ORIGINAL ARTICLE

3D disorganization and rearrangement of genome provide insights into pathogenesis of NAFLD by integrated Hi-C, Nanopore, and RNA sequencing

Lina Xu[a], Lianhong Yin[a], Yan Qi[a], Xuemei Tan[a], Meng Gao[a], Jinyong Peng[a,b,c,*]

[a]College of Pharmacy, Dalian Medical University, Dalian 116044, China
[b]Key Laboratory for Basic and Applied Research on Pharmacodynamics Substances of Traditional Chinese Medicine of Liaoning Province, Dalian Medical University, Dalian 116044, China
[c]National-Local Joint Engineering Research Center for Drug Development (R&D) of Neurodegenerative Diseases, Dalian Medical University, Dalian 116044, China

---

Abbreviations: 3C, chromosome conformation capture; 3D, three-dimensional; *Abcg5*, ATP-binding cassette sub-family G member 5; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BWA, Burrows-Wheeler Aligner; Camk1d, calcium/calmodulin-dependent protein kinase type 1D; Chr, chromosome; CNV, copy number variation; DEG, differentially expressed gene; DEL, deletion; DI, directionality index; DUP, duplication; *Elovl6*, elongation of very long chain fatty acids protein 6; FDR, false discovery rate; FFA, free fatty acid; *Fgfr2*, fibroblast growth factor receptor 2; *Gadd45g*, growth arrest and DNA damage-inducible protein GADD45 gamma; *GCKR*, glucokinase regulator; GO, gene ontology; *Grm8*, metabotropic glutamate receptor 8; GSH, glutathione; *Gsta1*, glutathione *S*-transferase A1; H&E, hematoxylin-eosin; HFD, high-fat diet; Hi-C, high-throughput chromosome conformation capture; *HSD17B13*, hydroxysteroid 17-beta dehydrogenase 13; IDE, interaction decay exponent; INS, insertion; INV, inversion; IR, inclusion ratio; *IRGM*, immunity related GTPase M; *IRS4*, insulin receptor substrate 4; *Kcnma1*, calcium-activated potassium channel subunit alpha-1; KEGG, Kyoto Encyclopedia of Genes and Genomes; *LPIN1*, lipin 1; *MBOAT7*, membrane bound *O*-acyltransferase domain containing 7; MDA, malondialdehyde; NAFLD, non-alcoholic fatty liver disease; *NF1*, neurofibromin 1 NGS, next-generation sequencing; *NOTCH1*, notch receptor 1; ONT, Oxford Nanopore Technologies; PCA, principal component analysis; *Pde4b*, phosphodiesterase 4B; *Plce1*, 1-phosphat-idylinositol 4,5-bisphosphate phosphodiesterase epsilon-1; *Plxnb1*, Plexin-B1; *PNPLA3*, patatin like phospholipase domain containing 3; *PPP1R3B*, protein phosphatase 1 regulatory subunit 3B; *PTEN*, phosphatase and tensin homolog; *RB1*, RB transcriptional corepressor 1; RNA-seq, RNA-sequencing; *Scd1*, acyl-CoA desaturase 1; SD, standard deviation; SOD, superoxide dismutase; *Sugct*, succinate-hydroxymethylglutarate CoA-transferase; SV, structural variation; TAD, topologically associated domain; TC, total cholesterol; TG, triglyceride; *TM6SF2*, transmembrane 6 superfamily member 2; *TP53*, tumor protein p53 ; TRA, translocation; WGS, whole-genome sequencing.

*Corresponding author. Tel./fax: +86 411 86110411.
E-mail address: jinyongpeng2008@126.com (Jinyong Peng).

Abstract The three-dimensional (3D) conformation of chromatin is integral to the precise regulation of gene expression. The 3D genome and genomic variations in non-alcoholic fatty liver disease (NAFLD) are largely unknown, despite their key roles in cellular function and physiological processes. High-throughput chromosome conformation capture (Hi-C), Nanopore sequencing, and RNA-sequencing (RNA-seq) assays were performed on the liver of normal and NAFLD mice. A high-resolution 3D chromatin interaction map was generated to examine different 3D genome hierarchies including A/B compartments, topologically associated domains (TADs), and chromatin loops by Hi-C, and whole genome sequencing identifying structural variations (SVs) and copy number variations (CNVs) by Nanopore sequencing. We identified variations in thousands of regions across the genome with respect to 3D chromatin organization and genomic rearrangements, between normal and NAFLD mice, and revealed gene dysregulation frequently accompanied by these variations. Candidate target genes were identified in NAFLD, impacted by genetic rearrangements and spatial organization disruption. Our data provide a high-resolution 3D genome interaction resource for NAFLD investigations, revealed the relationship among genetic rearrangements, spatial organization disruption, and gene regulation, and identified candidate genes associated with these variations implicated in the pathogenesis of NAFLD. The newly findings offer insights into novel mechanisms of NAFLD pathogenesis and can provide a new conceptual framework for NAFLD therapy.

1. Introduction

Non-alcoholic fatty liver disease (NAFLD) is a common chronic liver disorder associated with accumulation of fats in the liver. The prevalence of NAFLD is high in obese individuals, and increases susceptibility to liver fibrosis, cirrhosis, and hepatocellular carcinoma, which is considered a major health problem worldwide[1,2]. Genetic factors are believed to contribute to 30%−50% of the risk for high-prevalence diseases such as NAFLD[3]. Efforts in understanding the molecular mechanisms of NAFLD initiation and progression have elucidated the significant role of genetic factors in disease susceptibility, with more information still emerging. Gene variants like polymorphism rs626283 in membrane bound *O*-acyltransferase domain containing 7 (*MBOAT7*) gene may affect glucose metabolism by modulating intra-hepatic fat content in Caucasian obese children and adolescents, increasing the risk of NAFLD as observed by Umano et al.[4], the I148M patatin like phospholipase domain containing 3 (*PNPLA3*) variant has been identified as the major common genetic determinant of NAFLD[5]. Moreover, genome-wide association studies have uncovered robust and reproducible associations between variations in genes such as transmembrane 6 superfamily member 2 (*TM6SF2*), glucokinase regulator (*GCKR*), hydroxysteroid 17-beta dehydrogenase 13 (*HSD17B13*), protein phosphatase 1 regulatory subunit 3B (*PPP1R3B*), immunity related GTPase M (*IRGM*), lipin 1 (*LPIN1*), and the natural history of NAFLD[6,7]. In view of the complexity of NAFLD pathogenesis, the variation of individual gene cannot play a decisive role in the occurrence of disease. Analysis of full genome and disclosure of NAFLD candidate gene genes will improve understanding of NAFLD pathophysiology.

RNA-sequencing (RNA-seq) and whole-genome sequencing (WGS) are now widely applied technologies to obtain deep insights into diseases. Major advances in sequencing technologies have enabled implementation of RNA-seq and WGS in determining the molecular pathogenesis of liver diseases[8,9]. Nanopore sequencing as a third-generation sequencing approach, promises high sequencing throughput and longer read length[10,11], allowing a comprehensive analysis of the entire genome with identification of large numbers of structural variations (SVs) and copy number variations (CNVs). Nanopore sequencing is widely used in "-omic" sciences like genomics, epigenomics, and transcriptomics to evaluate the mutational status of single or multiple genes involved in disease processes[11].

SVs are commonly found in the genome, however their functional impacts are often elusive. SV events may be accompanied by alternations in the chromosomal three-dimensional (3D) structures disrupting the gene structures, leading to disease[12]. Tissue-specificity and development require maintenance of precisely regulated gene expression programs in higher eukaryotes[13]. This tissue-specific gene regulation is closely associated with the highly folded and organized 3D conformation of the genome. This 3D organization of the chromatin plays critical roles in many processes like regulation of DNA replication, DNA damage repair, X chromosome inactivation[14−16]. Alternations in the 3D genome may affect numerous key biological functions and their mechanistic connections with disease development are increasingly apparent[17]. Several studies about the 3D cancer genome have been performed in the past few years. It has been uncovered that about 12% of genomic regions had A/B compartment switch in breast cancer compared with normal cells, and gene-rich chromosomes displayed a lower interaction frequency with each other than in normal cells[18]. In lung squamous carcinoma, deletion of the topologically associated domain (TAD) boundary resulted in the spread of active chromatin and subsequent insulin receptor substrate 4 (*IRS4*) overexpression[19]. In some tumor suppressor genes, including but not limited to phosphatase and tensin homolog (*PTEN*), tumor protein p53 (*TP53*), RB transcriptional corepressor 1 (*RB1*), notch receptor 1 (*NOTCH1*), and neurofibromin 1 (*NF1*), sequences or promoters can be frequently interrupted by translocation and inversion,

leading to inactivation and the subsequent onset and progression of cancer[20,21]. Changes in 3D conformation of genomes have also been reported in other disease such as heart failure, diabetes, and brain and chronic kidney diseases[22−26].

The development and application of imaging and chromosome conformation capture (3C) during the last decade, have advanced our capacity to explore the structures in 3D genome architecture[27]. High-throughput chromosome conformation capture (Hi-C) sequencing technology is a 3C-based method, which obtains 3D genome information of a disease by detecting changes in chromatin interactions and hierarchical structures[28], like alterations of A/B compartment, TAD boundary, and chromatin loop. By understanding of the role of 3D genome organization in the precise regulation of gene expression, the relationship between NAFLD pathogenesis and 3D genome needs to be elucidated.

In the present study, we performed an integrated analysis combing Hi-C, Nanopore sequencing, and RNA-seq data on mice with NAFLD, to better understand the molecular mechanisms of NAFLD. We investigated the spatial organization and alterations in the 3D genome interactions to reveal the relationship between 3D genome reorganization, SVs, and gene regulation in NAFLD. The integration of these technologies has the potential to bridge the gap between genome structures and gene function, to obtain deeper insights into NAFLD.

2. Materials and methods

2.1. Animals

Male C56BL/6 mice (18−22 g) obtained from Liaoning Changsheng Biotechnology Co., Ltd. [SCXK (Liao) 2015-0001, Benxi, China] were housed in the controlled environment at 23 ± 2 °C under a 12 h dark−light cycle with free access to food and water. After being acclimatized to laboratory conditions for 1 week, animal maintenance and experiments were performed. Mice were randomly divided into control and model groups ($n = 10$). Mice in control group were fed a normal chow diet (12% kcal fat content), and mice in model group were fed a high-fat diet (HFD, 45% kcal fat content) for 8 weeks. Then, the serum samples and liver tissues were obtained for the following pathological staining and assay. Animal experiments were performed referring to the ARRIVE guidelines and approved by the Animal Research Committee of Dalian Medical University in accordance with the Laboratory's Animal Ethics Guidelines [SYXK (Liao) 2014-0002, Dalian, China].

