Article

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

Lina Zheng [1] **& Wei Wang** [1,2,3] ✉

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[1,2]. 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)[3–6] 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[7,8]. Computational models have shown that histone modification signals are predictive of Hi-C contacts particularly for enhancer–promoter interactions[9], TAD boundaries[10] and compartments[11]. Histone modifications are tightly associated with transcriptional activity[12–14] 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].

[1]Bioinformatics and Systems Biology Program, University of California San Diego, La Jolla, CA 92093-0359, USA. [2]Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093-0359, USA. [3]Department 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 unfilled 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/)[19] and ENCODE portal (https://www.encodeproject.org/)[20] (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.

At a given step size, we identified the valley or minima of the H3K27ac peak profile that was smoothed using local polynomial fit in each chromosome and in each cell type (see "Methods"). These valleys demarcated the boundaries of the modular domains (called Regulation Associated Domains or 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).

One example of the identified RAMs in chr12 of the GM12878 cell by IGV software[21] is shown in Fig. 1c–e. Obviously, the RAM boundaries have lower signals of the active histone marks (H3K27ac, H3K4me1, H3K4me3, H3K36me3) and higher repressive marks (H3K9me3 and H3K27me3) compared to the within RAM regions. Consistently, they tend to align with the B compartment or subcompartments (B1, B2, B3). Furthermore, by counting the number of the 3D contacting neighbors for each locus using the 10 kb resolution Hi-C data in GM12878 (contacts with log(P-value) ≤ = −10), we found that the RAM boundaries tend to harbor many 3D contacts, indicated an enrichment with densely packed DNA sequences forming many spatial contacts (Fig. 1c).

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, HMEC, 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. 2j). Consistently, the cRAM boundaries are enriched with B-compartment but also with a significant portion of A compartments (Fig. 2j). Third, we checked the Lamin-B1 signals for the cRAM boundaries. Lamin-B1 is a scaffolding component of the nuclear envelope[22,23]. 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
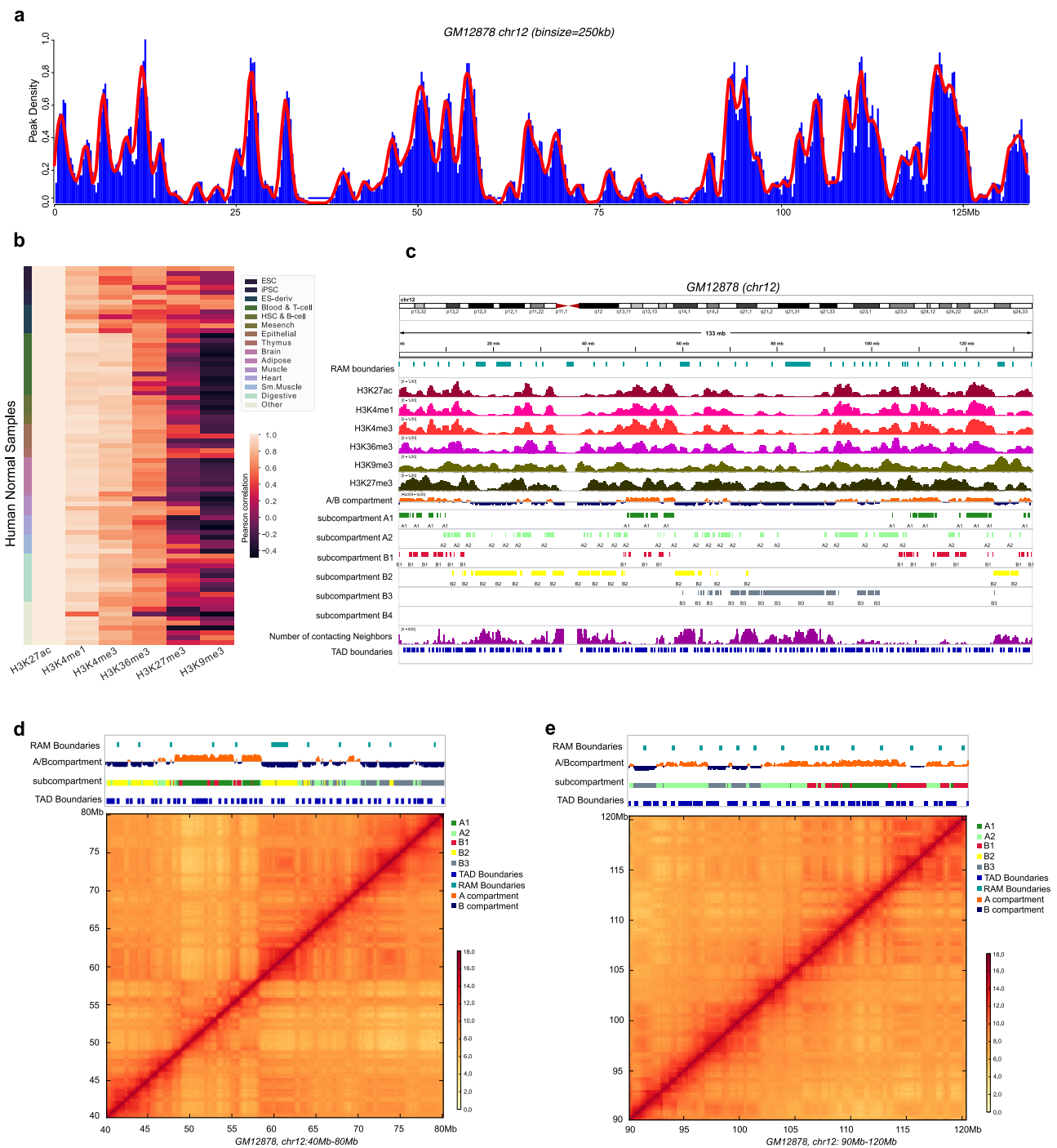
Fig. 1 | Regulation associated module (RAM) identification. a H3K27ac peaks density of chr12 in GM12878 (bin size=250 kb). **b** Pearson correlation between histone modification marks (H3K27ac, H3K4me1, H3K4me3, H3K36me3, H3K27me3, and H3K9me3) for the Roadmap samples. **c** Examples of histone modifications, A/B compartments, subcompartments, number of the 3D contacts, TAD boundaries and RAM boundaries in chr12 for GM12878. **d** The zoom-in genomic view for chr12:40Mb-80Mb in hg19. **e** The zoom-in genomic view for chr12:90Mb-120Mb in hg19.

