Nucleic Acids Research, 2024, 1–12
https://doi.org/10.1093/nar/gkae999
NAR Breakthrough Article

EXPRESSO: a multi-omics database to explore multi-layered 3D genomic organization

Liuyang Cai[1,†], Jun Qiao[1,†], Ruixin Zhou[1,†], Xinyi Wang[1,†], Yelan Li[1], Lei Jiang[6], Qiangwei Zhou[4,5,*], Guoliang Li[4,5,*], Tao Xu[2,3,*] and Yuliang Feng[1,*]

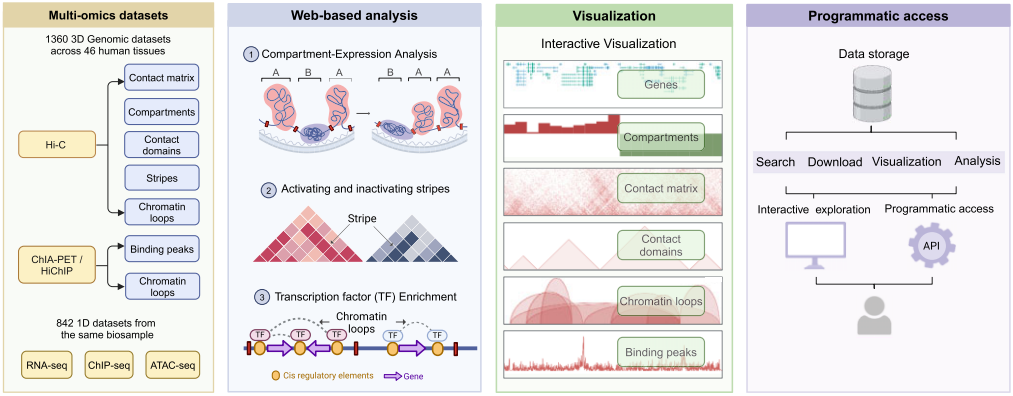
[1]Department of Pharmacology, Joint Laboratory of Guangdong-Hong Kong Universities for Vascular Homeostasis and Diseases, School of Medicine, Southern University of Science and Technology, Shenzhen 518055, China
[2]Inflammation and Immune Mediated Diseases Laboratory of Anhui Province, School of Pharmacy, Anhui Medical University, Hefei 230022, China
[3]The Second Affiliated Hospital of Anhui Medical University, Hefei 230601, China
[4]National Key Laboratory of Crop Genetic Improvement, Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan 430070, China
[5]Agricultural Bioinformatics Key Laboratory of Hubei Province, Hubei Engineering Technology Research Center of Agricultural Big Data, Key Laboratory of Smart Farming Technology for Agricultural Animals, 3D Genomics Research Center, College of Informatics, Huazhong Agricultural University, Wuhan 430070, China
[6]Guangdong Provincial Geriatrics Institute, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510080, China

[*]To whom correspondence should be addressed. Tel: +86 755 88012564; Fax: +86 755 88012564; Email: fengyl@sustech.edu.cn
Correspondence may also be addressed to Tao Xu. Email: xutao@ahmu.edu.cn
Correspondence may also be addressed to Guoliang Li. Email: guoliang.li@mail.hzau.edu.cn
Correspondence may also be addressed to Qiangwei Zhou. Email: qwzhou@mail.hzau.edu.cn
[†]Contributed equally to this study.

Abstract

The three-dimensional (3D) organization of the human genome plays a crucial role in gene regulation. EXPloration of Regulatory Epigenome with Spatial and Sequence Observations (EXPRESSO) is a novel multi-omics database for exploration and visualization of multi-layered 3D genomic features across 46 different human tissues. Integrating 1360 3D genomic datasets (Hi-C, HiChIP, ChIA-PET) and 842 1D genomic and transcriptomic datasets (ChIP-seq, ATAC-seq, RNA-seq) from the same biosample, EXPRESSO provides a comprehensive resource for studying the interplay between 3D genome architecture and transcription regulation. This database offers diverse 3D genomic feature types (compartments, contact matrix, contact domains, stripes as diagonal lines extending from a genomic locus in contact matrix, chromatin loops, etc.) and user-friendly interface for both data exploration and download. Other key features include REpresentational State Transfer application programming interfaces for programmatic access, advanced visualization tools for 3D genomic features and web-based applications that correlate 3D genomic features with gene expression and epigenomic modifications. By providing extensive datasets and tools, EXPRESSO aims to deepen our understanding of 3D genomic architecture and its implications for human health and disease, serving as a vital resource for the research community. EXPRESSO is freely available at https://expresso.sustech.edu.cn.

Graphical abstract



Received: August 15, 2024. **Revised:** September 26, 2024. **Editorial Decision:** October 6, 2024. **Accepted:** October 17, 2024

Introduction

The human genome is folded into hierarchical three-dimensional (3D) organization, which can be dissected into multiple levels. The development of chromosome conformation capture (3C) assay and its variants, including Hi-C (1), HiChIP (2) and chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) (3), and others, have greatly expanded our knowledge of 3D genomic organization. At the large scale, megabase-scale A/B compartments segregate the genome into activating (A) and inactivating (B) regions (1). At the submegabase scale, contact domains represent regions of the genome that interact more frequently with themselves than with neighboring regions (4), serving as fundamental units of genome organization that constrain enhancer-promoter interactions. Chromatin loops bring distant regulatory elements such as enhancers and promoters into close proximity (3). Recent studies have identified so-called stripes, which appear as as diagonal lines extending from a genomic locus in contact matrices and are believed to form through a loop extrusion process when a loop anchor interacts with the entire domain at high frequency (5,6). Stripes are often enriched in super enhancers (6), though repressive stripes also exist (7).

In recent years, the 3D genomic states of various cell lines, tissues and disease conditions have been elucidated. Large collaborative projects such as ENCODE (8) and 4D Nucleome (9) have generated a large number of 3D genomic data. Additionally, sequencing repositories such as GEO (10) database have seen a continuous increase in the number of 3D genomic datasets. Collecting these data to create a comprehensive human 3D genome database is a highly meaningful endeavor. Several initiatives have already been undertaken to address this challenge. For example, ChromLoops (11) and HiChIPdb (12) have collected public HiChIP and ChIA-PET data, while 3DIV (13) focuses on genomic structural variations and their impact on cancer 3D genomes. These studies provide rich resources for 3D genomic research.

