Circadian variation of gene expression in the liver is essential to temporally coordinate

liver indicates that conformation dynamics at different genomic scales are coupled with circadian transcriptional oscillations.

Circadian A/B chromatin compartments switch between open and closed configurations throughout the day

To study global genome architecture during a circadian cycle, we performed in-nucleus Hi-C (see "Methods") on mouse adult liver at four different timepoints of the circadian cycle (ZT0, 6, 12 and 18), with ZT0 and ZT12 being the start of the light and dark phase, respectively, in three biological replicates. Samples of individual livers were processed in parallel for RNA-seq. We produced high-quality Hi-C data sets with a high percentage of valid pairs (~ 80%), low PCR duplicates (less than 2%), and high cis:trans interaction ratios obtained (~ 80:20%) (Table S1). In total, we obtained ~ 2 billion valid Hi-C read pairs from mouse adult liver across a circadian cycle (Table S1).

To detect "open," transcriptionally active and "closed," silent genomic compartments (A and B compartments, respectively), we performed PCA analysis on Hi-C data at different timepoints throughout the circadian cycle, at 100-kb bin resolution. PCA analysis is used to analyze high dimensional data and we redefine them, with as few dimensions as possible that explain most of the variance in the data. As such, the 1st principal component (PC1) will explain the majority of the variance, followed by the second component (PC2) and so forth. When applying PCA to the normalized contact Hi-C matrix from individual chromosomes, important features can be identified. For most chromosomes, the PC1 value reflects two distinct interaction compartments that correspond to open and closed chromatin [10]. Changes in chromatin compartments have been associated with changes in transcription and chromatin states during cell differentiation and mouse early development [11, 12]. As expected, PC1 values partitioned the liver genome into chromatin compartments (Fig. 1a,b, Additional file 1: Figure S1A,B). We then compared the eigenvectors of the different timepoints and identified changes in the sign of regional PC1 values, indicative of compartment switching between all timepoint pairs (Fig. 1a,b, individual replicates and merged replicates, respectively, Additional file 1: Figure S1C one-way ANOVA p value < 2e – 16). These genomic regions, termed oscillatory chromatin compartments (OCCs) spanned 440.4 Mb of the mouse genome. The rest of the genome (82.7%) retained the same compartment identity during the 24-h cycle (Additional file 1: Figure S1A,B individual replicates and merged replicates, respectively, S1D). We found OCCs with compartment assignments ZT0 = A, ZT6 = A, ZT12 = B, ZT18 = A (AABA) being the most abundant type in the genome covering 194.7 Mb (Additional file 1: Figure S1E).

To anPCA matin comp41-8.8(g)-9(nts)-326.10001(have87096.20001(ZT0)-13ription)-309.70001(and).446,ngmati(d

