Supplemental Materials for

Diploid genome architecture revealed by multi-omic data of hybrid mice

Zhijun Han[1,2], Kairong Cui[3], Katarzyna Placek[3], Ni Hong[1], Chengqi Lin[2], Wei Chen[1], Keji Zhao[3*], Wenfei Jin[1*]

[1]Department of Biology, Southern University of Science and Technology, Shenzhen, Guangdong 518055, China;

[2]Institute of Life Sciences, Southeast University, Nanjing, Jiangsu 210096, China;

[3]Systems Biology Center, National Heart, Lung and Blood Institute, NIH, Bethesda, MD 20892, USA.

[*]To whom correspondence should be addressed. E-mail: jinwf@sustech.edu.cn (W.J.); zhaok@nhlbi.nih.gov (K.Z.)

Contents

Supplemental Methods

Mice and Isolation of Cells

Both C57BL/6J (B6) mice and CAST/EiJ (Cast) mice were purchased from Jackson Laboratories (Bar Harbor, ME, USA). This study was reviewed and approved by the Animal Care and Use Committee of the National Heart Lung and Blood Institute. All mice received humane treatment according to NIH guidelines and the ‘‘Guiding Principles for Research Involving Animals and Human Beings.’’

Male B6 and female Cast mice were crossbred to generate the F1 hybrid mice. Naïve CD4 T cells were purified from spleen and lymph nodes of B6, Cast and hybrid mice with CD4 microbeads (Miltenyi Biotech, $CD4^+CD62L^+$ T cell isolation kit II, mouse). The purity of naïve CD4 T cells was assessed by flow cytometry for $CD4^+$, $CD8^-$ and $CD62L^+$ using FACSCanto II (BD Biosciences). Over 98% purity of $CD4^+CD8^-CD62L^+$ cells was considered for further experiments. All experiments were conducted on 6 to 8-week-old mice. Antibodies including anti-CD4 (RM4–5) and anti-CD62L (MEL-14) for flow-cytometry analyses were purchased from eBiosciences. Dead cells were excluded by DAPI staining.

3e Hi-C

The multiple-enzyme Hi-C (3e Hi-C) was performed according to our previous studies (Ren et al. 2017; Hu et al. 2018). About 1,000,000 naive CD4 T cells were cross-linked with 1% formaldehyde for 10 mins. Cells were lysed and digested with 20 Units CviQ I (NEB), and 20 Units CviA II (NEB) at 25°C for 20 minutes, then 20 Units Bfa I (NEB) at 37°C for 20 minutes. The reaction was stopped by washing the cells twice with 600 mL wash buffer (10mM NaCl, 1mM EDTA, 0.1% triton-100). DNA ends were marked by biotin-14-dATP with Klenow (large) for 1h at 37°C. Blunt-end DNA fragments were ligated with T4 DNA Ligase overnight at 16°C. DNA was then reverse cross-linked and purified by phenol/chloroform extraction. Biotin was removed from unligated DNA-ends by T4 DNA polymerase for 2hs at 12°C. DNA was purified by phenol/chloroform and sheared to 300-500 bp by sonication followed by DNA-end repair and addition of adenosine. Biotin labeled DNA was pull-downed by streptavidin

beads followed by Illumina adaptor ligation and PCR amplification. DNA fragments of 300 to 700 bp were isolated 2% agarose gel and sequenced by paired-end sequencing on Illumina Hiseq 2500.

ChIP-seq

ChIP-seq experiments were performed as described previously (Barski et al. 2007). In brief, 1,000,000 of Naïve CD4 T cells were fixed for 10 min with 1% formaldehyde in complete medium, sonicated and chromatin immunoprecipitation was performed with antibodies against CTCF (07-729, Millipore), RNA Polymerase II (ab5408, Abcam), H3K4me3 (ab8580, Abcam) and H3K4me2 (ab32356, Abcam). The DNA was then ligated with the 'Y' shaped Illumina adaptor and amplified for 18 cycles using indexing primers as described. PCR products between 160-300bp were isolated on 2% E-gel for sequencing on Illumina Hiseq2500.

RNA-seq

Total RNA was extracted and purified with miRNeasy micro kit (217084; Qiagen) and DNase set (79254; Qiagen), followed by delusion with 10 mL of RNase-free water. Purified total RNA was reverse transcribed with the Ovation RNA-Seq System V2 (7102-08; NuGEN Technologies). cDNA was sonicated in a Diagenode Bioruptor (level M, for a total of 30 min of 20s on and 20s off) to obtain fragment in the size range of 100-400bp. Indexed libraries were prepared with a Multiplexing Sample Preparation Oligonucleotide Kit (1005709; Illumina) and DNA End-repair Kit (ER81050; Epcentre) according to the user's manual (Epcentre) and sample preparation guide (Illumina).

SNPs calling and cross validation

We downloaded the files containing mouse indels and SNPs information (about 21

million SNPs) (mgp.v5.merged.indels.dbSNP142.normed.vcf) from Mouse Genomes Project (ftp://ftp-mouse.sanger.ac.uk) (Keane et al. 2011; Yalcin et al. 2012) and examined the SNPs between C57BL/6J and CAST/EiJ using SNPsplit toolkit (Krueger and Andrews 2016). Next, we built the mm10 Bowtie 2 index with all SNP positions being masked by the ambiguity base 'N'. All reads from B6 and Cast Hi-C libraries were mapped to this N-masked reference using Bowtie 2 (Langmead and Salzberg 2012) (with parameter --very-sensitive -L 30 --score-min L,-0.6,-0.2 --end-to-end). For retrieving the highly confident SNPs, we re-called the genotype for each SNP locus in B6 and Cast Hi-C libraries using mpileup in samtools (Li et al. 2009) with reads MAPQ $\geq$ 30. The called SNPs showing identical states with that downloaded from mouse genome project were kept for further analyses. In total, we obtained 13 million highly confident SNPs for distinguishing Cast haploid reads and B6 haploid reads in hybrid mice.

