Supplementary text

Library saturation analysis

All the libraries were remained unsaturated for the combination of restriction fragments (Figure S3B). Given the data amount we have generated for *Drop-t* has already exceed the common practice for Hi-C, this unsaturation may imply the existence of rich information about the high order chromatin structure were missed by Hi-C.

Droplet statistics

In our dataset, there are about 3.9 million droplets, which can be roughly classified into two categories, one with very few restriction fragments loaded (average 4.46), and the other has substantial more restriction fragments (average 566) (Figure S1C). Although the former type contains about 62% of total droplets, less than 1.8% of complex were originated from those droplets (see below). In that regard, we did not distinguish the two type of droplet in the entire work.

Supplementary Methods

3C library construction. About 5 million K562 or GM12878 cells or *Drosophila* S2 cells were cultured in logarithmic phase, and crosslinked by formaldehyde (Sigma, final concentration is 1%) at room temperature for 10min. Moderate 1.25M glycine were added to quench formaldehyde (Sigma, final concentration is 0.125M) at room temperature for 5 min and then transferred to the ice for another 10 min. Cell pellets were washed by precooled 1× PBS once, then suspended in lysis buffer (10mM Tris-HCl pH8.0, 10mM NaCl, 0.2% NP-40) with proteinase inhibitor cocktail(Sigma, 100×), placed on ice for 1 h and mix it several times. Cell pellets were washed by lysis buffer once. After incubating the cell pellets with 50μl 0.5% SDS in 62°C for 10min, add 170μl mixture with 25μl 10% Triton X-100 and 145μl H_2O to quench SDS and incubated at 37°C for 15min with 1000 rpm. Cell pallets were digested by 6μl MboI (250,000 U/ml, NEB) at 37°C for 20h, and 65°C for 20min to inactive the enzyme, then ligation mix (120μl 10× T4 ligase buffer, 100μl 10% Triton X-100, 12μl BSA(10mg/ml),10μl 100mM ATP, 4000U T4 ligase, 698μl H_2O) were added and incubated at 16°C for 16h.

RNA was removed by 100mg RNase A (Thermo Fisher Scientific) at 37°C for

45min. 2500*g* 5min to collect the pallets, the nuclei was resuspended by 50μl proteinase K (NEB, 200mg/ml), 400μl 10mM Tris-HCl pH8.0 and 50μl 10% SDS. After incubating at 55°C for 30min, 55μl 5M NaCl was added to reverse cross-link at 68°C with 1000rpm. Then, Add 550μl phenol to the tube and vortex for 2 min. then spin for 5min at 10000*g*, Transfer the supernatant to a new 1.5ml tube and add same volume of phenol-chloroform-isopentanol (25:24:1, v/v/v), vortex for 1min, then spin for 5min at 10000*g*. Pipette supernatant to a new 2ml tube, add 4μl acryl carrier (Biotech), 1/10 volume of 3M sodium acetate, vortex and add 2.5 fold volume of precooling alcohol, place the tube at -80°C for 1h. centrifuge the tube for 20min at 4°C with maximum speed (about 14000*g*). Discard the supernatant and let the residual dry at air for 5min. Dissolve the precipitate in 70μl water. 3C DNA was purified by 1.8× AMPure XP beads and diluted in 150μl H_2O.

Human-Fly DNA Mixture for Drop-t Library Construction. A total of 1 ng of mixed high-molecular-weight 3C DNA, with a 1:4 mass ratio of human to Drosophila DNA, was used for Drop-*t* library preparation. The experimental procedure followed the established Drop-*t* library construction workflow. The final library was sequenced on the NovaSeq 6000 using 150 bp paired-end reads.

Algorithm: Mont Carlo Sampling for 3C Ligation Simulation

Initial V ← {restriction fragments, F ← V.

W ← fragments contact matrix from bulk Hi-C

Calculate probability matrix ***P*** according to Hi-C contact frequency matrix (Rao et al. 2014), $p_{ii}=0$, $p_{ij}=\frac{w_{ij}}{\sum_{v_k \in N_i} w_{ik}} (i \neq j)$ where w_{ij} is the contact frequency of v_i and v_j

while (True) do

Get free end fragments V from F

$V' \leftarrow \{\}, S \leftarrow \{\}, p_0 = 0,$

while (| V′ |<| V |) **do**

Select fragment v_i randomly from $V \setminus V'$,

$V' \leftarrow V' \cup \{v_i\}$,

calculate probability vector $\mathbf{p_i} = (\mathbf{p_f}, p_{iu})$ from matrix $\boldsymbol{P}$, where $\mathbf{p_f}$ is the probability vector of v_i and each free end fragments, p_{iu} is the sum of probability of v_i and all non-free fragments.

