A compendium of promoter-centered long-range chromatin interactions in the human genome

Inkyung Jung1,13*, Anthony Schmitt2,3,13, Yarui Diao2,4,13, Andrew J. Lee1, Tristin Liu2, Dongchan Yang5, Catherine Tan2, Junghyun Eom1, Marilynn Chan6, Sora Chee2, Zachary Chiang2, Changyoun Kim8,9, Eliezer Masliah8,9,10, Cathy L. Barr11, Bin Li1, Samantha Kuan2, Dongsup Kim5 and Bing Ren2,12*

A large number of putative cis-regulatory sequences have been annotated in the human genome, but the genes they control remain poorly defined. To bridge this gap, we generate maps of long-range chromatin interactions centered on 18,943 well-annotated promoters for protein-coding genes in 27 human cell/tissue types. We use this information to infer the target genes of 70,329 candidate regulatory elements and suggest potential regulatory function for 27,325 noncoding sequence variants associated with 2,117 physiological traits and diseases. Integrative analysis of these promoter-centered interactome maps reveals widespread enhancer-like promoters involved in gene regulation and common molecular pathways underlying distinct groups of human traits and diseases.

Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with human diseases and phenotypic traits, but molecular characterization of these genetic variants has been challenging because they are mostly noncoding and lack clear functional annotation. Recent studies have shown that these noncoding variants are frequently marked by chromatin signatures of cis-regulatory elements (cREs), leading to the hypothesis that a substantial fraction of variants may act by affecting transcriptional regulation. To formally test this hypothesis, it is critical to define the target genes of cREs in the human genome. However, inferring target genes of cREs based on linear genomic sequences is not straightforward, because cREs can regulate nonadjacent genes over large genomic distances. Such long-range regulation can take place because chromatin fibers are folded into a higher-order structure in which distant DNA fragments can be juxtaposed in space. Consequently, mapping spatial contacts between DNA has the potential to uncover target genes of cREs. To this end, chromosome conformation capture techniques such as 4C-seq, ChIA-PET and Hi-C, have been developed to determine chromatin interactions in a high-throughput manner. More recently, Hi-C combined with targeted capture and sequencing (capture Hi-C) has emerged as a cost-effective method to map chromatin interactions for specific regions at high resolution.

To systematically annotate candidate target genes for the cREs in the human genome, we performed capture Hi-C experiments (Fig. 1a and Supplementary Fig. 1) to interrogate chromatin interactions centered at well-annotated human gene promoters for 19,462 protein-coding genes (see Methods). We carried out these experiments with 27 human cell/tissue types including embryonic stem cells, four early embryonic lineages (mesoderm, mesenchymal stem cell, neural progenitor cells and trophoblast), two primary cell lines (fibroblast cells and lymphoblastoid cells) and 20 primary tissue types (hippocampus, dorsolateral prefrontal cortex, esophagus, lung, liver, pancreas, small bowel, sigmoid colon, thymus, bladder, adrenal gland, aorta, gastric tissue, left heart ventricle, right heart ventricle, right heart atrium, ovary, psoas, spleen and fat) for which reference epigenome maps have previously been produced as part of the Epigenome Roadmap project (Supplementary Fig. 2a and Supplementary Table 1). We designed and synthesized 12 capture probes for each promoter, six for each of the nearest HindIII restriction sites upstream and downstream of the transcription start site (TSS). Among 16,720 promoter-containing HindIII restriction DNA fragments, 14,357 (86%) contain a single promoter, but the 2,363 remaining HindIII fragments harbor multiple promoters (Supplementary Fig. 2b; see Methods). The robustness and the coverage of capture probe synthesis were validated by sequencing (Supplementary Fig. 2c–f). On average, each capture Hi-C experiment produced 65 million unique, on-target paired-end reads, yielding a total of 1.8 billion valid read pairs, ~30% of which were between DNA fragments >15 kilobase (kb) apart (Supplementary Table 2).

To identify the long-range chromatin interactions from the capture Hi-C data, two normalization steps were introduced. First, the biases in capture efficiency of each promoter (Supplementary Fig. 2g, h) were calibrated with the variable ‘capturability’ for each DNA fragment, defined as the fraction of total read counts mapped to the region, by use of a β-spline regression model (see Methods). Second, significant chromatin interactions were then identified after normalizing against the distance-dependent background signals (9% and 5% false discovery rate (FDR) for promoter–other and promoter–promoter interactions, respectively) and filtering for a minimum number of reads for mapping (100 reads) and overlap between interacting regions (15 kb).

1Department of Biological Sciences, KAIST, Daejeon, Korea. 2Ludwig Institute for Cancer Research, La Jolla, CA, USA. 3University of California, San Diego, Biomedical Sciences Graduate Program, La Jolla, CA, USA. 4Departments of Cell Biology and Orthopaedic Surgery, Regenerative Next Initiative, Duke University School of Medicine, Durham, NC, USA. 5Department of Bio and Brain Engineering, KAIST, Daejeon, Korea. 6University of California, San Francisco, San Francisco, CA, USA. 7Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA. 8Molecular Neuropathology Section, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. 9Department of Neurosciences, School of Medicine, University of California, San Diego, La Jolla, CA, USA. 10Department of Pathology, School of Medicine, University of California, San Diego, La Jolla, CA, USA. 11Krembil Research Institute, University Health Network, Toronto, and The Hospital for Sick Children, Ontario, Canada. 12Department of Cellular and Molecular Medicine, Center for Epigenomics, Institute of Genomic Medicine, and Moores Cancer Center, University of California, San Diego, La Jolla, CA, USA. 13These authors contributed equally: Inkyung Jung, Anthony Schmitt, Yarui Diao. E-mail: ijung@kaist.ac.kr; biren@ucsd.edu

