

ChIA-PET Evaluation of Data Quality

Results of Short-Read ChIA-PET Data

Table 5

Statistics of ChIA-PET library from human K562 cells.

Category	Amount
Percentage of same-linker PETs over total PETs	90.74%
Percentage of uniquely mappable PETs over total PETs	14.2%
Percentage of PETs after merging those mapped to the same positions exactly over uniquely mappable PETs	69.02%
Percentage of inter-ligation PETs over PETs after purification	40.28%
Percentage of intra-chromosomal inter-ligation PETs over inter-ligation PETs	53.94%
Number of peaks	6332
Number of interactions	31,011

- 1st line: the percentage of the same-linker PETs over the total PETs(varying from 60% to 99%), If such a percentage is above 75%, the library is good in the linker composition level.
- 2nd line: the percentage of the uniquely mappable PETs over the total PETs, which varies with the libraries.
- 3rd line: the percentage of the PETs after merging those mapped to the same positions exactly over the uniquely mappable PETs. If this percentage is too low (less than 30%), it means that there are more PETs from PCR amplification, and the data are already near saturated(there is no point in re-sequencing the library for more distinct PETs). if >70%, deeper sequencing can be applied to get more data for the library.
- 4th line: the percentage of inter-ligation PETs over all the PETs after purification. This is about the efficiency catching the interacting PETs. If this percentage is low, it means that the library has too few inter-ligation PETs and is not good enough for chromatin interaction detection.
- 5th line: the percentage of intra-chromosomal inter-ligation PETs over the total inter-ligation PETs. From the current understanding, most of the chromatin interactions are within the individual chromosomes. Therefore, there should be more intra-chromosomal PETs than inter-chromosomal PETs. If there are more inter-chromosomal PETs, it means that the proximity ligation introduces many random ligations.

6th line: the peak number. This depends on the transcription factor used, and should be compared with the background knowledge or available ChIP-Seq data.

7th line: the number of chromatin interactions. This depends on the transcription factors used. For RNAPII and CTCF, there are tens of thousands of interactions in a good ChIA-PET library from human and mouse.