frontiers | Frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 19 April 2022
doi: 10.3389/fcell.2022.845118

Extensive Chromatin Structure-Function Associations Revealed by Accurate 3D Compartmentalization Characterization

*Zi Wen [1,2], Weihan Zhang [1], Quan Zhong [1,2], Jinsheng Xu [1,2], Chunhui Hou [3], Zhaohui Steve Qin [4] and Li Li [1,2,5]**

[1]Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan, China, [2]3D Genomics Research Center, Huazhong Agricultural University, Wuhan, China, [3]Department of Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen, China, [4]Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA, United States, [5]Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan, China

OPEN ACCESS

Edited by:
Saket Jain,
University of California, San Francisco, United States

Reviewed by:
Kyle Eagen,
Baylor College of Medicine, United States
Shashank Srivastava,
Northwestern University, United States

***Correspondence:**
Li Li
li.li@mail.hzau.edu.cn

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Cell and Developmental Biology

Received: 29 December 2021
Accepted: 24 March 2022
Published: 19 April 2022

Citation:
Wen Z, Zhang W, Zhong Q, Xu J, Hou C, Qin ZS and Li L (2022) Extensive Chromatin Structure-Function Associations Revealed by Accurate 3D Compartmentalization Characterization. Front. Cell Dev. Biol. 10:845118. doi: 10.3389/fcell.2022.845118

A/B compartments are observed in Hi-C data and coincide with eu/hetero-chromatin. However, many genomic regions are ambiguous under A/B compartment scheme. We develop MOSAIC (MOdularity and Singular vAlue decomposition-based Identification of Compartments), an accurate compartmental state detection scheme. MOSAIC reveals that those ambiguous regions segregate into two additional compartmental states, which typically correspond to short genomic regions flanked by long canonical A/B compartments with opposite activities. They are denoted as micro-compartments accordingly. In contrast to the canonical A/B compartments, micro-compartments cover ~30% of the genome and are highly dynamic across cell types. More importantly, distinguishing the micro-compartments underpins accurate characterization of chromatin structure-function relationship. By applying MOSAIC to GM12878 and K562 cells, we identify *CD86*, *ILDR1* and *GATA2* which show concordance between gene expression and compartmental states beyond the scheme of A/B compartments. Taken together, MOSAIC uncovers fine-scale and dynamic compartmental states underlying transcriptional regulation and disease.

Keywords: A/B compartment, modularity, transcriptional regulation, chromatin architecture, heterochromatin

INTRODUCTION

Modularity, a widely recognized principle of living systems, has been observed in many aspects of biological organization (Wagner et al., 2007). In the eukaryotic genome, the highest level of physical modularity is the separation of genetic material into chromosomes that occupy discrete territories inside the nucleus. Chromosome conformation capture (3C)-based high-throughput technologies have revealed rich modular features in 3D chromatin architecture at fine scales (Gibcus and Dekker, 2013; Schoenfelder and Fraser, 2019). Plaid patterns were observed at the megabase scale in intra-chromosomal heatmaps generated by the first high-throughput chromosome conformation capture (Hi-C) experiments (Lieberman-Aiden et al., 2009) on human cells. These patterns suggested the

existence of two segregated groups of chromatin in which regions within the same group are in closer spatial proximity than regions across groups, a hallmark of modularity. The two groups, or structural modules, were named the A and B compartments, corresponding to transcriptionally active and silent chromatin, respectively. Principal component analysis (PCA) of the correlation heatmap yields the first eigenvector (EV1), whose sign designates whether a region is in compartment A or B. Due to its simplicity in implementation, straightforwardness in interpretation, and robustness to noise, eigenvector decomposition methods, including PCA and singular value decomposition (SVD), are the standard procedures for compartmentalization analysis (Schmitt et al., 2016b; Durand et al., 2016; Wolff et al., 2018).

Compartmental organization of chromosomes has been confirmed in other animals (Dixon et al., 2012; Rowley et al., 2017), plants (Dong et al., 2017), and archaea (Takemata et al., 2019). Although it disappears temporarily in certain stages of mitosis (Naumova et al., 2013) and zygotic genome activation (Du et al., 2017; Ke et al., 2017; Niu et al., 2021), chromosome compartmentalization is a widespread pattern in a variety of cell types and biological processes. Given the mechanistic connection to phase separation (Larson et al., 2017; Strom et al., 2017) and the distinction in gene activity between compartments, the role of compartmentalization in chromatin functions, especially transcriptional regulation, becomes an increasingly important question (Hildebrand and Dekker, 2020).

While assignment of A/B compartments by EV1 can capture the overall plaid pattern of individual intra-chromosomal heatmaps, studies based on trans contracts have further divided the human genome into six subcompartments (Rao et al., 2014; Xiong and Ma, 2019; Ashoor et al., 2020). Recently, a computational framework that integrates TSA-seq, DamID and Hi-C data provides spatial interpretation for Hi-C subcompartments relative to multiple nuclear compartments (Wang et al., 2021). Another method starts with compartmental domain identification, followed by hierarchical clustering on the domains, then obtains eight subcompartment types (Liu et al., 2021). An analysis based on HCT116 cells identifies three distinct compartments that are correlated with H3K27ac, H3K27me3, or H3K9me3, respectively (Nichols and Corces, 2021). These results suggest that simply partitioning chromatin into two compartments is insufficient to capture the full complexity of chromatin compartmentalization. In practice, however, when bipartite compartment analysis and subcompartment analysis are compared, the former is still more widely applied in practice (Sati et al., 2020; Belaghzal et al., 2021; Furlan-Magaril et al., 2021) considering its straightforward connection to chromatin activity and robustness to noise (Schmitt et al., 2016b). Thus, it is important to design a method which advance the eigenvector decomposition to more refined levels while keeping its advantages over the subcompartment methods.

In this study, we showed that modularity can be used to quantify chromatin compartmentalization. Based on this observation, we developed a modularity-based compartmentalization analysis framework named MOSAIC (MOdularity and Singular vAlue decomposition-based Identification of Compartments), using intra-chromosomal contact information from Hi-C. After performing dimension reduction by SVD, we systematically evaluated the structural and functional implications of additional EVs orthogonal to EV1. While EV1 showed the strongest correlation with active histone marks as known, we found that the second EV (EV2) strongly correlated with H3K27me3, highlighting facultative heterochromatin, which typically consists of small regions interspersed within large areas of the A compartment. A significant part of the genome does not belong to canonical A/B compartments. These initially ambiguous regions harboring marginal EV1 values can be clarified and segregated into two additional components, which totally account for ~30% of the genome. We termed them micro-compartments given their smaller size and genomic neighborhood with respect to canonical A/B compartments. Through the top EVs, MOSAIC intuitively connected 3D chromatin architecture to 1D chromatin states and activities, including histone modifications and transcription. More importantly, when applying MOSAIC to GM12878 and K562 cells, we identified much more genes that exhibit concordance between expression and compartmental states when compared to results obtained using conventional A/B compartments. Furthermore, the case studies on *HoxA* gene cluster and *GATA2* exemplify the power of MOSAIC in revealing such structurally delicate but biologically essential features, and the limitations of the A/B compartment scheme.

MATERIALS AND METHODS

A/B Compartment Identification

Compartment A and B are determined by the sign of the first eigenvector obtained from PCA of intra-chromosomal Hi-C contact maps. Regions with opposite EV1 signs in different samples are considered as compartment switching between A and B.

The MOSAIC Algorithm

Step 1: Remove centromere effect and matrix reconstruction.

Because the O/E matrix is a symmetric and non-negative adjacency matrix of the chromosomal contact map, SVD on O/E matrix M can get the eigenvalues and eigenvectors according to the following equation:

$$M = U\Sigma U^T \tag{1}$$

where

$$U = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{nn} \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \\ \vdots & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$$

Here U is a matrix composed of eigenvectors, and Σ contains the eigenvalues in descending order of magnitude.

The top eigenvectors are of particular importance because they reflect the dominant structure of the O/E matrix. In general, the top eigenvectors represent genome coverage, global or local pattern, centromere effect, etc. The eigenvector with strong bias between chromatin's p and q arm is considered to represent the centromere effect. Here, we define I to measure the effect using the following equation:

$$I = \frac{P + Q}{2} \tag{2}$$

where

$$P = \left| \overline{\sum_{i \in p} Sgn(x_i)} \right|$$

$$Q = \left| \overline{\sum_{i \in q} Sgn(x_i)} \right|$$

Here x_i is the value of the eigenvector at position i, and p and q represent the short and long arm, respectively. The Sgn function returns the sign of a number.

I ranges from 0 to 1. The higher the value of I, the stronger the arm effect. To disentangle the influence of arm effect from the compartmentalization pattern, the O/E matrix M was updated by modified eigenvectors and eigenvalues according to following equation:

$$\begin{cases} untreated, & I \le 0.4 \\ balance\ \boldsymbol{X}, & 0.4 < I < 0.5 \\ \lambda = 0, & I \ge 0.5 \end{cases} \tag{3}$$

When I is ≥ 0.5, we replace the eigenvalue corresponding to the eigenvector with 0. For I between 0.4 and 0.5, we balance the corresponding eigenvector of the long arm or the short arm by subtracting their arm-wise means. We reconstruct the new O/E matrix A using the processed eigenvectors and eigenvalues. Without the above correction on centromere effect, the results will have negligible effect for most chromosomes. For the other several chromosomes with particularly strong centromere effect, the identified A2 or B2 will be restricted to within the long or short arms of the chromosomes, respectively, which is of course does not reflect the compartmental pattern.

Step 2: Pick eigenvectors with the highest and second-highest modularity.

We perform SVD on the updated O/E matrix A to obtain the new eigenvectors. We then construct a Hi-C interaction graph by A and compute the modularity matrix, B, according to the following equation (Newman, 2006):

$$B_{ij} = \left(A_{ij} - r\frac{k_i k_j}{2m} \right) \Big/ 2m \tag{4}$$

where A_{ij} is the new O/E matrix, k_i is the degree of i, m is total number of links, and r is the resolution parameter ($r = 1$ is used to compute the modularity). Thus, we can obtain the modularity Q of the whole network from modularity matrix B according to the following equation:

$$Q = \sum_{i,j} \left[B_{i,j} \right] \delta_{i,j} \tag{5}$$

where $\delta_{i,j}$ is 1 if vertex i and vertex j are assigned to the same community, and 0 otherwise.

For each eigenvector, we can divide the chromosome corresponding to matrix A into two communities according to the signs of the values, so that we can calculate the modularity corresponding to each eigenvector. We take top two eigenvectors in terms of modularity for the next step of the analysis.

