Regulation associated modules reflect 3D genome modularity associated with chromatin activity

The 3D genome has been shown to be organized into modules including topologically associating domains (TADs) and compartments that are primarily defined by spatial contacts from Hi-C. There exists a gap to investigate whether and how the spatial modularity of the chromatin is related to the functional modularity resulting from chromatin activity. Despite histone modifications reflecting chromatin activity, inferring spatial modularity of the genome directly from the histone modification patterns has not been well explored. Here, we report that histone modifications show a modular pattern (referred to as regulation associated modules, RAMs) that reflects spatial chromatin modularity. Enhancer-promoter interactions, loop anchors, super-enhancer clusters and extrachromosomal DNAs (ecDNAs) are found to occur more often within the same RAMs than within the same TADs. Consistently, compared to the TAD boundaries, deletions of RAM boundaries perturb the chromatin structure more severely (may even cause cell death) and somatic variants in cancer samples are more enriched in RAM boundaries. These observations suggest that RAMs reflect a modular organization of the 3D genome at a scale better aligned with chromatin activity, providing a bridge connecting the structural and functional modularity of the genome.

Histone modifications are critical to shaping the chromatin structure and regulating gene expression\(^1\). Active marks such as H3K27ac and H3K4me3 open up chromatin to allow access to transcription factors (TFs) and transcription machinery to promoters or enhancers. Repressive marks such as H3K9me3 and H3K27me3 condensate chromatin to block TF binding and suppress gene expression. DNA marked by active and repressive histone modifications form euchromatin and heterochromatin that are distinct on the compactness. These observations suggest that histone modifications have an important impact on organizing the regional and global 3D genome.

Accumulating evidence has revealed the association of histone modifications with the topologically associating domains (TADs)\(^1\)\(^-\)\(^6\) and compartments\(^7\)\(^-\)\(^8\) derived from the Hi-C contact maps showing plaid patterns. TADs represent genomic domains forming dense internal contacts but fewer contacts with neighboring regions. The TAD boundaries are demarcated with CTCF sites or active transcribed DNA sequences. The Hi-C data also show that the 3D genome is partitioned into transcriptionally active (compartment A) and suppressed (compartment B) compartments. Active and repressive histone marks are enriched, but do not exclusively appear, in the A and B compartments, respectively\(^7\). Computational models have shown that histone modification signals are predictive of Hi-C contacts particularly for enhancer–promoter interactions\(^5\), TAD boundaries\(^6\) and compartments\(^7\). Histone modifications are tightly associated with transcriptional activity\(^9\)\(^-\)\(^14\) while transcription and proteins involved in transcriptional regulation including RNA polymerase and TFs have been shown to contribute to compartmentation and active promoters and enhancers tend to form clusters in the nucleus\(^15\)\(^-\)\(^18\).

---

1Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093-0359, USA. 2Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA 92093-0359, USA. E-mail: wei-wang@ucsd.edu
Despite the mechanisms underlying the interplay between histone modifications and chromatin organization remain elusive, histone modifications can indicate the spatial organization of the genome as readout signals for regulatory modules. However, the current analysis has been limited to associating histone marks to Hi-C derived TADs and compartments. An unfill gap is to use histone modifications to directly elucidate the modular organization of the 3D genome. We propose here to define the spatial module of the genome organization resulting from the chromatin activities reflected by histone modifications.

We find that the frequency profiles of the H3K27ac peaks present a modular structure (referred to as regulation-associated modules, RAMs). A large number of these modules are shared across cell types and can be independently derived using other active histone marks, including H3K4me3 and H3K4me1. We uncover several lines of evidence to support the hypothesis that the RAMs are spatial modules resulting from functional activities: the enhancer–promoter interactions, loop anchors and super-enhancer clusters dominantly occur within RAMs; the extrachromosomal DNAs (ecDNAs) tend to be originated from the same RAMs rather than split in multiple RAMs; RAMs are resistant to cohesin degradation. These properties of RAMs distinguish them from TADs and compartments. Furthermore, deletion of RAM boundaries is predicted to alter the chromatin organization more significantly than the deletion of TAD boundaries by a deep learning model. In addition, deletion of RAM boundaries likely interferes with cellular functions and even affects cell viability. Consistently, the somatic genetic variations in cancer patients are enriched in RAM boundaries, suggesting a possible mechanism of tumorigenesis involved in altering the chromatin modules.

