Empirical Bayes spline model learns multi-way genomic interactions from single cell 3D genome data

Kwangmoon Park[1] and Sündüz Keleş[1,2]

[1] Department of Statistics, University of Wisconsin, Madison, WI, USA
[2] Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI, USA

Abstract. A number of foundational analysis methods have emerged for single cell chromatin conformation (scHi-C) datasets capturing 3D organizations of genomes at the single cell resolution; however, these scHi-C datasets are currently under-utilized. The canonical uses of the existing scHi-C data encompass, beyond standard cell type identification through clustering and trajectory analysis, inference of chromosomal structures such as topologically associated domains, A/B compartments, and pairwise interactions. However, multi-way interactions, e.g., looping among multiple genomic elements such as multiple enhancers of a gene, are entirely overlooked. We introduce **ELECT**, an empirical Bayes modelling framework to **E**xtract mu**L**ti-way g**E**nomi**C** in**T**eractions by leveraging scHi-C data. ELECT builds on a dirichlet-multinomial spline model, incorporates well-known genomic distance bias of the chromatin conformation capture data, and yields multi-way interaction scores by leveraging corresponding pairwise interactions across cells of the same type. The multinomial-poisson transformation enables parameter estimation and inference for ELECT in a computationally feasible way for both low and high resolution single cell chromatin conformation data. ELECT yields well-calibrated p-values for controlling the false discovery rate and inferring multi-way interactions. We applied ELECT to both low and high resolution scHi-C datasets and carried out evaluations with external genomic and epigenomic data including data from DNA methylation, SPIRITE, scNanoHi-C, and DNA seqFISH+ assays. Application of ELECT to scHi-C data from human prefrontal cortex revealed multi-way interactions that involved GWAS SNPs associated with psychiatric disorders including autism and major depressive disorder, suggesting ELECT's potential for interrogating genomewide association studies for epistasis. ELECT is publicly available at https://github.com/keleslab/elect.

Keywords: Single cell chromatin conformation capture · Multi-way genomic interaction · Empirical Bayes · Spline model.

1 Introduction

Advances in high-throughput mapping of nuclear genome organization in a single nuclei by capturing the interactions between proximal genomic loci have enabled genome-wide characterization of 3D genome features, including loops, topologically associating domains, A/B compartments, and sub-compartments at the single nuclei level. However, these datasets have not yet been scrutinized for delineating multiple loci that interact simultaneously in the same nucleus (i.e., multi-way interactions). Identifying such interactions is critical for assessing the epistatic potential of risk SNPs from genomewide association studies (GWAS). Genes linked to diseases are often connected to multiple distant enhancers that span a considerable genomic range. A canonical example is oncogene MYC, epistatic enhancer interactions of which were recently studied by multiplexed CRISPRi screening in detail[1]. Genome-wide association studies have shown that more than 90% of disease-associated variants are noncoding variants, and can be dispersed over large distances. While single enhancer variants may pose relatively minor clinical risks, a combination of multiple variants can significantly intensify their impact on traits and diseases.

A number of experimental innovations emerged to generate multi-way chromatin contacts: GAM[2], ChIA-drop[3], SPRITE[4, 5], Tri-C55, multi-contact 4C[6], COLA[7], and Pore-C[8]. While these approaches provided proof-of-concept with applications to cell lines and mESCs, sn-m3C-seq (scHi-C) data collection emerging from the Brain Initiative presents an exciting opportunity to identify multi-way chromatin interactions in brain cell (sub)types. Thus far, all scHi-C data analysis focused on pairwise interactions and deployed, with the exception of snapHi-C [9, 10], pairwise interaction inference methods developed for *bulk* Hi-C data. This is a vast under-exploitation of scHi-C data in terms of elucidating how multiple enhancers interact with each other in regulating their target genes over long genomic distances to confer disease risks.

While single cell Hi-C methods are traditionally considered as limited since they measure pairwise interactions by proximity ligation, recently developed proximity ligation method scNanoHi-C[11] that leveraged long-read sequencing to explore high-order chromatin contacts within individual cells successfully captured multi-way interactions. Furthermore, direct comparisons of scNanoHi-C results with the SPRITE results [5] revealed that the latter contained most reads from indirect and trans-chromosomal ultra-high order interactions which may be mediated by subnuclear compartments. In addition, a recent study[12] found that the effect of long-distance, multi-way interactions as quantified by SPRITE on reconstructing the 3D genome is minimal. This is largely attributed to the fact that the quantity of these interactions is overshadowed by the much larger number of pairwise contacts, which potentially carry critical information for inferring multi-way interactions.

The problem of identifying multi-way interactions from scHi-C data can be recognized as a clique (i.e., a subset of nodes in a network where each node is connected to every other node in the subset) finding problem in a network where the nodes are genomic loci (e.g., bins/genomic-intervals created for quantifying scHi-C data, where the bin sizes vary from 10Kb (kilo base pair) to 1Mb (million base pair) depending on the sequencing depths of the cells)[13]. The edges between the nodes quantify the observed interaction frequency (e.g., by 3D chromatin methods) between the genomic loci. Then, the key inference problem can be formulated as assessing the significance of the cliques that represent candidate multi-way chromatin interactions. Despite the ligation limitations and data sparsity, we observed that scHi-C data yields large numbers of cliques (**Fig.** S1). This supports the notion that while scHi-C might be limited in capturing very higher-order multi-way interactions, it holds potentially critical information on lower-order (e.g., empirically ≥ 3 and ≤ 6) multi-way interactions. To capitalize on this observation, we developed ELECT as an empirical Bayes framework that evaluates the significance of cliques by aggregating pairwise interaction data within a multi-way interaction across multiple cells of the same type. We show that analytical manipulations with appropriate Dirichlet priors to a multinomial distribution lend itself to a computationally feasible generalized linear model with splines. Applications of ELECT to multiple datasets indicate that ELECT yields well-calibrated p-values for controlling false discovery rate. Evaluations of ELECT with external multi-way interaction data and data from methylation assays support the multi-way interactions identified by ELECT. More specifically, multi-way interactions between genes and multiple enhancers are also well supported by ABC scores representing the contributions of individual enhancers to the regulation of genes. Furthermore, when contrasting ELECT with a baseline approach based on state-of-the-art loop calling from scHi-C data, ELECT notably detects a greater number of multi-way interactions from human prefrontal cortex data, specifically involving gene and multiple enhancers listed as linked in the PsychENCODE gene-enhancer resource. Interrogation of GWAS association

SNPs of brain disorders with ELECT inferred multi-way interactions for brain cell types reveal higher-order interactions that involve genes and their multiple validated enhancers harboring GWAS risk SNPs.

2 Methods

2.1 ELECT model representation

ELECT, presented in **Fig.** 1, is an empirical Bayes model for inferring multi-way interactions from scHi-C contact matrices. It leverages pairwise contacts quantified in contact matrices to construct a test statistic for each candidate multi-way interaction, i.e., clique of bins/genomic-intervals), and evaluates these against a generative null distribution for contact matrices. Specifically, we consider cell k, $k \in [K]$, and its symmetric (*cis*-interaction) contact matrices (maps) with dimensions $l_{chr} \times l_{chr}$ for each chromosome $chr \in [Chr]$, where l_{chr} denotes the number of loci for chromosome chr and the bracket notation $[K]$ denotes a set of sequential numbers $\{1, \cdots, K\}$ (**Fig.** 1a top row). Each entry (edge) of a contact matrix (graph) quantifies the level of contact (edge weight) between two genomic loci (nodes), e.g., $n_1 \in [l_{chr}], n_2 \in [l_{chr}]$. None-zero pairwise contacts among more than two loci, e.g., a size 3 clique of loci n_1, n_2, n_3 (starred entries in **Fig.** 1a top row), provide evidence for a potential multi-way interaction among the loci n_1, n_2, n_3 (triangles in **Fig.** 1a bottom row) in a single cell. Under a noise free setting, all pairwise interactions stemming from a multi-way interaction would be observed and generate a clique. However, owing to sparsity levels of contact matrices (different edge widths in each triangle in **Fig.** 1a bottom row) and potential proximity ligation limitations, all pairwise interactions may not be observable within an individual cell. To address this issue for detecting multi-way interactions from contact matrices, we view each contact matrix of a cell as an independent sample of the true unknown contact matrix of a context, e.g., cell type, pseudo-time, or treatment. Then, we compute a clique score for each candidate clique $q \in [Q]$, by summarizing the pairwise interaction levels of the clique throughout the cells (**Fig.** 1e.). This score aggregates the evidence for the cliques across the cells of the same type by leveraging all pairwise interactions.

To assess the significance of the cliques, we evaluate whether the clique scores are higher than expected under a contact matrix generating model which considers counts in each upper triangular entry of the symmetric contact matrix of a cell k as a random draw from a multinomial distribution with the sequencing depth C_k of the cell as the number of trials. In order to take into account the well-known genomic distance bias, i.e., band bias indicating that contact counts between two loci systematically vary with the genomic distance between them, observed in both the bulk and single cell Hi-C data [15–17], we model the parameters of the multinomial distribution with a natural cubic spline (**Fig.** 1c). Specifically, we define the set of locus pairs within each off-diagonal of the contact matrix as a "band" (blue, green, and gold bars in **Fig.** 1a top panel), where the locus pairs $i \in [N_j]$ in the same band $j \in [J]$ have the same genomic distance d_j. Then, the contact counts y_{ijk} of locus pairs $i \in [N_j]$ of bands $j \in [J]$ of cell $k \in [K]$ follow a multinomial random vector $Y_k := (\boldsymbol{y}_{1k}^T, \cdots, \boldsymbol{y}_{Jk}^T)^T$ with total count C_k, where $\boldsymbol{y}_{1k} = (y_{ijk}, \cdots, y_{N_j jk})^T$. Without loss of generality, we dropped the chromosome subscript $chr \in [Chr]$ and cell type subscript $t \in [T]$ for multinomial sample Y_k; however, this model applies for each chromosome and for each cell type, separately. Additionally, while the contact counts from different cells can be assumed to be independent, dependency among contact counts of a clique within a single cell is plausible. This aspect is naturally incorporated by the multinomial model.

In addition to the multinomial likelihood that serves to estimate the band-level parameters $p(d_j)$ (green box in **Fig.** 1b), we further impose a dirichlet prior distribution (yellow box in **Fig.** 1b) on the locus-pair level contact probabilities $\tilde{p}_{ij}$. Hyper-parameters α_{ij} of this dirichlet prior are utilized to distribute the band-level contact probability p_j onto locus pairs within the band while taking into account the overall null hypothesis (**Fig.** 1d). This statistical model for the scHi-C contact counts, along with the null distributional assumption introduced in the next section, enables constructing a clique level normalized statistic Z_q, i.e., clique score, for each clique $q \in [Q]$. In addition, it provides the clique-level p-values, which enables false discovery rate (FDR) control (**Fig.** 2b, c).

