


RESEARCH ARTICLE

Open Access

Joint analyses of multi-tissue Hi-C and eQTL data demonstrate close spatial proximity between eQTLs and their target genes



Jingting Yu¹, Ming Hu^{2*} and Chun Li^{1,3*} 

Abstract

Background: Gene regulation is important for cells and tissues to function. It has been studied from two aspects at the genomic level, the identification of expression quantitative trait loci (eQTLs) and identification of long-range chromatin interactions. It is important to understand their relationship, such as whether eQTLs regulate their target genes through physical chromatin interaction. Although chromatin interactions have been widely believed to be one of the main mechanisms underlying eQTLs, most evidence came from studies of cell lines and yet no direct evidence exists for tissues.

Results: We performed various joint analyses of eQTL and high-throughput chromatin conformation capture (Hi-C) data from 11 human primary tissue types and 2 human cell lines. We found that chromatin interaction frequency is positively associated with the number of genes that have eQTLs and that eQTLs and their target genes tend to fall into the same topologically associating domain (TAD). These results are consistent across all tissues and cell lines we evaluated. Moreover, in 6 out of 11 tissues (aorta, dorsolateral prefrontal cortex, hippocampus, pancreas, small bowel, and spleen), tissue-specific eQTLs are significantly enriched in tissue-specific frequently interacting regions (FIRES).

Conclusions: Our data have demonstrated the close spatial proximity between eQTLs and their target genes among multiple human primary tissues.

Keywords: Hi-C, eQTL, Gene regulation, Primary human tissues, Human cell lines

Background

Gene regulation is important for cells and tissues to function. Differences in gene regulation are often responsible for cellular and morphological differences between cell lines and tissues. The advancement of high-throughput technologies such as DNA and RNA sequencing and SNP chips allows researchers to study gene regulation at the genomic level and from multiple perspectives. On the one hand, motivated by the likely functional importance of genetic variants in gene regulation, many studies have focused on identifying expression quantitative trait loci (eQTLs), which are genetic

variants statistically associated with gene expression across individuals [1–4]. eQTLs can regulate the expression of their target genes by altering *cis*-regulatory elements (CREs) such as enhancers, promoters, insulators, mediators, etc. [5–7]. On the other hand, analyses of chromatin spatial organization have established the importance of chromatin interaction in gene regulation [8–10]. For example, by forming long-range chromatin interactions, CREs can regulate the expression of their target genes hundreds of kilobases (Kb) away [11–13]. High-throughput chromatin conformation capture (Hi-C) has been widely adopted to provide a genome-wide view of chromatin interactions within a tissue or cell line [14–17]. Hi-C data are usually presented as a chromatin contact matrix, in which the genome is divided into equal-sized bins. The value of each element in the matrix represents the number of read pairs mapped to a pair of

* Correspondence: hum@ccf.org; cxl791@case.edu

²Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH, USA

¹Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA

Full list of author information is available at the end of the article



bins, which is called the chromatin interaction frequency (CIF).

These two complementary approaches focus on different aspects of gene regulation. eQTL results are statistical across individuals and require an associated SNP, while chromatin interactions are physical within a sample and do not require a polymorphism to be present. It is desirable to integrate the results of these two approaches to better understand their relationships, such as whether eQTLs regulate their target genes through chromatin interactions. Analyzing Hi-C data from human IMR90 fibroblasts and embryonic stem cells, Duggal et al. [18] showed that eQTLs are spatially close to their target genes, especially for those located within the same topologically associating domain (TAD) and overlapping with CREs, and that genomic regions containing eQTLs tend to have a higher CIF. Consistent with Duggal et al., using Hi-C data from human cell lines generated by Rao et al. [14], the Genotype-Tissue Expression (GTEx) study [3] has also shown that eQTLs that are enriched for CREs are in close spatial proximity with their target gene promoters. In both studies, the Hi-C data came only from cell lines, while the eQTL results were generated from both tissues and cell lines. Some other studies, for example, those aimed to predict enhancer-promoter interactions [19–21] and those aimed to detect regulatory SNPs [22], did not provide evidence to support the relationship between Hi-C data and eQTL results, but instead used this connection as ground truth to demonstrate the validity of their own results.

Although chromatin interactions have been widely believed as one of the main mechanisms underlying eQTLs, we are unaware of any direct evidence of this for tissues. It is well known that eQTLs are tissue specific [3]. Moreover, Schmitt et al. [23] recently identified hotspots of local chromatin interactions from Hi-C data, called frequently interacting regions (FIREs). FIREs are bins that frequently interact with nearby regions <200Kb, and they display strong tissue specificity. It is unclear how much overlap exists between tissue-specific FIREs and tissue-specific eQTLs.

Hi-C and eQTL data are now available for multiple human primary tissues and cell lines. For example, Schmitt et al. [23] generated Hi-C data for 14 human primary tissues and 7 human cell lines, and the GTEx study [3] performed genome-wide mapping of eQTLs across 48 human tissues. There are 11 tissues and 2 cell lines that overlap between these two sources (See Additional file 1: Supplementary Materials and Table S1), facilitating a direct evaluation of the relationship between chromatin interactions and eQTLs across multiple tissues and cell lines. Fig. 1 shows an example of the connection between Hi-C data and eQTL results.

