Comparison of computational methods for Hi-C data analysis

Mattia Forcato[1], Chiara Nicoletti[1], Koustav Pal[2], Carmen Maria Livi[2], Francesco Ferrari[2–4] & Silvio Bicciato[1,4]

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

Hi-C is a genome-wide sequencing technique used to investigate 3D chromatin conformation inside the nucleus. Computational methods are required to analyze Hi-C data and identify chromatin interactions and topologically associating domains (TADs) from genome-wide contact probability maps. We quantitatively compared the performance of 13 algorithms in their analyses of Hi-C data from six landmark studies and simulations. This comparison revealed differences in the performance of methods for chromatin interaction identification, but more comparable results for TAD detection between algorithms.

The identification of the 3D structure of chromatin inside the nucleus is crucial for deciphering how the spatial organization of DNA affects genome functionality and transcription. Methods based on chromosome conformation capture (3C)[1], such as Hi-C, combine proximity-based DNA ligation with high-throughput sequencing to assess the spatial proximity of potentially any pair of genomic loci[2]. Techniques like Hi-C investigate chromatin structures, such as interactions and topologically associating domains[3]. Chromatin interactions are contacts between regions far from each other on the linear DNA sequence but close in 3D space[4]. TADs are structural domains consisting of chromatin regions that are highly self-interacting but have limited interaction with regions in other domains[5–7].

Hi-C produces hundreds of millions of read pairs that are used to generate genome-wide maps containing millions of contacts between genomic loci pairs[8–10]. The analysis of this enormous amount of genomic data required the development of *ad hoc* algorithms and computational procedures. Different bioinformatics tools have recently been implemented to efficiently preprocess sequence reads (quality control, alignment, and filtering), remove biases (normalization of contact matrices), and infer chromatin structures[10,11]. As algorithmic choices severely impact the identification of chromatin structures[9,12,13], to ensure the reliability of results, it is essential to assess the relative performance of various tools.

Using experimental and simulated data, we quantitatively compared the performances of Hi-C data analysis methods in the identification of chromatin interactions[9,14–19] and TADs[5,9,14,20–24]. We also addressed elements of tool usability including running time and computational requirements. In general, we see that, depending on the tool, identified structures vary in terms of quantity and characteristics and are more reproducible for TADs than for interactions.

RESULTS

Tools and data preprocessing

We compared 13 methods for the analysis of Hi-C data (**Table 1**; **Supplementary Notes 1** and **2**) using experimental and simulated data. Experimental data were obtained from six landmark studies[2,5,7–9,25], from which we selected nine data sets for a total of 41 samples covering multiple protocol variations, data resolutions, and cell types (**Table 2** and **Supplementary Table 1**). We generated simulated data with a modified version of the model proposed by Lun and Smyth[19] (**Supplementary Note 3**). The various methods preprocess Hi-C data using different alignment and filtering strategies (**Fig. 1a** and **Supplementary Table 2**). Most interaction callers do not include an alignment step; and we used Bowtie[26], a full-read approach, for read mapping. However, HIPPIE[18], HiCCUPS[9,14], and diffHic[19] use chimeric alignment that also allows mapping of reads spanning the ligation junction. Each interaction caller adopts a specific filtering method—with the exception of Fit-Hi-C[15], for which we used GOTHiC[16] filtering. Most TAD callers require a fully preprocessed interaction matrix as input, and thus they do not provide specific approaches for alignment and filtering—TADbit[21] and Arrowhead[9,14] are the two exceptions. Thus, to maximize comparability, we applied a uniform preprocessing procedure (i.e., Bowtie[26] for alignment and hicpipe[27] for filtering) to create the interaction matrix for TAD identification.

On average, methods implementing chimeric alignment aligned 18.4% (chimeric STAR[28] in HIPPIE[18]), 27.4% (chimeric BWA[29] in HiCCUPS[9,14]), and 40.1% (chimeric Bowtie2 (ref. 30) in diffHic[19]) more reads than Bowtie[26]. As the read length increased, the difference in alignment rate between chimeric and full reads became more evident, ranging from 30.9% (at 36 bp) to 55.4% (at 101 bp) of additionally aligned reads (chimeric Bowtie2 (ref. 30); **Fig. 1b**).

[1]Department of Life Sciences, Center for Genome Research, University of Modena and Reggio Emilia, Modena, Italy. [2]IFOM, the FIRC Institute of Molecular Oncology, Milan, Italy. [3]Institute of Molecular Genetics, National Research Council, Pavia, Italy. [4]These authors contributed equally to this work. Correspondence should be addressed to S.B. (silvio.bicciato@unimore.it) or F.F. (francesco.ferrari@ifom.eu).

RECEIVED 16 DECEMBER 2016; ACCEPTED 8 MAY 2017; PUBLISHED ONLINE 12 JUNE 2017; DOI:10.1038/NMETH.4325

Table 1 | Methods for Hi-C data analysis used in this comparison

	Method	Availability	Programming language
Chromatin interactions	Fit-Hi-C[15]	http://noble.gs.washington.edu/proj/fit-hi-c	Python
	GOTHiC[16]	http://bioconductor.org/packages/release/bioc/html/GOTHiC.html	R
	HOMER[17]	homer.ucsd.edu/homer/download.html	Perl, R
	HIPPIE[18]	wanglab.pcbi.upenn.edu/hippie	Python, Perl, R
	diffHic[19]	https://bioconductor.org/packages/release/bioc/html/diffHic.html	R, Python
	HiCCUPS[9,14]*	https://github.com/theaidenlab/juicer/wiki/Download	Java
TADs	HiCseg[20]	https://cran.r-project.org/web/packages/HiCseg/index.html	R
	TADbit[21]	https://github.com/3DGenomes/TADbit	Python
	DomainCaller[5]	http://chromosome.sdsc.edu/mouse/hi-c/download.html	Matlab, Perl
	InsulationScore[22]	https://github.com/dekkerlab/crane-nature-2015	Perl
	Arrowhead[9,14]*	https://github.com/theaidenlab/juicer/wiki/Download	Java
	TADtree[23]	compbio.cs.brown.edu/projects/tadtree/	Python
	Armatus[24]	https://github.com/kingsfordgroup/armatus	C++

*HiCCUPS and Arrowhead[9,14] are the algorithms for interaction and TAD calling of the Juicer software suite.

After the filtering step, HiCCUPS[9,14] retained the largest number of aligned reads (**Fig. 1c**), although it is worth noting that HiCCUPS[9,14] filters only PCR duplicates without discarding other potential artifact reads. diffHic[19] filtered the highest proportion of aligned reads in most data sets (from 27% to 94%, depending on the data set); but, given its higher alignment rate, still retained a large number of reads (**Supplementary Table 3**). The different experimental protocols severely affected the percentage of filtered reads, and *in situ* Hi-C resulted in more reads passing the filtering step (>76%; **Fig. 1c**). The smaller fraction of retained reads observed in data generated with the simplified Hi-C protocol was mostly due to a larger amount of PCR duplicates (**Supplementary Table 3**).

Hi-C read counts are usually summarized at the level of genomic bins with a fixed width larger than the size of individual restriction fragments. For each data set, we used the same bin size (resolution) used in the original publication (of the given data set) to call interactions, whereas we used bins of at least 40 kb for TAD calling (**Table 2**).

When a method required a normalization step, we used its original normalization procedure; we applied hicpipe[27] to normalize the matrices for DomainCaller[5], InsulationScore[22], Arrowhead[9,14], Armatus[24], and TADtree[23] (**Fig. 1a**). In all cases, we did not evaluate the effect of different normalization strategies, as thorough comparisons of normalization methods have already been addressed[27,31,32].