2.2 ELECT model estimation and null distribution for multi-way interactions

ELECT first fits the contact matrix generative model by a poisson generalized linear model on the total band counts using the multinomial-poisson transformation [18] (**Fig.** 1c). This transformation reduces

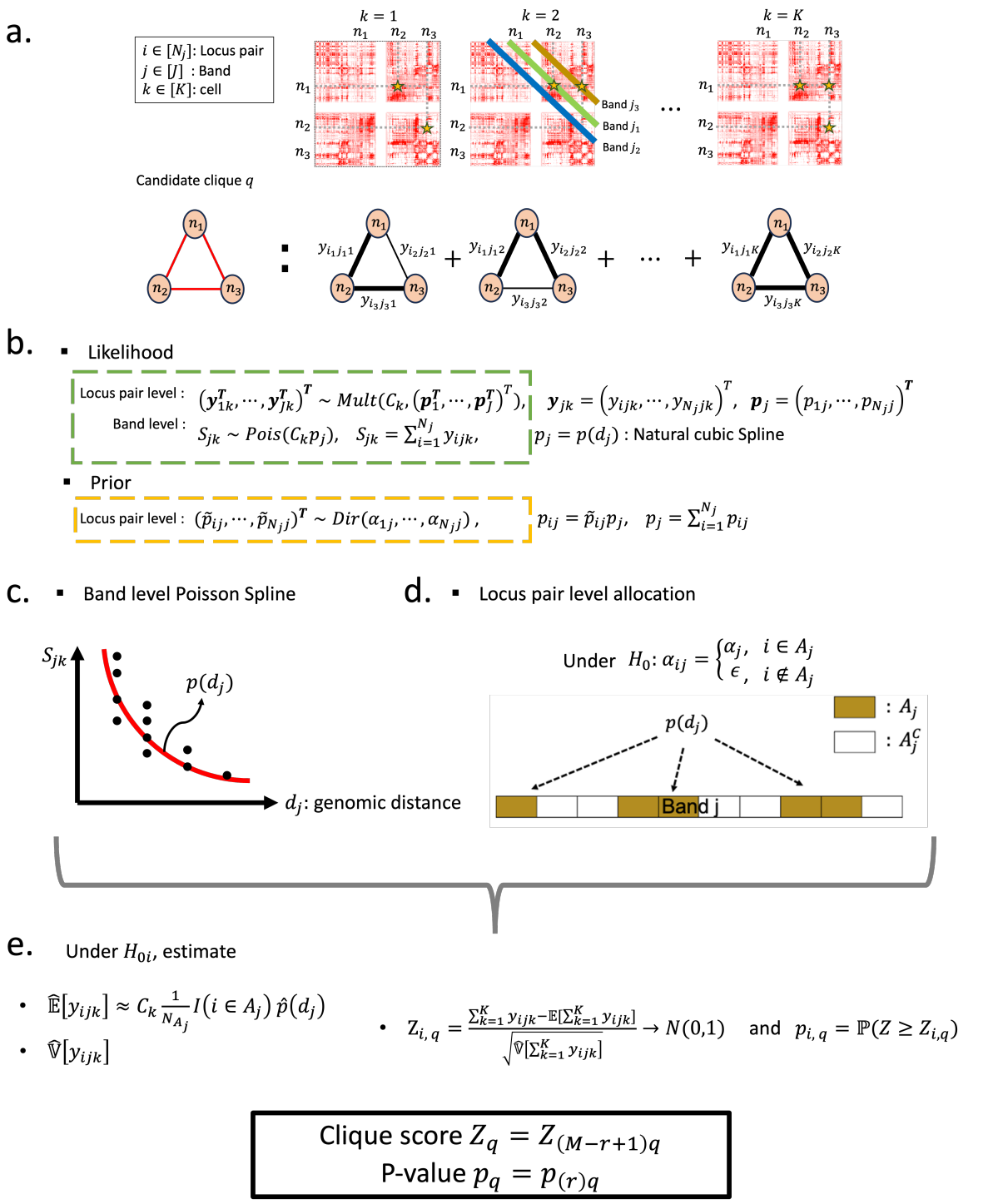
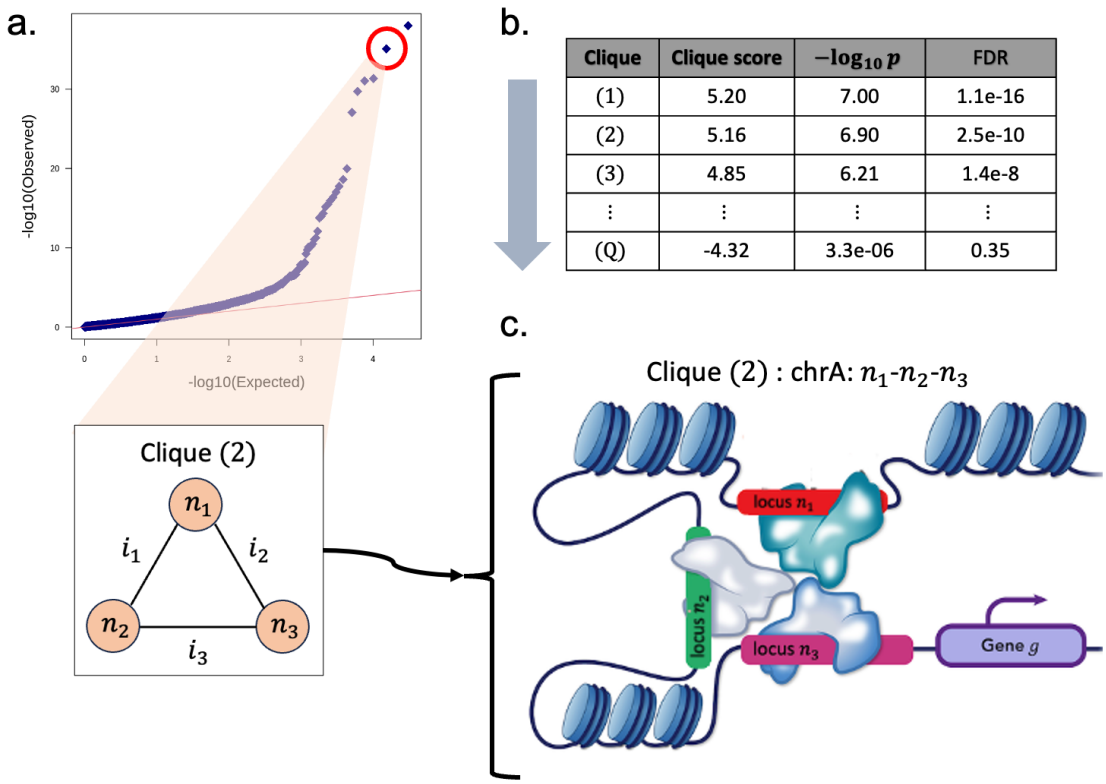


Fig. 1: Overview of ELECT. **a.** Top panel: scHi-C contact matrices of K cells for a single chromosome. The starred entries in each matrix depict the pairwise contacts among three genomic loci (or nodes) n_1, n_2, n_3, which are also displayed as an undirected graph in the bottom panel. The colored (blue, green, gold) bars along the off diagonals denote different bands. For each cell type, ELECT aggregates the pairwise contact counts (graphs corresponding to contact matrices) within a given candidate clique (the left most graph with red edges) across the cells of the cell type, and generates a clique-level statistic. **b.** ELECT adopts an empirical Bayes framework to generate clique-level scores as the standardized version of the clique statistic. Contact counts y_{ijk} of locus pairs, $i \in [N_j]$, across bands, $j \in [J]$, of a cell k follow a multinomial distribution with the cell sequencing depth C_k as the number of trials. **c.** The generative model is further summarized into band-level counts S_{jk}, where the count probabilities depend on genomic distance d_j by a natural cubic spline function . **d.** The dirichlet prior to the multinomial counts with hyper parameters α_{ij} imposes a prior distribution on the contact probabilities *within* a band. **e.** The mean and variance of the locus pair level contact y_{ijk} are estimated with this model under the null hypothesis, and the locus pair level test statistics $Z_{i,q}$ and p-value $p_{i,q}$ are constructed based on these estimates. Finally, ELECT clique scores (Z_q) are constructed by rth largest $Z_{i,q}$ value among M locus pairs within a clique, and the clique level p-values, p_q, are constructed as the rth smallest $p_{i,q}$ value of a clique.



Clique	Clique score	$-\log_{10} p$	FDR
(1)	5.20	7.00	1.1e-16
(2)	5.16	6.90	2.5e-10
(3)	4.85	6.21	1.4e-8
⋮	⋮	⋮	⋮
(Q)	-4.32	3.3e-06	0.35

Fig. 2: ELECT output. **a.** Quantile-Quantile plot (Q-Q plot) of the $-\log_{10}(p)$ values for size 3 cliques of chromosome 2 for cell type L23 from Lee *et al.* [14], with x- and y-axes denoting the theoretical values from $-\log_{10}$ transformed uniformly distributed random variable and empirical distribution of the observed values, respectively. **b.** The cliques are sorted based on their $-\log_{10}(p)$ values (or clique scores) and thresholded at the desired FDR level to identify significant multi-way interactions. **d.** Illustration of a significant clique representing a multi-way interaction among genomic loci n_1, n_2, n_3 harboring two enhancers and a gene promoter.

