standard Illumina paired-end sequencing, enabling researchers to query protein-directed chromatin conformation mediated by specific proteins of interest. By enriching for a DNA-biding factor of interest, the resulting libraries display topological features of the protein at lower sequencing depths than traditional Hi-C approaches.

Simple three day workflow from sample to sequencing-ready library

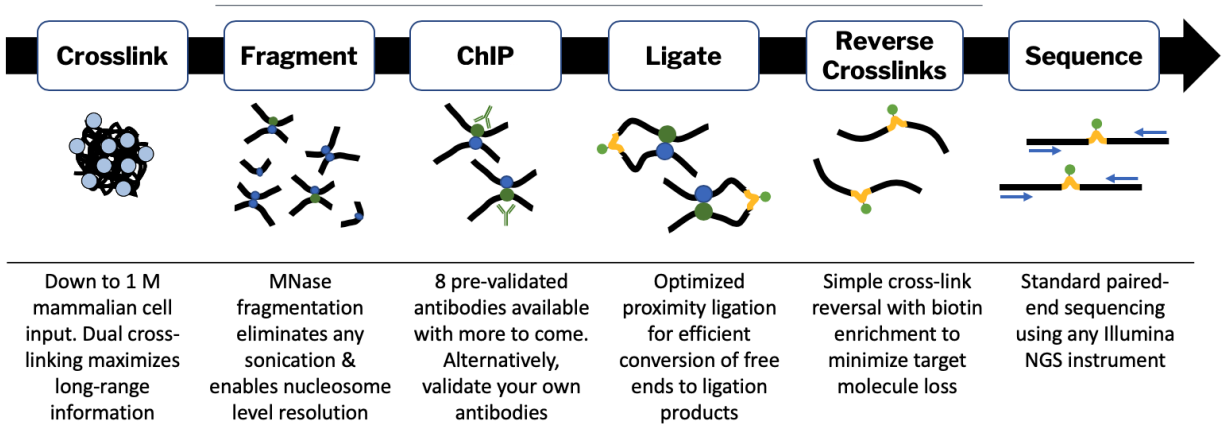
The Dovetail HiChIP workflow (**Figure 1**) offers several advantages over other comparable methods:

- Enzymatic and uniform fragmentation of chromatin
- Enrichment in long-range information from proximity ligation
- Robustness and reproducibility down to one million cell inputs
- Validated list of antibodies for common proteins of interest and guidelines for use

The unique combination of the Dovetail™ Micro-C Proximity Ligation Assay with the Dovetail HiChIP approach enables the use of micrococcal nuclease (MNase) to fragment chromatin uniformly and without sequence bias prior to proximity ligation, eliminating the need for finicky sonication procedures and offering the

Figure 1 – Overview of the Dovetail™ HiChIP *MNase* Kit Three Day Workflow.

A range of 1 M to 10 M cells is used as starting material . Day one consists of cross-linking, fragmentation and addition of ChIP antibody for O/N incubation. Day two includes chromatin immunoprecipition and proximity ligation, during which incorporation of a biotinylated bridge at ligation junctions mark chimeric molecules. Finally, cross-links are reversed, and chimeric molecules are enriched and made ready for NGS library preparation on day three.



maximal resolution (down to mono-nucleosome size) of chromatin interactions. The proximity ligation portion has been optimized for highly efficient conversion of free ends to ligation products, maximizing the capture of long-range information in the final library.

As an example, CTCF-directed HiChIP libraries generated 82% valid read-pairs (**Table 1**). These metrics are in line with high quality Hi-C libaries produced by the Dovetail™ Micro-C Kit. Furthermore, sequencing reads are highly enriched in protein-factor binding sites as evidenced by the signal-to-noise ratio for the HiChIP library.

A common limitation to other HiChIP assays is the necessary input quantity of cells to achieve acceptable reproducibility and signal-to-noise ratios. The Dovetail HiChIP assay requires only one million cells for validated antibody immunoprecipitations (**Table 2**) and is robust and reproducible, evidenced by the high overlap in both chomatin contacts and protein-binding signal for a variety of immunoprecipitations, with inputs ranging from 10 million to 1 million cells (**Figure 2**).

For ease of adoption, Dovetail currently has a validated set of 8 commonly applied antibodies and immunoprecipitation guidelines ready for use with the kit, with plans to continue expanding antibody options (**Table 2**).

Table 1 – Dovetail HiChIP libraries are enriched in proximity ligation informative reads.