sequences; 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 TADs[4,6,8,26] (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
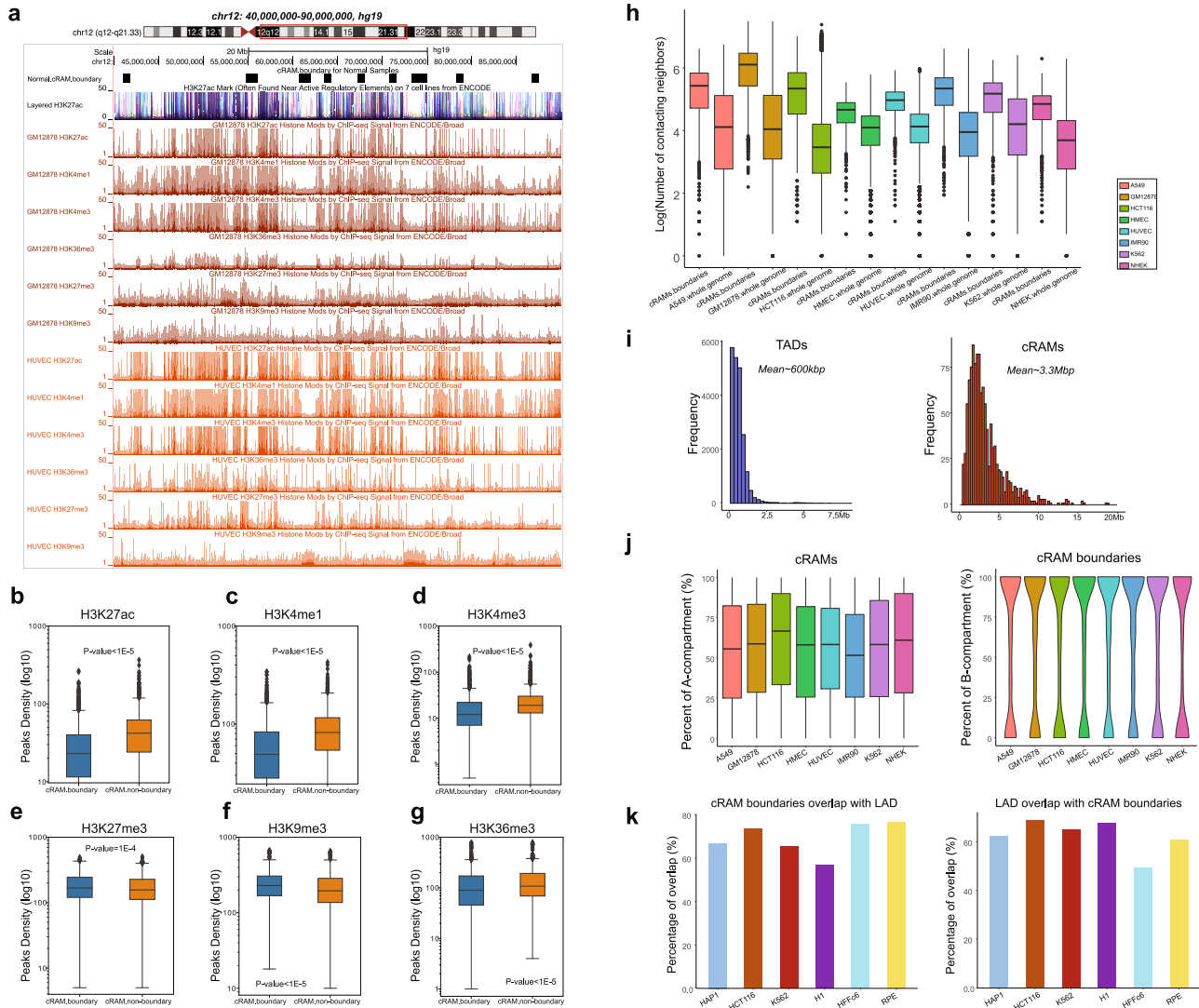
Fig. 2 | Characterization of the cRAMs and boundaries. a Genome browser examples of normal cRAM boundaries and the histone modifications in genomic region chr12:40Mb-90Mb in hg19 for GM12878 and the HUVEC cells. **b–g** The genome-wide enrichment of the histone modification marks ($n = 93$ samples/cells). **b** H3K27ac, (**c**) H3K4me1, (**d**) H3K4me3, (**e**) H3K27me3, (**f**) H3K9me3, and (**g**) H3K36me3 in cRAM boundaries and non-boundaries. The p-value was computed by two-sided Wilcoxon Rank Sum test. **h** The contacting neighbors distribution of cRAM boundaries and whole-genome locus in the 3D contact network in a diverse of the cell types ($n = 8$ cell lines). The bounds of all boxplots showed the 25 percentile, median, 75 percentile of the dataset; the maxima and minima were defined excluding the outliers. **i** Sizes of the TADs and cRAMs. **j** cRAM boundaries distribution over A/B compartments ($n = 8$ cell lines). The bounds of all boxplots showed the 25 percentile, median, 75 percentile of the dataset; the maxima and minima were defined excluding the outliers. **k** cRAM boundaries distribution over LaminB1 signals (LAD).