2.2. Biochemical analysis and pathological staining

The serum levels of alanine aminotransferase (ALT), aspartate aminotransferase (AST), total cholesterol (TC), triglyceride (TG), and the levels of free fatty acid (FFA), malondialdehyde (MDA), superoxide dismutase (SOD), and glutathione (GSH) in liver tissue ($n = 10$) were examined by the detection kits according to their protocols (Nanjing Jiancheng Institute of Biotechnology, Nanjing, China). At the end of the experiment, liver tissues were removed and fixed in 4% polyoxymethylene solution for more than 24 h. Then, the livers were embedded in paraffin wax and section for hematoxylin-eosin (H&E) staining to evaluate the pathological variation. The frozen liver sections were stained with oil red O to observe lipid droplets in the liver.

Three liver samples in each group were randomly chosen and simultaneously detected by RNA-seq, Hi-C, Nanopore sequencing, and qPCR methods.

2.3. Hi-C data processing

Three liver samples of mice without NAFLD (control group, C) and with NAFLD (model group, M) were mixed equally, respectively. The Hi-C library was prepared following the standard *in situ* Hi-C protocol[29], using Dpn II as the restriction enzyme. The ends of restriction fragments were labeled using biotinylated nucleotides, and then ligated in a small volume. After reversal of crosslinks, ligated DNA was purified and sheared to a target size of 300−700 bp fragments, and then ligation junctions were pulled down with streptavidin beads and prepped for high-throughput Illumina sequencing. Two Hi-C libraries were established for each group. After construction of the library, the concentration of the library and the size of the inserted fragment (Insert Size) were detected by Qubit 2.0 (Agilent Technologies, Santa Clara, USA) and Agilent 2100 (Agilent Technologies), respectively. The effective concentration of the library was quantified to ensure the quality of the library. After passing the library inspection, high-throughput sequencing was carried out on the Illumina platform (Illumina, Santiago, USA), and the sequencing read length was PE150. Clean reads were obtained by removing reads containing adapter, poly-N, and low quality reads from raw data. Clean paired-end reads were mapped to the sequences of the reference genome using Burrows-Wheeler Aligner (BWA, 0.7.10-r789 version)[30]. The mapped reads were analyzed using HiC-Pro (v2.10.0)[31], and the valid interaction pairs and invalid interaction pairs were identified. The corresponding standardized interaction matrix was obtained by HiC-Pro v2.10.0, and the Pearson correlation coefficient was calculated. Then, the resolution of Hi-C data was calculated based on the reported method[29], and the Hi-C resolution is 10 kb in this study. Then, the Hi-C data are standardized by HiC-Prov 2.10.0 software with 100 kb, and the relationship between interaction frequency and genome linear distance is calculated by HiCdat (HiCdatR_0.99.0 version)[32]. The slop of the curve is the interaction decay exponents (IDEs). Instructure simulation of the whole genome, the mds algorithm of PASTIS software was used to simulate the 3D location of chromatin[33].

2.4. A/B compartment analysis

We used normalized interaction matrices at 100 kb resolution to detect chromatin compartment types by HiTC (v1.24.0) software[34]. Positive or negative values of the first principal component separate chromatin regions into two spatially segregated compartments. The compartment with higher gene density and positive principal component analysis (PCA) value was assigned as A compartment, and the other compartment and negative PCA value was assigned as B compartment.

2.5. TAD analysis

We used TadLib (hitad 0.1.1-r1) software to identify TADs at 20-kb resolution[35]. Inclusion ratio (IR) value was calculated using HOMER (v4.10.1) software[36]. The TADs were obtained by filtering TAD with IR less than 0 and length less than 5 bin.

2.6. Chromatin loops analysis

Chromatin loops were identified with improved HICCUP method[32]. For the Hi-C matrix, HICCUP calculates several enrichment values over different local neighborhoods, and each enrichment value is associated with a false discovery rate (FDR) value for assessing statistical significance. We identified loops at a matrix resolution of 10 kb, only retaining loops that have an FDR no more than 0.1 every neighborhood. Then, we combined the loop results of each sample, and the original interaction frequency was used as the input of edgeR (v3.8.6)[37]. The P value was calculated by glmLRT function and adjusted by Benjamini-Hochberg. When the obtained loops with FDR less than 0.01, we recognized it as a differential loops.

2.7. Nanopore library preparation and sequencing

DNA libraries were prepared using the protocol provided by Oxford Nanopore Technologies (ONT, Oxford, UK). Briefly, 1 µg isolated DNA was sheared to ~8 kb fragments with g-tube using $4000 \times g$ and 1 min centrifuging parameters. Then, the sheared products were purified using 0.45-fold XP beads according to Agencourt protocol (Beckman Coulter, Brea, USA). DNA fragments were then subjected to FFPE DNA repair and end-repair (NEB, Ipswich, USA) steps. Then DNA was ligated to the adaptor using T4 DNA ligase (NEB). Following adaptor ligation, the products were purified by adding a 0.6-fold Agencourt XP beads according to the ONT purification protocol. Then, the final DNA libraries were added to FLO-MIN109 flowcells (Covaris, Woburn, USA). For each group, three biological reduplicated samples were loaded on PromethION-48 platform (ONT, Oxford, UK) and the data were collected. Basecalling was performed live on the compute module using Guppy (ONT)[38]. The original data was further filtered to remove adapters, short reads and low quality reads, and then the total clean reads were obtained. Sequence depth and alignment efficiency were counted by aligning the positions of clean reads at the reference genome after mapping with Minimap2 (v2.14) software[39,40].

2.8. Detection of SVs

SVs were identified using Sniffles (v1.0.10)[41], an ONT long-read SV mutation detection software. Four types of structural variants including deletion (DEL), insertion (INS), inversion (INV), and duplication (DUP) were annotated. The SV results of biological reduplicated samples were combined using SURVIVOR (v1.0.5) software[42], and the SV was re-identified for each sample based on the combined SV results. Difference analysis was performed on the re-identified SVs. If the variant genotypes identified between control group and model group were different, it was considered to be differential SVs.

2.9. Detection of CNVs

CNV detection was performed using QDNA-seq (QDNA-seqR_1.18.0) software[43]. The genome was divided into 15 kb-sized bins, and the alpha value of the circular binary segmentation of the DNA copy package less than 0.05 was accepted as the changed point. Gene annotation was performed in the region of identified CNV for each sample. For each bin, identify the copy number status under different experimental groups. If the status of the model group and the control group were inconsistent, the

difference was identified as the differential CNV, and the adjacent differential bins were combined.

2.10. RNA-seq transcriptome analysis

Total RNA was isolated and the RNA quality was assessed using the Agilent Bioanalyzer 2100 System (Agilent Technologies). The ribosomal RNA was removed using the Ribo-Zero rRNA Removal Kit (Epicentre, Madison, WI, USA), and the RNA sequencing libraries were generated using NEB Next RUltra™ Directional RNA Library Prep Kit (NEB) for Illumina following manufacturer's recommendations, and index codes were added to attribute sequences to each sample. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kitv3−cBot−HS (Illumina) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina platform and reads were generated. Clean data (clean reads) were obtained by removing reads containing adapter, reads containing poly-N, and low quality reads from raw data. These clean reads were then mapped to the mouse reference GRCm38 genome (ftp://ftp. ensembl.org/pub/release-95/fasta/mus_musculus/), and all the downstream analyses were based on the clean data with high quality.

2.11. qPCR

qPCR was performed to validate gene expressions in liver tissue. RNA was extracted from the liver by using RNAiso Plus (Trans-Gen Biotech, Beijing, China). Total RNA was reverse transcribed into cDNA with oligo (dR) primers by using TransScript All-in-One First-Strand cDNA Synthesis SuperMix (TransGen Biotech). qPCR were performed using the TransStart Top Green qPCRSuperMix (TransGen Biotech) and subjected to CFX96 PCR system (Bio-Rad Laboratories, Hercules, CA, USA). The levels of gene were normalized to those of β-actin and gene expressions were analyzed by $2^{-\Delta\Delta Ct}$ method. Each experiment was performed once with three technical triplicates per sample.

2.12. Statistical analyses

Data are presented as mean $\pm$ standard deviation (SD). Statistical analysis was performed using GraphPad Prism 8.0 Software (San Diego, CA, USA). Comparisons of statistical significant ($*P < 0.05$, $**P < 0.01$) between two groups were performed using a two-tailed unpaired Student's t-test.

3. Results

3.1. Reprogrammed liver transcriptome in mice with NAFLD

The mice were induced to develop NAFLD, and the results indicated significant increase in serum levels of ALT, AST, TC, TG, and FFA; with changes in levels of MDA, SOD, and GSH indicating a state of oxidative stress in the liver of model group mice (Fig. 1A). The histopathological changes and lipid accumulation in the hepatic tissues of the model group were observed by H&E and oil red O staining (Fig. 1B).