However, the above databases have some limitations. Firstly, they often feature limited assay types and thus limited data volumes, reflecting only specific aspects of the human 3D genome. Secondly, the availability of data types for download is limited, especially in Hi-C data, which depict various 3D genomic features such as compartments, contact domains, stripes and chromatin loops. Thirdly, understanding the genome's structure–function relationship requires integrating 3D genome structure with other omics data (e.g. RNA-seq, ChIP-seq and ATAC-seq). Given the cell- and tissue-specific nature of regulatory elements, using 1D data from the same biosample may better reflect these relationships. Fourthly, considering the diversity of data types, hierarchies, and volumes, developing a programmatic application programming interface (API) can facilitate more flexible data retrieval. Lastly, improving the visualization of multi-layered 3D genomic features remains an area requiring enhancement.

Here, we introduce EXPloration of Regulatory Epigenome with Spatial and Sequence Observations (EXPRESSO; https://expresso.sustech.edu.cn), a multi-omics database designed to explore and visualize multi-level 3D genomic features and their associations with the epigenome and gene expression. EXPRESSO integrates 1, 360 3D genomic datasets, including Hi-C, HiChIP and ChIA-PET, across 46 human tissues. These datasets were processed using standardized pipelines for each data type, providing the community with various 3D genomic features such as compartments, contact domains, stripes and chromatin loops. To investigate the relationship between the 3D genome and other epigenomic data, we processed 842 RNA-seq, ChIP-seq and ATAC-seq datasets from the same biosamples. To demonstrate the usage of multi-omics data integration, we developed three web-based applications: (a) the differential expression levels of genes located between A and B compartments; (b) the stripes marked by H3K27ac and H3K27me3 modifications; and (c) the enrichment of transcription factor (TF) motifs on chromatin loops. To facilitate data access, we created a set of REpresentational State Transfer (REST) APIs for programmatic queries and downloads. Additionally, we developed an associated 3D genome visualization browser to effectively visualize these 3D genomic features.

EXPRESSO provides a comprehensive platform to study the interplay between 3D genome architecture and gene regulation, aiming to advance our understanding of the genome's spatial organization and its implications for human health and disease.

Materials and methods

Database overview

We uniformly processed each 3D genomic dataset and uploaded the intermediate files to cloud storage. For Hi-C data, we provide various files including the contact matrices, compartments, contact domains, stripes, and chromatin loops. For HiChIP and ChIA-PET, we provide chromatin loops and coverage files (Figure 1). All relevant metadata is stored using MySQL format (https://www.mysql.com/). We utilized the Vue.js framework (https://vuejs.org/) to create the front-end pages for our database. Additionally, we developed an associated genome browser using Canvas (https://developer.mozilla.org/en-US/docs/Web/API/Canvas_API), D3.js (https://d3js.org/) and WebGL (https://get.webgl.org/) to display various 3D genomic features. Statistical graphs are plotted using ECharts (https://echarts.apache.org/en/index.html). The genome browser and statistical graphs are integrated into the front-end webpage. For each 3D genomic dataset, we attempted to collect corresponding 1D omics data of the same biosample and study, including RNA-seq, ATAC-seq and H3K27ac/H3K27me3/CTCF ChIP-seq data. Utilizing biosamples that have both 3D and 1D genomic data, we developed three web-based applications to explore the relations between 3D genomic organization and other epigenomic data: (a) compartment-expression analysis; (b) stripe atlas; and (c) TF enrichment. Additionally, we implemented a REST API using the Python framework FastAPI (https://fastapi.tiangolo.com/), which allows easy access to this resource from any programming language environment.

Data collection

We searched the GEO repository for 3D genomic datasets (Hi-C, RNAPII / H3K27ac ChIA-PET or HiChIP) datasets related to the species 'Homo sapiens'. In addition to 3D genomic datasets, we retrieved RNA-seq, CTCF ChIP-seq, H3K27ac ChIP-seq, H3K27me3 ChIP-seq and ATAC-seq datasets available for corresponding biosamples. In total, we acquired 1360 3D genomic datasets along with 334 corresponding RNA-seq, 169 H3K27ac ChIP-seq, 73 H3K27me3 ChIP-seq, 140 CTCF ChIP-seq and 126 ATAC-seq datasets.

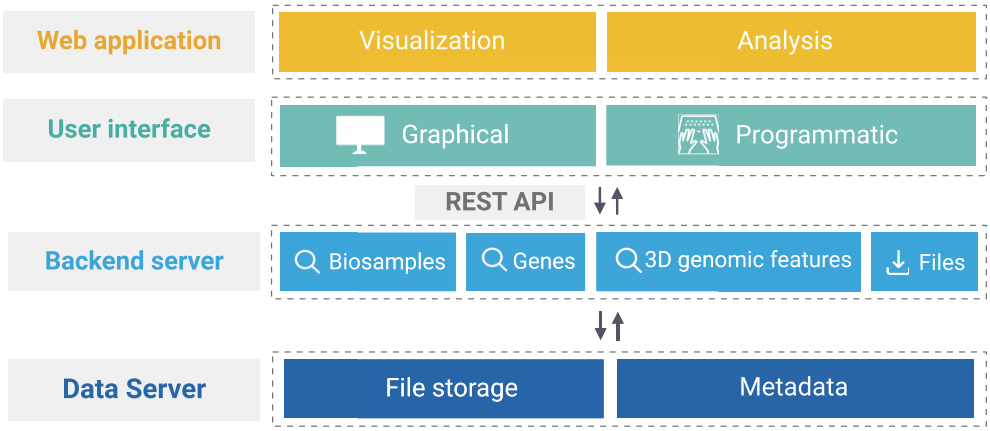


Figure 1. Overview of the functionalities in EXPRESSO database EXPRESSO integrates multiple types of 3D genomic datasets and 1D genomic datasets across various human tissues. The database provides a user-friendly interface for data exploration and visualization, as well as REST APIs for programmatic access.

Data processing

Hi-C and HiChIP data were processed using the HiC-Pro (14) pipeline with default parameters, adjusting the LIGATION_SITE parameter according to the specific restriction enzyme used. Specifically, pair-end reads were aligned to the human reference genome assembly version hg38. Multi-mapped, duplicated and other unvalid read pairs were discarded. Valid pairs were converted to *.hic* format with Juicer (15) or *.cool* format with cooltools (16). Hi-C contact matrices were normalized with cooler balance function (16). A and B compartments were identified using decomposed eigenvectors (E1) from contact matrices of 100 kb resolution with cooltools (16). The compartment scores (E1) were adjusted for GC content for each bin. Genome-wide contact insulation scores and domain boundaries were calculated using cooltools (16) insulation command. Chromatin loops were identified with cooltools (16) dots command at 10 kb resolution. ChIA-PET datasets were processed using the ChIA-PIPE (17) pipeline with default settings. RNA-seq data were mapped using STAR (18) version 2.5.3a, converted to BigWig format with the bamCoverage command from deepTools (19). Gene expression data with TPM normalization were quantified using featureCounts (20). ChIP-seq and ATAC-seq data were mapped using Bowtie2 (21). The resulting BAM files were deduplicated using Picard (https://broadinstitute.github.io/picard/) version 1.26. Reads with a quality score below 20 were filtered out, and the remaining reads were sorted and indexed using SAMtools (22). BAM files were then converted to BigWig format using the bamCoverage command from deepTools (19), and peaks were called using the macs2 (23) callpeak command.