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 were within the same RAMs 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 occurred in the same RAM 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 (cancer 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 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 database[27] (Supplementary Fig. 4b). These observations further illustrated that
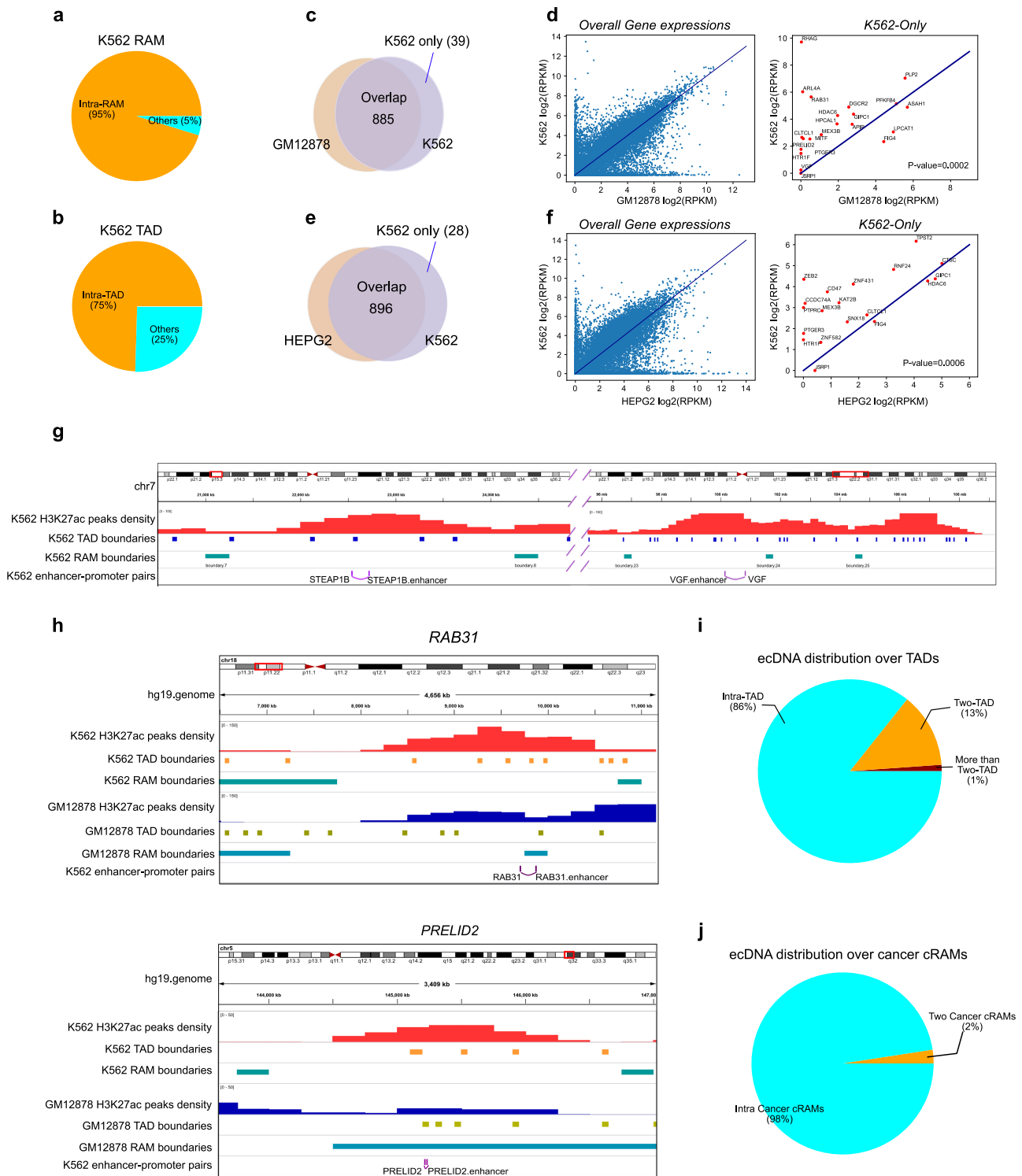
Fig. 3 | RAM is a functional unit. a K562 enhancer–promoter pairs distribution over K562 RAMs. **b** K562 enhancer–promoter pairs distribution over K562 TADs. **c** K562 enhancer–promoter pairs distribution over K562 RAMs and GM12878 RAMs. **d** Genes regulated by the 39 enhancer–promoter interactions only within K562 RAMs tend to have higher expressions in K562 compared to GM12878. The *p*-value was derived from hypergeometric test. **e** K562 enhancer–promoter pairs distribution over K562 RAMs and HEPG2 RAMs. **f** Genes regulated by the 28 enhancer–promoter interactions only within K562 RAMs tend to have higher expressions in K562 compared to HEPG2. The *p*-value was derived from hypergeometric test. **g** Examples of K562 enhancer–promoter pairs relative to K562 TAD and RAM boundaries. **h** Examples of K562 enhancer–promoter pairs relative to K562 and GM12878 RAM boundaries. **i** ecDNA distribution over TADs. **j** ecDNA distribution over cancer cRAMs.