Hi-C, ChIP-seq and RNA-seq data processing

The paired-end reads from Hi-C, also known as paired-end tags (PETs), were mapped to N-masked mm10 reference with Bowtie 2 separately. Both ends of the PETs with MAPQ $\geq$ 30 were kept for further analyses. The PETs with the same start and the same end were treated as redundant PETs, among which only one PET was kept for further analysis. Intra-chromosomal PETs within 10kb were considered as self-ligations and were filtered out. The interaction matrices were constructed at different resolution, from 5kb to 100kb. In general, the resolutions for LBS analyses, ID analyses, heatmap plot and compartment analyses were 5kb, 20kb and 100kb, respectively. The matrices were normalized using iterative correction algorithm ICE (Imakaev et al. 2012) , and further optimized in HiC-Pro (Servant et al. 2015). A|B compartments were calculated as described in Lieberman-Aiden et al (Lieberman-Aiden et al. 2009). PETs from hybrid mice Hi-C library were split into haploid using SNPsplit using highly confident SNPs, and the downstream processing was similar to normal Hi-C samples.

ChIP-seq reads were mapped and filtered similarly to Hi-C reads, and extraction of reads for generation of haploid ChIP-seq data was similar to the generation of haploid Hi-C. Peaks were called using MACS2 (Zhang et al. 2008). For comparison of the differences of histone modifications and TF bindings between Fb6 and Fcast, we

counted the allele-specific reads located in the peaks using *intersect* in BEDTools (Quinlan 2014). The significances of the differentially binding sites were calculated by DESeq (Love et al. 2014). ChIP-seq reads were extended to 150bp and normalized by sample size, then converted to bedGraph for visualization.

RNA-seq reads were mapped to N-masked mm10 reference by TopHat2 (Kim et al. 2013), and reads with MAPQ ≥30 were kept for further analyses. The batch effect of replicates was adjusted by ComBat from R packa ge sva (Leek et al. 2012), and differentially expressed genes were identified by DESeq2 (Love et al. 2014) with FDR < 0.05 and fold change > 1.8. Allele-specific reads in turn were grouped by SNPsplit, whereas allele-specific gene expression was counted and quantified as normal.

PCA analysis of inter-chromosomal interaction matrix

In order to analyze the inter-chromosomal interaction patterns in the nucleus of hybrid mouse, we retrieved all the inter-chromosomal PETs with both ends containing confident SNPs in hybrid mouse Hi-C data. We generated chromosomal resolution interaction matrix for all the 40 paternal and maternal chromosomes. To eliminate the bias induced by chromosome size, the PETs number was normalized by formula (1) or ICE normalization (Imakaev et al. 2012). We found the first two PCs contribute majority of inter chromosomal variations after we conduced PCA on the allelic specific interaction matrix. In order to illustrate the different interaction patterns between active region and inactive region on X Chromosome, the full-length X Chromosome was separated into X-a and X-i at the Dxz4 loci. To remove the effects from interactions between homologous chromosomes, the interaction counts between homologous chromosomes were set to zero when conduct PCA.

$$M_{i,j} = C_{i,j} * \left(\sum C_{i,} + \sum C_{,j}\right) \Big/ \sum C_{i,} * \sum C_{,j} \; ; \qquad N_{i,j} = M_{i,j} \Big/ \sum_i \sum_j M_{i,j} * \sum_i \sum_j C_{i,j} \qquad (1)$$

Where $C_{i,j}$ is the raw interaction count for each chromosome pair i and j, $M_{i,j}$ is the intermediate count after normalization and $N_{i,j}$ is the final normalized count which is scaled to the original sample size.

Modeling the 3D nucleus in hybrid mouse

For construction of 3D nucleus of hybrid mouse, we developed an iteratively weighted adjusting algorithm to infer the relative positioning of each chromosome in 3D space. The purpose of iterative process of this model is to continuously minimize the sum of errors between coordinate-based distance and the 'real' distance which was converted from the allele-specific interaction matrix (Fig S1A). In brief, we first initialized random xyz values for each chromosome, then iteratively adjusted the xyz values based on the distance errors between that chromosome and the other chromosomes. We noted that this model would reach the local optimum and steady state after finite iterations (Fig S3A). The mathematical formulae of this model is:

$$X = X_0;$$
$$X_i = \sum_j \left(\frac{D_{ij}}{|| X_i - X_j ||} * \left(X_i - X_j \right) + X_j \right) * W_{ij}; \quad (2)$$
$$S = \sum | \left(D_{ij} - || X_i - X_j || \right) * W_{ij} |$$

Where X_0 is randomly initialized 3D coordinates (xyz values) for each chromosome, D_{ij} is the 'real' distance matrix converted from the allele-specific interaction matrix based on the fitted PETs count to distance function (Fig S1A), W_{ij} represents the weight matrix for each chromosome pair converted from the allele-specific interaction matrix. The aim of the iteration is to achieve the smallest S.

Since the interaction between homologous chromosomes may introduce bias to the 3D model, we reset the interactions between homologous chromosome pairs to the average interaction density of the corresponding chromosome to all non-homologous chromosomes. We also constructed the 3D model with split active and inactive X Chromosome (Fig S3D) and without X Chromosomes (Fig S3E).

