

spatzie: an R package for identifying significant transcription factor motif co-enrichment from enhancer–promoter interactions

Jennifer Hammelman ^{1,2,†}, Konstantin Krismer ^{2,3,†} and David K. Gifford ^{1,2,3,4,*}

¹Computational and Systems Biology Program, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA, ²Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar Street, Cambridge, MA 02139, USA, ³Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA and ⁴Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Received June 03, 2021; Revised January 07, 2022; Editorial Decision January 10, 2022; Accepted January 29, 2022

ABSTRACT

Genomic interactions provide important context to our understanding of the state of the genome. One question is whether specific transcription factor interactions give rise to genome organization. We introduce *spatzie*, an R package and a website that implements statistical tests for significant transcription factor motif cooperativity between enhancer–promoter interactions. We conducted controlled experiments under realistic simulated data from ChIP-seq to confirm *spatzie* is capable of discovering co-enriched motif interactions even in noisy conditions. We then use *spatzie* to investigate cell type specific transcription factor cooperativity within recent human ChIA-PET enhancer–promoter interaction data. The method is available online at <https://spatzie.mit.edu>.

INTRODUCTION

Genome organization plays an important role in the function of the genome in development (1–3) and disease (4). Specific transcription factor cooperation is a potential explanation for the cell type specificity of genomic interactions, especially those that tether enhancers to promoters. Recent methods seek to detect such transcription factor cooperativity by generating models to predict enhancer–promoter interactions and measuring the importance of model features (5,6). However, these methods can be difficult to interpret due to the complexity of the computational models, the use of shrinkage techniques that could eliminate correlated features. Furthermore, an implementation of the

motif co-enrichment analysis is either not available (6) or not straightforward to use, as motif sites have to be provided by the user in a proprietary format (5).

Here, we introduce *spatzie*, an R/Bioconductor package named after the German diminutive for sparrow and inspired by the long-range geographical patterns of their songs (7), a reference to the long-range genomic interactions of transcription factor cooperativity. Within *spatzie* we implement a collection of statistical methods to identify transcription factor co-enrichment in experimental data obtained by protein-centric chromatin conformation methods such as ChIA-PET (8) and HiChIP (9). We demonstrate the utility of *spatzie* by discovering the co-enrichment of transcription factor binding motifs simulated from ChIP-seq data. Furthermore, we apply *spatzie* to investigate cell type-specific interactions from RAD21-targeted ChIA-PET experiments across 24 human cell lines.

MATERIALS AND METHODS

ChIP-seq data for simulated co-enrichment

We simulate a cooperative relationship where binding of USF1 at promoters is co-dependent on binding of ELF1 at enhancers. Raw ChIP-seq data for USF1 and ELF1 from MEL mouse cells was downloaded from ENCODE (Supplementary Table S1). Reads were trimmed for adaptors and low-quality positions using TrimGalore (Cutadapt v0.6.2) (10) and aligned to the mouse genome (mm10) with bwa mem (v0.7.1.7) (11) with default parameters. Duplicates were removed with samtools (v1.7.2) (12) markdup, and ChIP binding events were called with GPS (v3.4) (13) with default parameters.

*To whom correspondence should be addressed. Tel: +1 617 253 6039; Email: gifford@mit.edu

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

ChIA-PET datasets

ChIA-PET interaction data was downloaded as processed files from Grubert *et al.* (Supplementary Material). Raw data for all experiments is accessible from GEO (Supplementary Table S1).

Genomic annotations

For simulated data, mm10 promoter annotations were downloaded from the UCSC browser using the R package *GenomicFeatures* (14). For human RAD21 ChIA-PET, hg19 promoter ensemble gene annotations were also downloaded from the UCSC browser using the same method. Within interaction data, regions that were within 2.5 kb of a promoter were classified as promoter regions. All other regions were classified as gene-distal enhancer regions.

Statistical cooperativity calculation with spatzie

We implement three methods to measure the relationship between transcription factor binding motifs in enhancer and promoter regions of genomic interactions. Each method takes as input two vectors, $\mathbf{x} = (x_1, x_2, \dots, x_{n-1}, x_n)$ and $\mathbf{y} = (y_1, y_2, \dots, y_{n-1}, y_n)$, where n is the number of enhancer–promoter interactions. x_i is a set that contains all PWM scores for motif a in the enhancer region of interaction i . y_i , in contrast, contains those scores for motif b in the promoter region of interaction i .