computational burden significantly (e.g., from a spline GLM fit with the number of experimental units as $\left(\binom{N}{2}\right.$ locus pairs)x(number of cells for the cell type) to only (number of bands)x(number of cells for the cell type) units) (Supp. Section S1.1). Once the band-level estimates of $p(d_j)$ are obtained, ELECT distributes these estimates to locus pairs within the band j under an appropriate null distribution (**Fig.** 1d). For the band-level distribution of contact probabilities with α_{ij}, we first considered $H_{0i} : \alpha_{ij} = \alpha_j$ for each $i \in [N_j]$, which yielded the expected contact counts as $\mathbb{E}[y_{ijk}] = C_k \frac{p(d_j)}{N_j}$. This intuitively indicates that the expected contact count is equal (or uniform) throughout all the locus pairs within a band j. As a result, the cliques harboring locus pairs with significantly larger observed contact counts than expected will yield large clique scores for detection. For a more calibrated selection on the null hypothesis, if we were to know that locus pairs within set A_j of band j have credibly non-zero contact counts, while the others in the band are significantly more likely to have zero counts (e.g., in the case of heterochromatin regions with low mappability and/or high GC, Supp. Section S1.2), we can set the null hypothesis as

$$H_{0i} : \alpha_{ij} = \begin{cases} \alpha_j, & \text{if } i \in A_j \\ \epsilon, & \text{if } i \notin A_j, \end{cases}$$

where $\epsilon > 0$ is a pseudocount with very small value, e.g., $\epsilon = 1e-05$. This results in expected contact counts as $\mathbb{E}[y_{ijk}] = C_k p(d_j) \frac{\alpha_j \mathbf{1}(i \in A_j) + \epsilon \mathbf{1}(i \notin A_j)}{N_{A_j} \alpha_j + N_{A_j^C} \epsilon} \approx C_k \frac{p(d_j)}{N_{A_j}} I(i \in A_j)$, where N_{A_j} denotes the cardinality of the set A_j. Here, the locus pairs outside of A_j will have expected count as zero and the ones in A_j will share the non-zero uniform expected contact counts $C_k p(d_j)$. Under this null hypothesis of a locus pair, based on Maximum Likelihood Estimation, we can estimate required parameters (mean and variance) to construct a Z

test statistics for each locus pair i within a clique q as

$$Z_{i,q} = \frac{\sum_{k=1}^{K} y_{ijk} - \hat{\mathbb{E}}[\sum_{k=1}^{K} y_{ijk}]}{\sqrt{\hat{\mathbb{V}}[\sum_{k=1}^{K} y_{ijk}]}}, \tag{1}$$

where j denotes the band containing locus pair i. Then, the $Z_{i,q}$ asymptotically follows standard normal distribution by the Lindeberg-feller Central Limit Theorem under the null hypothesis. We define the corresponding p-value of each locus pair as $p_{i,q} := 1 - \Phi(Z_{i,q}) = \mathbb{P}(Z \geq Z_{i,q})$, where $\Phi(\cdot)$ denotes standard normal cumulative distribution function. Likewise, under the locus pair null distribution, $p_{i,q} \sim \text{Uniform}(0,1)$.

Finally, based on the set of locus pair level statistics of a (size N) clique q, $\{Z_{i,q} | i \in [M]\}$ (or $\{p_{i,q} | i \in [M]\}$) with $M := \binom{N}{2}$, we construct the clique level summary as the rth largest (smallest) order statistics (p-values) of the clique specific set. Specifically, we define

$$\text{Clique score } Z_q := Z_{(M-r+1)q}$$
$$\text{p-value } p_q := p_{(r)q},$$

where $Z_{(M-r+1)q}$ and $p_{(r)q}$ are defined as the rth largest value of $\{Z_{i,q} | i \in [M]\}$ and rth smallest value of $\{p_{i,q} | i \in [M]\}$. The rationale behind using the rth order statistics is to ensure that a clique is deemed as representing a multi-way interaction when at least r of its locus pairs have significantly large contact levels. With all the locus pair level statistics within a clique q from the null distribution, we can approximate $p_{(r)q} \sim B(r, M-r+1)$, where $B(a,b)$ is Beta distribution with parameters $a, b > 0$. As a result, a p-value for each clique can be obtained from this parametric null distribution to obtain and facilitate FDR control. Further details on the model and the selection of r can be found in Supp. Section S1.1.

For a given dataset, ELECT first obtains the initial set of candidate cliques for evaluation by applying a clique finding algorithm on each individual scHi-C contact matrix. It excludes cliques that are present only in a small number of cells (e.g., less than 3 cells) from evaluation. Further details on the initial set of candidate cliques are provided in Supp. Section S2.

3 Results

3.1 Application and validation of ELECT with both low and high resolution scHi-C data from human cell lines and mouse embryonic stem cells

We first applied ELECT on scHi-C datasets from human cell lines [19] GM12878, K562, Hela, Hap1, and mouse embryonic stem cells (mESC, Serum/Lif) [20] at resolutions 500Kb and 1Mb, respectively. The choice of the datasets and bin sizes were driven by availability of external datasets that can be used for validation and the sequencing depths of the datasets (Supp. Section S3). For both collections, cliques with sizes 3-6 were evaluated. This resulted in about 50K-170K (size 3), 36K-150K (size 4), 10K-50K (size 5), 1K-17K (size 6) candidate cliques throughout cell types for the Ramani *et al.* collection [19] and 134-159K (size 3),129K-187K (size 4), 21K-39K (size 5), 3K-9K (size 6) cliques for the Li *et al.* [20] dataset. After obtaining the clique scores from ELECT for these low resolution data (**Fig.** 2a), we considered multiple approaches to evaluate whether these scores agreed with external data supporting multi-way interactions. Specifically, we considered (i) external cell type specific bulk Hi-C contact matrices[21]; (ii) SPRITE enrichment scores of the cliques[4]; and (iii) SPRITE-derived contact matrices[4]. While SPRITE directly quantified multi-way interactions, this assay has not been widely applied (to the best of our knowledge, of the scHi-C datasets we have, only GM12873 and mESC have corresponding SPRITE data). Consequently, we also relied on external cell type specific bulk pairwise contact matrices which aggregate cells of the same type experimentally, and hence, are expected to exhibit high contact counts for pairwise interactions within cliques.

Fig. 3a,b display the clique scores (Z_q, $q \in [Q]$), against the $\left(\binom{N-1}{2} + 1\right)$th largest (normalized) contact counts among the locus pairs within each clique $q \in [Q]$ derived from external GM12878 and K562 bulk Hi-C datasets [21], respectively. Specifically, we chose the order number as $\binom{N-1}{2} + 1$ for N way cliques to ensure that all the N loci are participating to form a connected graph. These comparisons reveal that ELECT clique scores have strong positive correlations with the Hi-C contact counts for both cell types. We further quantified

this relationship by testing whether the regression coefficients from regressing average contact counts on the clique scores were significantly larger than 0, stratified by clique size and chromosome (**Fig.** S2). For both cell types, all correlations between the two modalities were significantly positive (all adjusted p-values for testing the regression slopes $\ll 0.01$; results for cell lines Hela, Hap1 in **Fig.** S3).

Next, we repeated a similar analysis with the SPRITE-derived pairwise contact matrices of GM12878 cells (**Fig.** 3c) and observed similarly high positive correlation between the clique scores and the average SPRITE contact levels of locus pairs harbored in each clique q of GM12878 cells (all clique sizes and chromosomes, $p_{adj} \ll 0.01$, **Fig.** S2).

We further explored the enrichment scores of the SPRITE multi-way interactions introduced in Quinodoz *et al.* [4]. These scores aim to denoise raw SPRITE counts by comparing observed counts of a multi-way interaction against counts of randomly constructed multi-way interactions of the same size while matching the genomic distances within the random multi-way interactions and the observed one. Higher enrichment scores are interpreted as strong evidence for a multi-way interaction [4]. We grouped the cliques from scHi-C into two based on p-values of their clique scores using the parametric distribution of the p-values $B(r, M - r + 1)$ and the Benjamini-Hotchberg (BH) p-value adjustment [22]. Cliques with FDR < 0.01 were labelled as significant multi-way interactions. For both the GM12878 and mESC datasets, for which SPRITE data was available, cliques labelled as significant by ELECT exhibited significantly larger enrichment scores compared to cliques that were not significant (**Fig.** 3d,e; Wilcoxon ranksum test p-values < 0.01). Further investigation with clique size stratification revealed that significant cliques from ELECT exhibit a larger enrichment score mode than that of the insignificant cliques for both GM12878 and mESC (**Fig.** S2b,c).

Next, we turned our attention to higher resolution (10Kb) mESC (2i/Lif) scHi-C data [23] and identified about 131K 3-way, 174K 4-way, 163K 5-way multiway interactions at FDR of 0.01 based on 3,686 cells. We leveraged the recent long-read based single cell high throughput chromatin conformation (scNanoHi-C) data of mESC (2i/Lif) [11] for validation of these cliques. scNanoHi-C assay generated long sequencing reads that contained multiple DNA monomers, referred to as "concatemers". Consequently, concatemers that include DNA monomers originating from three or more different genomic locations signify interactions that are multi-way. We investigated the genomic distance and depth normalized concatemer counts of multi-way interactions (**Fig.** 3f) and observed that multi-way interactions inferred from ELECT have significantly higher scNanoHi-C counts across clique sizes compared to cliques that are not deemed as multi-way interactions (one-sided Wilcoxon rank sum test $p < 2e - 16$). Further stratification of the cliques based on their within clique maximum genomic distance revealed that cliques with maximum within clique genomic distance less than 200Kb are well supported by the scNanoHi-C data (**Fig.** S5), suggesting an empirical genomic distance constraint for post-processing the significant cliques from ELECT.

Collectively, these results suggest that ELECT can identify multi-way interactions in high and low resolution scHi-Data from cell lines. In what follows, we showcase the applicability of ELECT in the high-resolution setting and with more recent human and mouse brain scHi-C datasets.

3.2 Application and validation of ELECT with high-resolution scHi-C data from human and mouse brain

We next applied ELECT to infer multi-way interactions in human prefrontal cortex cell types (excitatory neurons: L2/3, L4, L5, L6 and inhibitory neurons: Nndf, Pvalb, Sst, Vip) using scHi-C data from sn-m3C-seq assay which simultaneously profiles chromatin conformation capture and methylation at the single cell level [14]. Data for each cell type was binned at 10Kb resolution to generate contact matrices for each cell. While gold standard multi-way interactions for this system are not available via other technologies, we designed an approach to assess whether the resulting clique scores associated with multi-way interactions by leveraging the interplay between methylation and chromatin conformation capture at the single cell level. Specifically, we utilized the well-established observation that DNA methylation level correlations (either positive or negative) are indicative of regulatory relationships among the loci [25–27]. In fact, recent work established that DNA methylation profiles and chromatin accessibility as measured by ATAC-seq at the single nuclei level validated each other at the cell-subtype level [28]. Hence, genomic loci within a multi-way interaction can be expected to exhibit a dependency structure in their DNA methylation profiles. To assess such dependence, we leveraged partial correlation coefficient and quantified the level of conditional dependence between pairs of genomic loci given the other loci in the multi-way interactions. For example, for a three way interaction, if any of the pairs