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)[28]. As the HCC1599, MB157,

and REC1 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[29], 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[30,31]. 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[32–34]. 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[35], 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[4] 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[36] (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 5.1641E−19. There are 12 genes involved in this pathway: *ATF1, BMP5, BMP6, EGFR, FGFR1, GLI2, IGF1, IL6, JUN, KITLG, PDGFA, PDGFRA*, which are known important for cell proliferation and cancer pathogenesis. For example, *EGFR* is a driver of tumorigenesis[37]. Deregulation of the oncogenic FGFR signaling has been frequently observed in multiple types of cancers[38]. The PDGF-mediated signaling has been reported to be involved in the cell proliferation and invasion[39]. 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[40] (https://github.com/jzhoulab/orca) as it is prohibitive to perform hundreds of Hi-C experiments with sufficient resolution. We took the ORCA model pre-trained on the high-

resolution Hi-C and Micro-C data in H1-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 H1-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 H1-hESC cells 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[41]. 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 PCAWG 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)[42]. 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[4] covering 6.5% 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[43]. 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[44] to identify the occurrences of the enriched motifs that counted for 25.8% and 30.1% in base pairs,
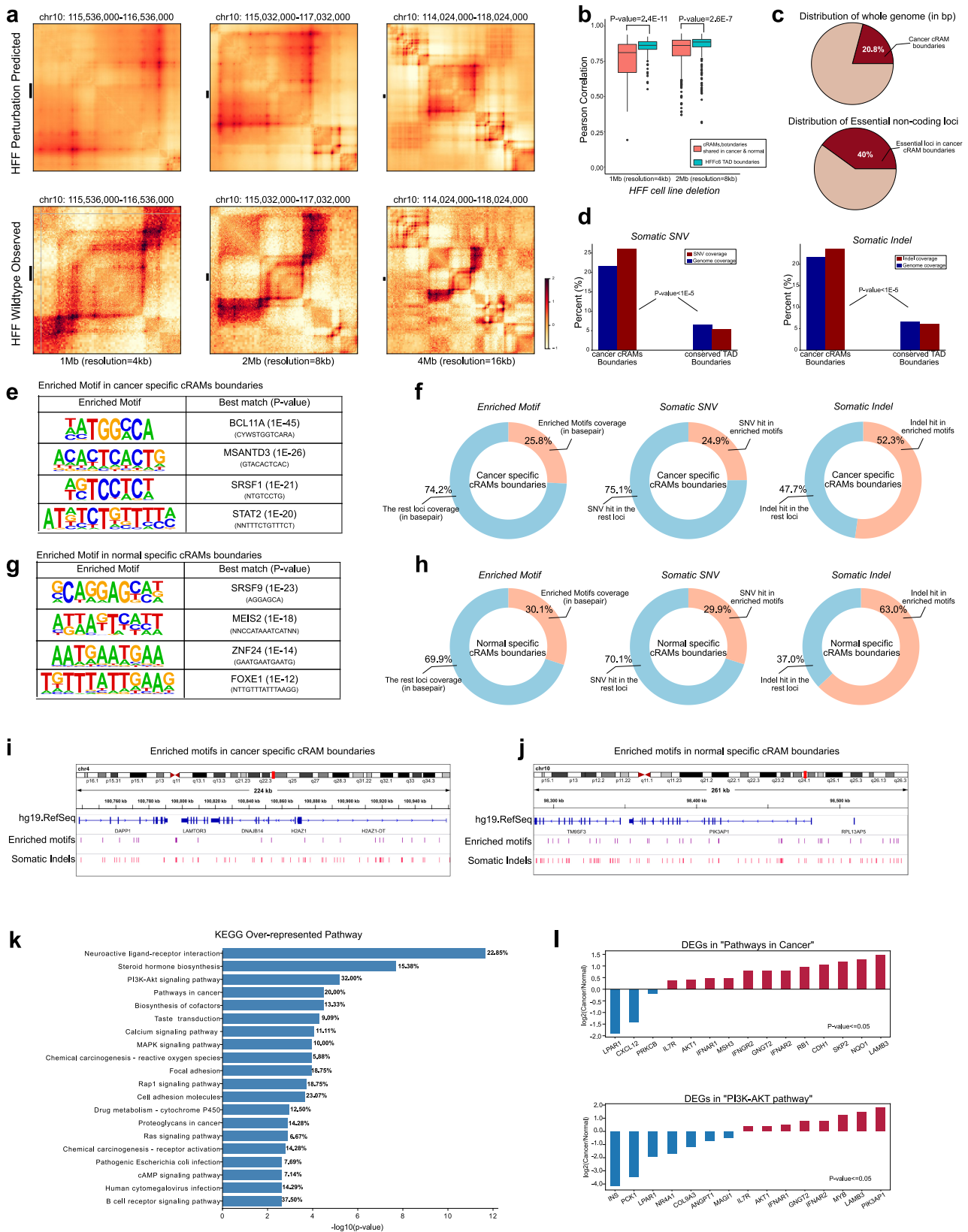
respectively, in the cancer and normal-specific cRAM boundaries (Fig. 4f, h). We next mapped the PCAWG somatic SNVs and indels onto the cancer and normal-specific cRAM boundaries. While somatic SNVs do not show preferred occurrences within the enriched motifs (24.9% and 29.9% for cancer and normal-specific 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