Comparison of Dovetail HiChIP MNase (88 M read-pairs) and Dovetail Micro-C (1.7 B read-pairs) library statistics. Dovetail HiChIP against CTCF is enriched in CTCF target sites (Signal-To-Noise ratio is 2.5) and contains long-range information (82% valid read-pairs). Signal-To-Noise = Mean of coverage (top 25th percentile) over target sites / mean of coverage (top 25th percentile) over non-target sites, normalized to IgG HiChIP. Percentages of valid reads is defined as trans + cis > 1 kbp of the total alignments.

Target	CTCF HiChIP	Micro-C
Input Cells	1 M	1 M
Total Read-Pairs	88 M	1.3 B
PCR Duplicates	2%	19%
% *cis* Read-Pairs	60%	68%
% Valid Read-Pairs	82%	94%
Signal-To-Noise	2.5	0.9

Table 2 – Dovetail™ HiChIP validated antibodies.

For ease of use, the Dovetail HiChIP MNase Kit has been validated for use against 8 proteins of interest with additional antibodies expected to validated in the near future. The protocol describes antibody catalog numbers and guidelines for use with sample inputs down to one million cells.

Validated Antibodies	*Future Antibodies*
CTCF	*Sox2*
H3K27ac	*Nanog*
H3K4me3	*Oct4*
H3K27me3	*KLF4*
H3K36me3	*PRC*
H3K4ac	*Suz12*
PolII	*Smc1*
Smc3	

Map protein binding sites and long-range protein-directed interactions

The Dovetail HiChIP Kit enables study of protein-directed 3-D chromatin interactions. Akin to ChIP-seq, the Dovetail HiChIP data captures sequences directly bound by the protein of interest (**Figure 3**). Direct comparison of CTCF ChIP-seq and HiChIP datasets generated using the same antibody confirms that sequences enriched in the ChIP-seq data are similarly represented in the HiChIP data, illustrated by the respective bedgraphs.

However, additional long-range interactions mediated by CTCF factors and bound at different sites are captured and visualized by the arc tracks below in the HiChIP data, highlighting the added value of HiChIP over ChIP-seq, and the use of MNase for chromatin fragmentation provides nucleosome-level resolution. As shown in the zoomed in view of the CTCF binding site, the periodic pattern observed is reflective of nucleosome positioning, with the peak being the center of a nucleosome and the valley the intervening nucleosome linker DNA. Together, this data demonstrates capture protein-mediated chromatin contacts at a nucleosome resolution.

Figure 2 – Robust performance to varying cell input amounts.

Sample input from 10 M to 1 M GM12878 cells were tested with 4 antibodies (CTCF, H3K4me3, H3K27ac, H3K36me3). Comparison of the data confirms near equivalency at all inputs tested as demonstrated by the distribution of reads in the bedgraphs.

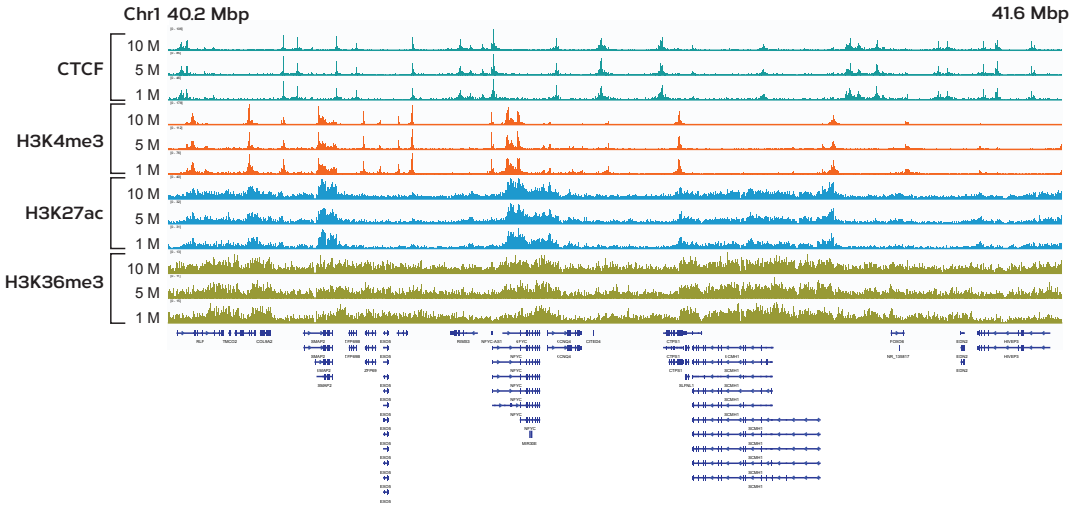
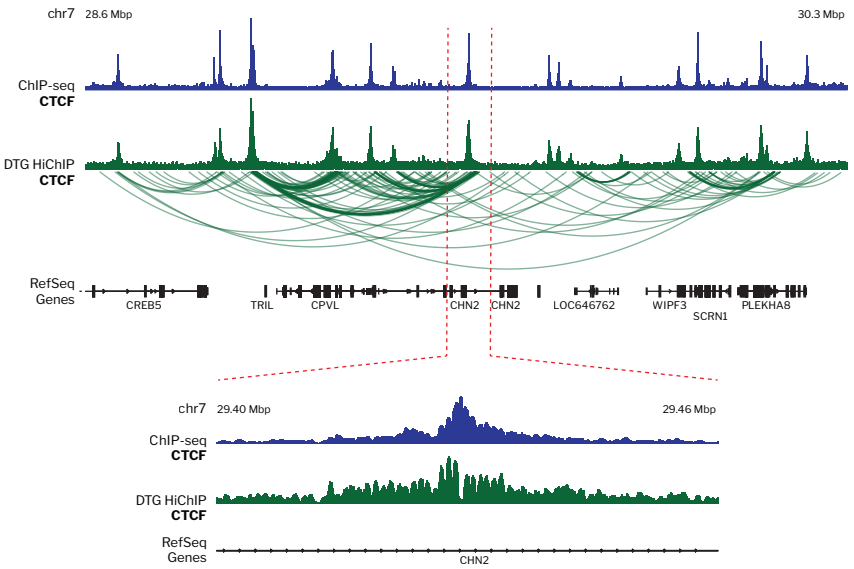