On comparing gene expression levels by RNA-seq between mice without NAFLD (control group, C) and with NAFLD (model group, M), each sample retained more than 116 million
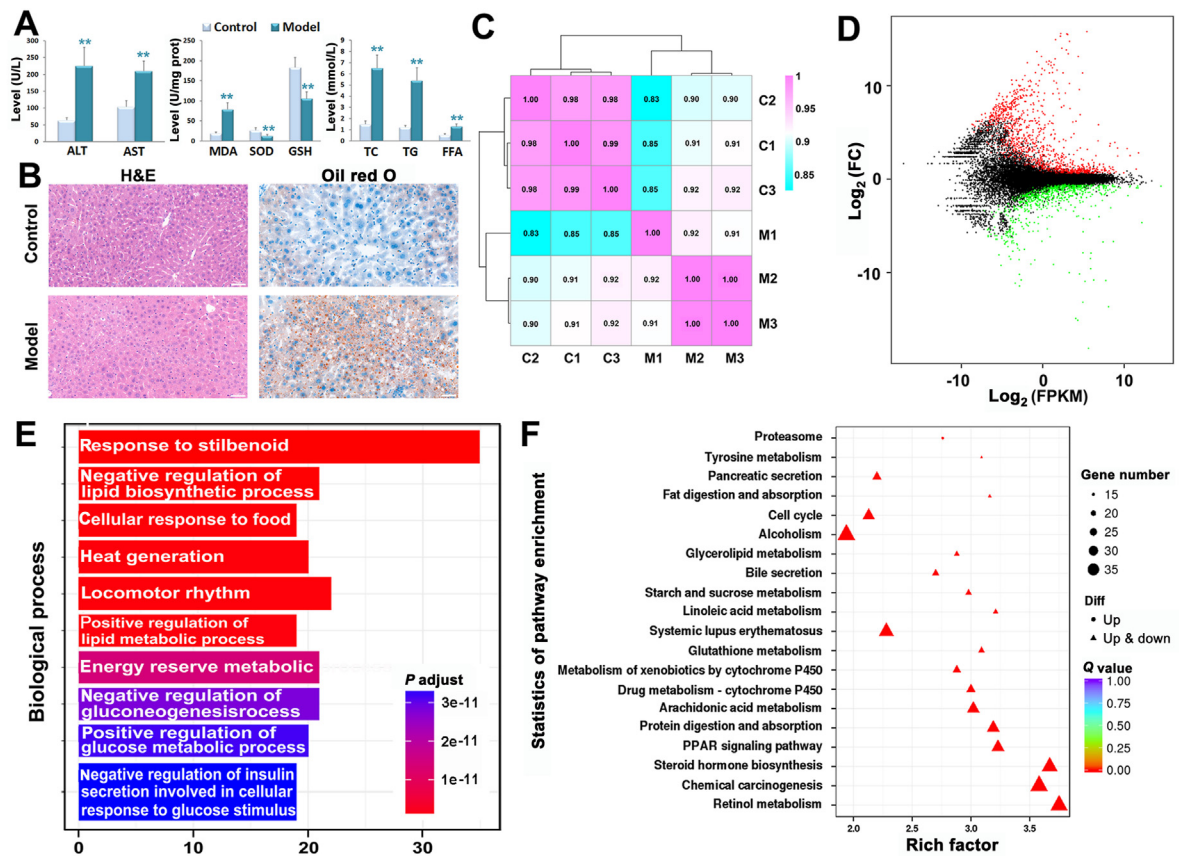
Figure 1 The reprogrammed liver transcriptome in C57BL/6 mice with NAFLD. (A) The levels of serum marker of mice in normal and NAFLD group, including ALT, AST, MDA, SOD, GSH, TC, TG, and FFA (mean ± SD, $n = 10$/group, **$P < 0.01$ vs. control). (B) The morphological changes in the livers of NAFLD mice assessed by H&E and oil red O staining. Scale bar: 50 μm. (C) Heatmap showing the Pearson correlation of gene expressions in RNA-seq across liver samples from control (C1, C2, and C3) and NAFLD model (M1, M2, and M3) groups. (D) The MA plot of differentially expressed genes (DEGs) identified by RNA-seq between control and NAFLD groups, with absolute value of $\log_2$ (fold change) > 1 and adjusted $P < 0.01$ vs. control (mean ± SD, $n = 3$). Green dots represent significantly downregulated genes in model group, red dots represent significantly upregulated genes in model group, and black dots represent genes with on significant difference in expression between two groups. (E) The top 10 most significant biological process terms in GO enrichment within DEGs. Abscissa represents the number of DEGs enriched in the corresponding biological process, while color of the column represents the adjusted P value by hypergeometric test. (F) The top 20 enriched KEGG pathways within DEGs. Abscissa represents the rich factor, which is the ratio of the proportion of DEGs annotated to the pathway to all genes annotated to the pathway. The ordinate is $-\log_{10}$ (Q value) that represents the enrichment significance of DEGs in this pathway, where Q is the adjusted P value after multiple hypothesis test.

(from 116 to 205 million) high quality clean reads after quality filtering (Supporting Information Table S1). A total of 91.44% reads (from 89.03% to 93.92%) were successfully mapped to the mouse reference genome. Coverage depth maps of the mapped reads were drawn on the reference genome by counting the position distribution of reads on each chromosome (Supporting Information Fig. S1A). The high degrees of similarity between different biological replicates of the liver samples are shown in Fig. 1C and Fig. S1B. Considering all replicates, we identified 2163 differentially expressed genes by FPKM analysis [absolute value of $\log_2$ (fold change) > 1 and adjusted P value < 0.01] in NAFLD mice (Fig. 1D). Of the differentially expressed genes, 2100 genes were used in the study and can be annotated as differentially expressed genes (DEGs), including 1251 upregulated and 849 downregulated genes in NAFLD mice (Supporting Information Table S2). Enrichment analysis on the obtained differentially expressed genes was performed using GO and

KEGG database and the top 10 most significantly enriched GO terms and the top 20 KEGG pathways are shown in Fig. 1E and F and Fig. S1C and D, respectively. Among the GO terms, "negative regulation of lipid biosynthetic process", "cellular response to food", "positive regulation of lipid metabolic process", "energy reserve metabolic process", "negative regulation of gluconeogenesis", and "positive regulation of glucose metabolic process", associated with glycolipid metabolism were enriched in the biological process. Furthermore, KEGG pathways including "fat digestion and absorption", "glycerolipid metabolism", and "PPAR signaling pathway" were enriched in the model group compared with the control group (Supporting Information Table S3). These results indicate a large proportion of perturbations in the transcriptome of a NAFLD liver, with the majority of dysregulated protein-coding genes associated with glycolipid metabolism. We selected a subset of the differentially expressed genes for validation by qPCR, and the

expression levels of the selected genes were consistent with RNA-seq results (Fig. S1E).

3.2. Mapping 3D chromatin conformation

To get maps of 3D chromatin conformation, Hi-C libraries of liver samples were sequenced and more than 477 Gb clean data were acquired for each group in total. More than 802 million pairs of valid Hi-C data were obtained for each sample after removing PCR duplicates, and the mappability of the reads was 68.56% and 68.96% at a depth of sequencing that enabled examination of interactions with a 10 kb resolution for C and M groups, respectively (Supporting Information Table S4). The standardized interaction matrix was obtained using HiC-Pro v2.10.0, and the Pearson's correlation coefficient between the sub-libraries of each group was 0.95 (Fig. S2A). Cis-interaction refers to a read with both ends aligned to the same chromosome, whereas trans-interaction indicates the ends of the read aligned to different chromosomes. There was no significant difference in the number of cis- and trans-interaction in the Hi-C libraries between control and model samples (Fig. 2A and B, and Supporting Information Table S5). The ratio of cis- to trans-interaction revealed about three-quarters of interactions

occurring within chromosomes and a quarter of the interactions between chromosomes (Fig. 2A). The interaction frequency between chromosomes can be calculated by comparing the observed read number of interactions between any two chromosomes by Hi-C and the number of expected interactions, this frequency can reveal the interaction strength between chromosomes in the genome and the relative position of the chromosomes in the nucleus. The interaction frequency heatmaps between chromosomes are shown in Fig. 2C. Differences in the interaction frequency were identified between chromosome (Chr) 2 and Chr 5, by color change in the heatmap from blue in the control group to light red in the NAFLD group, suggesting relatively closer distance and stronger interaction between the two chromosomes in NAFLD group. In contrast, the observed color changed from red to blue revealing a farther spatial distance between chromosomes and weaker interaction in the NAFLD group, such as the interaction between Chr 7 and Chr 11. The two Hi-C libraries from each group were combined in the subsequent comparative analyses between control and model group. Based on the normalized interactive signal strength of Hi-C read pairs covered between any two bin, the whole-genome interaction matrixes were drawn (Fig. 2D), and the intra-chromosomal interaction matrixes are shown in Fig. 2E and Fig. S2B.
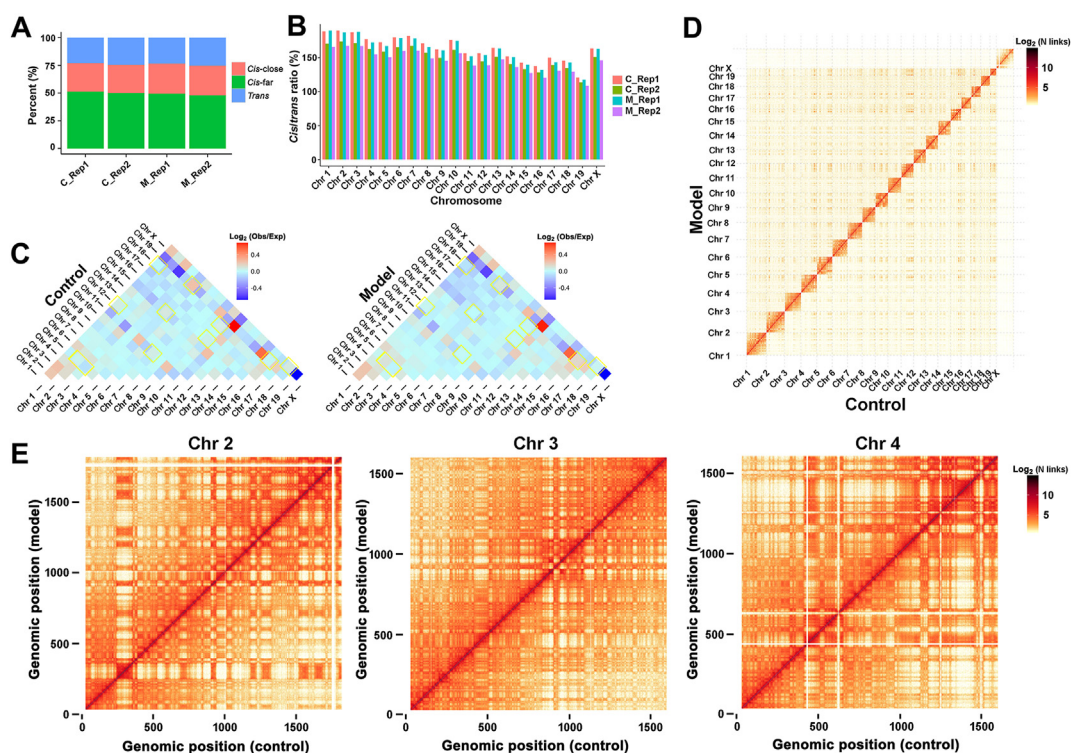

Figure 2 Mapping 3D chromatin conformation of the liver tissues of mice in control and NAFLD group by Hi-C. (A) The distributing proportion of cis- and trans-interaction in the Hi-C libraries obtained from control and NAFLD mice. Liver tissues from 3 mice in each group were mixed equally, and two Hi-C libraries were established for each group. Cis-close, the cis-interactions with the linear distance between two terminals was less than 20 kb; cis-far, the cis-interaction with the linear distance between two terminals was more than 20 kb; Trans, the ends of the read aligned to different chromosomes. (B) The cis/trans-interaction ratio in each chromosome in the established Hi-C libraries. (C) The interaction frequency heatmaps between chromosomes of the livers from mice in control and model groups. Red color in heatmap suggests that the distance between the two chromosomes is relatively close and there is a stronger interaction; while blue color indicates a closer distance between the two chromosomes and a weaker interaction. Some obvious differences between control and model groups are outlined in yellow. (D) Hi-C heatmap of genome-wide interaction pattern of liver samples from control and NAFLD mice. (E) The representative Hi-C interaction matrices of Chr 2, Chr 3, and Chr 4.