Fig. 4 | The association of the cRAM boundaries with the 3D chromatin structure and cancer somatic variants. a An example of the Hi-C contact change upon deletion of the cRAM boundary (chr10:115,940,000-116,040,000 in hg38) in HFF cells predicted by a deep learning model ORCA. **b** Pearson correlations between the predicted Hi-C contacts before and after cRAM boundary and TAD boundary deletion in HFF cells. A lower correlation indicates a larger perturbation to the wild-type chromatin structure upon deletion. $n = 399$ loci in 1 Mb resolution and $n = 989$ loci in 2 Mb resolution data for HFF cell line. The p-value was derived from two-sided Wilcoxon Rank Sum. The bounds of all boxplots showed the 25 percentile, median, 75 percentile of the dataset; the maxima and minima were defined excluding the outliers. **c** Essential non-coding loci enriched in the cancer cRAM boundaries. **d** Somatic SNV and indels enrichment in cancer cRAM boundaries and TAD boundaries. Genome coverage: the total base pairs of the cancer cRAM boundaries or TAD boundaries in the whole genome; SNV coverage: the percentage of the SNVs in the cancer cRAM boundaries or TAD boundaries in the whole genome; INDEL coverage: the percentage of the indels in cancer cRAM boundaries or TAD boundaries in the whole genome. **e** Examples of the enriched motifs in the cancer-specific cRAM boundaries. The motif enrichment p-value was derived from cumulative binomial distribution with default settings in Homer pipeline. **f** Overlaps of somatic SNVs and indels with the enriched motifs in the cancer-specific cRAM boundaries. **g** Examples of the enriched motifs in the normal-specific cRAM boundaries. The motif enrichment p-value was derived from cumulative binomial distribution with default settings in Homer pipeline. **h** Overlaps of somatic SNVs and indels with the enriched motifs in the normal-specific cRAM boundaries. **i** The zoom-in genomic view for the enriched motifs with somatic indels in cancer-specific cRAM boundaries. **j** The zoom-in genomic view for the enriched motifs with somatic indels in normal-specific cRAM boundaries. **k** KEGG over-represented pathways for the genes within 2.5 kb from the enriched motifs overlapping with somatic indels. The p-value was derived from hypergeometric test with Benjamini–Hochberg False Discovery Rate approach at FDR cutoff 0.05 in g:Profiler software. **l** Differentially expressed genes (p-value ≤ 0.05 from two-sided Wilcoxon Rank Sum test) in KEGG "Pathways in cancer" and "PI3K-AKT pathway".