Score-based correlation coefficient. We assume motif scores follow a normal distribution and are independent between enhancers and promoters. We can therefore compute how correlated scores of any two transcription factor motifs are between enhancer and promoter regions using Pearson's product–moment correlation coefficient:

$$r = \frac{\sum(x'_i - \bar{x}')(y'_i - \bar{y}')}{\sqrt{\sum(x'_i - \bar{x}')^2 \sum(y'_i - \bar{y}')^2}},$$

where the input vectors \mathbf{x} and \mathbf{y} from above are transformed to vectors \mathbf{x}' and \mathbf{y}' by replacing the set of scores with the maximum score for each region:

$$x'_i = \max x_i$$

x'_i is then the maximum motif score of motif a in the enhancer region of interaction i , y'_i is the maximum motif score of motif b in the promoter region of interaction i , and \bar{x}' and \bar{y}' are the sample means.

Significance is then computed by transforming the correlation coefficient r to test statistic t , which is Student t -distributed with $n - 2$ degrees of freedom.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

All P -values are calculated as one-tailed P -values of the probability that scores are greater than or equal to r .

Count-based correlation coefficient. Instead of calculating the correlation of motif scores directly, the count-based correlation metric first tallies the number of instances of a given

motif within an enhancer or a promoter region, which are defined as all positions in those regions with motif score P -values of less than 5×10^{-5} , which tends to work well for human and mouse motifs (15,16). Formally, the input vectors \mathbf{x} and \mathbf{y} are transformed to vectors \mathbf{x}'' and \mathbf{y}'' by replacing the set of scores with the cardinality of the set:

$$x''_i = |x_i|$$

And analogous for y''_i . Finally, the correlation coefficient r between \mathbf{x}'' and \mathbf{y}'' and its associated significance are calculated as described above.

Instance co-occurrence. Instance co-occurrence (or match association) uses the presence or absence of a motif within an enhancer or promoter to determine a statistically significant association, thus \mathbf{x}''' and \mathbf{y}''' are defined by:

$$x'''_i = \mathbb{1}_{x''_i > 0}$$

The significance of instance co-occurrence is determined by the hypergeometric test:

$$p = \sum_{k=I_{ab}}^{E_a} \frac{\binom{E_a}{k} \binom{n-E_a}{P_b-k}}{\binom{n}{P_b}},$$

where I_{ab} is the number of interactions that contain a match for motif a in the enhancer and motif b in the promoter, E_a is the number of enhancers that contain motif a ($E_a = \sum_i x'''_i$), P_b is the number of promoters that contain motif b ($P_b = \sum_i y'''_i$), and n is the total number of interactions, which is equal to the number of enhancers and to the number of promoters.

Multiple hypothesis testing. While the R package spatzie supports several methods to adjust P -values, three to control the family-wise error rate or FWER (Holm's method (17), Hochberg's method (18), Bonferroni's method (19)) and two to control the false discovery rate or FDR (Benjamini and Hochberg's method (20), and Benjamini and Yekutieli's method (21)), all P -values presented in this work were corrected using the method of Benjamini and Hochberg.

RESULTS

spatzie tests transcription factor motifs for co-enrichment in enhancer–promoter interactions

The goal of spatzie is to identify pairs of transcription factor motifs which have a relationship such that the presence of motif A in an enhancer is associated with the presence of motif B in the promoter, indicating these transcription factors may be cooperating to drive enhancer–promoter interactions (Figure 1A). Given an input of interacting genomic loci, we select only those interactions which contain one locus that is gene-distal, which we label enhancer, and one locus overlapping a gene transcription start site, which we label promoter. Then, we scan these regions for transcription factor motifs using a database of DNA-binding motifs identified by ChIP-seq experiments, such as HOCOMOCO (22), HOMER (23) or JASPAR (24). In order to limit hypothesis testing, spatzie provides a function to filter transcription