3.3. Identification of alternations in A/B compartment and related DEGs in NAFLD

Besides gene interactions, the high-resolution resource generated by Hi-C in this study can explore physical environment changes of any genomic *locus* in the liver cells with NAFLD at multiple scales, like A/B compartments, TAD, and chromatin looping. To improve the resolution of Hi-C, two libraries of each sample were combined in subsequent analysis. The chromatin was segregated into A/B compartments by PCA to understand the differences in chromatin interaction patterns in the NAFLD liver genome. As euchromatin and heterochromatin regions are spatially separated within a chromosome territory, the chromatin can be divided into A and B compartments, corresponding to higher and lower gene expressions, respectively. The distribution of A/B compartments on each chromosome of each sample are shown in Fig. 3A and Supporting Information Fig. S3. In total, 714 A (42.93%) and 722 B (49.65%) compartments were identified in liver genome of control sample, whereas 819 A (42.64%) and 821 B (49.92%) compartments were identified in NAFLD sample (Supporting Information Table S6). Very few differences in the ratio of A/B compartments in each chromosome in NAFLD were observed (Fig. 3B). On combining the genomic and transcriptome data, the gene number and

expression distributed in the A compartment were significantly higher than those in the B compartment (Fig. 3C and D), with the results consistent with the feature of A/B compartments. The compartments were highly correlated between the normal and NAFLD conditions, with switch of only 2.34% of the genome following NAFLD inducement (Fig. 3E). The distributions of A/B switching compartments in each chromosome are shown in Fig. 3F. A total of 447 A/B compartment switching events occurred in all chromosomes, with 226 A to B switches and 221 B to A switches, with the highest number of compartment switches occurring on Chr 9 and Chr 13 (Supporting Information Fig. S4A−C and Table S7). Comparison of gene expression (RNA-seq) with the A/B switching compartments defined by Hi-C identified correlation between A/B compartment changes and the gene expressions (Fig. S4D). Screening 38 DEGs associated with the A/B compartment switching between the experimental groups identified upregulation of 94.12% DEGs in B-to-A compartment and downregulation of 33.33% DEGs in A-to-B compartment in the NAFLD group. We screened 16 upregulated DEGs related to B-to-A compartment and 4 downregulated DEGs related to A-to-B compartment switch (Fig. S4E and Table S8). SUGCT is a mitochondrial enzyme synthesizing glutaryl-CoA from glutarate in tryptophan and lysine catabolism. SUGCT dysfunction is accompanied by lipid accumulation
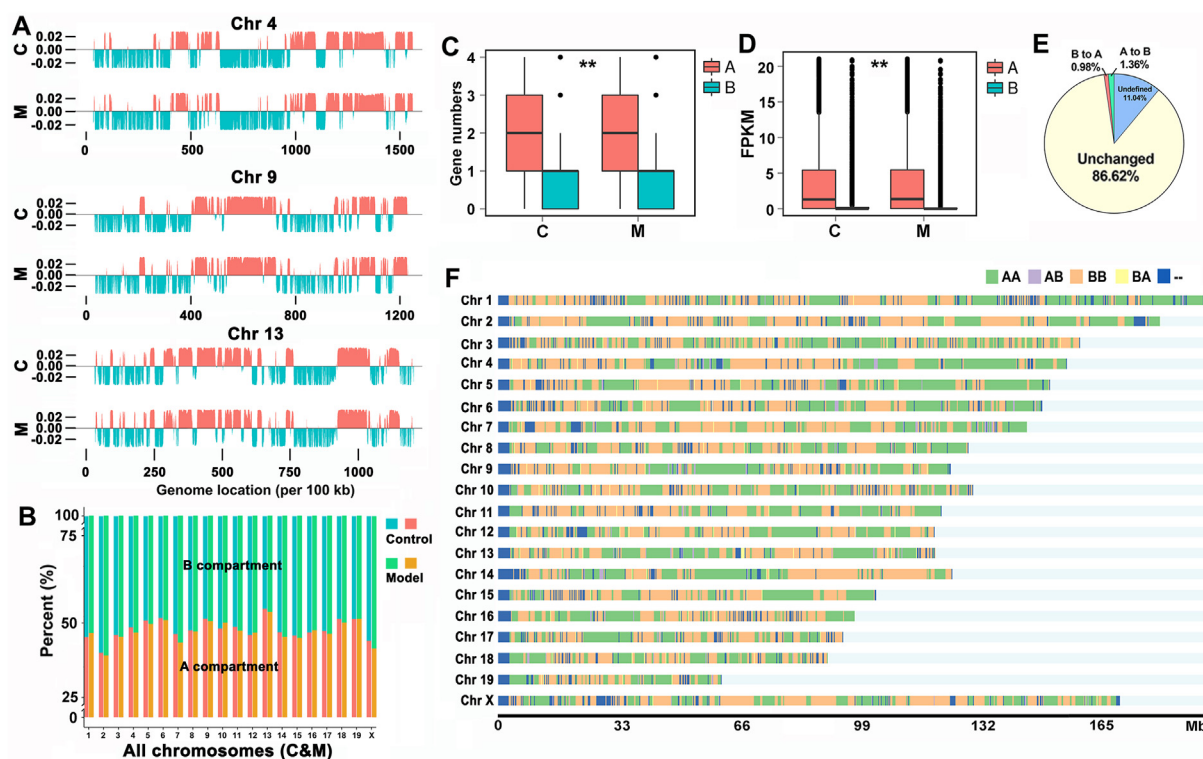

Figure 3 The changes in A/B compartment and the related DEGs in the livers of mice with NAFLD by Hi-C and RNA-seq. (A) The distributions of A/B compartment on Chr 4, Chr 9, and Chr 13. Red represents A compartment and blue represents B compartment. (B) The distribution ratio of A/B compartment on each chromosome in control and model group. Red represents A compartment and blue represents B compartment of control group, whereas green represents A compartment and orange represents B compartment of NAFLD model group. (C) The gene number distributed on A and B compartments in the livers of mice in control (C) and NAFLD (M) groups. **$P < 0.01$ *vs.* A compartment (mean ± SD, $n = 3$). (D) The gene expressions distributed on A and B compartments in RNA-seq. **$P < 0.01$ *vs.* A compartment (mean ± SD, $n = 3$). (E) The changes of A/B compartment in the livers of mice with NAFLD. Undefined means chromosome segment that impossible to be identified as A or B compartment due to the insufficient coverage or the inconsistency in biological duplication of the Hi-C data. (F) The distributions of A/B switching compartment in each chromosome. "−" represents undefined compartment.

in the liver, and is an important contributing factor in metabolic disorders[44]. Fibroblast growth factor receptor 2 (*FGFR2*), a tumorigenesis-related gene is upregulated in advanced NAFLD[45], and was also upregulated in our study, and may be associated with potential tumorigenic acquirement in NAFLD.

3.4. Identification of alternations in TADs and related DEGs in NAFLD

We identified the TADs using the TadLib software at 20 kb resolution and TADs with IR > 0 and length >5 bin were retained as the final TADs. On comparison of TADs in the control and model groups, we identified 8179 TADs in the control group and 7818 TADs in the NAFLD group, and their median TAD sizes were 180 kb (Supporting Information Table S9). Analysis of TAD architecture demonstrated TADs to be stable features of chromatin structure. Directionality index (DI) measures imbalance in the directionality of contacts from a given region of interest[46], and is related to TAD boundaries which show a signature of strong shift from upstream to downstream DI (Fig. 4A). We identified 9390 and 9246 TAD boundaries in control and NAFLD, respectively, of which 538 are special boundaries of control and 394 are of model groups (Fig. 4B–D and Supporting Information Table S10). TADs were highly correlated between control and NAFLD groups, with only approximately 5% TAD boundaries changed upon NAFLD development (Fig. 4C and D). TADs play an important role that in the regulation of gene expression, hence, we mapped the DEGs

associated with the differential TAD boundaries and analyzed their numbers. DEGs associated with the unique TAD boundary of each group and the unchanged boundary in both groups are shown in Fig. 4E. Gene expression in NAFLD was affected by changes in TAD boundary such as the chromosome region with red shadow in Fig. 4F, as well as alternation of the domains shown as the blue shadow in Fig. 4F. We identified 42 upregulated and 30 down-regulated genes (Supporting Information Fig. S5 and Table S11). These results indicate altered transcriptional profiles due to formation of new and disappearance of the original TAD boundaries associated with NAFLD development.