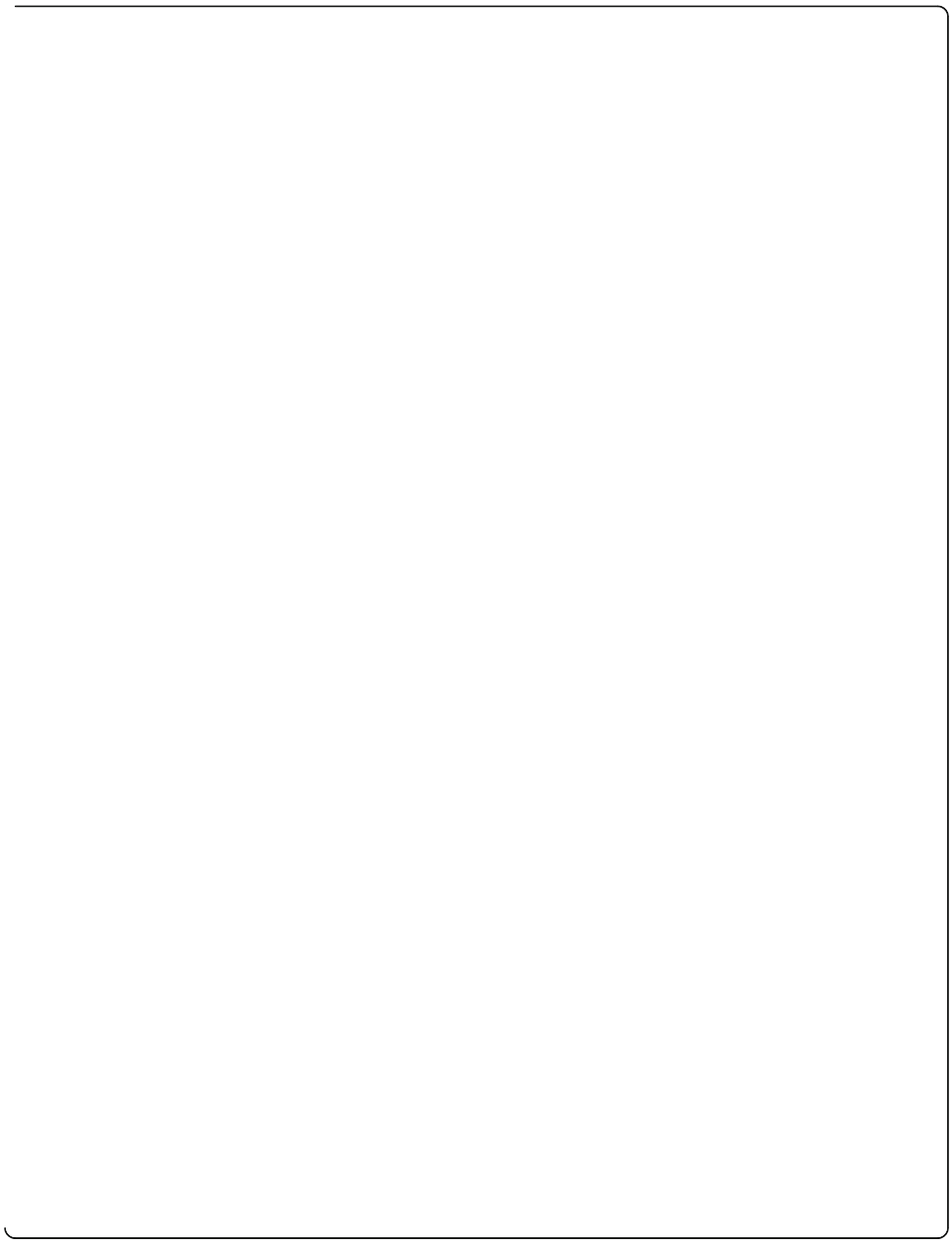
expected oscillatory expression pattern for examples of both core-clock and output circadian genes in the liver (Additional file 2: Figure S2C,F). Gene Ontology and KEGG pathway analysis identified circadian rhythm and metabolism as significantly enriched categories in our identified circadian gene set (Additional file 2: Figure S2D,E)

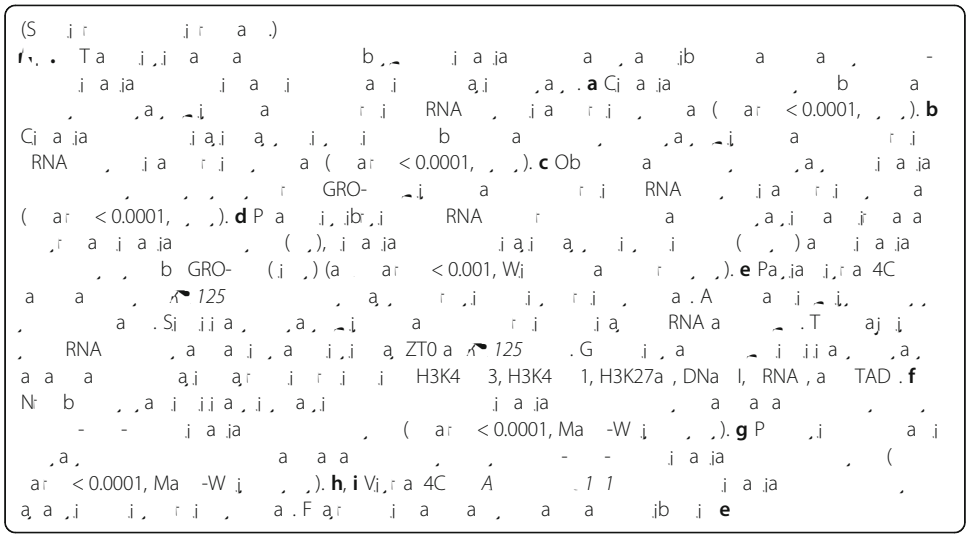
properties, independent of the timepoint examined (Fig. 2b right panel and Additional