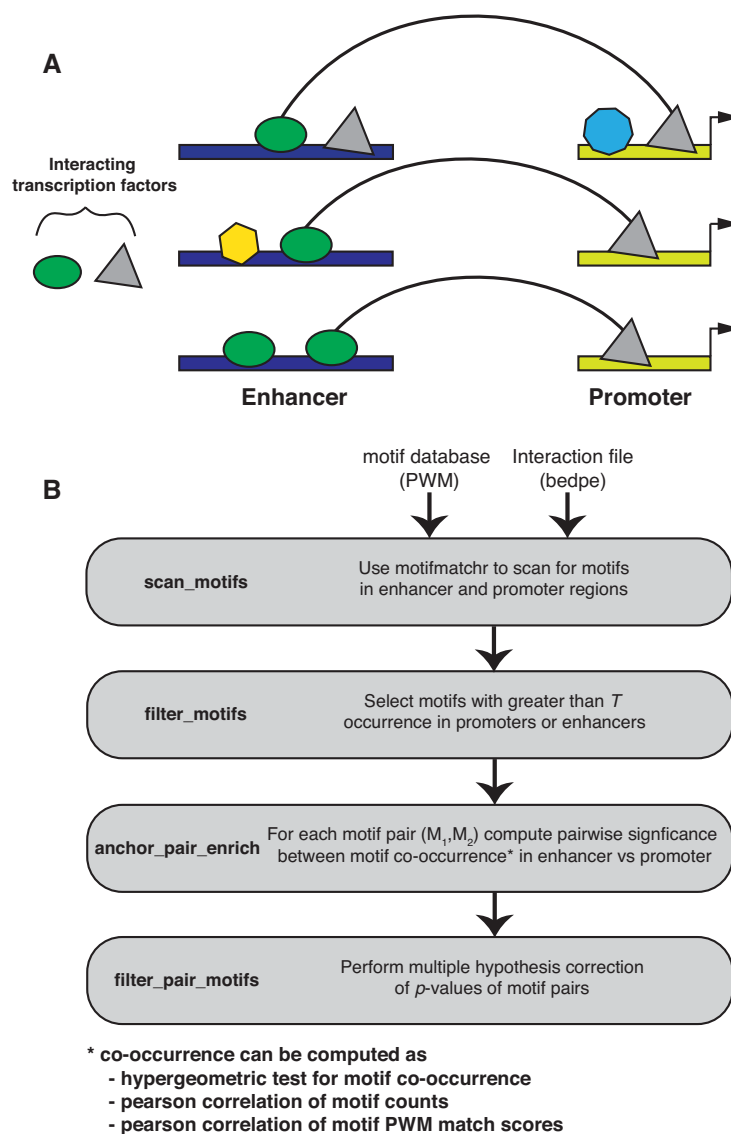


Figure 1. spatzie identifies motif pairs underlying enhancer–promoter interactions using co-occurrence and correlation statistics. (A) spatzie is designed to identify transcription factors which are facilitating interactions between enhancers and promoters based on detecting co-enrichment relationships between the presence of DNA-binding motifs in enhancer–promoter pairs. (B) Given input of a database of transcription factor DNA-binding motifs and a set of enhancer–promoter interactions, scan interactions for motifs that are frequently present within the interactions of interest. Next we compute pairwise significance of motif co-occurrence in the enhancer and promoter data. Finally, we filter motif pairs that significantly co-occur under multiple hypothesis correction.

factor motifs to those present in some threshold number of interactions. After filtering, we test transcription factor motifs pairwise for co-enrichment between enhancer and promoter pairs. Since the relationship between the DNA-binding motif and transcription factor activity is complex, spatzie provides three possibilities: (i) the strength of the transcription factor motif match (i.e. the PWM score), (ii) the number of motif sites within the sequence and (iii) the presence or absence of motif sites. These definitions result in three different statistical tests for co-enrichment: the significance tests of (i) score-based or (ii) count-based correlation coefficients and (iii) the hypergeometric test for co-occurrence over-representation (Figure 1B). Finally, we adjust the significance of these co-enrichment scores to ac-

count for multiple hypothesis testing and report significant transcription factor pairs.

spatzie identifies co-enrichment from simulated data

We validate spatzie by simulating a co-enrichment relationship where binding of USF1 at promoters is co-dependent on binding of ELF1 at enhancers. Using ELF1 and USF1 ChIP-seq data from ENCODE, we aligned and called binding events with GPS (13). We then filtered USF1 binding sites to those that overlapped annotated promoters and filtered ELF1 binding events to any event that did not overlap a promoter. Then, we matched the most significant USF1 promoter event to the most significant ELF1 en-

