ChiCMaxima: a robust and simple pipeline for detection and visualization of chromatin looping in Capture Hi-C

Additional File 1: Figures S1-S8, and Tables S2 and S3

Yousra Ben Zouari, Anne M Molitor, Natalia Sikorska, Vera Pancaldi and Tom Sexton
A

CHiCAGO

793208

225743

3318

Jaccard Index: 0.003

ChiCMaxima

66793

64915

9215

Jaccard Index: 0.07

B

Tet2– chr3: 132509612–133907246

Running mean QN signal

Genomic coordinates

CTCF

H3K27ac
A

Genomic coordinates

Running mean QN signal

ChiCMaxima

+ CHiCAGO

Fold enrichment over background

Enhancer Polycomb

CTCF H3K27ac H3K4me1 Ring1B Suz12

Total ChiCMaxima

not in CHiCAGO

Enhancer Polycomb

CTCF H3K27ac H3K4me1 Ring1B Suz12

Fold enrichment over background

B

Fold enrichment over background

Enhancer Polycomb

CTCF H3K27ac H3K4me1 Ring1B Suz12

Total ChiCMaxima

in CHiCAGO

not in CHiCAGO

C

Chn2– chr6: 53501685–54694204

Running mean QN signal

Genomic coordinates

CTCF

Creb5 Tril Cpvl Chn2 Prr15 Wipf3 Scm1 Plekha8

9130019P16Rik PkBp14

ChiMaxima

ChiMaxima + CHiCAGO
Ben Zouari et al. Fig S5

A. M0 macrophage HEY1--chr8:80300703--81296252

B. M1 macrophage PICALM--chr11:85400223--86197866

C. M2 macrophage NFIL3--chr9:93405708--94694530

D. Megakaryocyte DAAM1--chr14:59300201--60199421

E. Monocyte IPO9--chr1:201303313--202198741

Naive CD4 T cell IL7R--chr5:35505210--36199584
SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Under-sampled CHi-C datasets confound analyses at single restriction fragment level. a Venn diagrams of interactions from two biological replicates of mES CHi-C, called separately by CHICAGO (left) or ChiCMaxima (right). Although at a single-replicate level, CHICAGO calls about ten-fold more interactions than ChiCMaxima, fewer total interactions are conserved across two biological replicates at the single restriction fragment level. Note that despite giving a 20-fold improvement in reproducibility over CHICAGO, fewer than 10% of interactions called by ChiCMaxima are conserved across two replicates at the single restriction fragment level. b The CHi-C profile centered on the bait Tet2 promoter is shown for two mES replicates (blue and red); the quantile-normalized raw reads in the top panel, and the quantile-normalized running mean (over ten fragment windows) in the lower panel. Less reproducible signal “spikes” at single fragments in the raw reads are called as interactions by replicate-specific CHICAGO analyses (red and blue stripes for each replicate, respectively; CHICAGO score ≥ 5); an interaction is called by ChiCMaxima (gray stripe) over both replicates at a CTCF-rich region, based on consistently higher CHi-C signal over consecutive restriction fragments. The gene positions (blue) and mES ChIP-seq profiles for CTCF and H3K27ac (green) are shown below the CHi-C profiles.

Figure S2. Testing parameters of ChiCMaxima. a Boxplot showing distribution of read numbers for interactions called by ChiCMaxima with different settings of $w$ within a subset of mES CHi-C data (no extra filters applied). Total numbers of called interactions is given below the boxes. Increasing $w$ increases the number of supporting reads within fewer overall called interactions. All $w$ settings tested gave interaction calls with overall greater numbers of reads
than the calls made by CHiCAGO on the same dataset. b Boxplot showing distribution of interaction distances for the same ChiCMaxima calls as in a. Large $w$ has a strong bias towards shorter-range interactions. c Boxplot showing distribution of read numbers for interactions called by ChiCMaxima on the same dataset as a, with $w = 20$, and applying different filters (None – no filter; Geo – filtering for reads greater than geometric mean within 20 kb genomic separation bins; Lin – filtering for reads greater than log-linear fit to same geometric mean distribution; Cub – filtering for cubic fit to log-distance of same geometric mean distribution). d Boxplot showing distribution of interaction distances for the same ChiCMaxima calls as in c. We note that applying a geometric mean filter to smaller $w$ introduces much less bias to shorter-range interactions than applying no filter to large $w$.