either ZT0 or ZT12 Hi-C data suggesting large agreement of CTCF chromatin occupancy and insulation properties between mESCs and adult liver tissue (Additional file 4

dots represent circadian gene promoters). Nevertheless, circadian promoters establish significantly more contacts among them compared to non-circadian gene promoters (Additional file 6: Figure S6B. Above, comparison of the number of edges formed between circadian promoters compared to a random set of non-circadian gene promoters. Below, Z-scores compared to the random sampling of non-circadian promoters). Next we looked at the reads supporting significant interactions between circadian gene promoters in contrasts with interactions between non circadian gene promoters. The result shows that circadian promoter-promoter contacts are more robust compared to non-circadian promoter-promoter contacts (Additional file 6: Figure S6C, p values < 0.001, Mann-Whitney test). Finally, we compared the time of maximal mRNA abundance of

corresponding to genes oscillating at the intronic level (see “Methods”), we found a significant preference for circadian gene promoters to contact with enhancers both detected by eRNA transcription or histone modifications, as well as regions occupied by core clock TF (Fig. 3





The Tcfap2c/AP-2 gamma binding motif was found to be highly enriched at dynamic interacting regions of circadian gene promoters. In the liver, Tcfap2c has been associated with repression of fatty acid synthesis pathways [37] and was identified as a key TF involved in lipid droplets biogenesis [38]. Fos:Jun/AP1 binding motifs were found in genomic regions forming stable contacts with circadian promoters. AP1 factors are a well-characterized immediate-early transcription factors induced in response to signals in the serum and that regulate the expression of circadian genes in liver and cultured cells [39] as well as the suprachiasmatic nucleus [40–42]. Recently, AP-1 was shown to bring together key genes and enhancers through stable and dynamic loops during macrophage development bringing together key macrophage genes and enhancers [43].

In summary, a set of DNA binding motifs for distinct liver nuclear receptors and immediate early genes are enriched in regions contacting circadian promoters and could function in the wiring of the circadian promoter 3D interactome in the liver.

Circadian gene promoters interact with diurnal and nocturnal enhancers in the nuclear space

A set of enhancers are transcribed in a circadian fashion in the mouse liver [27]. We found that these enhancers preferentially contact circadian gene promoters, suggesting that rhythmically transcribed genomic regions, protein-coding and non-coding, interact with each other in the nuclear space (Fig. 4a). This is also true for our subset of circadian genes oscillating at the intronic level (see "Methods") and for circadian genes detected through GRO-seq [27] reflecting primary transcriptional oscillation (Fig. 4b,c).

We then compared the transcriptional phases between promoters of circadian genes and their corresponding contacted enhancer elements with rhythmic transcription. To do so, we separated the circadian gene promoters into diurnal and nocturnal depending on their transcriptional acrophase and then analyzed the time of maximal RNA expression of their contacted enhancer elements. We found a significant contact preference between diurnal promoters and diurnal enhancers as well as nocturnal promoters and nocturnal enhancers (Fig. 4d, left). These preferences were more pronounced when analyzing our circadian intronic gene set reflecting primary transcription (see

"Methods") (Fig. 4d, center) or detected by GRO-seq (Fig. 4d, right, all p values < 0.0001 Wilcoxon signed rank test). For example, the Rnf125 circadian gene promoter with peak transcription at ZT0 contacts 12 rhythmically expressed enhancers with acrophases between 19 and 1 h during the circadian cycle. Furthermore, as it can be observed, the number of contacts with the enhancers, increase during the acrophase (Fig. 4e).