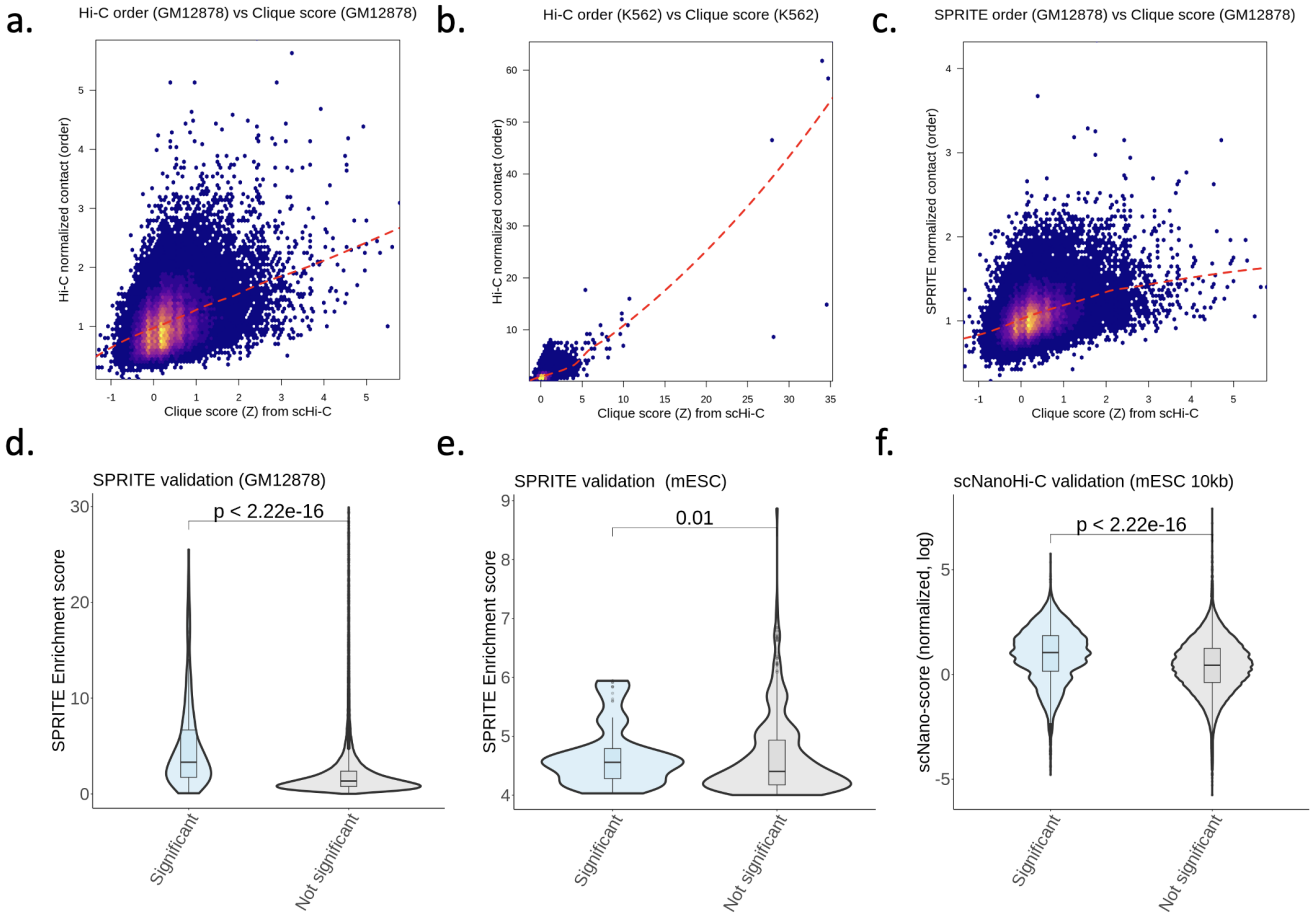


Fig. 3: Validation of ELECT cliques from low resolution scHi-C datasets. **a,b.** Hexbin plots of ELECT clique scores (x-axis) against the $\left(\binom{N-1}{2}+1\right)$th band normalized bulk Hi-C contact counts of locus pairs within cliques (y-axis) for GM12878 and K562 cell lines, respectively. Cliques at all sizes (sizes 3-6) and all chromosomes are displayed together. **c.** Hexbin plots of ELECT clique scores (x-axis) for GM12878 and K562 cells against the $\left(\binom{N-1}{2}+1\right)$th band normalized SPRITE-derived contact count of locus pairs in GM12878 cells. For **a-c**, points with brighter colors correspond to higher density levels and the dashed red lines represent locally estimated scatter plot smoothing (LOESS) fits. **d,e.** Comparison of SPRITE multi-way interaction enrichment scores [4] between the cliques identified as multi-way interactions ($FDR < 0.01$) by ELECT and those with insignificant clique scores ($FDR \geq 0.01$) for GM12878 cells [24] and mESC cells [20], respectively. **f.** Comparison of normalized scNanoHi-C [11] concatemer counts between the cliques identified as multi-way interactions ($FDR < 0.01$) by ELECT and those with insignificant clique scores ($FDR \geq 0.01$).

exhibit low partial correlation in absolute value given the third locus in the candidate interaction, this would indicate that pairwise dependencies within the candidate multi-way interaction are not supported by the co-methylation patterns. With this line of reasoning, we then compared clique scores with the average DNA methylation-based partial correlations between all the loci of individual cliques, at clique sizes of 3 and 4 and across all the eight cell types (**Fig.** 4a-b). This analysis revealed that cliques with larger clique scores tend to have larger multi-way dependency levels across all clique sizes for each cell type. We further assessed the association between the clique scores and partial-correlation based multi-way dependencies (slopes of the regression lines, **Fig.** 4c). In this quantification, majority of the comparisons yielded a significant and positive association between the clique scores and partial-correlation based evidence for multi-way interaction.

Next, we utilized ELECT to infer multi-way interactions from mouse cortex and hippocampus scHi-C data [29]. Data for each cell type was binned at 25Kb resolution to match the resolution of the publicly available mouse brain DNA seqFISH+ data [30] for validation. The DNA seqFISH+ dataset harbored spatial coordinate information of targeted 1.5Mb genomic region of each chromosome and enabled calculation of spatial distances of loci pairs. We considered the cell types common to both data resources (scHi-C and DNA seqFISH+) and compared the clique scores from ELECT against the $\left(\binom{N-1}{2}+1\right)$th closest spatial distance

between loci pairs within the clique (**Fig.** 4d,e). This comparison indicated a consistent inverse relationship (high ELECT score and low spatial distance) across the cell types (Asctrocyte ,Microglia, and Oligodendrocyte Progenitor) and clique sizes (3,4). Assessment of this inverse association with slopes of the regression lines (**Fig.** 4f) revealed that majority of the comparisons yielded a significant and negative association between the clique scores and the spatial distance-based evidence for multi-way interactions. Collectively, these results support that ELECT can identify multi-way interactions by leveraging high resolution scHi-C data from tissues.

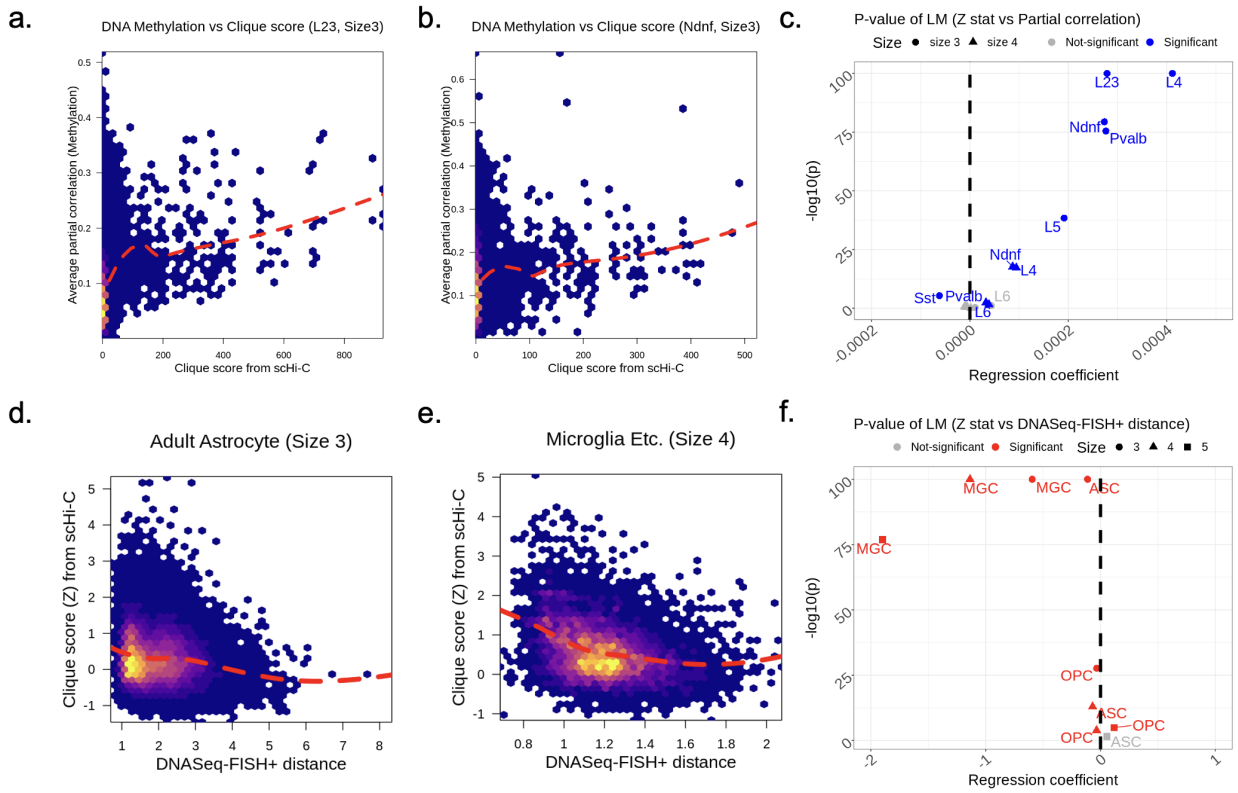


Fig. 4: Validation of ELECT cliques from high resolution scHi-C datasets and comparison with a baseline approach. **a-b**. Hexbin plots of clique scores (x-axis) against average absolute partial correlations of locus pairs within each clique based on single cell DNA methylation (y-axis) for cell types L23 and Ndnf, respectively (**Fig.** S4 for other cell types). **c**. $-\log_{10}(p)$-values (y-axis) of the estimated regression coefficients (x-axis) from regressing the average absolute partial correlations (DNA methylation support for multi-way dependency) on clique scores. Each point represents the result for one cell type at a fixed clique size. The dashed vertical line marks $x = 0$. Data points with BH-adjusted p-values less than 0.05 are labelled as significant. **d-e**. Hexbin plots of clique scores (y-axis) against the $\left(\binom{N-1}{2}+1\right)$-th closest DNA seqFISH+ [30] distance of locus pairs within each clique (x-axis) for cell types Adult Astrocyte, Microglia, etc., respectively (**Fig.** S4 for other clique sizes and cell types). For **a,b,d,e**, points with brighter colors correspond to higher density levels and the dashed red lines represent locally estimated scatter plot smoothing (LOESS) fits. **f**. $-\log_{10}(p)$-values (y-axis) of the estimated regression coefficients (x-axis) from regressing the ELECT clique scores on the clique level DNA seqFISH+ distances. Each point represents the result for one cell type at a fixed clique size. The dashed vertical line marks $x = 0$. Data points with BH-adjusted p-values less than 0.05 are labelled as significant.