*Corresponding author.
Focusing on the HindIII fragments over 15 kb away and within 2 megabase (Mb) of each promoter, we determined a total of 892,014 chromatin interactions (431,141 unique interacting pairs) in one or more of the 27 human cell/tissue types (Fig. 1b; Supplementary Fig. 3a; Supplementary Tables 3–5). A total of 18,943 promoter regions were involved in at least one significant chromatin interaction in one or more cell/tissue types analyzed in this study. The median distance between the interacting DNA pairs was 158 kb, which is within a similar range of previously reported chromatin loops and eQTL associations (Supplementary Fig. 3b and Supplementary Table 6). The slight discrepancy between promoter capture Hi-C (pcHi-C) interactions and eQTL associations may be attributed to different experimental approaches, but nevertheless, the two methods give complementary information to each other. Between 13% and 45% pcHi-C interactions detected in a cell or tissue type were unique to that cell/tissue type (Supplementary Fig. 3c). As expected, most of the detected chromatin interactions were within the topologically associating domains (TADs) defined in the corresponding tissue/cell type (Supplementary Fig. 3d,e).
To demonstrate that phC-Hi-C could effectively and reproducibly capture long-range chromatin interactions as detected by whole-genome in situ Hi-C, we compared the phC-Hi-C data with the in situ Hi-C data obtained from four distinct biosamples, including two cell lines (IMR90 lung fibroblast cell line and GM12878 lymphoblastoid cell line) and two primary tissues—dorsolateral prefrontal cortex and hippocampus (see Methods). Results of phC-Hi-C experiments accurately recapitulated chromatin loops identified from in situ Hi-C assays in all samples, with the area under the receiver operating curve (ROC) ranging between 0.84 and 0.91 (Supplementary Fig. 4a–e; see Methods). Additionally, we found high reproducibility of phC-Hi-C chromatin interactions between two biological replicates (average ROC score = 0.85; the average Spearman’s rank correlation between replicates = 0.4; Supplementary Fig. 4f–j and Supplementary Table 7; see Methods) and between two independent studies (Supplementary Fig. 4k). The observation that interactions identified in both replicates exhibited the strongest interaction signals, whereas interactions identified in one replicate were moderately strong but moderately weak in the other replicate (Supplementary Fig. 4l–m), suggests that the interactions that are specific to one replicate may be due to undersampling of the other replicate.

The chromatin interactome maps allowed us to assign candidate target genes for 70,329 putative cREs, defined based on H3K27ac signals in each tissue/cell type profiled previously, for 17,295 promoters. Each promoter was putatively assigned to 25 cREs on average (Supplementary Fig. 5a), and 45% of cREs were assigned to one candidate target gene (Supplementary Fig. 5b), similar to the previous observation with DNase I hypersensitivity analysis across diverse human cell types. We took advantage of the existing chromatin datasets collected for the same tissue/cell types and examined the relationship of the chromatin states between the cREs and the target promoters (see Methods). As expected, the fragments that extensively interact with multiple promoters were often found at active chromatin regions, such as transcription factor (TF) binding clusters or superenhancer regions (Supplementary Fig. 5c–i and Supplementary Tables 8–10; see Methods). Furthermore, integrative analysis with the ChromHMM model revealed that active promoters interact three times more frequently with DNA fragments harboring active enhancers than the bivalent promoters (Fig. 1c). However, the bivalent promoters interact five times more frequently with genomic regions associated with polycomb repressor complexes than the active promoters (Fig. 1c). Further analysis based on a refined 50-chromatin-state ChromHMM model for five cell lines also supports our conclusion (Supplementary Fig. 6).

Three lines of evidence support that the above promoter-centered chromatin interactions contain information on regulatory interactions at each promoter in the corresponding cell/tissue types. First, we compared the chromatin interactions at promoters with regulatory relationships inferred from expression quantitative trait loci (eQTL) in 14 matched tissue types that were recently reported by the GTEx consortium (see Methods) (Fig. 2a and Supplementary Fig. 7a–e). For each tissue and cell type, the previously reported eQTLs were highly enriched in the chromatin interactions identified in the corresponding tissue, with enrichment up to five-fold (ovary) (Supplementary Fig. 7d,e). A total of 42,627 eQTL associations were detected by P-0 phC-Hi-C chromatin interactions, whereas only 21,362 were expected by random chance after controlling for linear genomic distances (Supplementary Tables 11 and 12). Second, there is significant correlation between activities of cis-regulatory sequences and the assigned candidate target gene expression across multiple tissues and cell types, consistent with the purported regulatory relationships. Specifically, the levels of H3K27ac in these cREs were significantly correlated with both the promoter H3K27ac levels (Supplementary Fig. 8a) and transcription levels of the predicted target genes (Supplementary Fig. 8b) across these tissues/cell types. For example, POU3F3 expression (second column in Fig. 2b) was highly correlated with H3K27ac signals in the distal cRE (first column in Fig. 2b) connected by a tissue-specific chromatin interaction (last column in Fig. 2b). Lastly, cell/tissue-specific cRE–promoter pairs connected by phC-Hi-C interactions are significantly associated with active cREs and genes that are specific to the same cell/tissue types. For example, hippocampus-specific cRE–promoter chromatin interactions are significantly associated with active cREs (Fig. 2c) and highly expressed genes, albeit modestly, (Supplementary Fig. 8c) in hippocampus. Significant associations of cell/tissue-specific phC-Hi-C interactions in active cREs and highly expressed genes are found in other cell/tissue types as well (Fig. 2d–f; see Methods). The above results, taken together, strongly suggest that the predicted cRE–promoter pairs could uncover regulatory relationships between the cRE and target genes in diverse tissues and cell types.