The core clock gene promoter contacts

Finally, we focused on the genomic interactions formed by circadian core clock gene promoters including Npas2, Clock, Arntl, Cry1, Cry2, Per1, Per2, Rorc, Nr1d1, and Nr1d2 as defined by [44]. Notably, all core clock genes displayed fewer overall contacts compared to a random set of the same number of other circadian genes in the liver (12 vs 19.6 mean number of contacts for core-clock vs other circadian genes, Fig. 4f, $p < 0.0001$, t test). However, the contacts formed by the core clock gene promoters were more dynamic than a random set of the same number of contacts for other circadian genes in the liver (42.3% vs 26.8% mean proportion of dynamic contacts for core clock vs other circadian genes, Fig. 4g, $p < 0.0001$, t test). For instance, the Arntl circadian gene promoter engages in contacts with two enhancer elements at ZT18, the time when Arntl expression increases, in three contacts at ZT0, at time of maximal transcriptional output, and does not engage in contacts at either ZT6 or ZT12, when Arntl1 transcription decreases (Fig. 4h, see Additional file 3: Figure S3 for the expression profile). The Nr1d1 gene promoter engages in more contacts at the gene's maximal time of expression, around ZT6 (Fig. 4i, see Additional file 3: Figure S3 for the expression profile). In contrast to core clock contact profiles (additional examples are shown for Rorc, Nr1d2, Npas2, and Per2 (Additional file 7: Figure S7A-D, see Additional file 3: Figure S3 for expression profiles), promoters of circadian output genes engage in numerous contacts that are constant during the day as exemplified by Dhr3 and Ppp1r3c gene promoters (Additional file 7: Figure S7E and F, see Additional file 3: Figure S3 for

and cell cycle [11, 12, 45]. However, our results reveal that dynamic changes between compartment states occur also within hours and without cells dividing or changing their identity (Fig. 5).

Inside cTADs, circadian genes tend to be alone or sharing the TAD with other circadian genes and regulatory elements that are transcribed at similar times during the day,

circadian genes in the liver (Fig. 5). This is in line with recent candidate-scale 4C chromosome conformation capture experiments for two core clock and output genes [46

Promoter Capture in nucleus Hi-C (Chi-C)

Promoter Capture was performed as previously described [21–23]. Briefly, Biotinylated 120-mer RNA baits were designed to target both ends of HindIII restriction fragments overlapping the Ensembl promoters of protein-coding and noncoding transcripts and UCEs as described in detail in [22]. Promoter Capture was carried out using in nucleus Hi-C libraries derived from three biological replicates at ZT0 0, 6, 12, and 18 with the SureSelect target enrichment system and the biotinylated RNA bait library according to the manufacturer's instructions (Agilent Technologies). After library enrichment, a post-capture PCR amplification step was carried out using the PE PCR 1.0 and PE PCR 2.0 primers (Illumina) with 4–6 PCR amplification cycles as required. In nucleus Hi-C and CHi-C libraries were sequenced on the Illumina HiSeq 2000 platform.

ChIP-seq

For ChIP-seq, liver tissue for two biological replicates at ZT0 and ZT12 was dissected as processed as for Hi-C and then fixed in 1% formaldehyde for 5 min. Chromatin immunoprecipitation was performed as described [21] using 10 μg of α-CTCF (Millipore, 07-729). DNA was purified using Zymo Research DNA purification columns. Sequencing libraries were prepared with the NEBNext ChIP-seq library prep kit (NEB) according to the manufacturer's instructions. DNA was purified using AMPure beads (Agencourt). For quantitative chromatin immunoprecipitation experiments (ChIP-qPCR), liver tissue, collected at ZT0, ZT6, ZT12, and ZT18 per duplicate, was fixed with a 1% formaldehyde for 10 min.