Calculate the relations between genes and 3D genomic features

Gene annotation files were downloaded from GENCODE v29 (https://www.gencodegenes.org/human/release_29.html). For each of the 3D genomic features (compartment, contact domain, stripe and chromatin loop), we computed its relationships with gene promoters, defined as the region extending 2.5 kb upstream and downstream of the transcription start site. Specifically, for each dataset, we determined whether each gene promoter is located in A or B compartment, within a contact domain or at its boundary, at a chromatin loop anchor, or within a stripe. The intersections were calculated using bedtools (24).

Genome browser development

We developed a genome browser to visualize multi-level 3D genomic features, including compartments, contact domains, contact matrices, domain boundaries, chromatin loops, stripes and coverage files. This browser is a pure front-end library implemented using the Vue.js framework. Visualizations were created using Canvas (https://developer.mozilla.org/en-US/docs/Web/API/Canvas_API) and D3.js (https://d3js.org/) and WebGL (https://get.webgl.org/). We also designed a navigation bar that allows users to navigate and zoom through the genome. Additionally, each track type has specific configuration options to adjust the track's color, lines and other attributes.

Web-based applications

We developed three applications to facilitate multi-omics integration.

Compartment-expression analysis

For biosamples that include both Hi-C and RNA-seq data, we examined the relationship between 3D genome organization and gene expression by determining the compartment (A or B) in which each gene resides. By comparing the expression levels of genes in the A compartment versus those in the B compartment, we can identify significant expression differences that suggest a correlation between a gene's spatial positioning within the genome and its transcriptional activity. This application provides insights into how the structural organization of the genome influences gene regulation, highlighting potential mechanisms by which 3D genome architecture affects cellular function and phenotype.

Stripe atlas

Stripe is a recently identified 3D genomic feature linked to both active transcription and gene silencing (6). We identify stripes across all Hi-C datasets using Stripenn (6) and further classify them based on the presence of activating and inactivating histone marks, in biosamples that contain both 3D genomic data and H3K27ac or H3K27me3 ChIP-seq data. This

Table 1. Track type and file format for EXPRESSO visualization module

Track Type	File format
Chromatin loops	bedpe
Compartment	bigwig
Coverage	bigwig
Contact domain and domain boundary	bed
Contact matrix	hic
Gene track	UCSC gene annotation

classification allows us to distinguish between stripes associated with active transcription and those related to transcriptional repression. The Stripe Atlas application reveals the dual roles that stripes play in gene regulation and helps elucidate the complex regulatory landscape within the genome.

Transcription factor enrichment

By analyzing Hi-C and ATAC-seq data from the same biosample, we can map TF-binding motifs within chromatin loops and identify that TFs might be involved in loop formation. A total of 126 biosamples with both Hi-C and ATAC-seq data were used for the analysis. ATAC-seq peaks located within chromatin loop anchors were identified and motif scan was then conducted using Homer 4.11 (25). We used the HOCOMOCOv11 core human motif database (26) ('HOCOMOCOv11_core_HUMAN_mono_homer_format_0.0001.motif') for motif analysis.. This application enables us to distinguish common TFs that mediate loops across different conditions from specific TFs unique to subset of samples.

API development

To facilitate programmatic access to EXPRESSO's resources, we developed a suite of REST APIs. These APIs allow users to access metadata, search for 3D genomic data features, search for genes and download files. The 'metadata API' provides detailed information about the biosample, including their data type, biomaterial type, health status and other experimental conditions. The '3D genomic features search API' enables users to find specific chromatin structures such as chromatin loops, contact domains, and insulation boundaries. The 'gene search API' allows users to query specific genes and retrieve associated 3D genomic data. Finally, the 'download API' enables users to download relevant data files, thus streamlining the data acquisition process for further analysis.

Results

Graphical interface of the EXPRESSO database

We have developed a user-friendly web interface to facilitate exploration of EXPRESSO, featuring the following sections:

(1) Data Portal: This module allows users to search for specific datasets based on various criteria, filter results and navigate to individual dataset pages.
(2) Gene Search: This functionality enables users to search for specific genes. It provides detailed information about the genes and their associated 3D genomic features across all the samples.
(3) Multi-Omics Analysis: We have created three web-based applications that integrate 3D genomic data with other omics data types, enabling users to perform multi-omics analyses. Detailed introductions to each application are provided in the 'Applications' section below.
(4) Download: This section allows users to download all available datasets, including contact matrices, compartments, contact domains, chromatin loops, stripes and coverage files. Datasets associated with individual biosamples can also be downloaded from their respective webpages.
(5) API: This section introduces the APIs of EXPRESSO. It provides documentation and examples for using the API to retrieve metadata and download datasets programmatically, enabling more flexible and advanced access.
(6) Documentation: Comprehensive documentation for the database is available in this section. It includes guides, tutorials and reference materials to help users understand and utilize the database effectively.

Data portal page

In the 'Data Portal' section, we presented a bar plot illustrating the number of datasets available from each tissue (Figure 2A). We also provided a comprehensive Summary Table (Figure 2B) that allows users to filter samples based on various criteria such as body sites, health status, biomaterial and assay type. Users can access individual sample pages for more detailed information by selecting the appropriate samples after filtering.

Individual biosample page

Each biosample contains its own page, consisting of four sections: (1) Sample Information (Supplementary Figure S1A), which provides basic information about the dataset, including details like the sample's body site, health status, biomaterial and assay type; (2) File Download (Supplementary Figure S1B), where users can directly download various 3D genomic features associated with the biosample for further analysis. (3) Genome Browser (Figure 3A–H), an embedded module that allows users to visualize the 3D genomic features of the dataset, including gene tracks, compartments, contact matrices, domains, loops and stripes; and (4) File Table (Supplementary Figure S1C), which details the relationships between genes and the above 3D genomic features, such as whether a gene is located in the A or B compartment, within a contact domain or boundary, within a stripe or at a chromatin loop anchor.