**Results**

**Regulation associated modules (RAMs) detected by the histone modification peaks**

We analyzed the density profile of H3K27ac peaks (i.e. peak count in a sliding window) from chromatin immunoprecipitation assays with sequencing (ChIP-Seq) experiments as using the peak density instead of read count density can better remove the noise from the background signals. We downloaded ChIP-seq data of 93 normal and 19 cancer samples from Roadmap Epigenomics Project (http://www.roadmapepigenomics.org/) and ENCODE portal (https://www.encodeproject.org/) (Supplementary Table 1). Using a sliding window (a fixed flanking size of 500kbp and step size varying from 10kbp to 500kbp), we computed the H3K27ac peak densities in the linear genome. Regardless of the step size, the H3K27ac peak densities were not evenly distributed and showed a modular pattern (Fig. 1a). The active marks of H3K4me1, H3K4me3, and H3K36me3 showed similar peak density profiles to H3K27ac in the 93 samples, indicated by high Pearson correlations between them, whereas the repressive marks of H3K27me3 and H3K9me3 had less consistent patterns (Fig. 1b). Given the highly correlated active mark patterns, we focused on analyzing the H3K27ac signals as the other active marks show similar modular structures (referred to as regulation-associated modules, RAMs). We varied the step size from 10kbp to 500kbp and fixed the window size to 500kbp. It is not surprising that, with the increasing step size, the RAM size increased and a higher percentage of RAMs were shared between samples (Supplementary Figs. 1, 2). We observed that the number of common RAMs in all the chromosomes reached a plateau at 250kbp step size in both normal and cancer samples, which indicates the identified RAMs are most conserved across diverse cell types (Supplementary Fig. 3a, b). We thus used this step size of 250kbp for the remaining analyses. A RAM boundary is called a consensus-RAM (cRAM) boundary if it is shared by >25% of the samples (see “Methods”). This way, 711 cRAMs were detected in the normal samples and 771 cRAMs in the cancer samples (see “Methods”). On average, 60% of the RAMs in a cell type are consensus (referred to as cRAMs) and the remaining cell-type specific (see “Methods”, Supplementary Fig. 3c).

**Characterization of the consensus regulation associated modules (cRAMs)**

As cRAMs are largely shared between diverse cell lines, we further characterized them. Among the cell lines that have both active and repressive marks (all are normal cells), as expected, we found that the cRAM boundaries have a lower peak density of the active marks (H3K27ac, H3K4me3, H3K4me1, and H3K36me3) and slightly higher peak densities of repressive marks (H3K27me3 and H3K9me3) than the cRAM regions (see examples in GM12878 and HUVEC cell lines in Fig. 2a). To quantify the difference of the histone modifications among the cRAM boundaries and the non-boundary regions, we counted the peak density using a sliding window, and compared the histone modifications enrichment across 93 normal samples. The p-value < 0.05 from the Wilcoxon Rank Sum test indicated that cRAM boundaries have significantly lower active marks and higher repressive marks than the non-boundaries of cRAMs (Fig. 2b–g). Furthermore, using the available 10kb-resolution Hi-C data in the K562, GM12878, A549, IMR90, NHEK, HUVEC, HMVEC, and HCT116 cell lines, we found that the cRAM boundaries have significantly more Hi-C contacts (intrachromosomal contacts with log(p-value) ≤ −10) compared to the whole genome (Fig. 2h), which is consistent with the genome browser view for any RAM boundary in Fig. 1c. These observations suggested that the cRAM boundaries are formed by densely packed DNA sequences harboring many 3D contacts.

We next investigated how RAMs are related to the previously identified chromatin modules. First, the median size of cRAMs (~3.3Mbp) is larger than TADs (~600kbp) and one RAM often spans across multiple TADs (Fig. 2i). Second, using the Hi-C data, we identified the A/B compartments at 250 kb resolution (see “Methods”). We calculated the percentage of the A and B compartments in each cRAM (250 kb bin size) across the cell types. While the A compartments account around 50–75% in each of the cRAM, a single cRAM is largely composed of a mixture of A and B compartments, indicating a distinction between cRAMs and compartments (Fig. 3j). Consistently, the cRAM boundaries are enriched with B-compartment but also with a significant portion of A compartments (Fig. 3j). Third, we checked the Lamin-B1 signals for the cRAM boundaries. Lamin-B1 is a scaffolding component of the nuclear envelope. A positive signal for Lamin-B1 suggests a close distance to the nuclear lamina, which could be used to define lamina-associated domain (LAD). When aligning the cRAM boundaries with the Lamin-B1 signals (see “Methods”), we found on average 69% of the cRAM boundaries overlapped with Lamin-B1 signals across the cell types and meanwhile on average 62.7% of the LADs identified from each cell type overlapped with cRAM boundaries (Fig. 2k), indicating that LADs and cRAMs are also different. Taken together, the cRAM boundaries are formed by densely packed DNA.
quences; while they are enriched with B compartment and Lamin-B1 signals, RAMs are clearly distinct from previously reported domain structures such as TADs, LADs, and A/B compartments.