Step 3: K-means clustering.

We perform K-means clustering on EV1 and EV2 weighted by eigenvalues by setting k equal to 3, 4, 7, or 8 according to SSE with k of SSE (**Supplementary Figure S1G**). The best results were obtained when k was set to 4, and these results were consistent overall among chromosomes.

Step 4: Refine clustering by Louvain algorithm.

To optimize the clustering results, we employed the Louvain algorithm to maximize modularity. The Louvain algorithm relies on an iterative method to quickly converge to the maximum in Q. For each iteration t, individual node movement maximizes the increase in modularity Q, ΔQ, according to the following equation:

$$\Delta Q = Q_{t+1} - Q_t = \left(\sum_{i,j} \left[B_{i,j} \right] \delta_{i,j} \right)_{t=t+1} - \left(\sum_{i,j} \left[B_{i,j} \right] \delta_{i,j} \right)_{t=t} \tag{6}$$

Initially, each node is assigned to the community according to the result of the K-means clustering. There are two ways to start the iteration: the reassignment process of the nodes in the first iteration is random, or the reassignment process of the nodes in the second iteration follows the order of their positions along the chromosome. We can choose any one of these two approaches for iteration and repeat until a local maximum of modularity is reached. The choice of approach has minor impact on the results. But to keep the results consistent, we used the second option.

Data Preprocessing of RNA-Seq

We used StringTie (Pertea et al., 2015) to obtain the FPKM (Fragments per KB per million mapped reads) of each gene of GM12878 and K562, and then used Ballgown (Frazee et al., 2015) to identify DEGs between the two cell lines.

Data Preprocessing of Epigenomic Tracks

We use bwtool (Pohl and Beato, 2014) to calculate the mean value of the histone and replication time signals in each 100 kb or 10 kb bin of downloaded bigWig tracks. For the histone signal analysis in **Figures 1E,F**, we adopt the Z-score to normalize the signal values.

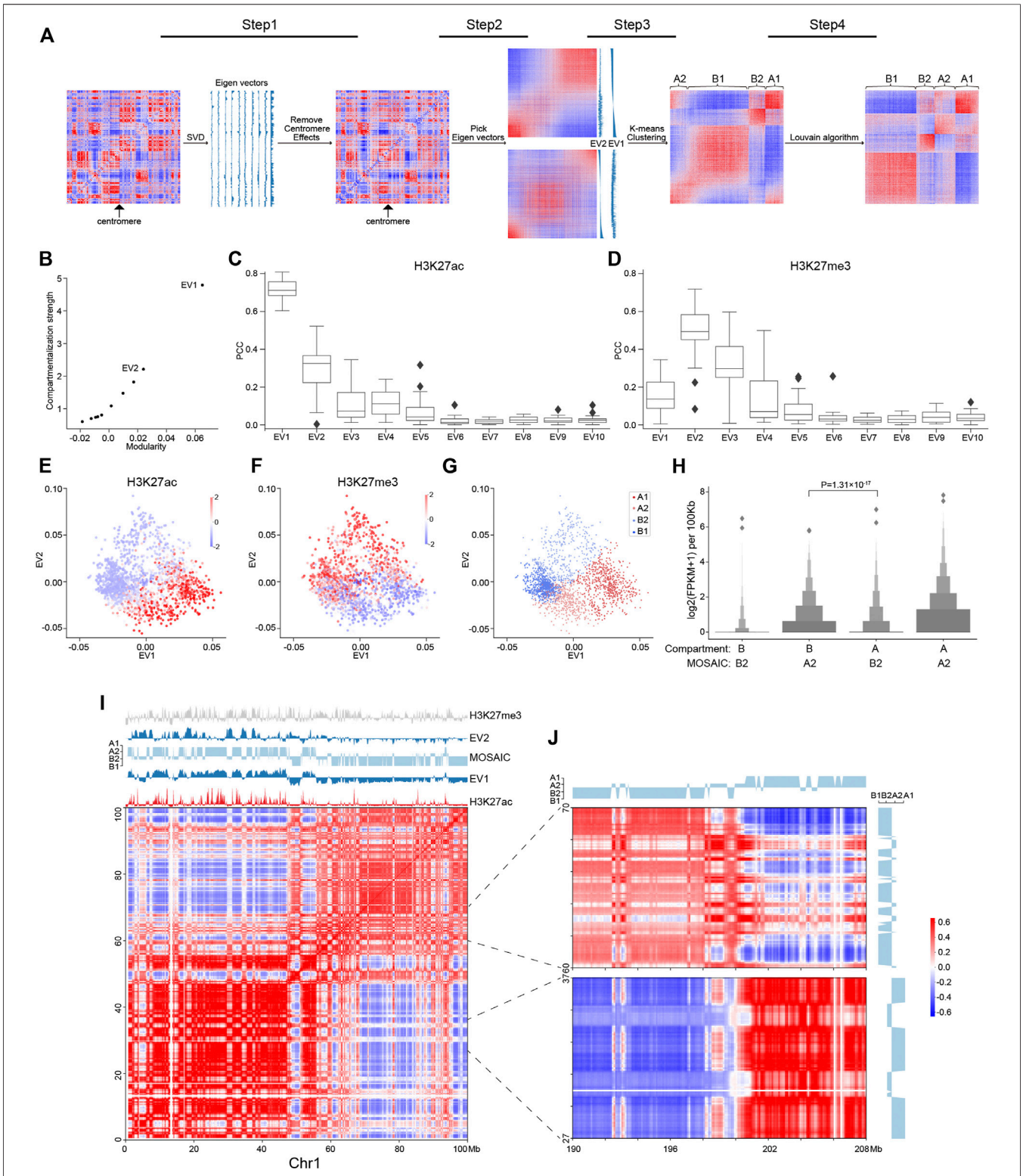

FIGURE 1 | Workflow of MOSAIC and overview of compartmental states. **(A)** MOSAIC workflow starts with an individual intra-chromatin O/E matrix. After two rounds of SVD, potential centromere effect is removed from the matrix, and the two eigenvectors that can best reflect the structural features are selected for K-means clustering. Finally, a modified Louvain algorithm is applied to refine the results of K-means clustering. **(B)** The average compartmentalization strength of all chromosomes versus the average modularity of all chromosomes for the top-ten EVs in terms of modularity for each chromosome. **(C,D)** Pearson correlation coefficients between the top-ten EVs sorted by modularity and H3K27ac **(C)** and H3K27me3 **(D)**, respectively. **(E,F)** Enrichment of H3K27ac **(E)** and H3K27me3 **(F)** for all 100 kb bins of

(*Continued*)

FIGURE 1 | chromosome 1 in relation to EV1 and EV2, respectively. The enrichment values are ZSCOR-normalized. **(G)** Compartmental states of all 100 kb bins of chromosome 1 in relation to EV1 and EV2. **(H)** Gene expression for the A2 and B2 regions identified by the MOSAIC and A/B compartment schemes (*p* value calculated by one-tailed Mann–Whitney rank test). **(I)** Hi-C correlation heatmap of Chr1: 0–100 Mb along with EV1, EV2, H3K27me3, H3K27ac, and compartmental states. **(J)** Hi-C correlation heatmap of Chr1: 27–37, 60–70 Mb and the distal region Chr1: 190–208 Mb.

Enrichment Analysis of Epigenomic Signal and Chromatin State

For analysis of epigenomic signal enrichment and chromatin state enrichment, we adapted a previously described method (Rao et al., 2014). Enrichment of the epigenomic signal can be calculated as the median within the cluster divided by the total median at a resolution of 100 kb. For chromatin state enrichment calculations, we first binned the annotations into 100 kb bins and then calculated the proportion of each annotation within each bin. Finally, we divide the mean value of the proportion within the cluster by the mean value of the proportion across all bins. To facilitate the display of annotations in **Figure 3B**, we abbreviate the annotations defined by ChromHMM, replacing "Transcription Associated" with "Transcription", "Low activity proximal to active states" with "Low Activity", "Candidate Strong Enhancer" with "Strong Enhancer", "Candidate Weak enhancer/DNase" with "Weak Enhancer", "Heterochromatin/Repetitive/Copy Number Variation" with "Heterochromatin" and "Distal CTCF/Candidate Insulator" with "Insulator", respectively.

Compartment Border Evaluation

Nine histone modification marks representing euchromatin, heterochromatin, and Polycomb-repressed chromatin were used to evaluate the effect of insulation between compartmental states. Briefly, we divided 10 bins on the left and right side of each compartment border and recorded the mean value of the histone modification signal in each 10 kb bin. Then, a N*20 matrix was generated in which the 20 columns represent the signal strength of two continuous compartments, and the N rows represent borders.

Gene Ontology Analysis

Because A1 and A2 represent euchromatin, whereas B1 and B2 represent heterochromatin and Polycomb-repressed chromatin, respectively, we performed GO analysis on the transcribed protein-coding genes in A1 and A2, and the non-transcribed protein-coding genes in B1 and B2. B1 and B2 contain 1,385 and 1,335 non-transcribed protein-coding genes, respectively, whereas A1 and A2 contain 7,605 and 1,667 transcribed protein-coding genes, respectively. Because common GO analysis software, including Metascape, limits the number of input genes to 3,000 or fewer, we randomly selected 3,000 genes from A1 as input to Metascape to search for enriched GO terms.

Compartmentalization Strength

We obtained the corresponding eigenvectors after SVD of the O/E matrix M. Each eigenvector can separate the O/E matrix M into two parts through its sign. We define the compartmentalization strength of each eigenvector based on the compartmentalization (Schwarzer et al., 2017) defined as:

$$C_s = \frac{\overline{\sum_{i \in p90, j \in p90} M_{ij}} + \overline{\sum_{i \in p10, j \in p10} M_{ij}}}{2 * \overline{\sum_{i \in p90, j \in p10} M_{ij}}} \tag{7}$$

where $p90$ represents the location of the values above the 90th percentile in this eigenvector, and $p10$ represents the location of the values below the 10th percentile.

Clustering Metrics

The Silhouette coefficient is defined as:

$$S_i = \frac{b_i - a_i}{\max(b_i, a_i)} \tag{8}$$

where b_i is the mean nearest-cluster distance for all other clusters, and a_i is the mean intra-cluster distance for genomic bin i. We calculated the Silhouette coefficient for each genomic bin; the mean Silhouette coefficient of all genomic bins was used to evaluate the quality of the clustering. Cosine distance was used as our distance measure.