Genome browser and file table sections

To facilitate intuitive exploration of multi-layered 3D data, we developed a web-based genome browser integrated with EXPRESSO. The browser fetches data from the file server and renders it on the client side. It supports widely used data formats for visualizing various levels of 3D genomic features. The supported track types and file formats are as follows (Table 1):

The genome browser includes a toolbar for seamless navigation to any genomic region (Figure 3A) and a configuration panel for each track to customize style and appearance (Figure 3B). Users can also enable a vertical line to assist the navigation.

For each dataset, we calculated the relations between genes and various 3D genomic features, including compartments, contact domains, stripes and chromatin loops (See 'Materials and methods' section); and stored the results in File Table section. The File Table indicates whether each gene is in the A

Figure 2. Biosample search features in EXPRESSO. (**A**) Bar plot displaying the count of 3D genomic datasets across various tissues. (**B**) An interactive summary table for filtering biosamples under various criteria.

or B compartment, inside a contact domain or at its boundary, whether it is involved in a stripe, and if it serves as a chromatin loop anchor.

Examples of 3D genome organization comparison across samples

Using the large datasets in EXPRESSO, users can compare and visualize 3D genomic features of related samples. We present three examples to demonstrate this strategy.

Switched compartment in prostate cancer cells

Human 3D genome contact maps display checkerboard patterns that represent the spatial compartmentalization of two main types of chromatin: activating and open A compartments and inactivating and more closed B compartments. Analyzing changes in gene compartments across multiple samples helps understand the impact of the 3D genome on the transcriptional regulation. For instance, *Wnt5a* (Wnt Family Member 5A) plays a crucial role in cell proliferation, differentiation, migration, adhesion and polarity. It can co-regulate

Figure 3. Individual biosample page overview, (**A**) Navigation bar of the EXPRESSO browser. (**B**) Configuration panel of the gene annotation track. (**C** and **D**). Example of compartment switch between healthy prostate cells (**C**) and PCa cell (**D**). The gene *Wnt5a* is highlighted. (**E** and **F**) Example of neo contact domain formation in DCM heart (**F**) compared with healthy heart (**E**). The neo domain that wrapped *NPPA* and *NPPB* gene is highlighted. (**G** and **H**) Example of multi-level 3D genome organization of *AFP* gene. Panel **G** displayed compartment and contact domain in Hi-C data. Panel **H** displayed chromatin interactions and 1D coverage using RNAPII ChIA-PET data.

the transactivation function of mutated androgen receptors and promote the proliferation of prostate cancer (PCa) cells in an autocrine manner (27). Another study found a substantial enrichment of *Wnt5a* in circulating tumor cells (CTCs) from enzalutamide-resistant PCa patients (28). In the EXPRESSO database, *WNT5A* shifts from the B compartment in healthy prostate cells (GSM3564252) to the A compartment in PCa cells (GSM4908718) (Figure 3C-D; Compartment track). Meanwhile, there are also specific chromatin loops at the *Wnt5a* site in PCa genome, consistent with its active state (Figure 3C-D; Loop track). Such compartmental changes can affect the accessibility of genomic regions to TFs or other regulatory proteins (29).

Neo contact domain in dilated cardiomyopathy

Changes in contact domain positioning or shifting of domain boundaries could reveal altered interactions of neighboring promoter-enhancer regions. We provided an example compar-

ing human healthy heart sample (GSM5029738, Figure 3E) and Dilated Cardiomyopathy (DCM) sample (GSM5029740, Figure 3F) to demonstrate *de novo* domain formation. In this example, the cardiac hypertrophy marker genes (30) *NPPB* (Natriuretic Peptide B), *NPPA* (Natriuretic Peptide A) and its antisense gene *NPPA-AS1* show strong Hi-C signals and form a neo contact domain in the DCM sample (Figure 3E and F; Domain Track, Boundary Track and Matrix Track), which are absent in the healthy heart. This observation aligns with study suggesting that the *NPPA-AS1* promoter may function as an enhancer to regulate *NPPA* and *NPPB* transcription in DCM (31).

Multi-level 3D genome organization of *AFP* locus in liver cancer

Liver cancer cell Hi-C data (GSM7656195) showed that the hepatocellular carcinoma marker gene *AFP* (α-fetoprotein) (32) localizes in a contact domain with *ALB* (Albumin) (Figure 3G). Hi-C matrix identified strong intra-TAD contact signals, indicating chromatin interactions between *AFP* and *ALB* (Figure 3G; Domain Track and Matrix Track). Consistent with Hi-C dataset, RNAPII ChIA-PET data (GSM5379365) further showed *AFP* is associated with strong RNAPII-binding (Figure 3H; Peak Track) and chromatin looping with *ALB* and an upstream enhancer element (Figure 3H; Loop Track). Hi-C excels in identifying compartment and domain-level patterns, whereas RNAPII ChIA-PET provides high-resolution insights into transcription-related enhancer-promoter interactions. By integrating Hi-C and RNAPII ChIA-PET data, users can obtain a comprehensive, multi-level view of a specific gene.

Search genes and individual gene page

We provide two methods for selecting genes: users can either directly input the gene name or filter for genes contained within a genomic region. For each gene, we display its relationships with four types of 3D genomic features: (1) the number of samples in which the gene is located in the A or B compartment, along with corresponding metadata, which can indicate if the gene resides in a specific compartment in certain tissues or diseases; (2) the number of samples in which the gene is located within contact domains or boundaries; (3) the samples in which the gene is found within a stripe; and (4) the samples in which the gene participates in chromatin interactions. The above results can suggest potential regulatory roles of the gene in a subset of samples. As an illustration, we ranked the gene *NEAT1* by the compartment E1 score (Figure 4A). The results show that *NEAT1* is in the A compartment in most samples (Figure 4B); however, it belongs to the B compartment in the cervical cancer cell line Hela (Figure 4C), indicating 3D genome is reorganized in cervical cancer cells.

Applications

Gene expression and compartment change

Compartmentalization in the genome refers to the spatial organization of chromatin into distinct regions, typically classified as A (activating) and B (inactivating) compartments, which are associated with differential gene expression and chromatin modifications (33), although some regions in A compartments are not transcribed (34). To visualize the relationship between gene expression and their compartmentalization, we used a box plot to compare expression levels of genes located in the A compartment versus those in the B compartment.