Metaplots The metaplots were created using python custom scripts. Briefly, the script takes a feature of interest and calculates the frequency of interactions around it using as input the KR normalized Obs/Exp Hi-C matrices at different resolutions (10, 25, or 50 kb) from different timepoints (ZT0,6,12,18). The final metaplot is the median value of all the plots for the list of anchors. For the TAD-anchored metaplots, Hi-C normalized matrices at 50 kb resolution were used. Each TAD (see TAD calling) was scaled to fit into 5 bins, and only 1000 TADs from all datasets and using all chromosomes were randomly chosen to reduce computing time. For the CTCF- anchored metaplots, Hi-C normalized matrices at 10 kb resolution were used. Each CTCF peak (see ChIP-Seq analysis) was scaled to fit into a single bin, and only 1000 CTCF peaks identified at ZT0 and ZT12 were randomly chosen to reduce computing time. The matrices generated were plotted using heatmap.2 from the package plots.

TAD calling

For all timepoints, we retrieved Knight-Ruiz normalized contact matrices from Juicer for all chromosomes at 25 kb and 50 kb resolution. TADs were identified using TADtool [50] with the insulation score algorithm. To find appropriate parameters for TAD identification, we called TADs for chromosome 1 across all timepoints using contact matrices at 25 kb and 50 kb resolution and a window size of 100, 150, 155, 175, 195, and 200 kb over threshold values from 70 to 200. For all data sets at 50 kb resolution, we called TADs with a window size value of 200 kb and a threshold value of 140 while for all data sets at 25 kb resolution, we called TADs with a window size value of 100 kb and a threshold value of 76. We found that these parameters show good agreement between identified TADs and visual inspection of Hi-C datasets in Juicer. Of note, visual inspection of Hi-C datasets with TADs identified at 25 kb resolution reveals that these represent sub-TADs contained within TADs identified at 50 kb resolution.

TAD analysis

module of Intervene [51]. A TAD was considered shared between timepoints if more than 80% of the genomic domain region overlapped with a domain from a different Hi-C data set.

Compartment analysis

Compartments were identified applying PCA to the normalized interaction matrices at a 100 kb resolution using Juicer [48]. PCA1 was used to assign A and B compartments. To verify the reproducibility of the compartment call, the PCA analysis was applied on the separate replicates and just the merged data was used for downstream analysis. A custom script and publicly available ChIP-seq BAM files for H3K4me3 [13] were used to set the sign to the compartments identified by Juicer. A total of ~20,000 compartments were identified at each timepoint. We identified significantly changing compartments as those genomic regions with a change in PCA1 across different timepoints consistently in the three biological replicates through a one-way ANOVA test.

Transcription in A and B compartments To relate compartments A and B with transcription, we calculated the log2 RPM (reads per million) values for all regions assigned to compartments A and B per timepoint (ZT0,6,12,18) using SeqMonk (https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/) and RNA-seq BAM files (see "RNA-seq data processing") per timepoint as input and to applied a Kruskal Wallis test for all compartments and a Mann-Whitney test for OCCs. The distribution of log2 RPM values per compartment type at each timepoint is presented as violin plots.

Correlation with histone marks To relate changes in compartment status with the enrichment of histone post-translational modifications, we calculated the RPM (reads per million) values for all regions assigned to compartment A and B per timepoint (ZT0,6,12,18) using SeqMonk and publicly available ChIP-seq datasets for the histone post-translational modifications H3K4me3 and H3K4me1 [28] per timepoint as input and applied a one-way ANOVA test and a Tukey post hoc test.

Correlation with HDAC3 To relate changes in compartment status with the enrichment of HDAC3, we calculated the log2 RPM (reads per million) values for all regions assigned to Compartment A at ZT0 and that change to Compartment B at Z12 using SeqMonk and publicly available ChIP-seq datasets for the histone deacetylase HDAC3 [52] and applied a Wilcoxon test.

Promoter CHi-C

The sequenced reads were processed using HiCUP [47]. The filtering and identification of significant interactions were performed with CHiCAGO [24]. To identify differential interactions, the script implemented by [21] was used. This script can identify the differential interactions from a Promoter Capture Hi-C dataset, using the edgeR package [53] to statistically quantify changes in reads for the interactions. To increase the confidence in dynamic interactions, we filtered the dataset only including baits overlapping circadian genes. To account for the

distance bias in the read count, we also divided the CHi-C interactions into greater or less than 150 kb groups. These preliminary results were filtered by FDR and fold change; both distance regimes were combined. To plot long-range interactions, we used the Washington Epigenome Browser (http://epigenomegateway.wustl.edu/browser/) using the mouse genome version mm9 and as input properly formatted CHiCAGO output files.