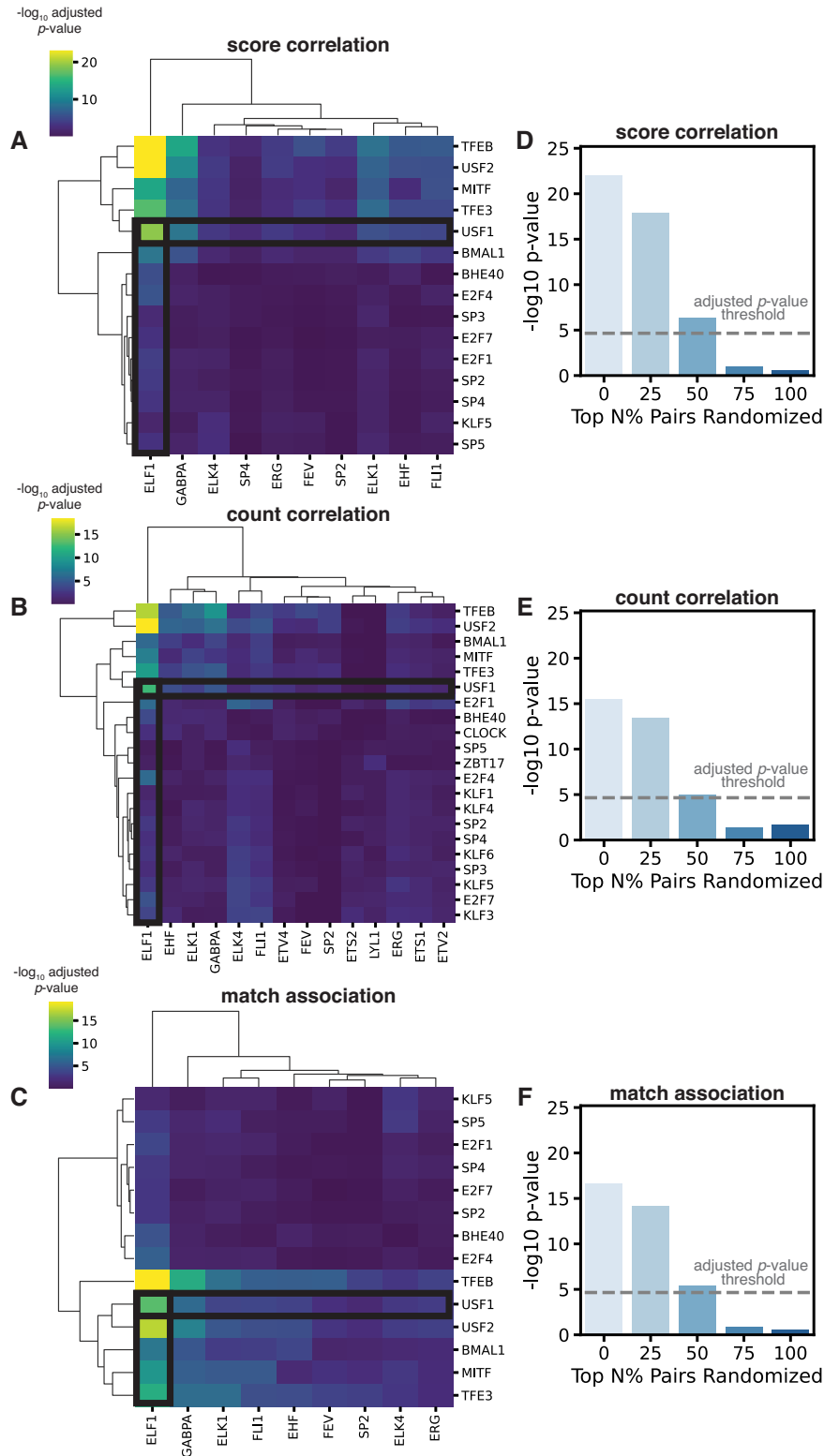


Figure 2. spatzie validates co-enrichment of ELF1 and USF1 on simulated enhancer–promoter interaction data. (A) spatzie cooperativity estimation computed using correlation of motif scores shows significant enhancer–promoter interactions for USF1 and ELF1 motifs. (B) spatzie cooperativity estimation computed using correlation of counts shows strongest enrichment between USF1 and ELF1 motifs. (C) spatzie cooperativity estimation computed using motif instance co-occurrence shows significant co-enrichment of USF2 and ELF1 motifs. Adjusted *P*-values were corrected with the Benjamini-Hochberg procedure. Randomization experiments where top *N*% of enhancer events are randomly permuted shows shrinking significance of co-enrichment under noisy data for (D) score correlation, (E) count correlation and (F) hypergeometric co-enrichment. Dashed line represents the significance threshold at $p < 0.05$ under Bonferroni correction.

hancer event, thus creating a simulated enhancer–promoter interaction data set where the strongest USF1 promoter events are matched with ELF1 enhancer events. We found that the three described methods (score-based correlation, count-based correlation, and motif presence/absence association) all result in significant co-occurrence between the USF1 motif and the ELF1 motif (Figure 2A–C). The DNA binding motifs of all other transcription factors with significant co-enrichment are highly similar to USF1 and ELF1 (Supplementary Figure S1).