Figure S3. CHi-C interaction calling across biological replicates. a The CHi-C profile (quantile-normalized running mean over ten fragments) centered on the bait $Ebf1$ promoter is shown for two mES replicates (blue and red). A selected interaction with a CTCF site is highlighted, with the exact position of the called ChiCMaxima interaction shown for each replicate (replicate 1 in blue; replicate 2 in red). This interaction is not reproduced across the two replicates at the individual restriction fragment level, but the two calls are within 5650 bp/4 HindIII fragments of each other, and centered on the same major peak. The gene positions (blue) and mES ChIP-seq profile for CTCF (green) are shown below the CHi-C profiles. b Histogram showing the distributions for the genomic distances between interacting regions called from one replicate and the closest interaction called from the other replicate. c Cumulative frequency plot for the same genomic distance distribution as b; nearly 40% of the interactions of one replicate are within 20 kb of the interactions of the other replicate.
**Figure S4.** Epigenomic enrichments from alternative interaction calling methods. a Bar chart showing fold enrichment over genomic background for different ChIP-seq peaks within all the promoter-interacting sequences determined by ChiCMamaxma with different parameters. Whichever ones are used, the enrichments are consistently higher for enhancer marks than either CHiCAGO- or GOTHiC-called interactions, and comparable to CHiCAGO for Polycomb marks. b Bar chart showing fold enrichment over genomic background for different ChIP-seq peaks within all the promoter-interacting sequences determined by ChiCMamaxma, compared to the subsets that are or are not conserved with interactions called by CHiCAGO. c The CHi-C profile (quantile-normalized running mean over ten fragments) centered on the bait Chn2 promoter, with interactions called by ChiCMamaxma alone denoted with green stripes, and interactions called by both ChiCMamaxma and CHiCAGO denoted with orange stripes. Gene position (blue) and the mES CTCF ChIP-seq profile (dark green) are shown below the CHi-C profile. Interactions with additional CTCF sites are called by ChiCMamaxma, which are not recapitulated by CHiCAGO.

**Figure S5.** Improved stringency of ChiCMamaxma over CHiCAGO when applied to human primary hematopoietic cell CHi-C data. Several CHi-C profiles are plotted for different cell type/gene promoters, with interactions called by ChiCMamaxma denoted as gray stripes, and interactions called by CHiCAGO denoted as red spots. Gene positions (blue) are shown below the profiles. a HEY1 in M0 macrophages; b PICALM in M1 macrophages; c NFIL3 in M2 macrophages; d DAAM1 in megakaryocytes; e IPO9 in monocytes; f IL7R in naïve CD4 T cells.
Figure S6. ChiCM Maxima is not just a more stringent version of CHiCAGO. Bar charts showing fold enrichment over genomic background for ChIP-seq peaks (a H3K27ac; b H3K4me1) within all the promoter-interacting sequences determined by ChiCM Maxima, compared to the equal number of highest-scoring CHiCAGO-called interactions, computed across nine human primary hematopoietic cell CHi-C datasets. The numbers below the bar chart denote the threshold CHiCAGO score applied to obtain the required number of interactions. ChiCM Maxima nearly always outperforms the matched CHiCAGO calls.

Figure S7. Scatter plot of chromatin assortativity against relative feature abundance for different chromatin features within the ChiCM Maxima-called interaction network derived from the mES CHi-C dataset.

Figure S8. Flexibility in handling replicates in ChiCBrowser. The CHi-C profiles plotted from two mES CHi-C biological replicates (called “N.1” and “N.2” in the original input file) by ChiCBrowser for a 1 Mb window centered on the bait Sall1 promoter. Left: The replicates are plotted side by side by giving N.1 and N.2 different levels (1 and 2, respectively) in the Set Conditions menu, and renaming them “rep1” and “rep2” in the Plot Conditions menu. Right: The mean of the two replicates is plotted by giving N.1 and N.2 the same level (1) in the Set Conditions menu, and renaming the combined plot “mES” in the Plot Conditions menu.
Table S2: Overview of CHi-C interactions called by ChiCMaxima with different parameters.

|                         | ChiCMaxima (standard) | c = 5 Mb | w = 10 | w = 50 |
|-------------------------|-----------------------|----------|--------|--------|
| Number of called interactions | 23583                | 21947    | 47932  | 14212  |
| Mean number of called interactions per bait | 1.4                   | 1.37     | 2.5    | 1.05   |

Table S3: Overview of putative mES enhancers found within CHi-C interactions called by ChiCMaxima with varying parameters.

|                         | Putative mES enhancers in called interaction set | Total called interactions / interactions with putative enhancers |
|-------------------------|--------------------------------------------------|---------------------------------------------------------------|
| ChiCMaxima (standard)   | 16.8% (3235)                                     | 7.3                                                           |
| c = 5 Mb                | 16.1% (3092)                                     | 7.1                                                           |
| w = 10                  | 33.0% (6341)                                     | 7.6                                                           |
| w = 50                  | 9.7% (1868)                                      | 7.6                                                           |