For a set of data E that has been clustered into k clusters of size n_E, the Calinski–Harabasz score is defined as the ratio of the sum of inter-cluster dispersion and intra-cluster dispersion for all clusters:

$$s = \frac{tr(B_k)}{tr(W_k)} \times \frac{n_E - k}{k - 1} \tag{9}$$

where $tr(B_k)$ is the trace of the inter-cluster dispersion matrix, and $tr(W_k)$ is the trace of the intra-cluster dispersion matrix, defined by:

$$B_k = \sum_{q=1}^{k} n_q (c_q - c_E)(c_q - c_E)^T$$

$$W_k = \sum_{q=1}^{k} \sum_{x \in C_q} (x - c_q)(x - c_q)^T$$

where C_q is the set of points in cluster q, c_q is the center of cluster q, c_E is the center of E, and n_q is the number of points in cluster q.

We used the silhouette_score and calinski_harabaz_score functions from scikit-learn package version 0.19.0 for the calculations.

RESULTS

MOSAIC Overview and Eigenvector Exploration

To identify compartmental components in an individual chromosome, we started with the 100 kb-binned intra-chromosomal "observed over expected" (O/E) matrix (Lieberman-Aiden et al., 2009) and used Hi-C data from the

GM12878 cell line (Rao et al., 2014). First, we applied SVD to the O/E matrix (**Figure 1A**). The EV2 and EV3 obtained by dimension reduction of the genome-wide interaction matrix are able to reveal the enrichment of centromere-centromere and telomere-telomere contacts (Imakaev et al., 2012). For some chromosomes, one of the top EVs of intra-interactions reflects the long and short arms (Schmitt et al., 2016b), potentially due to the barrier role of centromeres in suppressing contacts across chromosome arms (Muller et al., 2019). To disentangle the impact of the centromere effect on compartmentalization analysis, we quantified EVs in terms of arm bias, and the O/E matrices showing a strong centromere effect were corrected accordingly (**Section 2**).

With the O/E matrices corrected for centromere effect, we performed another round of SVD and explored the structural and functional implications of the top EVs. From the perspective of modularity, compartmentalization is a modular feature, with the A/B compartments representing two structural modules of the genome. To quantify an EV in terms of structural segregation, a modularity score (Newman, 2006), termed Q_{EV}, can be assigned, supposing a bipartition based on its sign (Newman, 2013) along the chromosome. The top-ten EVs were re-ranked by their Q_{EV} in descending order for each chromosome. EV1, which corresponds to the A/B compartments, gave the highest Q_{EV} (**Supplementary Figure S1A**). We also adapted compartmentalization strength as previously described (Schwarzer et al., 2017) to quantify the structural segregation represented by each eigenvector, and found that the compartmentalization strength was quantitatively consistent with modularity (**Figure 1B** and **Supplementary Figure S1B**). As expected, EV1 possesses both the highest compartmentalization and the highest Q_{EV}, followed by EV2. Therefore, MOSAIC adopts modularity as the target function for compartment identification.

To quantify the epigenomic and functional implications of EVs, their correlation to various histone modifications was calculated and denoted as R_{EV}. Due to arbitrary sign designation, we used the absolute value to measure the correlation strength. For the top-ten EVs, Q_{EV} and R_{EV} were highly correlated (**Figure 1C** and **Supplementary Figures S1C,D**), indicating concordance between chromatin structure and epigenetic states. In other words, EVs with high modularity were also highly correlated to histone modification signals. Such a monotonic relationship held for all histone modifications, with H3K27me3 to be the only exception (**Figure 1D** and **Supplementary Figure S1C**); for H3K27me3, the EV with the highest correlation was EV2 rather than EV1 (**Figure 1D** and **Supplementary Figure S1D**). In contrast to the marginal correlation between EV1 and H3K27me3, EV2 and H3K27me3 had a Pearson correlation coefficient (PCC) around 0.5. This prominent distinction prompted us to interrogate EV2 and its relation to H3K27me3.

To further explore the relationship between EVs and histone modifications, we made scatter plots using EV1 and EV2, and examined the patterns of histone modifications of all 100 kb-binned regions in chromosome 1 (**Figures 1E,F**). For H3K27ac (**Figure 1E**), as expected, the regions with the most enriched and depleted signals lay within the two extremes along EV1, corresponding to the typical A and B compartments, respectively. For H3K27me3, which lacked a clear level gradient along EV1, EV2 became a strong indicator (**Figure 1F**). Moreover, the span of EV2 was largely orthogonal to EV1, as indicated by the convex outline of the data points. According to the algebraic interpretation of EVs, the magnitude of EV for a region represents the strength of the noted pattern. As shown in **Figures 1E,F**, regions with extreme EV2 values also have marginal EV1, indicating that the regions ambiguous under EV1 are in fact structurally distinctive, as pinpointed by EV2 (**Supplementary Figures S1E,F**). Taken together, these findings indicate that EV1 and EV2 are complementary in characterizing the structural and functional states of chromatin.

Based on above results, EV1 and EV2 were most representative and informative in terms of both structural modularity and epigenomic relatedness. Therefore, we chose them to conduct K-means clustering to identify compartmental states. We employed the sum of squared errors (SSE) to determine the optimal number of clusters (**Supplementary Figure S1G**). Accordingly, cluster number of four were adopted in subsequent analysis. Because compartmentalization is a global feature, we sought to optimize the identification of compartmental states by Louvain algorithm (Blondel et al., 2008) to maximize modularity and obtained the final four compartmental states (**Figure 1A**; **Section 2**).

MOSAIC Captures Structural and Functional Features More Accurately Than A/B Compartments

The distribution of the four modules for chromosome 1 is shown in **Figure 1G**. The two modules lying to the left and right, well separated by EV1, correspond to canonical A/B compartments, and were therefore denoted as A1 and B1 (**Supplementary Figure S1H**). The other two modules on the top and bottom were mostly ambiguous in the A/B compartment scenario. After expanding to the 2D space of EV1 and EV2, they are clearly segregated along the dimension of EV2 and coincide with by H3K27me3 levels. Because of the repressive nature of H3K27me3, the modules with and without H3K27me3 signals were denoted as B2 and A2, respectively. To further confirm the activity distinction of A2 and B2, we checked the discrepant regions between our partition and the A/B partition based on only the sign of EV1 (**Supplementary Figure S1H**). In the matching matrix, the regions originally considered B, but now identified as A2, are transcriptionally more active than the regions originally considered A but now identified as B2 (**Figure 1H**), justifying the discriminating power of EV2 in these regions ambiguous under EV1. In summary, each individual chromosome of the genome was partitioned into four compartmental states, labeled A1, B1, A2, and B2 (**Supplementary Figure S2A**).

To examine our partition results, we visually inspected intra-chromosomal heatmaps along with compartmental states identified by MOSAIC. As an example, **Figure 1I** shows the correlation heatmap of the initial 100 Mb of chromosome 1. This

area can be coarsely split from near the center, as noted by the two large predominantly red squares along the diagonal of the heatmap. Default A/B compartments marked by the sign of EV1 correctly reflect this coarse-grained pattern, with the two halves of EV1 profile having opposite signs (**Figure 1I**). Nevertheless, interspersed on each half, we can see plentiful secondary and fine plaid patterns on the heatmap. These patterns are visualized as fine bands in alternating white and red that are lighter than the surrounding background of the large red squares. Notably, on the EV1 profile, most of these bands just correspond to local depressions without sign switching. That is, most of these fine patterns are not identified as compartment-switching between conventional A and B compartment. Strikingly, such moderate contrasts, although secondary to the distinction of A/B compartments, are highlighted by EV2 with sharp peaks or plateaus on the positive or negative side. Furthermore, the plaid pattern extends and spans the whole chromosome, including distal regions on the other arm (**Figure 1J**), confirming the global compartmental features of the newly identified modules. Such patterns of the interspersed regions, which also correspond to the points on the top and bottom areas of **Figure 1G**, ground the structural significance of EV2 and the rationale of distinguishing these regions from the surrounding typical A/B compartments.

Consistent with their strong correlation with H3K27me3 (**Figure 1D**), the EV2 peaks were specifically marked by H3K27me3 signals (**Figure 1F**). Such coincidence points to an intuitive interpretation of EV2 and supports the heterochromatic nature of the corresponding regions. Therefore, these regions were denoted as B2 to differentiate them from B1, which represents constitutive heterochromatin. Taken together, the area on chromosome 1 confirms the orthogonality of EV1 and EV2 in capturing the patterns in heatmaps, and supports the refined partition of the chromosome into four compartmental states, with the newly identified A2 and B2 which are typically small regions surrounded by large areas of B1 and A1, respectively.

Whole Genome Characterization of Micro-Compartment and Compartment Borders

To examine our findings on all chromosomes, we conducted statistical analyses on the whole genome. The newly identified A2 and B2 regions exhibit features distinct from those of A1 and B1, which correspond to the conventional A and B compartments. In terms of genomic constitutions, A1 and B1 sum to 70.9% of the genome. A2 and B2 constitute the other 29.1% (**Figure 2A**). Nevertheless, the number of regions in each compartmental state is comparable (**Supplementary Figure S2B**), while their length distributions differ dramatically (**Figure 2B**). Consistent with the example shown in **Figure 1I**, regions of A2 and B2 are significantly shorter than those of A1 and B1. The ratios and lengths of these four compartmental states are consistent with those of 100 kb at 50 kb resolution (**Supplementary Figures S2C,D**). Therefore, we denoted A2 and B2 as micro-compartment.

To characterize the degree of compartmental conservation, we examined the results of A/B compartmentalization in 16 cell lines (Schmitt et al., 2016a) and defined a compartmental conservation score for each region by counting the A/B compartment assignment in the cell lines. As shown in **Figure 2C**, A1 and B1 correspond to consensus A/B compartments. By contrast, A2 and B2 tend to switch between A and B compartments, thus correspond to dynamic and cell-type specific compartmental regions from the perspective of A/B compartments.

Given the four compartmental states, there will be six combinations in terms of genomic neighborhood. Among the six neighboring types, majority are A1–B2 and A2–B1 (**Figure 2D**). These two types are highly enriched relative to the results of compartment identity permutations, whereas the other four neighboring types are under-represented (**Figure 2E**). The ratios of these six neighboring types are consistent with those of 100 kb at 50 kb resolution (**Supplementary Figures S2E,F**). Combined with the region length distributions (**Figure 2B**), these features coincide with the observations on chromosome 1 that B2 and A2 are embedded within large areas of A1 and B1, respectively. The depletion of A1–A2 and B1–B2 means that subtypes of euchromatin/heterochromatin do not tend to be next to each other. Notably, the depletion of A1–B1 suggests that drastic switches of compartmental states are rarer than expected. Instead, genomic neighborhood favors moderate switches between euchromatin heterochromatin, e.g., A1–B2 and A2–B1.