We also test that spatzie performs under noisy experimental conditions. We take the top N percent of enhancer interactions and randomly permute them such that they are paired with new promoters and then run spatzie using score-based correlation, count-based correlation, and motif presence/absence association (Figure 2D–F). We find that all methods collapse after 75% and 100% of the enhancers have been randomly permuted, indicating that co-occurrence is a result of the strength of the co-occurrence of the DNA binding motifs underlying the simulated enhancer–promoter interaction data.

spatzie identifies germ layer and tissue-specific enhancer–promoter transcription factor interactions

Finally, we investigate enhancer–promoter interactions from RAD21 ChIA-PET of 24 human cell lines from ENCODE (25). Based on the most significant co-enrichment scores on simulated data coming from the score correlation method, we chose to use score correlation to investigate transcription factor motif co-occurrence in enhancer–promoter interactions. After evaluating with spatzie the score correlation of 50 286 enhancer–promoter interactions that were present in at least one cell type, we find interactions cluster by tissue and germ line (Figure 3A–C and Supplementary Figure S3 for scatterplots with PC1, PC2 and PC3), and that correlation of discovered motif pairs increases among cell types from the same germ layer and tissue (Figure 3D). While previous work has shown that cohesin-mediated genomic interactions are similarly stratified by germ layer and tissue (25), our analysis with spatzie shows that there is sufficient information within the co-enrichment of motifs underlying enhancer and promoter interactions to reproduce biologically meaningful germ layer and tissue layer organization. We found the tissue-level correlation between spatzie-discovered co-enriched motifs was reproducible with interaction calls using CID (Supplementary Figure S2), which was previously shown to recover more reproducible interactions from ChIA-PET data (26). We then extracted enhancer–promoter interactions that had the highest germ layer-specific expression and found that these include transcription factors such as Fox family members, which have a known role in endoderm development (27,28) and ectoderm development (29,30), and Nfatc4 in mesoderm interactions, which is a known T cell (31) and myogenic (32,33) differentiation factor. (Figure 3E). Similarly, we examined tissue-level specific interactions (Figure 3F) and found instances that were cell type specific, which is indicative of a potential trade-off in partnering of Sta5 at the enhancer with either Nfatc4 or Nanog at the promoter in blood or in liver tissues,

respectively, or of AP2B at the promoter with Nfia or Zsc31 at the enhancers in breast or blood tissues, respectively (Figure 3F). Supplementary Figure S4 shows examples of motif co-occurrence in enhancer–promoter interactions. Furthermore, Supplementary Table S2 contains results from a ChIA-PET reproducibility analysis using IDR2D (34), showing that the ChIA-PET experiments from two (Jurkat and MCF7) out of the three cell lines (Jurkat, MCF7 and LX) that do not cluster with other cell lines from the same germ layer (Figure 3A) suffer from poor reproducibility.

DISCUSSION

Overall, spatzie contributes to a growing field of tools for the analysis of enhancer–promoter interaction data by providing a collection of statistical tests to identify transcription factor motif co-enrichment. While other methods such as PEP-Motif (5) and the graphical lasso approach taken in Pliner *et al.* (6) may identify such co-enrichment relationships, they spend computational power to identify motifs that predict the activity of enhancers and promoters independent of their interactions, whereas spatzie focuses exclusively on identifying motifs which share a co-enrichment relationship between enhancer–promoter interactions. We validate spatzie on experimental data where we use real ChIP-seq data to simulate enhancer–promoter interactions between USF1 binding at promoters and ELF1 binding at enhancers. We show that spatzie’s three modes of motif pair relationship (motif score correlation, motif count correlation, and instance co-occurrence) all successfully identify USF1:ELF1 motif co-occurrence relationships even under noisy conditions, with motif score correlation achieving the most robust results. We also apply spatzie to data from 24 human cell lines and are able to show transcription factor co-enrichments that are discovered by spatzie cluster at germ-layer and tissue level, indicating these co-enrichment relationships are related to the organization of these cell types by germ layer and tissue. Furthermore, the identified germ layer and tissue-specific transcription factor interactions contain lineage-determining transcription factors, indicating that transcription factor co-enrichment between enhancers and promoters contains transcription factors that are known to play a role in differentiation and may point to their function as players in the structural organization of the genome. One concern with motif co-enrichment approaches is that the primary effects discovered can be attributed to the activity of cell type-specific transcription factors without a cooperative relationship. However, as evidence to the contrary we find examples of tissues that share enhancer or promoter motifs with different partners, indicating spatzie is identifying co-enrichment beyond general transcription factor activity. This combined with evidence that spatzie does not discover co-enrichment when we entirely randomize the relationship between binding for simulated interactions from USF1 and ELF1 ChIP-seq suggests that spatzie effects are not dominated by the general over-enrichment of motifs, but instead are based on a dependent relationship between a pair of transcription factors underlying enhancer–promoter interactions. In sum, we hope spatzie provides biological insight into the cell

