Application Note

Low Input Micro-C for Chromatin Conformation Analysis

Key Takeaways

- Successful Micro-C libraries can be generated with inputs as low as 10 K cells.
- Low input samples are limited in their capacity to support deep sequencing.
- The ability to call TADs and loops is proportional to sequencing depth.
- High-quality Micro-C low input libraries can be generated with the Dovetail® Micro-C Assay, however, discovery may be limited to larger and/or more highly supported topology features.

Introduction

To robustly call chromatin loops, chromatin conformation analysis requires deep sequencing, typically more than 1 billion paired-end reads. For traditional Hi-C protocols, this requires 5 million cells or more to have sufficient input material. However, for many biological questions, samples may be limiting preventing chromatin conformation studies from being performed.

The Dovetail® Micro-C Assay offers superior resolution, signal-to-noise, and a lower overall sequencing burden compared to traditional Hi-C approaches making it the ideal solution for studying chromatin conformation. However, the standard protocol still requires a minimum of 1 million cells to generate sufficient material for

	Molecular QC									Library QC		
Stage	**Stage 1**		**Stage 2**				**Stage 3**	**Stage 5**		**Shallow Sequence QC (2M Read Pairs)**		
Input Amt	**MNase Dil.**	**Vol. MNase (µL)**	**Chromatin Yield (ng/ µL)**	**Total Lysate (ng)**	**% Mono-nucleosome**	**Chromatin Bound (ng)**	**Crosslink Reversal Yield (ng)**	**Final Library Yield (ng)**	**Number of Libraries Supported**	**Valid Read Pairs**	**Long-range Cis Read Pairs**	**Unique Read Pairs @ 300M**
1 M	N/A	0.5	7.2	1,680	36%	1,000	428	1,344	3	91.6%	69.1%	281 M
100 K	1:10	0.5	0.8	188	34%	188	104	1,446	1	91.4%	71.1%	280M
10 K	1:100	0.5	0.2	37	20%	37	36	531	1	94.0%	76.0%	278 M

Table 1 QC metrics for wet lab and computation indicate successful low-input library creation. Stage 1, MNase dilution (MNase: Nuclease-Free Water) and volume of dilution used for chromatin digestion are provided for the various input amounts. Stage 2,3, & 5 capture the molecular biology QC metrics. Library QC metrics such as % valid read pairs (successful proximity ligation), % cis > 1kb (long-range enrichment), and complexity (how deeply a library can be sequenced) are summarized under computation.