We performed a series of joint analyses on the relationship between Hi-C data and eQTL results. We found that CIF is positively associated with the number of eGenes identified from the GTEx study (an eGene is defined as a gene in which the expression is significantly associated with an eQTL), and that eQTLs and their target genes are more likely to co-localize within the same TAD than randomly generated control datasets. All these results are consistent across all tissues and cell lines we evaluated. Since both eQTLs and FIREs are known to be highly tissue specific [3, 23], we also studied the relationship between tissue-specific eQTLs and tissue-specific FIREs and found that majority of the tissues demonstrate a positive association between them.

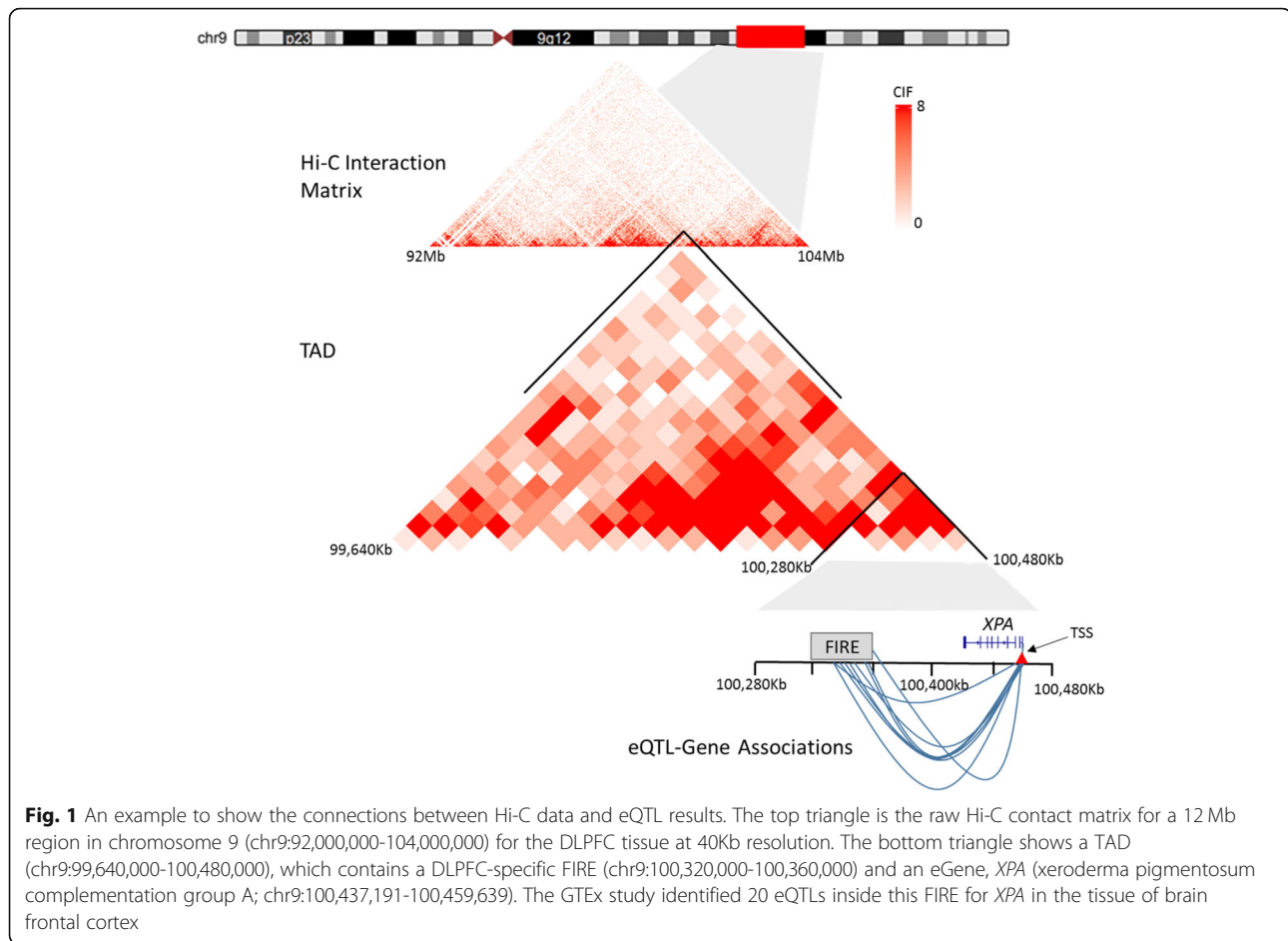
To the best of our knowledge, our study is the first to demonstrate the relationship between chromatin interactions and eQTLs across multiple human primary tissues, and to study the relationship between tissue-specific eQTLs and tissue-specific FIREs. These results help improve our understanding of the roles of chromatin interactions and eQTLs in gene regulation mechanisms.