Identification of chromatin interactions

On experimental data, the total number of interactions called by each method increased with the number of reads retained by the filtering step for all tools at any resolution, although the rate of increase varied from tool to tool (**Fig. 2a**). Consistent with the expectation that 3D interactions mostly occur within chromosomes (*cis*) rather than between chromosomes (*trans*), all methods detected more *cis* than *trans* interactions. In most data sets, GOTHiC[16] called the highest number of *cis* interactions (**Supplementary Fig. 1a**) and, in general, diffHic[19] found the largest number of *trans* interactions (**Supplementary Fig. 1b**). For all tools, the rate of increase of the number of interactions with the number of retained reads was higher for *cis* than for *trans* interactions (**Supplementary Fig. 1c**). HiCCUPS[9,14], which aggregates nearby peaks into a single interaction, identified fewer interactions than all other tools.

When measuring the distance between the interacting points in *cis*, GOTHiC[16] found interactions at shorter mean distance at both 5- and 40-kb resolutions (**Fig. 2b** and **Supplementary Fig. 2**). At 5 kb, Fit-Hi-C[15] called interactions at an average distance of more than 10 Mb; which was expected, as Fit-Hi-C[15] is designed to call midrange interactions. At a resolution of 1 Mb, with the exception of HIPPIE[18], all tools detected interactions with an average distance between 10 Mb (HiCCUPS[9,14] and GOTHiC[16]) and 53 Mb (diffHic[19]) (**Supplementary Fig. 2**).

The differences in the number of interactions and the distance between the interacting points identified by the various methods are immediately evident in the visual representation of the contact matrices (**Fig. 2c**).

To compare the reproducibility of interactions called in different replicates, we calculated the similarity coefficient of Jaccard

Table 2 | Hi-C experimental data

	Cell type				Restriction enzyme								
Study	LCL[a]	H1-hESC	IMR90	Fly embryo	Hi-C protocol[b]	HindIII (6 bp)	NcoI (6 bp)	DpnII (4 bp)	MboI (4 bp)	Read length (bp)	Median read count (per replicate, in millions)	Resolution (kb)[d]	No. of replicate samples
Lieberman-Aiden[2]	✓				Dilution	✓	✓			76	11	1,000	4
Sexton[7]				✓	Simplified			✓		36	362	40	1
Dixon[5] (2012)		✓	✓		Dilution	✓				36/100[c]	328	40	4
Jin[8]		✓	✓		Dilution	✓				36/50[c]	440	5/40	7
Rao[9]	✓		✓		*In situ*			✓	✓	101	240	5/40	23
Dixon[25] (2015)		✓			Dilution	✓				36/50[c]	999	5/40	2

[a]LCL, lymphoblastoid cell lines (GM06990 in Lieberman-Aiden [2]and GM12878 in Rao[9]). [b]Dilution, simplified, and in situ refer to the Hi-C protocols presented in Lieberman-Aiden *et al.*[2], Sexton *et al.*[7], and Rao *et al.*[9], respectively. [c]Samples have been sequenced with different read lengths in the same study. [d]Resolution refers to the resolution used in this comparison. In the case of two values, the first refers to the resolution used for chromatin interactions, the second for TADs.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

(Jaccard Index, JI) as a measure of the overlap between sets of interactions. For most data sets, the reproducibility among replicates of the same data set (intra-data set) was low at all resolutions (**Fig. 2d** and **Supplementary Fig. 3a**), yet it was significantly higher than random sets of interactions (*P* values ≤ 0.001; **Supplementary Fig. 3b**). Surprisingly, the concordance was higher for *trans* (median JI of 0.19) than for *cis* interactions (median JI < 0.03). At low resolution, GOTHiC[16] had the highest concordance, most likely because it called a large number of short-range interactions in every sample replicate. Conversely, in almost all data sets at high resolution, the interactions found by HiCCUPS[9,14] were the most conserved among replicates. JI quantification that considered only the top 1,000 *cis* interactions (called by each method in each replicate of Rao *et al.*[9] IMR90) resulted, with the exception of Fit-Hi-C[15], in no overall significant improvement of the concordance (*q* value > 0.05 in a one-tail Wilcoxon test with Benjamini–Hochberg correction; **Supplementary Fig. 4a**). Instead, when grouping samples based on increasing number of reads, the reproducibility increased with the number of reads, especially for HiCCUPS[9,14] and GOTHiC[16] (**Supplementary Fig. 4b**). The interactions identified by HiCCUPS[9,14] and GOTHiC[16] were also the most reproducible when using the overlap coefficient, a similarity measure more robust to imbalanced numbers of interactions between the compared replicates (**Supplementary Fig. 4c**).

The intra-data set reproducibility remained similar when comparing replicates of the same cell line processed using different restriction enzymes (**Supplementary Fig. 5**). However, the inter-data set reproducibility—i.e., the concordance between interactions called in samples of the same cell line in different data sets (using different protocols and enzymes)—was much lower (median JI < 4×10^{-4}; **Supplementary Fig. 6**).

We then evaluated the performance of each tool in detecting interactions associated to chromatin states related to transcriptional regulation. In particular, for each data set and cell type, we classified interactions (**Supplementary Table 4**) based on the respective chromatin states at their anchoring points[33,34]. Considering all methods and the data at 5-kb resolution; on average, 16% of all detected *cis* interactions were classified as promoter–enhancer, 23% as interactions connecting heterochromatin or quiescent states, and 3% as biologically less expected (i.e., connecting promoter or enhancer to heterochromatin or quiescent states) (**Fig. 2e**). At this resolution, HiCCUPS[9,14] and HOMER[17] called the highest proportion of promoter–enhancer interactions, although not the highest absolute number (**Supplementary Fig. 7a**). In data sets at 40-kb resolution, all methods detected larger proportions of promoter–enhancer interactions on account of the higher probability that larger bins will contain an enhancer or a promoter (**Supplementary Fig. 7b**). On the contrary, the proportion of *trans* interactions classified as promoter–enhancer was very low for all tools in almost all data sets (**Supplementary Table 5**). diffHic[19] returned the highest quantity and percentage of interactions connecting heterochromatin or quiescent states, even though, in some data sets, the proportion of this type of interaction was extremely high for all tools. Irrespective of the method and the resolution, less than 8% of all *cis* interactions were classified as biologically less expected. For all tools, the enrichment of the number of promoter–enhancer interactions over random expectation tends to be higher in data sets at higher resolution (*P* value ≤ 0.01 in a hypergeometric test for most data sets at 5 kb; **Supplementary Table 6**).