**RAMs are functional units**

If RAMs are functional modules, we reasoned that the majority of the promoter-enhancer interactions should occur within the same RAMs. We downloaded 970 high-confidence promoter-enhancer interactions in the K562 cell line that were experimentally validated in refs. 24, 25 and 95% of them are located within the same RAMs, compared to 75% of them in the same TADs4,6,8,26 (Fig. 3a, b). Two examples of promoter-enhancer interactions are shown in Fig. 3g: the enhancer–promoter interactions of *STEAP1B* and *VGF* are across multiple TADs but within the same K562 RAMs marked by continuous strong H3K27ac peaks. This observation suggests that RAMs may represent regulatory modules and RAM boundaries insulate promoter-enhancer contacts across RAMs at a scale more appropriate than TADs to capture the functional modularity of chromatin activity.

We further investigated how the modularity defined by RAMs affects gene expression. To this end, we examined whether the
enhancer–promoter pairs located within the same K562 RAMs but separated by RAM boundaries in other cells would specifically impact gene expression in K562. When comparing K562 to the normal cell line GM12878, we found 885 K562 enhancer–promoter interactions with in the same RAM in both cell lines and 39 only in K562 (Fig. 3c). The majority of the genes regulated by the 39 enhancer–promoter interactions are upregulated in K562 compared to in GM12878 (p-value = 0.0002 by Hypergeometric Test, see "Methods") (Fig. 3d), indicating that the RAM organization facilitates promoter-enhancer interactions to activate gene expression. For example, the RAB31 promoter interacts with an enhancer that is located within the same RAM in K562 but in a RAM boundary in GM12878 where the enhancer would be silenced in GM12878; the PRELID2 promoter and its interacting enhancer are located within the same K562 RAM but reside in a GM12878 RAM boundary indicating suppression in GM12878 (Fig. 3h). In fact, the RAB31 and PRELID2 normalized expression levels are 34.29 and 5.37 folds higher in K562 than in GM12878, respectively. All the upregulated gene expressions involved in enhancer–promoter interactions that occurred in the same RAM in K562 but in different GM12878 RAMs are shown in Supplementary Table 2. We had a similar observation by comparing K562 and HEPG2: while the majority of the promoter-enhancer pairs are intra-RAM in both cell lines, 28 of them only occur in K562 but in different HEPG2 RAMs (Fig. 3e) and the corresponding genes have higher expressions in K562 than in HEPG2 (p-value = 0.0006 by Hypergeometric Test) (Fig. 3f, Supplementary Table 3). Furthermore, we examined the K562 enhancer–promoter pairs in K562-specific RAMs and the cancer consensus RAMs (cRAMs). 750 pairs were identified as intra-RAM interactions in both K562 RAMs and cancer cRAMs (Supplementary Fig. 4a). 76 genes were involved in the 174 pairs that are only intra-RAM interactions in K562, and 19 of them were detected as K562 specifically highly expressed genes across 92 cancer cell lines (hypergeometric test p-value = 0.03) documented in the Harmonizome database27 (Supplementary Fig. 4b). These observations further illustrated that
RAMs represent modularity directly associated with functional activity of the chromatin. Furthermore, we reasoned that the majority of loop anchors and super-enhancer clusters should occur within the same RAMs. We downloaded the loop anchor data (HCC1599_DMSO, GSI, MB157_DMSO, GSI, WO, and REC1_DMSO, GSI) and the contact domains (HCC1599_DMSO, HCC1599_GSI, MB157_DMSO, MB157_GSI, MB157_WO, REC1_DMSO, and REC1_GSI). As the HCC1599, MB157, and
and RECl cell lines are cancer cells we examined cancer cRAMs for them. We found that on average 80% of loop anchors occur within the same cRAMs (intra-domain loop anchors), compared to only 42% on average within the same TADs (Supplementary Fig. 4c). Similarly, for the super-enhancers identified across 86 human cell and tissue samples, 81% and 84% of super-enhancers are located within the same cancer and normal cRAMs, respectively. (Supplementary Fig. 4d). Taken together, these observations suggested that cRAMs represent modularity of the genome at a scale better aligned with the chromatin activities.