Analysis of the compartment borders revealed 2,205 MOSAIC-specific borders, 270 A/B compartment-specific borders and 1,849 overlapping borders (**Supplementary Figure S2G**). 99.2% of MOSAIC-specific borders were associated with A2/B2 (**Figure 2F**), confirming that the MOSAIC-specific borders are the consequence of the newly identified micro-compartments. We found that the MOSAIC-specific borders have distinct patterns of histone modification transition depending on the border types (**Figure 2F**). For example, the transition patterns of H3K27me3 and H3K36me3 at the A1–B2 borders are different from those at the A2–B1 borders. For the overlapping borders, MOSAIC revealed that they can be distinguished and divided into six types with distinct epigenetic patterns, which the A/B compartment scheme cannot discern. Together, these results tell us that micro-compartments are fragmented regions embedded in large areas of A1 and B1. These architectural features are reinforced by the distinct histone modification transitions at corresponding compartment borders.

Compartmental States Show Distinct Epigenetic and Functional Features

Our analysis of Hi-C data revealed that individual chromosomes segregate into four compartmental components. To characterize their epigenetic and functional features, we explored the pattern of histone modifications and replication timing using a previously described method (Rao et al., 2014) (**Figure 3A**). The results revealed that A1 and A2 regions are both enriched in histone modifications representing euchromatin or transcriptional activities, with A1 enriched to a higher degree than A2. B2 regions are specifically

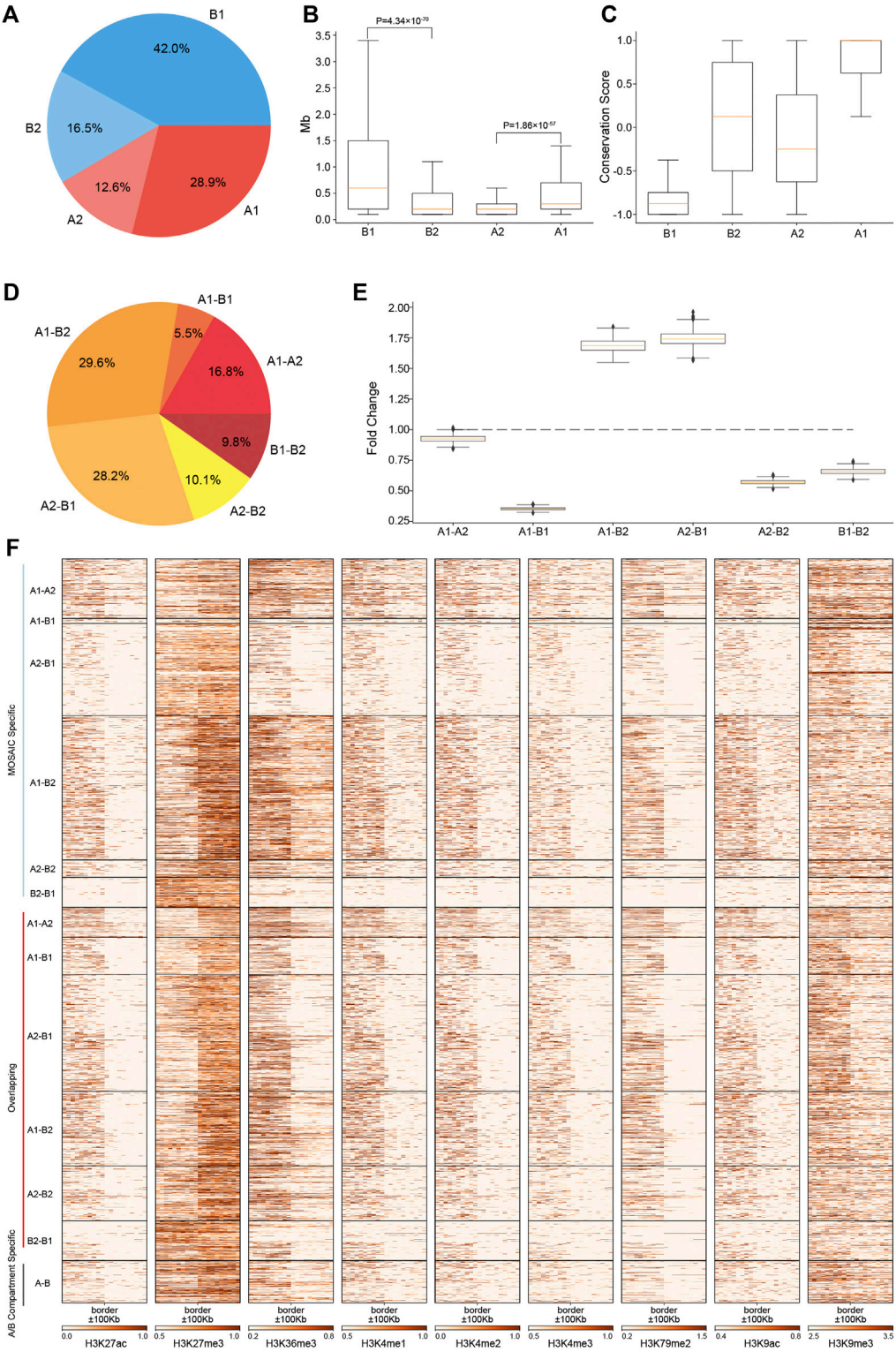

FIGURE 2 | Regions of A2 and B2 are short, constitute a minor proportion of the genome and are embedded within B1 and A1, respectively. **(A)** Proportions of four compartmental states throughout the genome. **(B)** Length distributions of four compartmental states throughout the genome (*p* value calculated by one-tailed Mann–Whitney rank test). **(C)** Compartmental conservations of four compartmental states in 16 cell lines. The values range from −1 to 1. A value of 1 means that the region is in the A compartment in all 16 cell types; a value of −1 means that it is in the B compartment in all 16 cell types; and a value of 0 means that it is in A compartment in 8 cell lines and in B compartment in the other eight cell lines. **(D)** Proportions of border types in terms of neighboring compartmental states. **(E)** Distribution of enrichment of true border types relative to random border types with neighborhoods shuffled 1,000 times. **(F)** Heatmaps of various histone modifications at the grouped border types.

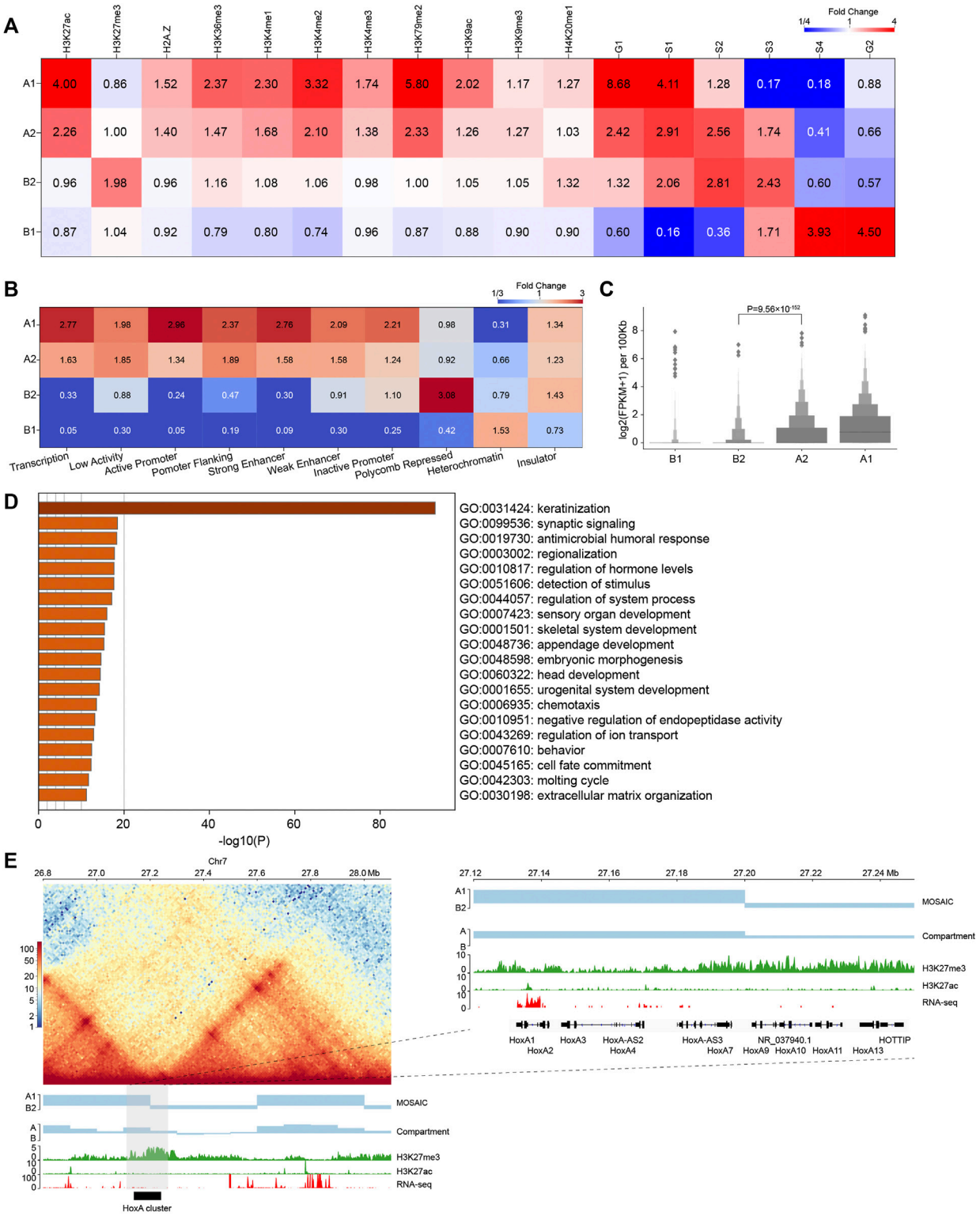

FIGURE 3 | Compartmental states show distinct epigenomic features. **(A)** Heatmap of histone mark and replication timing enrichment for four compartmental states throughout the genome. **(B)** Heatmap of ChromHMM annotation enrichment for four compartmental states throughout the genome. **(C)** Gene expression of four compartmental states throughout the genome (*p* value calculated by one-tailed Mann–Whitney rank test). **(D)** Enriched GO terms of genes in B2. **(E)** Left panel: Hi-C interaction heatmap along with compartmental states, A/B compartment, H3K27me3, H3K27ac, and RNA-seq around the HoxA cluster. Right panel: zoom-in view of the region where the HoxA cluster is located.

enriched in H3K27me3 signals. B1 regions are barely enriched in any histone modifications, indicative of constitutive heterochromatin. Consistent with this, the replication timing enrichment of S3, S4, and G2 in these regions reveals that they are late replicating.