Characterization of interacting regions To characterize the type of genomic element that promoters contact derived from our Promoter CHi-C, we calculated the observed/expected number of overlaps between the other ends (the genomic segment interacting with a promoter) and a set of genomic regions occupied by Transcription Factors or enriched for histone post-translational modifications using a custom python script. The

cycle) [27] as well as the promoter of the CHi-C datasets using our RNA-seq analysis. To make eRNA phases [27] more comparable to the circadian promoters identified by our RNA-seq analysis, we grouped them into eight groups each containing three time-points. First, we mapped the osceRNAs to the other ends of the CHi-C, then retrieved the bait fragment associated with that eRNA and filtered the fragments overlapping with the set of circadian promoters identified by our RNA-seq analysis. Then, we di-

view). Bam files were sorted (samtools sort) and indexed (samtools index). Duplicates were removed with Pickard. Bam files were imported to deepTools v3.3.1 [58] to create signal tracks with bamCovera

29. S b JA, K i I, A i T, Ra a S, Ca a D, Gi a i F, a. T a i i a r a i i r a i r i . PL S Bi . 2017;15(4): 2001069. :// i. /10.1371/j r a . bi .2001069.
30. Bai TL, J J, G a CE, N b WS. T MEME S i . N i A i R . 2015;43:39–49.
31. C F, Ma L, Da PA, Si a CJ, S a E, G a FJ, a. Li r -1 ia i - a i - i i bi a i - a i ba r a i a i a ir - bi a i a . J Bi C . 2003;278(22):19909–16. :// i. /10.1074/jb .M207903200.
32. C HK, Bi i J, S Y-K, Xi X, O b TF. G - i a a i a i LRH-1 a a bi i a r a i r a i i i ab i i i FXR. BMC G i . 2012;1(13):51.
33. Ma r r a KE, Wa L, B MK, O b TF. A a r a i r -1 i a i a i a a i a b i X . J Bi C . 2007;282(28):20164–71. :// i. /10.1074/jb .M702895200.
34. W N, Ki KH, Z r Y, L J, K NM, Ma JL, a. S a i a (NR0B2) i a r i i ai a i a ia i i . M E i . 2016;30(9):988–95. :// i. /10.1210/ .2015-1295.
35. S a J, Sa P a T, G A, D J, L i M, B r T, a. P a a i a LRH-1/N 5a2 i ibi i i i -i a a i r i i a a a a ia i a a i i . C D a Di . 2020;11(2): 154. :// i. /10.1038/ 41419-020-2348-9.
36. Mi a a DA, K a WC, Ca a -Ga i A, S a a M, E r a H, F JC, a. LRH-1 r a a i i i a i a ai ai a a i i i i i a i i i . JCI I i . 2018;3(5): 96151. :// i. /10.1172/j i i i .96151.
37. H D, K b P, W i T, E A, B A, H S, a. T a i i T a 2 /AP-2 a a r i i ai r a i i a a ia. PL S O . 2011;6(7): 22034. :// i. /10.1371/j r a . .0022034.
38. S CC, V i S, R r J, G r b J. TFAP2 a i i a a r a i i bi i . ELi . 2018;7: 36330. :// i. /10.7554/ Li .36330.
39. Ba a b A, Da i a F, S ib U. A r i r i a ia i i a a ia i r r r . C . 1998;93(6):929–37. :// i. /10.1016/S0092-8674(00)81199-X.
40. C Y, Ya a r i Y, S r i T, D i M, O a r a H. E ai i -F i i r a ia a i r r r j a i i . A a Hi C . 2018;51(2):73–80. :// i. /10.1267/a .18001.