3.5. Identification of changes in chromatin looping and related DEGs in NAFLD

In addition to short-range chromatin interactions, long-range looping in 3D genome organization brings physical proximity between two linearly separated *loci*. We annotated chromatin loops through the Hi-C contact matrices using an improved HICCUP method with 10 kb resolution. The results demonstrated a decrease in the total number of loops (from 39,670 to 38,469, Fig. 5A), without a change in loop size (loop median length at 70 kb, Supporting Information Fig. S6A, in association with NAFLD development. The distributions of loop in each chromosome are shown in Fig. 5B and C, with the highest number of loop distributions on Chr 1, with 3193 and 3099 loops in control and model group, respectively (Fig. 5B). Although there is no
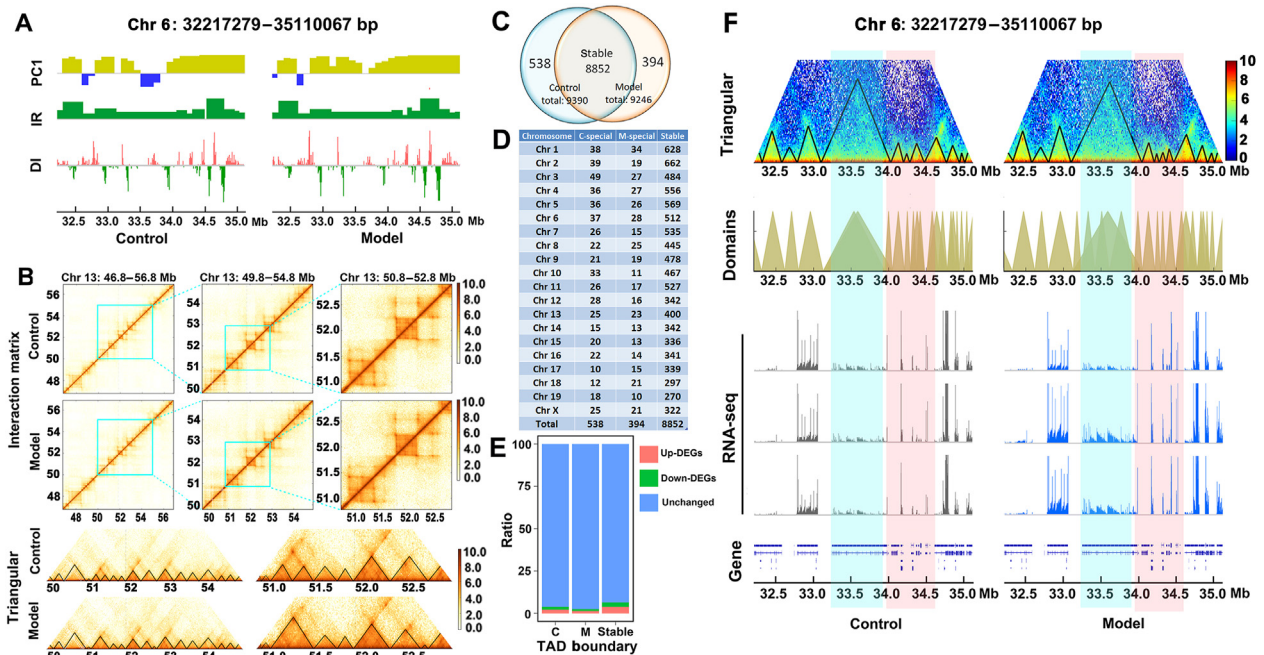

Figure 4 Changes in TADs and the TAD-impacted DEGs in the livers of mice with NAFLD identified by Hi-C and RNA-seq. (A) The PC1, IR, and DI scores of Chr 6: 32217279−35110067 bp in control and NAFLD groups generated by Hi-C. (B) Hi-C contact matrix and TADs (black triangle) in Chr 3 in control and NAFLD groups. (C) Number of TAD boundaries identified by Hi-C in control and NAFLD groups. (D) The distribution number of TAD boundaries in each chromosome in control (C) and NAFLD (M) groups. (E) The DEGs and unchanged-genes ratios associated with the unique TAD boundary in each group and the unchanged boundary in both samples. (F) Example depicting relationship between gene expression and TAD on Chr 6: 32217279−35110067 bp. Black triangles in triangular contact matrix indicate the TADs. Red shadow displays changes in TAD boundary affected gene expression in NAFLD, and blue shadow shows that gene expression was also related to the alternation of domains in addition to TAD boundary.
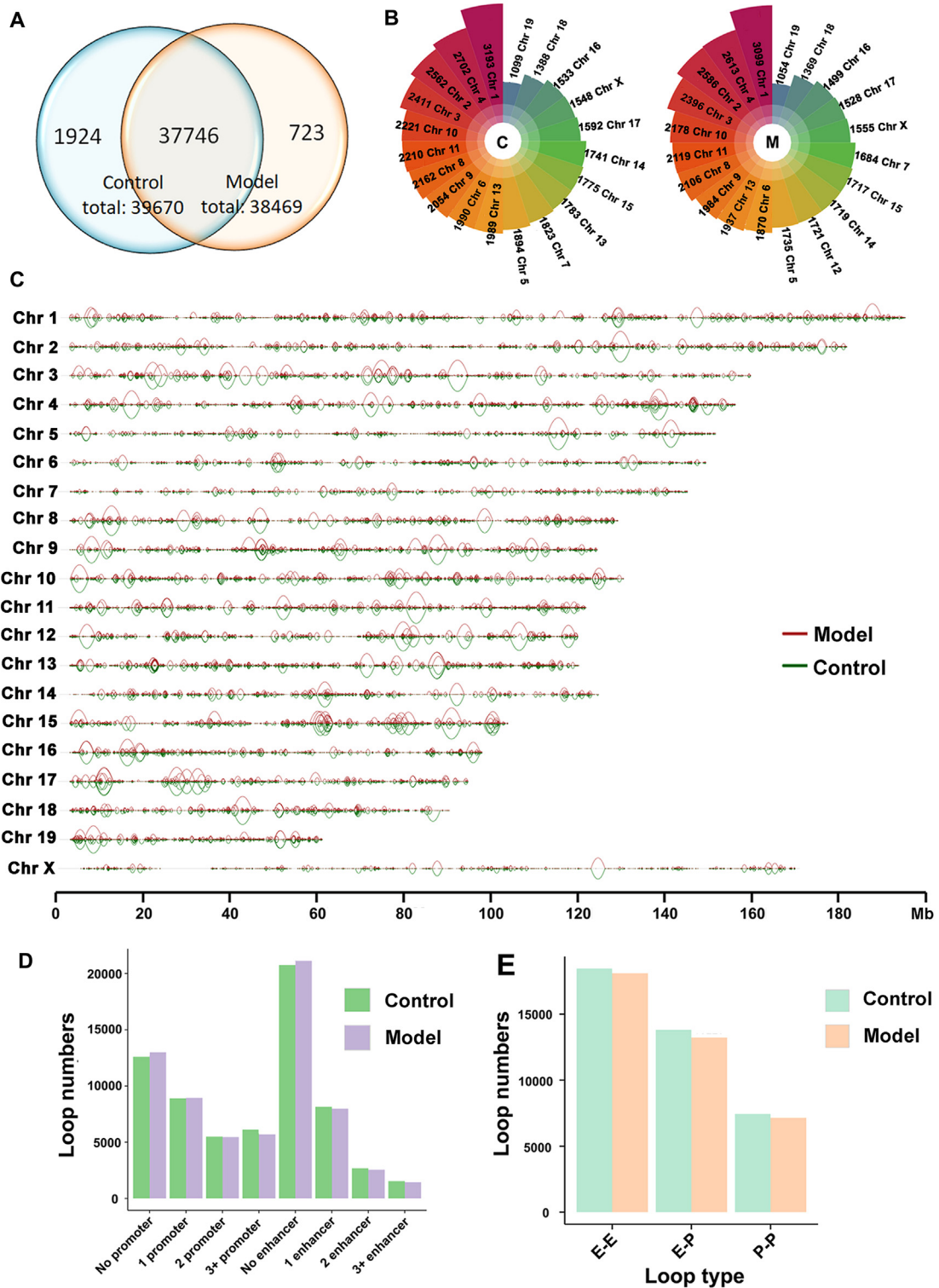
Figure 5 Changes in chromatin looping and the related DEGs in NAFLD identified by Hi-C and RNA-seq. (A) Number of chromatin loop identified by Hi-C in control and NAFLD groups. (B) The distribution number of TAD boundaries in each chromosome in control (C) and NAFLD (M) groups. (C) Distribution of chromatin loop in each chromosome in control (green) and NAFLD (red) groups. (D) Number of different types of chromatin looping. (E) Number of chromatin looping with anchor located on enhancer (E) and promoter (P).

Figure 6 Changes in structure variants (SV) in NAFLD identified by Nanopore sequencing. (A) The representative distribution (C1 group) of sequencing depth. Distributions of the other samples are shown in Supporting Inform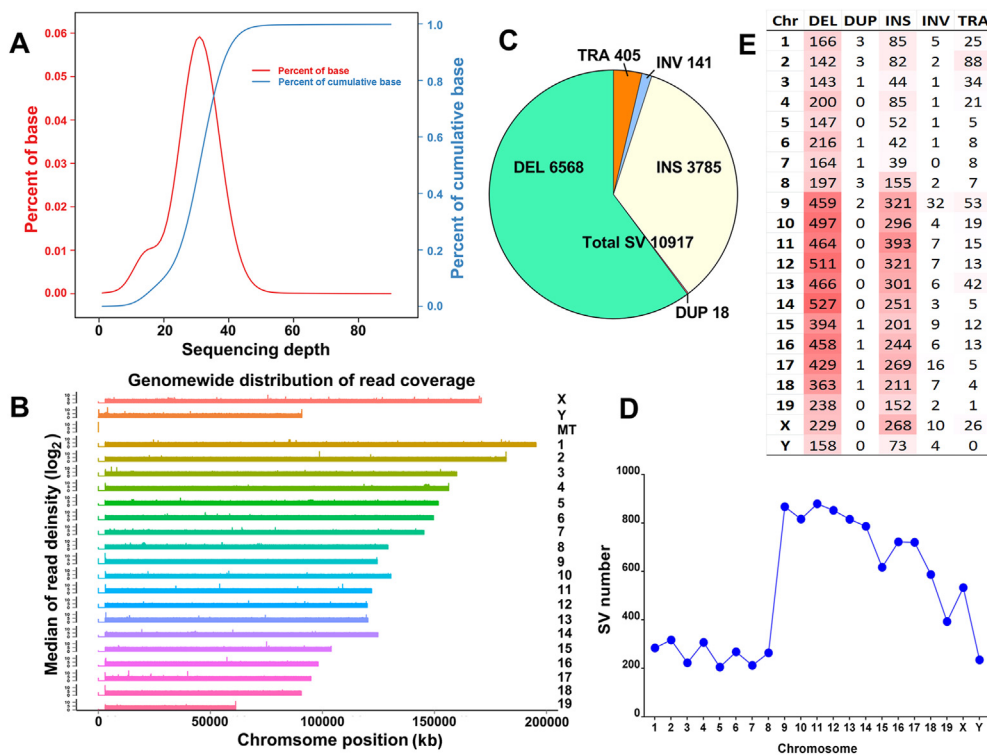ation Fig. S7B. (B) The representative distribution (C1 group) of read coverage. Distributions of the other samples are shown in Supporting Information Fig. S8. (C) Number of different type of SV identified by Nanopore sequencing, including deletion (DEL), insertion (INS), translocation (TRA), inversion (INV), and duplication (DUP). (D) Changes in the SV number in each chromosome. (E) Distributions of different type of SV in each chromosome.

significant change in loop number, the disappearance of original loop and formation of new loop is observed in NAFLD group (Fig. 5C). On classifying the identified chromatin loops in each group, statistical analysis indicated increase in number of loops without promoter and enhancer, and decrease in number of loops with anchor located on enhance and promoter in the NAFLD group compared to the control group (Fig. 5D and E). We confirmed 1924 specific loops in the control group and 723 specific loops in the model group with an FDR threshold of 0.1 (Fig. 5A and Supporting Information Table S12). The relationship between loop variants and gene expression was investigated combing the RNA-seq data, and 98 mapped DEGs were regulated by variants at chromatin loops, including 48 upregulated and 50 downregulated genes (Supporting Information Fig. S6B and Table S13). GO enrichment and KEGG analysis of these loop-impacted DEGs (Fig. S6C and D) identified hallmark genes like acyl-CoA desaturase 1 (*Scd1*) and elongation of very long chain fatty acids protein 6 (*Elovl6*) involved in *de novo* lipogenesis, and their expression was downregulated in NAFLD. On the other hand, levels of ATP-binding cassette transporters like ATP-binding cassette sub-family G member 5 and 8 (*Abcg5* and *Abcg8*) that form a heterodimer and facilitate the efflux of cholesterol into bile[47], and *Fgfr2* as reported in advanced NAFLD[45] were upregulated in NAFLD in our study. These finding indicate the potential role of chromatin loops in cell metabolic events in NAFLD.