Results

Chromatin interaction frequency is positively associated with the number of eGenes

If chromatin spatial organization affects how eQTLs regulate their target genes, one would expect that a pair of genomic loci mapped with eQTL-gene associations would interact frequently. To test this hypothesis, we fitted negative binomial regression models to evaluate the relationship between the number of eGenes and CIF between two loci at the 40Kb bin resolution. In our analysis, we only considered chromatin interactions between different bins, and eQTL-gene pairs that fall into different bins (see Methods). After adjusting for genomic distance between loci, the number of eGenes showed significantly positive effects on CIF across all tissues and cell lines (Fig. 2a). For example, in spleen, the effect of the number of eGenes is estimated to be 0.20 (p value $< 2.2 \times 10^{-16}$), indicating that CIF would be 1.22 ($= e^{0.20}$) times higher for every extra eGene in a 40Kb bin pair. The magnitude of the effects varies across tissues and cell lines, ranging from 0.02 to 0.20. As expected, the effect sizes are similar for tissues from the same organ, such as the two brain tissues, DLPFC and hippocampus. Moreover, genomic distance has a significant negative effect on chromatin interaction and the effects are similar across all tissues and cell lines. This is expected because CIF between two genomic regions tends to decrease as their genomic distance increases [16].

Lieberman-Aiden et al. [16] have discovered the A and B compartments, which are associated with relatively high and low gene density, respectively, and showed that two regions within the same type of compartment (A vs.



A or B vs. B) have a higher CIF than regions within different types of compartment (A vs. B). This result indicates that gene density may play a role in Hi-C interaction, and motivated us to add the absolute difference in the number of tested genes between the two bins as a covariate in our regression model (Fig. 2b). As expected, the difference in gene density has a significant negative effect on CIF in all tissues and cell lines. The difference in gene density also has little correlation with the number of eGenes (Pearson correlation ≤ 0.06 in all tissues and cell lines). Moreover, the number of eGenes is still positively correlated with CIF, and its estimated effects are slightly increased compared to the model without the difference in gene density. Furthermore, we also stratified the data by the difference in gene density (Additional file 1: Figure S1). The difference in the number of eGenes varied from 0 to 8 and the sample sizes were relatively small for the strata with difference > 5 (Additional file 1: Table S2). For most strata, the results showed similar patterns as those in Fig. 1; in some strata, the estimated effects for the number of eGenes were negative but they were not statistically significant (Additional file 1: Figure S1).

We also repeated the analyses with the number of eGenes replaced by the number of genes without any associated eQTL and by the number of genes not expressed in the corresponding tissue or cell line (Additional file 1: Figure S2). The latter two had mostly opposite effects from the number of eGenes. For spleen and GM12878, the effects of the number of genes without eQTL were positive, but at a much smaller magnitude than those for the number of eGenes. These results clearly demonstrate that CIF is associated with the number of eGenes, not the total number of genes.

To ensure the conclusions are not sensitive to the choice of models, we also performed alternative analyses by taking the log-transformation of covariates (see Methods). The results still support the conclusion of positive association between CIF and the number of eGenes (Additional file 1: Supplementary Materials).

eQTL-gene associations are enriched in TADs

Since genomic regions within the same TAD are known to interact more frequently than those in different TADs [24, 25], we next examined whether