Data normalization for inferring chromatin organization

Data normalization is very important for analyzing and presenting Hi-C data. We normalized the interaction matrix using iterative correction algorithm ICE in the original analysis and the mappability was considered in the normalization. Mouse autosomes are telocentric chromosome which are rod-like with one centromere-end and

one telomere-end. We separated each chromosome into three equal length sections: centromere-end, middle, telomere-end, in which centromere region and telomere region only account 1.5%-4.9% and 0.1%-0.3% of the chromosomes, respectively (Table S3). Furthermore, only 864 and 909 reads were mapped to centromere regions, which are ignorable compared to millions of reads mapped to centromere-end or telomere-end. Therefore, either centromere regions or telomer regions have unlikely significant impacts on the results due to that they only account for a few reads in the study.

Local boundary score (LBS) analysis

LBS is defined as the logarithm of ratio of local interactions to inter-local interactions in a given region (Fig 4A). In details, genome was separated into many small bins with equal size and several continuous bins made up a locus (for example, bin size=2kb and locus=150kb). For a given bin, the interactions within its left locus (A_1 in Fig 4A) and its right locus (A_2 in Fig 4A) are defined as its intra-loci interactions. The interactions between these two loci (150kb x 150kb, B in Fig 4A) are inter-loci iterations, then the log2-transformed ratio of intra-loci interactions to inter-loci interactions was calculated as LBS of the bin. To avoid false positive in sparse interactions, chromosome-wide average value was set as the initial background for intra- and inter-loci interactions. In this way, the peak of LBS is the relative inter-loci interaction in local region reaches the maximum value, which indicated the presence of the TAD boundary. Peakdet (http://billauer.co.il/peakdet.html) was used to call local peaks of LBSs. To compare TAD boundary shift between samples, LBS biases were calculated using ROSE that initially for distinguishing super enhancers from typical enhancers (Whyte et al. 2013).

We compared the TADs called by LBSs with that by HiCExplorer (Wolff et al. 2018), with default parameters. Results showed that more than 80% of the TAD boundaries called from two methods were exactly the same (Fig S5B), and the median sizes of TADs at the given resolution are 300kb for both methods (Fig S5C), indicating that LBS is efficiently for the identification of TADs at very high resolution. B6 and Cast hybrid mice had been used for investigation of cis- and trans-effect of gene regulation and the B6 haploid and Cast haploid could be distinguished from each other based on strain specific alleles (Goncalves et al. 2012).

CTCF motifs in shifted TAD boundaries

We identified 4,197 TAD boundaries in B6 and 4,917 TAD boundaries in Cast, with 648 boundary shifts between B6 and Cast. Thus about 84.6% TADs in B6 are conserved in Cast, or 86.8% TADs in Cast are conserved in B6. The results indicated the TADs between B6 and Cast are much more highly conserved than the TAD conservation between other cell types showed by the previous study (Dixon et al. 2012), in which 30-40% TAD boundaries are conserved between human and mouse and 50-70% TAD boundaries are conserved between different cell types.

The 648 shifted boundaries between B6 and Cast represent the difference of chromatin status, which could be caused by either genetic changes or epigenetic dynamics. We identified a total of 2,460 CTCF binding sites with CTCF motif in these 648 shifted boundaries (CTCF binding motifs were searched by fimo referring MA1102.2). While only 97 (4%) of these 2,460 motifs contain strain specific-SNPs, with only 8 motifs potentially impacting the binding of CTCF. Thus, we think the genetic changes of CTCF sites may play only a limited role in the TAD boundary shift.

References

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129**: 823-837.

Goncalves A, Leigh-Brown S, Thybert D, Stefflova K, Turro E, Flicek P, Brazma A, Odom DT, Marioni JC. 2012. Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. *Genome research* **22**: 2376-2384.

Hu G, Cui K, Fang D, Hirose S, Wang X, Wangsa D, Jin W, Ried T, Liu P, Zhu J et al. 2018. Transformation of Accessible Chromatin and 3D Nucleome Underlies Lineage Commitment of Early T Cells. *Immunity* **48**: 227-242 e228.

Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, Dekker J, Mirny LA. 2012. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature methods* **9**: 999-1003.

Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M et al. 2011. Mouse genomic variation and its effect on phenotypes and gene

regulation. *Nature* **477**: 289-294.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**: R36.

Krueger F, Andrews SR. 2016. SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes. *F1000Res* **5**: 1479.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**: 357-359.

Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**: 882-883.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**: 289-293.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**: 550.

Quinlan AR. 2014. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* **47**: 11 12 11-34.

Ren G, Jin W, Cui K, Rodrigez J, Hu G, Zhang Z, Larson DR, Zhao K. 2017. CTCF-Mediated Enhancer-Promoter Interaction Is a Critical Regulator of Cell-to-Cell Variation of Gene Expression. *Molecular cell* **67**: 1049-1058 e1046.

Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, Heard E, Dekker J, Barillot E. 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome biology* **16**: 259.

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. 2013. Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell* **153**: 307-319.

Wolff J, Bhardwaj V, Nothjunge S, Richard G, Renschler G, Gilsbach R, Manke T, Backofen R, Ramirez F, Gruning BA. 2018. Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis,

quality control and visualization. *Nucleic acids research* **46**: W11-W16.