3.6. SVs in NAFLD identified by Nanopore sequencing

Nanopore sequencing was used to sequence the variations in the NAFLD mice genome. After quality control, 643.97 Gb of clean data was obtained from the control and model samples ($n = 3$ for each group), with an average clean data volume of 107.33 Gb. The N50 value was 9.89 kb, and the average length of the reads was more than 6.95 kb (Supporting Information Table S14), with the reads length distribution of each sample given in Supporting Information Fig. S7A. The mapping ratio in each sample was over 99.04% in each sample according to the reference genome (Supporting Information Table S15). The distribution of sequencing depth and read coverage showed acceptable randomness of sequencing (Fig. 6A and B, Supporting Information Figs. S7B and S8). The Sniffles software was used to detect SVs in Nanopore sequencing and the SUR-VIVOR software was used to identify differential SVs between the NAFLD and control groups; which led to the discovery of 10,917 differential SVs, including 6568 DEL, 3785 INS, 405 TRA, 141 INV, and 18 DUP (Fig. 6C and Supporting Information Table S16). A majority of differential variations identified were DEL and INS, with a higher distribution number in Chr 10 to Chr 17 (Fig. 6D and E). All annotated transcribed features related to differential SVs were retrieved, to evaluate the functional impact of the differential SVs. Between the two
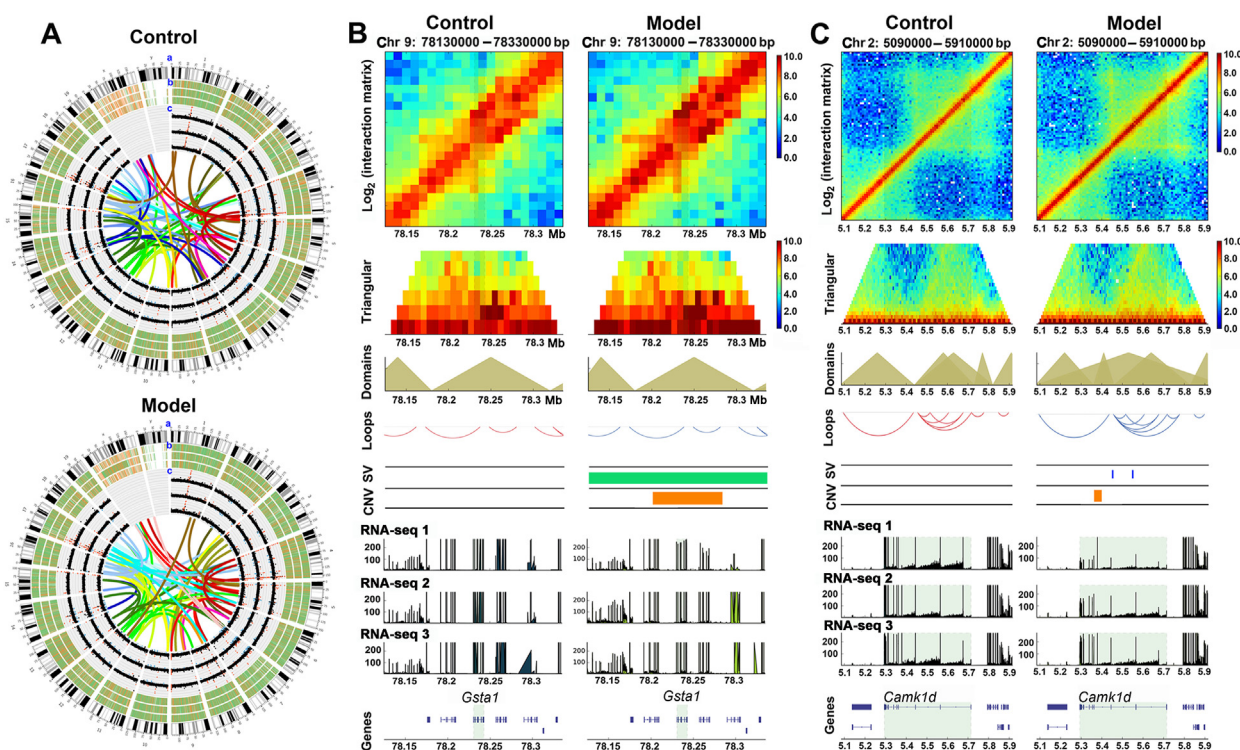
Figure 7 *Gsta1* and *Camk1d* as the candidate pathogenic genes of NAFLD caused by reorganization of chromatin structure identified by the comprehensive analysis of Hi-C, Nanopore, and RNA sequencing. (A) Summary of the 100 highest chromatin interactions in Hi-C and SV data in Nanopore sequencing in control and NAFLD groups: (a) The chromosome number; (b) The distribution of SVs (DEL is shown in green, INS is shown in orange, and DUP is shown in red); (c) Copy number data from Nanopore sequencing (gain and amplification are shown in red dot, loss and deletion are shown in blue dot, and the normal CVs are shown in gray dot). In the innermost layer, lines represent the highest 100 chromatin contact identified by Hi-C. (B) Mapping of local neighborhood interactions of *Gsta1*. (C) Mapping of local neighborhood interactions of *Camk1d*. The results of contact matrix, triangular, domain, and loop reflect the changes in spatial organization generated by Hi-C. SV and CNV were obtained from Nanopore sequencing. The abundance of gene expressions was generated by RNA-seq.

experimental groups, 381 DEGs overlapped with the differential SVs, of which 201 DEGs overlapped with DEL, 99 DEGs overlapped with INS, and 188 DEGs overlapped with INV (Supporting Information Table S17). Based on gene function, we found 14 SV-impacted DEGs involved in lipid transport and metabolism (*Acaca*, *Acsl5*, *Cbr3*, *Ces1f*, *Ces1g*, *Cryl1*, *Dgka*, *Dgkb*, *Dgkk*, *Idi1*, *Nlgn1*, *Nlgn3*, *Sugct*, and *Hao2*), and 8 SV-impacted DEGs associated with energy production and conversion (*Bckdhb*, *Me1*, *Hao2*, *Mt-co3*, *Mt-nd4*, *Nnt*, *Slc1a1*, and *Slc1a2*). Some of the SV-impacted DEGs are associated with NAFLD, however no association with NAFLD has been reported for some SV-impacted DEGs and may present as new molecular candidates in NAFLD. These results implicate SVs in the development of NAFLD through impact on gene regulation and chromatin organization.

3.7. CNVs in NAFLD identified by Nanopore sequencing

The deletion or duplication of DNA fragments causes copy number variations (CNVs), which are an important source of both normal and pathogenic variations in the genome. In this study, approximately 150 CNVs were found in each sample including loss, gain, and amplification (Supporting Information Fig. S9). Differential CNVs between the control and NAFLD groups were identified by combining with the results of RNA-seq, and 22 DEGs with differential CNVs including 12 upregulated and 10 downregulated genes were associated with the development of NAFLD (Supporting Information Table S18). CNV variants in

Cntnap2, *Mrgpra2b*, *Hist1h4n*, *Hist1h2br*, *Hist1h2ao*, *Pde4b*, *Gsta1*, *Gm3776*, *Kcnh1*, and *Camk1d* were simultaneously identified on comparing SV-impacted DEGs. *Camk1d* may have a potential role in liver gluconeogenesis[48], and a difference in CNV in the *Camk1d* locus was observed between the two groups, with a downregulation of gene expression in NAFLD group. A CNV difference was found in the *locus* of *Gsta1*, with increased expression in the NAFLD group, as previously reported and is associated with oxidative stress in NAFLD[49]. It is thus evident that mechanisms of gene regulation in NAFLD are associated with a difference in chromatin microenvironment.

3.8. Pathogenic gene candidates due to reorganization of chromatin structure in NAFLD

The primary goal of the study was to identify variations in genomic and spatial features to reveal the underlying pathogenesis of NAFLD. We comprehensively assessed the Hi-C and whole genome sequencing results and extracted the structure variations data by Nanopore sequencing, the 100 highest chromatin interactions in Hi-C, and the significant changes in structure variation and chromatin interaction between the control and NAFLD group (Fig. 7A). The occurrence of NAFLD is the results of these factors. Hence, we integrated Hi-C, Nanopore data with DEGs in NAFLD, and all DEGs that impacted by 3D disorganization or rearrangement were screened by comprehensive analysis of RNA-seq, Hi-C, and Nanopore sequencing results (Supporting Information Table S19). There were totally 555

Table 1 The candidate genes that might contribute to NAFLD screened from the comprehensive analysis of the results of RNA-seq, Hi-C, and ONT in this study.