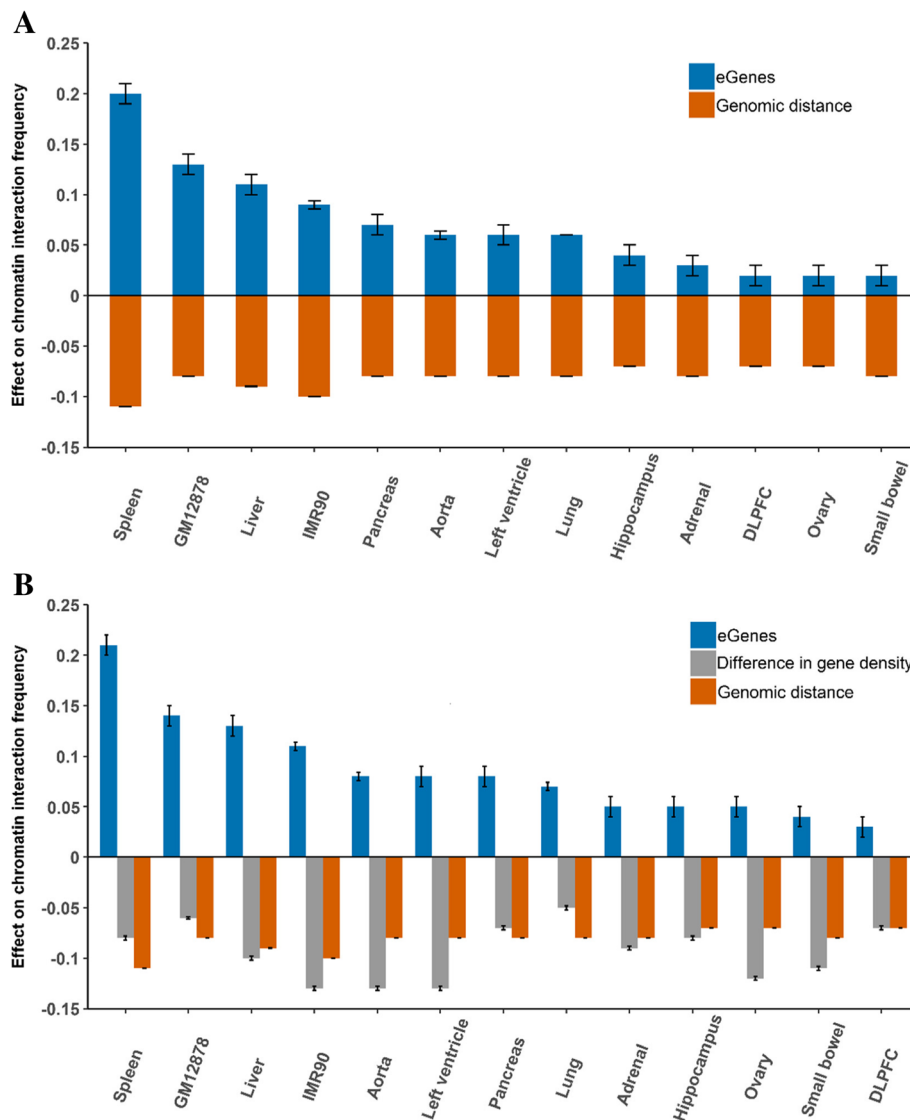


Fig. 2 Effect of the number of eGenes on CIF. Two negative binomial regression models were fitted to estimate the effect of the number of eGenes (blue) on CIF, one with adjustment for genomic distance (orange) (a), and the other with adjustment for both genomic distance (orange) and the difference in gene density (grey) (b). The error bars are \pm standard error

eQTL-gene associations are enriched within TADs. For each tissue and cell line, we simulated a pseudo SNP-gene pair to match every real eQTL-gene pair by keeping the location of the TSS of the gene but flipping the SNP position to the opposite side of the TSS (details in Methods). Most eQTL-gene pairs stayed in the same TAD after flipping, but a significant number of them changed from being inside the same TAD to falling in different TADs (range 10–15% across tissues and cell lines), while none changed the other way (McNemar’s test p -value $< 2.2 \times 10^{-16}$ for all tissues and cell lines). The real data also had a significantly higher fraction of eQTL-gene pairs falling in the same TAD than the simulated data (Fisher’s exact test

p -value $< 2.2 \times 10^{-16}$ for all tissues and cell lines; Fig. 3a). For example, 74.0% of the real eQTL-gene associations and 62.8% of the simulated pairs in GM12878 were inside TADs.

The associations also hold when we take the genomic distance into account, both in a joint analysis with genomic distance as a covariate and in stratified analyses with genomic distance as the stratifying variable (see Methods). In the logistic regression, the odds for a SNP-gene pair to be in the same TAD is significantly higher for the real data than for the pseudo data, and the results are consistent across all tissues and cell lines (Fig. 4). The results for our stratified analyses are in Fig. 3b (for GM12878 and hippocampus) and Additional file 1: Figure S4.

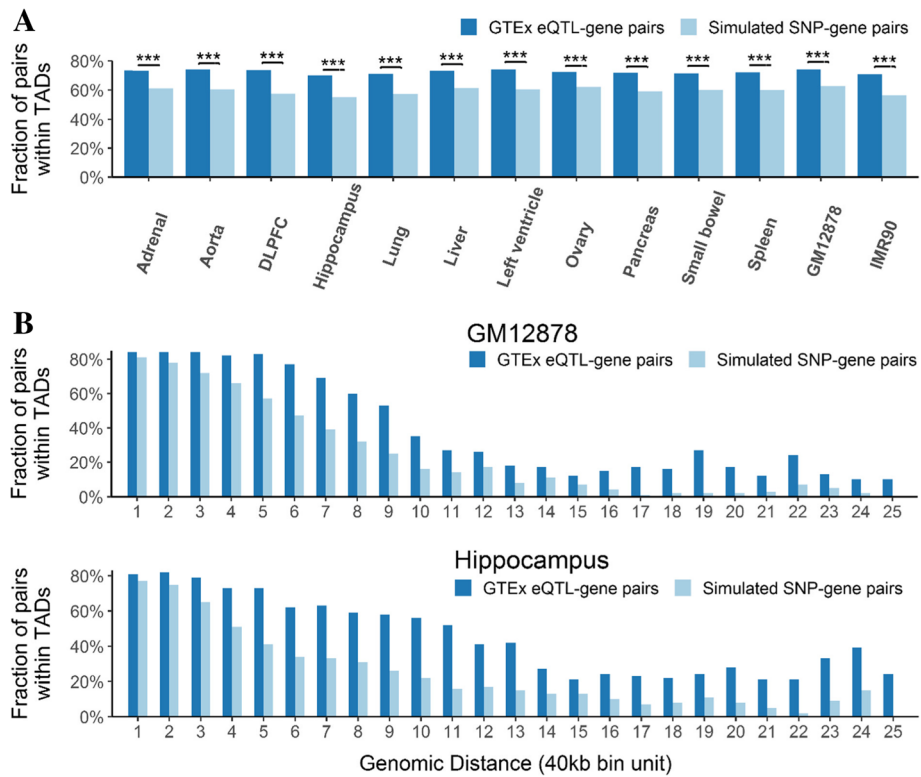


Fig. 3 Enrichment of eQTL-gene associations in TADs. (a) The fraction of SNP-gene pairs within TADs for the real data (dark blue) and the simulated data (light blue). (***) $p < 0.001$ for Fisher’s exact test. (b) Two examples, GM12878 and hippocampus, showing more detailed results after stratifying by the distance between eQTL and TSS of target genes (x-axis, in 40kb). All comparisons have $p < 0.001$