if $(\mathbf{p_f} = \mathbf{0})$ **then**

$p_0 \leftarrow p_0 + 1$

continue

else do

select one of fragment v_j from $V \cup \{\text{'non} - \text{free'}\}$ with probability $\mathbf{p_i}$

if (v_j is 'non − free') **then**

continue

elif (v_j not in V') **then**

$V' \leftarrow V' \cup \{v_j\}$

$S \leftarrow S \cup \{(v_i, v_j)\}$

else do

continue

end while

if ($p_0 = |V|$) **then**

break

update F according to the ligations in S.

end while

K562 Pore-C sequencing data analysis

K562 Pore-C data analysis was performed following the methodology outlined in [1]. The reads were then partitioned based on the MboI restriction sites (GATC) and mapped to the hg19 human reference genome using BWA-MEM. Reads with a mapping quality (MAPQ) of ≤30 were discarded. For reads mapped to multiple locations, the location with the highest frequency was selected.

Sequencing saturation evaluation

Sequencing saturation was assessed using the Preseq [2] Library complexity was evaluated with the 'c_curve' command, and library size was estimated using 'lc_extrap' with the '-Q' parameter.

Converting high-order chromatin conformation data into pairwise contacts.

Hi-C, C-walk, SPRITE, Pore-C, and HiPore-C were benchmarked (Table S1). For each technique, HCCs interactions were extracted into fragment pairs using the 'juicer_tools pre' command and converted into '.hic' files. Cooler files are created for Drop-t, Pore-

C, HiPoreC and C-walks data using hicConverFormat[3]. Pairwise contact maps across multiple resolutions were generated (5kb, 10kb, 25kb, 50kb, 100kb, 250kb, 500kb, 1Mb and 2.5Mb).

Pairwise contact map analysis

The similarity of pairwise contact maps was assessed using stratum-adjusted correlation coefficient(SCC) calculated by HiCRep[4]. The similarity of p(s) curves were assessed using Jensen-Shannon distances.

The 3D structure of genome was most analyzed by juicer tools [5] and visualized by Juicebox[6]. A/B compartment value was calculated using "juicer_tools eigenvector" command, and the pearson's correlation map was created using "juicer_tools pearsons" command for each chromosome at 500kb and 1Mb resolution. Insulation score and TAD boundary was obtained using "cooltools insulation" at resolutions 5000, 10000, 25000, 50000, 100000 respectively. TAD boundaries are obtained using from the "is_boundary_" column from the output of "cooltools insulation". Treating TADs as partitions of the genome, weighed similarity (WS) between two partitions was calculated to evaluate the similarity of TADs[7]. Aggerated peak analysis (APA) was performed by "juicer_tools apa".

The pseudo code for algorithm: Backtracking d-LHCC from Drop-t sequencing data

Input: droplet set $[\mathrm{bc}_1, \mathrm{bc}_2, \ldots \mathrm{bc}_N]$, where bc_i = {DNA restriction fragments with barcode i}.
Output: d-LHCC for all droplets.

Let x, y, and z be parameters;
Using reads from all droplets create the global graph $G_f = (V_f, E_f)$, where node set V_f represent restriction fragments and edge set E_f represent ligations. The weight of edge represents the ligation frequency;
delete 4000 nodes with the largest degree in G_f;
for i **in** $1..N$
    Create local graph $G_c = (V_c, E_c)$ using reads in bc_i;
    Create $G_i = (V_i, E_i)$, where $V_i = V_c$ and E_i contains all the edges in E_f that connects nodes in V_i;
    Find all connected components in G_i and ascendingly sort by size $\mathrm{C} \leftarrow [\mathrm{c}_1, \mathrm{c}_2, \ldots \mathrm{c}_m]$;
    for i **in** $1..m$
        $\mathrm{c_i}'$ = {all fragments within genomic flanking *x* fragments of any node in $\mathrm{c_i}$};
        n_{i} = {all neighbors in $\mathrm{G_f}$ for all nodes in $\mathrm{c_i}'$};
    repeat
    for i **in** $1..m$
        if $|\mathrm{c_i}| \geq \mathrm{z}$

break

$C_{overlap} \leftarrow []$;

$N_{overlap} \leftarrow []$;

for c_j **in** $\{c_{i+1}, \ldots, c_m\}$

if $(n_i \cap n_j) = \emptyset$

continue

else

Append c_j to $C_{overlap}$;

Append $n_i \cap n_j$ to $N_{overlap}$;

End if

Repeat

$C_{overlap} \leftarrow \{c_j$ with the most shared neighbors in $C_{overlap}\}$;

$C_{overlap} \leftarrow \{c_j$ of the smallest size in $C_{overlap}\}$;

$W \leftarrow \{w_1, \ldots, w_k\}$, where w_j is the sum of weight of edges from any node in $n_i \cap n_j$ to any node in c_i or c_j;