Yalcin B, Adams DJ, Flint J, Keane TM. 2012. Next-generation sequencing of experimental mouse strains. *Mamm Genome* **23**: 490-498.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**

Supplemental Resource Table.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-mouse CD4 APC	eBioscience	Cat# 17-0042-82
anti-mouse CD62L Pacific Blue	Biolegend	Cat# 104424
H3K4me1	Abcam	Cat# ab8895
H3K4me2	Abcam	Cat# ab32356
H3K4me3	Abcam	Cat# ab8580
CTCF	Millipore	Cat# 07-729
RNA Polymerase II	Abcam	Cat# ab5408
Experimental Models: Organisms/Strains		
Mouse: CAST/EiJ	Jackson Laboratory	Stock No: 000928
Mouse: C57BL/6	Jackson Laboratory	Stock No: 000664
Software and Algorithms		
Bowtie 2	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml	NA
Samtools	samtools.sourceforge.net/	NA
TopHat2	ccb.jhu.edu/software/tophat/index.shtml	NA
BEDTools	https://sourceforge.net/projects/bedtools/	NA
SNPsplit	www.bioinformatics.babraham.ac.uk/projects/SNPsplit/	NA
HOMER 4.7	http://homer.ucsd.edu/homer/interactions	NA
EdgeR	http://bioconductor.org/packages/release/bioc/html/edgeR.html	NA
WashU epigenome browser	http://wiki.wubrowse.org	NA
HiCExplorer	https://hicexplorer.readthedocs.io	NA
Peakdet	http://billauer.co.il/peakdet.html	
Hi-Corrector1.2	http://zhoulab.usc.edu/Hi-Corrector	NA
ROSE	https://bitbucket.org/young_computation/rose	NA
Fit-HiC	https://noble.gs.washington.edu/proj/fit-hi-c	NA
R	https://www.r-project.org	NA
DAVID	https://david.ncifcrf.gov/	NA

Supplemental Tables

Table S1. Sequence libraries generated in this study.

Library type	Library	Sequenced reads	Mapped reads
3e Hi-C	B6-rep1	237,868,627	225,734,453
3e Hi-C	B6-rep2	224,838,109	212,017,072
3e Hi-C	B6-rep3	748,962,123	714,185,833
3e Hi-C	B6-rep4	539,145,243	515,696,896
3e Hi-C	Cast-rep1	25,553,829	23,855,757
3e Hi-C	Cast-rep2	429,371,157	398,759,778
3e Hi-C	Hybrid-rep1	12,298,602	11,627,183
3e Hi-C	Hybrid-rep2	311,860,870	311,850,615
ChIP-seq	H3K4me2	9,311,172	7,766,416
ChIP-seq	H3K4me3	10,940,581	8,830,815
ChIP-seq	PolII	15,529,391	11,490,569
ChIP-seq	CTCF	13,132,563	9,421,527
RNA-seq	B6-rep1	13,871,558	10,425,217
RNA-seq	B6-rep2	18,000,576	14,700,817
RNA-seq	B6-rep3	15,687,279	13,041,608
RNA-seq	Cast-rep1	13,324,093	10,452,689
RNA-seq	Cast-rep2	15,862,646	12,269,036
RNA-seq	Cast-rep3	13,729,557	11,041,403
RNA-seq	Hybrid-rep1	15,568,013	12,403,659
RNA-seq	Hybrid-rep2	16,674,430	13,616,182
RNA-seq	Hybrid-rep3	12,880,554	10,612,284

Table S2. Statistics of Hi-C Libraries.

Libraries	Sequenced PETs	Mapped PETs	MAPQ≥30 & unique	IntraChr	IntraChr <20kb	IntraChr >20kb
B6-rep1	237,868,627	225,734,453	116,738,507	90,748,029	61,127,208	29,620,821
B6-rep2	224,838,109	212,017,072	105,633,075	82,547,299	54,737,136	27,810,163
B6-rep3	748,962,123	714,185,833	361,080,940	288,170,772	202,080,319	86,090,453
B6-rep4	539,145,243	515,696,896	268,409,206	219,774,900	162,285,965	57,488,935
Cast-rep1	25,553,829	23,855,757	12,735,265	10,766,993	8,243,486	2,523,507
Cast-rep2	429,371,157	398,759,778	178,472,596	139,252,207	103,959,226	35,292,981
Hybrid-rep1	12,298,602	11,627,183	6,494,488	5,132,872	3,418,062	1,714,810
Hybrid-rep2	311,860,870	311,850,615	264,075,948	198,312,044	125,864,362	72,447,682
Fb6	--	--	39,058,806	30,554,583	8,634,185	21,920,398
Fcast	--	--	35,862,641	27,577,727	7,770,693	19,807,034

Table S3. Length ratio of centromere and telomere for each chromosome.

Chr	Centromere	Telomere
chr1	1.53%	0.10%
chr2	1.65%	0.11%
chrX	1.75%	0.12%
chr3	1.87%	0.12%
chr4	1.92%	0.13%
chr5	1.98%	0.13%
chr6	2.00%	0.13%
chr7	2.06%	0.14%
chr10	2.30%	0.15%
chr8	2.32%	0.15%
chr14	2.40%	0.16%
chr9	2.41%	0.16%
chr11	2.46%	0.16%
chr13	2.49%	0.17%
chr12	2.50%	0.17%
chr15	2.88%	0.19%
chr16	3.05%	0.20%
chr17	3.16%	0.21%
chr18	3.31%	0.22%
chr19	4.88%	0.33%