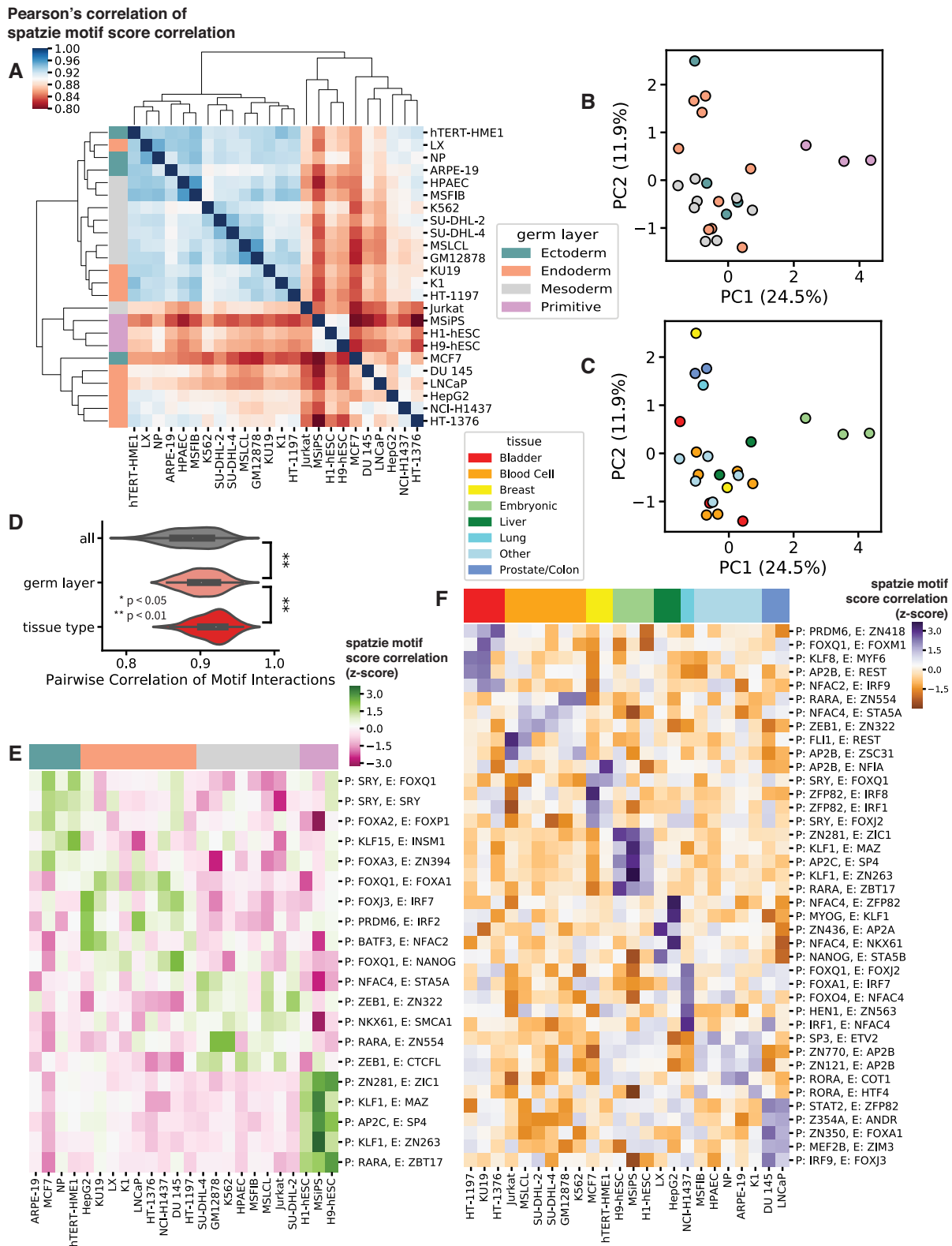


Figure 3. spatzie identifies transcription factor cooperation underlying interactions that are germ line and tissue-specific. (A) Pearson correlation of pairwise interactions shows similarity between related tissues. PCA on spatzie discovered transcription factor interactions shows (B) germ layer and (C) tissue type clustering. (D) Pairwise correlation of spatzie transcription factor motif interactions shows increasing relatedness of germ layer and tissue type. Significance computed by Wilcoxon rank-sum test. (E) Extraction of germ layer-specific transcription factor motif interactions include relevant lineage-determining transcription factor families, such as Fox in endoderm and ectoderm. Spatzie correlation scores are z-scores normalized by row. Each column represents an independent cell type, the colors above denote the germ layer. (F) Extraction of tissue-specific transcription factor motif interactions include potential transcription factor trade-offs at the promoter and enhancer that may mediate tissue-specific enhancer-promoter interactions. Spatzie correlation scores are z-scores normalized by row.

type-specific rules of transcription factor cooperativity underlying enhancer–promoter interactions.

DATA AVAILABILITY

