Targeted Hi-C: Uncovering Local Genome Architecture in Regions of Interest

Application Note

1. Introduction

Genome-wide conformational studies using Hi-C proximity ligation approaches can be expensive due to the total sequencing burden required. For example, ~300 million read pairs are required to investigate topologically associating domains (TADs) at a 10kb resolution in human samples. This cost increases when considering experiments involving multiple samples or requiring a higher resolution view. In theory, this sequencing burden, and thus the overall experimental cost, can be reduced in two ways.

First, since the sequence read distribution generated from a Hi-C library is skewed, chemistry improvements that provide more even coverage are expected to reduce the final sequencing requirements. At its core, Hi-C chemistry utilizes restriction endonucleases, either singly or combined, to fragment chromosomal DNA at regular intervals prior to proximity ligation. Due to the non-random distribution in the genome of restriction enzyme cut sites, this results in an over-representation of sequences in the immediate vicinity of and under-representation of sequences more distal to the sites cleaved. As a direct result, sequence distant to the cut sites require greater total reads to meet minimal read depth requirements.

Second, many experiments do not *a priori* require a genome wide view. Target enrichment approaches in use with standard next-generation sequencing (NGS) libraries provide an opportunity to interrogate conformation only at sites of interest thus reducing Hi-C library complexity. In fact, HiChIP, PLACseq, and CaptureC, have all been used to target sites of interest in proximity ligation libraries (Mumbach *et al.*, 2016; Fang *et al.,* 2016; Schoenfelder *et al.*, 2015).

Dovetail Genomics' Omni-C™ Proximity Ligation Kit replaces restriction-site endonucleases with a sequence-independent endonuclease. This less biased approach to Hi-C has been shown to produce significantly more even sequence coverage across the genome and, as a result, reduces the sequencing depth required to reach a desired genomic coverage. Here, we are interested in understanding whether overall sequencing costs can be further reduced by combining targeted enrichment with Omni-C™ chemistry. Our results demonstrate the use of a hybrid capture panel with the Omni-C™ Kit on cell lines and cancerous tissues resulting in Omni-C Proximity Ligation libraries enriched for various genes implicated in cancer.