Widespread promoter–promoter (P–P) interactions have been reported in cultured mammalian cells and a few primary tissues. The promoter-centered interaction maps obtained from 27 diverse tissues and cell types allowed us to test whether this is a general phenomenon. Indeed, consistent with previous reports, a significant fraction of the chromatin interactions was found between two promoters (9%, n = 79,989, Fisher’s exact test P < 2.2 × 10−16, Supplementary Fig. 9a). The physical proximity of these promoters is accompanied by a strikingly high correlation in chromatin modification state between the pair of promoters across diverse cell/tissue types (Fig. 3a,b). Previously, several promoter loci have been shown to function as enhancers to regulate distal genes. In support of the functional significance of enhancer-like promoters identified in the current study, 6,127 eQTLs match P–P interaction pairs, whereas only 2,722 eQTLs were expected by random chance (Fig. 3c). For instance, strong chromatin interactions were found between the DACT3 and AP2S1 promoter regions, and one significant eQTL, rs78730097 (NC_000019) for DACT3 was located in the AP2S1 promoter in the dorsolateral prefrontal cortex (Supplementary Fig. 10a). Notably, this eQTL does not show any meaningful genetic association with the adjacent downstream gene (AP2S1) or nearby genes but is exclusively associated with DACT3 (Supplementary Fig. 10b), suggesting regulatory potential of the AP2S1 promoter region in distal DACT3 regulation. To validate the function of enhancer-like promoters, we deleted two core promoter regions, where the downstream gene is not expressed but the promoter region shows active chromatin marks, using CRISPR-mediated system (Supplementary Fig. 10c,d and Supplementary Table 15; see Methods). Deletion of the ARHIZOS core promoter resulted in marked downregulation of the distal target gene (FDR-adjusted P = 0.02), NCKIPSD, identified by long-range chromatin interactions (Fig. 3d) with no significant or moderate effect on nearby genes (Supplementary Fig. 10e). Importantly, single guide RNA (sgRNA)-induced mutations in selected eQTLs proximal to transcriptional start sites demonstrated a significant downregulation effect on distal target genes but no significant effect on nearby gene expression in H1-hESC (Fig. 3e, Supplementary Fig. 10f and Supplementary Table 16; see Methods). Our results strongly suggest the genome-wide presence of enhancer-like promoters in the human genome and provide additional insight into their potential function in distal gene regulation.

The above promoter-centered chromatin interaction maps allowed us to infer the target genes of sequences harboring disease-associated variants and to understand the molecular basis of human disease. We focused on 42,633 putative disease/trait-associated genetic variants from a recent public repository of GWAS catalog. Consistent with previous reports, a significant portion
Fig. 2 | Inference of target genes of cis-regulatory sequences from pcHi-C data. a, Illustrative LocusZoom plot of eQTLs for VLDLR (top) and pcHi-C interactions in aorta tissue (bottom). Highlighted in translucent yellow are the VLDLR promoter and an eQTL connected by a pcHi-C interaction. Dots represent the P values of the SNPs’ association with VLDLR expression levels in the aorta (data obtained from GTEx). Dots are also color coded based on their linkage disequilibrium scores ($r^2$) with a tagging SNP. The blue bars indicate the recombination rate. b, Browser snapshots of the POU3F3 locus, showing positive correlation between the H3K27ac signals at a distal cRE (bottom left) and expression levels (bottom middle) of the promoter connected by long-range chromatin interactions (bottom right). The significant chromatin interaction between the POU3F3 promoter and a distal cRE is shown at the top (translucent yellow). c, Boxplots illustrating the H3K27ac signals at linked cREs ($n = 7,712$) connected by hippocampus (HC, colored by blue)-specific pcHi-C interactions. These cREs are marked by higher levels of H3K27ac in hippocampus than in other cell/tissue types (one-sided KS test $P < 0.005$). For the boxplots, the box represents the IQR and the whiskers correspond to the highest and lowest points within 1.5 × IQR. d–f, Heatmaps demonstrate the enrichment of pcHi-C interactions (column in d), z-score-transformed H3K27ac reads per kilobase million (RPKM) values at cREs (column in e) and z-score-transformed RNA-seq (fragments per kilobase of exon per million (FPKM) values at the cREs’ putative target genes (column in f) for given cell/tissue-specific cRE–promoter pairs in the corresponding cell/tissue type (rows in d–f). The KS test was performed between pcHi-C interaction frequencies, z-score-transformed H3K27ac RPKM values, and z-score-transformed RNA-seq FPKM values in the matched cell/tissue types (values in diagonal of each heatmap), and those in other cell/tissue types (values in off diagonal in each heatmap), demonstrating significant association of cRE–promoter pairs with cell/tissue-specific cRE H3K27ac signals and gene expression (two-sided KS test $P < 2.2 \times 10^{-16}$). H1, embryonic stem cell; ME, mesendoderm; MSC, mesenchymal stem cell; NPC, neural progenitor cell; HC, hippocampus; FC, dorsolateral prefrontal cortex; IMR90, fibroblast; LG, lung; LI, liver; PA, pancreas; SB, small bowel; TH, thymus; GA, gastric tissue; LV, left heart ventricle; PO, psoas; RV, right heart ventricle; SX, spleen; AD, adrenal gland; AO, aorta; LCL, lymphoblastoid.
of single nucleotide polymorphisms (SNPs; 30%, Fisher’s exact test $P < 2.2 \times 10^{-16}$) were found in putative cREs, emphasizing the importance of target gene identification of cREs in functional interpretation of disease-associated genetic variants. Because the causal SNPs are unknown in most cases, we also included SNPs that lie outside the previously defined cREs for further analysis. In total, each SNP was assigned to between one and three candidate target genes in each cell/tissue type, with the caveat that the precise importance of target predictions based on the promoter-centered chromatin interaction maps found to be the closest gene to the sequence variant (Supplementary Fig. 11d). To evaluate the validity of target predictions based on the promoter-centered chromatin interaction maps, we focused on seven GWAS variants that overlap with previously annotated cREs and eQTLs in the human lymphoblastoid cell line GM12878. We introduced deletions to these elements in GM12878 cells using CRISPR–Cas9 genome editing tools and examined the expression of predicted target genes, using quantitative PCR with reverse transcription (RT–qPCR) in the mutant cells and controls. For five of the seven tested cREs, genetic perturbation led to downregulation of the predicted distal target genes.
Fig. 4 | Analysis of human diseases and physiological traits based on the putative target genes of GWAS SNPs. a, Browser snapshots showing multiple cREs harboring GWAS SNPs (translucent yellow with a scissors symbol) and their common target gene, NT5DC2 (translucent green), together with signals of H3K27ac (ChIP-seq) and chromatin accessibility (DNase I) (left). The DNA fragments containing these cREs interact with the NT5DC2 gene promoter region as evidenced by pcHi-C analysis (arcs). The relative mRNA expression levels of NT5DC2 were quantified by RT–qPCR (right). Error bars indicate s.d. of two mutant clones with technical triplicates and n = 687, based on similarities of the putative target genes of trait-associated SNPs and SNPs in high linkage disequilibrium (LD). The color intensity of each dot indicates Pearson correlation coefficient (PCC) of the nearest gene similarity between GWAS-mapped traits ordered by hierarchical clusters (C1–C40). Each entry indicates –log10(P value) of gene ontology biological processes in the corresponding cluster obtained from DAVID. Several representative biological processes are highlighted.
Many diseases and traits could be linked to common molecular pathways, and the identification of these shared molecular pathways can be beneficial in understanding disease pathogenesis and developing treatment. To uncover the common molecular pathways underlying different diseases and physiological traits, we first determined the diseases/traits that share a significant number of common target genes predicted from their respective GWAS-associated SNPs. We grouped 687 traits and diseases into 40 clusters (Fig. 4b, Supplementary Fig. 12a–c and Supplementary Table 19; see Methods). Many physiological traits with known connections are found to be clustered together. For example, C5 clusters oxygen transport-related traits together; C6 groups together traits related to renal functions and C20 includes vascular function-associated traits (Fig. 4b). The above grouping is made possible thanks to the promoter-centered chromatin interactome maps, because the similarities among related traits observed in Fig. 4b were much less evident when we used either GWAS SNPs or nearest genes of the GWAS SNPs to compute the similarities as control experiments (Fig. 4c,d and Supplementary Fig. 12d). Our result suggests the power of target gene identification of GWAS variants to uncover trait-association networks.