Random pick c_j in $C_{overlap}$ with probability proportional to w_j;

$c_{ij} \leftarrow \{c_i\} \cup \{c_j\}$;

if $|c_{ij}| < z$ **and** $(|c_i| < y$ **or** $|c_j| < y)$

$C \leftarrow C \setminus \{c_i, c_j\}$;

$n_{ij} \leftarrow n_i \cup n_j$;

Add c_{ij} to C , n_{ij} to N while maintaining the order of C;

end if

repeat

repeat

d-LHCC is defined as the final connected components in C;

We took *x = 5, y = 8,* and *z= 50* for K562 dataset and *x = 4, y = 7,* and *z= 40,* for GM12878 dataset in this work as they yielded modest false connection rate(~5%) and relatively sufficient connection (Figure S3).

Single-cell transcriptome analysis

Sequencing data were processed using the 10X scRNA-seq tool cell-ranger with "count" command and default parameters[8]. The expression count matrix was normalized to counts per million (CPM) using python package AnnData[9]. Gene expression level was defined as the mean value of normalized expression across all cells. Gene expression noise was quantified as the coefficient of variation (CV) of gene expression, excluding genes expressed in fewer than five cells.

Leaders and dictator analysis

Leader genes with highest POH is defined as all genes int the class with highest POH(top 20%) in all groups. Dictator genes with highest EPIE is defined as all genes int the class with highest EPIE(top 20%) in all groups.

Motif enrichment was performed using AME[10] in meme suite[11] with default parameters. Sequences of all promoters (enhancers) were selected as background when analyzing promoters (enhancers). The number of motifs was performed using fimo[12] in meme suite[11] with default parameters. GO analysis was performed using metascape[13] including GO terms of "molecular function", "biological process" and "cellular components".

Supplementary Figure Legends

Figure S1. Monte carlo simulation and AFM imaging showing DNA entanglement.

(A) Agarose gel electrophoresis of 3C ligation experiment. Gels showing the 3C DNA length distribution after HindIII digestion (lane 1), ligation (lane 2) and fragment selected by BluePippin DNA size selection system (lane 3). (B) Comparison of pairwise interaction maps generated from Drop-t (left) and monte carlo simulated 3C (right). Chromosome 11 at 500 kb resolution was shown as an example. (C) Length distribution of the simulated 3C ligation products. The histogram (left panel) and cumulative frequency curve (right panel) were shown. (D) AFM images of 3C heavy products with DNA concentration of 1 ng/μl (left) and 0.5 ng/μl (right). The colors represent relative height. (E) AFM images of unligated DNA after digestion with concentration of 1 ng/μl. (F) AFM images of linear plasmid DNA with concentration of 1 ng/μl.

Figure S2. Drop-t accurately recapitulates fine-scale pairwise chromosomal topologies of Hi-C. (A) The pairwise inter-chromosomal interaction maps of K562 in situ Hi-C, Drop-t, Pore-C and C-walks showing the similarity between different libraries. The color indicates the number of total contacts between the corresponding

chromosomes in log scale. (B) Contact frequency decay curves of intra-chromosomal pairwise interactions for GM12878 Drop-t, HiPore-C, Pore-C and in situ Hi-C. The x-axis represents genomic distance and the y-axis represents the normalized contact frequency. (C) Similarity of chromatin compartmentalization for in situ Hi-C vs Drop-t, Pore-C, C-walks libraries in K562. Chr7 was shown as an example with 500kb resolution. The pearson correlation matrices of the interaction maps together with the first eigenvector profiles were shown. The lower diagonal shows Hi-C maps and the upper diagonal shows Drop-t, HiPore-C and C-walks pairwise contacts from left to right. (D) Genome-wide correlation between Drop-t and Hi-C insulation scores calculated using contact map at 10kb resolution. The colors indicate the frequency of the entry. (E) Comparison of insulation score profile between Drop-t or Hi-C for the 100-105Mb on chromosome 1 calculated using contact map at 10kb resolution.

Figure S3. Design and parameter determination for the graph-based algorithm. (A) Saturation curves showing the number of detected distinct reads (y-axis) as a function of sequencing depth (x-axis) for both Pore-C and Drop-t. The current sequencing depth in our study is indicated by arrows of corresponding color. The dotted line represents the predicted trend of sequencing saturation. (B) Distribution of number of fragments in each droplet. The x-axis represents the number of restriction fragments in the droplet in log scale and the y-axis represents the number of droplets. (C)(D) Deciding parameters for d-LHCC backtracking algorithm of K562(C) and GM12878(D). The probability of having shared neighbors is calculated for connected

component pairs from different droplets of different size combinations. 1000 component pairs were tested for each size combination. The y-axis represents the probability of not having shared neighbors. The x is the parameter explained in the "Deciding parameters" section in Methods. (E) Size distribution of GM12878 d-LHCC (drop-t) and hp-LHCC (Hi-PoreC). The x-axis represents DNA mass in HCC and y-axis represents the normalized frequency.