In addition, we used chromatin state annotation by ChromHMM (Ernst and Kellis, 2012) to examine their relation to compartmental states (**Figure 3B** and **Supplementary Figure S2H**). A1 is most enriched in the chromatin states "Transcription" and "Active Promoter", and A2 is most enriched in "Low Activity" and "Promoter Flanking". On the other side, B2 is highly enriched in "Polycomb Repressed" chromatin, and B1 is enriched in "Heterochromatin". These results are consistent with the epigenetic pattern enrichment shown above, also support that functionally similar regions are in close spatial proximity (Hnisz et al., 2017).

We then compared gene expression among compartmental states (**Figure 3C**). Consistent with the ChromHMM annotation results, A1 primarily contains expressed genes. Genes in A1 expressed at higher levels than A2. By contrast, B1 and B2 have minimal levels of transcription. Again, expression level of genes in A2 are significantly higher than those in B2, supporting our definition of the two types of compartments as euchromatin and heterochromatin, respectively. In summary, chromatin activities ascend in the order of B1, B2, A2, A1 in terms of gene expression. Then, we conducted Gene Ontology (GO) analyses using Metascape (Zhou et al., 2019). The result in **Figure 3D** shows that genes in B2 are enriched for developmental and cell type-specific functions. In particular, the late cornified envelope (LCE) gene cluster and keratin (KRT) gene cluster associated with keratinization are both located in B2 (**Supplementary Figure S3C**). Overall, A2 and B2 are enriched in regulation-related genes (**Supplementary Figures S3A,B**).

Considering the enrichment of regulation-related genes in A2 and B2, the compartmental states may contribute to their transcriptional regulation. Taking the genomic regions containing the *HoxA* cluster as an example, the corresponding EV1 values are positive, indicating that the whole *HoxA* cluster is in the conventional A compartment. In contrast, our scheme more accurately separated the cluster into A1 and B2, which is marked by H3K27me3 (**Figure 3E**) (Lopez-Delisle et al., 2020). Our result is also more consistent with the finding that H3K27me3 system forms repressive microenvironments at *Hox* gene clusters in the more active nuclear environment in differentiated cells (Vieux-Rochas et al., 2015). This example highlights the necessity of recognizing regions as heterochromatic B2 embedded in large area of conventional euchromatic A compartment. Together, these results reveal that compartments identified by structural modularity are in different chromatin states and contain genes that perform different classes of functions in the cell.

MOSAIC Outperforms Subcompartment Identification Methods in Structural Partition of Individual Chromosomes

Existing methods using inter-chromosomal contacts divide the genome into six subcompartments that are functionally and spatially distinct (Rao et al., 2014; Xiong and Ma, 2019; Ashoor et al., 2020). It is of interest to compare the results of subcompartment identification using inter-chromosomal interactions and MOSAIC based on intra-chromosomal interactions. For this purpose, we used clustering metrics and modularity to quantify the effects of different compartmental partitions. We adopted the Silhouette coefficient (Rousseeuw, 1987) and Calinski-Harabasz score (Calinski and Harabasz, 1974) as indicators of consistency within a cluster. The Silhouette coefficient ranges from −1 to 1, whereas the Calinski-Harabasz score, as a standard of variance ratio, has values greater than zero. In both cases, higher values indicate better clustering performance. Modularity is a system property for characterizing community structure in networks, i.e., dense connection within groups but sparse connections between groups (Newman, 2006). Modularity is utilized to measure the independency of clusters. Higher modularity indicates closer proximity of regions in the same cluster. The three available subcompartment identification methods are denoted as Rao_HMM (Rao et al., 2014), Xiong_SNIPER (Xiong and Ma, 2019), and Ashoor_SCI (Ashoor et al., 2020), respectively. As shown in **Figure 4A**, MOSAIC outperforms all subcompartment identification methods in the division of individual chromosomes, in terms of both cluster consistency and independence.

While O/E matrix reflects direct interactions, the correlation matrix measures the relationships of interaction patterns. We utilized both types of information at the intra-chromosomal level to measure the clustering effect (**Figures 4B,C** and **Supplementary Figures S4A,B**). The compartmental states identified by MOSAIC consistently exhibit preferential interactions within the same states and depleted interactions between different states. The only exception is the marginal enrichment (0.04) between A1 and A2 in the correlation matrix. By contrast, for subcompartments of Rao_HMM, A1–A2 and B2–B3 exhibit enriched interactions. When we compared the extent of segregation between compartmental states, we observed the strongest segregation for A1–B1, followed by A2–B2. This is consistent with the results of the SVD analysis (**Figure 1G**), in which A1 and B1 are separated along EV1, whereas A2 and B2 are separated along EV2. Given the strong depletion of A1–B1 and A2–B2 in terms of genomic neighborhood (**Figure 2D**), these two pairs tend to segregate in both 1D and 3D.

Because subcompartments are obtained using genome-wide inter-chromosomal data, they may not consistently reflect fine interaction patterns within individual chromosomes. As shown in **Figure 4D**, there are four shaded areas, three of which are 100–200 kb regions not discerned by subcompartment identification methods. MOSAIC distinguishes all three as A2. Histone modification and gene expression information support the distinction of these regions from their surroundings as active chromatin. Moreover, the inter-chromatin interaction matrix shows that the MOSAIC-specific compartmental state A2 exhibits a different interaction pattern than the

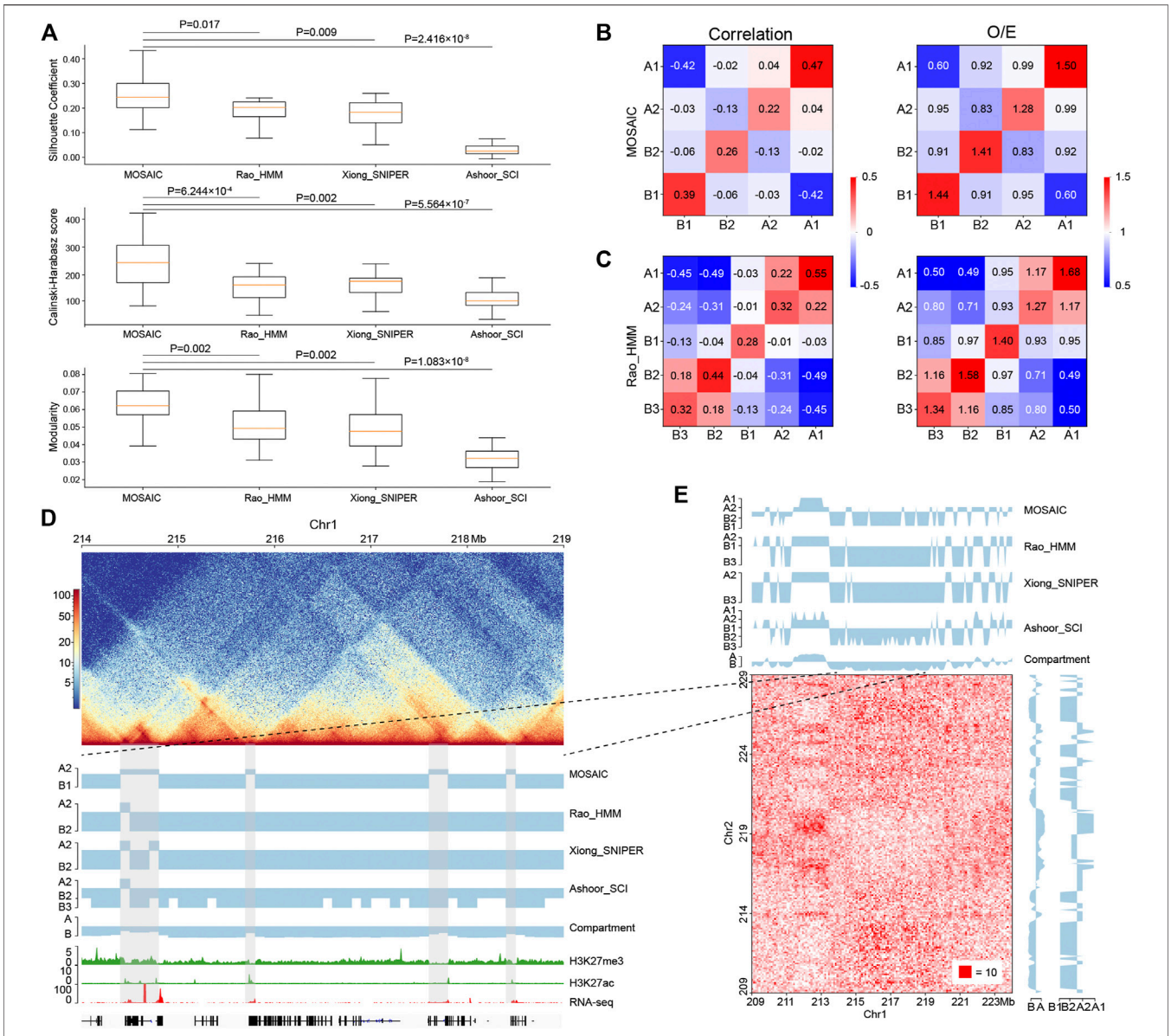

FIGURE 4 | Comparisons of MOSAIC with the subcompartment identification methods. **(A)** Silhouette coefficients (top), Calinski–Harabasz scores (middle), and modularity (bottom) were used to compare the results of MOSAIC, Rao_HMM, Xiong_SNIPER and Ashoor_SCI in the division of individual chromosomes (*p* values calculated by one-tailed Mann–Whitney rank test). **(B,C)** Heatmap of mean values of the correlation matrix (left panel) and mean values of the O/E matrix (right panel) of various compartmental state combinations identified by MOSAIC **(B)** and Rao_HMM **(C)**. Because subcompartment B4 was identified only on chromosome 19, it was excluded from this analysis. **(D)** An example showing that MOSAIC reveal more detailed and accurate compartmentalization at the intra-chromosomal level than subcompartments at the three shaded areas. **(E)** Inter-chromosomal heatmap showing that the specific compartmental states (shaded areas) identified by MOSAIC in **(D)** also have different interaction patterns relative to the surrounding regions.

surrounding B1 region. We can also see that the MOSAIC-specific compartmental states A2 in chromosome 1 have weak interactions with A2 on chromosome 2. The reason why this A2 was not detected by subcompartment identification methods might be that the inter-chromosomal interactions are not as pronounced as the intra-chromosomal interactions (**Figure 4E**). Taken together, MOSAIC exhibits higher accuracy and sensitivity than subcompartment identification methods in characterizing chromatin compartmentalization at intra-chromosomal level.