41. G i ME, G r D, D G i L, R b HA, R a B. Ci a ia a i r a i i ia - a i i a r a ia a i r r . N r i . 1999;90(2):555–71. :// i. /10.1016/S0306-4522(98)00467-9.
42. S a WJ, Ca i A, D La I ia HO, Ba R, K i DC, Na ab r Y, a. Di ia r a i a i i a a a ia r b i i i a r a ia a i r r . N r i . 2000;98(3): 535–47. :// i. /10.1016/S0306-4522(00)00140-8.
43. P a i DH, Va B K, S a D, H GT, Saa S a i M, Ma I, a. S a i a a i DNA AP-1 b r a i a i r b r i a a . M C . 2017;67(6):1037–48. :// i. /10.1016/j. .2017.08.006.
44. A a i RC, L Y, Sa TK, V a a a a A, Ra a a a C, Ka a i IH, a. Ma i a i i i CHRONO a a i a ia . PL S Bi . 2014;12(4): 1001840. :// i. /10.1371/j r a . bi .1001840.
45. Na a T, L bi Y, V ai C, D C, L r W, Ba a Y, a. C - a i a a i a i a i - r i . Na r . 2017;547(7661):61–7. :// i. /10.1038/ a r 23001.
46. M J, Y r J, Na F. O i a i a ab i r i a i i i a r i i a ia . PLOS G . 2021;17(2): 1009350. :// i. /10.1371/j r a . .1009350.
47. Wi S, E P, F a -Ma a i M, Na a T, S S, F a P, a. HiCUP: i i a i a i Hi-C a a. F1000R a . 2015; :// i. /10.12688/ 1000 a .7334.1.
48. D a NC, R bi JT, S a i MS, Ma I, M i JP, La ES, a. J i b i a i r a i a i Hi-C a a i r i i . C S . 2016;3(1):99–101. :// i. /10.1016/j. .2015.07.012.
49. D a NC, S a i MS, Ma I, Ra SSP, H MH, La ES, a. J i i a - i a a i - r i Hi-C i . C S . 2016;3(1):95–8. :// i. /10.1016/j. .2016.07.002.
50. K r K, H CB, H -R r B, Va r i a JM. TAD : i r a a a i i i a i TAD- a i a i . Bi i a i . 2016;32(20):3190–2. :// i. /10.1093/bi i a i /b 368.
51. K a A, Ma i A. I : a i i a i r a i a i r i i i . BMC Bi i a i . 2017;18(1):287. :// i. /10.1186/ 12859-017-1708-7.
52. F D, Li T, S Z, B A, M i a SE, A a T, a. A i a ia a b i a a 3 a i i i ab i . S i . 2011;331(6022):1315–9. :// i. /10.1126/ i .1198125.
53. M Ca DJ, C Y, S GK. Di ia i a a i r i a RNA-S i bi i a a ia i . N i A i R . 2012;40(10):4288–97. :// i. /10.1093/ a / 042.
54. Ha b A. A., S r , D. A., & S a P. J. E i r r a i a r i r i N X. 2008. :// . i . / i /S iP 2008/ a _2
55. Vi a P, G R, O i a TE, Hab a M, R T, C r a a D, a. S iP 1.0: r a a a i i i i r i i P . Na r M . 2020;17(3):261–72. :// i. /10.1038/ 41592-019-0686-2.
56. S ab , S., & P , J. S a : i a a i i a i i P . P . O 9 i i . 2010; :// a . r . /.
57. La a B, Sa b SL. Fa a - a ai B i 2. Na r M . 2012;9(4):357–9. :// i. /10.1038/ .1923.
58. Ra F, R a DP, G i B, B a aj V, a. T 2: a a i b - r i a a a a i . N i A i R . 2016;44(W1):W160–5. :// i. /10.1093/ a / 257.
59. T a i H, R bi JT, M i JP. I a i G i Vi (IGV): i - a i a a i r a i a i a i . B i Bi i a i . 2013;14(2):178–92. :// i. /10.1093/bib/bb 017.
60. Z a Y, Li T, M CA, E r J, J DS, B i BE, a. M -ba a a i C IP-S (MACS). G Bi . 2008;9(9):R137. :// i. /10.1186/ b-2008-9-9- 137.