To further understand the common molecular pathways affected in various human diseases, we carried out gene ontology analysis for the predicted target genes of the GWAS SNPs within each cluster (Supplementary Table 20; see Methods). The enriched gene ontology biological processes suggest potential shared molecular pathways for disease and trait types in each cluster (Fig. 4e, Supplementary Fig. 12e and Supplementary Table 21), including unexpected connections between specific traits. For example, C39 exposes a link between the susceptibility to infectious and autoimmune diseases and the risk of chemotherapeutic toxicity by carboplatin and cisplatin. In support of such link, a putative target gene associated with the response to carboplatin and cisplatin is \( ABCF1 \), which is involved in inflammatory response\(^3\). Although speculative, the shared molecular pathways uncovered by our analyses may provide new leads for investigation of the molecular basis of complex traits and disease phenotypes.

In summary, we have generated promoter-centered chromatin interactome maps across diverse human cell/tissue types. Our analysis covers a broad range of human tissue types and provides prediction of target genes for over 70,000 putative \( cис \)-regulatory elements and 27,000 GWAS SNV variants. This resource enables a systematic approach to understanding the molecular pathways dysregulated in distinct diseases and traits\(^1\). In future studies, delineation of disease-specific chromatin interactions with clinical samples by comparing our reference chromatin interaction maps could greatly improve the functional interpretation of many disease- and trait-associated genetic variants.

It should be noted that the current study only surveys a limited number of human tissues and cell types and assigned target genes for a small fraction of the putative \( cис \)-regulatory elements annotated in the human genome. Furthermore, the heterogeneous nature of the tissue samples used in this study prevents us from accessing the cell types in which the identified chromatin interactions occur, except for a few cell lines. Nevertheless, this resource lays the groundwork for further understanding of human disease pathogenesis and development of new treatment strategies.