**RAMs are insensitive to cohesin degradation**

Previous studies showed that cohesin degradation would disrupt loop domains and topological associated domains (TADs) but largely not change histone modifications and gene expression. Therefore, RAMs are not expected to be affected by cohesin degradation. For confirmation, we identified RAMs using H3K27ac data in the HCT-116 RAD21-mAC cells untreated and treated for 6 h with IAA. The RAM patterns for each chromosome were highly correlated between treated and untreated cells (Supplementary Fig. 5a) and the recall rate for the RAM boundaries was 0.9 on average for all the chromosomes (Supplementary Fig. 5b). This observation indicates that the RAM formation is independent of cohesin, distinguishing RAMs from TADs and loop domains.

**Extrachromosomal DNA (ecDNA) from cancer patients majorly originated from intra-RAM**

Circular extrachromosomal DNAs (ecDNAs) are prevalent in tumors and their length ranges from 100kbp to megabases, and the genes encoded in the ecDNAs are often amplified in cancers. We reason that if RAMs are functional modules, ecDNAs would form within RAMs because RAM boundaries are highly condensed DNAs that would restrain the transcription of genes residing in ecDNAs. To test this hypothesis, we downloaded the ecDNAs identified from cancer patients, and filtered the ecDNAs corresponding to the median size of the cancer cRAM length (2.5 Mb), i.e. only ecDNAs with size < 2.5 Mb were kept (i.e. 78% of all the ecDNAs). We found that 98% of 2459 ecDNAs were located within the individual RAMs. As a comparison, we performed the same analysis on TADs. We took the conserved TADs defined in the Dixon et al. study and only kept the ecDNAs shorter than 880kbp (68% of all the ecDNAs), the median size of the TAD length. We found that 86% of the 2150 ecDNAs were within individual TADs (Fig. 3j, k). Because the ecDNAs were filtered to have comparable length with the RAM and TAD sizes, respectively, this lower intra-domain percentage for TAD compared to RAM is not due to the larger size of RAMs. Furthermore, GREAT analysis (http://great.stanford.edu/public/html/) on the ecDNAs that fall into intra-cancer cRAMs but split in TADs revealed that they are highly enriched in “positive regulation of DNA replication” with a P-value of 5.364E-19. There are 12 genes involved in this pathway: ATF1, BMP5, BMP6, EGR1, FGFRI, GLI2, IGFI, IL6, JUN, KITLG, PDGFA, PDGFR, which are known important for cell proliferation and cancer pathogenesis. For example, EGR1 is a driver of tumorigenesis. Deregression of the oncogenic FGF signaling has been frequently observed in multiple types of cancers. The PDGF-mediated signaling has been reported to be involved in the cell proliferation and invasion. The observations that ecDNAs tend to originate from intra-RAMs suggest that RAM is a functional module.

**Deletion of the cRAM boundaries is predicted to alter the 3D chromatin structures**

To systematically examine the impact of deleting cRAM boundaries on the chromatin structure, we resorted to computational predictions using a deep learning model ORCA (https://github.com/jzhoulab/orca) as it is prohibitive to perform hundreds of Hi-C experiments with sufficient resolution. We used the ORCA model pre-trained on the high-resolution Hi-C and Micro-C data in Hi-hESC and HFF cell lines to predict 3D chromatin architecture from kilobase to whole-chromosome scale using DNA sequences. It also provided perturbation predictions if certain sequences were deleted. cRAM boundaries shared between cancer and normal samples are apparently important, therefore we selected all 418 of them that are located at least 16 Mb away from the centromere to predict their impacts on Hi-C contacts if deleted. As a comparison, we also included 298 HI-hESC and 187 HFF TAD boundaries (length of the TAD boundaries > = 100kbp; non-overlap with selected cRAM boundaries; away from centromere at least 16 Mb) in the computational perturbations screening. Considering that the cRAM boundaries are often larger than the TAD boundaries, we only deleted the center 100kbp of cRAM and TAD boundaries to avoid bias introduced by deletion size.

To measure the similarity between the deletion and wildtype Hi-C contact matrices, we calculated the Pearson correlation between them. Compared to the TAD deletions, deleting cRAM boundaries obviously resulted in lower correlation coefficients, indicating larger chromatin alterations. At the highest resolutions the ORCA model could predict (4 kb and 8 kb resolutions with Wilcoxon Rank Sum test p-values of 2.4E-11 and 2.6E-7, respectively) (Fig. 4b). Deletion of the cRAM boundary (chr10:115,940,000-116,040,000, in hg38) on Hi-C contacts in HFF and HI-hESC cell lines is shown as an example (Fig. 4a, Supplementary Fig. 6). The 3D contacts are severely weakened by deleting the cRAM boundary in both cell types.