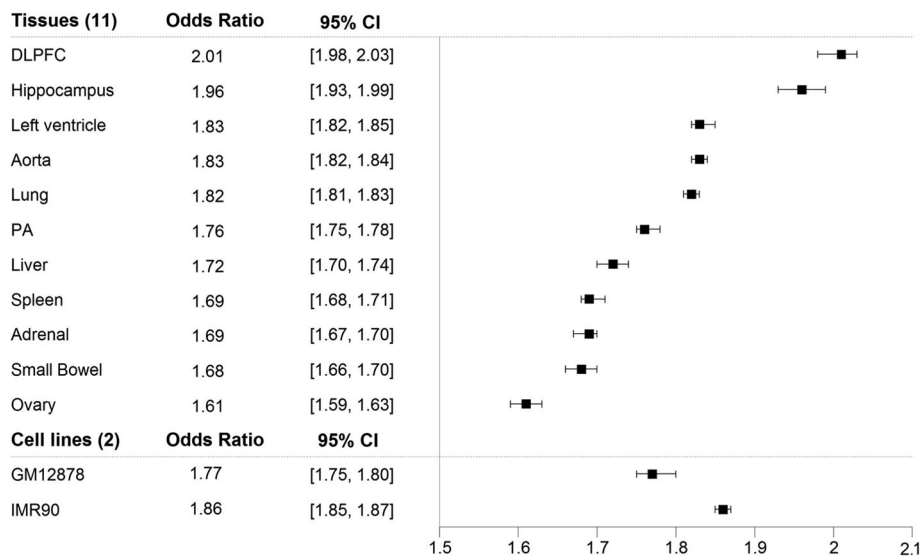


Fig. 4 Odds ratio and 95% confidence interval (CI) for SNP-gene pairs mapping at the same TAD between real and pseudo data

Tissue-specific eQTLs are enriched in tissue-specific FIREs in the majority of tissues

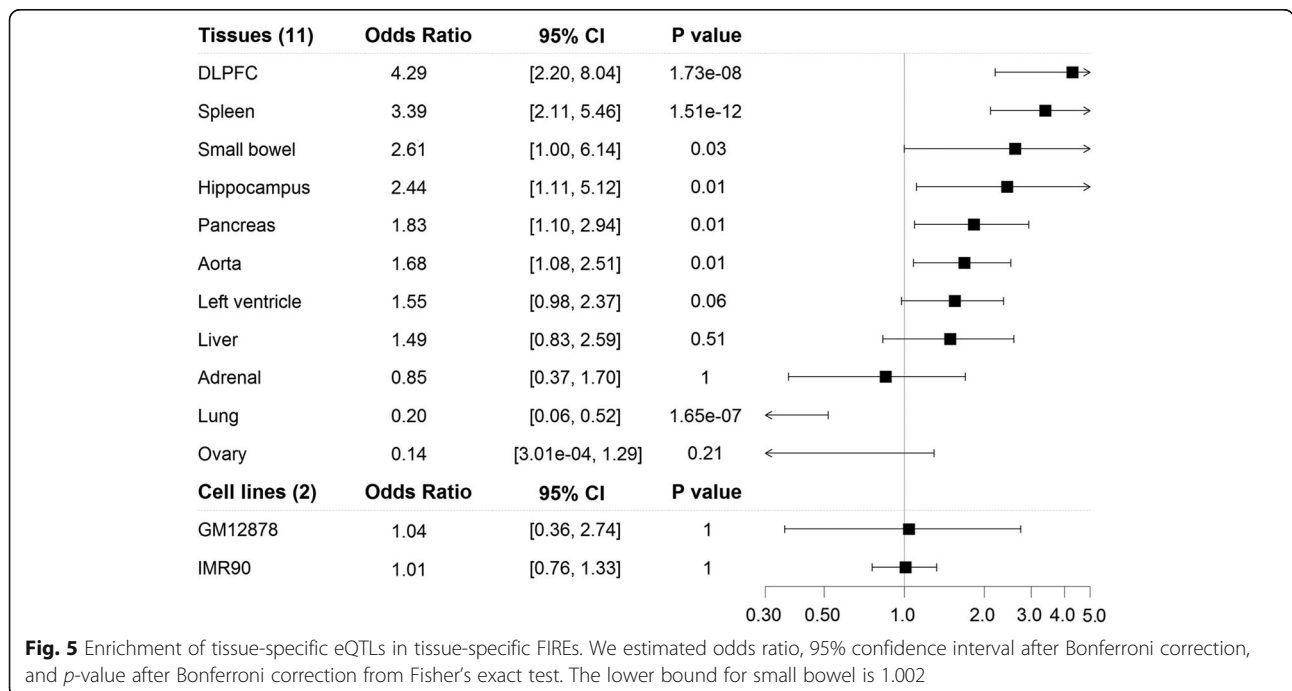
Since eQTLs also have high tissue specificity [3], we then examined whether tissue-specific eQTLs are enriched in tissue-specific FIREs. For the 11 tissues we considered, a total of 349,311 eQTLs were tissue specific. Eight (73%) of the 11 tissues had estimated odds ratio > 1, indicating enrichment of tissue-specific eQTLs in tissue-specific FIREs. Among them, six (55%) tissues (DLPFC, spleen, small bowel, hippocampus, pancreas and aorta) were statistically significant after Bonferroni correction (Fig. 5), and left ventricle almost met the cut-off. This significant enrichment of tissue-specific eQTLs in tissue-specific FIREs suggest that tissue-specific eQTLs may function through chromatin interactions that are also tissue specific, at least in the six tissues that were significant. However, a significant negative association was found in lung. There could be multiple reasons for this result: 1) a more complicated relationship might exist between eQTLs and chromatin spatial organization in lung than in some other tissues; 2) the biospecimen used to generate the eQTL and Hi-C data were collected from different people and might have been sampled from different locations of the lung; 3) those samples might be heterogeneous, consisting of different ratios of cell types. Further experimental data will be needed to help evaluate these potential factors. The results for other tissues and the two cell lines were not significant (Fig. 5). Although the results did not have a consistent direction across all tissues, they did suggest that more tissues may have a positive association

between tissue-specific eQTLs and tissue-specific FIREs than a negative association.