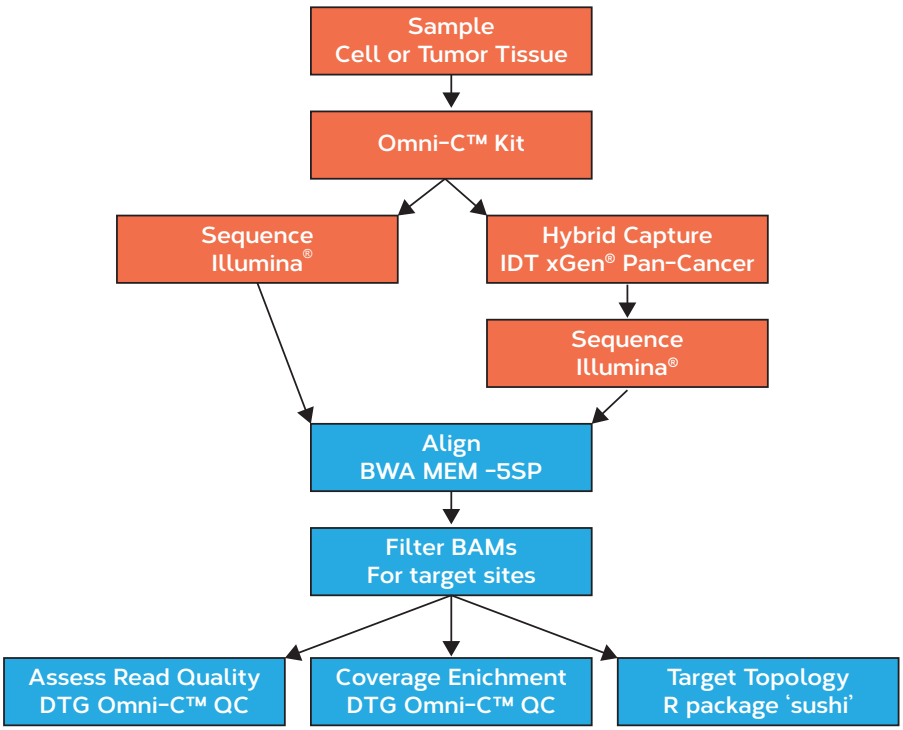
2. Methods

A total of seven samples were selected for Omni-C™ Proximity Ligation library preparation and subsequent hybrid capture enrichment. The samples included three cell lines and four tumor tissues obtained from AMSBio **(Table 1)**. The samples were processed according to **Figure 1**. Briefly, Omni-C™ Proximity Ligation libraries were prepared following the protocol guidelines. Aliquots of the resulting libraries were then subjected to IDT xGen® Pan-Cancer Panel hybrid capture as per the manufacturer's instructions. Libraries were then sequenced with Illumina® technology (2 x 150 bp). Omni-C™ libraries and hybrid capture (HC) libraries were sequenced to ~3 million read pairs. Sequences were then aligned to the human genome (hg38) using BWA MEM with the –5SP flag enabled. Dovetail Genomics' Omni-C™ QC Pipeline was used to assess library quality (read type and cumulative distribution frequency; CDF) over targeted regions by the capture panel. Coverage was calculated and summed across all targets. Topological interactions were visualized in the R package 'sushi'

Table 1 – Sample description and sequencing. Sample used in this study are classified by sample type and cancer type.

Sample	Sample Name	Sample Type	Cancer Type
GM12878	GM12878	Cell	NA
HCC1187	HCC1187	Cell	Breast
HCC1187-BL	HCC1187-BL	Cell	Breast
AMS195	Breast Tumor 1	Tissue	Breast
AMS660	Breast Tumor 2	Tissue	Breast
AMS931	Breast Tumor 3	Tissue	Breast
AMS326	Colon Tumor 1	Tissue	Colon

Figure 1 – Library generation and analysis workflow.



by drawing an arc between two sets of read pairs with one end anchored to a captured target site and the second mapping to a distal sequence as a function of a proximity ligation event.

3. Results and Discussion

All Omni-C™ libraries analyzed contained the key features expected of proximity ligation assays. When subjected to hybrid capture, Omni-C™ libraries contained similar proportions of read types across all samples **(Figure 2A).** Of interest, although a reproducible decrease in trans-reads and increases in cis reads < 1 kbp was observed between Omni-C™ and Omni-C™ HC libraries, both libraries had very similar average cis reads > 1 kbp across all target sites - 44% and 43% for Omni-C™ and Omni-C™ HC libraries respectively. The similarity between Omni-C™ and the paired Omni-C™ HC libraries can also be observed in the CDF plot of the breast cancer cell line HCC1187 **(Figure 2B).** These data confirm that the cis long-range proximity ligation properties inherent to Omni-C™ libraries and critical for detecting chromatin topological features are preserved during the hybrid capture enrichment procedure.

As expected, target sites are enriched in HC libraries compared to their corresponding Omni-C™ libraries. For similar sequencing costs, on average, 20X coverage compared to 0.3X coverage across all target sites were achieved for Omni-C™ HC and Omni-C™ libraries respectively **(Figure 3).** This corresponds to a 375X enrichment in Omni-C™ HC libraries over their Omni-C™ library counterparts. As a result, a significantly increased signal across target sites is evident when comparing the Omni-C™ libraries to the HC libraries in the Integrative Genome Viewer (IGV); **(Figure 3)**. Investigation of the RPL5 gene locus demonstrates a roughly 1X coverage of the HCC1187 Omni-C™ library whereas the corresponding Omni-C™ HC library shows significantly more reads in the targeted region. Taken together, these results indicate that the hybrid capture approach is enriching for desired targets.

It is important to note that, as a result of proximity ligation, Omni-C™ HC libraries will not only enrich for regions targeted but also regions that are in physical proximity to the targeted sites. With this in mind, topological interactions of a target site and the surrounding genomic neighborhood were mapped using the Omni-C™ HC libraries **(Figure 4)**. The high frequency of the interaction with regions directly upstream and downstream, shown in an Omni-C™ contact matrix, can be visualized in the contact arcs drawn between one end of a paired-read mapping to a target site with the second read mapping to an off-target region. Consistent to the behavior observed

Figure 2 – Library characteristics.

A) Sample read type is plotted as a proportion of the filtered alignment. Read types are as follows: proportion of long-range cis reads (>1 kbp), proportion of short cis reads (<1 kbp) and proportion of trans reads.
B) CDF plot example of HCC1187 Omni-C™ libraries (blue) and HC libraries (grey).