**Online content**
Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41588-019-0494-8.

**Received:** 15 May 2019; **Accepted:** 31 July 2019; **Published online:** 9 September 2019

**References**
1. Welter, D. et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 42, D1001–D1006 (2014).
2. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190–1195 (2012).
3. Hindorff, L. A. et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* 106, 9562–9567 (2009).
4. Lettice, L. A. et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* 12, 1725–1735 (2003).
5. Ulu, V. V. et al. Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nat. Genet.* 46, 753–758 (2014).
6. Claussnitzer, M. et al. FTO obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* 373, 895–907 (2015).
7. Smemo, S. et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* 507, 371–375 (2014).
8. Yu, M. & Ren, B. The three-dimensional organization of mammalian genomes. *Annu. Rev. Cell Dev. Biol.* 33, 265–289 (2017).
9. de Wit, E. et al. The plurigene tumor genome in three dimensions is shaped around pluripotency factors. *Nature* 501, 227–231 (2013).
10. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* 489, 109–113 (2012).
11. Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* 518, 331–336 (2015).
12. Jin, F. et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* 503, 290–294 (2013).
13. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680 (2014).
14. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–382 (2012).
15. Tang, Z. et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* 163, 1611–1627 (2015).
16. Sahlen, P. et al. Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. *Genome Biol.* 16, 156 (2015).
17. Jager, R. et al. Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nat. Commun.* 6, 6178 (2015).
18. Mifsud, B. et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* 47, 598–606 (2015).
19. Dryden, N. H. et al. Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. *Genome Res.* 24, 1854–1868 (2014).
20. Martin, P. et al. Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. *Nat. Commun.* 6, 10069 (2015).
21. Javieer, B. M. et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* 167, 1369–1384 e19 (2016).
22. Freire-Pritchett, P. et al. Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic stem cells. *eLife* 6, e21926 (2017).
23. Siersbaek, R. et al. Dynamic rewiring of promoter-anchored chromatin loops during adipocyte differentiation. *Mol. Cell* 66, 420–435 e5 (2017).
24. Rubin, A. J. et al. Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat. Genet.* 49, 1522–1528 (2017).
25. Orlando, G. et al. Promoter capture Hi-C-based identification of recurrent noncoding mutations in colorectal cancer. *Nat. Genet.* 50, 1375–1380 (2018).
26. Leung, D. et al. Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature* 518, 350–354 (2015).
27. Consortium, G. T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multisite gene regulation in humans. *Science* 348, 648–660 (2015).
28. Schmitt, A. et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* 17, 2042–2059 (2016).
29. Thurman, R. E. et al. The accessible chromatin landscape of the human genome. *Nature* 489, 75–82 (2012).
30. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell-identity genes. *Cell* 153, 307–319 (2013).
31. Zhang, Y. et al. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature* 504, 306–310 (2013).
32. Rajagopal, N. et al. High-throughput mapping of regulatory DNA. *Nat. Biotechnol.* 34, 167–174 (2016).
33. Diao, Y. et al. A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat. Methods* 14, 629–635 (2017).
34. Engreitz, J. M. et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* 539, 452–455 (2016).
35. Dao, L. T. M. et al. Genome-wide characterization of mammalian promoters with distal enhancer functions. Nat. Genet. 49, 1073–1081 (2017).
36. Roadmap Epigenomics, C. et al. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330 (2015).
37. Richard, M., Drouin, R. & Beaulieu, A. D. ABC50, a novel human ATP-binding cassette protein found in tumor necrosis factor-alpha-stimulated synoviocytes. Genomics 53, 137–145 (1998).

Acknowledgments
We thank members of the Ren laboratory for critical suggestions in the course of this work. We thank N. Nariai (UCSD) for sharing LD information. This work was funded in part by the Ludwig Institute for Cancer Research (to B.R.), NIH (grant no.1R01ES024984 to B.R.), the Ministry of Science, ICT and Future Planning through the National Research Foundation in Republic of Korea (grant no. 2017R1C1B2008838 to I.J.), Korean Ministry of Health and Welfare (grant no. HI17C0328 to I.J.) and SUHF Fellowship (to I.J.).

Author contributions
I.J., A.S., Y.D. and B.R. conceived the study. I.J., A.S. and Y.D. performed experiments with assistance from T.L., C.T. and S.C., I.J., A.J.L. and D.Y. performed data analysis with assistance from J.E., M.C., Z.C. and C.L.B., D.K. supervised data analysis by D.Y., C.K., E.M. and C.L.B. contributed to provide human brain tissue samples. B.L. and S.K. contributed to sequencing and initial data processing. I.J. prepared the manuscript with assistance from A.S., Y.D., A.J.L., J.E. and B.R. All authors read and commented on the manuscript.

Competing interests
B.R. is a co-founder of Arima Genomics, Inc. A.S. is an employee of Arima Genomics.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0494-8.
Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to I.J. or B.R.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
**Methods**

Human tissue samples. Esophagus, lung, liver, pancreas, small bowel, sigmoid colon, thymus, bladder, adrenal gland, aorta, gastric, left heart ventricle, right heart ventricle, right heart atrium, ovary, testis, spleen and fat tissues were obtained from deceased donors at the time of organ procurement at Barnes-Jewish Hospital as described in our previous study[1]. The same tissue types from different donors were combined together during downstream data analysis. Human dorsolateral prefrontal cortex (dlPFC rep1) and hippocampus (HC rep1) tissues were obtained from the National Institute of Child Health and Human Development Brain Bank for Developmental Disorders. These two sets of samples were from a healthy 31-year-old male donor. Ethics approval was obtained from the University Health Network and The Hospital for Sick Children for the use of these tissues. Another set of human dorsolateral prefrontal cortex (dlPFC rep2) and hippocampus (HC rep2) tissues were obtained from the Shirley-Marcos Alzheimer’s Disease Research Center. These two samples were from a healthy 80-year-old female donor. Institutional Review Board approval was obtained from Korea Advanced Institute of Science and Technology (KAIST) for the use of these tissues.