Supplemental Figures

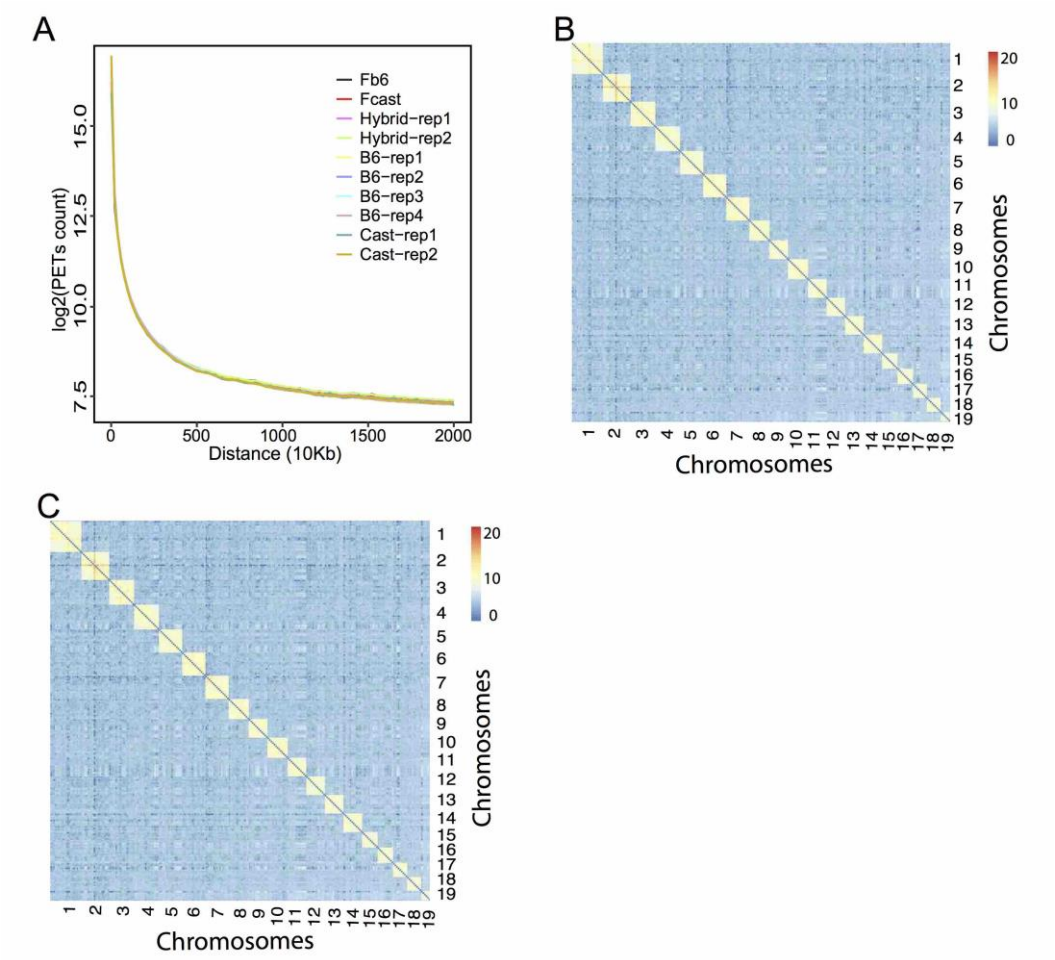

Figure S1. Distributions of PETs distance and heatmap of interaction in hybrid mice system. (A) Distribution of PET along distances of PETs for Hi-C replicates and haploid in hybrid mice. (B) Heatmap of genome-wide interaction of Fb6 at 10Mb resolution. (C) Heatmap of genome-wide interaction of Fcast at 10Mb resolution.