Stripe atlas

Stripes were recently observed in 3D contact maps and proposed as evidence for the loop extrusion model. Recent studies suggest that the stripe is associated with both activating and inactivating chromatin states (7). In this study, we detected 1, 891, 697 stripes across all Hi-C datasets with a median count of 1361 stripes per dataset. To further characterize these stripes, we analyzed their overlap with peak regions of H3K27ac, H3K27me3 and CTCF within the same biosample. This analysis identified 250 650 stripes associated with H3K27ac, 30 631 with H3K27me3 and 279 087 with CTCF. To illustrate the utility of the Stripe Atlas, we highlighted a stripe located in the *CCAT1* and *MYC* gene region of the cervical cancer cell line HeLa. The dashed box in the Hi-C matrix track highlighted the locations of the stripe starting from the gene *CCAT1* and extending to MYC gene (Figure 5A; Top track). *MYC* aberrations or upregulation of *MYC*-related pathways occur in the vast majority of cancers (35), and studies have demonstrated that *CCAT1* enhances cervical cancer cell proliferation and invasion (36). Interestingly, the stripe appears to be absent in colorectal cancer cells (Figure 5A; Bottom track), underscoring its potential regulatory significance in cervical cancer cells.

Transcription factor mediated chromatin loops

TFs play a crucial role in organizing the 3D genome by binding to specific DNA motifs and facilitating the formation of chromatin loops, which bring distant regulatory elements into proximity with their target genes. Therefore, understanding both common and specific TFs involved in mediating the 3D genome is of great interest. We identified the enrichment of TFs in chromatin loop anchors across 126 biosamples and provided various visualization methods to display the results.

Firstly, we ranked TFs based on the number of samples in which they are enriched and provided an interactive bar plot (Figure 6A). Notably, CTCF (3) and YY1 (37), known to be key TFs for 3D genome organization in tissues, ranked highly on the list. Interestingly, IKZF1's binding motif was enriched in 74 out of 126 samples (Figure 6A). IKZF1 encodes the TF IKAROS, a zinc finger DNA-binding protein with a key role in lymphoid lineage development. IKAROS is recently reported to be involved in inter-TAD interactions that harbor lymphoid-specific genes (38). Our enrichment result is consistent with IKAROS's role to assemble lineage-specific domains. Further study is needed to examine the exact locations of IKAROS-bound interaction anchors. Furthermore, we performed clustering of both TFs and biosamples, and presented the results using a heatmap. These visualizations enhance our understanding of the common and specific roles of different TFs in mediating chromatin loops. Future research could focus on TFs that are enriched in a smaller number of samples, as these may be related to cell- or tissue-specific 3D genome organization.

When clicking the TF names in the bar plot or heatmap, the Motif Table will display the exact motif logo and associated statistics in each biosample where the TF is enriched (Figure 6B).

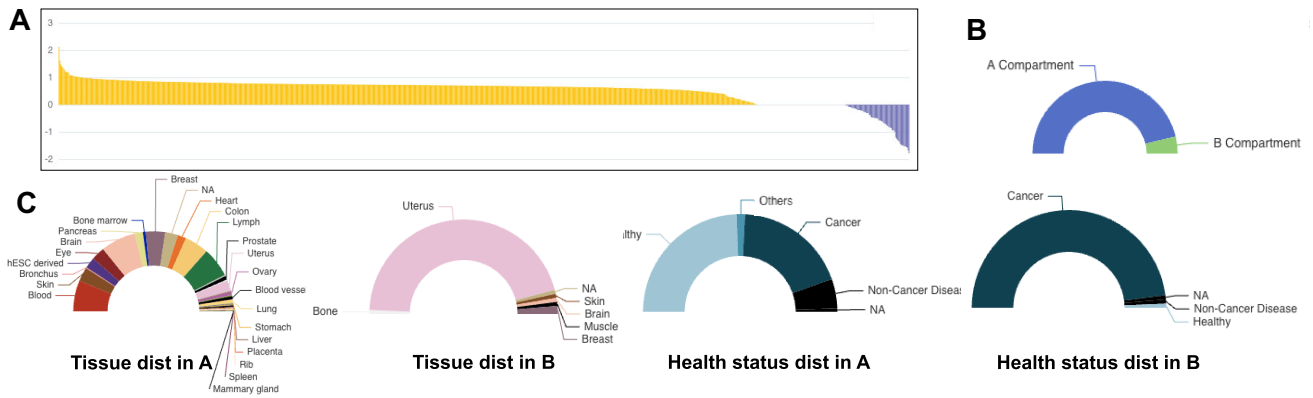


Figure 4. Gene search functionality. (**A**) Bar plot demonstrating the compartment E1 score of *NEAT1* across Hi-C samples, where E1>0 represents A compartment and E1<0 represents B compartment.. (**B**) Proportion of samples where *NEAT1* is in the A or B compartment. (**C**) Corresponding sample metadata, indicating the number of samples by tissue where *NEAT1* is in the A or B compartment, as well as the number of samples by health status where *NEAT1* is in the A or B compartment.

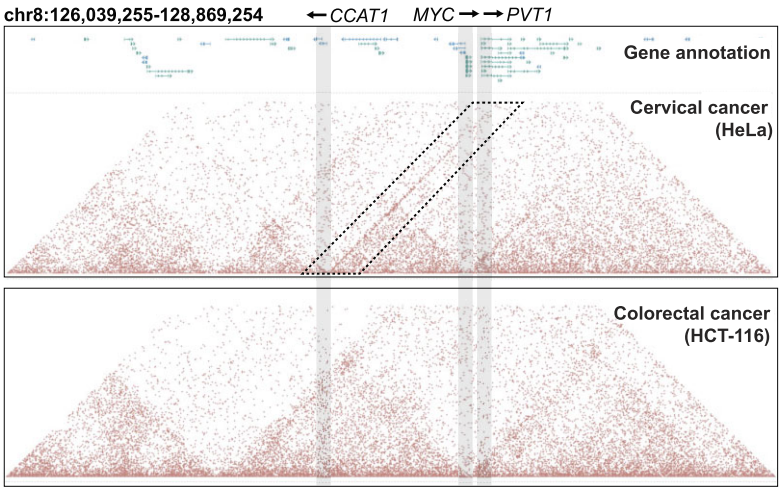


Figure 5. Visualization of CCAT1-MYC associated stripes. Screenshot of a stripe associated with *CCAT1* and *MYC* gene in cervical cancer cells.

Programmatic access

We provide a comprehensive API to programmatically access EXPRESSO (Table 2), with functionalities encompassing four main areas: (1) search metadata: filter samples by sample ID or metadata such as health status, tissue and biomaterial; (2) Search 3D genomic features: query genes contained in compartments, domains, boundaries, stripes and chromatin loops within specific samples; (3) search genes: explore the relationships of genes across samples with compartments, contact domains, stripes and chromatin loops; (4) download files: each file has a unique API URL, allowing users to download data directly by entering the URL. We provided examples of various usages on the Documentation page. Users can conveniently use the API within their programming environment (e.g. R).