Hi-C library on human tissue samples and early embryonic cell types. Human tissue samples were flash frozen and pulverized before formaldehyde cross-linking. Fibroblasts (IMR90) and lymphoblastoid cell lines (GM12878 and GM19240) were cultured, and 5 million cells were formaldehyded cross-linked for each Hi-C library. Hi-C was then conducted on the samples as previously described with HindIII for Hi-C library preparation[2]. Previously constructed Hi-C libraries[1] were used for human ES cells (H1) and early embryonic cell types including mesendoderm, mesenchymal stem cell, neural progenitor cells and trophoblast-like cells. Generation of capture RNA probes. To perform pcHi-C, we computationally designed RNA probes that capture promoter regions of previously annotated human protein-coding genes. Capture regions were selected for 19,462 well-annotated protein-coding gene promoters across 22 autosomes and X chromosome according to GENCODE v19 annotation with confidence levels 1 and 2. The annotation confidence levels 1 and 2 comprise genes that are accurately annotated with sufficient validation and manual annotation by combining the manual gene annotation from the Human and Vertebrate Analysis and Annotation group, automatic gene annotation from Ensembl and validating by cap-analysis of gene expression (CAGE). Due to the variability of capture efficiency, 19,328 promoter regions (%) were captured in this study. Among them, 18,943 promoter regions were involved in pcHi-C interactions in one or more cell/tissue types analyzed in this study. For each transcription start site, the two nearest left-handed and right-handed HindIII restriction sites were selected. Six capture oligonucleotide sequences were designed to be of 120 nucleotide (nt) length and to have 30 nt tiling overlap. Oligonucleotides were designed at 300 basepairs (bp) upstream and downstream of each restriction site. As two restriction sites were chosen for each transcription start site, a total of 12 capture oligonucleotides were designed to target each promoter region. Capture sequences that overlap with directly adjacent HindIII restriction sites were removed. GC contents of 94% capture sequences ranged from 25% to 65%. Some promoters shared the same HindIII fragment with at least one other promoter. These two or more HindIII fragments (86%) were uniquely assigned to one promoter. The effect of the DNA fragments harboring multiple promoters on the quality of our analytical findings is modest because only 15% of pcHi-C interactions emanated from the promoter-sharing DNA fragments, and eliminating these fragments results in no significant changes in our conclusion for both eQTL enrichment test and gene set enrichment analysis. Further, strong correlation of GWAS trait associations remains even after excluding unresolvable promoters. In total, our capture oligonucleotide design generated 280,445 unique probe sequences including randomly selected capture regions (that is gene deserts). Single-stranded DNA oligonucleotides were then synthesized by CustomArray Inc. Single-stranded DNA oligonucleotides contained universal forward and reverse primer sequences (total length 31 nt), whereby the forward priming sequence contained a truncated SP6 recognition sequence that was completed by the universal capturability of two DNA fragments. During normalization, we defined universal capturability as the summation of all adjacent upstream and downstream restriction sites, respectively. We merged adjacent DNA fragments if the total length of the DNA fragments was less than 3 kb. As a result, 510,045 DNA fragments were defined with a median length of 4.8 kb. After that, we calculated raw interaction frequencies at DNA fragment resolution and performed normalization to remove experimental biases caused by intrinsic DNA sequence biases (GC contents, mappability and effective fragment lengths), RNA probe synthesis efficiency bias and RNA probe hybridization efficiency bias. Highly variable RNA probe synthesis efficiency would greatly complicate the control of experimental bias. However, if the efficiency bias was reproducible, the bias can be computationally removed. To prove such bias reproducibility, we performed four RNA-seq samples (four pcHi-C libraries) that were synthesized independently. The RNA-seq results can quantitatively measure the amount of synthesized RNA probes, which is an indicator of the probe synthesis efficiency. We observed highly reproducible RNA-seq results (Pearson correlation coefficient = 0.98), indicating reproducible probe synthesis efficiency. To address the high complexity of different types of experimental biases, we defined a new term named ‘capturability’, which refers to the probability of the region being captured. We assumed that capturability represents all combined experimental biases and can be estimated by the total number of capture reads spanning a given DNA fragment divided by the total number of captured reads in cis. We found that ‘capturability’ in each DNA fragment is highly reproducible across samples with a Pearson correlation coefficient of 0.99 between samples on average. Therefore, we defined universal capturability as the summation of all capturability defined in each sample and normalized raw interaction frequencies by considering apturability of two DNA fragments. During normalization, we processed promoter-promoter interactions and promoter-other interactions independently because promoter regions tend to show very high capturability because our capture probes were designed to target promoter regions. Also, we only considered promoter-centered long-range interactions over 15 kb and within 2 Mb from TSS of each gene. We denoted Yi to represent the raw interaction frequency between DNA fragment i and j, and Cj to represent capturability defined in DNA fragment j. We assumed Yi to follow a negative binomial distribution with mean μi and variance μi + αμi, where α is a parameter to measure the magnitude of overdispersion. We then fitted a negative binomial regression model as follows: log(Yi) = β0 + β1 Bi(Cj) + β2 Bi(Cj) + Bi(Cj), where Bi is an expected interaction frequency between DNA fragment i and j with coverage C and Cj, and the residual Hj = Yi/exp(β0 + β1 Bi(Cj) + β2 Bi(Cj)) defined as a normalized interaction frequency between DNA fragments i and j. BS represents a basis vector obtained from B-spline regression, which applied to a vector of values of input variable C during negative binomial regression model fitting for robustness and memory-efficient calculation.

Identification of P-P and P-o pcHi-C long-range chromatin interactions. To identify significant pcHi-C chromatin interactions, we removed distance-dependent background signals from normalized interaction frequencies. Here, we assumed that the normalized interaction frequency R follows a negative binomial distribution with mean μ and variance μ + αμ2. Similar to the interaction frequency normalization step above, we calculated the expected interaction frequency at a given distance by fitting it to αμ2. Here, α > 0 is a parameter to measure the magnitude of overdispersion. We then fitted a negative binomial regression model with this vector obtained from B-spline regression of the distance between two DNA fragments. We denoted Dij to represent the expected interaction frequency at a given distance d calculated from a negative binomial regression model. Distance-dependent background signals were removed by taking signal to background dependent background signals from normalized interaction frequencies. Here, we defined universal capturability as the summation of all distance normalized interaction frequencies with three-parameter Weibull distribution. Considering this, significant long-range chromatin interactions are
defined when observed interaction frequencies show lower than 0.01 P-value thresholds by fitting distance-background-removed interaction frequencies with three-parameter Weibull distribution. To eliminate false pHi-C interactions caused by the experimental noise, we applied the criterion of minimum raw interaction frequencies (having more than five raw interaction frequencies), which is chosen by investigating reproducibility between two independently prepared replicates with lymphoblastoid and mesenchymal stem cells. Note that as the interaction frequencies in pHi-C are mostly zeros or close to zero, the distribution of P-values do not follow the uniform distribution, violating the basic assumption of FDR calculation, which assumes that the null distribution follows uniform (0,1) distribution. Thus, we simulated normalized interaction frequencies that follow three-parameter Weibull distribution in a sample-specific manner and computed the estimated FDR through multiple permutations. The estimated FDR through multiple permutation (n=1,000) for P-O and P-P pHi-C interactions is 9% and 5% on average, respectively.