Discussion

Chromatin spatial organization and eQTLs are known to be involved in gene regulation. In this work, we systematically studied the relationship between eQTL-gene associations and chromatin interactions across 11 tissues and 2 cell lines. We found that CIF is positively associated with the number of eGenes. Moreover, we found that eQTL-gene associations are enriched in TADs. These results hold in all tissues and cell lines we evaluated. While these results may have been widely expected to hold, we have now provided solid statistical evidence across multiple tissues. These results suggest that eQTLs may regulate their target genes through chromatin interactions.

We also found that in six of the 11 tissues, tissue-specific eQTLs are significantly enriched in tissue-specific FIREs. This result suggests that tissue-specific eQTLs may function through chromatin interactions that are also tissue specific, at least in those six tissues. However, lung showed a significant negative association. This might be due to more complicated mechanisms in lung, or heterogeneity in sample sources, location, or tissue cell types because we matched the two sources of data simply by tissue name. The potential heterogeneity might have introduced noise in our analyses, and some of our results might have been stronger if the data had been more homogeneous.



The relationship between tissue-specific eQTLs and tissue-specific chromatin interactions is helpful for identifying genes regulated by eQTLs through chromatin interactions in the corresponding tissue. For example, the brain cortex tissue DLPFC has 2954 tissue-specific eQTLs and 323 tissue-specific FIREs. When both factors were considered, we identified 32 DLPFC-specific eQTLs located in those DLPFC-specific FIREs. These eQTLs are significantly associated with 4 genes, including *ADGRB2* (adhesion G protein-coupled receptor B2), *WASF3* (WAS protein family member 3), *SPEF2* (sperm flagellar 2), and *XPA* (xeroderma pigmentosum complementation group A). Among these genes, *ADGRB2*, which encodes a transmembrane signaling receptor [25], has a brain-specific developmental expression pattern and its expression level is increased as the development of the brain progresses [26]. The TSS of this gene (chr1:32,192,718) is ~47Kb from a DLPFC-specific FIRE (chr1:32,240,000-32,320,000).

Due to the relatively low depth in the Hi-C data currently available for multiple tissues, we performed our analyses at the 40Kb resolution to avoid data sparsity. It would be ideal to perform the analyses at a higher resolution. However the availability of high-resolution Hi-C data for tissues is currently very limited.

The eQTL results used in our study were available only for SNP-gene pairs that are within 1 Mb distance. The power of the eQTL analysis is largely determined by sample size. Because of these issues, we might have missed some SNP-gene associations in our analyses. In addition, our TAD enrichment analysis did not account for linkage disequilibrium (LD) between eQTLs. The effects of LD, if any, are probably canceled out between the real and simulated datasets.

Conclusions

In summary, we have demonstrated the close spatial proximity between eQTLs and their target genes across multiple human primary tissues. These results help us further understand the complementary effects of chromatin interactions and eQTLs in gene regulations.

Methods

Data description

We used Hi-C and eQTL data of 11 primary human tissues and 2 cell lines from Schmitt et al. [23] and the GTEx project [3], including the lymphoblastoid cell line GM12878, the fetal lung fibroblast cell line IMR90, and adrenal, aorta, dorsolateral prefrontal cortex (DLPFC), hippocampus, left ventricle, liver, lung, ovary, pancreas, small bowel and spleen tissues (Additional file 1: Table S1). We focus on the autosomes in all our analyses. The reference genome is hg19.

The Hi-C data contained over 2.9 billion raw intra-chromosomal unique read pairs on the 13 samples, out of which >1 billion have distance >15Kb. We used 40Kb bin resolution for the Hi-C data. We also downloaded the information for TAD boundaries and FIREs from Schmitt et al. [23]. On average, there are 2068 TADs and 3681 FIREs per sample. Schmitt et al. [23] study also published RNA-seq data for each tissue sample, measured by FPKM values. In the GTEx study, all tested SNP-gene pairs were within 1 Mb distance [3]. Details of data preprocessing are in the Additional file 1: Supplementary Materials.