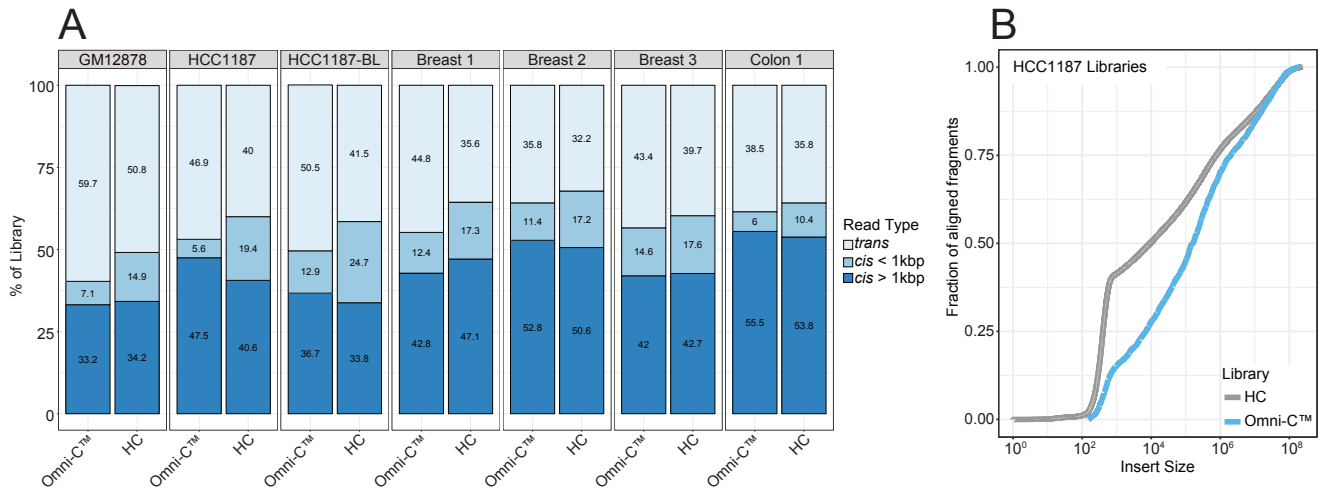
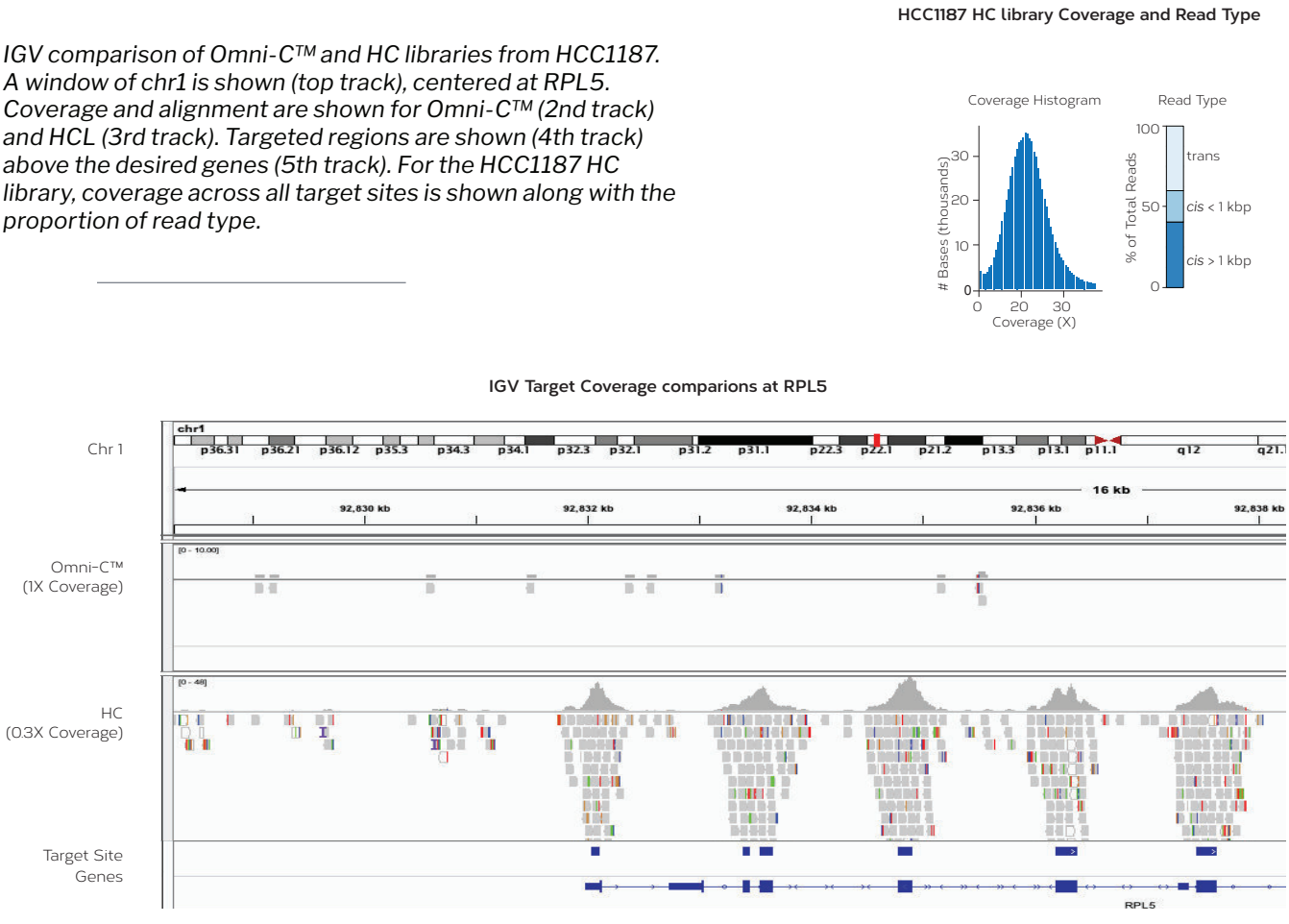