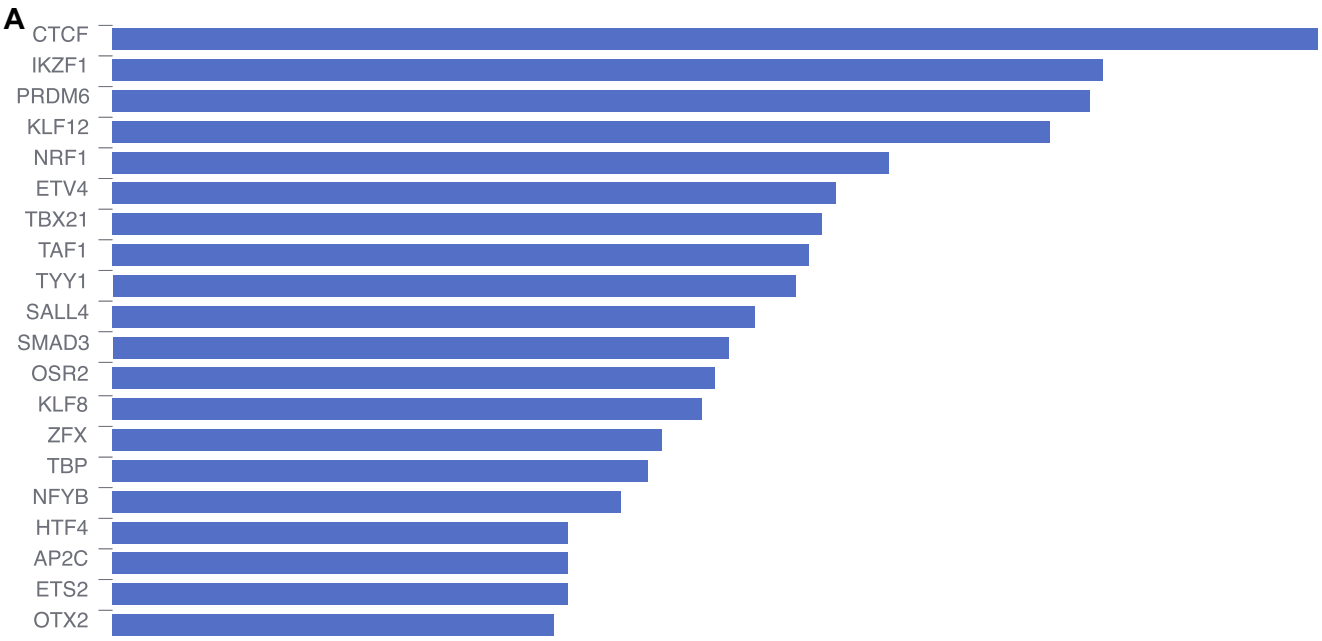
Discussion

A comprehensive database of multi-level 3D genome organization

In this study, we presented EXPRESSO, a comprehensive multi-omics database for exploring and visualizing multi-level 3D genomic features and their associations with the epigenome and gene expression across a wide range of human tissues. The database's ability to facilitate comparisons of 3D genomic features across different samples, as demonstrated in the examples of PCa cells, DCM and liver cancer, highlights its potential to uncover novel regulatory mechanisms and disease-associated genomic alterations. Furthermore, the integration of multi-omics data and the programmatic APIs enhance the database's utility for in-depth investigations. The Stripe Atlas and TF Enrichment analyses offer unique perspectives for exploring the functional implications of 3D genome organization. As research in this field continues to evolve, EXPRESSO is poised to become an invaluable resource for researchers seeking to unravel the complex interplay between 3D genome structure and gene regulation in both health and disease states.

A multi-omics database to study the relations between 3D genome and gene regulation

The 3D genome organization plays a crucial role in regulating gene expression by facilitating the interactions between genes and their regulatory elements. Understanding the intricate relationship between genome architecture and gene expression requires measuring the 3D genome, transcriptome



B

Showing 1 to 11 cell types of 74 in total.

Now filtering using transcription factor **IKZF1** Reload

Motif	Pvalue	LogP	TargetRatio	BackgroundRatio	BgSTD	Sample Name
	1.00e-95	-219.8	66.85%	41.08%	174.5bp (167.6bp)	GSM3395116
	1.00e-115	-266.8	63.69%	28.64%	280.1bp (274.9bp)	GSM3905152
	1.00e-35	-82.53	38.33%	9.30%	225.9bp (289.8bp)	GSM3905154
	1.00e-17	-39.38	10.86%	3.84%	163.4bp (302.4bp)	GSM3905156
	1.00e-16	-38.61	15.87%	0.75%	181.2bp (205.7bp)	GSM3905160
	1.00e-12	-27.72	7.21%	1.53%	218.6bp (268.2bp)	GSM4119020
	1.00e-35	-81.66	6.35%	1.15%	257.1bp (264.5bp)	GSM4453814
	1.00e-23	-54.09	76.23%	40.64%	166.3bp (202.5bp)	GSM4981509
	1.00e-29	-67.35	2.95%	0.16%	162.3bp (100.9bp)	GSM5098062
	1.00e-30	-71.06	9.95%	3.27%	135.6bp (156.0bp)	GSM5617594

‹ 1 2 3 4 5 6 7 8 ›

Figure 6. Exploration of transcription factor-mediated chromatin loops, (**A**) Bar plot of the top 20 TFs ranked by the prevalence of their enrichment across biosamples. (**B**) Detailed view of the top ten samples in *IKZF1* enrichment table, featuring the *IKZF1* motif logo and relevant statistics for each biosample. Meaning of the table columns: *P*-value, Probability of observing the motif by chance. logP, Logarithm of the *P*-value. Target Ratio, Proportion of target sequences containing the motif. Background Ratio: Proportion of background sequences containing the motif. bgSTD, Variability of motif occurrence in the background sequences.

Table 2. Application programmatic interfaces of EXPRESSO

Type	Required resource	URL
Search metadata	Sample	/dataset/{gsm_id}
	Samples	/datasets/tissue={string}&heatlth_status={string}&biomaterial={string}
	Files	/files/{sample_name}
Seach genomic features	Compartments	/compartments?sample_name={sample_name}
	Domains	/domains?sample_name={sample_name}
	Stripes	/stripes?sample_name={sample_name}
	Loops	/loops?sample_name={sample_name}
Search gene associated genomic features	Compartments	/gene/{gene_name}/compartment
	Domains	/gene/{gene_name}/domain
	Stripes	/stripes?gene_id={gene_id}
	Loops	/loops?gene_id={gene_id}
Download	File	/download/{sample_name}/{file_type}

and epigenome from the same biosample. For instance, ChIP-seq identifies binding sites of TFs and histone modifications, elucidating the regulatory landscape of the genome. ATAC-seq maps open chromatin regions, indicating genomic regions of regulatory potential. By combining 3D genomic data with these epigenomic datasets, researchers can gain a comprehensive view of how chromatin architecture influences gene regulation.