Regression analysis of chromatin interaction frequency

We first evaluated the relationship between CIF and eQTL results using regression analysis. We mapped every SNP-gene pair tested in the GTEx study to a bin pair. SNP-gene pairs mapped to the same 40Kb bin were excluded from our analysis; that is, if a tested gene falls in one bin, it must have a corresponding SNP or eQTL in the other bin. We defined the following features for every bin pair (i, j), where $i < j$: 1) the CIF (I_{Hi-C}), 2) the number of eGenes with TSS in bin i or bin j (G_{eGene}), 3) the number of tested genes with TSS mapped to bin i (G_i), 4) the number of tested genes with TSS mapped to bin j (G_j), and 5) the genomic distance between bin i and bin j ($D=|i-j|$). We focused on bin pairs that contain at least one tested gene (i. e., $G_i + G_j > 0$). Since in the GTEx study, all tested SNP-gene pairs were within 1Mb distance, the bin pairs in our analysis also had distance ≤ 1 Mb. In addition, because the samples in the two original studies came from different tissue sources, we focused on genes that expressed in both sources, specifically, genes that were tested in the GTEx study [3] and had FPKM > 1 in the corresponding tissue or cell line in Schmitt et al. [23].

We performed negative binomial regression of CIF on the number of eGenes in each tissue or cell line. Since the CIF between two loci is known to be affected by their genomic distance [16], we included distance as a covariate in our model:

$$\ln(I_{Hi-C}) \sim G_{eGene} + D \quad (1)$$

In addition, it is known that chromatin interactions are less frequent between a gene-dense compartment and a gene-poor compartment than those within the same compartment [16]. While the compartments are defined on multi-Mb scale, this result indicates that gene density may play a role in Hi-C interaction. We thus further adjusted for the unevenness in the distributions of genes between two bins. Specifically, we added $G_{Diff} = |G_i - G_j|$ to our model as another covariate:

$$\ln(I_{Hi-C}) \sim G_{eGene} + G_{Diff} + D \quad (2)$$

We also conducted stratified analyses over subsets stratified by G_{Diff} .

In Results we showed that there were significant positive associations between CIF and the number of eGenes. To further evaluate whether these associations are truly due to eGenes, we repeated the regression analysis for the number of non-eGenes (defined as tested genes without any eQTLs), and compared its effect with the effect of eGenes. We also performed a regression analysis for the number of not expressed genes (defined as genes that had FPKM ≤ 1 in Schmitt et al. [23] and not tested in GTEx in the corresponding tissue or cell line). Specifically, we performed the following regression analyses:

$$\ln(I_{Hi-C}) \sim G_{non-eGene} + G_{Diff} + D \quad (3)$$

$$\ln(I_{Hi-C}) \sim G_{not\ expressed\ gene} + G_{Diff} + D \quad (4)$$

where $G_{non-eGene}$ is the number of non-eGenes and $G_{not\ expressed\ gene}$ is the number of not expressed genes.

We also performed sensitivity analysis by fitting alternative regression models where the input variables enter the models on the log scale or as categorical variables (details in Additional file 1: Supplementary Materials, and in Additional file 2: Additional Results).

Enrichment analysis of eQTL-gene associations in TADs

We next evaluated if eQTL-gene associations are enriched in TADs for all the tissues and cell lines we considered. For each tested SNP-gene pair, we created a matched pseudo pair as a control: we kept the gene's TSS position but flipped the position of SNP to be on the opposite side of the TSS but with the same distance from the TSS. For example, if the SNP is 93Kb downstream of the TSS, the flipped position will be 93Kb upstream of the TSS. The real SNP-gene pairs and the pseudo SNP-gene pairs have the same overall distribution of gene locations and same overall distribution of SNP-TSS distance. If the flipped position fell outside of the chromosome, both the real and the matched pseudo pair were removed from analysis.

We categorized SNP-gene pairs by two features: whether the pair is a real pair and whether the SNP and the gene's TSS are in the same TAD. We then performed McNemar's test on the resulting 2×2 table to detect whether there was a significantly higher probability for SNP-gene pairs to change from being inside the same TAD to falling in different TADs after flipping the position of SNP than the opposite change. We also performed Fisher's exact test to evaluate the association between these two features.

In addition, we took distance into account by performing logistic regression $Y \sim X + D$, where Y indicates whether a SNP-gene pair is in the same TAD, X indicates whether the pair is real, and D is the distance between the SNP and the TSS of the gene. We also performed stratified analyses by stratifying the data by genomic distance ranging from 40Kb to 1 Mb.

Analysis of tissue-specific FIREs and tissue-specific eQTLs

Finally, we investigated the tissue specificity of Hi-C data. We defined tissue-specific FIREs and studied their enrichment surrounding genes and their association with tissue-specific eQTLs. For each of the 11 tissues, we defined tissue-specific FIREs as those FIREs detected only in that tissue and not in any of the other 10 tissues. Tissue-specific eQTLs were similarly defined using the GTEx meta-analysis results (details in Additional file 1: Supplementary Materials). For GM12878 and IMR90, cell line-specific FIREs and eQTLs were similarly defined using all 13 samples we considered. For example, GM12878-specific FIREs are the FIREs detected only in GM12878 and not in any of the other 12 samples.

For each tissue, we examined whether tissue-specific eQTLs are enriched in tissue-specific FIREs. For each tissue and cell line, we counted the number of eQTLs according to whether the eQTL is tissue specific and whether it falls in a tissue-specific FIRE. We computed the odds ratio as the ratio of the fraction of tissue-specific eQTLs mapped to tissue-specific FIREs to the fraction of those mapped to the other FIREs of the tissue. We also computed Bonferroni corrected p -value and confidence interval at the 95% level after Bonferroni correction (i.e. 99.54% nominal level, where $0.9954 = 1 - 0.05/11$).