Figure 3 – The Dovetail™ HiChIP Kit combines ChIP-seq data with Hi-C long-range information.

Comparison of HiChIP and ChIP-seq data using the same anti-CTCF antibody. The ChIP-seq data is preserved in the HiChIP dataset with the addition of long-range CTCF interactions as viewed in the arc track. The arcs connecting interacting genomic bins are visualized in the R package "Sushi" where the arc height is dependent on the interaction distance. Only cis contacts greater the 50 kbp in distance with an interaction frequency of greater than one were considered. Interactions extending beyond the plot window are not shown. Use of MNase for fragmentation enables nucleosome resolution in both Dovetail™ HiChIP and Dovetail™ Micro-C Kits. The zoomed in view on the bottom panel displays the nucleosome periodicity with the peaks centered on nucleosomes.



Produce high resolution contact maps with less sequencing

Enrichment of protein-directed chromatin features enables high-resolution contact map generation with less read depth. Compared to a high resolution RE-based Hi-C generated using over 6 billion read-pairs (Rao *et al.*), Dovetail HiChIP data enables visualization of higher-order chromatin features, such as loops and chromatin interactions, at a fraction of the read depth (only 88 million read-pairs for CTCF and 160 million read-pairs for H3K4me3; **Figure 4**).

As expected, CTCF HiChIP data displays interactions between pairs of CTCF factors, visualized by a concentration of reads at the diagonal intersection of the two CTCF binding sites (outlined by the dashed circles). Interestingly and in contrast, H3K4me3, a mark of active transcription, produces a more diffuse pattern suggestive of short-lived interactions emanating from the promoter in the direction of gene transcription.

Further highlighting the power of CTCF enrichment for chromatin loop detection and the associated cost savings, **Figure 5** displays a comparison of contact maps between Dovetail HiChIP and Multi-RE Based Hi-C contact maps centered on a loop feature. Immediately visible in the Dovetail HiChIP library is the enrichment in signal at this CTCF-mediated chromatin loop and depletion elsewhere. This comparsion illustrates the power of the Dovetail HiChIP Kit to increase signal-to-noise ratios in both proximity ligation and protein enrichment, which translates to cost savings for researchers focused on particular proteins of interest.

Figure 4 – High resolution contact matrices highlight contacts observed in CTCF and H3K4me3 HiChIP datasets.

Contact matrices from three individual Dovetail HiChIP experiments with H3K4me3 (160 M read-pairs), CTCF (88 M read-pairs), and IgG as a negative control (72 M read-pairs) in GM12878 cells. CTCF HiChIP data displays interactions between pairs of CTCF factors as demonstrated by a concentration of reads at the diagonal intersection of the two CTCF binding sites (outlined by the dashed circles). In contrast, H3K4me3, a mark of active transcription, produces a more diffuse pattern suggestive of short-lived interactions emanating from the promoter in the direction of gene transcription.