Compartmental States Accurately Reflect Gene Expression Dynamics

Since GM12878 and K562 are closely related and the characteristics of compartmental states identified by the

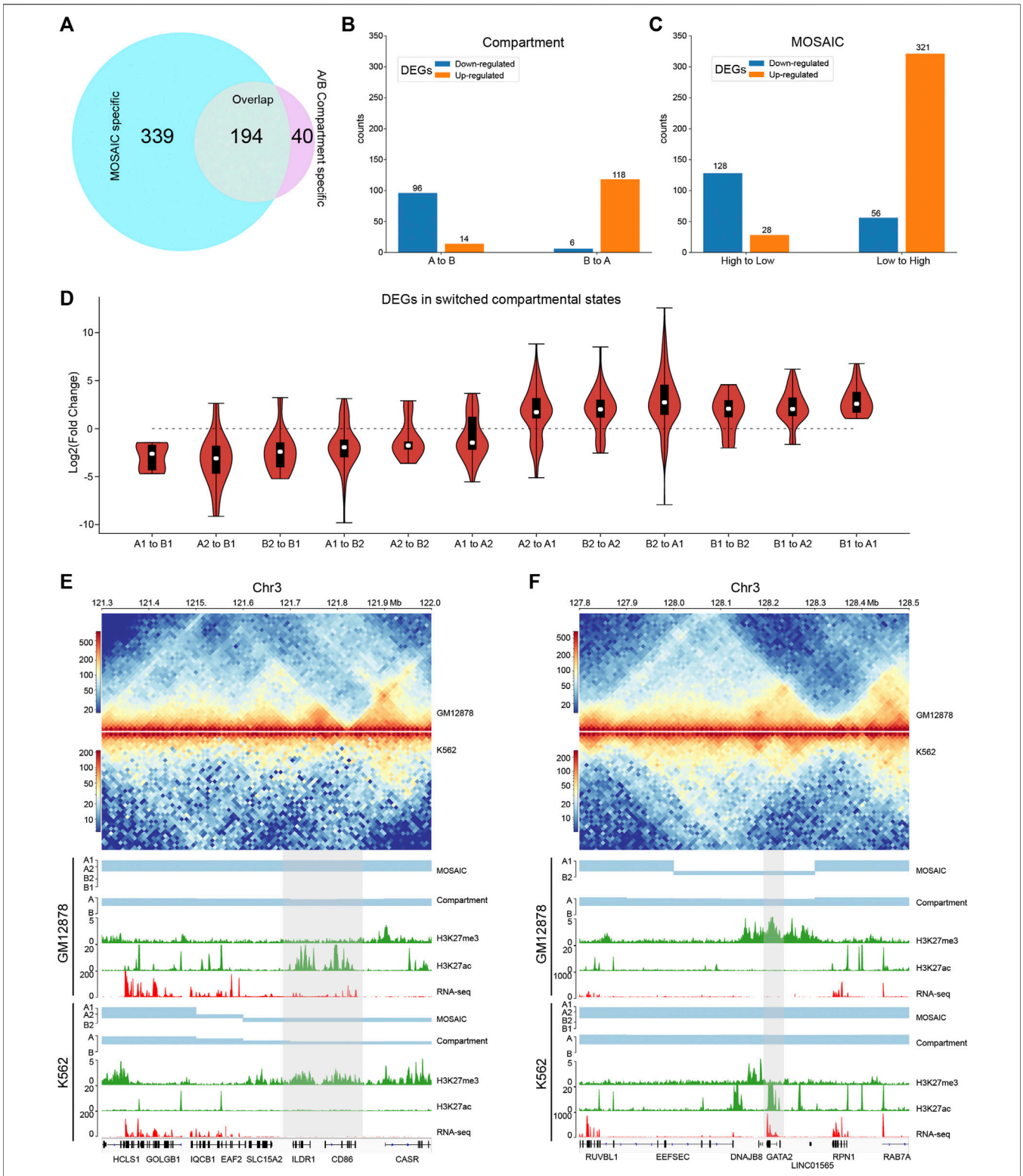

FIGURE 5 | MOSAIC reveals concordance of gene expression and compartmental state dynamics between K562 and GM12878 cells. **(A)** Venn diagram showing the overlap of DEGs that can be covered by the two schemes of compartmentalization analysis. **(B,C)** Boxplots showing categorization of DEGs with A/B compartment switches **(B)** and with compartmental state switches **(C)**. For DEGs, the ratio of gene expression level in GM12878 to K562 were calculated. Switches in compartmental state are grouped into descending and ascending order, in terms of transcription activity implications of compartmental states. **(D)** Fold change distribution of DEGs grouped by compartmental state switches. Fold changes were calculated as $\log_2(FPKM_{GM12878}/FPKM_{K562})$. **(E,F)** Two examples on chromosome 3 showing concordance between compartmental states identified by MOSAIC and expression of CD86, ILDR1 **(E)** and GATA2 **(F)** located in the shaded areas. In both cases, A/B compartment scheme does not identify switch of compartment.

MOSAIC on K562 are comparable to those of GM12878 (**Supplementary Figure S5**), we compared these two cell lines to investigate the relationship between compartmental state and gene expression. We identified the regions in which compartmental states change and then analyzed differentially expressed genes (DEGs). Among the 1,420 DEGs (>50% change in FPKM and p value < 0.05) (Djebali et al., 2012), 533 are in regions with compartmental changes using MOSAIC. By comparison, only 234 DEGs are in regions with switches in A/B compartments. Among the 533 DEGs, 194 are shared between the two methods, whereas the other 339 are MOSAIC-specific (**Figure 5A** and **Supplementary Figure S6A**).

Next, we examined the extent of concordance between gene expression and compartmental states. **Figure 5B** shows that only 214 DEGs are consistent with A/B compartment switches. By contrast, MOSAIC identified 449 DEGs that exhibit concordance between compartmental states and gene expression (**Figure 5C**). Particularly, as shown in **Figure 5D** and **Supplementary Figure S6B**, compartmental changes are accompanied by consistent changes in gene expression: genomic regions changing toward more active compartments contain genes that are upregulated, and vice versa. This suggests that MOSAIC scheme is much more sensitive than conventional A/B compartment scheme in characterizing the dynamics of transcriptional regulation. A/B compartment scheme might severely under-estimate the chromatin architectural basis of transcriptional regulation.

Figures 5E,F show two typical regions containing DEGs without A/B compartment switches. In the A/B compartment scheme, the regions shown in **Figure 5E** are constantly A compartment in both GM12878 and K562. However, RNA-seq data showed that two genes, *CD86* and *ILDR1*, were specifically expressed in GM12878. This differential expression was accompanied by dynamic H3K27ac and H3K27me3 signals. MOSAIC reconciles the dynamics by revealing that this region switches from A1 in GM12878 to B2 in K562. Interestingly, both *CD86* and *ILDR1* are associated with immunoglobulin, which is related to the function of GM12878 cells. A recent study on GM12878 cells showed that this region contains a super-enhancer that regulates both genes (Kleinstern et al., 2020). Moreover, a genome-wide association study validated one locus at 3q13.33 (rs9831894) that is significantly inversely associated with the risk of diffuse large B-cell lymphoma (DLBCL) (Kleinstern et al., 2020). Accurate compartmentalization characterization by MOSAIC might provide insight of this phenotypical difference from the perspective of fine-scale chromatin architecture dynamics.

As shown in **Figure 5F**, the entire region belongs to the A compartment in both K562 and GM12878 cells according to EV1 values. By contrast, MOSAIC identified the region containing *GATA2* as B2 in GM12878 and A1 in K562. *GATA2* encodes a transcription factor that is involved in stem cell maintenance and plays a significant role in hematopoietic development (de Pater et al., 2013). From a functional standpoint, *GATA2* should be specifically expressed in K562. Consistent with this, in GM12878 cells, the genomic region containing *GATA2* has an elevated level of H3K27me3 to repress its expression.

In summary, the capacity to interpret differential expression between cells based on changes in the A/B compartment was limited. By contrast, concordance between gene expression and compartmentalization was significantly improved in MOSAIC. The fine-scale compartmental dynamics involving micro-compartments revealed by MOSAIC might shed light on mechanisms of transcriptional regulation of genes important for cell-type specific function and disease.

DISCUSSION

We applied SVD to intra-chromosomal contacts by extending the conventional A/B bipartition which uses only EV1 to a comprehensive compartmentalization framework using further EVs. Although EV1 matches the overall plaid patterns, EV2 provides a striking separation of ambiguous regions with marginal EV1 values, also highlighted by corresponding bimodality of H3K27me3. The regions pinpointed by EV2 are typically interspersed with activities opposite to their surrounding environments. Specifically, A2 and B2, referred to as micro-compartments, are embedded in large areas of B1 and A1, respectively. Our scheme provides a fine-scale and dynamic view of chromatin compartmentalization. While EV1 characterizes the euchromatin (A1) and constitutive heterochromatin (B1) corresponding to the structurally stable parts of the chromatin, bipolar EV2 captures the dynamic aspects of chromatin that are structurally flexible and functionally regulated. **Figure 6** gives a schematic representation of fine-scale chromatin compartmentalization and its dynamics in the space spanned by the top two EVs.