In situ Hi-C experiments and validation of pHi-C long-range chromatin interactions. The visual inspection of normalized interaction frequencies between IMR90 pHi-C and high-resolution IMR90 Hi-C showed high consistency based on manual inspection despite pHi-C having only 10% sequencing depth compared to high-resolution Hi-C (Supplementary Fig. 4a). Next, we compared the identified pHi-C interactions with ‘loops’ defined from IMR90, GM12878, dorsolateral prefrontal cortex and hippocampus tissues, using in situ Hi-C experiments (Supplementary Fig. 4b-e). Although there is a huge discrepancy between the measured Hi-C and pHi-C Hi-C loops and pHi-C interactions, we consider ‘loops’ as a subset of high-confidence long-range chromatin interactions that involve ‘loop’ domains but cannot cover all promoter-mediated long-range chromatin interactions. Loops of IMR90 and GM12878 in situ Hi-C results were obtained from a previous publication. Loops of dorsolateral prefrontal cortex and hippocampus tissues were distributed with HiCUPS-X (ref. 16). The loops were called from Knight-Ruiz normalized 5kb, 10kb and 25kb resolution data, as these parameters were suggested for a medium-resolution Hi-C map by the authors of HiCUPS. As a result, 7,722 and 8,040 loops were identified from dorsolateral prefrontal cortex and hippocampus, respectively. We compared the identified pHi-C long-range chromatin interactions to loops of in situ Hi-C data and measured the reproducibility in terms of ROC curve, a plot of the true-positive rate against the false-positive rate at various threshold settings. Here, we set loops as true interactions. We ranked all tested pHi-C DNA fragment pairs in terms of P values and then calculated the fraction of true positives and false positives to draw the ROC curve. We only considered ‘loops’ emanating from promoters containing DNA fragments defined in pHi-C result. The rank of all pcHi-C interactions identified in one replicate is always higher than all other tested DNA fragment pairs according to pHi-C interaction P values.

Reproducibility of pHi-C chromatin interactions between biological replicates. The reproducibility of pHi-C chromatin interactions between biological replicates (two different donors for tissues and two independently cultured cells for cell lines) was measured from ROC curve in Supplementary Fig. 4f). Here, we compare pHi-C interactions identified in one replicate as true interactions. For the other replicate, we ranked all tested DNA fragment pairs in terms of P values and then calculated the fraction of true positive and false positive to draw ROC curve. The area under the ROC curve is defined as an ROC score, and an ROC score of 1 indicates that all tested DNA fragment pairs matched by loops are always higher than all other tested DNA fragment pairs according to pHi-C interaction P values.

Identification of extensively interacting DNA fragments. To identify DNA fragments that showed extensive long-range interactions with multiple promoters, we systematically defined these promiscuously interacting DNA fragments from P-P pHi-C interaction maps and P-O pHi-C interaction maps, respectively. For each cell or tissue type, we selected frequently interacting DNA fragments with multiple promoters in terms of 0.01 Poisson P-value cutoff.
enrichment analysis of superenhancers was conducted by generating random
genomic locations of the same size as superenhancers but at different genomic
coordinates. We also conducted the enrichment test with typical enhancers. We
revealed that P–O EIFs highly co-exist with superenhancer regions, rather than
typical enhancers and genomic background for most of the samples, except for
two samples, lymphoblastoid cell lines and gastric tissue. Note that half of the
lymphoblastoid P–O EIFs are co-occupied with typical enhancers that are classified
as superenhancers in other cell/tissue types.

Comparison between eQTL associations and P–O interactions. To test the
enrichment for P–O pcHi-C chromatin interactions in significant eQTL associations,
we compared P–O pcHi-C interactions to significant eQTL associations in the
matching tissue types. The eQTL associations were downloaded directly from GTEx
Portal (downloaded on 10 November 2017) for all matching tissue types (n = 14,
adrenal gland, aorta, dorsolateral prefrontal cortex, brain hippocampus, sigmoid
colon, esophagus, left heart ventricle, liver, lung, ovary, pancreas, small intestine
terminal ileum for small bowel, spleen and stomach for gastric). First, the significant
eQTLs defined by GTEx (q value ≤ 0.05) were filtered so that only the eQTL variants
within the fragments that involve P–O pcHi-C interactions remain for comparison.
Then, we removed pcHi-C interactions beyond 1 Mb in distance to match the range
of eQTL association and discarded eQTL associations with distance below 15 kb to
match the valid interaction cutoff. The filtered, significant eQTL associations were
compared with pcHi-C and randomized interactions in the same condition. Here,
we only considered P–O pcHi-C interactions with DNA fragments that do not
harbor multiple promoters. For the random expectation, we generated a simulated
pcHi-C interaction pool by creating all possible combinations of DNA fragments
with no TSS and the protein-coding genes that exist within the distance range. The
pcHi-C interactions that exist in any of the tissue/cell types were removed from
the control interaction pool for the enrichment analysis. To avoid variation caused
by the difference in distance between pcHi-C interactions and eQTL associations,
we created distance-matched control, in which the number of pcHi-C interactions
was stored at the interval of 40 kb, and the same number of interactions was drawn
randomly from the control interaction pool. The number of randomized interactions
drawn from each chromosome was matched to the pcHi-C interactions. The
standard deviation was obtained by permuting the random expectation with 1,000
iterations and was used to calculate the statistical confidence.

To illustrate the filtering process of the eQTL data, for example, the 549,763
significant eQTLs in adrenal gland were reduced to 237,181 after collecting eQTLs
located in the DNA fragments without TSS and discarding eQTL associations
with the distance below 15 kb and with a pseudogene target. This filtered set of
significant eQTL associations was used for an enrichment test for both pcHi-C and
randomized interactions. The number of total tested significant eQTL associations,
19,996 in the case of adrenal gland, in Supplementary Table 11, indicates the
number of significant eQTLs located in the DNA fragments that are associated
with the pcHi-C interactions in the corresponding cell/tissue type.