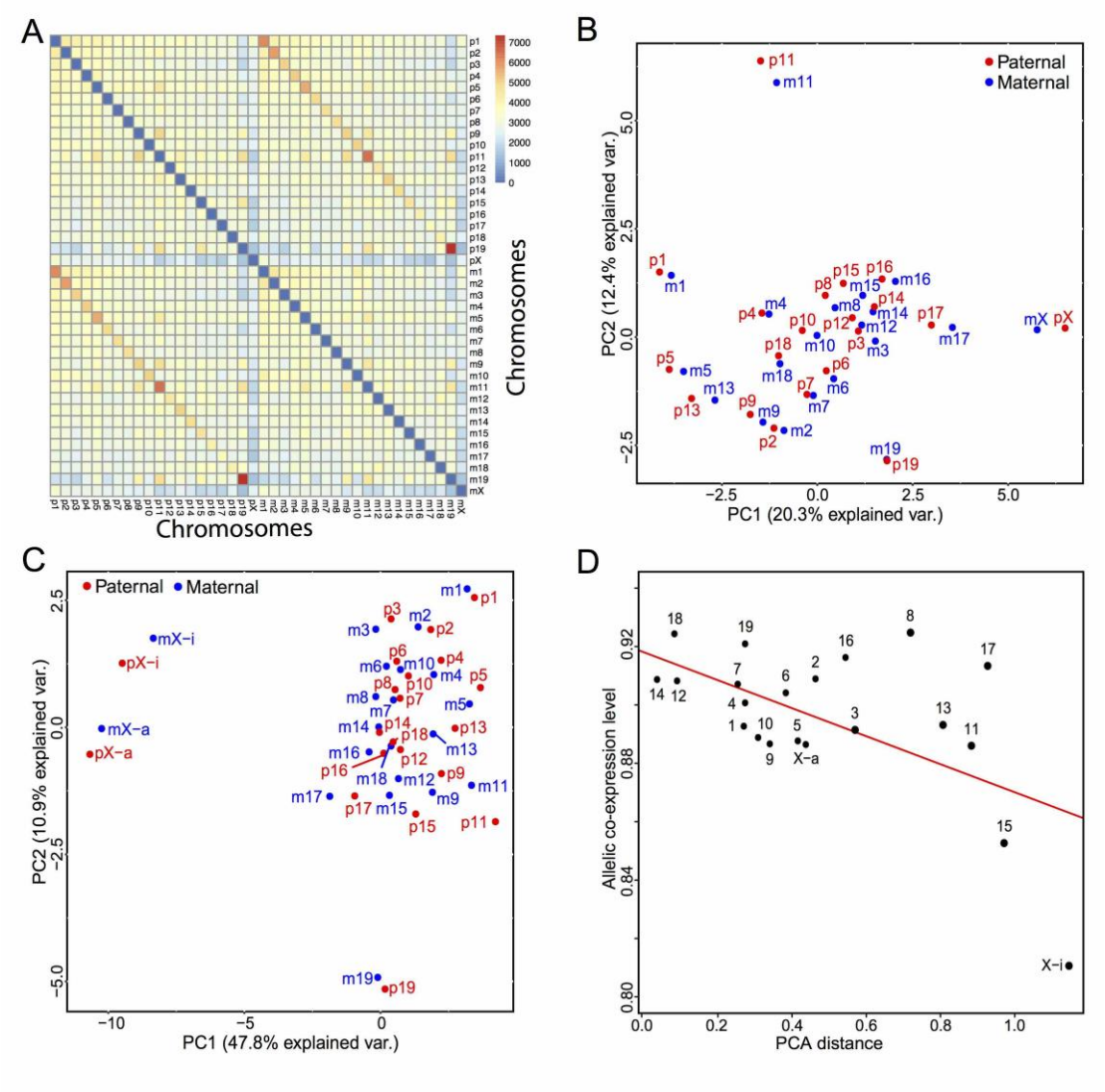

Figure S2. Homologous chromosomes show similar interaction patterns. p indicates paternal chromosome; m indicates maternal chromosome.

(A) Chromosome-level interaction heatmap of the 38 autosomes and 2 X Chromosomes in hybrid mice. (B) PCA analysis of chromosomal interactions by excluding PETs between homologous chromosomes. (C) PCA analysis of chromosomal interactions with X Chromosome being separated into X-a and X-i, partly because X-a is similar to autosomes while X-i is not. (D) Allelic co-expression between homologous chromosomes is correlated with the similarity of chromosomal interaction pattern, with X Chromosome being separated into X-a and X-I (R = 0.66, P value = 0.001).

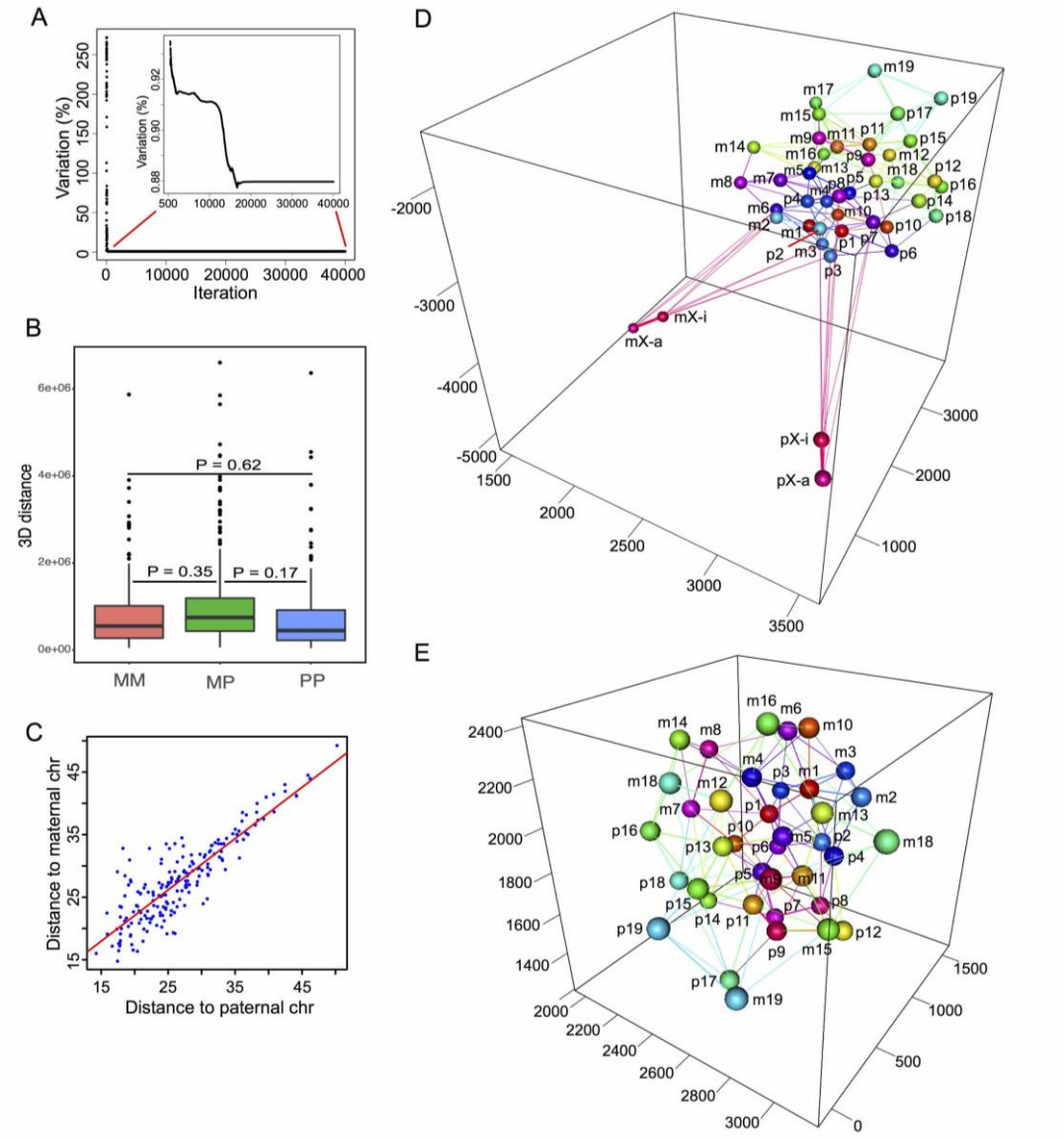

Figure S3. Constructing 3D nucleus of hybrid mice and identifying principles underlying 3D nucleus organization.