samples. Two examples of these motifs overlapping with indels are shown in Fig. 4i, j. We identified genes within 2.5 kb from the enriched motifs overlapping with somatic indels and analyzed the enriched pathways using g:Profiler[45]. The KEGG over-represented pathways are shown in Fig. 4k, and the gene ontology molecular functions are shown in Supplementary Fig. 7. Furthermore, we downloaded the normalized gene expressions of the TCGA and GTEx samples from Expression Atlas (https://www.ebi.ac.uk/gxa/home) and identified differentially expressed genes (DEGs) (p-value ≤ 0.05 by the two-sided Wilcoxon Rank Sum tests between cancer and normal samples). The top-ranked pathways associated with cancer and cell proliferation are highly enriched with the DEGs, such as *PIK3AP1, LABM3, AKT1, MYB* in "PI3K-AKT pathway" (32%) and "pathways in cancer" (20%) (Fig. 4l). These observations suggested that the motifs specifically enriched in the formation of cancer cRAM boundaries and disruption of normal cRAM boundaries are close to genes important for tumorigenesis, cell survival, and cell proliferation. Somatic indels can severely alter these motifs and may contribute to the cRAM boundary change, affecting the expressions of the nearby genes.

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

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[48] and affect nucleosome positioning[49,50]. Therefore, a possible mechanism can be that genomic DNAs become densely packed around cations such as Ca^{2+}, Mg^{2+} and Mn^{2+} 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 their interacting partners may recognize specific DNA sequences such as those motifs enriched in the cancer or normal-specific cRAM boundaries to facilitate locus-specific localization of cations. Interestingly, the most enriched molecular function of the genes close to (<2.5kbp) the enriched motifs overlapping with somatic indels in the cancer or normal-specific cRAM boundaries is calcium ion binding (Supplementary Fig. 7), and ~30% of them are differentially expressed in cancer and normal samples (Supplementary Table 5), implying a possible feedback mechanism. This hypothesis and the mechanistic details are awaiting future studies.

Methods
Regulatory associated modules (RAM) identification
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/)[19] 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[51]. 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 occuring percentage > =25% 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 GEO (GSE63525) and the ENCODE portal (ENCSR662QKG and GSE104333). All the raw fastq files were aligned to hg19 genome and then processed using Juicer pipeline (v1.5.6) with the default settings[7,52]. 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[53–55] (v2.2.1.1) and HiCplotter[56] (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 in all the cell lines 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[52]. POLR2A ChIP-seq data were obtained from the ENCODE[57] portal (https://www.encodeproject.org/). To determine A or B compartment, we calculated the correlation between the first eigenvector of each chromosome and the Pol II peak density[58]. As there was no Pol II ChIP-seq data available for HMEC, we used TSS density in the hg19 genome to assign A/B compartments in HMEC.

Topological associated domains (TAD). To identify topological associated domains (TAD), we applied the insulation score method[59] 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 line hic data were downloaded with accession number 4DNFIMROE6N4) and its TADs identified with insulation score method[59] at 50 kb resolution data. The H1 cell line TADs were downloaded from ref. 4.

Lamina-associated domains (LADs) data processing
The Lamin-B1 signal data in the K562, HCT116, H1, HAP1, RPE-hTERT, HFFc6 cell lines generated by the DamID technique were obtained from the 4DN portal[60]. We downloaded the mean of the replicates for each cell line. The Lamin-B1 signals at 50 kb resolution for the K562,

HCT116, H1, HAP1, RPE-hTERT, HFFc6 cell lines were lifted over from hg38 to hg19.

Cohesin degradation analysis
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 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[62], and then deduplicated the reads using PicardTools (Picard v2.17.0). The narrow peaks were called by comparing the associated input data using MACS2[63] (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 PCAWG[42,64]. The VCF files were transformed to bed files by BEDOPS vcf2bed tools (bedops v2.4.40) (https://bedops.readthedocs.io/en/latest/content/reference/file-management/conversion/vcf2bed.html)[65]. 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[43] with default parameters. The motif occurrence was called using FIMO[44] with p-value ≤ = 1E−4. (MEME Suite v5.4.1).