Furthermore, if the RAM formation is important, deletion of cRAM boundaries would likely severely interfere with cellular functions and even affect cell viability. Recently, we deleted 960 5kbp-long non-coding loci that form a large number of 3D contacts in the Hi-C experiments in K562 and 35 of them are essential for cell viability. Strikingly, 14 out of the 35 essential non-coding loci (40%) reside in the cancer cRAM boundaries (Two-sample proportional test p-value = 0.01) while the cancer cRAM boundaries only accounted for 20% of the genome (Fig. 4c. Supplementary Table 4) Note that the median size of the cRAM boundaries is 750kbp long and deletion of 5kbp is only a tiny portion, which highlights the importance of cRAM boundaries in maintaining the normal cellular functions.

**Somatic genetic variations enriched in RAM boundaries**

If RAMs are functional modules important for regulating functional activities, we reason that somatic mutations in cancers may target the RAM boundaries to disrupt the modular organization of chromatin leading to aberrant regulation of gene expression and resulting tumorigenesis. The FCWG study revealed consensus mutations and variations from thousands of cancer patients including ~20 millions of somatic single nucleotide variations (SNVs) and ~1.08 millions of indels (https://dcc.icgc.org/releases). We found that, while cancer cRAM boundaries cover ~21.6% of the genome, they host 25.9% of somatic SNVs and 23.4% somatic indels. As a comparison, the conserved TAD boundaries covered 6.5% of the genome contained 5.2% somatic SNVs and 6% somatic indels (Fig. 4d). The cancer cRAM boundaries are significantly enriched with both somatic SNVs and indels compared to the TAD boundaries (P-value < 1E-5 from Two-sample Proportion tests, see “Methods”), indicating a stronger association with cancer mutations.

To elucidate the sequence features associated with cRAM boundaries and investigate how the somatic mutations change such features, we performed motif analysis on the cRAM boundaries using Homer. We focused on the cRAMs that are common in cancer (normal) but not in normal (cancer) samples as cancer (normal)-specific cRAMs, as they represent changed modularity between cancer and normal samples. By comparing cancer-specific and normal-specific cRAMs, we found 73 and 74 motifs enriched only in cancer and normal-specific cRAM boundaries, respectively (example motifs shown in Fig. 4e, g). We employed FIMO to identify the occurrences of the enriched motifs that counted for 25.8% and 30.1% in base pairs, respectively.
respectively, in the cancer and normal-specific cRAM boundaries (Fig. 4f, h). We next mapped the PCAWG somatic SNVs and indels onto the cancer and normal-specific cRAM boundaries. While somatic SNVs do not show preferred occurrences within the enriched motifs (24.9% and 29.9% for cancer and normal-specific cRAM boundaries, respectively), the somatic indels overlapping with the cancer/normal-specific cRAM boundaries preferentially hit the enriched motifs (52.3% and 63% for cancer and normal-specific cRAM boundaries, respectively), more than twofold by chance, in the altered cRAM boundaries between normal and cancer samples. We speculate that the enriched motifs in cancer and normal-specific cRAM boundaries may respectively facilitate disruption and formation of cRAM boundaries in the normal
Regulatory associated modules (RAM) identification

**Methods**

**Data source.** The 93 normal and 19 cancer samples with processed narrow peaks (H3K27ac, H3K4me3, H3K4me1) and broad peaks interferes with cellular functions and even affects cell viability as revealed by the CRISPR-Cas9 screening experiments. Fourth, cRAM boundaries are also more enriched with somatic genetic variants of SNVs and indels than the TAD boundaries. In particular, the somatic indels tend to disrupt the motifs specifically enriched in cancer or normal-specific cRAM boundaries, suggesting a possible mechanism of tumorigenesis involved in altering the chromatin modularity.

To investigate the mechanisms underlying the RAM formation, we found that the RAMs are separated by densely packed DNA regions (as shown by their large number of Hi-C contacts) enriched with repressive histone modifications and lacking open chromatin, active histone marks, or transcriptional events. Furthermore, unlike TADs, RAMs are insensitive to cohesin degradation. Taken together, these observations clearly show that RAMs are distinct from loop domains and TADs. RAMs are also different from lamina-associated domains (LADs) defined by measuring the intermediate filament protein LMNB1 localization. The LADs are formed through interactions between chromatin and lamina, and they are located at the periphery of the genome. The RAM boundaries are demarcated by densely packed DNAs and many RAM boundaries are not located in regions interacting with lamina or overlapping with TAD boundaries, and thus the mechanism underlying RAM boundary formation should be different from other chromatin modules including TADs, LADs and compartments.