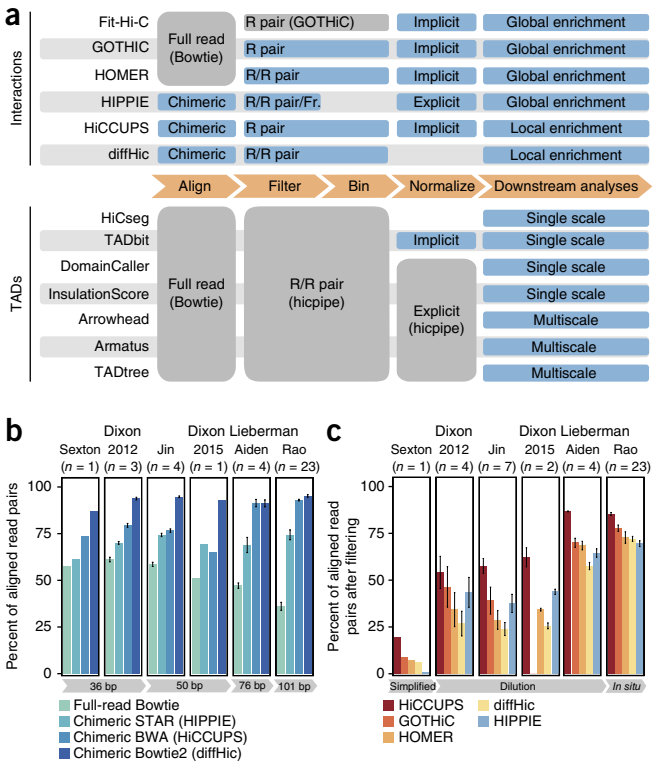

Figure 1 | Tools for Hi-C data analysis used in the comparison and performances in data preprocessing. (**a**) Tools for the identification of chromatin interactions and TADs from Hi-C data and key analysis steps (orange arrows). Blue boxes detail the strategy used in each analysis step by each tool. A gray box is used when an external ng together, a blue or gray box spanning both steps is used in the schematic workflowtool is required for a preprocessing step. Since most tools perform filtering and binni. For filtering, the following abbreviations are used: read-level filtering, R; read-pair-level filtering, R-pair; fragment-level filtering, Fr. (**b**) Percentage of aligned read pairs (alignment rate) for all data sets ordered by read length (gray arrows at the bottom). Data are shown as mean ± s.e.m. Samples with different or mixed read length were not used when calculating the alignment rate. (**c**) Percentage of mapped reads retained after filtering (fraction of usable reads) in each data set, ordered by experimental protocol (gray arrows at the bottom). Data are shown as mean ± s.e.m. GOTHiC[16] could not be applied to Dixon *et al.*[25], since the read-pairing step required an amount of memory larger than 1 TB of RAM.

All methods identified large proportions of convergent orientation of CTFC motifs—a distinctive feature of specific types of interactions[9]—among interactions with a single CTCF-binding motif in each of the two interacting bins (**Supplementary Note 4**).

When comparing the power to recall validated *cis* interaction evidences between algorithms (**Supplementary Table 7**), GOTHiC[16] recovered the largest number of true-positive interactions. HOMER[17] and Fit-Hi-C[15] performed comparably to GOTHiC[16], although they called a smaller number of total interactions (**Fig. 2f**). In high-resolution data sets, diffHic[19] recalled the highest number of true positives, although HOMER[17] identified more true positives than any other tool at comparable numbers of called interactions (**Supplementary Fig. 7c**). All tools recalled low proportions of true negatives in almost all data sets,

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

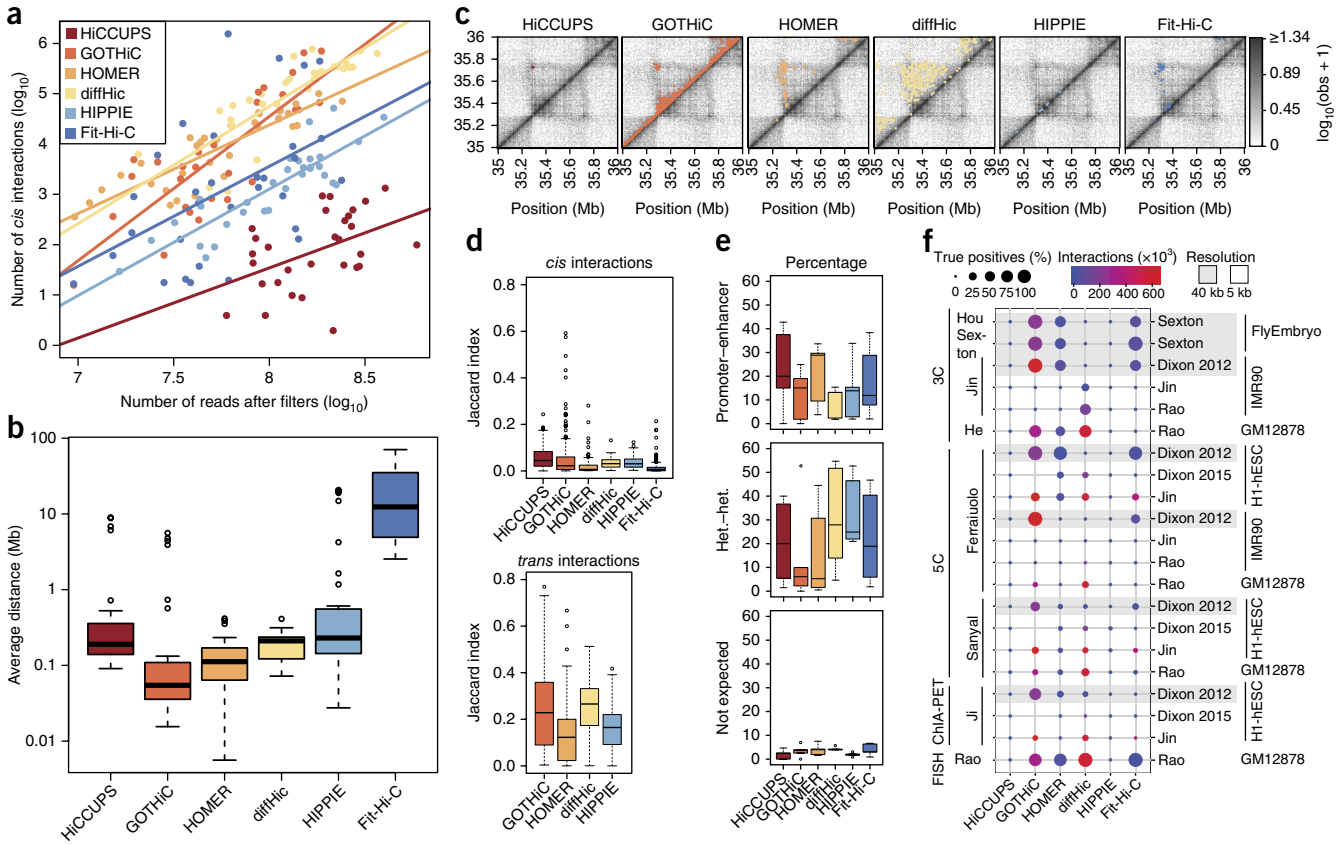

Figure 2 | Comparative results of methods for the identification of chromatin interactions. (**a**) Total number of *cis* interactions called by each method as a function of the number of reads retained by the filtering step (data sets at 5-kb resolution; see **Table 2**; $n = 32$). Dots, sample replicates; solid line, linear interpolation. (**b**) Box plot of average distances between anchoring points in *cis* interactions (data sets at 5-kb resolution; $n = 32$). (**c**) Heat map of the contact matrix of Rao *et al.*[9] GM12878 replicate H (chr21:35,000,000–36,000,000) at 5-kb resolution. Identified peaks are marked in different colors for the various methods. Obs, observed counts. (**d**) Box plots of the Jaccard Index for concordance of *cis* (upper) and *trans* (lower) interaction calls between sample replicates (intra-data set concordance) for all data sets with at least two replicates ($n = 39$; **Supplementary Table 1**). Fit-Hi-C[15] and HiCCUPS[9,14] do not return *trans* interactions. (**e**) Proportion of *cis* interactions classified on the base of the chromatin states at their anchoring points (data sets at 5-kb resolution). (**f**) Performances in the identification of true positive validated evidences of *cis* interactions. Each row represents the comparison between a list of true positives and the interactions called by each method in each data set. The dot size is proportional to the percentage of recalled true positives, and the dot color accounts for the number of total called interactions. Left side, validation technique and name of true positive lists; right side, data set used to call interactions (in gray if at 40-kb resolution). GOTHiC[16] was not applied to Dixon *et al.*[25] (see legend of **Fig. 1c**). Only *cis* interactions conserved in at least two replicates within each data set were used for **Figure 2e–f**, with the exception of the Jin *et al.*[8] H1-hESC and Sexton[7] data sets (both containing a single replicate; **Supplementary Table 4**).

although GOTHiC[16] was more prone to false positives in data sets at 40 kb (**Supplementary Fig. 7d**).