(A) Variation of the model rapidly decreased and reached steady-state in iteratively adjusting model. (B) Boxplot of 3D distances between MM, MP and PP in constructed 3D nucleus. The 3D distances between any pair of MM, MP and PP is not significantly different (*t*-test). (C) The distances of a chromosome to homologous chromosomes are highly correlated in 3D model. (R = 0.87, P value < 2.2×10^{-16}). (D) 3D nucleus with X Chromosome being separated into X-a and X-i. (E) 3D nucleus without X Chromosomes.

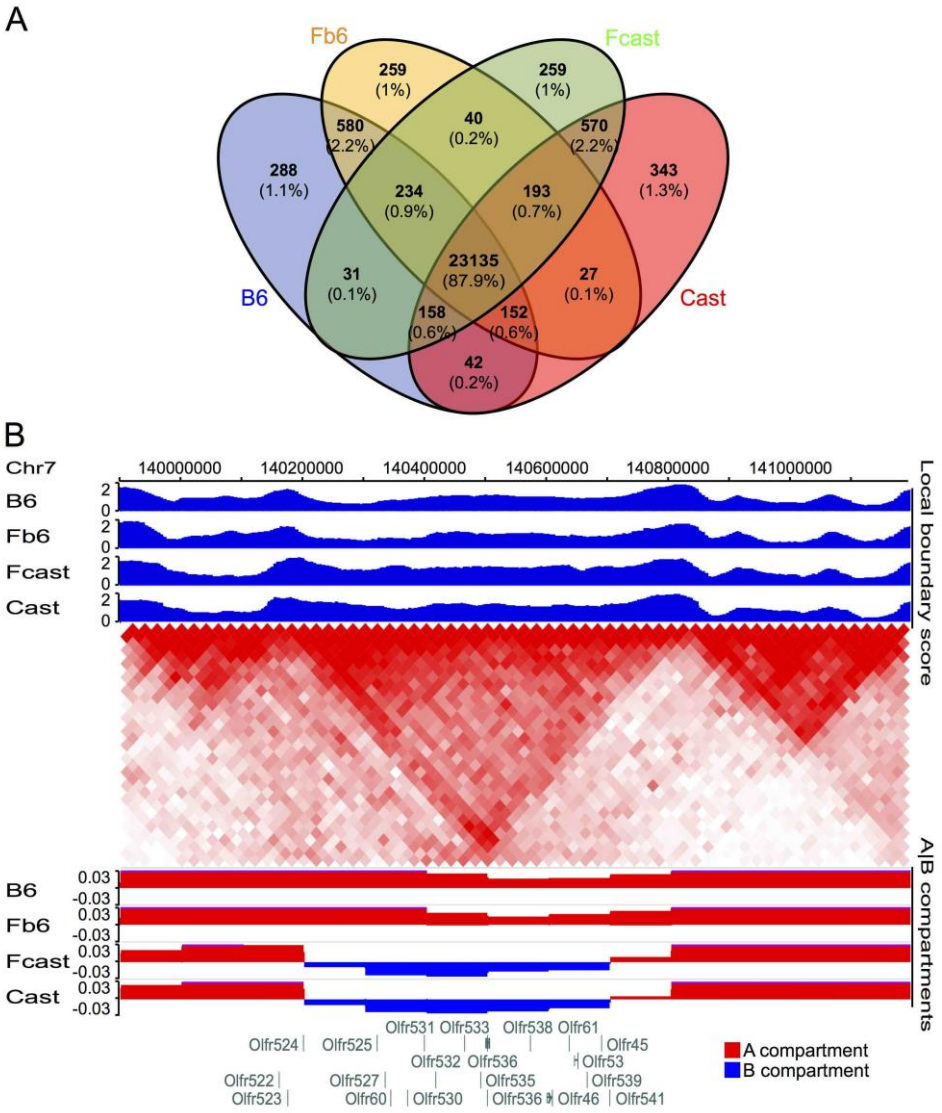

Figure S4. A|B compartment status among B6, Cast and two haploids in hybrid mice.

(A) Venn diagram showing significant overlaps of A|B compartment among parents and two haploids of hybrid. (B) The olfactory genes are located within a single TAD, with divergent A|B compartment between B6/Fb6 and Cast/Fcast. Resolution: LBS (5kb); Heatmap (20kb); compartment (100kb).