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/

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 GSE63525, ENCSR662QKG, 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, HAP1, RPE-hTERT, HFFc6 cell lines LAD data were downloaded through 4DNFIJHD22QE, 4DNFIA2LBQCD, 4DNFIJXADI29, 4DNFIDFCY3JN, 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).
19. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
20. Sloan, C. A. et al. ENCODE data at the ENCODE portal. *Nucleic Acids Res.* **44**, D726–D732 (2016).
21. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
22. Foisner, R. & Gerace, L. Integral membrane proteins of the nuclear envelope interact with lamins and chromosomes, and binding is modulated by mitotic phosphorylation. *Cell* **73**, 1267–1279 (1993).
23. Wydner, K. L., McNeil, J. A., Lin, F., Worman, H. J. & Lawrence, J. B. Chromosomal assignment of human nuclear envelope protein genes LMNA, LMNB1, and LBR by fluorescence in situ hybridization. *Genomics* **32**, 474–478 (1996).
24. Gasperini, M. et al. A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell* **176**, 1516 (2019).
25. Fulco, C. P. et al. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
26. Valton, A.-L. & Dekker, J. TAD disruption as oncogenic driver. *Curr. Opin. Genet. Dev.* **36**, 34–40 (2016).
27. Rouillard, A. D. et al. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* **2016**, baw100 (2016).
28. Petrovic, J. et al. Oncogenic notch promotes long-range regulatory interactions within hyperconnected 3D cliques. *Mol. Cell* **73**, 1174–1190.e12 (2019).
29. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
30. Cremer, M. et al. Cohesin depleted cells rebuild functional nuclear compartments after endomitosis. *Nat. Commun.* **11**, 6146 (2020).
31. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320.e24 (2017).
32. Turner, K. M. et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* **543**, 122–125 (2017).
33. Verhaak, R. G. W., Bafna, V. & Mischel, P. S. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nat. Rev. Cancer* **19**, 283–288 (2019).
34. deCarvalho, A. C. et al. Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma. *Nat. Genet.* **50**, 708–717 (2018).
35. Kim, H. et al. Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nat. Genet.* **52**, 891–897 (2020).
36. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
37. Sigismund, S., Avanzato, D. & Lanzetti, L. Emerging functions of the EGFR in cancer. *Mol. Oncol.* **12**, 3–20 (2018).
38. Babina, I. S. & Turner, N. C. Advances and challenges in targeting FGFR signalling in cancer. *Nat. Rev. Cancer* **17**, 318–332 (2017).
39. Sahraei, M. et al. MUC1 regulates PDGFA expression during pancreatic cancer progression. *Oncogene* **31**, 4935–4945 (2012).
40. Zhou, J. Sequence-based modeling of genome 3D architecture from kilobase to chromosome-scale. *bioRxiv* https://doi.org/10.1101/2021.05.19.444847 (2021).
41. Ding, B. et al. Noncoding loci without epigenomic signals can be essential for maintaining global chromatin organization and cell viability. *Sci. Adv.* **7**, eabi6020 (2021).
42. The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
43. 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).

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. Koltover, 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. Davey, C. A. & Richmond, T. J. DNA-dependent divalent cation binding in the nucleosome core particle. *Proc. Natl Acad. Sci. USA* **99**, 11169–11174 (2002).
50. 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).
51. Cleveland, W. S. & Grosse, E. Computational methods for local regression. *Stat. Comput.* **1**, 47–62 (1991).
52. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
53. 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).
54. Ramírez, F. et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189 (2018).
55. Wolff, J. et al. Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* **46**, W11–W16 (2018).
56. Akdemir, K. C. & Chin, L. HiCPlotter integrates genomic data with interaction matrices. *Genome Biol.* **16**, 198 (2015).
57. Yip, K. Y. et al. Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biol.* **13**, R48 (2012).
58. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F. & Chen, L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* **30**, 90–98 (2011).
59. Crane, E. et al. Condensin-driven remodeling of X-chromosome topology during dosage compensation. *Nature* **523**, 240 (2015).
60. van Schaik, T., Vos, M., Peric-Hupkes, D., Hn Celie, P. & van Steensel, B. Cell cycle dynamics of lamina-associated DNA. *EMBO Rep.* **21**, e50636 (2020).
61. Rao, S. S. P. et al. Cohesion loss eliminates all loop domains. *Cell* **171**, 305–320.e24 (2017).
62. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
63. 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. Neph, S. et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).

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.