To assess sensitivity and precision of the methods, we modified the model of Lun and Smyth[19] to generate simulated interaction matrices, and we analyzed the simulated data with HiCCUPS[9,14], HOMER[17], diffHic[19], and Fit-Hi-C[15]—the only tools that can take as input the sole interaction matrix. For a set of 40 samples, at eight levels of base-interaction strength, all tools called a much larger number of interactions than the 1,000 true interactions (**Supplementary Fig. 8a**). As for experimental data, Fit-Hi-C[15] called interactions at larger mean distance than the other three tools we analyzed (**Supplementary Fig. 8b,c**). The highest sensitivity was achieved by Fit-Hi-C[15], although all tools displayed an extremely high false discovery rate (FDR) (i.e., a low precision) (**Supplementary Fig. 8d,e**).

Identification of topologically associating domains

For TAD calling, we analyzed all experimental data at a resolution of 40 kb with the exception of Lieberman-Aiden *et al.*[2], for which we used the original 1-Mb resolution. Unlike interaction callers, the number of TADs did not increase with the number of reads retained after filtering for all tools, with the exception of Arrowhead[9,14] (**Fig. 3a**). The number of TADs identified varied from tool to tool and was, for most methods in most data sets, inversely proportional to their size (**Fig. 3b**). On average, in all data sets at 40-kb resolution, TADtree[23] called the largest (7,638) and Arrowhead[9,14] the smallest (636) number of TADs. Conversely, at 1-Mb resolution, InsulationScore[22] returned the largest number of TADs (**Supplementary Table 8**). The characteristics of the identified TADs are exemplified in the heat map representation of the contact matrices (**Fig. 3c**). Note that some methods (HiCseg[20], TADbit[21], InsulationScore[22]) partition chromosomes in a continuous set of TADs, whereas the others allow gaps between TADs. Arrowhead[9,14] and TADtree[23], which adopt multiscale approaches, returned nested TADs.

To compare the reproducibility of TADs, we calculated the JI as a measure of the overlap between TAD boundaries across

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

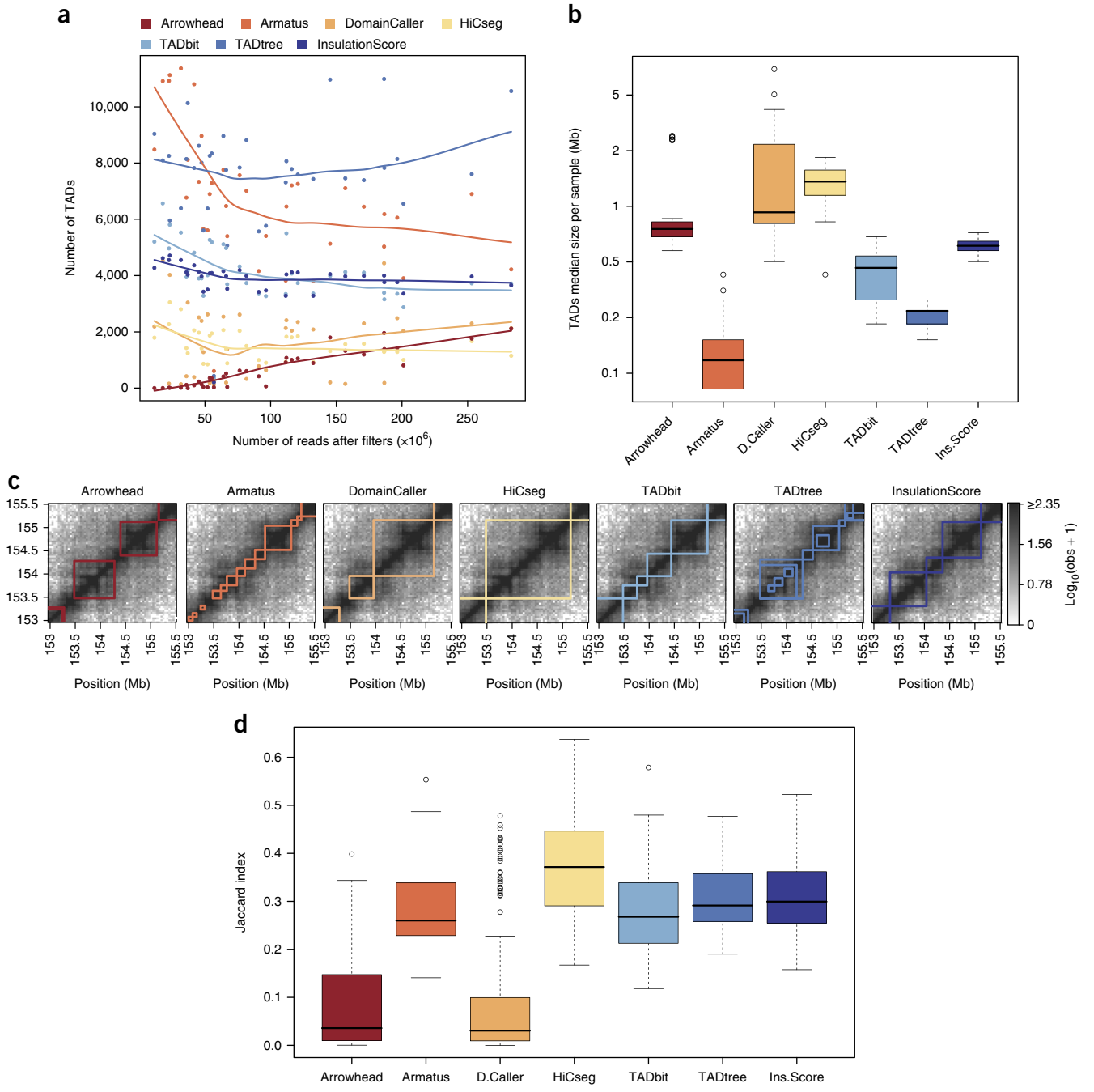

Figure 3 | Comparative results of methods for the identification of TADs. (**a**) Scatter plot of total number of TADs called by each method as a function of the number of reads retained by the filtering step in all data sets except Lieberman-Aiden *et al.*[2] and Jin *et al.*[8] H1-hESC ($n = 36$; **Supplementary Table 1**). Different points represent sample replicates. Loess interpolation for each method is shown as solid line. (**b**) Box plot of median TAD size in all replicates of all data sets (analyzed at 40-kb resolution) except Lieberman-Aiden *et al.*[2] and Jin *et al.*[8] H1-hESC ($n = 36$). (**c**) Heat map of the contact matrix of Rao *et al.*[9] GM12878 replicate H (chr1:153,000,000–155,500,000) at 40-kb resolution. Identified TADs are framed in different colors for the various methods. Obs, observed counts. (**d**) Box plots of the Jaccard Index for concordance of TAD boundaries between sample replicates of all data sets with at least two replicates ($n = 39$).

biological replicates. At all resolutions, HiCseg[20] had the highest reproducibility among replicates of the same data set (intra-data set; **Fig. 3d** and **Supplementary Fig. 9a**). In the majority of comparisons, the reproducibility of TAD boundaries was higher (median JI of 0.25) than what was observed for chromatin interactions. The reproducibility increased with the number of reads for all methods when grouping samples based on increasing number of reads (**Supplementary Fig. 9b**). TADs identified by HiCseg[20] were also the most reproducible when using the overlap coefficient (**Supplementary Fig. 9c**).