Gene	Description	Location (bp)	Regulation in NAFLD	Hi-C			ONT	
				A/B	TAD	Loop	SV	CNV
Camk1d	Calcium/calmodulin-dependent protein kinase type 1D	Chr 2: 5293456–5676046	Down	–	–	–	√	√
Elovl6	Elongation of very long chain fatty acids protein 6	Chr 3: 129532385–129638493	Down	–	√	√	–	–
Gadd45g	Growth arrest and DNA damage-inducible protein GADD45 gamma	Chr 13: 51846674–51848474	Down	–	–	–	√	–
Pde4b	Phosphodiesterase 4B	Chr 4: 102254938–102607262	Down	–	–	–	√	√
Plce1	1-Phosphatidylinositol 4,5-bisphosphate phosphodiesterase epsilon-1	Chr 19: 38524196–38785100	Down	√	–	–	√	–
Plxnb1	Plexin-B1	Chr 9: 109095435–109119915	Down	–	√	√	√	–
Fgfr2	Fibroblast growth factor receptor 2	Chr 7: 130162450–130266808	Up	√	–	–	√	–
Grm8	Metabotropic glutamate receptor 8	Chr 6: 27275120–28135146	Up	–	–	√	√	–
Gsta1	Glutathione *S*-transferase A1	Chr 9: 78230668–78242683	Up	–	–	–	√	√
Kcnma1	Calcium-activated potassium channel subunit alpha-1	Chr 14: 23298693–24004205	Up	√	–	–	√	–
Sugct	Succinate-hydroxymethylglutarate CoA-transferase	Chr 13: 16857472–17695553	Up	√	–	√	√	–

DEGs related to the SVs in A/B compartment, TAD, chromatin loop, and/or CNV. Moreover, we screened DEGs that were related to the pathogenesis of NAFLD in gene ontology annotation and existed variation in genomic structure as the potential candidate genes, which may contribute to NAFLD development (Table 1). Variations in only CNV and SV in the *locus* of the upregulated *Gsta1* gene, without change in reorganized spatial structure, were discovered (Fig. 7B). The expression of *Fgfr2* was upregulated in the NAFLD group, with no variations in TAD, loop, and CNVs except for A/B compartment, the switching in which may be associated with *Fgfr2* upregulation (Supporting Information Fig. S10A). *Plxnb1*, a suppressor of liver tumor formation[50], was downregulated in NAFLD group, with variations observed in its spatial organization (Table 1). The deregulation of tumor-related genes in NAFLD may increase the potential of tumorigenic acquirement. Downregulation of *Elovl6* was associated with the reorganization of 3D structure in TAD and loop (Fig. S10B). *Camk1d* expression was downregulated and comprehensively affected by alternations in CNV and SV, chromatin loop, domains, and interaction matrix (Fig. 7C). These results indicate the functional relevance of each level of the 3D genome hierarchy, and the establishment of 3D genome provides multiple regulatory layers to alter gene expression in NAFLD development.

4. Discussion

NAFLD is the most common liver disease worldwide and can progress to inflammatory hepatitis, cirrhosis, and hepatocellular carcinoma. Due to an imbalance between hepatic fatty acid uptake and removal, the hallmarks of NAFLD are metabolic disorders and lipotoxic hepatocyte damage. It is imperative to alleviate the conditions of NAFLD; however, the pathogenesis of NAFLD is complex and needs to be elucidated further. In the present study, we investigated the 3D genome of NAFLD and performed integrated analyses using Hi-C, Nanopore sequencing, and RNA-seq. Mice were fed a HFD to induce NAFLD, and physiologic measures of abnormally elevated lipid levels, serum transaminases, and oxidative stress state were observed. Additionally, inflammatory injury and visible accumulation of lipid droplets in the liver of NAFLD mice was observed. These metabolic and physiological changes reflect the occurrence of hepatic damage in NAFLD mice.

The hepatic injury in NAFLD is accompanied by alternations in gene expression and an underlying remodeling of the liver transcriptome. Evidence suggests functional relevance of the different 3D genome hierarchies, with the 3D genome organization providing multiple additional regulatory layers to control gene expression[51]. The 3D genome organization contains spatial compartmentalization, organization into TADs, and chromatin looping, which modulate several processes including transcription, DNA repair, and replication. Due to the essential role of 3D genome organization in physiological conditions, their derangements may lead to potential pathogenic consequences. We generated and characterized genome-wide chromosome conformation Hi-C interaction matrices from the liver of normal and NAFLD mice to link the spatial regulatory elements and gene expressions. We identified reorganization in the interactions within chromosomes and between different chromosomes in the livers of NAFLD mice, and the relationship between the reorganizations and abnormal gene expressions were analyzed hierarchically.

Chromosomes are partitioned into either the active and open chromatin named A compartment or the inactive and closed chromatin called B compartment. The characteristics of A compartment include high GC-content and gene-rich regions, with generally high transcription levels. Conversely, B compartment is

gene-poor and less transcriptionally active[52]. A/B compartment switching is associated with the epigenetic alterations in gene expression during disease development[51]. In this study, approximately 33% of genome regions switched between A and B compartment in the liver of NAFLD mice compared to normal mice. To clarify the relationship between A/B switched compartments and changes in gene expression, we screened the DEGs in NAFLD impacted by A/B compartment switching and observed the trend of transcriptional activity in the A/B compartment consistent with the change. The screening revealed *Sugct* and *Fgfr2,* associated with lipid metabolic disorder and tumorigenic acquirement potential, respectively, as two of the DEGs in NAFLD.

TADs are the basic units of chromosome and are self-contact domains conserved among tissues and species[53]. Smaller domains within TADs called contact domains can be observed by using high resolution Hi-C, and are detected as peaks on the interaction matrices and correspond to more frequent contact regions than flanking *loci*[29]. TAD organization is important for genomic integrity and disease, and the disruption of TADs can cause abnormal contacts between enhancers and genes, thereby activating or inhibiting disease-relevant gene[54]. TAD disruption-mediated oncogene activation is reported to be a key mechanism in several cancer types[55]. In NAFLD, on drawing a map of the TADs, most TADs were conservative, however approximately 5% of TAD boundaries were disrupted upon NAFLD inducement. DEGs mapped to the changed TADs were screened, and down-regulation of *Elovl6*, a hallmark gene involved in *de novo* lipo-genesis was observed in relation to TAD variation in NAFLD. *Plxnb1*, a suppressor of liver tumor formation, was a TAD-impacted downregulated DEG in NAFLD. The disruptions in TADs and affected DEGs provide a new insight into understanding the pathogenesis of NAFLD.

Chromatin loop is an intrinsic property of the chromatin fiber and a fundamental mechanism for building the 3D genome hierarchy[56]. Chromatin looping plays an important role in gene transcription and RNA splicing by bringing distant genomic regions between the gene promoter and other regulatory elements, physically adjacent[57]. Therefore, chromatin looping provides a mechanism to magnify the long-range transcriptional and epigenetic effects of noncoding genetic variants. We identified differences in chromatin loops between the two groups and 98 DEGs associated with changes in looping. The TAD-impacted genes, *Elovl6* and *Plxnb1*, were affected by variations in chromatin looping as well. It may be possible that loop extrusion is a mechanistic aspect giving rise to TADs[40].

The mechanisms of derangements in 3D genome may involve preternatural mutations of 3D genome organizers or genetic changes including SVs and CNVs in the genome[58]. It is not determined yet if 3D genome disruption is the cause or the consequence of genetic variations, however abnormal 3D genome organization and genetic variations together cause dysregulation of genome function and hence, the phenotypes of different diseases[17]. Gene dysregulation caused by TAD disruption is a novel mechanism of SV-induced congenital developmental diseases[59]. Genomic rearrangements associated with human limb malformations altered the TAD organization in the *WNT6/IHH/EPHA4/PAX3 locus*, as reported by Lupiáñez et al.[60]. The application of next-generation sequencing (NGS) has obviously improved the throughput of data generation, which can make genome sequencing much faster and cheaper[61]. However, there are several drawbacks and pitfalls of NGS, most notably their relatively short reads. There are commonly a large number of repeated sequences in genomes that are longer than the NGS reads, which may cause misassemblies and gaps. Large SVs are more challenging to detect and characterize for NGS[62,63], that is an important issue due to the relevance between SVs and many diseases. In addition, NGS methods rely on PCR causes difficulties with regions of extreme GC%[63]. Nanopore sequencing distinguishes itself from NGS, in that it directly detects the input molecule without DNA synthesis or amplification, so there is no apparent limit to the length of DNA that can be sequenced[63]. Nanopore sequencing can not only directly observe the canonical four nucleobases, but also base modifications such as methylations[64] (the methylation data of this study will be uncovered in our following research). The main drawback of long-read sequencing is the relatively high error rate compared to NGS. Nanopore sequencing improved data quality by sequencing both template and complementary strands to obtain a final more accurate consensus read. Considering the above factors, Nanopore sequencing, a long-read method, was employed in the current study. We identified all SVs and CNVs in the genome, and retrieved the DEGs related to differential SVs and CNVs in NAFLD. We identified 381 SV-impacted and 23 CNV-impacted DEGs associated with NAFLD development, and dysregulation of *Camk1d*, *Pde4b*, and *Gsta1* genes was due to presence of SVs and CNVs in their locus. Additionally, presence of genomic rearrangements (SV or CNV) around the *locus* of DEGs including *Plce1*, *Plxnb1*, *Fgfr2*, *Grm8*, *Kcnma1*, and *Sugct*, was accompanied by disruptions of different 3D genome hierarchies. This indicates the presence of shared mechanisms that underlie variations in 3D chromatin conformation, genome, and transcription regulation, and may influence hepatocyte phenotypes in NAFLD.

There are some limitations to the current work. The regulatory role of 3D genome in the nuclear processes needs to be further elucidated in NAFLD. Genome editing studies to determine the functional consequences of perturbing the various 3D organization features are required in the future to clarify the mechanisms of 3D genome disruption causing NAFLD. Additionally, deciphering the combination of epigenetic mechanisms mediated by histone modification, DNA methylation, and noncoding RNAs is required for precise determination of genetic rearrangements and spatial organization disruption, and may likely uncover new mechanisms of NAFLD and identify novel therapeutic targets.

5.   Conclusions

By combining transcriptome, long-read sequencing, and 3D genome analyses, we revealed occurrence of multiple levels of alternations including SV, CNV, spatial genome reorganization, and their influences on gene expression in NAFLD. Our results reveal gene regulation through genetic rearrangements and spatial organization disruption and identify candidate genes associated with the pathogenesis of NAFLD. These findings extend our knowledge of the spatial disorganization and genetic rearrangements of the liver genome of NAFLD and may have implications on clinical treatment and drug development for NAFLD.

Author contributions

Lina Xu and Jinyong Peng conceived this project and designed the study. Lina Xu performed the experiments and analyzed the data. Lianhong Yin, Yan Qi, Xuemei Tan, and Meng Gao gave methodological support and conceptual advice. Lina Xu and Jinyong Peng wrote and edited the manuscript. All authors have read and approved the final text and consent to its publication.

Conflicts of interest

The authors have declared that no competing interest exists.

Appendix A. Supporting information

Supporting data to this article can be found online at https://doi.org/10.1016/j.apsb.2021.03.022.