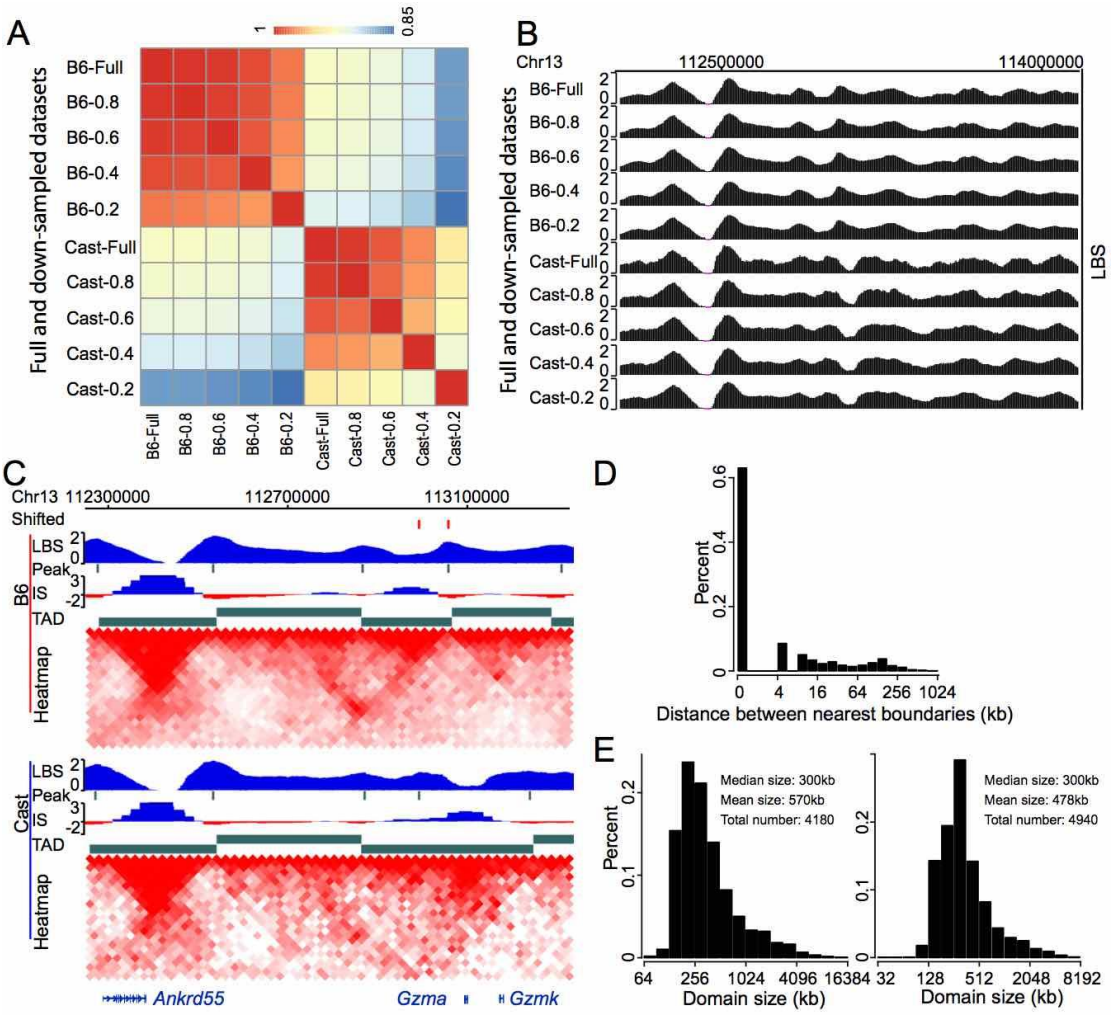

Figure S5. Robust of LBSs and TADs called by LBSs. Resolution: LBS (5kb); IS (5kb); Heatmap (20kb).

(A) The Pearson correlation of LBSs between full dataset and down-sampled datasets. (B) LBS patterns of down-sampled datasets are similar to that of their full dataset. (C) TAD boundaries called by LBS and HiC explorer insulation score (IS) are highly consistent. Abbreviation: Shifted, different TAD boundary between B6 and Cast. (D) Histogram of distance between TAD boundaries called by LBS and these nearest called by HiCExplorer. (E) Distribution of TADs size called by LBS (left) and HiCExplorer (right).

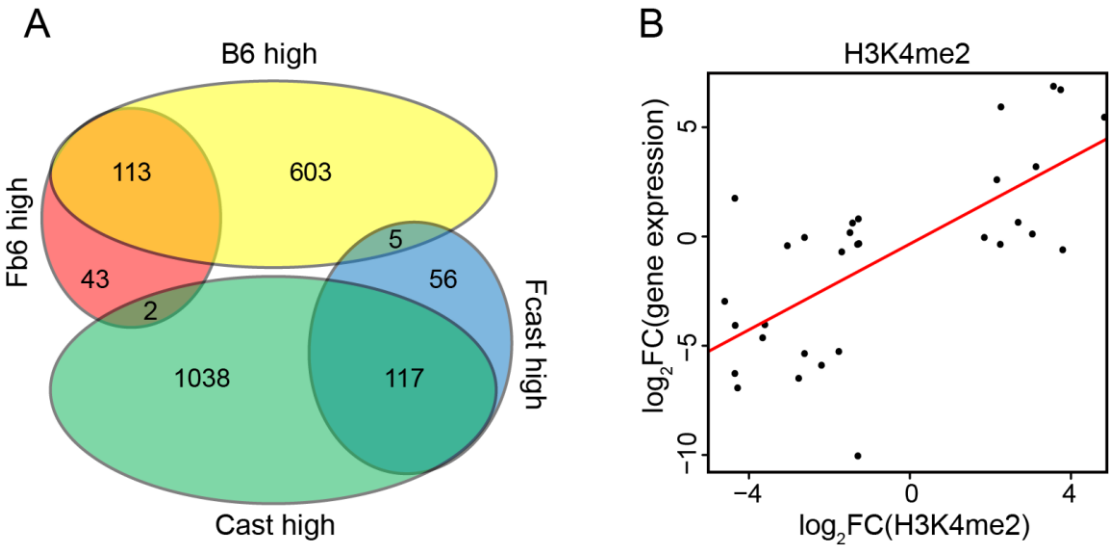

Figure S6. Allelic specific expressed genes and its association with allelic epigenetic modification.

(A) Venn diagram showing the overlap of DEGs between B6/Cast and Fb6/Fcast. (B) Allele specific H3K4me2 is positively correlated with allele-specific gene expression in hybrid mouse. Each point represents a biased ChIP-seq peak and its regulated gene ($R = 0.82$, P value $< 2.2\times10^{-16}$).