The intra-data set reproducibility remained similar for most tools when using different restriction enzymes for the same cell line (**Supplementary Fig. 10**). However, the inter-data set concordance (i.e., between TAD boundaries called in replicates of the

same cell line in different data sets obtained using different protocols and enzymes) was lower than the intra-data set reproducibility, with TADtree[23] showing the highest and Arrowhead[9,14] the lowest inter-data set concordance (**Supplementary Fig. 11**).

The various tools called TADs with consistent enrichment of insulators (e.g., CTCF or BEAF32; **Supplementary Table 9**) at the TAD boundaries. In almost all data sets, more than 50% of TAD borders overlapped CTCF peaks (**Supplementary Table 10**). Moreover, all tools identified TADs with an enrichment of CTCF peaks at the TAD borders, with Armatus[24] and TADtree[23] returning domains with a stronger CTCF enrichment at their borders (**Supplementary Fig. 12a**). In the Sexton *et al.*[7] data set, most tools returned TADs with a clear enrichment (at TAD borders) of BEAF32, an architectural protein reported to be more enriched than CTCF at TAD boundaries in *Drosophila*[7] (**Supplementary Fig. 12b**).

When using synthetic data, DomainCaller[5], TADbit[21], and InsulationScore[22] identified a number of TADs comparable to the number of simulated not overlapping TADs, irrespectively of the noise (**Supplementary Fig. 13a**). As with experimental data, HiCseg[20] called a small number of large TADs, whereas TADtree[23] identified a large number of small TADs (**Supplementary Fig. 13b**). The ability of both methods to identify the correct structures was strongly affected by the noise present in the data (**Supplementary Fig. 13c,d**). TADbit[21] and Armatus[24] had the highest sensitivity in recovering TAD boundaries, although TADbit[21] displayed a higher precision (low FDR) at all noise levels. These results hold similar when simulating a hierarchy of nested TADs, while the precision of TADtree[23], specifically designed to identify nested domains, ameliorated in the latter case (**Supplementary Fig. 13e,g**).

Other analyses

In additional analyses, we compared the performances of interaction callers using a common preprocessing procedure (**Supplementary Note 5 and Supplementary Fig. 14**) and we assessed the computational requirements, running time, and usability of all tools (**Supplementary Note 6** and **Supplementary Fig. 15**).

DISCUSSION

The performance of algorithms in the identification of chromatin interactions and TADs from Hi-C data have been, in most cases, compared using semiquantitative approaches[19,20,23,24]. Indeed, a robust quantification of performance in terms of specificity and sensitivity is hindered by the lack of ground-truth-positive and ground-truth-negative controls for chromatin architecture and by conceptual difficulties in designing simulators of Hi-C data. To overcome these limitations, we adopted a framework that uses a large set of experimental and synthetic data and exploits various metrics to quantitatively compare the performance of several tools currently available for the analysis of Hi-C data.

Based on this comparison framework, our results indicate that there is no algorithm that can be considered the gold standard to identify chromatin interactions. Independently of the data resolution, the choice of the method impacts the quantity and characteristics of the identified interactions.

While Hi-C replicates are commonly pooled before the analysis to generate a unique sample with higher number of reads, we here kept replicates separated to quantitatively assess the concordance of identified interactions. Surprisingly, interactions called in one replicate were poorly conserved in other replicates from the same cell type of the same study. The overall low reproducibility may be partly explained by the fact that biological replicates, being an ensemble of cells in different states and phases of the cell cycle, are not necessarily identical in terms of chromatin contacts, as hypothesized when quantifying reproducibility in terms of the co-occurrence of the same point interaction. Notwithstanding the limited reproducibility, all methods detected comparable, statistically significant proportions of *cis* promoter–enhancer looping interactions and a very small quantity of interactions classified as biologically less plausible.

In agreement with what recently reported by Dali and Blanchette[35], TAD callers returned different numbers of TADs with different mean sizes. However, predicted TADs were more comparable than loops among replicates and were characterized by enrichment in binding sites of known architectural proteins.

Overall, our comparison suggests that, although no single method outperforms others in all situations, TAD callers are methodologically more mature than interaction callers. Among TAD callers, TADbit[21], Armatus[24], and TADtree[23] had balanced performances for most metrics in experimental and simulated data. For interaction callers, HOMER[17] and HiCCUPS[9,14] yielded the highest proportion of interactions with a potential biological significance—although the potential of HiCCUPS[9,14] (e.g., in terms of absolute number of called interactions) could be fully exploited only in the analysis of very high-resolution data sets.

We encountered difficulty in reconciling the results obtained from experimental and synthetic data, especially for interaction callers. This can be most likely ascribed to the complexity of designing sound strategies to simulate Hi-C data sets with predefined features that represent well-defined and unambiguous true positives and negatives. Although several promising approaches are available from the biophysics of polymer folding modeling[36], no algorithm has been proposed so far to generate reads that fully mimic the distribution and biases observed in real Hi-C data. The availability of synthetic data will be essential to rationally tune any algorithm parameter, thus limiting the heuristics currently inherent in the choice of the best setting.

The various tools greatly differed in terms of usability, interoperability, stability of the implementation, and computing resources required to complete the analysis. Considering the pace of data production, priorities for developers should be the deployment of methods able to analyze larger and higher resolution data sets with reasonable amounts of computational resources and the adoption of common data formats to easily exchange inputs and outputs among the various tools[37].

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by AIRC Special Program Molecular Clinical Oncology "5 per mille" (to S.B.); by AIRC Start-up grant 2015 N.16841 (to F.F.); and by Italian Epigenomics Flagship Project (Epigen) (to S.B.). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (grant agreement no. 670126-DENOVOSTEM to S.B. and M.F.) and from CINECA (ISCRA Class C project

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

HP10CDMGT8 to M.F.). C.M.L. is supported by SIPOD (Structured International Post Doc program of SEMM), a Marie Curie cofunded fellowship. We thank A. Lun (University of Cambridge) for sharing the code used to simulate Hi-C data in the diffHic article. We thank F. Fanelli (Dept. of Life Sciences, University of Modena and R. Emilia) and the center for scientific computing of the University of Modena and R. Emilia for the use of GPUs. We thank M. Cordenonsi (Dept. of Molecular Medicine, University of Padova), P. Maiuri (The FIRC Institute of Molecular Oncology, IFOM), E. Sebestyen (The FIRC Institute of Molecular Oncology, IFOM), and M. Morelli (Center for Genomic Science, Istituto Italiano di Tecnologia IIT) for critical feedback on the manuscript. We would also like to thank the authors of all the tools compared for providing support for their methods and for prompt replies to our inquiries.