EXPRESSO, incorporating 842 1D omics datasets, provides a comprehensive view of the dynamic interplay between the 3D genome organization and the epigenetic landscape, revealing how spatial organization collectively influences gene expression. To demonstrate the usage of multi-omics integration, we created three web-based applications: compartment-expression analysis, stripe atlas, and TF enrichment. The first application compares the expression of genes located in A and B compartments, providing a general view of the impact of 3D genome organization on gene expression. The second application identifies and annotates a comprehensive resource of architectural stripes in the human genome. Further exploration of large-scale stripes collected by EXPRESSO can help better understand the grammar of stripe formation. The third application identified enriched TFs in chromatin loops across different tissues and cell types. Understanding the role of TFs in chromatin looping provides valuable insights into the regulatory mechanisms that shape 3D genome organization and influence gene expression patterns, offering potential targets for therapeutic intervention.

API access and data visualization

Other key features of EXPRESSO include a user-friendly suite of REST APIs, crucial for a database of this nature. Users can develop tools for specific programming languages based on these APIs to integrate EXPRESSO's data. As the 3D genome itself is a multi-layered structure, clearly visualizing these structures is also a critical component. The visualization module we developed is a lightweight framework that has been deeply integrated with EXPRESSO.

Future directions

In the future, we plan to continue updating EXPRESSO with specific plans that include the addition of mouse data, as mice are a crucial model organism in 3D genomics, and comparing mouse and human 3D genomes can reveal similarities and differences in transcriptional regulation between species. We also plan to incorporate Micro-C data (39), with its global and high-resolution characteristics, complements Hi-C, HiChIP and ChIA-PET data effectively. Furthermore, we aim to expand our omics data types, adding whole-genome sequencing data, exome sequencing genetic data and epigenetic data such as DNA methylation. Additionally, we will offer users the capability to annotate their own 3D genomic data, using our extensive reference dataset to aid them in understanding their data.

Data availability

EXPRESSO is a database with online and open access, available at https://expresso.sustech.edu.cn. Any constructive comments and suggestions are welcome to send to Prof. Yuliang Feng at email address fengyl@sustech.edu.cn.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank the group members in Feng Lab for providing feedback on the database. We appreciate Prof. Siim Pauklin in the University of Oxford for the suggestions on revising our text. This work is supported by Center for Computational Science and Engineering at Southern University of Science and Technology.

Funding

Natural Science Foundation of China Excellent Young Scientists Fund (Overseas) [K241141101 to Y.F.]; Guangdong Basic and Applied Basic Research Foundation for Distinguished Young Scholars [2024B1515020047 to Y.F.], Shenzhen Pengcheng Peacock Plan, Shenzhen Basic Research General Projects of Shenzhen Science and Technology Innovation Commission [JCYJ20230807093514029 to Y.F.]; National Natural Science Foundation of China [82373932 to T.X.]; Natural Science Foundation of Anhui Province [2208085MH203 to T.X.]; Department of Education of Anhui Province Outstanding Young Teacher Training Project [YQZD2023023 to T.X.]; Research Fund of Anhui Institute of Translational Medicine [2022-zhyx-C09 to T.X.]; National Natural Science Foundation of China [32370630,

32250710678 to G.L.]; National Key Research and Development Program of China [2021YFC2701201 to G.L.]; China Postdoctoral Science Foundation [2023M741534 to L.C.].

Conflict of interest statement

None declared.