Many studies (such as in refs. 46, 47) have shown that multivalent cations such as calcium, magnesium, and manganese can reduce the electrostatic repulsions between the DNA chains and induce DNA condensation. Furthermore, these cations may bind to specific DNA sequences and affect nucleosome positioning. Therefore, a possible mechanism can be that genomic DNAs become densely packed around cations such as Ca²⁺, Mg²⁺ and Mn²⁺ to form RAM boundaries even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina. Proteins such as calcium binding proteins that carry many cations or even if they are not marked by H3K9me3 or interacting with lamina.

**Discussion**

In this study, we analyzed the peak density profiles of histone modification data and found they show modular patterns. These modules are clearly defined by active marks such as H3K27ac, H3K4me1, and H3K4me3, indicating their association with the functional activity of the genome, and are thus termed regulation-associated modules (RAMs). While TADs and compartments are identified from the 3D contacts measured by Hi-C, RAMs are delineated by histone modifications that are directly related to chromatin accessibility and gene expression. We showed that RAMs are obviously distinct from TADs, compartments and LADs although some RAM boundaries do overlap with TAD, compartment or LAD boundaries.

By surveying 93 normal and 19 cancer samples, we found the following evidence to support that RAMs are spatial modules resulting from functional activities. First, we observed that on average 60% of the RAMs (i.e. consensus RAMs) are largely shared across samples, while some of them are sample specific. Compared to TADs, consensus RAMs host a higher percent of experimentally confirmed promoter-enhancer contacts, loop anchors, and super-enhancer clusters (i.e. within the same RAMs), suggesting RAMs represent a modularization of the genome at a scale better aligned with transcriptional regulation. Second, ecDNAs detected from cancer patients tend to originate from the same RAMs rather than across multiple RAMs, supporting the insulation effect of RAM boundaries. Third, deletion of the CRAM boundaries would result in more severe chromatin alteration than the TAD boundaries based on in silico predictions of Hi-C contacts, suggesting the importance of CRAM boundaries in maintaining the chromatin structure. In addition, deletion of RAM boundaries likely excludes the outliers. A lower correlation indicates a larger perturbation to the wild-type chromatin structure upon deletion.

In this study, we analyzed the peak density profile of chromatin structure using g:Profe

**Data source.** The 93 normal and 19 cancer samples with processed narrow peaks (H3K27ac, H3K4me3, H3K4me1) and broad peaks...
(H3K27me3, H3K9me3, H3K36me3) in hg19 were downloaded from Roadmap Epigenomics portal (https://egg2.wustl.edu/roadmap/web_portal/) and ENCODE portal (https://www.encodeproject.org/). All the downloaded data met the ENCODE standards. Supplementary Table 1 lists the samples used in this study. All the data were obtained by GNU Wget 1.14.

**RAM identification in individual samples.** We calculated the H3K27ac narrow-peak density using a sliding window of step size equal to 10 kb, 50 kb, 100 kb, 250 kb, and 500 kb, respectively and 500 kb flanking size of each window in every sample. The H3K27ac narrow-peak profiles were then smoothed by a local polynomial regression fitting. The RAM boundaries (valley/minima on the smoothing curves) and peaks (summit/maxima on the smoothing curve) in the smoothed curves were then detected using the “findpeaks” function in R package “pracma” (R v4.1.2).

**Consensus-RAM (cRAM) identification.** We first identified RAM boundaries in the 93 normal or 19 cancer samples using different step sizes (10 kb, 50 kb, 100 kb, 250 kb, and 500 kb) and then counted the percentage of a genomic region identified as RAM boundary in the 93 normal or 19 cancer samples. A genomic region with occurring percentage >10% was considered as a consensus-RAM (cRAM) boundary in the normal or cancer samples. We merged cRAM boundaries if they are located <250 kb apart from each other and required cRAMs have a size >250 kb. All the analysis was done in Python v3.8.