AUTHOR CONTRIBUTIONS

M.F., C.N., and K.P. collected the experimental data and implemented the computational pipelines. M.F., C.N., K.P., and C.M.L. analyzed the Hi-C data sets. M.F. and C.N. compiled the list of interaction evidences. F.F. generated the simulated data. M.F., F.F., and S.B. designed the experiments and analyzed the results. M.F., C.N., F.F., and S.B. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
2. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
3. Pombo, A. & Dillon, N. Three-dimensional genome architecture: players and mechanisms. *Nat. Rev. Mol. Cell Biol.* **16**, 245–257 (2015).
4. Cavalli, G. & Misteli, T. Functional implications of genome topology. *Nat. Struct. Mol. Biol.* **20**, 290–299 (2013).
5. Dixon, J.R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
6. Nora, E.P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
7. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
8. Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290–294 (2013).
9. Rao, S.S.P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
10. Schmitt, A.D., Hu, M. & Ren, B. Genome-wide mapping and analysis of chromosome architecture. *Nat. Rev. Mol. Cell Biol.* **17**, 743–755 (2016).
11. Ay, F. & Noble, W.S. Analysis methods for studying the 3D architecture of the genome. *Genome Biol.* **16**, 183 (2015).
12. Mora, A., Sandve, G.K., Gabrielsen, O.S. & Eskeland, R. In the loop: promoter-enhancer interactions and bioinformatics. *Brief. Bioinform.* **17**, 980–995 (2016).
13. Shavit, Y., Merelli, I., Milanesi, L. & Lio', P. How computer science can help in understanding the 3D genome architecture. *Brief. Bioinform.* **17**, 733–744 (2016).
14. Durand, N.C. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
15. Ay, F., Bailey, T.L. & Noble, W.S. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res.* **24**, 999–1011 (2014).
16. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).
17. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
18. Hwang, Y.C. *et al.* HIPPIE: a high-throughput identification pipeline for promoter interacting enhancer elements. *Bioinformatics* **31**, 1290–1292 (2015).
19. Lun, A.T.L. & Smyth, G.K. diffHic: a Bioconductor package to detect differential genomic interactions in Hi-C data. *BMC Bioinformatics* **16**, 258 (2015).
20. Lévy-Leduc, C., Delattre, M., Mary-Huard, T. & Robin, S. Two-dimensional segmentation for analyzing Hi-C data. *Bioinformatics* **30**, i386–i392 (2014).
21. Serra, F., Baù, D., Filion, G. & Marti-Renom, M.A. Structural features of the fly chromatin colors revealed by automatic three-dimensional modeling. Preprint at http://dx.doi.org/10.1101/036764 (2016).
22. Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).
23. Weinreb, C. & Raphael, B.J. Identification of hierarchical chromatin domains. *Bioinformatics* **32**, 1601–1609 (2016).
24. Filippova, D., Patro, R., Duggal, G. & Kingsford, C. Identification of alternative topological domains in chromatin. *Algorithms Mol. Biol.* **9**, 14 (2014).
25. Dixon, J.R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
26. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
27. Yaffe, E. & Tanay, A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat. Genet.* **43**, 1059–1065 (2011).
28. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
29. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
30. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
31. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
32. Sauria, M.E.G., Phillips-Cremins, J.E., Corces, V.G. & Taylor, J. HiFive: a tool suite for easy and efficient HiC and 5C data analysis. *Genome Biol.* **16**, 237 (2015).
33. Roadmap Epigenomics Consortium. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
34. Ho, J.W.K. *et al.* Comparative analysis of metazoan chromatin organization. *Nature* **512**, 449–452 (2014).
35. Dali, R. & Blanchette, M. A critical assessment of topologically associating domain prediction tools. *Nucleic Acids Res.* **45**, 2994–3005 (2017).
36. Imakaev, M.V., Fudenberg, G. & Mirny, L.A. Modeling chromosomes: beyond pretty pictures. *FEBS Lett.* **589**, 3031–3036 (2015).
37. Dekker, J. *et al.* The 4D nucleome project. Preprint at: http://dx.doi.org/10.1101/103499 (2017).

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

ONLINE METHODS

Hi-C data analysis tools. We chose algorithms that were (i) specifically designed for the identification of chromatin interactions and TADs and (ii) implemented as a publicly available tool at the time of our survey (July 2016). An extended description of the methods is provided in **Supplementary Notes 1** and **2**.

Among the tools used to identify chromatin interactions, Fit-Hi-C[15] uses spline models to estimate expected contact probabilities as a function of distance. Statistical significance of interactions is calculated using a binomial distribution and *P* values corrected for multiple testing. Fit-Hi-C[15] requires as input a raw count interaction file and a bias file calculated with an implementation of ICE, the iterative correction technique from Imakaev *et al.*[31]. As output, Fit-Hi-C[15] returns only *cis* interactions characterized by contact count, *P* value, and FDR. Significant interactions have been selected based on the FDR.

In GOTHiC[16], significant chromatin interactions are identified using a binomial test followed by Benjamini–Hochberg multiple testing correction. GOTHiC[16] takes aligned reads as input and performs read-pair-level filtering and square root of vanilla coverage normalization (a type of implicit normalization). For all interactions (*cis* and *trans*), the algorithm outputs the $\log_2$ ratio of observed to expected interactions, *P* value, FDR, and the number of supporting read pairs. Here, we used FDR and contact counts to identify significant interactions[38].

HOMER[17] performs a binomial test to find significant interactions. The input file is in the form of aligned reads; filtering is at read- and read-pair level; the implicit normalization method is based on region coverage and distance between regions. All interactions (*cis* and *trans*) are characterized in terms of *P* value, FDR, number of supporting read pairs (both observed and expected), and interaction distance. Significant interactions are called setting a threshold on the *P* value.

HIPPIE[18] implements an approach similar to the one presented in Jin *et al.*[8] to call interactions. Significant interactions are detected by fitting a negative binomial distribution, where the expected random contact frequency (mean) is estimated from GC content, mappability, fragment length, and distance; and the overdispersion parameter is fixed and derived from Jin *et al.*[8]. HIPPIE[18] starts from sequencing reads and performs chimeric alignment; read-, read-pair- and fragment-level filtering; and explicit normalization without binning. The output is a set of restriction-fragment-based interactions (interchromosomal and intrachromosomal) with an associated *P* value. Significant interactions have been selected setting a threshold on the *P* value.

diffHic[19] takes raw sequencing data as input and performs chimeric alignment as well as read- and read-pair-level filtering. Significant interactions (*cis* and *trans*) are identified from the raw contact matrix using a *local* approach—i.e., searching for bin pairs that have substantially more reads than their neighbors, an approach conceptually similar to HiCCUPS[9,14]. The enrichment value for each interaction is calculated as the log-fold change between the abundance (number of read pairs) of the target bin pair and the region of the neighborhood with the largest abundance. Here, we set thresholds on the enrichment, on the number of supporting reads, and on the distance from the diagonal to call interactions. When calling interactions on individual samples, no statistical test is performed, and no significance value is returned.

HiCCUPS[9,14] is part of the Juicer software suite, a pipeline used to process and analyze Hi-C data that starts from the raw sequencing files and generates normalized contact matrices at several resolutions. The pipeline aligns raw reads from FASTQ files using the Burrows–Wheeler aligner (BWA[29]) algorithm, pairs the reads, handles chimeras, and merges and sorts the reads to filter out PCR duplicates. Juicer Tools Pre is used to create the normalized Hi-C contact matrix (.hic file) from the filtered read pairs. HiCCUPS[9,14] takes as input the normalized Hi-C contact matrix to identify chromatin interactions. Specifically, HiCCUPS[9,14] calls only *cis* interactions detecting pixels enriched with respect to four neighboring areas given the width of the peak and the window size as described in Rao *et al.*[9]. It returns the centroid of the clusters of significant peaks called using a modified Benjamini–Hochberg FDR.