In comparison with other state-of-the-art compartment identification tools (Durand et al., 2016; Wolff et al., 2018) based on PCA or SVD, MOSAIC provides automatic and accurate facilities allowing universal identification of potentially diverse compartmentalization patterns. It is recognized that the first PC in PCA may not represent compartmental features for all chromosomes (Schmitt et al., 2016b). Consequently, manual curation is required for proper PC selection in current A/B compartments identification tools. Using modularity as the criterion, we demonstrated the power of MOSAIC for automatically selecting EVs that correctly reflect structural and functional features of chromatin. MOSAIC scheme is also robust in the sense that it accommodates scenarios with structure-function relations that might differ among diverse species and cell types. For instance, the layout of some chromosomes may have more than four compartmental states, or have histone modifications other than H3K27me3, which is indicative of its fine-scale structure. MOSAIC allows detection of such compartmental structures by proper, automatic EV selection and robust clustering. Considering the A/B compartments based on the leading EV as a first-order approximation of the true compartmental states, MOSAIC represents a natural extension that uncovers finer-scale patterns by including the second EV. MOSAIC framework also allows generalization to EVs that further

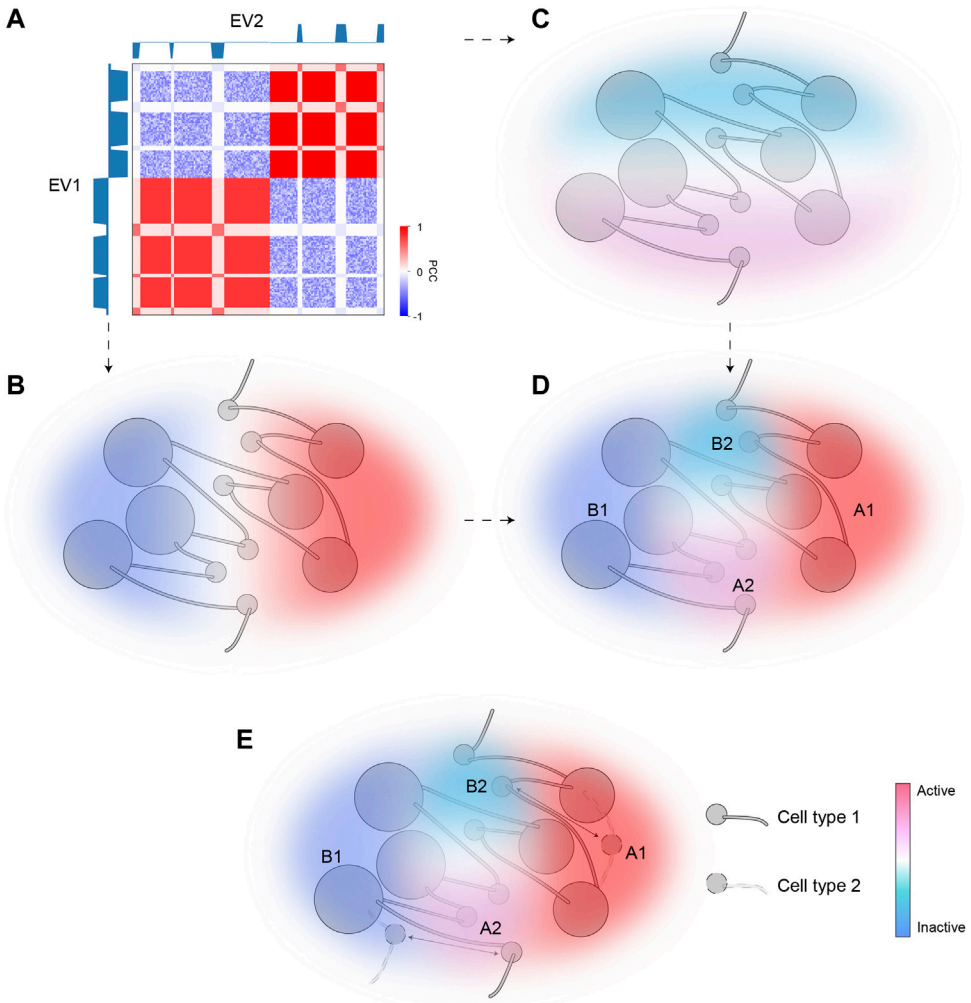

FIGURE 6 | A model of dynamic chromatin compartmentalization highlighting genomic and architectural features of micro-compartments. **(A)** The schematic Hi-C heatmap of an area of the genome containing A1, A2, B1, and B2 compartment, identified based on EV1 and EV2. EV1 can separate this segment into two major categories (B1, left; A1, right), but cannot distinguish between the small regions (A2/B2) embedded in each major category, whereas EV2 marks A2 and B2. **(B)** The sign of EV1 can separate A1 and B1. Color represents activities (red for active; blue for inactive). **(C)** The sign of EV2 can separate A2 and B2. Color represents activities (red for active; blue for inactive). **(D)** The combination of EV1 and EV2 can accurately identify all four compartmental states and accurately reflect their activities. **(E)** A diagram showing the dynamic regions between the two cell types are the micro-compartment A2 and B2. Arrows indicate cell-type specific compartmental states for certain regions.

qualify the structural and functional analysis described above.

Currently, conventional A/B compartment scheme is widely adopted for comparative analysis on chromatin compartmentalization. However, comparison between GM12878 and K562 cells using MOSAIC revealed extensive and fine-scale concordance between compartmental state and gene expression which are missed by A/B compartments. The rationale behind this finding is that transcriptional regulation through switching between constitutive euchromatin and heterochromatin is rarer than expected. Instead, our results show that the majority of compartmental neighborhood and dynamics are A1–B2 and B1–A2. Incapability of A/B compartment scheme in capturing such types of compartmental neighborhood and dynamics leads to its limitations. Consequently, the analysis based on bipartite compartments is prone to under-estimate the role of compartmentalization in transcriptional regulation, exemplified by the cases of *CD86*, *ILDR1* and *GATA2* shown above.

Spatial segregation of regions marked by H3K27me3 has been observed in various systems and loci (Vieux-Rochas et al., 2015; Du et al., 2020; Johnstone et al., 2020). For example, a study focusing on *Hox* gene clusters revealed clustering of these regions with other H3K27me3 targets (Vieux-Rochas et al., 2015). Another investigation of mouse development revealed Polycomb-associating domains and local compartment-like structures (Du et al., 2020). In addition, a study (Johnstone et al., 2020) comparing primary and tumor cells found an

intermediate compartment highlighted by DNA hypomethylation and H3K27me3. Based on this observation, the authors proposed a three-compartment model. Generalizing the above findings, our work provides a global and comprehensive view of the role of H3K27me3 in compartmentalization within the framework of MOSAIC. In each chromosome, EV2 correlates with H3K27me3. The strong modularity of the resultant B2 compartment supports the aggregation of the H3K27me3-marked regions. Furthermore, our results show that the typical H3K27me3-marked regions are embedded in large areas of the conventional A compartment, consistent with the contrasting active nuclear environment of repressive *Hox* gene clusters previous observed (Vieux-Rochas et al., 2015).

A recent study (Kundu et al., 2017) based on 5C using ESCs revealed PRC1-mediated self-interacting domains that are similar to TADs but smaller in size. Our observations of *HoxA* cluster in GM12878 cells suggest specific compartmental states corresponding to H3K27me3-marked regions (**Figure 3E**). While these results support the role of H3K27me3 in chromatin architecture at various layers, the interplay between domain and compartmental states, and the underlying histone modifications, remains to be elucidated. Based on the close connection between H3K27me3 and micro-compartments, it is promising that compartmentalization might play a significant role in histone mark spreading.

While TADs represent local segregation of adjacent genomic regions, compartments represent global aggregation of 1D-separated regions with similar genomic features and epigenetic and functional states. Under the current working model, TADs influence transcription by providing a local environment that facilitates enhancer-promoter looping within the same TAD and insulates enhancer-promoter contact across TADs (Andrey et al., 2013). However, the perturbation experiments (Rao et al., 2017; Schwarzer et al., 2017) removing TADs and loops did not lead to proportionally differential gene expression. It is important to note that these perturbations reinforced compartmentalization, pointing to the potential role of compartmentalization in transcriptional regulation, which might be under-estimated by conventional comparative analysis based on A/B compartments as shown above. The extensively improved concordance between expression and compartmental states revealed by MOSAIC supports transcriptional regulation through compartmentalization as an important mechanism.

Hierarchical modularity is a general organizing principle of living systems. From the perspective of hierarchical chromatin architecture in the nucleus (Gibcus and Dekker, 2013), compartmental states add another layer to hierarchy of its spatial modular organization. In this hierarchy, the well-known chromosome territories represent the highest level of modularity, highlighted by chromosome-specific radial distribution within the nucleus. Under this top-level circumstance, A/B compartments form the second level of organization which spatially segregate chromatin with different activities at a large scale. Embedded in large areas of A or B compartments, fine-scale regulation of individual genes toward the opposite activities, represent the third level of architecture, in which loci are spatially segregated into micro-compartments. The precise and delicate regulation of genome function necessitates this hierarchical compartmentalization of chromatin architecture. Corresponding perturbation experiments might provide exact answer to the mechanistic connection between chromatin architecture and function.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

LL conceived the project. ZW and LL designed the algorithm. ZW implemented the algorithms, developed software, and performed computational analysis. WZ, QZ and JX contribute to Hi-C data collection and pre-processing. CH and ZQ advised on data analysis. LL and ZW wrote the manuscript with input from all authors.

ACKNOWLEDGMENTS

The authors would like to thank Victor Corces for insightful comments on the manuscripts. We acknowledge financial support from the National Natural Science Foundation of China (No. 31771430 to LL, No. 31571347 to CH), Huazhong Agricultural University Scientific and Technological Self-innovation Foundation (to LL), Hubei Hongshan Laboratory (to LL) and Shenzhen Science and Technology Innovation Commission (No. JCYJ20170412152835439 to CH).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.845118/full#supplementary-material

REFERENCES

Andrey, G., Montavon, T., Mascrez, B., Gonzalez, F., Noordermeer, D., Leleu, M., et al. (2013). A Switch between Topological Domains Underlies HoxD Genes Collinearity in Mouse Limbs. *Science* 340, 1234167. doi:10.1126/science.1234167

Ashoor, H., Chen, X., Rosikiewicz, W., Wang, J., Cheng, A., Wang, P., et al. (2020). Graph Embedding and Unsupervised Learning Predict Genomic Sub-compartments from HiC Chromatin Interaction Data. *Nat. Commun.* 11, 1173. doi:10.1038/s41467-020-14974-x

Belaghzal, H., Borrman, T., Stephens, A. D., Lafontaine, D. L., Venev, S. V., Weng, Z., et al. (2021). Liquid Chromatin Hi-C Characterizes Compartment-dependent Chromatin Interaction Dynamics. *Nat. Genet.* 53, 367–378. doi:10.1038/s41588-021-00784-4

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast Unfolding of Communities in Large Networks. *J. Stat. Mech.* 2008, P10008. doi:10.1088/1742-5468/2008/10/P10008