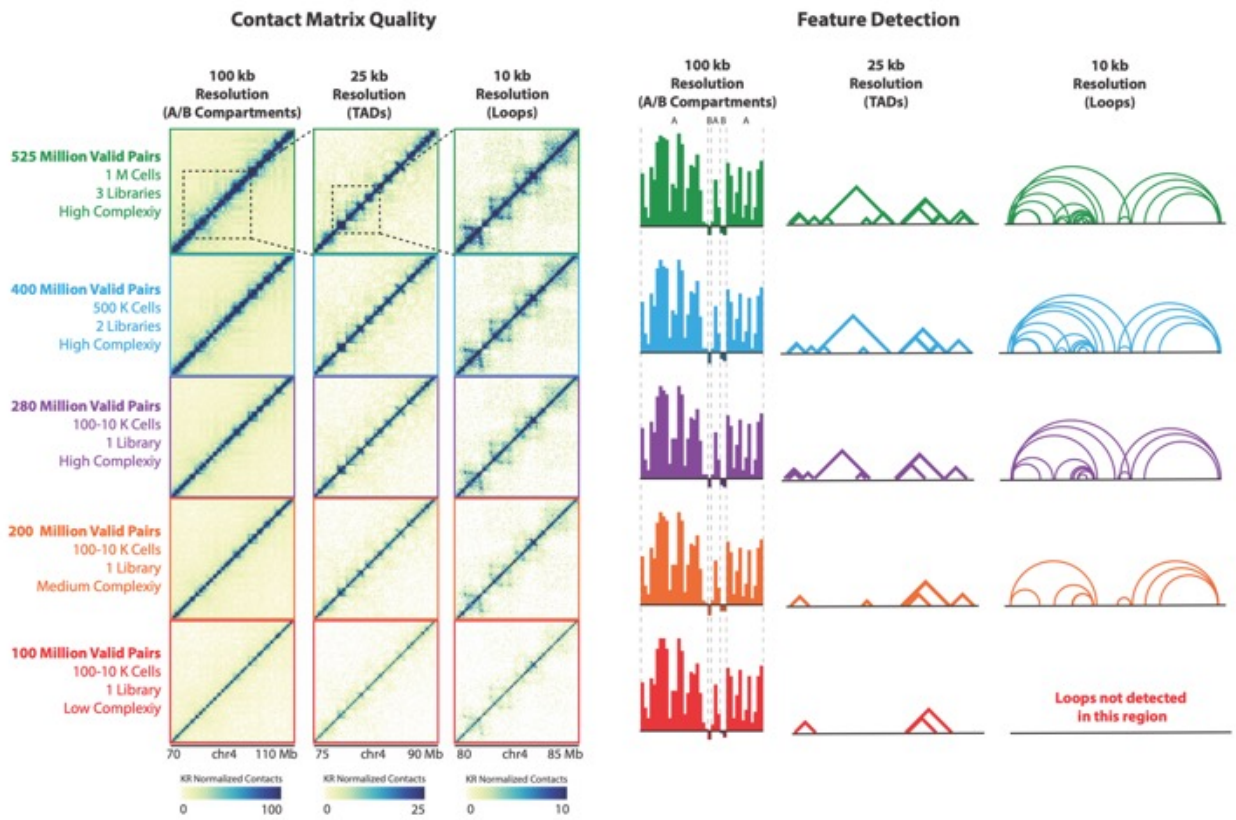


Figure 1 Low input Dovetail® Micro-C libraries preserve conformation signal despite reduction in feature calling efficiency. KR normalized contact matrices were visualized at three resolutions, reflecting feature calling size of A/B compartments (100 kb), TADs (25 kb), and loops (10 kb). Contact scales were held consistent across data sets but varied across resolutions to account for changes in signal strength at higher resolutions. Called features are plotted in the corresponding order. As read depth is limited by input, moderate to smaller features are no longer called, leaving larger or more strongly supported TAD or loops to be detected.

the > 500 million no-duplicate read pairs (800 million total read pairs split across 3 libraries) necessary to support high resolution loop calling.

Here, we provide guidance for low input experiments using the Dovetail® Micro-C Kit, including:

- Optimized Stage 1 digestion for different input amounts.
- Updated QC criteria when working with low input samples.
- Discovery potential for topological features when valid read pairs are limiting.
- Recommendations with low input proximity-ligation data.

Our findings demonstrate that high quality Micro-C libraries can be successfully generated with low-input samples, however, these may be limited to large and/or highly supported topology features due to limiting valid read pairs.

Chromatin Digestion & Molecular Biology QC

Varying input amounts (1 M, 100 K, and 10 K cells), were processed as described in the Dovetail® Micro-C protocol with a single variation to the workflow – during the Stage 1 chromatin digestion, the MNase enzyme was diluted for lower inputs (Table1). Stage 2 lysate QC values scaled with the number of cells and, in all instances, resulted in an acceptable percentage of mononucleosomes. The Stage 3 yield of crosslink reversed (XLR) DNA also correlated with cell input amount.

To support deep sequencing, our

Library Information						Topological Feature Recall						
Input			Complexity & Sequencing Recommendations			TADS (Count [% Recall])			Loops (Count [% Recall])		A/B Comp. [% Recall]	
Valid Pairs	Cell Qty	Supported Libraries	Complexity Per Library	Read Pairs Per Library	Total Pairs Per Sample	5 kb	10 kb	25 kb	5 kb	10 kb	100 kb	1 Mb
525 M	1 M	3	>250 M	300 M	800-900 M	1,342 [100%]	3,566 [100%]	3,383 [100%]	8,650 [100%]	10,851 [100%]	[100%]	[100%]
400 M	500 K	2	>250 M	300 M	500-600 M	872 [65%]	2,802 [95%]	3,218 [95%]	5,684 [66%]	9,956 [98%]	[100%]	[100%]
280 M	100-10 K	1	>250 M	300 M	300 M	383 [29%]	1,748 [49%]	2,702 [80%]	2,705 [37%]	8,566 [79%]	[100%]	[100%]
200 M	100-10 K	1	175 M - 250 M	300 M	300 M	144 [11%]	814 [23%]	2,183 [65%]	1,090 [13%]	6,922 [64%]	[100%]	[100%]
100 M	100-10 K	1	<175 M	200 M	200 M	5 [0.4%]	88 [3%]	892 [26%]	N/A	31 [0.3%]	[100%]	[100%]

Table 2 Summary of down sampled valid read pairs and supported topology features. Number of valid pairs reflects different inputs and library quality as determined by number of supported libraries and library complexity. These metrics assume that proximity ligation was successful and long-range cis pairs >40% of the valid read pairs. Number of topology features (TADs, loops, A/B compartments) and their recall rates are displayed as a percentage of the 525M data set.

recommendation is to generate 3 libraries from each XLR reaction, with each library requiring between 100-150 ng of XLR DNA input. While all Stage 5 library DNA yields indicated successful library conversion, only the 1M cell input reaction provided sufficient XLR DNA to support three library conversions. The 100 K and 10 K samples only yielded enough for a single library conversion.