References

1. Younossi ZM. Non-alcoholic fatty liver disease—a global public health perspective. *J Hepatol* 2019;**70**:531−44.
2. Yan T, Yan N, Wang P, Xia Y, Hao H, Wang G, et al. Herbal drug discovery for the treatment of nonalcoholic fatty liver disease. *Acta Pharm Sin B* 2020;**10**:3−18.
3. Anstee QM, Seth D, Day CP. Genetic factors that affect risk of alcoholic and nonalcoholic fatty liver disease. *Gastroenterology* 2016;**150**:1728−44.
4. Umano GR, Caprio S, Di Sessa A, Chalasani N, Dykas DJ, Pierpont B, et al. The rs626283 variant in the *MBOAT7* gene is associated with insulin resistance and fatty liver in caucasian obese youth. *Am J Gastroenterol* 2018;**113**:376−83.
5. Eslam M, Valenti L, Romeo S. Genetics and epigenetics of NAFLD and NASH: clinical impact. *J Hepatol* 2018;**68**:268−79.
6. Trépo E, Valenti L. Update on NAFLD genetics: from new variants to the clinic. *J Hepatol* 2020;**72**:1196−209.
7. Jonas W, Schürmann A. Genetic and epigenetic factors determining NAFLD risk. *Mol Metab* 2021;**50**:101111.
8. Chai C, Cox B, Yaish D, Gross D, Rosenberg N, Amblard F, et al. Agonist of RORA attenuates nonalcoholic fatty liver progression in mice *via* up-regulation of microRNA 122. *Gastroenterology* 2020;**159**:999−1014.
9. Kan Z, Zheng H, Liu X, Li S, Barber TD, Gong Z, et al. Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res* 2013;**23**:1422−33.
10. CretuStancu M, van Roosmalen MJ, Renkens I, Nieboer MM, Middelkamp S, de Ligt J, et al. Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat Commun* 2017;**8**:1326.
11. Gilpatrick T, Lee I, Graham JE, Raimondeau E, Bowen R, Heron A, et al. Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat Biotechnol* 2020;**38**:433−8.
12. Taberlay PC, Achinger-Kawecka J, Lun AT, Buske FA, Sabir K, Gould CM, et al. Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. *Genome Res* 2016;**26**:719−31.
13. Yu M, Ren B. The three-dimensional organization of mammalian genomes. *Annu Rev Cell Dev Biol* 2017;**33**:265−89.
14. Marchal C, Sima J, Gilbert DM. Control of DNA replication timing in the 3D genome. *Nat Rev Mol Cell Biol* 2019;**20**:721−37.
15. Pope BD, Ryba T, Dileep V, Yue F, Wu W, Denas O, et al. Topologically associating domains are stable units of replication-timing regulation. *Nature* 2014;**515**:402−5.
16. Crane E, Bian Q, McCord RP, Lajoie BR, Wheeler BS, Ralston EJ, et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* 2015;**523**:240−4.
17. Krijger PH, de Laat W. Regulation of disease-associated gene expression in the 3D genome. *Nat Rev Mol Cell Biol* 2016;**17**:771−82.
18. Barutcu AR, Lajoie BR, McCord RP, Tye CE, Hong D, Messier TL, et al. Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. *Genome Biol* 2015;**16**:214.
19. Weischenfeldt J, Dubash T, Drainas AP, Mardin BR, Chen Y, Stütz AM, et al. Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nat Genet* 2017;**49**:65−74.
20. Quigley DA, Dang HX, Zhao SG, Lloyd P, Aggarwal R, Alumkal JJ, et al. Genomic hallmarks and structural variation in metastatic prostate cancer. *Cell* 2018;**174**:758−69.
21. Wang WJ, Li LY, Cui JW. Chromosome structural variation in tumorigenesis: mechanisms of formation and carcinogenesis. *Epigenet Chromatin* 2020;**13**:49.
22. Rosa-Garrido M, Chapski DJ, Schmitt AD, Kimball TH, Karbassi E, Monte E, et al. High-resolution mapping of chromatin conformation in cardiac myocytes reveals structural remodeling of the epigenome in heart failure. *Circulation* 2017;**136**:1613−25.
23. Miguel-Escalada I, Bonàs-Guarch S, Cebola I, Ponsa-Cobas J, Mendieta-Esteban J, Atla G, et al. Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. *Nat Genet* 2019;**51**:1137−48.
24. Fasolino M, Goldman N, Wang W, Cattau B, Zhou Y, Petrovic J, et al. Genetic variation in type 1 diabetes reconfigures the 3D chromatin organization of T cells and alters gene expression. *Immunity* 2020;**52**:257−74.
25. Sun JH, Zhou L, Emerson DJ, Phyo SA, Titus KR, Gong W, et al. Disease-associated short tandem repeats co-localize with chromatin domain boundaries. *Cell* 2018;**175**:224−38.
26. Brandt MM, Meddens CA, Louzao-Martinez L, van den Dungen NAM, Lansu NR, Nieuwenhuis EES, et al. Chromatin conformation links distal target genes to CKD *loci*. *J Am Soc Nephrol* 2018;**29**:462−76.
27. McCord RP, Kaplan N, Giorgetti L. Chromosome conformation capture and beyond: toward an integrative view of chromosome structure and function. *Mol Cell* 2020;**77**:688−708.
28. Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* 2017;**544**:59−64.
29. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014;**159**:1665−80.
30. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;**25**:1754−60.
31. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* 2015;**16**:259.
32. Schmid MW, Grob S, Grossniklaus U. HiCdat: a fast and easy-to-use Hi-C data analysis tool. *BMC Bioinf* 2015;**16**:277.
33. Varoquaux N, Ay F, Noble WS, Vert JP. A statistical approach for inferring the 3D structure of the genome. *Bioinformatics* 2014;**30**:i26−33.
34. Servant N, Lajoie BR, Nora EP, Giorgetti L, Chen CJ, Heard E, et al. HiTC: exploration of high-throughput 'C' experiments. *Bioinformatics* 2012;**28**:2843−4.
35. Wang XT, Cui W, Peng C. HiTAD: detecting the structural and functional hierarchies of topologically associating domains from chromatin interactions. *Nucleic Acids Res* 2017;**45**:e163.
36. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;**38**:576−89.
37. Robinson MD, McCarthy DJ, Smyth GK, edgeR. A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;**26**:139−40.

38. Wick RR, Judd LM, Holt KE. Performance of neural network base-calling tools for Oxford Nanopore sequencing. *Genome Biol* 2019;**20**: 129.

39. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;**34**:3094−100.

40. Mirny LA, Imakaev M, Abdennur N. Two major mechanisms of chromosome organization. *Curr Opin Cell Biol* 2019;**58**:142−52.

41. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* 2018;**15**:461−8.

42. Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun* 2017;**8**:14061.

43. Euskirchen P, Bielle F, Labreche K, Kloosterman WP, Rosenberg S, Daniau M, et al. Same-day genomic and epigenomic diagnosis of brain tumors using real-time nanopore sequencing. *Acta Neuropathol* 2017;**134**:691−703.

44. Niska-Blakie J, Gopinathan L, Low KN, Kien YL, Goh CMF, Caldez MJ, et al. Knockout of the non-essential gene *SUGCT* creates diet-linked, age-related microbiomedisbalance with a diabetes-like metabolic syndrome phenotype. *Cell Mol Life Sci* 2020;**77**:3423−39.

45. Hotta K, Kitamoto T, Kitamoto A, Ogawa Y, Honda Y, Kessoku T, et al. Identification of the genomic region under epigenetic regulation during non-alcoholic fatty liver disease progression. *Hepatol Res* 2018;**48**:E320−34.

46. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012;**485**:376−80.

47. Lee MH, Lu K, Hazard S, Yu H, Shulenin S, Hidaka H, et al. Identification of a gene, *ABCG5*, important in the regulation of dietary cholesterol absorption. *Nat Genet* 2001;**27**:79−83.

48. Rausch JC, Lavine JE, Chalasani N, Guo X, Kwon S, Schwimmer JB, et al. Genetic variants associated with obesity and insulin resistance in hispanic boys with nonalcoholic fatty liver disease. *J Pediatr Gastroenterol Nutr* 2018;**66**:789−96.

49. Lee DH, Han DH, Nam KT, Park JS, Kim SH, Lee M, et al. Ezetimibe, an NPC1L1 inhibitor, is a potent Nrf 2 activator that protects mice from diet-induced nonalcoholic steatohepatitis. *Free Radic Biol Med* 2016;**99**:520−32.

50. Song CQ, Li Y, Mou H, Moore J, Park A, Pomyen Y, et al. Genome-wide CRISPR screen identifies regulators of mitogen-activated protein kinase as suppressors of liver tumors in mice. *Gastroenterology* 2017; **152**:1161−73.

51. Krumm A, Duan Z. Understanding the 3D genome: emerging impacts on human disease. *Semin Cell Dev Biol* 2019;**90**:62−77.

52. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009;**326**:289−93.

53. Gong Y, Lazaris C, Sakellaropoulos T, Lozano A, Kambadur P, Ntziachristos P, et al. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nat Commun* 2018;**9**:542.

54. Lupiáñez DG, Spielmann M, Mundlos S. Breaking TADs: how alterations of chromatin domains result in disease. *Trends Genet* 2016; **32**:225−37.

55. Peifer M, Hertwig F, Roels F, Dreidax D, Gartlgruber M, Menon R, et al. Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* 2015;**526**:700−4.

56. Xiang JF, Yin QF, Chen T, Zhang Y, Zhang XO, Wu Z, et al. Human colorectal cancer-specific *CCAT1-L* lncRNA regulates long-range chromatin interactions at the MYC *locus*. *Cell Res* 2014;**24**:513−31.

57. Nuebler J, Fudenberg G, Imakaev M, Abdennur N, Mirny LA. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc Natl Acad Sci U S A* 2018;**115**:E6697−706.

58. Wu P, Li T, Li R, Jia L, Zhu P, Liu Y, et al. 3D genome of multiple myeloma reveals spatial genome disorganization associated with copy number variations. *Nat Commun* 2017;**8**:1937.

59. Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* 2016;**538**:265−9.

60. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 2015; **161**:1012−25.

61. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 2016;**17**: 333−51.

62. Salzberg SL, Yorke JA. Beware of mis-assembled genomes. *Bioinformatics* 2005;**21**:4320−1.

63. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The third revolution in sequencing technology. *Trends Genet* 2018;**34**:666−81.

64. Rand AC, Jain M, Eizenga JM, Musselman-Brown A, Olsen HE, Akeson M, et al. Mapping DNA methylation with high-throughput nanopore sequencing. *Nat Methods* 2017;**14**:411−3.