Figure 3 – HCC1187 coverage of hybrid capture target sites.

IGV comparison of Omni-C™ and HC libraries from HCC1187. A window of chr1 is shown (top track), centered at RPL5. Coverage and alignment are shown for Omni-C™ (2nd track) and HCL (3rd track). Targeted regions are shown (4th track) above the desired genes (5th track). For the HCC1187 HC library, coverage across all target sites is shown along with the proportion of read type.



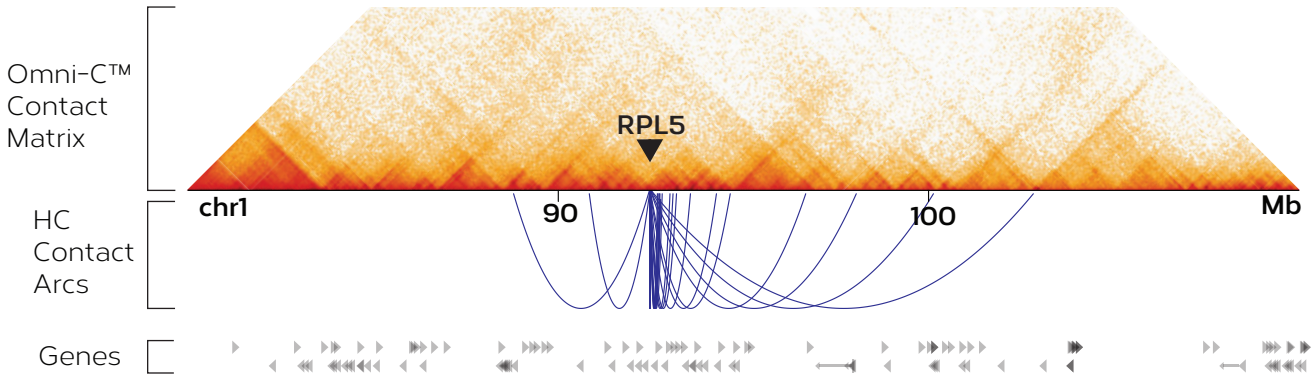
with Omni-C™ libraries, as the distance spanning topological interaction increases, the frequency of the interaction decreases in Omni-C™ HC libraries. Moreover, HC libraries contact arcs mirror the topological boundaries observed in the Omni-C™ contact matrix indicating that the topologically driven chromatin interactions are preserved through the hybrid capture process. The strong correlation between the HC library arcs and the Omni-C™ contact matrix demonstrates that hybrid capture with Omni-C™ libraries can be used to investigate genome conformation at a targeted site.

4. Conclusion

Through this work, we have demonstrated the use of hybrid capture techniques to enrich for target sites of interest using Omni-C™ Proximity Ligation libraries. During this process, proximity ligation events are preserved enabling topological interactions between targeted sites and other regions of the genome to be observed. Combining target enrichment approaches selecting for sites of interest with a non-biased, sequence independent Omni-C™ chemistry now opens up new avenues for investigating the interplay between SNPs, topology, and even large structural variants - all while minimizing sequencing costs. Studies that were previously out of reach due to the high costs associated with running large numbers of samples and/or the need for high-resolution topological data now become feasible.

Figure 4 – Topological interactions of the HC library compared to the Omni-C™ contact matrix.

The Omni-C™ contact matrix was generated from a deeply sequenced library. A region of chr1 centered at RPL5 (black arrow). Below the contact matrix, contact arcs are plotted from the HC library. Briefly, arcs are drawn between two sets of read pairs with one end anchored to a captured target site and the second mapping to a distal sequence as a function of a proximity ligation event. Gene locations are plotted on the bottom track.



References

[1]Mumbach *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. 2016. Nature Methods.
[2]Fang *et al.* Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. 2016. Cell Research.
[3]Schoenfelder *et al.* The pluripotent regulation circuitry connecting promoters to their long-range interacting elements. 2015. Genome Research.