Statistics. We used the KS test to compare distributions between two groups
as a nonparametric test without assumptions of normality. We also used the
permutation test to calculate empirical P values, which does not make any
assumptions on the underlying distribution of the data.

Reporting Summary. Further information on research design is available in the
Nature Research Reporting Summary linked to this article.

Data availability
All raw and processed data have been deposited in the GEO database under
accession number GSE86189. Visualization of processed pcHi-C data is available at
http://www.3div.kr/capture_hic.

Code availability
Code for pcHi-C interaction detection can be made available on request. For other
data analysis, we used publicly available software.

References
38. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range
interactions reveals folding principles of the human genome. Science 326,
289–293 (2009).
39. Yang, D. et al. 3DIV: A 3D-genome interaction viewer and database. Nucleic
Acids Res. 46, D52–D57 (2018).
40. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9,
R137 (2008).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

| Data collection | Data analysis |
|-----------------|---------------|
| fastq-dump: Download sequencing data deposited at GEO/SRA including in situ Hi-C for IMR90 and GM12878 | FastQC: A quality control tool for high throughput sequence data |
| TF ChIP-seq for GM12878 and H1-hESC: Downloaded from ENCODE | Samtools 1.3: Process aligned sequencing results |
| TAD annotation: Downloaded from 3DIV database | BWA 0.7.13-r1126: Alignment of sequencing results |
| ChromHMM chromatin state: Downloaded from Roadmap Epigenomics Project | HiCCUPs (distributed with Juicer v1.7.6) : Loop domain annotation from in situ Hi-C data |
| Super-enhancer list: Downloaded from 3DIV database | MACS2-2.1.1: ChIP-seq peak calling |
| eQTL association: Downloaded from GTEx Portal on Nov. 10th 2017 | Bedtools v2.13.3: Analysis of ChIP-seq data |
| GWAS-SNPs: Downloaded from GWAS catalogue database version 1.0.1 | In house script: manual pairing of BWA-mem alignment results |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All raw and processed data are deposited into GEO database under accession number GSE86189. All processed data is available at http://www.3div.kr/capture_hic

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☑ Life sciences  ☐ Behavioural & social sciences  ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | No sample size calculation was performed. To generate promoter-centered chromatin interaction maps across diverse cell/tissue types, we decided to use the the samples that were selected as reference epigenome maps of these samples have previously been produced as part of the Epigenome Roadmap Project. In this regard, we collected 27 human cell/tissue types with a total of 40 biological samples including biological replicates. |
|---|---|
| Data exclusions | No data exclusion during data analysis. |
| Replication | We observed high reproducibility of pcHi-C long-range chromatin interactions between donors, evidenced by substantial area under the receiver operating curve (ROC), ranging between 0.73 and 0.99. We also compared the pcHi-C data with the in situ Hi-C data obtained from four distinct biosamples, including two cell lines (IMR90 lung fibroblast cell line and GM12878 lymphoblastoid cell line) and two primary tissues (dorsolateral prefrontal cortex and hippocampus) and found that the pcHi-C experiments accurately recapitulated chromatin loops identified from in situ Hi-C assays in all samples, evidenced by ROC score, ranging between 0.84 and 0.91. |
| Randomization | Randomization of samples is not applicable to our study design as we collect the same samples used in Roadmap Epigenomics. |
| Blinding | As we collect all samples from Roadmap Epigenomics, blinded group allocation is not applicable to this study. |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | n/a | Involved in the study |
|---|---|---|
| ☑ | Antibodies |
| ☑ | Eukaryotic cell lines |
| ☑ | Palaeontology |
| ☑ | Animals and other organisms |
| ☑ | Human research participants |
| ☑ | Clinical data |

| Methods | n/a | Involved in the study |
|---|---|---|
| ☑ | ChIP-seq |
| ☑ | Flow cytometry |
| ☑ | MRI-based neuroimaging |

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s) | GM12878 and GM19240 from Coriell Institute (https://www.coriell.org). IMR90 from ATCC. |
|---|---|
| Authentication | The cells were not authenticated |
| Mycoplasma contamination | The cells were not tested for mycoplasma contamination |
Human research participants
Policy information about studies involving human research participants

Population characteristics
Esophagus, lung, liver, pancreas, small bowel, sigmoid colon, thymus, bladder, adrenal gland, aorta, gastric, heart, ovary, psoas, spleen, and fat tissues were obtained from deceased donors at the time of organ procurement at Barnes-Jewish Hospital (St. Louis, USA) as described in our previous study. The same tissue types from different donors were combined together during downstream data analysis. Human dorsolateral prefrontal cortex (DLPFC rep1) and hippocampus (HC rep1) tissues were obtained from the National Institute of Child Health and Human Development (NICHD) Brain Bank for Developmental Disorders. These two samples were from a healthy 31-year-old male donor. Ethics approval was obtained from the University Health Network and The Hospital for Sick Children for the use of these tissues. Another set of human dorsolateral prefrontal cortex (DLPFC rep2) and hippocampus (HC rep2) tissues were obtained from the Shiley-Marcos Alzheimer’s Disease Research Center (ADRC). These two samples were from a healthy 80-year-old female donor. Institutional Review Board (IRB) approval was obtained from KAIST for the use of these tissues.

Recruitment
Tissues were obtained from the National Institute of Child Health and Human Development (NICHD) Brain Bank for Developmental Disorders and the Shiley-Marcos Alzheimer’s Disease Research Center.

Ethics oversight
Ethics approval was obtained from the University Health Network and The Hospital for Sick Children for the use of DLPFC rep1 and HC rep1 tissues. Institutional Review Board (IRB) approval was obtained from KAIST for the use of DLPFC rep2 and HC rep2 tissues.

Note that full information on the approval of the study protocol must also be provided in the manuscript.