Library QC

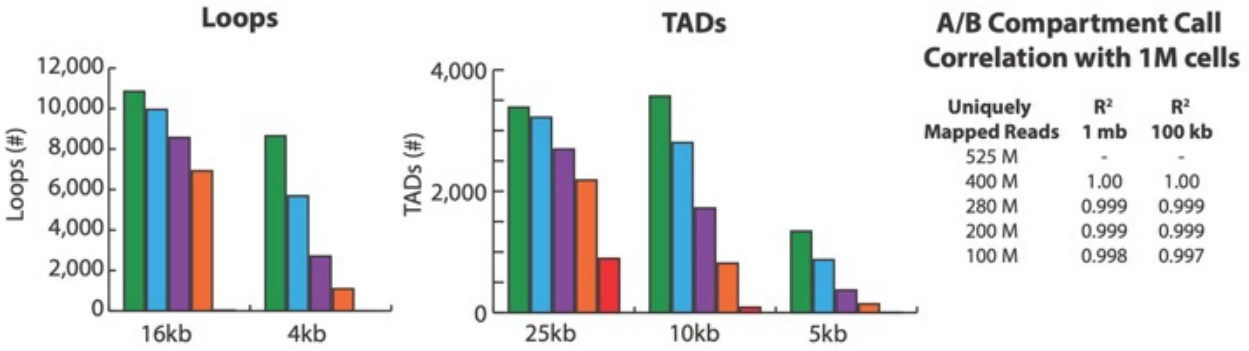
Libraries were sequenced to 2 million read pairs (2x75 bp) for QC. QC metrics are summarized in Table 1. All libraries passed our QC thresholds, with no-duplicate read pairs constituting over 70% of the total 2 M read pairs. Valid pairs (cis > 1 kb + trans) accounted for over 90% of no-duplicate read pairs and 65% of valid pairs were long-range informative. All libraries contained enough complexity to support sequencing to ~280 M read pairs. For the 1 M cell input, this would translate to 810 M pairs across the three libraries, while the 100K and 10 K samples, limited to a single library each, only support 280 M read pairs (Table 1 & Table2).

Impact of Valid Pairs on Contact Matrix Quality and Topological Feature Calling

To assess the effectiveness of limited read pair data sets in capturing conformation signals for various topological features, we down sampled a deeply sequenced data set. This enabled us to simulate the spectrum of achievable valid pairs expected from varying input amounts and library qualities (Table 2), with the 280 M valid read pair dataset representing a high-quality, low-input sample.

As expected, our matrix became sparser with diminishing valid read pairs. However, the general conformation signal was preserved across data sets and resolutions (Figure 2A).

Again, as expected, the number of identified features scales with usable read pairs. The largest feature in contact maps, A/B compartments, remains unaffected by the decrease in usable data with reduced input. Mid-sized contact features and highly supported small features are also detected at all input amounts (Figure 2B). While TADs (> 10 kb) and loops (> 5 kb) are detectable at lower inputs, finer-scale and more transient features are lost. (Figure 3). Feature counts and recall rates are summarized in Table 2.



A/B Compartment Call Correlation with 1M cells

Uniquely Mapped Reads	R^2 1 mb	R^2 100 kb
525 M	-	-
400 M	1.00	1.00
280 M	0.999	0.999
200 M	0.999	0.999
100 M	0.998	0.997

Cells	Libraries	Uniquely Mapped Reads	Complexity	Small Loops (<10kb res)	Big Loops (>10kb res)	Samll TADS (<25 kb res)	Large TADS (>25 kb res)	A/B Comp.
1M	3	525 M	High	+++	+++	+++	+++	+++
500K	2	400 M	High	++	+++	++	+++	+++
100-10K	1	280 M	High	++	++	++	+++	+++
100-10K	1	200 M	Medium	+	+	+	++	+++
100-10K	1	100 M	Low	O	O	O	+	+++

Figure 2 Feature calling discovery rates scale with valid read pairs and recommendations for low input conformation features. Both TAD and loop calls decrease with increased resolution and reduced valid pairs (reflecting low input). A/B compartments are unaffected as determined by eigenvector correlation with a deeply sequenced data set generated from 1 M cells. Finally, the table below highlights the effectiveness of low-input samples at identifying conformation features. Pluses indicate level of feature calling efficiently, three plus '+++' = optimal, two plus '++' =adequate, one plus '+' = minimal, and 'O' as not supported.

Recommendations

The Dovetail® Micro-C Assay can be used effectively with low input samples, however, expectations need to be adjusted. Moderate to high complexity low input samples enable the visualization of contact maps at resolutions of 25-10 kb (Figure 2A).

The Stage 1 MNase enzyme-to-chromatin ratio will need to be adjusted to ensure optimal digestion. Provided this minor adjustment to the protocol is made, chromatin topology features can be robustly called, albeit, with reduced TAD and loop discovery rates. In summary, the Dovetail Micro-C Assay can be used to generate actionable biological insights can be generated from low input samples.

For more information, visit https://dovetailgenomics.com/products/micro-c-product-page/