References

1. Rao,S.S.P., Huntley,M.H., Durand,N.C., Stamenova,E.K., Bochkov,I.D., Robinson,J.T., Sanborn,A.L., Machol,I., Omer,A.D., Lander,E.S., *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665–1680.
2. Mumbach,M.R., Rubin,A.J., Flynn,R.A., Dai,C., Khavari,P.A., Greenleaf,W.J. and Chang,H.Y. (2016) HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods*, **13**, 919–922.
3. Tang,Z., Luo,O.J., Li,X., Zheng,M., Zhu,J.J., Szalaj,P., Trzaskoma,P., Magalska,A., Wlodarczyk,J., Ruszczycki,B., *et al.* (2015) CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell*, **163**, 1611–1627.
4. Dixon,J.R., Selvaraj,S., Yue,F., Kim,A., Li,Y., Shen,Y., Hu,M., Liu,J.S. and Ren,B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.
5. Vian,L., Pękowska,A., Rao,S.S.P., Kieffer-Kwon,K.-R., Jung,S., Baranello,L., Huang,S.-C., El Khattabi,L., Dose,M., Pruett,N., *et al.* (2018) The energetics and physiological impact of cohesin extrusion. *Cell*, **173**, 1165–1178.
6. Yoon,S., Chandra,A. and Vahedi,G. (2022) Stripenn detects architectural stripes from chromatin conformation data using computer vision. *Nat. Commun.*, **13**, 1602.
7. Xu,J., Song,F., Lyu,H., Kobayashi,M., Zhang,B., Zhao,Z., Hou,Y., Wang,X., Luan,Y., Jia,B., *et al.* (2022) Subtype-specific 3D genome alteration in acute myeloid leukaemia. *Nature*, **611**, 387–398.
8. Moore,J.E., Purcaro,M.J., Pratt,H.E., Epstein,C.B., Shoresh,N., Adrian,J., Kawli,T., Davis,C.A., Dobin,A., Kaul,R., *et al.* (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*, **583**, 699–710.
9. Dekker,J., Alber,F., Aufmkolk,S., Beliveau,B.J., Bruneau,B.G., Belmont,A.S., Bintu,L., Boettiger,A., Calandrelli,R., Disteche,C.M., *et al.* (2023) Spatial and temporal organization of the genome: current state and future aims of the 4D nucleome project. *Mol. Cell*, **83**, 2624–2640.
10. Clough,E., Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M., *et al.* (2024) NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. *Nucleic Acids Res.*, **52**, D138–D144.
11. Zhou,Q., Cheng,S., Zheng,S., Wang,Z., Guan,P., Zhu,Z., Huang,X., Zhou,C. and Li,G. (2022) ChromLoops: a comprehensive database for specific protein-mediated chromatin loops in diverse organisms. *Nucleic Acids Res.*, **51**, D57–D69.
12. Zeng,W., Liu,Q., Yin,Q., Jiang,R. and Wong,W.H. (2022) HiChIPdb: a comprehensive database of HiChIP regulatory interactions. *Nucleic Acids Res.*, **51**, D159–D166.
13. Kim,K., Jang,I., Kim,M., Choi,J., Kim,M.-S., Lee,B. and Jung,I. (2020) 3DIV update for 2021: a comprehensive resource of 3D genome and 3D cancer genome. *Nucleic Acids Res.*, **49**, D38–D46.
14. Servant,N., Varoquaux,N., Lajoie,B.R., Viara,E., Chen,C.-J., Vert,J.-P., Heard,E., Dekker,J. and Barillot,E. (2015) HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.*, **16**, 259.
15. Durand,N.C., Shamim,M.S., Machol,I., Rao,S.S.P., Huntley,M.H., Lander,E.S. and Aiden,E.L. (2016) Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.*, **3**, 95–98.
16. Abdennur,N., Abraham,S., Fudenberg,G., Flyamer,I.M., Galitsyna,A.A., Goloborodko,A., Imakaev,M., Oksuz,B.A., Venev,S.V. and Xiao,Y. (2024) Cooltools: enabling high-resolution Hi-C analysis in Python. *PLoS Comput. Biol.*, **20**, e1012067.
17. Lee,B., Wang,J., Cai,L., Kim,M., Namburi,S., Tjong,H., Feng,Y., Wang,P., Tang,Z., Abbas,A., *et al.* (2020) ChIA-PIPE: a fully automated pipeline for comprehensive ChIA-PET data analysis and visualization. *Sci. Adv.*, **6**, eaay2078.
18. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
19. Ramírez,F., Dündar,F., Diehl,S., Grüning,B.A. and Manke,T. (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.*, **42**, W187–W191.
20. Liao,Y., Smyth,G.K. and Shi,W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.
21. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
22. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
23. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W., *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
24. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
25. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
26. Vorontsov,I.E., Eliseeva,I.A., Zinkevich,A., Nikonov,M., Abramov,S., Boytsov,A., Kamenets,V., Kasianova,A., Kolmykov,S., Yevshin,I.S., *et al.* (2024) HOCOMOCO in 2024: a rebuild of the curated collection of binding models for human and mouse transcription factors. *Nucleic Acids Res.*, **52**, D154–D163.
27. Takahashi,S., Watanabe,T., Okada,M., Inoue,K., Ueda,T., Takada,I., Watabe,T., Yamamoto,Y., Fukuda,T., Nakamura,T., *et al.* (2011) Noncanonical Wnt signaling mediates androgen-dependent tumor growth in a mouse model of prostate cancer. *Proc. Natl. Acad. Sci. USA*, **108**, 4938–4943.
28. Miyamoto,D.T., Zheng,Y., Wittner,B.S., Lee,R.J., Zhu,H., Broderick,K.T., Desai,R., Fox,D.B., Brannigan,B.W., Trautwein,J., *et al.* (2015) RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science*, **349**, 1351–1356.
29. Dixon,J.R., Jung,I., Selvaraj,S., Shen,Y., Antosiewicz-Bourget,J.E., Lee,A.Y., Ye,Z., Kim,A., Rajagopal,N., Xie,W., *et al.* (2015) Chromatin architecture reorganization during stem cell differentiation. *Nature*, **518**, 331–336.
30. Dirkx,E., da Costa Martins,P.A. and De Windt,L.J. (2013) Regulation of fetal gene expression in heart failure. *Biochim. Biophys. Acta*, **1832**, 2414–2424.
31. Feng,Y., Cai,L., Hong,W., Zhang,C., Tan,N., Wang,M., Wang,C., Liu,F., Wang,X., Ma,J., *et al.* (2022) Rewiring of 3D chromatin topology orchestrates transcriptional reprogramming and the development of human dilated cardiomyopathy. *Circulation*, **145**, 1663–1683.
32. Lu,X., Deng,S., Xu,J., Green,B.L., Zhang,H., Cui,G., Zhou,Y., Zhang,Y., Xu,H., Zhang,F., *et al.* (2023) Combination of AFP vaccine and immune checkpoint inhibitors slows hepatocellular carcinoma progression in preclinical models. *J. Clin. Invest.*, **133**, e163291.

33. Zhao,Y., Ding,Y., He,L., Zhou,Q., Chen,X., Li,Y., Alfonsi,M.V., Wu,Z., Sun,H. and Wang,H. (2023) Multiscale 3D genome reorganization during skeletal muscle stem cell lineage progression and aging. *Sci. Adv.*, **9**, eabo1360.
34. Jordan Rowley,M. and Corces,V.G. (2018) Organizational principles of 3D genome architecture. *Nat. Rev. Genet.*, **19**, 789–800.
35. Dhanasekaran,R., Deutzmann,A., Mahauad-Fernandez,W.D., Hansen,A.S., Gouw,A.M. and Felsher,D.W. (2022) The MYC oncogene — the grand orchestrator of cancer growth and immune evasion. *Nat. Rev. Clin. Oncol.*, **19**, 23–36.
36. Shen,H., Wang,L., Xiong,J., Ren,C., Gao,C., Ding,W., Zhu,D., Ma,D. and Wang,H. (2019) Long non-coding RNA CCAT1 promotes cervical cancer cell proliferation and invasion by regulating the miR-181a-5p/MMP14 axis. *Cell Cycle*, **18**, 1110–1121.
37. Weintraub,A.S., Li,C.H., Zamudio,A.V., Sigova,A.A., Hannett,N.M., Day,D.S., Abraham,B.J., Cohen,M.A., Nabet,B., Buckley,D.L., *et al.* (2017) YY1 is a structural regulator of enhancer-promoter loops. *Cell*, **171**, 1573–1588.
38. Hu,Y., Salgado Figueroa,D., Zhang,Z., Veselits,M., Bhattacharyya,S., Kashiwagi,M., Clark,M.R., Morgan,B.A., Ay,F. and Georgopoulos,K. (2023) Lineage-specific 3D genome organization is assembled at multiple scales by IKAROS. *Cell*, **186**, 5269–5289.
39. Krietenstein,N., Abraham,S., Venev,S.V., Abdennur,N., Gibcus,J., Hsieh,T.-H.S., Parsi,K.M., Yang,L., Maehr,R., Mirny,L.A., *et al.* (2020) Ultrastructural details of mammalian chromosome architecture. *Mol. Cell*, **78**, 554–565.

Received: August 15, 2024. **Revised:** September 26, 2024. **Editorial Decision:** October 6, 2024. **Accepted:** October 17, 2024