3.3 ELECT discovers multi-way interactions of enhancers with genes

After establishing that multi-way interactions inferred by ELECT are supported by external data such as SPIRITE, DNA seqFISH+, and scNanoHi-C, we turned our attention to specific multi-way interactions that involve genes and enhancers. For these investigations, we leveraged (i) Activity-By-Contact (ABC) scores [31] that infer gene-enhancer interactions based on genomewide profiling of chromatin accessibility and activating histone mark H3K27ac; and (ii) gene-enhancer interactions inferred from multi-modal single cell profiling of chromatin accessibility and gene expression [32].

We considered the mESC scHi-C dataset [23] cliques evaluated by ELECT and leveraged the ABC scores from Narita *et al.* [33] to asses whether multi-way interactions with at least one gene are supported by the ABC scores of the involved enhancers. A total of 6,030 cliques harbored genes regulated by multiple enhancers within the same clique, i.e., cliques containing a full $g - e_1 - e_2 - \cdots - e_N$ structure, where g refers to a gene with ABC scores evaluated for enhancers $e_1, e_2, \cdots, e_N$. Specifically, we required the cliques to have one locus overlapping with a gene and have the other clique loci each overlap with at least one enhancer for which the gene has a reported ABC score. For each clique satisfying this requirement and each $g - e_1 - e_2 - \cdots - e_N$ structure in the clique (6,584 cases in total), we took the average of the ABC scores assigned for all $g - e_j$ links ($j \in \{1, \cdots, N\}$) to summarize ABC scores at the clique level. Then, we assessed if the multi-way interactions inferred from ELECT (FDR cutoff of 0.05) are supported by the multi-way gene regulation level, which is represented by the clique level ABC score. This comparison revealed a significant difference in summarized ABC scores of significant and insignificant cliques (**Fig.** 5a, one-sided Wilcoxon rank sum test $p < 2e - 16$). A less stringent version of this analysis that did not require the $g - e_1 - e_2 - \cdots - e_N$ structure revealed similar results (Supp. Section S4).

Next, for the Lee *et al.* human brain scHi-C data analyzed at 10Kb resolution, we conducted a similar analysis for the cliques with a full $g - e_1 - e_2 - \cdots - e_N$ structure, with the enhancers e_js from the set of prefrontal cortex enhancers identified by an integrative analysis of snATAC-seq and scRNA-seq data (further details in **Fig.** 2 of [32]). Specifically, Li *et al.* [32] provides a set of $g - e_j$ links referred to as positively correlated gene-cCRE (*cis*-regulatory elements) connections. We compared the ELECT $-\log_{10}$(p-values) of the cliques containing a $g - e_1 - e_2 - \cdots - e_N$ (with positively correlated gene-cCRE connections) against those of randomly selected cliques and observed that cliques supported by the gene-cCRE connections tend to have markedly stronger ELECT multi-way interaction levels (**Fig.** 5c,d). A consistent result can be observed across cell types based on a more systematic comparison with multiple replicates of the randomly selected cliques (**Fig.** 5b).

Collectively, these results highlight that ELECT can be employed to identify multi-way interactions of genes with multiple enhancers.

3.4 ELECT identifies significantly more multi-way interactions of genes and enhancers in human prefrontal cortex compared to a baseline approach

To the best of our knowledge, ELECT is the first approach for inferring multi-way interactions from scHi-C data. We compared ELECT to a baseline approach that we devised from standard analysis of scHi-C data. Specifically, we considered multi-way interactions in the form of cliques from the significant loop calls of SnapHiC [9] (i.e., significant pairwise interactions) based on cell type specific scHi-C data. We deployed this baseline approach on the human prefrontal cortex scHi-C data and observed that it did not yield higher order cliques than size 3, hence we retained the comparisons to cliques of size 3. We thresholded the candidate cliques from ELECT at FDR 0.05 to obtain significant 3-way interactions.

Next, we employed external PsychENCODE data [34] to evaluate the resulting multi-way interactions of ELECT and the baseline method. Specifically, we utilized the compiled neuronal cell type gene-enhancer interactions which encompassed one or more interactions per gene. After filtering the cliques of each method so that at least one locus in the clique harbored at least one gene, we quantified how many of the method cliques were of the form $g - e_1 - e_2$ structure (v-structure), where g refers to the gene with enhancers e_1 and e_2 in the gene-enhancer link list from the PsychENCODE source. Specifically, for each gene g within a locus of a clique q, we assigned the clique q a score of 1, if the other loci in the clique individually overlapped the PsychENCODE enhancers of the gene g. Under this scoring scheme, a clique could have score s if it contains s number of $g - e_1 - e_2$ v-structures. Next, we quantified proportion of method cliques with scores $s \geq 1$ (**Fig.** 5e). This calculation was carried out separately for each cell type. We observe that ELECT outperforms the baseline approach in terms of capturing the $g - e_1 - e_2$ structures throughout all the cell types except for cell types L4 and L5. Collectively, these results suggest that ELECT identifies multi-way interactions from the high-resolution scHi-C data with better power than a natural baseline approach.

3.5 ELECT detects multi-way interactions with potential GWAS association SNP interactions

We next sought to utilize results from ELECT to ask more directed biological questions. First, we asked whether genes involved in significant multi-way interactions exhibited different expression patterns than those

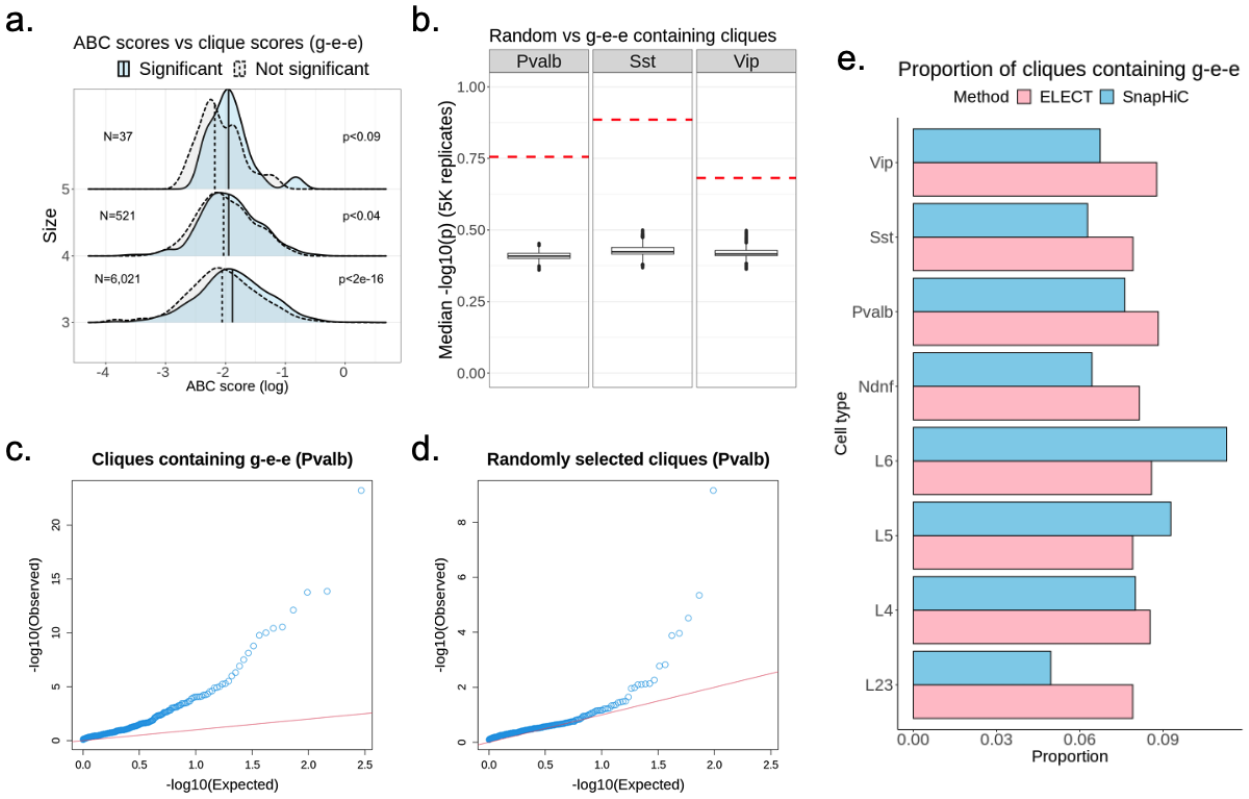


Fig. 5: **a**. Density plots of the cliques containing full $g - e_1 - e_2 - \cdots - e_N$ structures across the entire loci. x-axis denotes the clique level summarization of the ABC scores [33] with log_e transformation, and the y-axis denotes the clique size. For each given clique size, the cliques that are called as significant (BH FDR<0.05) by ELECT are colored as blue, and gray otherwise. The p-values on the right hand side of the density plots denote the one-sided Wicoxon rank sum test p-values. The N values denote the number of (clique,gene)-pair used for the violin plots. **b**. Boxplots summarizing the median ELECT $-\log_{10}(p)$ values of the Q randomly sampled cliques across 5K replicates, where Q is the number of cliques overlapping with the full $g - e_1 - e_2$ structure obtained from the positively correlated gene-cCRE (*cis*-regulatory elements) connections derived from human brain snATAC-seq [32]. Each panel corresponds to the result for each cell type common for both the Lee *et al.* and the gene-cCRE lists, and the red horizontal line on each panel denotes the median $-\log_{10}(p)$ values of the Q cliques containing full $g - e_1 - e_2$ structure. **c,d**. QQ-plot of the Q number of cliques (of cell type Pvalb) containing full $g - e_1 - e_2$ with positive correlation, and that of the Q number of randomly selected cliques. Red line denotes the $y = x$ line. **e**. Proportion of cliques with the full $g - e_1 - e_2$ v-structure (listed in PsychENCODE result) among the cliques harboring at least one gene. ELECT cliques are set by $FDR < 0.05$. The baseline approach identified multi-way interactions by enumerating cliques from the list of SnapHiC-derived significant loops (pairwise interactions listed in Supplementary data[9]) for each cell type.

that did not. Specifically, we grouped the cliques of human prefrontal cortex into two as those with and without significant clique scores at FDR of 0.1 within each cell type. Then, utilizing SMART-seq data [35] of the same system (Supp. Section S3), we compared expression levels of the genes harbored within each group. **Fig.** 6a compares the expression distribution of the genes within the two groups and yields that the cliques with significant multi-way interactions harbor genes with, on average, significantly higher expression levels compared to the other cliques throughout most of the cell types (Wilcoxon rank sum test p-values $\leq$ 0.003). This agrees well with the findings from the long-read based scHi-C which unravelled a similar result for GM12878 cells [36].