Since most of the tools to identify TADs lack the preprocessing steps, to maximize comparability we used a common pipeline based on the scripts of hicpipe[27] to align, filter, and normalize the data used as input to the TAD callers.

HiCseg[20] performs a 2D segmentation based on a maximum likelihood approach to partition each chromosome in its constituent TADs directly from raw or normalized contact matrices. Here, we applied HiCseg[20] to the raw Hi-C data.

TADbit[21] implements a breakpoint detection algorithm that identifies the optimal segmentation of the chromosome under a Bayesian information criterion (BIC)-penalized likelihood. TADbit[21] requires as input the observed read counts, which are then normalized using a modified implementation of ICE[31]. Read counts were obtained using hicpipe[27], even though TADbit[21] includes an alignment module (based on the genome multitool (GEM) mapper for iterative alignment) and implements several filters.

DomainCaller[5] is a single-scale algorithm that identifies TADs using a hidden Markov model on the directionality index. The directionality index is a score quantifying the bias in downstream, as compared to upstream, contact probabilities for each bin within a user-defined window of maximum distance. No preprocessing step is directly implemented by DomainCaller[5], which thus requires an external preprocessing tool to prepare the normalized contact matrix.

The InsulationScore[22] is a segmentation algorithm that identifies TADs within normalized Hi-C matrices using a sliding square (insulation square). It combines contact signals inside the square and assigns an insulation score to each bin along the diagonal, thus obtaining a 1D insulation vector. TAD boundaries are then identified based on the insulation vector.

Arrowhead[9,14] is part of the Juicer suite of tools for Hi-C data analysis and visualization. The tool is based on the Arrowhead[9,14] transformation of the Hi-C contact matrix, which results in translating the patterns of TAD domains from 'squares' along the diagonal to 'triangles' of high or low signal. For each pair of loci, the algorithm computes specific scores for the 'triangles' designed around the pair to assess their potential as TAD boundaries, thus exploring the definition of TADs at multiple scales.

Like Arrowhead[9,14], TADtree[23] can also identify nested TADs. It is based on a 1D boundary index similar to the one developed by Sauria *et al.*[32]. The algorithm is based on the observation that the average enrichment of intra-TAD contacts grows linearly with distance; but when a TAD lies inside another one, its enrichment

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

doi:10.1038/nmeth.4325

grows at a faster rate. The best TAD hierarchy is determined using a dynamic programming algorithm. No preprocessing step is directly implemented in TADtree[23], which thus requires an external preprocessing and normalization pipeline.

Armatus[24] adopts a multiscale approach that can identify a consensus set of domains across various resolutions. It is based on a score function that quantifies the quality of a domain based on its local density of interactions. Since Armatus[24] does not directly implement a preprocessing step, it requires a complete preprocessing pipeline to generate the normalized contact matrix.

For each method, we used the default statistical thresholds or the values suggested in the accompanying documentation to identify chromatin interactions or TADs (*P* values or FDR). Only in the case of HIPPIE[18], to guarantee a statistical significance comparable to that of the other tools, we adopted a threshold (*P* value < 0.01) more conservative than the one suggested in the original publication (*P* value < 0.1; see **Supplementary Note 1**).

GOTHiC[16] and HiCseg[20] were run in R (version 3.1.3), while for diffHic[19] (which requires at least R (version 3.2.0)) we used R (version 3.2.0). We used Python version 2.7.

Experimental Hi-C data. We selected nine Hi-C data sets from six studies obtained with three protocols in overlapping cell types and analyzed at different resolutions, primarily determined by the restriction enzyme and sequencing depth ($n = 41$ samples; **Table 2** and **Supplementary Table 1**). Data were generated using dilution Hi-C (i.e., the original Hi-C protocol published in Lieberman-Aiden *et al.*[2]), simplified Hi-C (introduced in Sexton *et al.*[7]), and *in situ* Hi-C (developed by Rao *et al.*[9]). Samples comprise human cell lines from various tissues (embryonic stem cells, H1-hESC; fetal lung fibroblasts, IMR90; lymphoblastoid cell lines (LCL), GM12878 and GM06990) and *D. melanogaster* embryos. All data were obtained using 6-bp or 4-bp cutter restriction enzymes. Some replicate samples from Lieberman-Aiden *et al.*[2] and Rao *et al.*[9] GM12878 were processed with both restriction enzymes.

All biological replicates have been analyzed separately. In particular, the Rao *et al.*[9] GM12878 data set contained 26 samples obtained with *in situ* protocol and MboI restriction enzyme and divided into a primary (16 technical replicates of one sample) and a replicate experiment (10 biological and technical replicates; see Supplementary **Table 1** of Rao *et al.*[9]). Here, we selected the replicate with the highest number of sequenced reads from the primary experiment (i.e., SRR1658572, originally labeled as HIC003 and renamed here as replicate H) and all the *in situ* samples of the replicate experiment. Moreover, we analyzed as separate samples the technical replicates of the replicate experiment, since the authors defined as technical replicates also those samples for which cells were crosslinked together but processed independently (**Supplementary Table 1** and Supplementary Table 1 of Rao *et al.*[9]). In the Jin *et al.*[8] study, it must be noted that the H1-hESC sample—originally composed of SRR639047, SRR639048, and SRR639049 and here renamed as replicate A—is the same H1-hESC sample used by Dixon *et al.*[5] (which was composed of SRR442155, SRR442156, and SRR442157 and is here renamed as replicate B) (**Supplementary Table 1**). Both H1-hESC samples from Jin *et al.*[8] and Dixon *et al.*[5] were analyzed with chromatin interaction callers at their original resolutions (5 and 40 kb, respectively); while we used only the H1-hESC sample from Dixon *et al.*[5] for the TAD analysis, which was conducted at 40 kb for all data sets.

Preprocessing of experimental data. For most of the interaction callers, we used the specific preprocessing procedure incorporated in the tool. Instead, with the exception of TADbit[21] and Arrowhead[9,14], all TAD callers require as input a fully preprocessed interaction matrix. Thus, for a fair comparison of methods performances, we used the same preprocessing procedure to prepare the data for all TAD callers.

Reads were aligned to the hg19 build of the human genome or dm3 of the fly genome using (i) Bowtie[26] (v.1.1.1) in single-end mode with parameters: -m 1 -a–best–strata–chunkmbs 200; (ii) Bowtie 2 (ref. 30) (v2.2.4) as implemented by diffHic[19], (iii) STAR[28] (v2.4.0) as implemented by HIPPIE[18], and (iv) BWA[29] (v0.7.15) as implemented by HiCCUPS[9,14]. Bowtie[26] performs full read alignment; whereas diffHic[19], HIPPIE[18], and HiCCUPS[9,14] implement different approaches for chimeric alignment (**Supplementary Note 1**). Reads aligned with Bowtie[26] were used as input to those interaction callers lacking a specific aligner and to all TAD callers. In particular, for interaction callers, this choice of input was dictated by constraints in the type of input required by GOTHiC[16] and HOMER[17] that hampered the use of chimeric aligners. After alignment, samples composed of more than one run were merged with SAMtools[39].