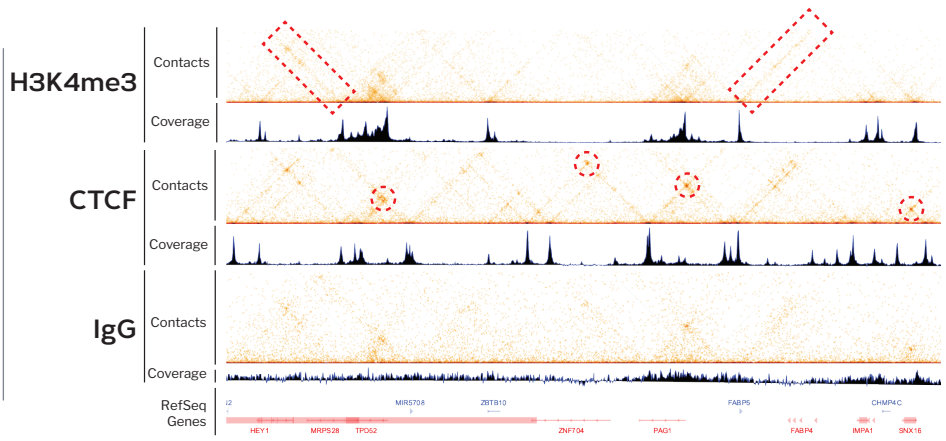
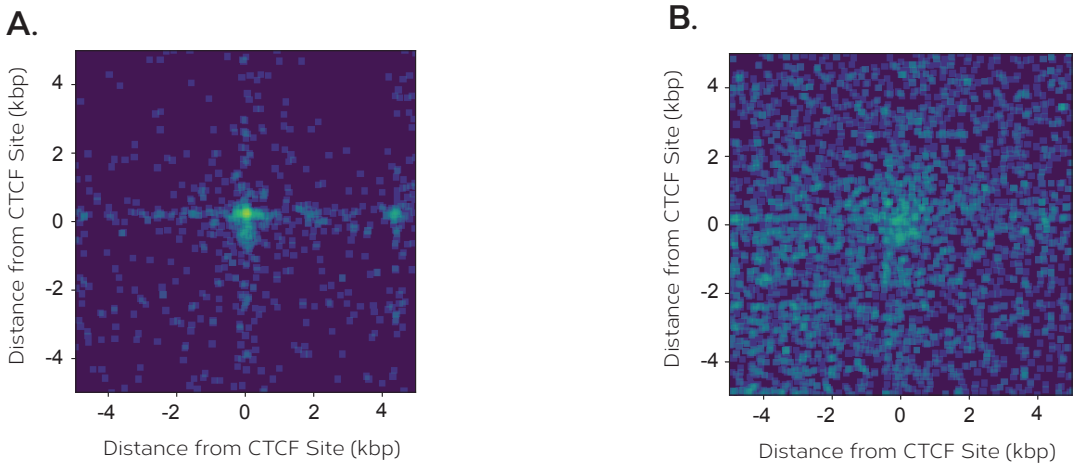


Figure 5 – High signal-to-background chromatin interaction maps enable visualization of higher-order features for a fraction of the sequencing cost.

*Comparison of contact maps centered on Rao et al. loop position between **A)** Dovetail HiChIP MNase Kit data against CTCF and **B)** Multi-Re Hi-C in GM12878 cells. CTCF immunoprecipitation with Dovetail HiChIP increases signal of CTCF-mediate chromatin interactions, such as loops, relative to all background chromatin interactions present in Multi-RE Hi-C. The Dovetail HiChIP anti-CTCF dataset contain 150 M total read pairs while the genom-wide RE-based Hi-C contains 800 M read pairs translating to a ~$2,700 cost differential in sequencing on an Illumina NovaSeq 6000SP - 300 cycle flow cell).*



Summary

In addition to the genome-wide mapping of protein DNA-binding sites (similar to standard ChIP-seq approaches), the Dovetail HiChIP Kit enables the simultaneous capture of long-range protein-directed interactions that involve chromatin architecture and/or gene regulation. The Dovetail HiChIP workflow combines our improved proximity ligation molecular biology with protein enrichment and is structured for ease of use with a set of pre-validated antibodies and without the need for sonication. This new tool enables the discovery and mapping of protein-directed topological features, such as TADs or loops, responsible for positioning transcription factor binding sites and gene promoters within close proximity to each other.

References

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., Lieverman Aiden, E. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. (2014). 159: 1665-1680.