Calinski, T., and Harabasz, J. (1974). A Dendrite Method for Cluster Analysis. *Comm. Stats. - Theor. Methods* 3, 1–27. doi:10.1080/03610927408827101

de Pater, E., Kaimakis, P., Vink, C. S., Yokomizo, T., Yamada-Inagawa, T., van der Linden, R., et al. (2013). Gata2 Is Required for HSC Generation and Survival. *J. Exp. Med.* 210, 2843–2850. doi:10.1084/jem.20130751

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature* 485, 376–380. doi:10.1038/nature11082

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., et al. (2012). Landscape of Transcription in Human Cells. *Nature* 489, 101–108. doi:10.1038/nature11233

Dong, P., Tu, X., Chu, P.-Y., Lü, P., Zhu, N., Grierson, D., et al. (2017). 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments. *Mol. Plant* 10, 1497–1509. doi:10.1016/j.molp.2017.11.005

Du, Z., Zheng, H., Huang, B., Ma, R., Wu, J., Zhang, X., et al. (2017). Allelic Reprogramming of 3D Chromatin Architecture during Early Mammalian Development. *Nature* 547, 232–235. doi:10.1038/nature23263

Du, Z., Zheng, H., Kawamura, Y. K., Zhang, K., Gassler, J., Powell, S., et al. (2020). Polycomb Group Proteins Regulate Chromatin Architecture in Mouse Oocytes and Early Embryos. *Mol. Cel* 77, 825–839. e7. doi:10.1016/j.molcel.2019.11.011

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., et al. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cel Syst.* 3, 95–98. doi:10.1016/j.cels.2016.07.002

Ernst, J., and Kellis, M. (2012). ChromHMM: Automating Chromatin-State Discovery and Characterization. *Nat. Methods* 9, 215–216. doi:10.1038/nmeth.1906

Frazee, A. C., Pertea, G., Jaffe, A. E., Langmead, B., Salzberg, S. L., and Leek, J. T. (2015). Ballgown Bridges the gap between Transcriptome Assembly and Expression Analysis. *Nat. Biotechnol.* 33, 243–246. doi:10.1038/nbt.3172

Furlan-Magaril, M., Ando-Kuri, M., Arzate-Mejía, R. G., Morf, J., Cairns, J., Román-Figueroa, A., et al. (2021). The Global and Promoter-Centric 3D Genome Organization Temporally Resolved during a Circadian Cycle. *Genome Biol.* 22, 162. doi:10.1186/s13059-021-02374-3

Gibcus, J. H., and Dekker, J. (2013). The Hierarchy of the 3D Genome. *Mol. Cel* 49, 773–782. doi:10.1016/j.molcel.2013.02.011

Hildebrand, E. M., and Dekker, J. (2020). Mechanisms and Functions of Chromosome Compartmentalization. *Trends Biochem. Sci.* 45, 385–396. doi:10.1016/j.tibs.2020.01.002

Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K., and Sharp, P. A. (2017). A Phase Separation Model for Transcriptional Control. *Cell* 169, 13–23. doi:10.1016/j.cell.2017.02.007

Imakaev, M., Fudenberg, G., McCord, R. P., Naumova, N., Goloborodko, A., Lajoie, B. R., et al. (2012). Iterative Correction of Hi-C Data Reveals Hallmarks of Chromosome Organization. *Nat. Methods* 9, 999–1003. doi:10.1038/nmeth.2148

Johnstone, S. E., Reyes, A., Qi, Y., Adriaens, C., Hegazi, E., Pelka, K., et al. (2020). Large-Scale Topological Changes Restrain Malignant Progression in Colorectal Cancer. *Cell* 182, 1474–1489. e23. doi:10.1016/j.cell.2020.07.030

Ke, Y., Xu, Y., Chen, X., Feng, S., Liu, Z., Sun, Y., et al. (2017). 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. *Cell* 170, 367–381. e20. doi:10.1016/j.cell.2017.06.029

Kleinstern, G., Yan, H., Hildebrandt, M. A. T., Vijai, J., Berndt, S. I., Ghesquières, H., et al. (2020). Inherited Variants at 3q13.33 and 3p24.1 Are Associated with Risk of Diffuse Large B-Cell Lymphoma and Implicate Immune Pathways. *Hum. Mol. Genet.* 29, 70–79. doi:10.1093/hmg/ddz228

Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J. T., Sadreyev, R. I., et al. (2017). Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. *Mol. Cel* 65, 432–446. e5. doi:10.1016/j.molcel.2017.01.009

Larson, A. G., Elnatan, D., Keenen, M. M., Trnka, M. J., Johnston, J. B., Burlingame, A. L., et al. (2017). Liquid Droplet Formation by HP1α Suggests a Role for Phase Separation in Heterochromatin. *Nature* 547, 236–240. doi:10.1038/nature22822

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* 326, 289–293. doi:10.1126/science.1181369

Liu, Y., Nanni, L., Sungalee, S., Zufferey, M., Tavernari, D., Mina, M., et al. (2021). Systematic Inference and Comparison of Multi-Scale Chromatin Sub-compartments Connects Spatial Organization to Cell Phenotypes. *Nat. Commun.* 12, 2439. doi:10.1038/s41467-021-22666-3

Lopez-Delisle, L., Rabbani, L., Wolff, J., Bhardwaj, V., Backofen, R., Grüning, B., et al. (2020). pyGenomeTracks: Reproducible Plots for Multivariate Genomic Datasets. *Bioinformatics* 37, 422–423. doi:10.1093/bioinformatics/btaa692

Muller, H., Gil, J., and Drinnenberg, I. A. (2019). The Impact of Centromeres on Spatial Genome Architecture. *Trends Genet.* 35, 565–578. doi:10.1016/j.tig.2019.05.003

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B. R., Mirny, L. A., et al. (2013). Organization of the Mitotic Chromosome. *Science* 342, 948–953. doi:10.1126/science.1236083

Newman, M. E. J. (2006). Modularity and Community Structure in Networks. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8577–8582. doi:10.1073/pnas.0601602103

Newman, M. E. J. (2013). Spectral Methods for Community Detection and Graph Partitioning. *Phys. Rev. E* 88, 042822. doi:10.1103/PhysRevE.88.042822

Nichols, M. H., and Corces, V. G. (2021). Principles of 3D Compartmentalization of the Human Genome. *Cel Rep.* 35, 109330. doi:10.1016/j.celrep.2021.109330

Niu, L., Shen, W., Shi, Z., Tan, Y., He, N., Wan, J., et al. (2021). Three-dimensional Folding Dynamics of the Xenopus Tropicalis Genome. *Nat. Genet.* 53, 1075–1087. doi:10.1038/s41588-021-00878-z

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie Enables Improved Reconstruction of a Transcriptome from RNA-Seq Reads. *Nat. Biotechnol.* 33, 290–295. doi:10.1038/nbt.3122

Pohl, A., and Beato, M. (2014). Bwtool: a Tool for bigWig Files. *Bioinformatics* 30, 1618–1619. doi:10.1093/bioinformatics/btu056

Rao, S. S. P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K.-R., et al. (2017). Cohesin Loss Eliminates All Loop Domains. *Cell* 171, 305–320. e24. doi:10.1016/j.cell.2017.09.026

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 159, 1665–1680. doi:10.1016/j.cell.2014.11.021

Rousseeuw, P. J. (1987). Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Mathematics* 20, 53–65. doi:10.1016/0377-0427(87)90125-7

Rowley, M. J., Nichols, M. H., Lyu, X., Ando-Kuri, M., Rivera, I. S. M., Hermetz, K., et al. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol. Cel* 67, 837–852. e7. doi:10.1016/j.molcel.2017.07.022

Sati, S., Bonev, B., Szabo, Q., Jost, D., Bensadoun, P., Serra, F., et al. (2020). 4D Genome Rewiring during Oncogene-Induced and Replicative Senescence. *Mol. Cel* 78, 522–538. e9. doi:10.1016/j.molcel.2020.03.007

Schmitt, A. D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C. L., et al. (2016a). A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cel Rep.* 17, 2042–2059. doi:10.1016/j.celrep.2016.10.061

Schmitt, A. D., Hu, M., and Ren, B. (2016b). Genome-wide Mapping and Analysis of Chromosome Architecture. *Nat. Rev. Mol. Cel Biol* 17, 743–755. doi:10.1038/nrm.2016.104

Schoenfelder, S., and Fraser, P. (2019). Long-range Enhancer-Promoter Contacts in Gene Expression Control. *Nat. Rev. Genet.* 20, 437–455. doi:10.1038/s41576-019-0128-0

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., et al. (2017). Two Independent Modes of Chromatin Organization Revealed by Cohesin Removal. *Nature* 551, 51–56. doi:10.1038/nature24281

Strom, A. R., Emelyanov, A. V., Mir, M., Fyodorov, D. V., Darzacq, X., and Karpen, G. H. (2017). Phase Separation Drives Heterochromatin Domain Formation. *Nature* 547, 241–245. doi:10.1038/nature22989

Takemata, N., Samson, R. Y., and Bell, S. D. (2019). Physical and Functional Compartmentalization of Archaeal Chromosomes. *Cell* 179, 165–179. e18. doi:10.1016/j.cell.2019.08.036

Vieux-Rochas, M., Fabre, P. J., Leleu, M., Duboule, D., and Noordermeer, D. (2015). Clustering of Mammalian Hox Genes with Other H3K27me3 Targets within an Active Nuclear Domain. *Proc. Natl. Acad. Sci. U.S.A.* 112, 4672–4677. doi:10.1073/pnas.1504783112

Wagner, G. P., Pavlicev, M., and Cheverud, J. M. (2007). The Road to Modularity. *Nat. Rev. Genet.* 8, 921–931. doi:10.1038/nrg2267

Wang, Y., Zhang, Y., Zhang, R., van Schaik, T., Zhang, L., Sasaki, T., et al. (2021). SPIN Reveals Genome-wide Landscape of Nuclear Compartmentalization. *Genome Biol.* 22, 36. doi:10.1186/s13059-020-02253-3

Wolff, J., Bhardwaj, V., Nothjunge, S., Richard, G., Renschler, G., Gilsbach, R., et al. (2018). Galaxy HiCExplorer: a Web Server for Reproducible Hi-C Data Analysis, Quality Control and Visualization. *Nucleic Acids Res.* 46, W11–W16. doi:10.1093/nar/gky504

Xiong, K., and Ma, J. (2019). Revealing Hi-C Subcompartments by Imputing Inter-chromosomal Chromatin Interactions. *Nat. Commun.* 10, 5069. doi:10.1038/s41467-019-12954-4

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets. *Nat. Commun.* 10, 1523. doi:10.1038/s41467-019-09234-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wen, Zhang, Zhong, Xu, Hou, Qin and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.