**Hi-C data analysis**

**Hi-C processing.** The Hi-C data for the wildtype K562, GM12878, A549, IMR90, NHEK, HUVEC, HMEC, and the HCT116 cell lines were downloaded from ENCODE (GSE63525) and the ENCODE portal (ENCSCR662QKG and GSE104433). All the raw fastq files were aligned to hg19 genome and then processed using Juicer pipeline (v1.5.6) with the default settings. The contact reads in a given cell line were further normalized by vanilla coverage (VC) normalization using the Juicer pipeline. The significance for a given fragment contact was computed by Poisson distribution with VC-normalized expected contact reads versus the VC-normalized observed contact reads. We then used HiCExplorer (v3.1.5) (v2.2.1.1) and HiCplotter (v0.8.1) software to visualize the Hi-C data.

**A/B compartment**

We performed A/B compartment analysis at 250 kb resolution. The eigenvectors for each chromosome involved in the Hi-C data analysis were extracted from the VC-normalized Hi-C counts processed by the Juicer pipeline with the default parameters. POLR2A ChIP-seq data were obtained from the ENCODE portal (https://www.encodeproject.org/). To determine A or B compartment, we calculated the correlation between the eigenvector of each chromosome and the Pol II peak density. As there was no Pol II ChIP-seq data available for HMEC, we used TSS density in the downloaded data. The H1 cell line TADs were downloaded from ref. 4.

**Topological associated domains (TAD).** To identify topological associated domains (TAD), we applied the insulation score method to the Hi-C data at 50 kb resolution in the K562, GM12878, A549, IMR90, NHEK, HUVEC, HMEC, HCT116, and HFF cell lines. The HFF cell lines generated by the DamID technique were obtained from GEO (GSE104888). We processed the H3K27ac data same as ref. 61. In brief, we aligned the raw data to the hg19 human genome using the BWA (v0.7.17) software, and then deduplicated the reads using PicardTools (Picard v2.17.0). The narrow peaks were called by comparing the associated input data using MACS2 (v2.2.7.1). All the parameters were set to the defaults.

**Enrichment of upregulated genes in enhancer-promoter pairs occurring in the same RAM of K562 but split in the background cell type by hypergeometric test**

The hypergeometric test was employed to measure the significance of the upregulated genes involved in the enhancer-promoter pairs occurring in the same RAM in K562 (foreground cell type) but in different RAMs of the background cell type. The population size N was the overall genes involved in the K562 and the compared cell type. The population success size M was the number of all upregulated genes in the K562 compared to the background cell type. The sampling size n was the number of the genes involved in the enhancer-promoter pairs occurring in the same RAM of K562 but in different RAMs of the background cell type, and the sampling success size m was the upregulated genes involved in the enhancer-promoter pairs occurring in the same RAM of K562 but in different RAMs of the background cell type. Enrichment was considered significant if p-value < 0.05.

**Enrichment of the somatic variants in cancer cRAM boundaries compared to the TAD boundaries assessed by Two-sample Proportion Tests**

The two-sample proportion test null hypothesis was to test the equal proportion of the number of the somatic variants relative to the genome coverage (in base pair) in the cancer cRAM boundaries and TAD boundaries. The two proportions were calculated separately by the number of the somatic variants divided by the boundary length (in base pair) for cancer cRAM boundaries and TAD boundaries. Enrichment was considered significant if p-value < 0.05 at two-sided test.

**Somatic mutations and structural variation analysis for cancer patients from PCAWG**

The consensus somatic SNV and indels were downloaded from PCAGW22,44. The VCF files were transformed to bed files by BEDOPS vc2bed tools (bedops v2.4.40) (https://bedops.readthedocs.io/en/latest/content/reference/file-management/convert/vcf2bed.html). The number of somatic SNV and indels overlapping with the RAM and TAD boundaries were then counted.

**Motif analysis**

The motif analysis was done using the Homer (v4.11.1) pipeline with default parameters. The motif occurrence was called using FIMO44 with p-value < 0.05 at two-sided test.

**Reporting summary**

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

The data that support this study are available from the corresponding author upon reasonable request. The 93 normal and 19 cancer samples with processed narrow peaks (H3K27ac, H3K4me3, H3K4me1) and broad peaks (H3K27me3, H3K9me3, H3K36me3) in hg19 were downloaded from Roadmap Epigenomics portal [https://egg2.wustl.edu/Research/Repository/Article/hc09].
roadmap/web_portal/ and ENCODE portal [https://www.encodeproject.org/]. Accession codes for these datasets are listed in Supplementary Table 1. The Hi-C data for the wildtype K562, GM12878, A549, IMR90, NHEK, HUVEC, HMEC, HCT116 and HFF cell lines were downloaded from GSE63325, ENCSR662OKG, GSE104333, and 4DNFIMROE6N4. The H3K27ac ChIP-seq data in the untreated HCT-116 RAD21-mAC cells and HCT-116 RAD21-mAC cells treated for 6 h with IAA were downloaded from GSE104888. The consensus somatic SNV and indels were downloaded from PCAWG. The K562, HCT116, H1, HAPI, RPE-hiTERT, HFFc6 cell lines LAD data were downloaded through 4DNFIJHD22QE, 4DNFIT9W77EE, 4DNFIUVTO2H3, 4DNFIDFCY3JN, 4DNFIA2LBQCD, 4DNFIJXADI29, 4DNFIUVTO2H3, 4DNFIT9W77EE.