To further showcase how ELECT results can be explored for other types of discovery, we overlapped GWAS association SNPs from psychiatric diseases including autism spectrum disease (ASD) [37–52] and major depressive disorder (MDD) [53–61] with ELECT inferred multi-way interactions of the human prefrontal cortex. **Fig.** 6 showcases the types of results emerging from this analysis. The 3-way clique in the top panel of **Fig.** 6b displays a multi-way interaction inferred from both the inhibitory and the excitatory neurons. All the loci within this clique harbor SNPs significantly associated with ASD. Furthermore, the three loci (n_1, n_2, n_3) within the clique harbor gene HLA-DRB1 (n_2), an enhancer named GH06J032548 (n_3) of the gene and several intergenic regions. The HLA-DRB1 gene is a member of the HLA class II beta chain paralogues, and previous investigations [62–64] for this class of genes indicated significant association between these

genes and the ASD status. Moreover, the enhancer GH06J032548 was validated as an active enhancer by ChIP-STARR-seq massively parallel reporter assay[65] and its target genes included HLA-DRB1 that is within the multi-way interaction ELECT identified.

While the 3-way interaction in the top panel of **Fig.**6b was identified for all the excitatory and inhibitory neurons, the second clique displayed in the right panel of **Fig.** 6b was significant exclusively for the inhibitory (Sst and Vip) cell types. This clique has one node containing the PGBD1 gene (n_6), another node with the ZSCAN26 (n_5) gene, and the third node (n_4) with the ZSCAN31. The three genes in each of the nodes n_4, n_5 and n_6 are shown to significantly associate with MDD [66, 67]. Overall, these results showcase the ability of ELECT in proffering valid multi-way interactions among genomic loci that harbor genes and enhancers.

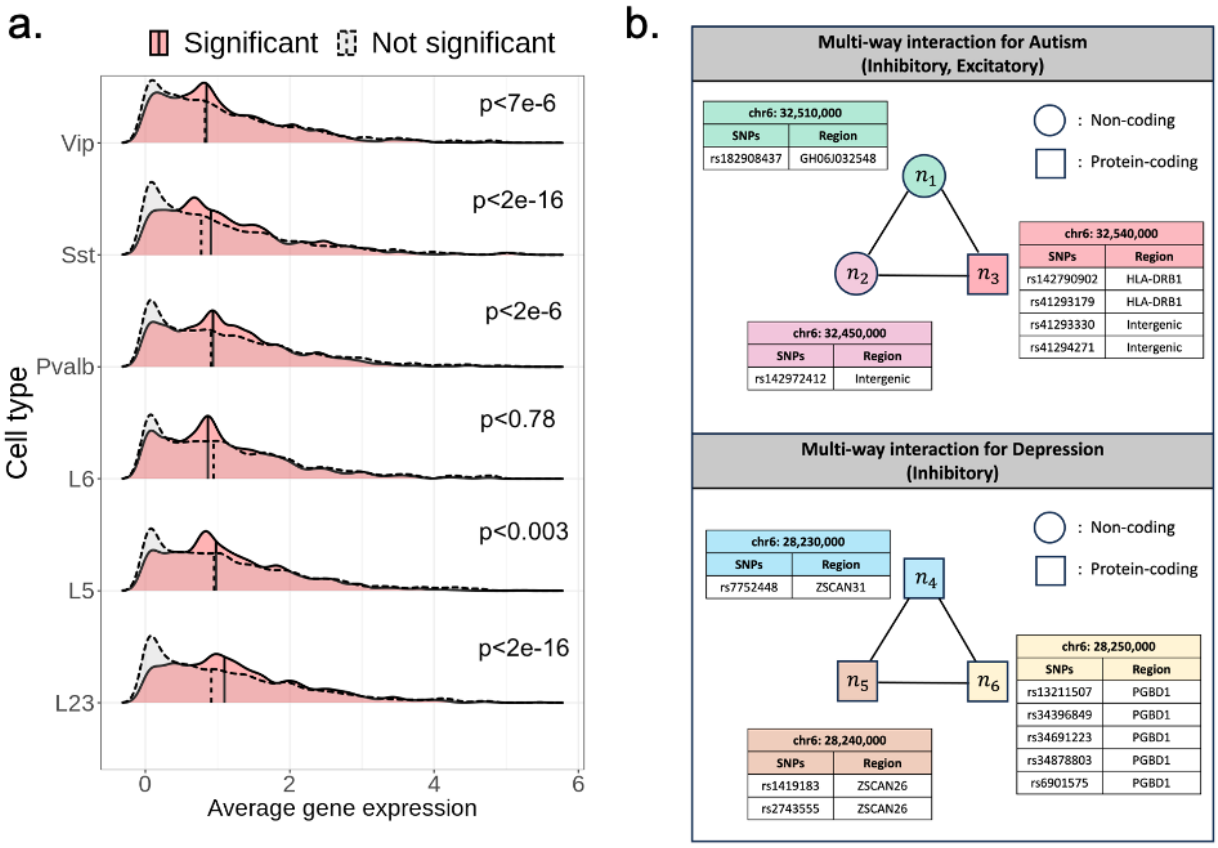


Fig. 6: Applications of ELECT for specific biological queries. **a.** Comparison of average (normalized) gene expression levels among cliques with ($FDR < 0.1$) and without significant multi-way interactions ($FDR \geq 0.1$) across the cell types in human brain prefrontal cortex[14], for all clique sizes and chromosomes. p-values on the top right corners are based on one-sided Wilcoxon rank sum test. **b.** Multi-way interactions overlapping GWAS SNPs from ASD (left) and MDD (right). For each panel, an undirected graph is used to display the multi-way interaction. A square node corresponds to a locus with significant GWAS SNPs (for the corresponding disease) overlapping with protein coding genes within the locus. Likewise, a circle node denotes a locus with significant GWAS SNPs overlapping the non-coding region in the locus. Tables around the nodes with matching colors display the lists of significant GWAS SNPs in the locus and the genes (or the enhancers) overlapping with the listed SNPs.

4 Discussion

We presented ELECT as an empirical Bayes framework to infer multi-way interactions from scHi-C data. Despite the limitations due to data sparsity and proximity ligation, scHi-C data harbors cliques representing potential multi-way interactions (albeit of lower sizes ≤ 6) and presents an exciting opportunity to learn higher-order chromatin interactions. ELECT builds on a dirichlet-multinomial model that takes into account inherent genomic distance bias of scHi-C data by a natural spline model. We developed a computationally

efficient estimation strategy for ELECT and showed that null generative model of ELECT results in well calibrated p-values, enabling appropriate false discovery rate control (**Fig.** S7, S8, S9, S10, S11). Output from ELECT includes clique scores generated by aggregating interactions from pairs of loci within a clique and across cells of the same cell type. Evaluations of ELECT using a wide range of real scHi-C datasets yielded multiple-levels of support for ELECT multi-way interactions from independent external genomic data resources. In addition, cliques characterized by stronger gene regulation via multiple enhancers typically exhibit a higher degree of ELECT multi-way interaction, as opposed to cliques with weaker levels of gene regulation. This finding highlights the potential of ELECT to infer multi-way gene regulation. While there are not existing computational methods aimed at identifying multi-way interactions from scHi-C data, we considered a baseline approach by utilizing state-of-the art scHi-C loop calling tool SnapHi-C[9] and generated multi-way interactions as cliques from significant loop calls. Applications of the baseline approach and ELECT on human prefrontal cortex cells yielded that ELECT is able to identify markedly more multi-way interactions involving at least one gene and two enhancers among the set of PsychENCODE-derived gene-enhancer links. We explored how multi-way interactions from ELECT can be leveraged for other types of analysis. Specifically, we showcased how ELECT multi-way interactions harbor loci, each of which harbor GWAS risk SNPs, suggesting potential interactions between genetic variants. We expect that with the continued development of scHi-C technologies, ELECT analysis of scHi-C data will complement data from other multi-way interaction assays.

While ELECT applications presented here relied on known cell type labels, this is not an inherent limitation because there is a plethora of computational tools for inferring cell type labels from scHi-C data [9, 17, 68–73]. Furthermore, once we quantify the uncertainty of label assignments (e.g., by soft-threshold in clustering of low-dimensional embeddings of single cell data), these uncertainity estimates can be naturally incorporated into the generalized linear model fit that ELECT employs. More generally, from a methodological point of view, the two-stage (band level and locus pair level) modeling by ELECT offers a computationally useful framework for any type of high-dimensional and sparse single cell genomics/epigenomics data where exponential family of distributions is a valid generative distribution choice. Specifically, if a group of genomic loci (or genes) can be bundled into a larger class (analogous to a band in ELECT) for a GLM fitting, then the original locus level estimation (analogous to locus pair level estimation in ELECT) can be carried out based on estimate allocation procedure introduced in **Fig.** 1d.

5 Acknowledgement

We thank Dr. Longzhi Tan from Stanford University for insightful discussions regarding the method validation, Dr. Jiansen Lu from Peking University for sharing detailed information and discussions about the scNanoHi-C data, Dr. Takeo Narita from Copenhagen University for sharing the ABC scores of the mESC system, and Siqi Shen from the University of Wisconsin-Madison for sharing processed Nagano *et al.* and Tan *et al.* scHi-C data. We also thank Dr. Ye Zheng from the University of Texas MD Anderson Cancer Center for a useful discussion regarding the baseline method comparison.