The functionality of spatzie is bundled as an R/Bioconductor package with the same name (<https://doi.org/doi:10.18129/B9.bioc.spatzie>). The source code of the R package is hosted on GitHub (<https://github.com/gifford-lab/spatzie>). The core functionality is also available online at <https://spatzie.mit.edu>, which includes enhancer–promoter motif co-enrichment analysis with HOCOMOCO or user-defined motifs on interactions data mapped to either hg38, hg19, mm9 or mm10.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Michael Closser for helpful discussions in the conception of this project. We thank members of the Gifford and Wichterle labs for insightful suggestions and discussions.

Author contributions: Conceptualization, J.H.; Methodology, J.H. and K.K.; Software, J.H. and K.K.; Formal Analysis, J.H., K.K. and D.K.G.; Investigation, J.H. and K.K.; Resources, D.K.G.; Data Curation, J.H. and K.K.; Writing - Original Draft, J.H. and K.K.; Writing - Review & Editing, J.H., K.K. and D.K.G.; Visualization, J.H.; Supervision, D.K.G.; Funding Acquisition, J.H. and D.K.G.

FUNDING

NIH [1R01HG008754 to D.K.G., 1R01NS109217 to D.K.G.]; National Science Foundation Graduate Research Fellowship [1122374 to J.H.]. Funding for open access charge: MIT.

Conflict of interest statement. None declared.