Most interaction callers implement their own filtering, binning, and normalization strategy (**Supplementary Note 1**). The filtering step is used to remove low-quality reads, reads that may originate from unspecific ligation events or which are not informative. We grouped filters in three major categories: read level, read-pair level, and fragment level. Read-level procedures filter reads based on read mapping quality (AQ) and restriction site proximity (RSP). Read-pair-level filters remove PCR duplicates (PD), spikes (S, reads aligning on a region with an abnormally high quantity of reads), and read pairs that derive from undigested chromatin (UC). This last filter can also consider strand orientation to identify potential self-ligation or no ligation events (UC + SLF). Restriction-site-proximity filter can also be performed at read-pair level. Finally, fragment-level filters (FLF) discard fragments based on the restriction-site proximity of their reads. Reads have been filtered according to the strategy implemented by each tool. We also filtered out reads aligning on chrY and chrM for hg19 and on chr4, chrY, and all heterochromatic chromosomes for dm3.

In almost all cases, we set the bin size equal to the highest resolution reported in the original publications. However, due to severe computational requirements, we analyzed the Jin *et al.*[8] data set and GM12878 samples of Rao *et al.*[9] at 5 kb with interaction callers and all data sets originally binned at less than 40 kb (Jin *et al.*[8], Rao *et al.*[9], and Dixon *et al.*[25]) at 40 kb with TAD callers.

All tools were run using default or suggested values for preprocessing parameters, filters, and normalization type. In some cases, parameters were adjusted according to the adopted resolution, following suggestions from the software documentation or directly from the developers (**Supplementary Notes 1** and **2**). Some of the steps in the preprocessing workflow have been adapted to the requirements of the specific tools. In particular, since Fit-Hi-C[15] requires as input raw interactions, we used GOTHiC[16]—whose output format can be easily adapted to Fit-Hi-C[15] input to perform filtering and binning. The binning step was not required

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

for HIPPIE[18], which calls interactions directly at the restriction-fragment level. However, when using diffHic[19] for calling interactions in individual samples, the normalization step was not performed, since it was not required.

For all TAD callers, we used hicpipe[27] for filtering and binning. hicpipe[27] was also used for normalization in all TAD tools, with the exceptions of TADbit[21] (which requires the use of its internal normalization method) and HiCseg[20] (which was applied to the raw interaction matrix) (see **Supplementary Note 2**).

Simulated Hi-C data. We generated the simulated data using a modification of the procedure proposed by Lun and Smyth[19] for a total of 65 samples obtained by varying the level of base interaction strength (for interactions only) and of noise (for TADs only; **Supplementary Note 3**). The simulated Hi-C count matrices were used as input to the interaction callers (HiCCUPS[9,14], HOMER[17], diffHic[19], and Fit-Hi-C[15]) and to HiCseg[20] and TADbit[21], which require raw count as input. For all other TAD callers, which require observed-over-expected normalized data, the raw count matrices were converted to vanilla coverage matrices as described in Lieberman-Aiden *et al.*[2].

Performance metrics. To assess the performance of interaction callers, we considered several metrics, including the total number of called interactions; the distance between the interacting points in *cis*; the concordance of results within and between data sets when analyzing different biological replicates; and the type of associated chromatin states. To determine a further basis for comparison, we searched the literature for interactions that had been demonstrated to be present (or absent) in the same cell types of the Hi-C data sets. Namely, we selected interactions validated using other 3C techniques (for example, 3C, 5C, ChIA-PET) and 3D-FISH, or reported in the literature to be specific to given cell types at a given physiological state (interaction evidences). Moreover, we calculated the sensitivity (true positive rate) and precision of the methods in identifying interactions from simulated data.

To compare TAD callers on experimental data, we considered the total number of called TADs, the TAD size, the concordance of TAD boundaries within and between data sets when analyzing biological replicates, and the enrichment at TAD boundaries of known boundary elements (i.e., CTCF and BEAF32).

Comparative analyses. The intersection of the results from different replicates has been generated using the R package ChIPpeakanno.

For both interactions and TAD boundaries, the Jaccard Index of two replicates has been defined as the ratio between the size of the intersection and the size of the union of interactions and TAD boundaries called in the replicates. Jaccard Index empirical *P* values were estimated with random permutations of interactions. Namely, for each data set, cell type, and data analysis method we defined, for each sample, a random set of *cis* interactions by keeping constant the sample-specific number of interactions and the sample-specific distribution of distances between anchoring points. The first of the two anchoring points for each interaction was randomly selected from the pool of detectable anchoring points, defined as any genomic bin that was called as anchoring point in any sample from the same data set and cell type. The second anchoring point was randomly defined by sampling from the observed distribution of anchoring point distances. The resulting sets of random interactions were then used to compute random Jaccard Index values in pairwise comparisons. The random sampling of interactions was repeated 1,000 times to obtain a null distribution of randomly expected Jaccard Index values for each pairwise comparison. The empirical *P* value is estimated as the probability of observing a random Jaccard Index value larger than or equal to the observed one.

Rao *et al.*[9] GM12878 replicates were divided into four groups of samples with increasing numbers of filtered read pairs. Specifically, replicates B2, B1, A2, A1, and G1 constituted the group of samples with less than 40 million reads; A3, D, B, and G2 the group with more than 40 and less than 100 million reads; C2, C1, F, and A the group of samples with a number of filtered reads comprised between 100 and 180 million reads; and E1 and E2 the group of samples with more than 180 million reads. Replicate H was not included in any of the above groups.

The overlap coefficient of two replicates was defined as the ratio between the size of the intersection and the size of the minimum set of interactions or TAD boundaries called in the replicates.

For interactions and TAD boundaries identified in simulated data, we defined sensitivity as the ratio of correctly identified features to all true features, and we defined precision as the ratio of correctly identified features to all called features (1 – FDR).

All comparative analyses were run using R-3.1.3. All box plots were generated with the R box plot function and default parameters.

Selection of validated interaction evidences. From the literature, we constructed a list of interactions that (i) had been demonstrated to be present (or absent) in the same cell types of the Hi-C data sets using other 3C techniques (e.g., 3C, 5C, ChIA-PET) and 3D-FISH or (ii) are known to exist in specific cell types at a given physiological state (interaction evidences). Altogether, we selected 2,439 validated true-positive cell-specific interactions, 389 validated true negatives, 61 true-positive evidences, and 138 true-negative evidences (**Supplementary Table 7**). True-positive and true-negative interactions were mapped to the bin level (at 40 kb and 5 kb resolution) and counted only if between bins that were not adjacent.

Integration with genomic data. Chromatin states for IMR90, H1-hESC and GM12878 (15-states model) were downloaded from the Roadmap Epigenomics Consortium[33], and chromatin states for fly late embryos (16 states) from modENCODE[34] (details in **Supplementary Note 7**; https://www.encodeproject.org/comparative/chromatin/). CTCF and BEAF32 ChIP-seq peaks were retrieved from ENCODE[40] and modENCODE[34] (**Supplementary Table 9**). In particular, we considered peaks generated by the uniform analysis pipeline of the ENCODE Analysis Working Group and peaks obtained from combined replicates for modENCODE data.

We used the R package ChIPpeakanno to compare chromatin interactions with chromatin states and TAD boundaries with CTCF and BEAF32 peaks.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

doi:10.1038/nmeth.4325

Code availability. Examples of how to run each tool and functions to analyze results, calculate general statistics, and performance metrics have been deposited in https://bitbucket.org/mforcato/hictoolscompare.

Data availability statement. The Hi-C experimental data used in this study were downloaded from the Sequence Read Archive (SRA) under the accession numbers listed in **Supplementary Table 1**. The Hi-C simulated data are available at https://bitbucket.org/mforcato/hictoolscompare. All data used to generate **Figures 1–3** and **Supplementary Figures 1–15** are provided as source data files. Any other data supporting the findings of this study are available from the corresponding author upon request.

38. Schoenfelder, S. *et al.* The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* **25**, 582–597 (2015).
39. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
40. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.