Code availability
The RAM identification package (“findRAM”) is available on GitHub (https://github.com/Wang-lab-UCSD/findRAM).

References
1. Kouzarides, T. Chromatin modifications and their function. Cell 128, 693–705 (2007).
2. Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. Cell 128, 707–719 (2007).
3. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485, 381–385 (2012).
4. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380 (2012).
5. Sexton, T. et al. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell 148, 458–472 (2012).
6. Crane, E. et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature 523, 240–244 (2015).
7. 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).
8. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289–293 (2009).
9. Zhu, Y. et al. Constructing 3D interaction maps from 1D epigenomes. Nat. Commun. 7, 10812 (2016).
10. Huang, J., Marco, E., Pinello, L. & Yuan, G.-C. Predicting chromatin organization using histone marks. Genome Biol. 16, 162 (2015).
11. Ashoor, H. et al. Graph embedding and unsupervised learning predict genomic sub-compartments from HiC chromatin interaction data. Nat. Commun. 11, 1173 (2020).
12. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. Cell Res 21, 381–395 (2011).
13. Vermeulen, M. et al. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell 142, 967–980 (2010).
14. Bartke, T. et al. Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell 143, 470–484 (2010).
15. Liu, Z. et al. 3D imaging of Sox2 enhancer clusters in embryonic stem cells. Elife 3, e04236 (2014).
16. Cisse, I. I. et al. Real-time dynamics of RNA polymerase II clustering in live human cells. Science 341, 664–667 (2013).
17. Cho, W.-K. et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science 361, 412–415 (2018).
18. Pancholi, A. et al. RNA polymerase II clusters form in line with surface condensation on regulatory chromatin. Mol. Syst. Biol. 17, e10272 (2021).
44. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. Bioinformatics 27, 1017–1018 (2011).
45. Raudvere, U. et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 47, W191–W198 (2019).
46. Bloomfield, V. A. Condensation of DNA by multivalent cations: considerations on mechanism. Biopolymers 31, 1471–1481 (1991).
47. Koltner, I., Wagner, K. & Safinya, C. R. DNA condensation in two dimensions. Proc. Natl Acad. Sci. USA 97, 14046–14051 (2000).
48. Srivastava, A. et al. Structure-guided DNA-DNA attraction mediated by divalent cations. Nucleic Acids Res. 48, 7018–7026 (2020).
49. Yang, Z. & Hayes, J. J. The divalent cations Ca2+ and Mg2+ play specific roles in stabilizing histone-DNA interactions within nucleosomes that are partially redundant with the core histone tail domains. Biochemistry 50, 9973–9981 (2011).
50. Cleveland, W. S. & Grosse, E. Computational methods for local regression. Stat. Comput. 1, 47–62 (1991).
51. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3, 95–98 (2016).
52. Wolff, J. et al. Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. Nucleic Acids Res. 48, W177–W184 (2020).
53. Yuan, Y. et al. Cohesion loss eliminates all loop domains. Cell 171, 305–320.e24 (2017).
54. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595 (2010).
55. Liu, T. Use model-based Analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. Methods Mol. Biol. 1150, 81–95 (2014).
64. Jiao, W. et al. A deep learning system accurately classifies primary and metastatic cancers using passenger mutation patterns. Nat. Commun. 11, 728 (2020).
65. Nep, S. et al. BEDOPS: high-performance genomic feature operations. Bioinformatics 28, 1919–1920 (2012).

Acknowledgements
The authors are grateful to the labs of Drs. Bas van Steensel, Erez Lieberman Aiden and Marion Cremer for data sharing. This project was partially supported by NIH (R01HG009626 to W.W.).

Author contributions
W.W. conceived and supervised the project. L.Z. performed the bioinformatics analysis. L.Z. and W.W. wrote the manuscripts.

Competing interests
The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-32911-y.

Correspondence and requests for materials should be addressed to Wei Wang.

Peer review information Nature Communications thanks Xiaochen Bo and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022