REFERENCES

- Kosak, S.T. and Groudine, M. (2004) Form follows function: the genomic organization of cellular differentiation. *Genes Dev.*, **18**, 1371–1384.
- Dekker, J. (2008) Gene regulation in the third dimension. *Science*, **319**, 1793–1794.
- Lancôt, C., Cheutin, T., Cremer, M., Cavalli, G. and Cremer, T. (2007) Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat. Rev. Genet.*, **8**, 104–115.
- Chakraborty, A. and Ay, F. (2019) The role of 3D genome organization in disease: from compartments to single nucleotides. *Semin Cell Dev. Biol.*, **90**, 104–113.
- Yang, Y., Zhang, R., Singh, S. and Ma, J. (2017) Exploiting sequence-based features for predicting enhancer–promoter interactions. *Bioinformatics*, **33**, i252–i260.
- Pliner, H.A., Packer, J.S., McFaline-Figueroa, J.L., Cusanovich, D.A., Daza, R.M., Aghamirzaie, D., Srivatsan, S., Qiu, X., Jackson, D. and Minkina, A. (2018) Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol. Cell*, **71**, 858–871.
- Searfoss, A.M., chun Liu, W. and Creanza, N. (2020) Geographically well-distributed citizen science data reveals range-wide variation in the chipping sparrow's simple song. *Anim. Behav.*, **161**, 63–76.
- Fullwood, M.J. and Ruan, Y. (2009) ChIP-based methods for the identification of long-range chromatin interactions. *J. Cell. Biochem.*, **107**, 30–39.
- Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J. and Chang, H.Y. (2016) HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods*, **13**, 919–922.
- Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, **17**, 10–12.
- Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv doi: <https://arxiv.org/abs/1303.3997>, 26 May 2013, preprint: not peer reviewed.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
- Guo, Y., Papachristoudis, G., Altshuler, R.C., Gerber, G.K., Jaakkola, T.S., Gifford, D.K. and Mahony, S. (2010) Discovering homotypic binding events at high spatial resolution. *Bioinformatics*, **26**, 3028–3034.
- Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T. and Carey, V.J. (2013) Software for computing and annotating genomic ranges. *PLoS Comput. Biol.*, **9**, e1003118.
- Schep, A.N., Wu, B., Buenrostro, J.D. and Greenleaf, W.J. (2017) chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods*, **14**, 975–978.
- Korhonen, J., Martinmäki, P., Pizzi, C., Rastas, P. and Ukkonen, E. (2009) MOODS: fast search for position weight matrix matches in DNA sequences. *Bioinformatics*, **25**, 3181–3182.
- Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scand. J. Stat.*, **6**, 65–70.
- Hochberg, Y. (1988) A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, **75**, 800–802.
- Dunn, J. and Dunn, O.J. (1961) Multiple comparisons among means. *Am. Stat. Assoc.*, **56**, 52–64.
- Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B (Methodological)*, **57**, 289–300.
- Benjamini, Y. and Yekutieli, D. (2001) The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, **29**, 1165–1188.
- Kulakovskiy, I.V., Vorontsov, I.E., Yevshin, I.S., Sharipov, R.N., Fedorova, A.D., Rumynskiy, E.I., Medvedeva, Y.A., Magana-Mora, A., Bajic, V.B., Papatsenko, D.A. et al. (2018) HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.*, **46**, D252–D259.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
- Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranašić, D. et al. (2020) JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **48**, D87–D92.
- Grubert, F., Srivas, R., Spacek, D.V., Kasowski, M., Ruiz-Velasco, M., Sinnott-Armstrong, N., Greenside, P., Narasimha, A., Liu, Q., Geller, B. et al. (2020) Landscape of cohesin-mediated chromatin loops in the human genome. *Nature*, **583**, 737–743.
- Guo, Y., Krismer, K., Closser, M., Wichterle, H. and Gifford, D.K. (2019) High resolution discovery of chromatin interactions. *Nucleic Acids Res.*, **47**, e35.
- Faial, T., Bernardo, A.S., Mendjan, S., Diamanti, E., Ortmann, D., Gentsch, G.E., Mascetti, V.L., Trotter, M. W.B., Smith, J.C. and Pedersen, R.A. (2015) Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells. *Development*, **142**, 2121–2135.
- Cernilogar, F.M., Hasenoder, S., Wang, Z., Scheibner, K., Burtscher, I., Sterr, M., Smialowski, P., Groh, S., Evenroed, I.M., Gilfillan, G.D. et al. (2019) Pre-marked chromatin and transcription factor co-binding

- shape the pioneering activity of Foxa2. *Nucleic Acids Res.*, **47**, 9069–9086.
29. Monaghan,A.P., Kaestner,K.H., Grau,E. and Schutz,G. (1993) Postimplantation expression patterns indicate a role for the mouse forkhead/HNF-3 alpha, beta and gamma genes in determination of the definitive endoderm, chordamesoderm and neuroectoderm. *Development*, **119**, 567–578.
30. Neilson,K.M., Klein,S.L., Mhaske,P., Mood,K., Daar,I.O. and Moody,S.A. (2012) Specific domains of FoxD4/5 activate and repress neural transcription factor genes to control the progression of immature neural ectoderm to differentiating neural plate. *Dev. Biol.*, **365**, 363–375.
31. Macian,F. (2005) NFAT proteins: key regulators of T-cell development and function. *Nat. Rev. Immunol.*, **5**, 472–484.
32. Horsley,V. and Pavlath,G.K. (2002) NFAT: ubiquitous regulator of cell differentiation and adaptation. *J. Cell Biol.*, **156**, 771–774.
33. Bushdid,P.B., Osinska,H., Waclaw,R.R., Molkenin,J.D. and Yutzey,K.E. (2003) NFATc3 and NFATc4 are required for cardiac development and mitochondrial function. *Circ. Res.*, **92**, 1305–1313.
34. Krismer,K., Guo,Y. and Gifford,D.K. (2020) IDR2D identifies reproducible genomic interactions. *Nucleic Acids Res.*, **48**, e31.