Additional files

Additional file 1: Supplementary description, tables, and figures. (DOCX 1137 kb)

Additional file 2: Additional results. (XLSX 18 kb)

Abbreviations

ADGRB2: Adhesion G protein-coupled receptor B2; CIF: Chromatin interaction frequency; CRE: Cis-regulatory element; DLPFC: Dorsolateral prefrontal cortex; eQTLs: Expression quantitative trait loci; FIRE: Frequently interacting region; FPKMs: Fragments per kilobase of exon per million reads; GTEx: Genotype-Tissue Expression study; Hi-C: High-throughput chromatin conformation capture; Kb: Kilobase; LD: Linkage disequilibrium; Mb: Megabase; SNP: Single nucleotide polymorphism; *SPEF2*: Sperm flagellar 2; TAD: Topologically associating domain; *WASF3*: WAS protein family member 3; *XPA*: Xeroderma pigmentosum complementation group A

Acknowledgements

Not applicable.

Funding

MH was partially supported by NIH grant U54DK107977.

Availability of data and materials

The Hi-C data were downloaded from the GEO (GEO accession number: GSE87112); eQTL results were downloaded from GTEx Portal (Version 7).

Authors' contributions

JY performed computational analyses, generated figures and drafted the manuscript. CL and MH initiated and guided the research. JY and CL wrote the manuscript and MH critically edited it. All authors read and approved the final manuscript.

Ethics approval and consent to participate

All data and materials used in this paper were obtained in de-identified forms from their sources.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA. ²Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH, USA. ³Cleveland Institute for Computational Biology, Cleveland, OH, USA.

Received: 28 January 2019 Accepted: 16 April 2019

Published online: 30 April 2019

References

- Cheung VG, Conlin LK, Weber TM, Arcaro M, Jen KY, Morley M, et al. Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat Genet.* 2003;33(3):422–5.
- The Genotype-Tissue Expression (GTEx) pilot analysis. Multitissue gene regulation in humans. *Science.* 2015;348:648–60.
- GTEx Consortium, et al. Genetic effects on gene expression across human tissues. *Nature.* 2017;550(7675):204–13.
- Gilad Y, Rifkin SA, Pritchard JK. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.* 2008;24(8):408–15.
- Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet.* 2015;16(4):197–212.
- Gaffney DJ, Veyrieras JB, Degner JF, Pique-Regi R, Pai AA, Crawford GE, et al. Dissecting the regulatory architecture of gene expression QTLs. *Genome Biol.* 2012;13(1):R7.
- Brown CD, Mangravite LM, Engelhardt BE. Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. *PLoS Genet.* 2013;9(8):e1003649.
- Heidari N, Phanstiel DH, He C, Grubert F, Jahanbani F, Kasowski M, et al. Genome-wide map of regulatory interactions in the human genome. *Genome Res.* 2014;24(12):1905–17.
- Dean A. In the loop: long range chromatin interactions and gene regulation. *Brief Funct Genomics.* 2011;10(1):3–10.
- Bonev B, Cavalli G. Organization and function of the 3D genome. *Nat Rev Genet.* 2016;17(11):661–78.
- Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell.* 2012;148(1–2):84–98.
- Hwang YC, Zheng Q, Gregory BD, Wang LS. High-throughput identification of long-range regulatory elements and their target promoters in the human genome. *Nucleic Acids Res.* 2013;41(9):4835–46.
- Tolhuis B, Palstra R-J, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active β -globin locus. *Mol Cell.* 2002;10(6):1453–65.
- Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell.* 2014;159(7):1665–80.
- Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science.* 2002;295:1306–11.
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long range interactions reveals folding principles of the human genome. *Science (New York, NY).* 2009;326(5950):289–93.
- Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature.* 2013;503(7475):290–4.
- Duggal G, Wang H, Kingsford C. Higher-order chromatin domains link eQTLs with the expression of far-away genes. *Nucleic Acids Res.* 2014;42(1):87–96.
- Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, Cairns J, Wingett SW, Várnai C, Thiecke MJ, Burden F. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell.* 2016 Nov 17;167(5):1369–84.
- He B, Chen C, Teng L, Tan K. Global view of enhancer-promoter interactome in human cells. *Proc Natl Acad Sci.* 2014;111(21):E2191–9.
- Ron G, Globerson Y, Moran D, Kaplan T. Promoter-enhancer interactions identified from Hi-C data using probabilistic models and hierarchical topological domains. *Nat Commun.* 2017;21;8(1):2237.
- Das A, Morley M, Moravec CS, Tang WH, Hakonarson H, Ashley EA, Brandimarto J, Hu R, Li M, Li H, Liu Y. Bayesian integration of genetics and epigenetics detects causal regulatory SNPs underlying expression variability. *Nat Commun.* 2015 Oct 12;6:8555.
- Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* 2016;17(8):2042–59.
- Dixon JR, Gorkin DU, Ren B. Chromatin domains: the unit of chromosome organization. *Mol Cell.* 2016;62(5):668–80.
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature.* 2012;485(7398):376–80.
- Kee HJ, Koh JT, Kim M-Y, Ahn KY, Kim JK, Bae CS, et al. Expression of brain-specific angiogenesis inhibitor 2 (BAI2) in normal and ischemic brain: involvement of BAI2 in the ischemia-induced brain angiogenesis. *J Cereb Blood Flow Metab.* 2002;22(9):1054–67.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