Figure S4. Human-fly hybrid data reveal low false ligation rate introduced by HCC assembly algorithm. (A) The frequency of human and fly connected components obtaining from the c-graph as a function of DNA mass. (B) Proportion of human-fly hybrid d-LHCCs with different parameter combinations. (C) Parameter estimation similar to Figure S3(C) and (D). The y-axis represents the proportion of components pairs having no shared neighbors. (D) Distribution of human fragment proportion in the d-LHCCs containing fly fragments. The x-axis represents the human fragment proportion and the y-axis represents the d-LHCC frequency.

Figure S5. Key distinguishes between d-LHCC and s-HCC. (A) Distribution of restriction fragment number in ultra-large (>=10) d-LHCC and s-HCC. The horizontal axis represents the restriction fragment number in HCC in log scale and the vertical axis represents the frequency. (B) Pie-chart showing the proportion of d-LHCCs and s-HCCs spanning different numbers of chromosomes. (C) The difference between gap

length distribution of intrachromosomal d-LHCCs and s-HCCs for different size on a chromosome. The color indicates the difference of frequency in Figure 4E. Dash lines indicate the threshold for proximal, middle and distal gaps. (D) Change of number of sub-clusters with the increase of HCC size for intrachromosomal d-LHCCs (left) and s-HCCs (right). Different colors represent different number of sub-clusters and the height of the bars represents corresponding proportion. (E) Similar to (D), except for different colors representing sub-cluster size. (F) Proportion of gene, typical-enhancer (TE) and super-enhancer (SE) fragment for d-LHCC (d) and s-HCCs (s) with different size.

Figure S6. Parameters range of groups related Figure 5 and Figure 6. (A) POT, number of EPI(#EPI) range of groups in Figure 5(C) and (E). (B) POT, POC, #EPI range of groups in Figure 5(D) and (F). (C) POT, POC, #EPI, EPIE, POH range of groups in Figure 6(A), (B) and (E). (D) POT, POC, #EPI, EPIE, POH range of groups in Figure 6(C), (D) and (F).

Figure S7. Competition and cooperation between EPIs associate with the strength and stability of gene expression. (A) Enrichment of housekeeping genes in the corresponding sub-group of Figure 5(C) and (E). The colors represent corresponding different sub-groups and the height of the bar represent -log10(p-value) of hypergeometric distribution test operated for each group. (B) Mixed trends (similar number of red and blue arcs) between POH and noise for dictators. (C) Mixed trends

between EPIE and expression for leaders. (D) Motifs enriched for enhancers regulating dictators with highest EPIE (left) and leaders with highest POH(right).

References

1. Tavares-Cadete, F., et al., *Multi-contact 3C reveals that the human genome during interphase is largely not entangled.* Nat Struct Mol Biol, 2020. **27**(12): p. 1105-1114.
2. Timothy Daley and Andrew D Smith, *Predicting the molecular complexity of sequencing libraries.* Nat Methods, 2013(10(4)): p. 25-27.
3. Ramirez, F., et al., *High-resolution TADs reveal DNA sequences underlying genome organization in flies.* Nat Commun, 2018. **9**(1): p. 189.
4. Yang, T., et al., *HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient.* Genome Res, 2017. **27**(11): p. 1939-1949.
5. Durand, N.C., et al., *Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments.* Cell Syst, 2016. **3**(1): p. 95-8.
6. Durand, N.C., et al., *Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom.* Cell Syst, 2016. **3**(1): p. 99-101.
7. Li, A., et al., *Decoding topologically associating domains with ultra-low resolution Hi-C data by graph structural entropy.* Nat Commun, 2018. **9**(1): p. 3265.
8. Zheng, G.X., et al., *Massively parallel digital transcriptional profiling of single cells.* Nat Commun, 2017. **8**: p. 14049.
9. Virshup, I., et al., *anndata: Annotated data.* JOSS, 2021.
10. McLeay, R.C. and T.L. Bailey, *Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data.* BMC Bioinformatics, 2010. **11**(1): p. 165.
11. Bailey, T.L., et al., *The MEME Suite.* Nucleic Acids Research, 2015. **43**(W1): p. W39-W49.
12. Grant, C.E., T.L. Bailey, and W.S. Noble, *FIMO: scanning for occurrences of a given motif.* Bioinformatics, 2011. **27**(7): p. 1017-1018.
13. Zhou, Y., et al., *Metascape provides a biologist-oriented resource for the analysis of systems-level datasets.* Nature Communications, 2019. **10**(1): p. 1523.