Bibliography

[1] Lin, X., Liu, Y., Liu, S., Zhu, X., Wu, L., Zhu, Y., Zhao, D., Xu, X., Chemparathy, A., Wang, H., Cao, Y., Nakamura, M., Noordermeer, J.N., La Russa, M., Wong, W.H., Zhao, K., Qi, L.S.: Nested epistasis enhancer networks for robust genome regulation. Science 377(6610), 1077–1085 (Sep 2022), https://www.science.org/doi/10.1126/science.abk3512

[2] Beagrie, R.A., Scialdone, A., Schueler, M., Kraemer, D.C.A., Chotalia, M., Xie, S.Q., Barbieri, M., De Santiago, I.and Lavitas, L.M., Branco, M.R., Fraser, J., Dostie, J., Game, L., Dillon, N., Edwards, P.A.W., Nicodemi, M., Pombo, A.: Complex multi-enhancer contacts captured by genome architecture mapping. Nature 543(7646), 519–524 (Mar 2017), http://www.nature.com/articles/nature21411

[3] Zheng, M., Tian, S.Z., Capurso, D., Kim, M., Maurya, R., Lee, B., Piecuch, E., Gong, L., Zhu, J.J., Li, Z., Wong, C.H., Ngan, C.Y., Wang, P., Ruan, X., Wei, C.L., Ruan, Y.: Multiplex chromatin interactions with single-molecule precision. Nature 566(7745), 558–562 (Feb 2019), http://www.nature.com/articles/s41586-019-0949-1

[4] Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y., Trinh, V., Aznauryan, E., Russell, P., Cheng, C., Jovanovic, M., Chow, A., Cai, L., McDonel, P.and Garber, M., Guttman, M.: Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. Cell 174(3), 744–757.e24 (Jul 2018), https://linkinghub.elsevier.com/retrieve/pii/S0092867418306366

[5] Arrastia, M.V., Jachowicz, J.W., Ollikainen, N., Curtis, M.S., Lai, C., Quinodoz, S.A., Selck, D.A., Ismagilov, R.F., Guttman, M.: Single-cell measurement of higher-order 3D genome organization with scSPRITE. Nature Biotechnology 40(1), 64–73 (Jan 2022), https://www.nature.com/articles/s41587-021-00998-1

[6] Vermeulen, C., Allahyar, A., Bouwman, B.A.M., Krijger, P.H.L., Verstegen, M.J.A.M., Geeven, G., Valdes-Quezada, C., Renkens, I., Straver, R., Kloosterman, W.P., De Ridder, J., De Laat, W.: Multi-contact 4C: long-molecule sequencing of complex proximity ligation products to uncover local cooperative and competitive chromatin topologies. Nature Protocols 15(2), 364–397 (Feb 2020), http://www.nature.com/articles/s41596-019-0242-7

[7] Darrow, E.M., Huntley, M.H., Dudchenko, O., Stamenova, E.K., Durand, N.C., Sun, Z., Huang, S., Sanborn, A.L., Machol, I., Shamim, M., Seberg, A., Lander, E.S., Chadwick, B.P., Aiden, E.L.: Deletion of *DXZ4* on the human inactive X chromosome alters higher-order genome architecture. Proceedings of the National Academy of Sciences 113(31) (Aug 2016), https://pnas.org/doi/full/10.1073/pnas.1609643113

[8] Zhong, J.Y., Niu, L., Lin, Z.B., Bai, X., Chen, Y., Luo, F., Hou, C., Xiao, C.L.: High-throughput Pore-C reveals the single-allele topology and cell type-specificity of 3D genome folding. Nature Communications 14(1), 1250 (Mar 2023), https://www.nature.com/articles/s41467-023-36899-x

[9] Yu, M., Abnousi, A., Zhang, Y., Li, G., Lee, L., Chen, Z., Fang, R., Wen, J., Sun, Q., Li, Y., Ren, B., Hu, M.: SnapHiC: a computational pipeline to map chromatin contacts from single cell Hi-C data. (12 2020), https://www.biorxiv.org/content/10.1101/2020.12.13.422543v1

[10] Li, X., Lee, L., Abnousi, A.and Yu, M., Liu, W., Huang, L.and Li, Y., Hu, M.: SnapHiC2: A computationally efficient loop caller for single cell Hi-C data. Computational and Structural Biotechnology Journal 20, 2778–2783 (2022), https://linkinghub.elsevier.com/retrieve/pii/S2001037022002021

[11] Li, W., Lu, J., Lu, P., Gao, Y., Bai, Y., Chen, K., Su, X., Li, M., Liu, J., Chen, Y., Wen, L., Tang, F.: scNanoHi-C: a single-cell long-read concatemer sequencing method to reveal high-order chromatin structures within individual cells. Nature Methods 20(10), 1493–1505 (Oct 2023), https://www.nature.com/articles/s41592-023-01978-w

[12] Olshen, A.B., Segal, M.R.: Does multi-way, long-range chromatin contact data advance 3D genome reconstruction? BMC Bioinformatics 24(1), 64 (Feb 2023), https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-023-05170-x

[13] Zhang, R., Ma, J.: MATCHA: Probing Multi-way Chromatin Interaction with Hypergraph Representation Learning. Cell Systems 10(5), 397–407.e5 (May 2020), https://linkinghub.elsevier.com/retrieve/pii/S2405471220301472

[14] Lee, D.S., Luo, C., Zhou, J., Chandran, S., Rivkin, A., Bartlett, A., Nery, J.R., Fitzpatrick, C., O'Connor, C., Dixon, J.R., et al.: Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. Nature methods 16(10), 999–1006 (2019)

[15] Ay, F., Bailey, T.L., Noble, W.S.: Statistical confidence estimation for hi-c data reveals regulatory chromatin contacts. Genome research 24(6), 999–1011 (2014)

[16] Stansfield, J.C., Cresswell, K.G., Dozmorov, M.G.: multihiccompare: joint normalization and comparative analysis of complex hi-c experiments. Bioinformatics 35(17), 2916–2923 (2019)

[17] Zheng, Y., Shen, S., Keleş, S.: Normalization and De-noising of Single-cell Hi-C Data with BandNorm and scVI-3D (2022), to appear in *Genome Biology*. https://www.biorxiv.org/content/10.1101/2021.03.10.434870v1

[18] Baker, S.G.: The Multinomial-Poisson Transformation. The Statistician 43(4), 495 (1994), https://www.jstor.org/stable/2348134?origin=crossref

[19] Ramani, V., Deng, X., Qiu, R., Gunderson, K.L., Steemers, F.J., Disteche, C.M., Noble, W.S., Duan, Z., Shendure, J.: Massively multiplex single-cell Hi-C. Nature methods 14(3), 263–266 (2017)

[20] Li, G., Liu, Y., Zhang, Y., Kubo, N., Yu, M., Fang, R., Kellis, M., Ren, B.: Joint profiling of DNA methylation and chromatin architecture in single cells. Nature methods 16(10), 991–993 (2019)

[21] Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al.: A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159(7), 1665–1680 (2014)

[22] Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) 57(1), 289–300 (1995)

[23] Nagano, T., Lubling, Y., Várnai, C., Dudley, C., Leung, W., Baran, Y., Mendelson Cohen, N., Wingett, S., Fraser, P., Tanay, A.: Cell-cycle dynamics of chromosomal organization at single-cell resolution. Nature 547(7661), 61–67 (2017)

[24] Ramani, V.and Deng, X., Qiu, R., Gunderson, K., Steemers, F., Disteche, C., Noble, W., Duan, Z., Shendure, J.: Massively multiplex single-cell Hi-C. Nature Methods 14 (01 2017)

[25] Akulenko, R., Helms, V.: Dna co-methylation analysis suggests novel functional associations between gene pairs in breast cancer samples. Human molecular genetics 22(15), 3016–3022 (2013)

[26] Wang, F., Xu, H., Zhao, H., Gelernter, J., Zhang, H.: Dna co-methylation modules in postmortem prefrontal cortex tissues of european australians with alcohol use disorders. Scientific reports 6(1), 19430 (2016)

[27] Kachroo, P., Morrow, J.D., Kho, A.T., Vyhlidal, C.A., Silverman, E.K., Weiss, S.T., Tantisira, K.G., DeMeo, D.L.: Co-methylation analysis in lung tissue identifies pathways for fetal origins of copd. European Respiratory Journal 56(4) (2020)

[28] Liu, H., Zhou, J., Tian, W., Luo, C., Bartlett, A., Aldridge, A., Lucero, J., Osteen, J.K., Nery, J.R., Chen, H., Rivkin, A., Castanon, R.G., Clock, B., Li, Y.E., Hou, X., Poirion, O.B., Preissl, S., Pinto-Duarte, A., O'Connor, C., Boggeman, L., Fitzpatrick, C., Nunn, M., Mukamel, E.A., Zhang, Z., Callaway, E.M., Ren, B., Dixon, J.R., Behrens, M.M., Ecker, J.R.: DNA methylation atlas of the mouse brain at single-cell resolution. Nature 598(7879), 120–128 (Oct 2021), https://www.nature.com/articles/s41586-020-03182-8

[29] Tan, L., Ma, W., Wu, H., Zheng, Y., Xing, D., Chen, R., Li, X., Daley, N., Deisseroth, K., Xie, X.S.: Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. Cell 184(3), 741–758 (2021)

[30] Takei, Y., Zheng, S., Yun, J., Shah, S., Pierson, N., White, J., Schindler, S., Tischbirek, C.H., Yuan, G.C., Cai, L.: Single-cell nuclear architecture across cell types in the mouse brain. Science 374(6567), 586–594 (2021)

[31] Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V., Grossman, S.R., Anyoha, R., Doughty, B.R., Patwardhan, T.A., et al.: Activity-by-contact model of enhancer–promoter regulation from thousands of crispr perturbations. Nature genetics 51(12), 1664–1669 (2019)

[32] Li, Y.E., Preissl, S., Miller, M., Johnson, N.D., Wang, Z., Jiao, H., Zhu, C., Wang, Z., Xie, Y., Poirion, O., et al.: A comparative atlas of single-cell chromatin accessibility in the human brain. Science 382(6667), eadf7044 (2023)

[33] Narita, T., Higashijima, Y., Kilic, S., Liebner, T., Walter, J., Choudhary, C.: Acetylation of histone h2b marks active enhancers and predicts cbp/p300 target genes. Nature Genetics 55(4), 679–692 (2023)

[34] Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F.C., Clarke, D., Gu, M., Emani, P., Yang, Y.T., et al.: Comprehensive functional genomic resource and integrative model for the human brain. Science 362(6420), eaat8464 (2018)

[35] Bakken, T.E., Jorstad, N.L., Hu, Q., Lake, B.B., Tian, W., Kalmbach, B.E., Crow, M., Hodge, R.D., Krienen, F.M., Sorensen, S.A., et al.: Comparative cellular analysis of motor cortex in human, marmoset and mouse. Nature 598(7879), 111–119 (2021)

[36] Li, W., Lu, J., Lu, P., Gao, Y., Bai, Y., Chen, K., Su, X., Li, M., Liu, J., Chen, Y., et al.: scnanohi-c: a single-cell long-read concatemer sequencing method to reveal high-order chromatin structures within individual cells. Nature Methods pp. 1–13 (2023)

[37] Peyrot, W.J., Price, A.L.: Identifying loci with different allele frequencies among cases of eight psychiatric disorders using cc-gwas. Nature genetics 53(4), 445–454 (2021)

[38] Matoba, N., Liang, D., Sun, H., Aygün, N., McAfee, J.C., Davis, J.E., Raffield, L.M., Qian, H., Piven, J., Li, Y., et al.: Common genetic risk variants identified in the spark cohort support ddhd2 as a candidate risk gene for autism. Translational Psychiatry 10(1), 265 (2020)

[39] Wu, Y., Cao, H., Baranova, A., Huang, H., Li, S., Cai, L., Rao, S., Dai, M., Xie, M., Dou, Y., et al.: Multi-trait analysis for genome-wide association study of five psychiatric disorders. Translational psychiatry 10(1), 209 (2020)

[40] Kuo, P.H., Chuang, L.C., Su, M.H., Chen, C.H., Chen, C.H., Wu, J.Y., Yen, C.J., Wu, Y.Y., Liu, S.K., Chou, M.C., et al.: Genome-wide association study for autism spectrum disorder in taiwanese han population. PLoS One 10(9), e0138695 (2015)

[41] Al-Sarraj, Y., Al-Dous, E., Taha, R.Z., Ahram, D., Alshaban, F., Tolfat, M., El-Shanti, H., Albagha, O.M.: Family-based genome-wide association study of autism spectrum disorder in middle eastern families. Genes 12(5), 761 (2021)

[42] Almandil, N.B., AlSulaiman, A., Aldakeel, S.A., Alkuroud, D.N., Aljofi, H.E., Alzahrani, S., Al-Mana, A., Alfuraih, A.A., Alabdali, M., Alkhamis, F.A., et al.: Integration of transcriptome and exome genotyping identifies significant variants with autism spectrum disorder. Pharmaceuticals 15(2), 158 (2022)

[43] Guo, W., Samuels, J., Wang, Y., Cao, H., Ritter, M., Nestadt, P., Krasnow, J., Greenberg, B., Fyer, A., McCracken, J., et al.: Polygenic risk score and heritability estimates reveals a genetic relationship between asd and ocd. European Neuropsychopharmacology 27(7), 657–666 (2017)

[44] Liu, X., Shimada, T., Otowa, T., Wu, Y.Y., Kawamura, Y., Tochigi, M., Iwata, Y., Umekage, T., Toyota, T., Maekawa, M., et al.: Genome-wide association study of autism spectrum disorder in the east asian populations. Autism Research 9(3), 340–349 (2016)

[45] Chaste, P., Klei, L., Sanders, S.J., Hus, V., Murtha, M.T., Lowe, J.K., Willsey, A.J., Moreno-De-Luca, D., Timothy, W.Y., Fombonne, E., et al.: A genome-wide association study of autism using the simons simplex collection: Does reducing phenotypic heterogeneity in autism increase genetic homogeneity? Biological psychiatry 77(9), 775–784 (2015)

[46] Rao, S., Baranova, A., Yao, Y., Wang, J., Zhang, F.: Genetic relationships between attention-deficit/hyperactivity disorder, autism spectrum disorder, and intelligence. Neuropsychobiology 81(6), 484–496 (2022)

[47] Baranova, A., Wang, J., Cao, H., Chen, J.H., Chen, J., Chen, M., Ni, S., Xu, X., Ke, X., Xie, S., et al.: Shared genetics between autism spectrum disorder and attention-deficit/hyperactivity disorder and their association with extraversion. Psychiatry Research 314, 114679 (2022)

[48] Qi, G., Chatterjee, N.: Heritability informed power optimization (hipo) leads to enhanced detection of genetic associations across multiple traits. PLoS genetics 14(10), e1007549 (2018)

[49] Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R., et al.: Identification of common genetic risk variants for autism spectrum disorder. Nature genetics 51(3), 431–444 (2019)

[50] Meta-analysis of gwas of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24. 32 and a significant overlap with schizophrenia. Molecular autism 8, 1–17 (2017)

[51] of the Psychiatric Genomics Consortium, C.D.G., et al.: Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. The Lancet 381(9875), 1371–1379 (2013)

[52] Lee, P.H., Anttila, V., Won, H., Feng, Y.C.A., Rosenthal, J., Zhu, Z., Tucker-Drob, E.M., Nivard, M.G., Grotzinger, A.D., Posthuma, D., et al.: Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. Cell 179(7), 1469–1482 (2019)

[53] Mei, L., Gao, Y., Chen, M., Zhang, X., Yue, W., Zhang, D., Yu, H.: Overlapping common genetic architecture between major depressive disorders and anxiety and stress-related disorders. Progress in Neuro-Psychopharmacology and Biological Psychiatry 113, 110450 (2022)

[54] Wang, H., Yi, Z., Shi, T.: Novel loci and potential mechanisms of major depressive disorder, bipolar disorder, and schizophrenia. Science China Life Sciences 65(1), 167–183 (2022)
[55] Yao, X., Glessner, J.T., Li, J., Qi, X., Hou, X., Zhu, C., Li, X., March, M.E., Yang, L., Mentch, F.D., et al.: Integrative analysis of genome-wide association studies identifies novel loci associated with neuropsychiatric disorders. Translational psychiatry 11(1), 69 (2021)
[56] Clements, C.C., Karlsson, R., Lu, Y., Juréus, A., Rück, C., Andersson, E., Boberg, J., Pedersen, N.L., Bulik, C.M., Nordenskjöld, A., et al.: Genome-wide association study of patients with a severe major depressive episode treated with electroconvulsive therapy. Molecular psychiatry 26(6), 2429–2439 (2021)
[57] Yin, B., Wang, X., Huang, T., Jia, J.: Shared genetics and causality between decaffeinated coffee consumption and neuropsychiatric diseases: A large-scale genome-wide cross-trait analysis and mendelian randomization analysis. Frontiers in Psychiatry 13, 910432 (2022)
[58] Tirozzi, A., Quiccione, M.S., Cerletti, C., Donati, M.B., de Gaetano, G., Iacoviello, L., Gialluisi, A.: A multi-trait association analysis of brain disorders and platelet traits identifies novel susceptibility loci for major depression, alzheimer's and parkinson's disease. Cells 12(2), 245 (2023)
[59] Guindo-Martínez, M., Amela, R., Bonàs-Guarch, S., Puiggròs, M., Salvoro, C., Miguel-Escalada, I., Carey, C.E., Cole, J.B., Rüeger, S., Atkinson, E., et al.: The impact of non-additive genetic associations on age-related complex diseases. Nature communications 12(1), 2436 (2021)
[60] Suppli, N.P., Andersen, K.K., Agerbo, E., Rajagopal, V.M., Appadurai, V., Coleman, J.R., Breen, G., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Pedersen, C.B., et al.: Genome-wide by environment interaction study of stressful life events and hospital-treated depression in the ipsych2012 sample. Biological psychiatry global open science 2(4), 400–410 (2022)
[61] Zhang, F., Cao, H., Baranova, A.: Shared genetic liability and causal associations between major depressive disorder and cardiovascular diseases. Frontiers in Cardiovascular Medicine 8, 735136 (2021)
[62] Chien, Y.L., Wu, Y.Y., Chen, C.H., Gau, S.S.F., Huang, Y.S., Chien, W.H., Hu, F.C., Chao, Y.L.: Association of hla-drb1 alleles and neuropsychological function in autism. Psychiatric genetics 22(1), 46–49 (2012)
[63] Bennabi, M., Gaman, A., Delorme, R., Boukouaci, W., Manier, C., Scheid, I., Si Mohammed, N., Bengoufa, D., Charron, D., Krishnamoorthy, R., et al.: Hla-class ii haplotypes and autism spectrum disorders. Scientific reports 8(1), 7639 (2018)
[64] Mostafa, G.A., Shehab, A.A., Al-Ayadhi, L.Y.: The link between some alleles on human leukocyte antigen system and autism in children. Journal of neuroimmunology 255(1-2), 70–74 (2013)
[65] Barakat, T.S., Halbritter, F., Zhang, M., Rendeiro, A.F., Perenthaler, E., Bock, C., Chambers, I.: Functional dissection of the enhancer repertoire in human embryonic stem cells. Cell stem cell 23(2), 276–288 (2018)
[66] Li, H., Chang, H., Song, X., Liu, W., Li, L., Wang, L., Yang, Y., Zhang, L., Li, W., Zhang, Y., et al.: Integrative analyses of major histocompatibility complex loci in the genome-wide association studies of major depressive disorder. Neuropsychopharmacology 44(9), 1552–1561 (2019)
[67] Levey, D.F., Stein, M.B., Wendt, F.R., Pathak, G.A., Zhou, H., Aslan, M., Quaden, R., Harrington, K.M., Nuñez, Y.Z., Overstreet, C., et al.: Bi-ancestral depression gwas in the million veteran program and meta-analysis in> 1.2 million individuals highlight new therapeutic directions. Nature neuroscience 24(7), 954–963 (2021)
[68] Zhou, J., Ma, J., Chen, Y., Cheng, C., Bao, B., Peng, J., Sejnowski, T.J., Dixon, J.R., Ecker, J.R.: Robust single-cell Hi-C clustering by convolution-and random-walk–based imputation
[69] Kim, H.J., Yardımcı, G.G., Bonora, G., Ramani, V., Liu, J., Qiu, R., Lee, C., Hesson, J., Ware, C.B., Shendure, J., et al.: Capturing cell type-specific chromatin compartment patterns by applying topic modeling to single-cell Hi-C data. PLoS computational biology 16(9), e1008173 (2020)
[70] Zhang, R., Zhou, T., Ma, J.: Multiscale and integrative single-cell Hi-C analysis with Higashi. Nature biotechnology 40(2), 254–261 (2022)
[71] Park, K., Keleş, S.: Joint tensor modeling of single cell 3d genome and epigenetic data with muscle. bioRxiv (2023)
[72] Li, X., Feng, F., Pu, H., Leung, W.Y., Liu, J.: scHiCTools: A computational toolbox for analyzing single-cell Hi-C data. PLoS computational biology 17(5), e1008978 (2021)
[73] Li, X., Zeng, G., Li, A., Zhang, Z.: DeTOKI identifies and characterizes the dynamics of chromatin TAD-like domains in a single cell. Genome biology 22(1), 1–26 (2021)